

ความสัมพันธ์เชิงปริมาณระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพและโมเลกุลาร์ดีออกกิ่ง
ของสารต้านมาลาเรียอนุพันธ์ไพโรไซคลิก 1,2,4-ไพโรออกเซน



นางสาวกุลวดี รัตนศักดิ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเคมี ภาควิชาเคมี

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2546

ISBN 974-17-5447-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP
AND MOLECULAR DOCKING OF ANTIMALARIAL TRICYCLIC
1,2,4-TRIOXANE DERIVATIVES



Miss Koonwadee Ratanasak

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย
A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Chemistry

Department of Chemistry

Faculty of Science

Chulalongkorn University

Academic Year 2003

ISBN 974-17-5447-7

Thesis Title Quantitative Structure Activity Relationship and Molecular Docking of Antimalarial Tricyclic 1,2,4-Trioxane Derivatives

By Miss Koonwadee Ratanasak

Field of Study Chemistry

Thesis Advisor Associate Professor Sirirat Kokpol, Ph.D.

Thesis Co-Advisor Associate Professor Vudhichai Parasuk, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree.

..... Dean of the Faculty of Science
(Professor Piamsak Menasveta, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Associate Professor Siri Varothai, Ph.D.)

..... Thesis Advisor
(Associate Professor Sirirat Kokpol, Ph.D.)

..... Thesis Co-Advisor
(Associate Professor Vudhichai Parasuk, Ph.D.)

..... Member
(Assistant Professor Pornthep Sompornpisut, Ph.D.)

..... Member
(Pongchai Harnyuttanakorn, Ph.D.)

กุลวดี รัตนศักดิ์: ความสัมพันธ์เชิงปริมาณระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพและโมเลกุลควิลาร์ดีออกกิ่งของสารต้านมาลาเรียอนุพันธ์ไตรไซคลิก 1,2,4-ไตรออกเซน (QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP AND MOLECULAR DOCKING OF ANTIMALARIAL TRICYCLIC 1,2,4-TRIOXANE DERIVATIVES) อาจารย์ที่ปรึกษา: รศ. ดร. ศิริรัตน์ ก๊กผล, อาจารย์ที่ปรึกษาร่วม: รศ. ดร. วุฒิชัย พาราสุข, 127 หน้า. ISBN 974-17-5447-7

ศึกษาสารต้านมาลาเรียไตรไซคลิก 1,2,4-ไตรออกเซนจำนวน 32 สาร โดยวิธี QSAR แบบดั้งเดิม 3D-QSAR (CoMFA) และโมเลกุลควิลาร์ดีออกกิ่ง ทำการคำนวณปรับโครงสร้างของสารทั้งหมดด้วยระเบียบวิธีแอบ อินิซิโอ ที่ระดับ HF/3-21G ใช้ฤทธิ์ทางชีวภาพที่ได้จากการทดสอบกับเชื้อมาลาเรียสายพันธุ์ NF54 แบบจำลองที่ได้จากเทคนิค QSAR แบบดั้งเดิม และ 3D-QSAR (CoMFA) ให้ผลเป็นที่น่าพอใจในเชิงค่าทางสถิติและประสิทธิภาพในการทำนาย ประจุของสารเหล่านี้ได้จากการคำนวณด้วย 6-31G(d) เบซิสเซต และประจุของฮีมีได้จากการคำนวณด้วย 6-311G(d,p) เบซิสเซต ผลที่ได้จากการศึกษาด้วยวิธีโมเลกุลควิลาร์ดีออกกิ่ง บ่งชี้ว่า Fe^{2+} จะเข้าจับกับสารในกลุ่มไตรไซคลิก 1,2,4-ไตรออกเซน ที่อะตอม O_1 หรือ O_2 ซึ่งขึ้นอยู่กับตำแหน่งของหมู่แทนที่ของสาร ผลที่ได้เสนอแนะว่ากลไกการออกฤทธิ์ของสารนี้คล้ายกับกลไกของอาร์ติมิซินิน ผลการศึกษาทั้งหมดสามารถนำไปใช้สำหรับเสนอแนวทางในการออกแบบและทำนายสารใหม่ที่มีฤทธิ์ยับยั้งเชื้อมาลาเรียที่สูงขึ้น

สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา...เคมี.....	ลายมือชื่อนิติ.....
สาขาวิชา...เคมี.....	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา...2546.....	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4472219523: MAJOR CHEMISTRY

KEYWORD: QSAR / CoMFA / Molecular Docking / Tricyclic 1,2,4-Trioxanes

KOONWADEE RATANASAK: QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP AND MOLECULAR DOCKING OF ANTIMALARIAL TRICYCLIC 1,2,4-TRIOXANE DERIVATIVES. THESIS ADVISOR: ASSOC. PROF. SIRIRAT KOKPOL, Ph.D., THESIS CO-ADVISOR: ASSOC. PROF. VUDHICHAH PARASUK, Ph.D., 127 pp. ISBN 974-17-5447-7

A set of 32 antimalarial tricyclic 1,2,4-trioxane compounds was subjected to classical QSAR, 3D-QSAR(CoMFA) and molecular docking studied. All compounds were geometry optimized with *ab-initio* method at the HF/3-21G level. The activities measured against *NF54* strain of malarial parasites were used. The models obtained from classical QSAR and 3D-QSAR(CoMFA) techniques are satisfied based on both statistical significance and predictive ability. For docking studies, The atomic charges of these compounds were obtained from single point calculations using 6-31G(d) basis set and the atomic charges of heme were calculated at 6-311G(d,p) level. The results derived from the structure-based design using molecular docking study reveal that Fe²⁺ approaches tricyclic 1,2,4-trioxane compounds at the O₁ or O₂ atoms depending on the substituted position of the compound. These results suggested similar mechanism of action to that of artemisinin. All results can support each other and can be used as a guideline to design and predict the new compounds with increasing malarial inhibitory activities.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department.....Chemistry.....	Student's signature.....
Field of study.....Chemistry.....	Advisor's signature.....
Academic year....2003.....	Co-advisor's signature.....

ACKNOWLEDGEMENT

This thesis would not have been accomplished without contributions from many persons who helped me to overcome extensive hindrances and difficulties during my past 3 years of study. Truly, I cannot acknowledge the names of all individuals who either directly or indirectly give me a help, and would like to apologize to them whom I fail to mention.

I wish to express my deepest appreciation and grateful thank to my thesis advisor, Associate Professor Dr. Sirirat Kokpol for her concern, encouragement and valuable advice throughout my period of study.

I would like to express my grateful thank to my co-advisor, Associate Professor Dr. Vudhichai Parasuk for generously guiding, kindly helping and encouraging throughout the course of this thesis.

I also would like to thank the thesis committee, Associate Professor Dr. Siri Varothai, Assistant Professor Dr. Pornthep Sompornpisut, Dr. Pongchai Harnyuttanakorn for their kind attention, valuable construction, criticism and helpful suggestions.

I also would like to thank the National Electronics and Computer Technology Center (NECTEC), Thailand, the Austrian-Thai Center for Computer Assisted Chemical Education and Research (ATC) and Computational Chemistry Chemical Unit Cell, Chulalongkorn University, Thailand for computer resources and other facilities. Gratefully thanks to Professor Dr. Karl Peter Wolschann for computer program TSAR software.

My sincere gratitude and deep appreciation are expressed to Associate Professor Dr. Supot Hannongbua and Dr. Tawun Remsungnen.

My work would not have been succeeded without the kindness of Dr. Somsak Tonmunphean.

I wish to express my infinite gratitude to my family and my friends for their everlasting love, understanding and encouragement.

Finally, I gratefully acknowledge the financial supports from the Graduate School, Chulalongkorn University and the Ratchdaphisek Somphot Endowment grant (CE scholarship).

CONTENTS

	Pages
ABSTRACT IN THAI.....	iv
ABSTRACT IN ENGLISH.....	v
ACKNOWLEDGEMENT.....	vi
CONTENTS.....	vii
LIST OF FIGURES.....	x
LIST OF TABLES.....	xii
LIST OF ABBREVIATIONS.....	xiv
CHAPTER 1 INTRODUCTION.....	1
1.1 Life Cycle of Malaria parasite.....	2
1.2 Situation of malaria in the World and in Thailand.....	3
1.2.1 Malaria in the World.....	3
1.2.2 Malaria in Thailand.....	4
1.3 Antimalarial drugs.....	5
1.3.1 Classification of Antimalarial Drugs.....	5
1.3.2 Antimalarial Therapy for Tomorrow.....	6
1.3.2.1 Drug Combination.....	6
1.3.2.2 Vaccines	7
1.3.2.3 Old Targets, New Compounds	7
1.3.3 1,2,4-Trioxane Antimalarials.....	8
1.4 Aim of Our Study.....	9
CHAPTER 2 THEORETICAL BACKGROUND IN QUANTUM CHEMISTRY	11
2.1 The Schrödinger Equation	11
2.2 Molecular Orbital Theory	12
2.3 The Hartree-Fock Approximation	14
2.4 Basis Set.....	20
2.4.1 Minimal Basis Set.....	21
2.4.2 Extended Basis Set.....	22

2.4.3 Polarized Basis Set.....	22
2.4.4 Basis Set Incorporating Diffuse Function.....	23
2.5 Population Analysis.....	23
CHAPTER 3 STRUCTURE AND BIOLOGICAL DATA	25
3.1 Chemical Structure of Tricyclic 1,2,4-trioxane	25
3.2 Structures and Antimalarial Activities of Tricyclic 1,2,4-trioxane Derivatives	25
3.3 Geometry Optimization and Atomic Charge Calculation.....	29
CHAPTER 4 CLASSICAL-QSAR.....	33
4.1 Introduction.....	33
4.1.1 Classical QSAR.....	33
4.1.1.1 Hydrophobicity Properties	34
4.1.1.2 Steric Properties.....	35
4.1.1.3 Electronic Properties.....	36
4.1.2 Statistical Analysis for QSAR Study	38
4.2 Calculations of Properties	43
4.3 Results and Discussions.....	44
4.4 QSAR Summary.....	51
CHAPTER 5 THREE DIMENSIONAL QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP.....	52
5.1 Theoretical Background of CoMFA	53
5.1.1 Alignment Rule.....	53
5.1.2 Interaction Energy Calculation.....	53
5.1.3 Statistical Analysis using Partial Least-Squares Method.....	55
5.1.4 Interpretation of CoMFA Results.....	59
5.2 Computational Methods.....	61
5.2.1 Conformation and Alignment Rule.....	61
5.2.2 CoMFA Calculations.....	64

	Pages
5.3 CoMFA Results and Discussions.....	65
5.3.1 Effect of Conformation and Alignment Rule.....	65
5.3.2 Effect of Type of Probe Atom.....	66
5.3.3 Effect of Steric and Electrostatic Cut-offs	67
5.3.4 CoMFA results.....	68
5.4 CoMFA Summary.....	78
CHAPTER 6 MOLECULAR DOCKING.....	80
6.1 Introduction.....	80
6.2 Docking Theory.....	82
6.2.1 Automated Docking.....	85
6.3 Computational Methods	87
6.4 Determination of Suitable Docking Parameters	89
6.4.1 Grid Dimension	89
6.4.2 Grid Spacing.....	91
6.4.3 Starting Temperature (T_s), Final Temperature (T_f) and Temperature Reduction Rate	92
6.5 Docking of All 32 compounds	96
6.6 Docking Summary	101
CHAPTER 7 CONCLUSIONS.....	103
REFERENCES.....	106
APPENDICES.....	114
CURRICULUM VITAE.....	126

LIST OF FIGURES

Figures	Pages
1.1 <i>Plasmodium falciparum</i> life cycle within human	2
1.2 Stereochemistry and atomic numbering scheme of artemisinin.....	8
2.1 Sequence of program step required for the solution of the Roothaan-Hall equations for closed-shell, Self Consistent Field procedure.....	19
2.2 The cusp of Slater function.....	21
3.1 Stereochemistry and atomic numbering scheme of tricyclic 1,2,4-trioxanes.....	25
4.1 Flow chart of the QSAR methodology in this study.....	42
4.2 Comparison between actual and predicted activities for 25 compounds in the training set by the best QSAR model.....	50
4.3 Comparison between actual and predicted activities for 4 compounds in the test set by the best QSAR model.....	51
5.1 Cross-validate procedure.....	56
5.2 Flow chart of CoMFA methodology in this study.....	60
5.3 Superimposition of all tricyclic 1,2,4-trioxane compounds using the alignment 13 and compound 31 as template (hydrogen atoms were omitted for clarity).....	61
5.4 Definition of 15 alignment rules used in CoMFA studies.....	64
5.6 Comparison between actual and predicted activities for 25 compounds in the training set.....	71
5.7 Comparison between actual and predicted activities for 4 compounds in the testing set.....	72
5.8 CoMFA S.D.*coeff. Steric contour maps, (A) compounds 3 and (B) compounds 14 which have substituents at C ₃ position (R ₁) are represented.....	73
5.9 CoMFA S.D.*coeff. Steric contour maps, compound 24 which has substituent at C ₄ position (R ₂) is represented.....	73
5.10 CoMFA S.D.*coeff. Steric contour maps, compound 31 which has substituent at C _{8a} position (R ₃) is represented.....	74

Figures	Pages
5.11 CoMFA S.D.*coeff. Steric contour maps, compound 21 which has substituent at C ₁₂ position (R ₄) is represented.....	75
5.12 CoMFA S.D.*coeff. Electrostatic contour maps, compound 14 which has substituent at C ₃ position (R ₁) is represented.....	76
5.13 CoMFA S.D.*coeff. Electrostatic contour maps, compound 24 which has substituent at C ₄ position (R ₂) is represented.....	77
5.14 CoMFA S.D.*coeff. Electrostatic contour maps, compound 31 which has substituent at C _{8a} position (R ₃) is represented.....	77
5.15 CoMFA S.D.*coeff. Electrostatic contour maps, compound 21 which has substituent at C ₁₂ position (R ₄) is represented.....	78
6.1 Proposed mechanisms for the Fe(II)-induced activation of simplified 1,2,4-trioxanes.....	81
6.2 Graph showed relationship between internuclear distance (r) and well-depth (ε).....	82
6.3 Grid base energy evaluation.....	88
6.4 Superimposed docking configurations between heme and 32 derivative of 1,2,4-trioxane compounds (without hydrogen atoms).....	98
6.5 Docking configurations between heme and compound 15 (without hydrogen atoms).....	99
6.6 Docking configurations between heme and compound 19 (without hydrogen atoms).....	99

LIST OF TABLES

Tables		Pages
1.1	Antimalarial drugs exploited since 1930.....	4
3.1	Structures and biological data of compounds number 1-15.....	27
3.2	Structures and biological data of compounds number 16-21.....	28
3.3	Structures and biological data of compounds number 22-24.....	28
3.4	Structures and biological data of compounds number 25-32.....	29
3.5	Comparison of important structural parameters of 1,2,4-trioxane as obtained from the X-ray and HF/3-21G optimization method.....	30
4.1	Predicted activities and residuals of 28 compounds in the training set by the QSAR model (eg. 4.25).....	45
4.2	The correlation coefficients between each pair of variables.....	46
4.3	Relationship between the angle of O ₁₃ -C ₃ -C ₄ , A(13-3-4) and activity.....	46
4.4	Relationship between the torsion angle T(1-2-3-13) and activity.....	47
4.5	Relationship between the C _{8a} atomic charge and activity.....	48
4.6	Relationship between the C ₁₂ atomic charge and activity.....	48
4.7	Predicted activities and residuals of 25 compounds in the training set by the best QSAR model (eg. 4.26).....	49
4.8	Predicted activities and residuals of 4 compounds in the test set by the best QSAR model.....	50
5.1	Atoms selected for the definition of alignment rules.....	62
5.2	CoMFA results of different conformation and alignment rules.....	66
5.3	Effect of Type of Probe Atom.....	67
5.4	CoMFA results with different steric and electrostatic cut-offs.....	68
5.5	CoMFA results of all 28 compounds.....	68
5.6	Predicted activities and residuals of 28 compounds in the training set.....	69
5.7	CoMFA results of 25 compounds*.....	69
5.8	Predicted activities and residuals of 25 compounds in the training set.....	70
5.9	Predicted activities and residuals of 4 compounds in the testing set.....	71
6.1	Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with three different grid dimensions.....	90

Tables	Pages
6.2 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with two different grid spacing.....	91
6.3 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different starting temperature (T_s).....	93
6.4 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different final temperature (T_f).....	94
6.5 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different Temperature Reduction Rate.....	95
6.6 The docking results of heme and 32 derivatives of 1,2,4-trioxane compounds.....	100



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF ABBREVIATIONS

2D	: Two Dimensional
3D	: Three Dimensional
CADD	: Computer-Aided Drug Discovery
CoMFA	: Comparative Molecular Field Analysis
DF	: Degree of Freedom
HF	: Hartree-Fock Method
HOMO	: Highest Occupied Molecular Orbital
LUMO	: Lowest Unoccupied Molecular Orbital
k	: Number of Variables
MLR	: Multiple Linear Regression
MPA	: Mulliken Population Analysis
n	: Number of Compounds
onc	: Optimal Number of Component
PLS	: Partial Least Squares
PRESS	: Predictive Residual Sum of Squares
q^2	: Cross-Validated r^2
QSAR	: Quantitative Structure-Activity Relationship
r	: Correlation Coefficient
RMS	: Root Mean Square
rmsd	: Root-Mean-Square-Deviation
s	: Standard Deviation
SSE	: Sum of Squared Errors
SSR	: Sum of Squared Regression
SST	: Sum of Squared Total
WHO	: World Health Organization

CHAPTER 1

INTRODUCTION

Drug discovery and development has long been the largest research area in life science. Drug research apparently plays an important role in improving quality of human health. Modern drugs are highly effective form of treatment. They are developed from organized research system discover precise treatment for diseases. The discovery of pharmaceutical agent has gone through an evaluation over the years and new technologies has been adding to this increasing complex process.

A modern approach to drug discovery deals with knowledge of biochemical pathway involving the diseases and mechanism of drug actions. It is important to know about the detail structural properties of drug molecules. Many drugs are substrate-like molecule since they hit precisely to specific target enzyme. Thus, knowledge of three dimensional structures contains useful information in designing new drugs. Rational drug designs are nowadays carried out by computer programs. This computational approach is generally known as Computer Aided Drug Design (CADD).

Quantitative Structure Activity Relationship (QSAR), one of CADD techniques, has been used for decades. QSAR constructs a mathematical expression derived from statistical relationship between chemical structure and biological activities of a series of compounds. Then, the model can be used to make an initial guess of biological activity before synthesizing the compounds. Moreover, the model obtained from QSAR method gives information on how to modify chemical property to increase the biological activity. Another technique of CADD is molecular docking which is used to predict possible interactions between drug and receptor. The knowledge obtained from the docking could be used for accurately structural modifications of drug to enhance the interactions.

In the CADD, quantum mechanical methods are increasingly used to calculate molecular and electronic properties due to some advantages over the experimental works. There are two main advantages. Firstly, they are cheaper and more convenient while give very reliable values as compared to those of experiments. It is, therefore, reduce number of synthesized compounds. Furthermore, the power in terms of hardware and software is increasing while the cost of computing is steadily

decreasing. Secondly, they are able to calculate some properties that are very difficult or impossible to measure experimentally, such as electronic properties. Moreover, from these methods, it is possible to derive properties that depend upon the electronic distribution and in particular to investigate chemical reactions in which bonds are being broken and formed.

In this thesis, ligand-based drug design approaches using classical QSAR and 3D-QSAR as well as structure-based drug design procedures using molecular docking were applied to antimalarial 1,2,4-trioxane compounds.

1.1 Life Cycle of Malaria parasite

Malaria is transmitted to human by a bite of an infected female anophele mosquito. The parasite first invades and develops within hepatocytes, and then colonizes in erythrocytes where it achieves asexual multiplications (Figure 1.1). Within erythrocytes, *Plasmodium* digests main part of host hemoglobin (20-80%) and use released amino-acids to build its own proteins. The synchronous intraerythrocytic maturation of parasites leads to cell burst going together with fever access and deep anemia. At each red blood cell burst (each 48 hours for *P. falciparum*), non mature parasites called merozoites are released for further erythrocytic reinvasion, and also gametocytes that will achieve a sexual development of the parasite within mosquito after reabsorption by biting an infected man.

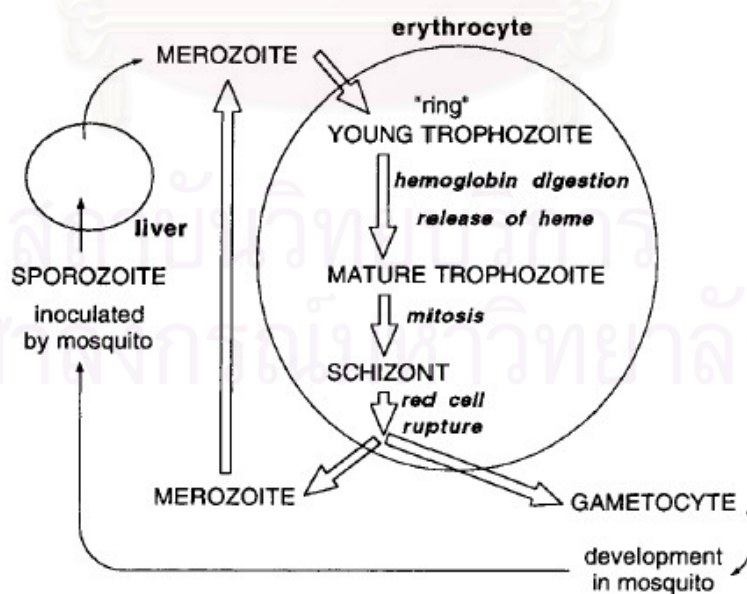


Figure 1.1 *Plasmodium falciparum* life cycle within human.¹

1.2 Situation of malaria in the World and in Thailand

1.2.1 Malaria in the World

Malaria is one of the leading causes of morbidity and mortality in tropical and subtropical regions. The disease causes about 1.22 million deaths per year according to the World Health Organization 2003 report,² which is a low estimate. In fact, obtaining accurate information on the rate of mortality and morbidity is not easy, since symptoms of acute malaria are similar to those of many other acute infectious diseases, and facilities for investigation of suspect cases are not available, leading to an under estimation of malarial cases. Four *Plasmodium* species are responsible for malaria in human. *P. malariae* and *P. ovale* are universally susceptible to the cheap and well tolerated drug chloroquine. *P. vivax* is widely extended and some cases of resistance appeared, however it is not lethal. *P. falciparum* is responsible for all the malignant cases. Severe malaria (cerebral malaria) is an encephalopathy that, even with a suitable treatment, can lead to death in 2-3 days (10-30% of cases). This latter species is also involved in all main cases of drug resistant malaria.³

The control of malaria nowadays largely depends on drug therapies. Several classes of drugs can be considered according to the target parasite stage. Schizontocides are able to inhibit the proliferation of schizonts within red blood cells. Among them, quinine is still a powerful drug. This 4-quinoline-methanol has been used as a lead structure for the design of synthetic antimalarial agents, between chloroquine and mefloquine are well known. Gametocytocides are active against sexual stages of the parasite. Few 8-aminoquinolines exhibit this kind of reactivity and in this class, primaquine was found to be efficient, including against intrahepatic stages.

During the World War II to the 1980s, a wide use of chloroquine for prophylaxis, associated with a fight against mosquitoes by draining wet lands and using insecticides such as DDT, led to a reduction of malaria in endemic areas. In the 1950s, the WHO optimistically targeted malaria for eradication. The United States and Europe got clear of malaria in the 1960s. The first case of resistance to chloroquine was reported in 1959 in South America, followed quickly by a report from South-East Asia. Since then, resistance has spread throughout those areas where *P. falciparum* is endemic. Unfortunately, a rapid development of an anopheline resistant to pesticides that have been widely used (for instance soaked bed nets, resulting in

some areas in a substantial reduction of child mortality) and the resistance of *P. falciparum* to main classes of drugs (chloroquine is no longer useful in most of the endemic areas) have quickly made the eradication out of sight. Furthermore, drug-resistant malaria is rapidly spreading and many regions that are now suitable might become contaminated in the next decades.

Although *falciparum* malaria is of course a public health scourge in endemic areas, northern countries, including large parts of Europe and the United States, might not be out of concern due to possible climate variations and increasing travels in endemic regions.

1.2.2 Malaria in Thailand

In Thailand, malaria is found mostly in the Thai-Cambodian border and the western border of Thailand with Myanmar and approximately 100,000 people are infected and around 800 people die from the disease annually.⁴ Malaria in Thailand was first documented during the reign of King Narai the Great of Ayutthaya (B.E. 2203-2230).⁵ Today malaria still remains an important health problem in Thailand. On the Thai-Myanmar border there is a very high level of drug resistance, with evidences both *in vitro* and *in vivo* for *P. falciparum* parasites that are highly resistant to chloroquine, sulfadoxine-pyrimethamine and mefloquine and to a lesser extent to quinine. Antimalarial drug resistance has spread and intensified over the past 40 years leading to a dramatic decline in the efficacy of the most affordable antimalarial drugs (Table 1.1).

Table 1.1 Antimalarial drugs exploited since 1930.

Old drugs	New drugs
- Cinchona alkaloids, pamaquine	- Artemisinin
- Mepacrine, chloroquine	- Artesunate, artemether, arteether
- Proguanil, amodiaquine	- Pyronaridine
- Pyrimethamine, primaquine	- Mefloquine, halofantrine
- Pyrimethamine-sulfa combinations	- Atovaquone-proguanil
	- Artemether-lumefantrine
	- Tafenoquine

New antimalarial drugs have been investigated in recent years in Thailand. Atovaquone, a hydroxynaphthoquine, was evaluated and the compound alone was

proved to be safe and effective. All patients treated had clinical cure, however, one third of the patients had late recrudescence (RI). When atovaquone was combined with proguanil, the cure rate increased to 100%.^{6,7} This combination has been developed as a fixed combination drug (Malarone®). Artemisinin derivatives such as artesunate, artemether, arteether and dihydroartemisinin were also tested. Arteether, a WHO/TDR supported drug, has been evaluated in a hospital and was recently has been licensed as Artemotil® for the use in severe malaria.⁸ Other combinations (artemisinin derivatives combined with lumefantrine or doxycycline and mefloquine combined with tetracycline or doxycycline) have also been evaluated with improvement of cure rates.^{9,10} Recently, a fixed combination of artemether and lumefantrine (Coartem®) has proven to be a safe and effective drug for the treatment of falciparum malaria.^{11,12} At present, studies with combinations of artemisinin derivatives plus mefloquine are being investigated. In general, artemisinin derivatives combined with mefloquine has been a standard regimen for the treatment of multidrug-resistant falciparum malaria in Thailand. Until proven otherwise, drug combinations are still remaining the recommended agents for treating patients suffering from acute uncomplicated falciparum malaria contracted in multidrug-resistant areas.

1.3 Antimalarial drugs

1.3.1 Classification of Antimalarial Drugs

A classification of antimalarial drugs can be done in many ways depending on a criteria used, such as chemical structure, drug target, and drug action. However, a biological classification, based on parasite stage in which drug mediates its action, is widely used.¹³ According to this classification, 5 categories are defined as follows.

1.3.1.1 Primary Tissue Schizontocides (causal prophylaxis drugs). The drugs belonging to this class, e.g., proguanil and chloroquanide, exert a lethal effect on the pre-erythrocytic stages of parasite (primary tissue forms or primary exo-erythrocytic forms). Thus, they completely prevent an invasion of parasites to red blood cells and also a further transmission of malaria to mosquitoes.

1.3.1.2 Secondary Tissue Schizontocides (radically curative drugs). The drugs, e.g., primaquine, eradicate exoerythrocytic stages or tissue forms of *P. vivax* and *P. ovale* and thus able to achieve radical cure of these infections. Individuals living in

endemic areas are not suitable candidates for radically curative therapy due to the considerable likelihood of reinfection. Normally, the treatment is usually reserved for persons who experience relapsing vivax malaria after leaving malarious regions.

1.3.1.3 Schizontocides (blood schizontocides or schizontocidal drugs). The drugs act rapidly on erythrocytic stages (schizon) of parasites in red blood cells. By interrupting an asexual reproduction of malarial parasite, the clinical attack is terminated. Continuing use of schizontocides for a longer period than a life-span of the infection can completely eliminate malarial parasites from the body. Chloroquine, quinine, mefloquine, halofantrine, artemisinin, and antifolate compounds are belonging to this class.

1.3.1.4 Gametocytocides (gametocytocidal drugs). Agents in this category, e.g., primaquine, chloroquine, and quinine, destroy all sexual forms of malarial parasites in the human blood including those of *P. falciparum*. Thus, they eliminate the reservoir from which mosquitoes are reinfected. They also act on the development stages of malarial parasites in Anopheles, thus some of them form the next group of drugs.

1.3.1.5 Sporontocides (sporontocidal drugs or antisporegonic drugs). Drugs in this category, e.g., primaquine and pyrimethamine, prevent or inhibit the formation of oocysts and sporozoites in Anopheles. Therefore, they interfere with the transmission of malaria.

1.3.2 Antimalarial Therapy for Tomorrow

1.3.2.1 Drug Combination

It is now the WHO policy to develop a use of artemisinin derivatives as first-intention drugs to treat severe malaria. A combination of two antimalarial drugs should allow to improve a treatment efficacy and to avoid an emergence of resistant strains. If one parasite is resistant to a drug A in a population of 10^9 parasites, and one parasite is resistant to a drug B in a population of 10^9 parasite, only one in a population of 10^{18} parasites will be resistant to both drugs A and B. Each ill person being carrier of 10^8 to 10^{12} parasites so a probability of simultaneous resistance to two drugs acting with different modes is close to zero.¹⁴

Currently, the most widely used combinations are pyrimethamine + sulfadoxine, chloroquine + proguanil, and atovaquone + proguanil. However, these

combinations are poorly active against strongly chloroquine resistant strains. Artemisinin (or its derivatives) associated with mefloquine are currently used against multi-resistant strains of *P. falciparum*.¹⁵ Furthermore, it appeared that this combination is able to stop a progression of mefloquine resistance.¹⁶

1.3.2.2 Vaccines

A lot of efforts have been made for elaboration of malaria vaccines without success.¹⁷ None of the available *in vitro* assays are predictive of functional immunity *in vivo*, and there is no reliable animal model. Furthermore, the life cycle of *Plasmodium* is complex, several parasite stages in human are morphologically and antigenically distinct, and that obliges to conceive multivalent antigen vaccines. The near availability of a complete sequencing of the *P. falciparum* genome should improve the chances of existing of a vaccine.¹⁸ and will also allow the identification of new parasite proteins to be inhibited. But time is long from discovering targets to develop new therapeutic agents, and high costs are likely for this approach.¹⁹

1.3.2.3 Old Targets, New Compounds

An alternative strategy is an exploitation of known targets: recently discovered targets as the phospholipid (PL) metabolism of infected erythrocytes or old classical targets as the free heme in the food vacuole. The PL metabolism of infected erythrocytes is an effective pharmacological target because of its specificity: malaria parasite needs large amounts of phospholipids, mainly phosphatidylcholine (PC) to grow and divide. A supply of PC is achieved via a choline carrier. Quaternary ammonium and bis-ammoniums salts, designed as choline analogues to target this choline carrier, are highly active *in vitro* even against multi-resistant isolates.²⁰ The free heme liberated in the parasite food vacuole is an “old” but always attractive pharmacological target: it is the most specific target that can be exploited since it comes from the hemoglobin digestion by the parasite, that occurs only in infected erythrocytes. Many chemical entities are directed toward this well-known target, among them are chloroquine and its derivatives, artemisinin and its derivatives. Many quinoline modifications have been investigated to obtain a molecule that is as affordable as chloroquine and active on resistant strains, e.g. by substitutions in the quinoline nucleus, variations in the side chain,²¹ synthesis of polyquinolines,²² and introduction of a ferrocenyl moiety.²³ Such modifications seem to be enough to make

a compound that is active on resistant strains but the request for a safe and effective chloroquine alternative is still going on. Artemisinin and derivatives (artemether, arteether, artesunate) are more and more used in Asia and Africa where multidrug resistant *P. falciparum* is prevalent. But this series of molecules, as well as other antimalarial derivative, is based on artemisinin itself, a molecule having a very short life time in plasma.²⁴ Furthermore, the artemisinin production is mainly limited to China and Vietnam, that is a handicap for other countries. An alternative is a development of synthetic trioxanes, simplified analogues of artemisinin retaining the crucial endoperoxide bridge, but up to now none of them has entered success fully in clinical trials.²⁵

1.3.3 1,2,4-Trioxane Antimalarials

In the 1960s, chinese chemists began to screen traditional herbal drugs in order to find new antimalarial drugs. Among the herbs tested was *Artemisia annua L.*, whose use dates back to 168 B.C.²⁶ In 1972, chinese scientists reported seven sesquiterpene compounds. The compound with principal antimalarial properties was named qinghaosu (artemisinin). Several total syntheses of artemisinin have been reported since its isolation. Artemisinin (Figure 1.2) was found to act on blood phase of *P. falciparum*. Artemisinin and its derivatives are effective against both chloroquine sensitive and chloroquine resistant strains of *P. falciparum*. Artemisinin derivatives have also proven to be useful for a treatment of severe cerebral malaria. The downside of artemisinin based antimalarials is high recrudescence rates, which is attributed to a rapid metabolic clearance. Moreover, artemisinin is poorly soluble in both water and oil and is not well absorbed by a gastrointestinal tract.²⁷

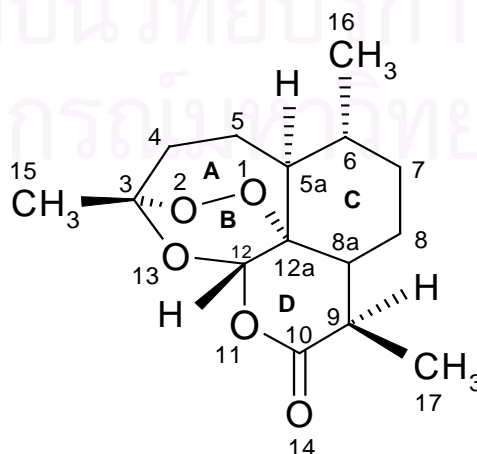


Figure 1.2 Stereochemistry and atomic numbering scheme of artemisinin.

Artemisinin and its hemisynthetic derivatives, artemether and artesunate, are highly efficient against multidrug-resistant parasite strains, but the cost of these naturally occurring drugs and the supply depending on contingencies are major drawbacks. The development of antimalarial synthetic trioxanes which are cheap and have a mode of action similar to that of artemisinin is essential.

The synthesis and testing of many simpler tricyclic trioxanes revealed that certain rings in artemisinin are redundant. Evidences from structure-activity relationship (SAR) investigations indicated the 1,2,4-trioxane as the critical pharmacophore and suggested that neither the peroxide function, nor the 1,2,4-trioxane ring alone, are sufficient for maximum efficacy.²⁸ However with the notation that ring A and lactone ring D (see Figure 1.2) are not essential²⁹ for antimalarial activity. Moreover, several other active tricyclic 1, 2, 4-trioxanes have been synthesized, tested, and confirmed that the full tetracyclic array of artemisinin is not required for high activity.^{30,31,32}

1.4 Aim of Our Study

Because the parasite's resistance to conventional drugs such as quinine, chloroquine, mefloquine and halofantraine is growing at an alarming rate, new efficient drugs are urgently needed. Therefore, more potent antimalarial drugs such as artemisinin, have been developed. Most of the antimalarial activity of artemisinin derivatives (but not artemisinin itself) is due to the metabolite dihydroartemisinin. Oral dihydroartemisinin treatment produces cure rates and parasite clearance time equivalent to historical controls treated with oral artesunate. Dihydroartemisinin can be manufactured more cheaply, however, its pharmacokinetic and pharmacodynamic properties vary greatly and it has some major drawbacks such as poor solubility, short plasma half-life, and complex and expensive total synthesis process. Therefore, simplified analogous of artemisinin retaining the crucial endoperoxide bridge, tricyclic 1,2,4-trioxanes have been developed. Advantages of tricyclic 1,2,4-trioxane are: a) ease of synthesise, which makes the compounds cheap, and b) some of their derivatives are more potent than artemisinin. Most of publications on simple tricyclic 1,2,4-trioxanes deal with synthesis but only few publications on theoretical studies of tricyclic 1,2,4-trioxanes exist.³³ Therefore, it is worth-while to explore antimalarial tricyclic 1,2,4-trioxanes compounds by some other techniques.

In this study, classical QSAR and 3D-QSAR (Comparative Molecular Field Analysis) were applied to 32 antimalarial tricyclic 1,2,4-trioxane compounds. Moreover, molecular docking calculations between tricyclic 1,2,4-trioxane compounds and heme were carried out to investigate their mechanism of action as well as to elucidate their intermolecular interactions. Relationships between biological activity and properties obtained from dock results such as binding energy, O₁-Fe distance, O₂-Fe distance, and O₁₃-Fe distance as well as other molecular properties were investigated. The obtained data are helpful for understanding the mechanism of action of tricyclic 1,2,4-trioxane compounds in more details which could aid in the development of new more effective antimalarial agents.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 2

THEORETICAL BACKGROUND IN QUANTUM CHEMISTRY

2.1 The Schrödinger Equation

The main goal of quantum chemical calculation is to compute energy and other properties of molecule. These properties can be obtained by solving the Schrödinger equation,

$$H\Psi = E\Psi \quad (2.1)$$

Here H is the Hamiltonian, a differential operator representing the total energy, which is the sum of kinetic and potential parts,

$$H = T + V \quad (2.2)$$

The first part is the kinetic energy operator (T), which is a sum of differential operators,

$$T = -\frac{\hbar^2}{8\pi^2} \sum_i \frac{1}{m_i} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \quad (2.3)$$

The sum is over all particles i (nuclei and electrons) and m_i is the mass of particles i . \hbar is the Plank's constant. The second part is potential energy operator (V), which is the Coulomb interaction,

$$V = \sum_{i < j} \left(\frac{e_i e_j}{r_{ij}} \right) \quad (2.4)$$

where the sum is over pairs of particles (i, j) with electric charges e_i, e_j separated by a distance r_{ij} . For electrons $e_i = -e$, while for a nucleus with atomic number Z_i , $e_i = +Z_i e$.

E is numerical value of the energy of the state; that is the energy relative to a state in which the constituent particles (nuclei and electrons) are at infinite separation and at rest. Ψ is the wave function. It depends on the cartesian coordinates of all particles and also on the spin angular momentum components in a particular direction. The square of the wave function, Ψ^2 , is interpreted as a measurement of the probability distribution of the particles within the molecule.

The acceptable solutions of equation (2.1) must be suitable symmetry under interchange of identical particles. For boson particles, the wavefunction is unchanged; that is *symmetric*, under such interchange. For fermion particles, the wavefunction must be multiplied by -1; that is *antisymmetric*. Electrons are fermions, so Ψ must be antisymmetric with respect to interchange of the coordinates of any pair of electrons. This is termed *antisymmetric principle*.

The Schrödinger equation for any molecule has many possible solutions, corresponding to different stationary states. The state with the lowest energy is the ground state.

In molecular system, the Hamiltonian for N electrons and M nuclei is

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.5)$$

In the above equation, M_A is mass of nucleus A and Z_A is the atomic number of nucleus A. The Laplacian operators ∇_i^2 and ∇_A^2 involve differentiation with respect to the coordinates of the *i*th electron and *A*th nucleus. r_{iA} is the distance between the *i*th electron and *A*th nucleus. R_{AB} is the distance between the *A*th nucleus and the *B*th nucleus.

2.2 Molecular Orbital Theory

Molecular orbital theory is an approach to molecular quantum mechanics, using one-electron function or orbitals to approximate the full wavefunction. The first major step in simplifying the general molecular quantum mechanics is the separation of the nuclear and electronic motions. This is possible because the nuclear masses are much greater than those of the electrons therefore, nuclei move much more slowly. The separation of the problem into two parts is called the adiabatic or *Born-Oppenheimer approximation*.³⁴ Therefore, the second term of equation (2.5), the kinetic energy of the nuclei, can be neglected and the last term of equation (2.5), the repulsion between the nuclei, can be considered to be constant. The remaining term in equation(2.5), are called the electronic Hamiltonian (H_{elec}) or Hamiltonian describing the motion of N electrons in the field of M point charges,

$$H_{elec} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (2.6)$$

The corresponding approximation to the total wavefunction is the multiplication product of electronic wavefunction $\psi_{elec}(\{r_i\}; \{R_A\})$, which describes the motion of the electrons that explicitly depends on the electronic coordinates, but parametrically depends on the nuclear coordinates and of nuclei wavefunction $\psi_{nucl}(\{R_A\})$, which describes the vibration, rotation, and translation of a molecule.

$$\Psi(\{r_i\}; \{R_A\}) = \psi_{elec}(\{r_i\}; \{R_A\}) \psi_{nucl}(\{R_A\}) \quad (2.7)$$

The solution to a Schrödinger equation involving the electronic Hamiltonian is:

$$H_{elec} \psi_{elec} = \epsilon_{elec} \psi_{elec} \quad (2.8)$$

The electronic energy, ϵ_{elec} , is also parametric on the nuclear coordinates,

$$\epsilon_{elec} = \epsilon_{elec}(\{R_A\}) \quad (2.9)$$

The total energy for fixed nuclei must also include the constant nuclear repulsion energy,

$$\epsilon_{tot} = \epsilon_{elec} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.10)$$

The total energy $\epsilon_{tot}(\{R_A\})$ provides a potential for nuclear motion. This function constitutes a potential energy surface.

The electronic Hamiltonian in equation (2.6) depends only on spatial coordinates of the electrons. To describe the behavior of an electron, it is necessary to specify its spin. In the context of nonrelativistic theory, there are two spin function $\alpha(\omega)$ and $\beta(\omega)$, corresponding to *spin up* and *spin down*, respectively. From each spatial, $\psi(r)$, one can form two different spin orbitals, $\chi(x)$

$$\chi(x) = \begin{cases} \chi(x) = \psi(r)\alpha(\omega) \\ or \\ \chi(x) = \psi(r)\beta(\omega) \end{cases} \quad (2.11)$$

The four coordinates of electron is denoted by x,

$$x = \{r, \omega\} \quad (2.12)$$

The wavefunction for an N-electron system is then a function of x_1, x_2, \dots, x_N . This is $\psi(x_1, x_2, \dots, x_N)$. Wavefunction must correspond to the antisymmetric or Pauli exclusion principle, a many electron wavefunction must be antisymmetric with respect to the interchange of the coordinates x (both space and spin) of any two electrons, i.e.,

$$\psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = -\psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N) \quad (2.13)$$

So the electronic wavefunction must be written as Slater determinant of spin orbital,

$$\Psi(x_1, x_2, \dots, x_N) = (N!)^{-1/2} \begin{vmatrix} \chi_1(x_1) & \chi_j(x_1) & \cdots & \chi_k(x_1) \\ \chi_1(x_2) & \chi_j(x_2) & \cdots & \chi_k(x_2) \\ \vdots & \vdots & & \vdots \\ \chi_l(x_N) & \chi_j(x_N) & \cdots & \chi_k(x_N) \end{vmatrix} \quad (2.14)$$

or short notation,

$$\Psi(x_1, x_2, \dots, x_N) = |\chi_1(x_1)\chi_j(x_2)\dots\chi_k(x_N)\rangle \quad (2.15)$$

where the factor $(N!)^{1/2}$ is a normalization factor.

2.3 The Hartree-Fock Approximation

The simplest antisymmetric wavefunction, which can be used to describe the ground state of the an N-electron system, Ψ_0 , is a single Slater determinant,

$$\Psi_0 = |\chi_1\chi_2 \dots \chi_a\chi_b \dots \chi_N\rangle \quad (2.16)$$

The variation principle states that the best wavefunction of this functional form is the one which gives the lowest possible energy,

$$E_0 = \langle \Psi_0 | H | \Psi_0 \rangle \quad (2.17)$$

where H is the full electronic Hamiltonian. The variational flexibility in the wavefunction (2.16) is in the choice of spin orbitals, one can derive eigenvalue equation, called the *Hartree-Fock equation* which determines the optimal spin orbitals of the form

$$f(i)\chi(x_i) = \varepsilon\chi(x_i) \quad (2.18)$$

where $f(i)$ is an effective one-electron operator, called the *Fock operator*, of the form

$$f(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + V^{\text{HF}}(i) \quad (2.19)$$

where $V^{\text{HF}}(i)$ is the average potential experienced by the i th electron due to the presence of the other electrons. The essence of the Hartree-Fock approximation is to replace the complicate many-electron problem by one-electron problem in which electron-electron repulsion is treated in an average way.

The Hartree-Fock potential $v^{\text{HF}}(i)$, or equivalently the "field" seen by the i th electron, depends on the spin orbitals of the other electrons. Thus the Hartree-Fock equation (2.18) is nonlinear and must be solved iteratively. The procedure for solving the Hartree-Fock equation is called the *self-consistent-field* (SCF) method.

From equation (2.19), the best (Hartree-Fock) spin orbitals is the Hartree-Fock integro-differential equation

$$h(1)\chi_a(1) + \sum_{b \neq a}^N \left[\int dx_2 |\chi_b(2)|^2 r_{12}^{-1} \right] \chi_a(1) - \sum_{b \neq a}^N \left[\int dx_2 \chi_b^*(2) \chi_a(2) r_{12}^{-1} \right] \chi_b(1) = \varepsilon_a \chi_a(1) \quad (2.20)$$

or

$$\left[h(1) + \sum_{b \neq a}^N J_b(1) - \sum_{b \neq a}^N K_b(1) \right] \chi_a(1) = \varepsilon_a \chi_a(1) \quad (2.21)$$

where $h(1)$ is core-Hamiltonian operator

$$h(1) = -\frac{1}{2} \nabla_1^2 - \sum_{A=1}^M \frac{Z_A}{r_{1A}} \quad (2.22)$$

$J_b(1)$ is the coulomb operator

$$J_b(1) = \int dx_2 |\chi_b(2)|^2 r_{12}^{-1} \quad (2.23)$$

$K_b(1)$ is the exchange operator

$$K_b(1) = \int dx_2 \chi_b^*(2) r_{12}^{-1} P_{12} \chi_b(2) \quad (2.24)$$

Therefore, the Fock operator $f(1)$ can be written as

$$f(1) = h(1) + \sum_{b=1}^N [J_b(1) - K_b(1)] \quad (2.25)$$

and the Hartree-Fock potential $V^{\text{HF}}(1)$,

$$V^{\text{HF}}(1) = \sum_{b=1}^N [J_a(1) - K_b(1)] \quad (2.26)$$

the orbital energy ε_a ,

$$\varepsilon_a = \int \chi_a |h| \chi_a d\tau + \sum_{b=1}^N \left[\int \chi_a \chi_a r_{12}^{-1} \chi_b \chi_b d\tau_1 d\tau_2 - \int \chi_a \chi_b r_{12}^{-1} \chi_b \chi_a d\tau_1 d\tau_2 \right] \quad (2.27)$$

and Hartree-Fock energy

$$E_0 = \sum_{a=1}^N \int \chi_a |h| \chi_a d\tau + \frac{1}{2} \sum_{a=1}^N \sum_{b=1}^N \left[\int \chi_a \chi_a r_{12}^{-1} \chi_b \chi_b d\tau_1 d\tau_2 - \int \chi_a \chi_b r_{12}^{-1} \chi_b \chi_a d\tau_1 d\tau_2 \right] \quad (2.28)$$

$$E_0 = \sum_{a=1}^N \varepsilon_a + \frac{1}{2} \sum_{a=1}^N \sum_{b=1}^N \left[\int \chi_a \chi_a r_{12}^{-1} \chi_b \chi_b d\tau_1 d\tau_2 - \int \chi_a \chi_b r_{12}^{-1} \chi_b \chi_a d\tau_1 d\tau_2 \right] \quad (2.29)$$

For closed-shell restricted Hartree-Fock wavefunction

$$\begin{aligned} |\Psi_0\rangle &= |\chi_1 \chi_2 \chi_3 \chi_4 \dots \chi_{N-1} \chi_N\rangle \\ &= |\psi_1 \bar{\psi}_1 \psi_2 \bar{\psi}_2 \dots \psi_{N/2} \bar{\psi}_{N/2}\rangle \end{aligned}$$

$$E_0 = 2 \sum_{a=1}^{N/2} \int \psi_a |h| \psi_a d\tau + \sum_{a=1}^{N/2} \sum_{b=1}^{N/2} \left[2 \int \psi_a \psi_a r_{12}^{-1} \psi_b \psi_b d\tau_1 d\tau_2 - \int \psi_a \psi_b r_{12}^{-1} \psi_b \psi_a d\tau_1 d\tau_2 \right]$$

or

$$E_0 = 2 \sum_{a=1}^{N/2} h_{aa} + \sum_{a=1}^{N/2} \sum_{b=1}^{N/2} (2J_{ab} - K_{ab}) \quad (2.30)$$

where $J_{ab} = \int \psi_a \psi_a r_{12}^{-1} \psi_b \psi_b d\tau_1 d\tau_2$ and $K_{ab} = \int \psi_a \psi_b r_{12}^{-1} \psi_b \psi_a d\tau_1 d\tau_2$

Equation (2.30) is the eliminated spin form, the calculation of molecular orbitals become equivalent to the problem of solving the spatial integro-differential equation

$$f(r_i)\psi(r_i) = \varepsilon_i\psi(r_i) \quad (2.31)$$

Solving this equation numerically, Roothaan introduces a set of known *basis functions* (ϕ_μ) and $\{\phi_\mu(r), \mu = 1, 2, \dots, K\}$ expand the unknown molecular orbitals (ψ_i) in the linear expansion of these functions,

$$\psi_i = \sum_{\mu=1}^K C_{\mu i} \phi_\mu \quad i = 1, 2, \dots, K \quad (2.32)$$

The quality of the molecular orbitals is related to the quality of the basis set, set of basis functions used.

From equation (2.31), the problem of calculating the Hartree-Fock molecular orbitals reduces to the problem of calculating the set of expansion coefficients $C_{\mu i}$.

By substitution the linear expansion (2.32) into Hartree-Fock equation (2.31) and using the index ν , gives

$$f(1) \sum_{\nu=1}^K C_{\nu i} \phi_\nu(1) = \epsilon_i \sum_{\nu=1}^K C_{\nu i} \phi_\nu(1) \quad (2.33)$$

then multiply by $\phi_\mu^*(1)$ on the left and integrate

$$\sum_{\nu=1}^K C_{\nu i} \int dr_1 \phi_\mu^*(1) f(1) \phi_\nu(1) = \epsilon_i \sum_{\nu=1}^K C_{\nu i} \int dr_1 \phi_\mu^*(1) \phi_\nu(1) \quad (2.34)$$

and define two matrices, are the overlap matrix S , has elements

$$S_{\mu\nu} = \int dr_1 \phi_\mu^*(1) \phi_\nu(1) \quad (2.35)$$

and the Fock matrix F , has elements

$$F_{\mu\nu} = \int dr_1 \phi_\mu^*(1) f(1) \phi_\nu(1) \quad (2.36)$$

Therefore, the integrated Hartree-Fock equation is written as

$$\sum_{\nu=1}^K F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu=1}^K S_{\mu\nu} C_{\nu i} \quad i = 1, 2, \dots, K \quad (2.37)$$

These are the *Roothaan-Hall equations*, proposed by Roothaan³⁵ and Hall³⁶ which can be written as the single matrix equation

$$FC = SC\epsilon \quad (2.38)$$

Where C is a $K \times K$ square matrix of the expansion coefficient $c_{\mu i}$

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & & \vdots \\ c_{K1} & c_{K2} & \cdots & c_{KK} \end{pmatrix} \quad (2.39)$$

and ε is a diagonal matrix of the orbital energies ε_i ,

$$\varepsilon = \begin{pmatrix} \varepsilon_1 & 0 & \cdots & 0 \\ & \varepsilon_2 & & \\ & & \ddots & \\ 0 & & & \varepsilon_k \end{pmatrix} \quad (2.40)$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu} \quad (2.41)$$

where $H_{\mu\nu}^{core}$ is the Hamiltonian matrix

$$H_{\mu\nu}^{core} = \int dr_1 \phi_\mu^*(1) h(1) \phi_\nu(1) \quad (2.42)$$

$G_{\mu\nu}$ is the two-electron part of the Fock matrix

$$G_{\mu\nu} = \sum_{\lambda=1}^{N/2} \sum_{\sigma=1}^{N/2} P_{\lambda\sigma} \left[(\mu\nu | \sigma\lambda) - \frac{1}{2} (\mu\lambda | \sigma\nu) \right] \quad (2.43)$$

$P_{\lambda\sigma}$ is the density matrix

$$P_{\mu\nu} = 2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^* \quad (2.44)$$

and $(\mu\nu | \lambda\sigma) = \int dr_1 dr_2 \phi_\mu^*(1) \phi_\nu(1) r_{12}^{-1} \phi_\lambda^*(2) \phi_\sigma(2) \quad (2.45)$

For solving the Roothaan-Hall equation, iterative process called a *Self Consistent Field* (SCF) procedure is required. The outline of mathematical steps to solve the Roothaan-Hall equations for a closed-shell system are shown in the Figure 2.1.

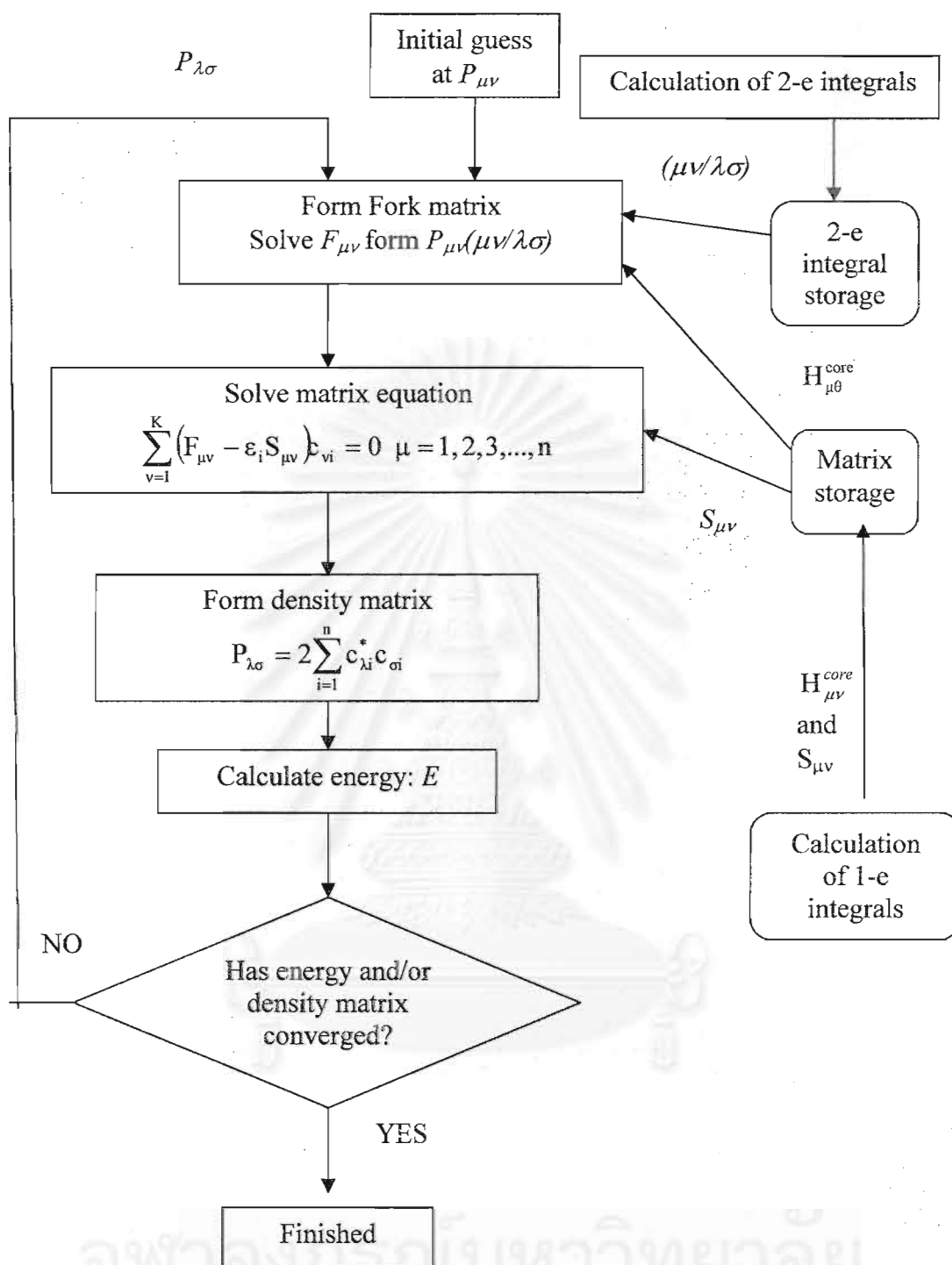


Figure 2.1 Sequence of program step required for the solution of the Roothaan-Hall equations for closed-shell, Self Consistent Field procedure.

2.4 Basis Set

The basis set most commonly used in quantum mechanical calculations are composed of atomic functions. The next approximation involves expressing the molecular orbitals as linear combinations of a predefined set of one-electron function. An individual molecular orbitals is defined as:

$$\psi_i = \sum_{\mu=1}^N c_{\mu i} \phi_{\mu} \quad (2.46)$$

where the coefficients $c_{\mu i}$ are known as molecular orbital expansion coefficients. The $\phi_1, \phi_2, \dots, \phi_N$ are the orthogonal function with known expression. The set of this function is called "basis set".

The best solution to the approximation of MO could theoretically be obtained by the use of an infinite and complete set of basis functions. The most often used mathematical expressions for the basis functions are the Slater-type orbital (STO) and the Gaussian-type orbitals (GTO).

The Slater-type orbitals has the form

$$\phi^{STO} = r^{n-1} e^{-\zeta r} Y_l^m(\theta, \phi) \quad (2.47)$$

where n, l, m are quantum numbers and ζ is orbital exponent. The r, θ, ϕ are the spherical polar coordinates. This Slater functions were introduced by Slater³⁷ and extremely close in form to the hydrogenic orbitals. In the past, they were mostly used for the calculations of small molecules. Their advantage is that only few functions are needed for a good description, but not suited to the numerical work, and their use in practical molecular orbital calculations has been limited.

The Gaussian-type orbitals has the form

$$\phi^{GTO} = x^a y^b z^c e^{-\alpha r^2} \quad (2.48)$$

where a, b, c are integer numbers and α is the orbital exponent. The x, y, z are the cartesian coordinates. The GTO were introduced into molecular orbital calculations by Boy.³⁸ They are less satisfactory than STO as representations of atomic orbital, particularly because they do not have a cusp at the origin (Figure 2.2). Due to this deficiency more than one GTOs are often required for a good description. At present,

most calculation performed uses GTOs instead of STOs since the computation of GTO is much faster than that of the STO. So, as many GTO functions, which yields the same quality to the STO can be computed in the fraction of time.

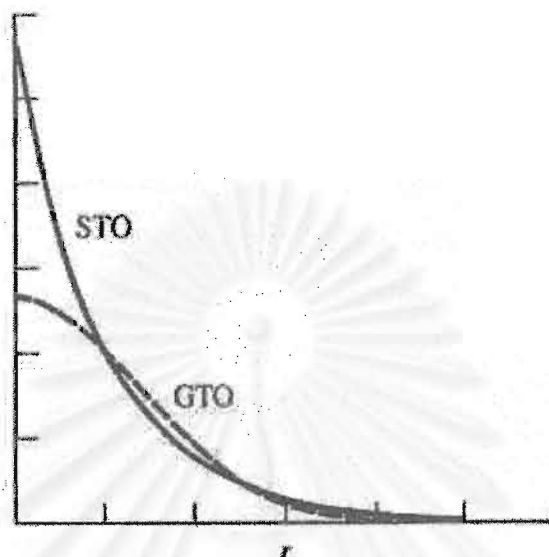


Figure 2.2 The cusp of Slater function.

2.4.1 Minimal Basis Set

A minimal basis set is a representation that contains just the number of functions that are required to accommodate all the filled orbitals in each atom. In practice, a minimal basis set normally includes all of the atomic orbitals in the shell. Thus, for hydrogen and helium a single s-type function would be required. For elements from lithium to neon the 1s, 2s and 2p functions are used and so on. The basis sets STO- n G for example, STO-3G, STO-4G, are all minimal basis sets in which n Gaussian functions are used to represent the Slater orbital. The minimal basis sets are well known to accompany with several deficiencies. There are particular problems with compounds containing atoms at the end of a period, such as oxygen or fluorine. Such atoms are described using the same number of basis functions as the atoms at the beginning of the period despite they have more electrons. A minimal basis set only contains one function per atomic orbital. Since the radial exponents are not allowed to vary during the calculation, the orbital cannot expand or contract in size in accordance with the molecular environment.

2.4.2 Extended Basis Set

The problems with minimal basis sets can be addressed if more than one function is used for each orbital. For STO, a basis set which doubles the number of functions in the minimal basis set is described as *double zeta* basis. The double or triple or more of STO minimal basis function allows the linear combination of the 'contracted' (large exponent) and the 'diffuse' (small exponent) functions which gives an overall result that is intermediate between the two. In other words, the size of orbital can be modified during the course of calculation. An alternative to the double zeta basis approach is to double the number of functions used to describe the valence electrons but to keep a single function for the inner shells called "split valence double zeta basis". For GTO, the similar notation to STO can also be used. The 3-21G exemplifies the notation used for such split valence double zeta basis sets. In this basis set, three Gaussian functions are used to describe the core orbitals. The valence electrons are also represented by three Gaussian; the contracted part by two Gaussian and the diffuse part by one Gaussian. The most commonly used split valence basis sets are 3-21G, 6-31G and D95V (Dunning's split valence double zeta basis).

2.4.3 Polarized Basis Set

Simply increasing the number of basis functions does not necessarily improve the model. This possibly leads to a wholly erroneous result, particularly for molecules with strongly anisotropy charge distribution. This distortion can be considered to correspond to mixing p-type character into the 1s orbital of the isolated atom, to give a form of sp hybrid. In a similar manner, the unoccupied d orbitals introduce asymmetry into p orbital. In other words, the addition of p function for H atom and d function for heavier atom enable orbitals on nuclei to polarize and form bond. These functions, p for hydrogen and d for 1st and 2nd row element, have a higher angular momentum and they are called the "polarization" function. An asterisk (*) indicates the use of polarization basis functions. Thus, 6-31G* refers to a 6-31G basis set with polarization functions on the heavy (non-hydrogen) atoms. Two asterisks, such as 6-31G** indicate the use of polarization functions on hydrogen and heavy atoms.

2.4.4 Basis set Incorporating Diffuse Function

Another deficiency of the basis sets is their inability to deal with species that have a significant amount of electron density away from the nuclear centers such as anions and molecules containing lone pair. This failure arises because the amplitudes of the Gaussian basis functions are rather low far from the nuclei. To remedy this deficiency highly diffuse functions *i.e.* functions with very small exponent, is added to the basis set. These basis sets are denoted using a '+'; thus the 3-21+G basis set contains an additional single set of diffuse s- and p-type Gaussian functions for heavy atoms. A '++' indicates that the diffuse functions are included for both hydrogen and heavy atoms such 6-31++G.

2.5 Population Analysis

The probability of finding an electron in various regions of space, $\rho(r)$, is called the charge density and is defined as,

$$\rho(r) = \sum_{\mu=1}^{N/2} \sum_{\nu=1}^{N/2} P_{\mu\nu} \phi_{\mu}(r) \phi_{\nu}^*(r) \quad (2.49)$$

Therefore, the total number of electrons (N) is

$$N = 2 \sum_{a=1}^{N/2} \int dr |\psi_a(r)|^2 \quad (2.50)$$

By substitution eq. (2.32) into eq.(2.50) and using the index μ , the equation(2.50) becomes

$$N = \sum_{\mu=1}^{N/2} \sum_{\nu=1}^{N/2} P_{\mu\nu} S_{\nu\mu} = \sum_{\mu=1}^{N/2} (PS)_{\mu\mu} = trPS \quad (2.51)$$

$(PS)_{\mu\mu}$ indicated the number of electrons to be associated ϕ_{μ} . This is called a *Mulliken population analysis*.³⁹ The net charge of the an atom (q_A) is given by

$$q_A = Z_A - \sum_{\mu \in A} (PS)_{\mu\mu} \quad (2.52)$$

where Z_A is the charge of atomic nucleus A.

One important property of molecule involving the electrons distribution is dipole moment, μ . The dipole moment of molecule can be calculated as

$$\mu = \left\langle \Psi_0 \left| - \sum_{i=1}^N r_i \right| \Psi_0 \right\rangle + \sum_A Z_A R_A \quad (2.53)$$

where the first term is the contribution of the electrons, of charge -1, and the second term is the contribution of the nuclei, Z_A to the dipole moment. The electronic dipole moment operator $-\sum_{i=1}^N r_i$, a sum of one-electron operators.

Therefore,

$$\mu = - \sum_{\lambda=1}^{N/2} \sum_{\sigma=1}^{N/2} P_{\lambda\sigma} (\lambda | r | \sigma) + \sum_A Z_A R_A \quad (2.54)$$

A vector equation with components (for example the x component) is

$$\mu_x = - \sum_{\lambda=1}^{N/2} \sum_{\sigma=1}^{N/2} P_{\lambda\sigma} (\lambda | x | \sigma) + \sum_A Z_A X_A \quad (2.55)$$

where

$$(\lambda | x | \sigma) = \int dr \phi_\lambda^*(r) x \phi_\sigma(r) \quad (2.56)$$

With the same manner, μ_y and μ_z can be also computed.

จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 3

STRUCTURE AND BIOLOGICAL DATA

3.1 Chemical Structure of Tricyclic 1,2,4-trioxane

An empirical formula of compound is $C_{12}H_{20}O_4$ with a molecular weight of 228.3, a melting point of 74-75 $^{\circ}C$ and a density of 1.275 g/cm^3 . Its stereochemistry and atomic numbering scheme according to the IUPAC of artemisinin is shown in Figure 3.1.⁴⁰

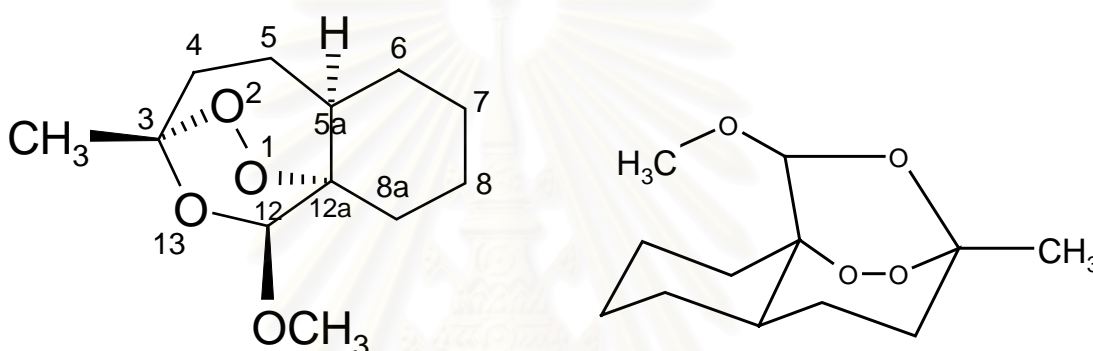


Figure 3.1 Stereochemistry and atomic numbering scheme of tricyclic 1,2,4-trioxanes.

3.2 Structures and Antimalarial Activities of Tricyclic 1,2,4-trioxane derivatives

Totally 32 derivatives of tricyclic 1,2,4-trioxane with significantly different structures and biological activities taken from literatures⁴¹⁻⁴⁴ were used in this study. The activities were measured *in vitro* against the NF54 strain of *Plasmodium falciparum*, a chloroquine-sensitive strain, as IC_{50} values, an inhibitory concentration of a compound required for 50% inhibition of parasitemia. Since the biological data arose from different literatures, there were an inconsistency from individual experimental testing procedures. For example, the activities of artemisinin measured in references 41, 42, 43 and 44, are 11, 9.2, 8.5 and 9.9 nM, respectively. Therefore, antimalarial activities of all compounds have to be converted into values in the same scale. For this purpose, the activity of artemisinin in reference 41 ($IC_{50} = 11$ nM) was set as a standard value. Then, the activity

of artemisinin from reference 42-44 was compared to the standard values to give a scaling factor as shown in equation (3.1). Subsequently, the activity of each compound in reference 42-44 was multiplied by the corresponding scaling factor. Finally, the antimalarial activity is calculated in logarithm unit by equation (3.2).

$$\text{Scaling Factor (S)} = \frac{\text{IC}_{50} \text{ of artemisinin in reference [41]}}{\text{IC}_{50} \text{ of artemisinin in reference [42 or 43 or 44]}} \quad \dots\dots (3.1)$$

$$\text{Activity} = \log \left(\frac{1}{\text{S} * \text{IC}_{50} \text{ (in molar unit)}} \right) \quad \dots\dots (3.2)$$

The structures and biological data of 32 compounds with the reference sources are given in Table 3.1-3.4. All the compounds have different substituent groups, which can be possible effects of structural differences in tricyclic 1,2,4-trioxane compounds concerning their biological activities. Compounds 1-15 have substituent groups at the C₃ position in which compounds 1-4 have alkyl substituent groups while compounds 5-15 have aryl substituent groups. Preliminary analysis of the raw dataset indicated interesting features as follows. Compounds with aryl substituent groups are more potent than those with alkyl substituents. Comparing compound 10 to compounds 11 and 12, a *p*-fluoro substituent in compound 12 enhances the antimalarial activity while an *o*-methyl substituent in compound 11 decrease the antimalarial potency. A comparison between compounds 6 and 12 showed that the nature of the halogen substituent, chloro, fluoro has significant effect. Remarkably, compounds 13-15, oxygen-containing benzylic trioxanes are potent antimalarials. Compounds 16-21 have substituent groups at the C₁₂ position. The sulfone trioxanes (compounds 19-21) have higher antimalarial potencies than the sulfides (compounds 16-18). Compounds 22-24 have substituent groups at the C₄ position. Compound 24 with *p*-fluorobenzyl ether substituent is the most active compound in this group. Compounds 25-32 have substituent groups at the C₄ and/or C_{8a} positions and these compounds have rather high activities. Comparisons of compounds 25 to 28, compounds 26 to 29 and compounds 27 to 31, in which each pair have the same substituent groups at the C_{8a} position but different substituent groups at the C₄ position,

show that changing of hydrogen atom at the C₄ position to a methyl group can increase antimalarial activities.

Table 3.1 Structures and biological data of compounds number 1-15.

Compound No.	R ₁	IC ₅₀ (nM)	Activity	Ref.
1	FCH ₂	160	6.718	42
2	(CH ₃) ₂ CHCH ₂ CH ₂	160	6.796	41
3	PhCH ₂ CH ₂ CH ₂	110	6.959	41
4	CF ₃ CH ₂ CH ₂	84	6.998	42
5	<i>p</i> -PhPh	68	7.090	42
6	<i>p</i> -ClPh	55	7.182	42
7	<i>p</i> -CF ₃ Ph	53	7.198	42
8	<i>p</i> -CH ₃ OCH ₂ Ph	51	7.215	42
9	1-naphthyl	44	7.279	42
10	Ph	38	7.342	42
11	<i>p</i> -F- <i>o</i> -CH ₃ Ph	34	7.391	42
12	<i>p</i> -FPh	30	7.445	42
13	<i>p</i> -(<i>p</i> '-FPhCH ₂ OCH ₂)Ph	23	7.561	42
14	<i>p</i> -CH ₃ C(O)OCH ₂ Ph	20	7.621	42
15	<i>p</i> -HOCH ₂ Ph	15	7.746	42

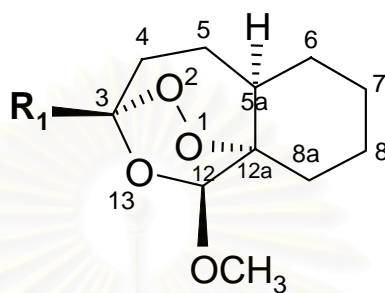


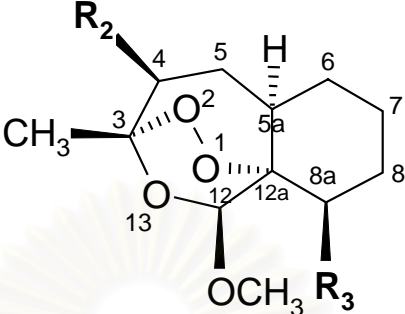
Table 3.2 Structures and biological data of compounds number 16-21.

Compound No.	R ₄	IC ₅₀ (nM)	Activity	Ref.
16	<i>p</i> -ClPhS	110	6.847	43
17	<i>p</i> -CH ₃ OPhS	89	6.939	43
18	PhS	56	7.140	43
19	PhSO ₂	33	7.370	43
20	<i>p</i> -CH ₃ OPhSO ₂	30	7.411	43
21	<i>p</i> -ClPhSO ₂	23	7.526	43

Table 3.3 Structures and biological data of compounds number 22-24.

Compound No.	R ₂	IC ₅₀ (nM)	Activity	Ref.
22	PhCH ₂	310	6.445	44
23	OHCH ₂	230	6.575	44
24	<i>p</i> -FPhCH ₂ OCH ₂	19	7.658	44

Table 3.4 Structures and biological data of compounds number 25-32.



Compound No.	R ₂	R ₃	IC ₅₀ (nM)	Activity	Ref.
25	H	(<i>p</i> -FPhCH ₂)OCH ₂ CH ₂	31	7.445	44
26	H	PhCH ₂ OCH ₂ CH ₂	25	7.538	44
27	H	(PhO) ₂ P(O)OCH ₂ CH ₂	14	7.790	44
28	CH ₃	(<i>p</i> -FPhCH ₂)OCH ₂ CH ₂	13	7.822	44
29	CH ₃	PhCH ₂ OCH ₂ CH ₂	11	7.895	44
30	CH ₃	OHCH ₂ CH ₂	7.7	8.050	44
31	CH ₃	(PhO) ₂ P(O)OCH ₂ CH ₂	6.9	8.097	44
32	PhCH ₂	OHCH ₂ CH ₂	8.3	8.081	41

3.3 Geometry Optimization and Atomic Charge Calculation

Quantum chemical calculations have been proven as helpful means to calculate directly molecular properties and geometries. The validity of these methods can be examined by comparing structures and properties with those of experimental results. As an experimental structure of most compounds used in this study is not available, quantum chemical calculations were applied to determine the geometry of all compounds. However, it is necessary to examine which level of calculation provides the more accurate geometry compared to the x-ray crystallographic data.

A starting geometry of tricyclic 1,2,4-trioxane was obtained from the crystallographic structure.⁴⁰ In our previous study,⁴⁰ artemisinin structure was optimized using CNDO, AM1, Hartree Fock (HF) with 3-21G and 6-31G(d,p) basis sets, and Density Functional Theory (DFT) with B3LYP functional and 6-31G(d,p) basis set. All optimized structures were compared to the X-ray structure. The results indicated that the Hartree Fock (HF) with 3-21G basis set (HF/3-21G) is the lowest level of theory that

gives geometrical parameters within acceptable accuracy of the X-ray data. Since tricyclic 1,2,4-trioxane has very similar structure to artemisinin, the HF/3-21G method should also be an appropriate method for geometry optimization of tricyclic 1,2,4-trioxane compounds. Therefore, the tricyclic 1,2,4-trioxane compound was optimized at the HF/3-21G level and then was compared to the X-ray structure in order to confirm our assumption (see Table 3.5).

Table 3.5 Comparison of important structural parameters of tricyclic 1,2,4-trioxane as obtained from the X-ray and HF/3-21G optimization method.

Parameter	X-ray	HF/3-21G
Bond Length (Å)		
O ₁ -O ₂	1.467	1.465
O ₂ -C ₃	1.443	1.442
C ₃ -C ₄	1.529	1.535
C ₄ -C ₅	1.529	1.538
C ₅ -C _{5a}	1.525	1.541
C _{12a} -O ₁	1.473	1.492
C _{12a} -C ₁₂	1.524	1.529
C ₁₂ -O ₁₃	1.428	1.434
C ₃ -O ₁₃	1.424	1.433
C _{5a} -C _{12a}	1.535	1.535
Bond Angle (°)		
C ₅ -C _{5a} -C _{12a}	113.4	113.2
C ₄ -C ₅ -C _{5a}	115.1	114.1
C ₃ -C ₄ -C ₅	114.0	112.3
O ₁₃ -C ₃ -C ₄	110.7	111.9
C ₁₂ -O ₁₃ -C ₃	113.3	114.3
C ₆ -C _{5a} -C _{12a}	113.1	112.3
C _{8a} -C _{12a} -C _{5a}	111.4	111.9

Table 3.5 (Continued)

Bond Angle ($^{\circ}$)	X-ray	HF/3-21G
O ₂ -C ₃ -C ₄	112.6	111.7
C _{5a} -C ₆ -C ₇	112.3	111.5
C ₈ -C _{8a} -C _{12a}	114.3	113.2
C _{12a} -C ₁₂ -O ₁₃	112.4	111.9
O ₁₃ -C ₃ -O ₂	107.7	106.8
C ₃ -O ₂ -O ₁	110.0	108.9
O ₂ -O ₁ -C _{12a}	112.6	112.3
O ₁ -C _{12a} -C ₁₂	106.2	107.4
Torsion Angle ($^{\circ}$)		
C _{12a} -C _{5a} -C ₅ -C ₄	41.8	44.1
C ₃ -C ₄ -C ₅ -C _{5a}	58.5	62.8
O ₁₃ -C ₃ -C ₄ -C ₅	25.6	21.6
C ₄ -C ₃ -O ₁₃ -C ₁₂	91.3	90.5
O ₁ -C _{12a} -C _{5a} -C ₅	73.7	72.7
C ₅ -C ₄ -C ₃ -O ₂	95.0	97.3
C ₇ -C ₆ -C _{5a} -C _{12a}	53.2	54.0
C ₈ -C _{8a} -C _{12a} -C _{5a}	48.2	51.3
O ₁ -O ₂ -C ₃ -O ₁₃	71.5	73.2
C ₃ -O ₂ -O ₁ -C _{12a}	39.9	44.9
O ₂ -O ₁ -C _{12a} -C ₁₂	23.0	18.2
O ₁₃ -C ₁₂ -C _{12a} -O ₁	61.5	58.8
C _{12a} -C ₁₂ -O ₁₃ -C ₃	32.8	33.0
C ₁₂ -O ₁₃ -C ₃ -O ₂	32.2	31.2
C _{5a} -C _{12a} -O ₁ -O ₂	102.6	105.4
C ₄ -C ₅ -C _{5a} -C _{12a}	41.8	44.1
C _{5a} -C ₅ -C ₄ -C ₃	58.5	62.8
O ₂ -C ₃ -C ₄ -C ₅	95.0	97.3
C ₄ -C ₃ -O ₂ -O ₁	50.8	47.9

The results show that HF/3-21G can reproduce most of the structural parameters very reliably in comparison to the X-ray structure. Thus, the HF/3-21G level of theory was used for the geometry optimizations of all tricyclic 1,2,4-trioxane derivatives. All calculations were done using the Gaussian 98 software.⁴⁶



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 4

CLASSICAL-QSAR

4.1 Introduction

Quantitative structure-activity relationship (QSAR) is a mathematical relationship linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. The mathematical methods, used in QSAR include various regression and pattern recognition techniques. QSAR is taken to be equivalent to chemometrics or multivariate statistical data analysis. QSAR study involves selecting the representations of molecules that can explain the activity induced in the biological system. These representations are generally referred to physicochemical parameters. Physicochemical properties which usually relate to activity of molecule are hydrophobic, steric and electronic properties, etc.

4.1.1 Classical QSAR

Classical QSAR techniques were original QSAR approach, attempting to relate physicochemical properties to pharmacological effect. The first QSAR study was done in 1893 by Charles Richet.⁴⁷ A few years later Meyer and Overton independently found linear relationships between lipophilicity (expressed as solubility or oil-water partition coefficient) and biological effect. In 1964, Hansch and Fujita⁴⁸ published their studies on quantitative relationships between physicochemical properties (i.e., lipophilicity substituent, electronic substituent, and reaction constants) and biological activities. Free and Wilson developed a model of additive group contributions to biological activity values.⁴⁹ In principle, the classical QSAR approaches are established by considering physicochemical properties related with the activities of a structurally homologous series of ligands and are solved by means of statistical techniques like linear and non-linear regression, cluster and discriminant analysis. The most commonly used method is multiple linear regression (MLR) analysis for creating QSAR model.

4.1.1.1 Hydrophobicity Properties

The hydrophobic or lipophilic character of a drug can be measured experimentally by testing the drugs's relative distribution in an octanol/water mixture. Hydrophobic molecules prefer to dissolve in 1-octanol layer of this two phase system, whereas hydrophilic molecules prefer the aqueous layer. The relative distribution is known as the partition coefficient (P) and is obtained from the following equation (4.1):

$$P = \frac{\text{Concentration of a solute in 1-octanol}}{\text{Concentration of a solute in aqueous phase}} \quad \dots\dots(4.1)$$

From the definition of P, it is obvious that hydrophobic compounds have a high P value, whereas hydrophilic compounds have a low P value. However, the main drawback of measuring P experimentally is that the compound has to be synthesized. Moreover, the measurement is sometimes not easy. Therefore, it is much better to calculate P theoretically.

Partition coefficient can be calculated by knowing the contribution that various substituent make to hydrophobicity. This contribution is known as the Hansch substituent hydrophobicity constant (π).⁵⁰ Partition coefficients are measured for a standard compound with and without a substituent (x). The hydrophobicity constant (π_x) for the substituent (x) is then obtained using the following equation (4.2):

$$\pi_x = \log P_x - \log P_H \quad \dots\dots(4.2)$$

where P_x and P_H are the partition coefficients for the standard compound with and without the substituent, respectively.

A positive value of π indicates that the substituent is more hydrophobic. These π values are characteristic for the substituents and can be used to calculate how the partition coefficient of a drug would be affected by adding these substituents. In addition, partition coefficient is additive constitutive parameter, like some other molecular properties. Log P is highly correlated with a diversity of biological activities and plays a significant role in the interactions between drugs and their receptors. The overall hydrophobicity of a molecule can be measured by its partition coefficient ($\log P$) in polar/nonpolar heterogeneous reference system. A

comprehensive study of partition experiments in the octanol/water system leads to the definition of hydrophobic contributions of single atom in their specific structural environment.⁵¹ These atomic partial values can be regarded as fragmental increments, f_i , to the total lipophilicity given by logP in equation (4.3).

$$\log P = \sum_i f_i \quad \dots\dots(4.3)$$

4.1.1.2 Steric Properties

The measure of steric factor can be obtained by several parameters e.g., molar refractivity (MR). This molar refractivity is calculated from the following equation (4.4):

$$MR = \frac{MW}{d} \cdot \frac{n^2 - 1}{n^2 + 2} \quad \dots\dots(4.4)$$

where n is refractive index, MW is a molecular weight, and d is a density. The term $\frac{MW}{d}$ defines a volume, while the $\frac{(n^2 - 1)}{(n^2 + 2)}$ term describes how easily the substituent

can be polarized. This is particularly significant if the substituent has π electrons or lone pair of electrons. MR is also an additive constitutive molecular property, like logP parameter. MR is correlated with lipophilicity, molar volume and steric bulk.

Due to its $\frac{MW}{d}$ component, it is indeed related to volume and size of a substituent.

The refractive index, related correlation term in MR, account polarizability and thus for the size and the polarity of a certain group. The larger the polar part of molecule is, the larger its MR value will be. Molar refractivity normally has significant contributions to the QSAR equations of ligand-enzyme interactions.

Other widely used steric descriptors for biological activity investigation are bond length, bond angle, and dihedral angle parameters, which can be simply measured from the structure, are members of this group. The topological indices, which are calculated using the chemical graph theory⁵² as the basis, are also widely used. Examples of these indices are the Wiener index,⁵³ molecular connectivity indices (Chi),⁵⁴ valence-modified molecular connectivity indices (ChiV),⁵⁴ and molecular shape indices (Kappa).⁵⁴ The Wiener index is the sum of distances between all pairs of heavy atoms in the molecule. The Chi and ChiV indices reflect

the atom identities, bonding environments, and number of bonding hydrogens. Molecules that are drawn without hydrogen atoms can be decomposed into fragments of length m , which may be divided into four different categories: Path, Cluster, Path/cluster, and Ring. The spread and numbers of substructure fragment membership for each category is determined by molecular connectivity. The main difference between these two types of indices is that only the valence electrons involved in skeletal bonding (sigma orbitals) are counted for the Chi indices whereas all the valence electrons are counted in the ChiV indices to take into account electron configuration of the atom. The kappa indices are molecular shape indices based on the assumption that the shape of a molecule is a function of the number of atoms and their bonding relationship (without considering hydrogen atoms). The values are derived from counts of one-bond (Kappa 1), two-bond (Kappa 2), and three-bond (Kappa 3) fragments, each count being relative to fragment counts in reference structures which possess a maximum and minimum value for that number of atoms. Therefore, the Kappa 1 shows the degree of complexity of a binding pattern. The Kappa 2 indicates the degree of linearity or star-likeness of bonding patterns. The Kappa 3 indicates the degree of branching at the center of a molecule. More details about topological indices can be found elsewhere.⁵⁵

4.1.1.3 Electronic Properties

Electronic properties of molecule can be described by a wide variety of different parameters such as the Hammett electronic constant (σ), the partial atomic charges, dipole moments (μ) and frontier molecular orbital energies.

The Hammett electronic constant (σ) was the first parameter used to describe electronic effects. However, it could account for only substituents on an aromatic ring. This disadvantage limits its use. Therefore, many new electronic parameters have been applied in the QSAR study, such as the partial atomic charges which are electronic charges of each atom in a molecule and are important descriptors for drug design.⁵⁶ The partial atomic charges can be calculated by quantum chemical wave functions.^{57,58} Wave functions either can be obtained using ab initio methods depending on the requested accuracy of the wave function and also on the available computational resources.⁵⁹ The molecular dipole moment is the dipole moment of the molecule taken as a whole. Dipole moment was calculated by using partial charge

information. The molecular dipole moment is a good indicator of the overall polarity of a molecule. Its value is equal to the vector sum of the individual bond dipole moments. This vector sum reflects both magnitude and direction of each individual bond dipole moment.⁶⁰ The individual bond dipole moment (μ), is defined as equation (4.5):

$$\mu = 4.8 \times \delta \times d \quad \dots\dots(4.5)$$

where 4.8 represent the charge on an electron, δ is the amount of charge separation on the two atoms, and d is the bond length.

The electronic properties of various substituents clearly have an effect on a drug's ionization or polarity. This in turn may have an effect on how easily a drug can pass through cell membranes or how strongly it can bind to a receptor. Moreover, parameters derived from quantum chemical calculation, e.g., orbital energies and partial atomic charges are important electronic descriptor in the frontier molecular orbitals. Two specific frontier molecular orbitals of particular interest in drug-receptor interaction are the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The HOMO energy is roughly related to the ionization potential of a molecule, while the LUMO energy is related to the electron affinity. The magnitudes of these quantities are measures of the overall susceptibility of the molecule to losing a pair of electrons to an electrophile or accepting a pair of electrons from a nucleophile.

4.12 Statistical Analysis for QSAR Study

After the desired physicochemical properties were calculated, the next step is to find relation with the biological activity in a quantitative manner. For this purpose, a statistical analysis is needed. The regression analysis is one of the most frequently used statistical analyses to find a correlation equation. The general form of multiple linear regression (MLR) models is depicted in equation (4.6). The assumption in regression analysis is that independent X variables, e.g., physicochemical properties, can be measured or determined more precisely than the dependent Y variables, e.g., biological activity. This is usually hold true for the recent QSAR studies because most of physicochemical properties can be calculated at a very high accuracy, hence with relatively much smaller error than that of the biological data, especially if the biological response is from an *in vivo* assay.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_mx_m \quad \dots(4.6)$$

where y = dependent variable

x_1, x_2, x_3, \dots = independent variables

$\beta_0, \beta_1, \beta_2, \dots$ = regression coefficients

One of the goals in QSAR studies is the ability to describe a biological activity of a compound from its physicochemical properties, it is important to achieve this ability by using a statistical analysis method that can minimize an error between actual and calculated biological values (ε). Therefore, a least-squares method, which has a strategy to minimize the residual sum of squares (sum of squares of the errors), is usually employed. Considering the simplest linear regression equation (equations 4.7 and 4.8) of n chemical compounds, a model with only one X variable, the regression coefficients could be evaluated as following.

$$y_{\text{obs}} = \beta_0 + \beta_1x + \varepsilon \quad \dots(4.7)$$

$$y_{\text{cal}} = \beta_0 + \beta_1x \quad \dots(4.8)$$

First, since the $\Sigma\varepsilon^2 = \Sigma\Delta^2 = \Sigma(y_{\text{obs}} - y_{\text{cal}})^2$ shall be a minimum, the derivative of the function $f = \Sigma(y_{\text{obs}} - \beta_0 - \beta_1x)^2$ with respect to β_0 and β_1 are set to zero, i.e., $df/d\beta_0 = df/d\beta_1 = 0$ (equations 4.9 and 4.10).

$$df/d\beta_0 = 2 \cdot \Sigma(y - \beta_0 - \beta_1x) \cdot (-1) = 0 \quad \dots(4.9)$$

$$df/d\beta_1 = 2 \cdot \Sigma(y - \beta_0 - \beta_1x) \cdot (-x) = 0 \quad \dots(4.10)$$

Second, the so-called normal equations (4.11) and (4.12) are then resulted from equations (4.9) and (4.10).

$$\Sigma y = n\beta_0 + \beta_1\Sigma x \quad \dots(4.11)$$

$$\Sigma(xy) = \beta_0\Sigma x + \beta_1\Sigma(x)^2 \quad \dots(4.12)$$

Finally, the regression coefficients β_0 and β_1 (equations 4.13 and 4.14) could be obtained by mathematical solving of the equations (4.11) and (4.12).

$$\beta_1 = \frac{n\Sigma(xy) - (\Sigma x) \cdot (\Sigma y)}{n\Sigma(x^2) - (\Sigma x)^2} \quad \dots(4.15)$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} \quad \dots(4.16)$$

where \bar{y} = mean of y variable

\bar{x} = mean of x variable

For the multiple linear regression equation (equation 4.6), a model with more than one X variables, the regression coefficients could be evaluated in the same manner. At this point it is necessary to have some indicators to justify the significance and quality of the correlation equations. The first indicator is the standard deviation, s , which is based on variance. It is defined as a sum of squared errors (SSE) per degree of freedom (DF) in a calculation (equation 4.15). The DF is calculated from $n-k-1$, where n is the number of compounds and k is the number of variables used in the equation. The lower the s value, the better is the regression model. This is because SSE is the variation that could not be explained by the regression equation. Another variation, sum of squared regression (SSR), is the variation that could be explained by the regression equation. Summation of the above two variances gives the total variance (sum of squared total, SST). The calculations of these three values are illustrated in equations (4.15) to (4.18).

$$s = \text{SSE} / (n-k-1) \quad \dots(4.15)$$

$$\text{SSE} = \sum \varepsilon^2 = \sum (y_{\text{observe}} - y_{\text{calculate}})^2 \quad \dots(4.16)$$

$$\text{SSR} = \sum (y_{\text{calculate}} - y_{\text{mean}})^2 \quad \dots(4.17)$$

$$\text{SST} = \text{SSR} + \text{SSE} = \sum (y_{\text{observe}} - y_{\text{mean}})^2 = \sum y^2 - (\sum y)^2/n \quad \dots(4.18)$$

The second and most popular indicator used to measure the quality of the QSAR model is the Pearson correlation coefficient, r (equation 4.19). The r statistics has a value between -1 and 1 ($-1 \leq r \leq 1$), where $r = 1$ implies a perfect positive correlation, $r = -1$ implies a perfect negative correlation, and $r = 0$ implies no correlation. Therefore, a value of r close to 1 or -1 indicates a strong degree of linear relationship.

$$r = \frac{\sum(xy) - (\sum x)(\sum y)/n}{\sqrt{\sum(x^2) - (\sum x)^2/n} \cdot \sqrt{\sum(y^2) - (\sum y)^2/n}} \quad \dots(4.19)$$

Generally, r^2 is used instead of the r itself, thus $0 \leq r^2 \leq 1$. The r^2 statistics is a ratio of the variance explained by the regression model to the total variance (equation 4.20). Therefore, it gives an information on how many percentage of the variation in

the biological activity (Y variable) can be explained by the physicochemical properties (X variables) presented in the equation. For example, in case of the sum of squared error (SSE, $\Sigma\epsilon^2$) goes to zero, r^2 goes to 1. Then, the equation can explain all 100% of the variation in the biological activity.

$$\begin{aligned} r^2 &= \frac{\Sigma(y_{\text{calculate}} - y_{\text{mean}})^2}{\Sigma(y_{\text{observe}} - y_{\text{mean}})^2} \\ &= \text{SSR} / \text{SST} = 1 - \text{SSE} / \text{SST} \end{aligned} \quad \dots(4.20)$$

The third indicator is the F value, which measures the level of statistical significance of the regression model. The F value can be calculated from the equation (4.21). In this case, the number of variables being included to derive the model has a stronger influence than that of the standard deviation. Only F values being larger than the 95 % significance limit prove the overall significance of a regression equation. With the same n and k values, the higher the F value is, the higher is the overall significance level of the model.

$$F = \frac{r^2 \cdot (n - k - 1)}{k \cdot (1 - r^2)} \quad \dots(4.21)$$

In general, the regression equation can be accepted in QSAR studies if the following four criteria are met. Firstly, the correlation coefficient r is around or better than 0.8 ($r^2 \geq 0.64$) for *in vivo* data or 0.9 ($r^2 \geq 0.81$) for *in vitro* data.⁶² Secondly, the standard deviation s is not much larger than the standard deviation of the biological data. Thirdly, the overall significance level is better than 95 % as indicated by the F value. Fourthly, the confidence intervals of all individual regression coefficients prove that they are justified at the 95 % significance level, i.e., their confidence intervals are smaller than the absolute values of the regression coefficients. In addition, there should be no fewer than five compounds for each chemical descriptor used in the final equation ($n > 5k$) to prevent the chance correlations. Moreover, the descriptors should not be intercorrelated, i.e., interdescriptor correlation coefficients should be less than 0.6.⁶³

Using the r^2 alone to justify the QSAR model is not recommended. But the predictability of the model should also be considered. This is because the r^2 gives information only on the reproducibility, how well the model reproduces the biological

activity of the compounds included in the model, not the predictability. The predictability, an ability to predict a biological activity of a new compound outside the model, could be measured by various approaches, e.g., cross-validation,⁶⁴ bootstrapping, random change of the values of the dependent variable, and dividing the original set into training and testing sets. However, the most widely used method is the cross-validation. In this method, the predictability of the model is estimated by repeatedly leaving out one (or more) compound(s) at a time until each compound is excluded exactly once. Using the reduced set of data, the model is derived and is used to predict the activity of the left out compound. During the cross-validation test, the sum of the squared prediction errors called the predictive residual sum of squares (PRESS), the cross-validated correlation coefficient (r_{cv}^2 or q^2), and the cross-validated standard error of estimate (s_{cv}) are evaluated. These values are calculated in the same manner as SSE, r^2 , and s , respectively (shown in equations 4.22 to 4.24). A smaller s_{cv} and a larger q^2 indicate the model's good predictability. Generally, a model with the q^2 value of greater than 0.50 is accepted as a good model.⁶⁵

$$\text{PRESS} = \sum (y_{\text{observed}} - y_{\text{predicted}})^2 \quad \dots\dots(4.22)$$

$$q^2 = 1 - \text{PRESS}/\text{SST} \quad \dots\dots(4.23)$$

$$s_{cv} = (\text{PRESS}/n)^{1/2} \quad \dots\dots(4.24)$$

The main goal for QSAR study is the ability to predict biological activity of other compounds outside the model rather than the ability to reproduce the biological activity of the compounds included in the model. Therefore, we should test the model by predicting the activity of the “new compound”, which is not included in the process of deriving the model. Therefore, the real predictive ability of the model could not be determined by the q^2 value. In order to investigate the real predictive ability, the compounds are randomly divided into 2 sets, training set and testing set. Compounds in the training set are used to derive the model. Subsequently, the obtained model is used to predict the biological activity of compounds in the testing set. By comparing between predicted and actual values, the real predictive power is obtained.

The overall steps of QSAR analysis in this study are summarized in Figure 4.1.

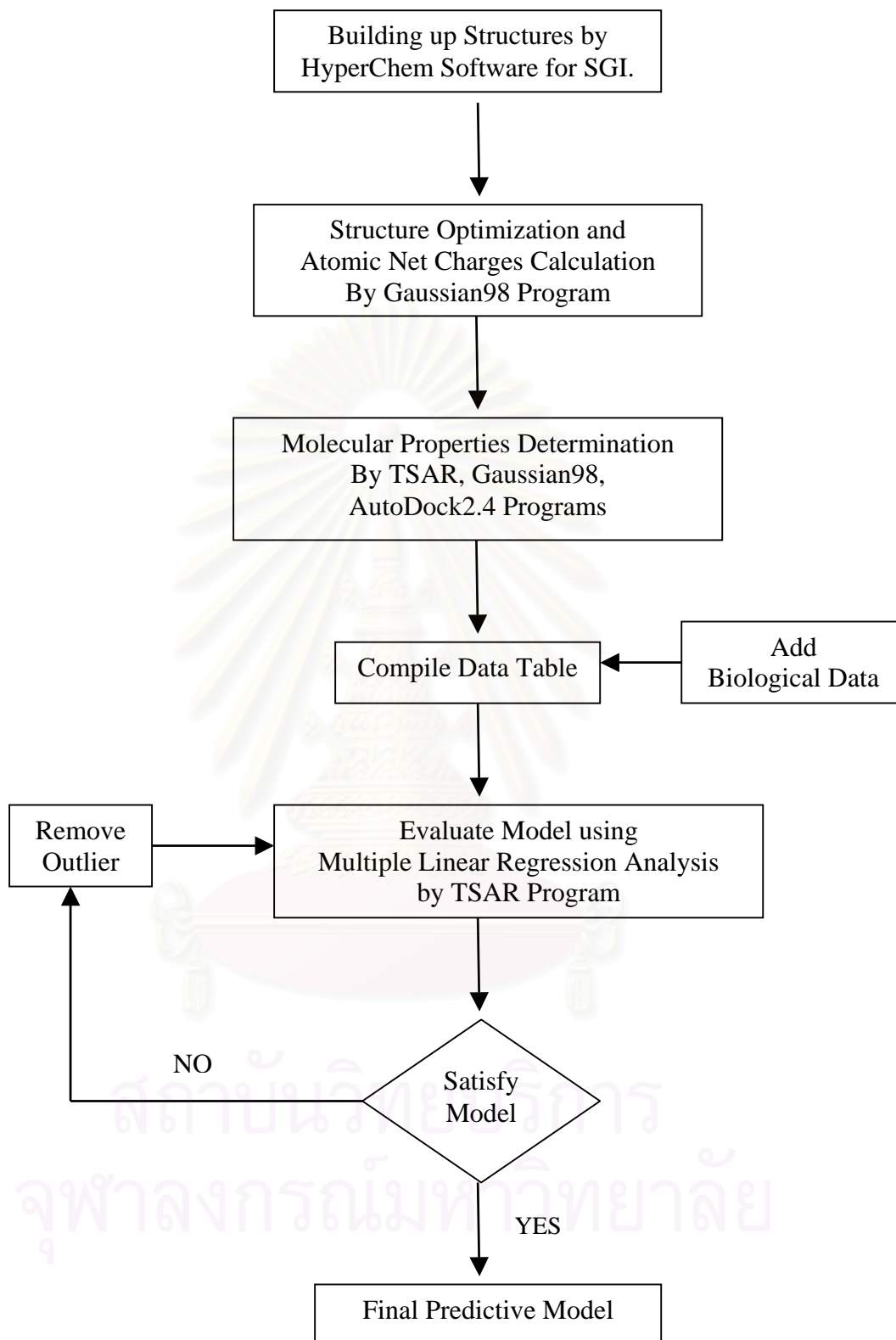


Figure 4.1 Flow chart of the QSAR methodology in this study.

4.2 Calculations of Properties

There are a lot of parameters that can be used in the field of QSAR study. Totally 58 physicochemical parameters were calculated using TSAR⁵⁷ and Gaussian 98 softwares.⁴⁶ These parameters can be grouped into three classes: hydrophobic, steric and electronic properties.

4.2.1 Hydrophobicity properties

For the hydrophobicity parameter, the log P was calculated using the TSAR software.

4.2.2. Steric properties

Structural parameters, 9 bond lengths (R), 10 bond angles (A), and 12 torsion angles (T), were taken from the HF/3-21G optimized structures. In order to represent these parameters, the atom number corresponding to the structure of tricyclic 1,2,4-trioxane in Figure 3.1 (Chapter 3) was given in the parenthesis. For example, the R(1-2) parameter represents the bond length between atom 1 and 2, the A(1-2-3) means the bond angle between atom 1, 2, and 3, and the T(1-2-3-4) is the torsion angle between atom 1, 2, 3, and 4. All structural parameters are as follows – R(1-2), R(2-3), R(3-4), R(4-5), R(3-13), R(12-13), R(8a-12a), R(8-8a), R(12a-12), A(4-5-5a), A(3-4-5), A(13-3-4), A(12-13-3), A(8a-12a-5a), A(2-3-4), T(4-5-5a-12a), T(3-4-5-5a), T(13-3-4-5), T(12-13-3-4), T(2-3-4-5), T(8-8a-12a-5a), T(1-2-3-13), T(3-2-1-12a), T(1-12a-12-13), T(12a-12-13-3), T(12-13-3-2) and T(4-3-2-1).

The molar refractivity (MR) was calculated using the TSAR software. Topological index after the Balaban method⁶⁶ and the following 6 connectivity indices were calculated using the TSAR software—Chi0 (atoms), ChiV0 (atoms), Chi1 (bonds), ChiV1 (bonds), Chi2 (path), and ChiV2 (path). In addition, three shape indices, i.e., Kappa 1, Kappa 2, and Kappa 3, were also computed.

4.2.3. Electronic properties

For the electronic parameters, atomic charges obtained by the Mulliken Population Analysis (MPA) method in the Gaussian98 software were used. Atomic charges of 13 atoms namely O₁, O₂, C₃, C₄, C₅, C_{5a}, C₆, C₇, C₈, C_{8a}, C₁₂, C_{12a} and O₁₃ were computed at the HF/6-31G(d) level.

Dipole moment, HOMO energy, and LUMO energy were also calculated at the HF/6-31G(d) level.

4.3 Results and Discussions

All the calculated physicochemical properties were related to the antimalarial activity by the multiple linear regression analysis using the stepwise procedure. The methodology of the stepwise method is to start with the best single variable to build the model and then add further significant variables, according to their contribution to the model. During the process, there is a proof whether already introduced variables are no longer significant at a later stage. If it is, this variable is excluded from the equation. The adding and proofing process continues until a static model is reached.

In order to access the real predictive ability, 32 compounds were divided into the training set (85%) and testing set (15%). Therefore, four compounds were randomly selected for the testing set, i.e., compound number 15, 20, 23 and 28 and the remaining 28 compounds, the training set, were used to derive the model. Four parameters were statistically selected into the model (equation 4.25).

$$\text{Activity} = - 0.434 * A(13-3-4) - 0.114 * T(1-2-3-13) + 4.632 * (C_{8a} \text{ charge}) + 0.504 * (C_{12} \text{ charge}) + 64.518 \quad \text{----- (4.25)}$$

$$n = 28, r^2 = 0.735, q^2 = 0.470, S = 0.235, F = 15.921$$

QSAR model in equation (4.27), has a correlation coefficient (r^2) 0.735 and a cross-validated correlation coefficient (q^2) 0.470. Both values are lower than the acceptable value of 0.81 and 0.50, respectively.^{62,65} Thus, the improvement of the q^2 and r^2 values could be attained by omitting some compounds with high residual values, i.e., compounds 22, 2 and 11 (Table 4.1). The best QSAR model was then obtained as shown in equation (4.26).

$$\text{Activity} = - 0.397 * A(13-3-4) - 0.138 * T(1-2-3-13) + 4.211 * (C_{8a} \text{ charge}) + 0.644 * (C_{12} \text{ charge}) + 61.998 \quad \text{----- (4.26)}$$

$$n = 25, r^2 = 0.832, q^2 = 0.697, S = 0.174, F = 24.776$$

The $r^2(0.832)$ and $q^2(0.697)$ values of equation (4.26) are quite good, i.e., both values are higher than the acceptable values.^{62,65} The correlation coefficients between each pair of variables were calculated (see Table 4.1) and the C_{8a} charge parameter

was found to have the highest relationship with activity. Moreover, all the interdescriptor correlation coefficients are less than 0.6, which indicates no intercorrelation between each pair of variables in the model, hence, model 2 (equation 4.26) is an acceptable model.⁶³

Table 4.1 Predicted activities and residuals of 28 compounds in the training set by the QSAR model (eg. 4.25).

Compound No.	Actual Activity	Predicted Activity	Residual
1	6.718	6.543	0.175
2	6.796	7.207	-0.411
3	6.959	7.166	-0.207
4	6.998	6.936	0.062
5	7.090	7.328	-0.238
6	7.182	7.229	-0.047
7	7.198	7.175	0.023
8	7.215	7.339	-0.124
9	7.279	7.305	-0.026
10	7.342	7.328	0.014
11	7.391	7.009	0.382
12	7.445	7.273	0.172
13	7.561	7.328	0.233
14	7.621	7.328	0.293
16	6.847	6.984	-0.137
17	6.939	7.043	-0.104
18	7.140	7.044	0.096
19	7.370	7.503	-0.133
21	7.526	7.270	0.256
22	6.445	6.989	-0.544
24	7.658	7.391	0.267
25	7.445	7.611	-0.166
26	7.538	7.670	-0.132
27	7.790	7.686	0.104
29	7.895	7.882	0.013
30	8.050	7.898	0.152
31	8.097	7.910	0.187
32	8.081	8.245	-0.164

Table 4.2 The correlation coefficients between each pair of variables.

	Activity	A(13-3-4)	T(1-2-3-13)	C _{8a} charge	C ₁₂ charge
Activity	1	-0.226	-0.147	0.722	0.294
A(13-3-4)	-0.226	1	-0.066	0.237	0.249
T(1-2-3-13)	-0.147	-0.066	1	0.067	0.534
C _{8a} charge	0.722	0.237	0.067	1	0.360
C ₁₂ charge	0.294	0.249	0.534	0.360	1

In QSAR study, the activity depends on all the variables presented in the model, therefore, using only one variable to predict the activity is not appropriate and it may not give a proper activity value. However, the analysis on each individual variable alone (assuming that the other variables are constant) could give useful information in a qualitative manner, which is very helpful for the design of new more effective drugs. Therefore, we performed the analysis for all parameters in this study.

The O₁₃-C₃-C₄ angle, A(13-3-4) (see Figure 3.1) is involved in the structural change during the C₃-C₄ bond breaking step in the reaction mechanisms (see Figure 6.1 in Chapter 6). Therefore, it has relationship with the activity and is presented in the model. Compound with smaller A(13-3-4) has higher activity as indicated by a minus sign in the equation. It can be seen that compounds having substituent group at the C₃ position (R₁), e.g., compounds 1, 4, 7, 12 and 13, the smaller of the angle A(13-3-4), the higher of the activity (see table 4.3). This may be explained by the facts that compound with smaller angle has more strain in the ring systems and so it is easier to achieve this structural change.

Table 4.3 Relationship between the angle of O₁₃-C₃-C₄, A(13-3-4) and activity

Compound	A(13-3-4)	Activity
1	112.1 ⁰	6.718
4	110.6 ⁰	6.998
7	110.0 ⁰	7.198
12	109.8 ⁰	7.445
13	109.7 ⁰	7.531

The torsion angle T(1-2-3-13) is involve to substituted carbon atoms (R_1 and R_2) and oxygen atoms of peroxide linkage. A minus sign of this torsion parameter refers that compound with smaller T(1-2-3-13) value has higher activity. This is because compounds with smaller T(1-2-3-13) value is easier to proceed the structural change during the reaction mechanism in which the oxygen-free radical is moved away from the ring after the attack of Fe at oxygen atom of the peroxide linkage. For example, compounds 2, 6, 8, 28, 30 and 32 have torsion angle T(1-2-3-13) 73.8^0 , 73.6^0 , 73.4^0 , 72.8^0 , 72.3^0 and 71.4^0 and activity 6.796, 7.182, 7.215, 7.822, 8.050 and 8.081, respectively, it is clearly that the larger of torsion angle T(1-2-3-13) the lower of the activity (see table 4.4).

Table 4.4 Relationship between the torsion angle T(1-2-3-13) and activity.

Compound	T(1-2-3-13)	Activity
2	73.8^0	6.796
6	73.6^0	7.182
8	73.4^0	7.215
28	72.8^0	7.822
30	72.3^0	8.050
32	71.4^0	8.081

C_{8a} atomic charge has the highest correlation coefficient ($r = 0.722$) as shown in Table 4.2. Therefore, this parameter has more contribution to activity than others that described in the model but rather to compounds 25-32 only because such compounds have substituents at C_{8a} position (R_3). Atomic charges of C_{8a} in those compounds show significantly different from the rest, e.g. compounds 25-32 have C_{8a} atomic charge in the range of -0.164 to -0.179, while compound 1-24 have C_{8a} atomic charge in the range of -0.318 to -0.335. A positive coefficient of C_{8a} parameter in the model indicates that compound with less negative C_{8a} are design to have higher activity. This is confirmed by our results as shown in the table 4.5, i.e. compound 31 has the less negative charge and hence the highest activity.

Table 4.5 Relationship between the C_{8a} atomic charge and activity.

Compound	C _{8a} charge	Activity
23	-0.318	6.575
25	-0.177	7.445
28	-0.175	7.822
32	-0.172	8.081
31	-0.164	8.097

From Appendix A, it is clear that the substituent R₃ influence the charge of O-atom of peroxide linkage, these compounds with R₃ at C_{8a} position has O₁ charge more negative than O₂ charge whereas the compounds without substituent at C_{8a} position show the opposite results. This suggest that compounds 25-32 should follow the pathway B of the reaction mechanism with Fe²⁺ bind to O₂ and follow by C₃-C₄ bond cleavage to form C₄ radical (see in Figure 6.1). This results also agree well with our the docking results (Chapter 6).

Similarly, C₁₂ atomic charge will describe better for compounds having substituents at C₁₂ than others. A plus sign in the model implies that compound with more positive C₁₂ charge will have higher antimalarial potency. As shown in table 4.6 compound 31 has the highest positive charge on C₁₂ atom and therefore the greatest activity. Having substituents at C₁₂ position (R₄) also leads to a different charge of carbon atoms in ring C of compounds 16-21, i.e., C₆ (-0.336 to -0.354), C₇ (-0.316 to -0.318), and C₈ (-0.336 to -0.344) compare to compounds without substituents at C₁₂ position (R₄) which have C₆ (-0.317 to -0.319), C₇ (-0.319 to -0.324), and C₈ (-0.318 to -0.330). This indicates that ring C is important feature of structure requirements in the reaction mechanism for the inhibition.

Table 4.6 Relationship between the C₁₂ atomic charge and activity.

Compound	C ₁₂ charge	Activity
4	0.574	6.998
9	0.575	7.279
11	0.576	7.391
26	0.579	7.538
29	0.580	7.895
31	0.584	8.097

The comparison between actual and predicted activity values of 25 compounds in the training set is given in Table 4.7. The predictions are good with the residual values in the range of only 0.001-0.269.

Table 4.7 Predicted activities and residuals of 25 compounds in the training set by the best QSAR model (eg. 4.26).

Compound No.	Actual Activity	Predicted Activity	Residual
1	6.718	6.714	0.004
3	6.959	7.208	-0.249
4	6.998	6.994	0.004
5	7.090	7.352	-0.262
6	7.182	7.259	-0.077
7	7.198	7.205	-0.007
8	7.215	7.367	-0.152
9	7.279	7.324	-0.045
10	7.342	7.352	-0.010
12	7.445	7.299	0.146
13	7.561	7.352	0.209
14	7.621	7.352	0.269
16	6.847	6.958	-0.111
17	6.939	7.015	-0.076
18	7.140	7.016	0.124
19	7.370	7.567	-0.197
21	7.526	7.272	0.254
24	7.658	7.468	0.190
25	7.445	7.588	-0.143
26	7.538	7.662	-0.124
27	7.790	7.674	0.116
29	7.895	7.894	0.001
30	8.050	7.911	0.139
31	8.097	7.920	0.177
32	8.081	8.259	-0.178

The relationships between actual and predicted values of activities for 25 compounds in the training set using the best model is shown in Figure 4.2.

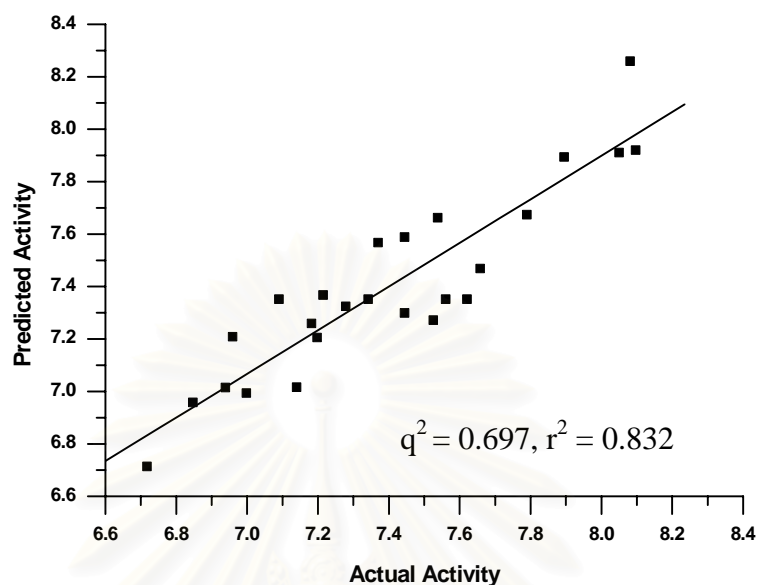


Figure 4.2 Comparison between actual and predicted activities for 25 compounds in the training set by the best QSAR model.

The comparison between actual and predicted values of 4 compounds in the testing set is shown in Table 4.8. The residual values indicated that the model is satisfied based on predictive ability.

Table 4.8 Predicted activities and residuals of 4 compounds in the test set by the best QSAR model.

Compound No.	Actual Activity	Predicted Activity	Residual
15	7.746	7.382	0.364
20	7.411	7.314	0.097
23	6.575	7.484	-0.909
28	7.822	7.870	-0.048

The comparison between actual and predicted activities of 4 compounds in the testing set is displayed in Figure 4.3.

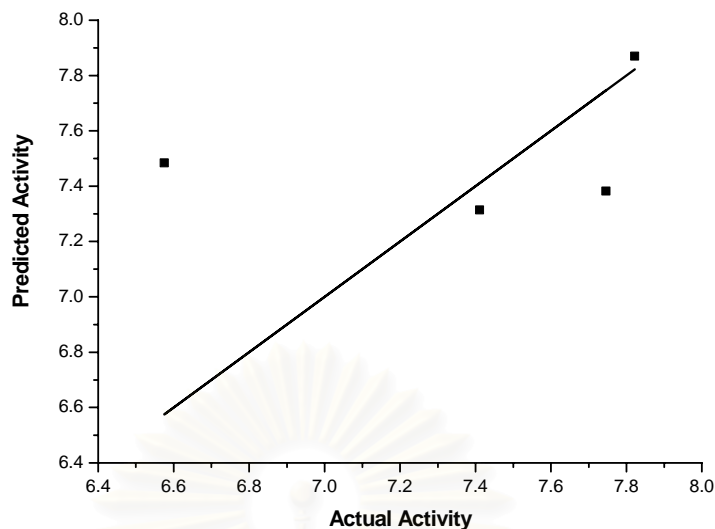


Figure 4.3 Comparison between actual and predicted activities for 4 compounds in the test set by the best QSAR model.

4.4 QSAR Summary

The relationship between antimalarial activities and molecular properties of tricyclic 1,2,4-trioxane compounds were investigated by QSAR technique. From the statistical analysis obtained for 28 compounds in training set, both $r^2(0.735)$ and $q^2(0.470)$ values are below the acceptable value. Statistically better model could be attained by omitting some compounds with high residual values, i.e., compounds 22, 2, 11 were omitted. The best QSAR model for 25 compounds was obtained with $r^2(0.832)$ and $q^2(0.697)$. Four parameters are the angle A(13-3-4), torsion angle T(1-2-3-13), C_{8a} and C_{12} atomic charges were statistically selected into the model and all the interdescriptor correlation coefficients are less than 0.6, which indicates no intercorrelation between each pair of variables in the model, hence, the model is acceptable model. The best QSAR model were obtained indicates compound with smaller angle A(13-3-4), smaller T(1-2-3-13), more positive C_{8a} and C_{12} charges values have higher activities.

Finally, the real predictive ability of the best QSAR model was judged by comparison between actual and predicted activities of compounds in the testing set. The obtained models can predict the activities very close to the experimental values. thus confirming their reliability.

CHAPTER 5

THREE DIMENSIONAL QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (3D-QSAR)

In classical QSAR study, most of physicochemical properties used do not represent three dimensional structure of compound directly. In reality, stereochemistry of a drug play an important role to its biological activity because drug's binding site is a chiral environment and performs discrimination between different enantiomers of an optically active drug. Drug must have a three dimensional structure complementary to its binding site to exert high affinity with the target which may result in high activity. Therefore, three-dimensional quantitative structure activity (3D-QSAR) approaches are very useful and plays an important role in drug design and development process. Several approaches to 3D-QSAR have been developed, e.g., Comparative Molecular Field Analysis (CoMFA).⁶⁷ These techniques are usually employed in drug discovery to find the common features that are important in binding to the biologically relevant target. They are based on the assumption that changes in binding affinities of ligands are related to changes in different fields surrounding the molecules. The 3D-QSAR models are usually generated by multivariate statistics using the Partial Least Squares (PLS) analysis⁶⁸ and can be used for predicting the binding affinity of new molecules.

In this study, the CoMFA technique was selected for 3D-QSAR study of tricyclic 1,2,4-trioxane analogues.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.1 Theoretical Background of CoMFA

Comparative molecular field analysis (CoMFA) is one of the most powerful 3D-QSAR techniques providing further insight into relationships between structure and function of drug. This methodology is based on assumption that non-covalent forces dominate receptor-drug interactions and that these forces can be described in terms of steric and electrostatic fields. The changes in biological activities of binding affinities of sample compounds correlate with changes in steric and electrostatic fields around these molecules. Practically, there are 4 main steps in CoMFA study: alignment, interaction energy calculation, statistical analysis, and interpretation. Detail for each steps are discussed as following.

5.1.1 Alignment Rule

An alignment of molecules is the most important input variable in CoMFA study since relative interaction energies depend strongly on relative molecular positions. The rationale for proper alignment is that differences in field values at each lattice point should reflect differences in structure only, not chance variations on model geometry. Three general methods used in molecular alignment are 1) *alignment based on pharmacophore*, this method assumes that all molecules that have activity at the same target must present their pharmacophoric groups in the same configuration in space. Then, one can search the conformational space of all molecules to identify a convergent distance map which defines the pharmacophore geometry common to all molecule. This method is particularly useful when the molecules to be analyzed represent a diverse set of structures rather than homologous series. 2) *Minimization of root mean square (rms) distance between specified pairs of atoms*, when aligned with this method, field differences are due to differences in functionality than to conformational variations. 3) *Field fit*, this method aligns a molecule so as to minimize the differences between its field values at a lattice point and those of some template field.

5.1.2 Interaction Energy Calculation

To calculate the electrostatic interaction, partial atomic charges were obtained from GAUSSIAN98. A grid spacing of 2 Å was used to generate a cubic lattice around

all molecules based on the molecular volume of the structures. Three dimensions ensured that the grid extended beyond the molecular size by 4 Å in all directions. Molecular interactions between probe atom and aligned molecules were then calculated. A sp³ carbon probe atom was placed at each lattice point, and the interactions of the steric and electrostatic fields with each atom in molecule were all calculated with CoMFA standard scaling and then put in a CoMFA QSAR table. The minimum sigma value was set to 2.0 kcal/mol and energy cutoff values of 30 kcal/mol were selected for both electrostatic and steric fields.

Steric Field

Steric interactions are calculated from the van der Waals potential functions using the Lennard-Jones 12-6 function, which can be described in the form below:

$$E_{\text{van der Waals}} = \sum_{i,j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

where A_{ij} is a coefficient depicting repulsive heteroatomic interaction with hydrogen (($A_i A_j$)^{1/2}).

B_{ij} is a coefficient depicting attractive heteroatomic interaction with hydrogen (($B_i B_j$)^{1/2}).

r_{ij} is a distance between atom i of drug molecule and probe atom j (Å).

Electrostatic Field

Electrostatic interactions are calculated from the Coulomb potential using a charge probe atom. Electrostatic properties of molecules are typically described by point charges at the center of atoms. The general form of electrostatic interaction between two molecules is given by

$$E_{ele} = \sum_{i,j} \left[\frac{q_i q_j}{r_{ij}} \right]$$

where q_i , q_j are atomic net charges of atom i of drug molecule and of probe

atom j, respectively.

r_{ij} is a distance between atom i of drug molecule and probe atom j (Å).

5.1.3 Statistical Analysis using Partial Least-Squares Method

A typical CoMFA data table usually contains hundreds or thousands of columns of interaction energy values and a number of compound included in the study is relatively much smaller than the number of the energy columns. Thus, a mathematical difficulty arises because of a large number of descriptors. For this reason, the multiple linear regression technique can not be used directly without the danger of chance correlation. Partial Least Square (PLS) was developed for such problems.⁶⁸ Not only it can be applied to solve an equation having hundreds or thousands of variables while involving only a small number of biological data, but also it can be simultaneously handled to derive models. In this latter case, some missing values can be handled without any problem. PLS is an iterative procedure which applies two criteria to produce its solution. First, for each iteration to generate a better set of coefficients (extract a new component), the criterion is to maximize the degree of commonality between all of the structural parameter columns collectively and the experimental data. Second, during the evaluation phase of the PLS iteration, the criterion for acceptance of the principal component just generated (and subsequent components as well) is the incremental improvement in the ability to predict, not to reproduce, the biological data.

The technique used in PLS to assess predictive ability of a QSAR model is cross-validation. Cross-validation,⁶⁹ also known as leave one out, rests on the intuitively satisfying, though more computationally demanding, concept that the best way to assess predictive performance is to do some predictions. The process of cross-validation is shown in Figure 5.1. In cross-validating, one pretends that one or more of the known experimental values is in fact unknown. The analysis being cross-validated is repeated, excluding the temporarily unknown compounds from the very beginning. The resulting equation is used to predict the experimental measurement for the omitted compounds, and the resulting individual squared errors of the prediction are accumulated. The cross-validation is the sum of the squared prediction errors, called the PRESS (Predictive

Residual Sum of Squares). For evaluation of the overall analysis, the PRESS is commonly expressed as a cross-validation r^2 (q^2) value.

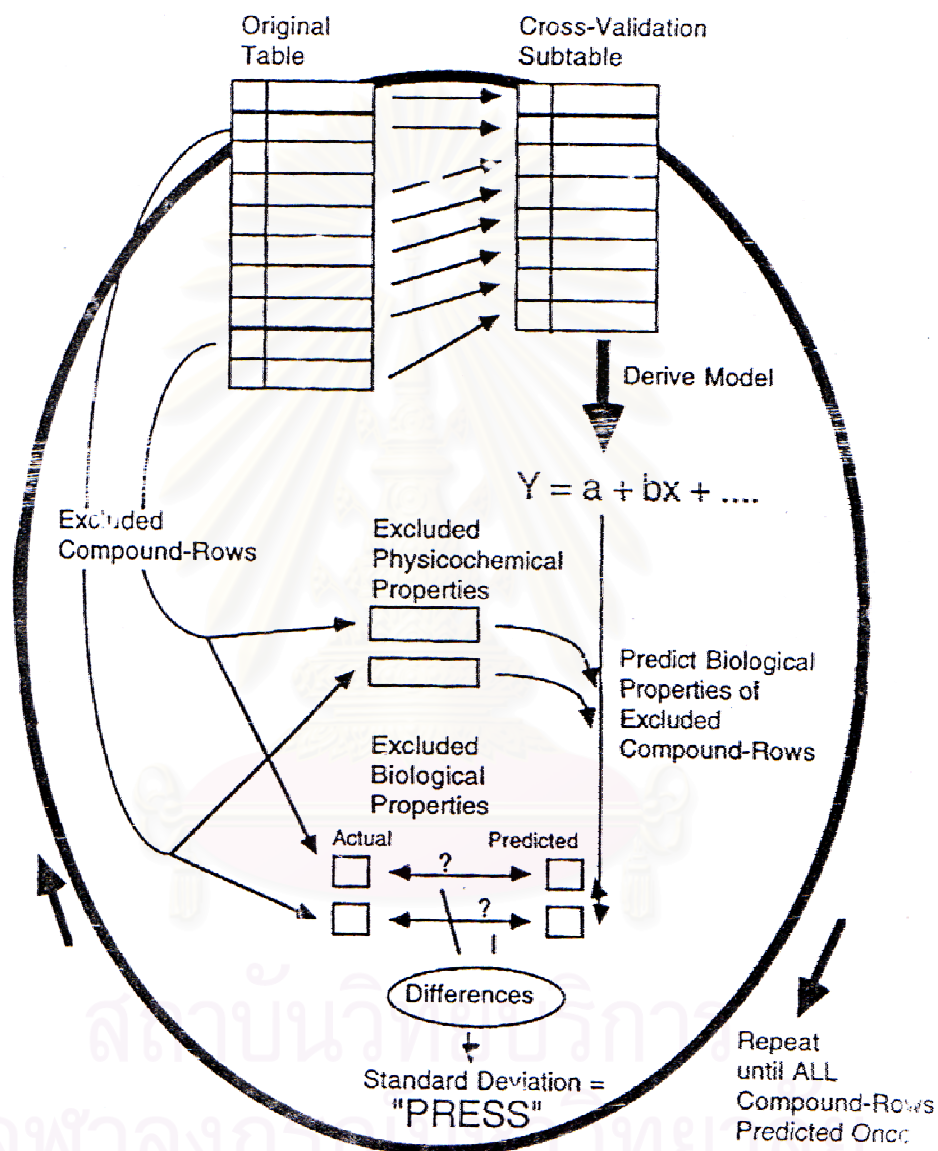


Figure 5.1 Cross-validate procedure.⁶⁹

The general form of the equation is;

$$\text{Activity} = C + \sum \sum c_{ij} S_{ij}$$

Where S_{ij} is the column in the matrix that corresponds to placing probe group j at grid point i , c_{ij} is the coefficient for each column, and C is the sum of the deviations of the observation from the fitted equation.

The PLS method expresses a dependent variable (y) in term of linear combination of the original independent (x) variables as the following

$$y = b_1 t_1 + b_2 t_2 + b_3 t_3 + \dots b_m t_m$$

$$\text{where } t_1 = c_{11} X_1 + c_{12} X_2 + c_{1p} X_p$$

$$t_2 = c_{21} X_1 + c_{22} X_2 + c_{2p} X_p$$

$$t_m = c_{m1} X_1 + c_{m2} X_2 + c_{mp} X_p$$

where t_1, t_2 etc. are called latent variables (or components) and are constructed in such a way that they form an orthogonal set. The maximum number of latent variables (m) is the smaller value between the number of x values or the number of molecules. However, there is an optimum number of latent variables in the model beyond which the predictive ability of the model does not increase. A PLS model is often evaluated according to its ability to predict the activity of compounds not used to derive the model. PLS generates iteratively one component at a time by maximizing the degree of commonality between all of the descriptor variables collectively and the biological data. The process stops when the requested number of components is extracted. The number of significant PLS component (latent variables) is determined by cross-validation test. Usually a model with less than the maximum number of components gives a better cross-validated sum of squared residuals than the full model. The model with optimal number of components has a higher standard error of fit than a full-rank model, but it is generally more reliable for

prediction. Thus the cross-validated PLS model is usually less subject to error of over-specification than the regression model.

Predictive Ability of the Model

The predictive ability of the model is tested by a cross-validation method. A cross-validation procedure must be used to select the model having the highest predictive ability. The predictive ability of the obtained model is expressed in term of r_{cv}^2 or q^2 , which is define as:

$$q^2 = \frac{(SSY - PRESS)}{SSY}$$

where SSY is the sum of square of deviation between the affinities of the fitted set and their mean affinity:

$$SSY = \Sigma(Y_{actual} - Y_{mean})^2$$

and $PRESS$ is the prediction error sum of squares obtained from the leave-one-out method:

$$PRESS = \Sigma(Y_{pred} - Y_{actual})^2$$

where Y_{actual} , Y_{mean} and Y_{pred} are actual, mean and predicted values of the target property, respectively. The standard of error of prediction (S_{PRESS}) is also considered and defined as:

$$S_{PRESS} = \left[\frac{PRESS}{n - k - 1} \right]^{1/2}$$

where n is the number of compounds used in the study, k is the number of variables in the model.

The real predictive ability of each analysis is determined form a set of compounds not included in the training set. These molecules are aligned, and their activities are

predicted by each PLS analysis. The predictive $r^2(r^2_{pred})$ value is based on molecules of the testing set only and is defined as:

$$r^2_{pred} = \frac{(SD - PRESS)}{SD}$$

Where SD is the sum of the squared deviations between biological activities of the test set and mean activity of the training set molecules and $PRESS$ is the sum of squared deviation between predicted and actual activity values for every molecule in the test set.

The optimum number of component is extracted from a previous cross-validated PLS analysis by examining the incremental change in q^2 with each additional components corresponding to the minimum S_{PRESS} is used for next non-cross-validated PLS analysis. PLS analyses with non-cross-validation are then run with the optimum number of components of each alignment to derive the final QSAR model and corresponding conventional r^2 and s .

5.1.4 Interpretation of CoMFA Results

Results of CoMFA are equations showing the contribution of energy fields at each lattice point. In order to facilitate the interpretation of the results, it is also displayed as coefficients (or standard deviation time coefficient or $stdev \cdot coeff$) contour plot showing the regions in space where specific molecular properties increase or decrease the potency. The coloring is standardized as followings:

- The contours are colored in green and yellow for positive and negative steric effect, respectively. Positive steric contours show the regions where substituents increase the biological potency if occupied and the negative steric contours show the area where substituents decrease the potency.

- The contours are colored in blue and red for positive and negative electrostatic effect, respectively. The positive electrostatic contours indicate the region where positive charges increase the potency, whereas the negative electrostatic contours display the region where negative charge increase the potency.

The CoMFA process is summarized in Figure 5.2.

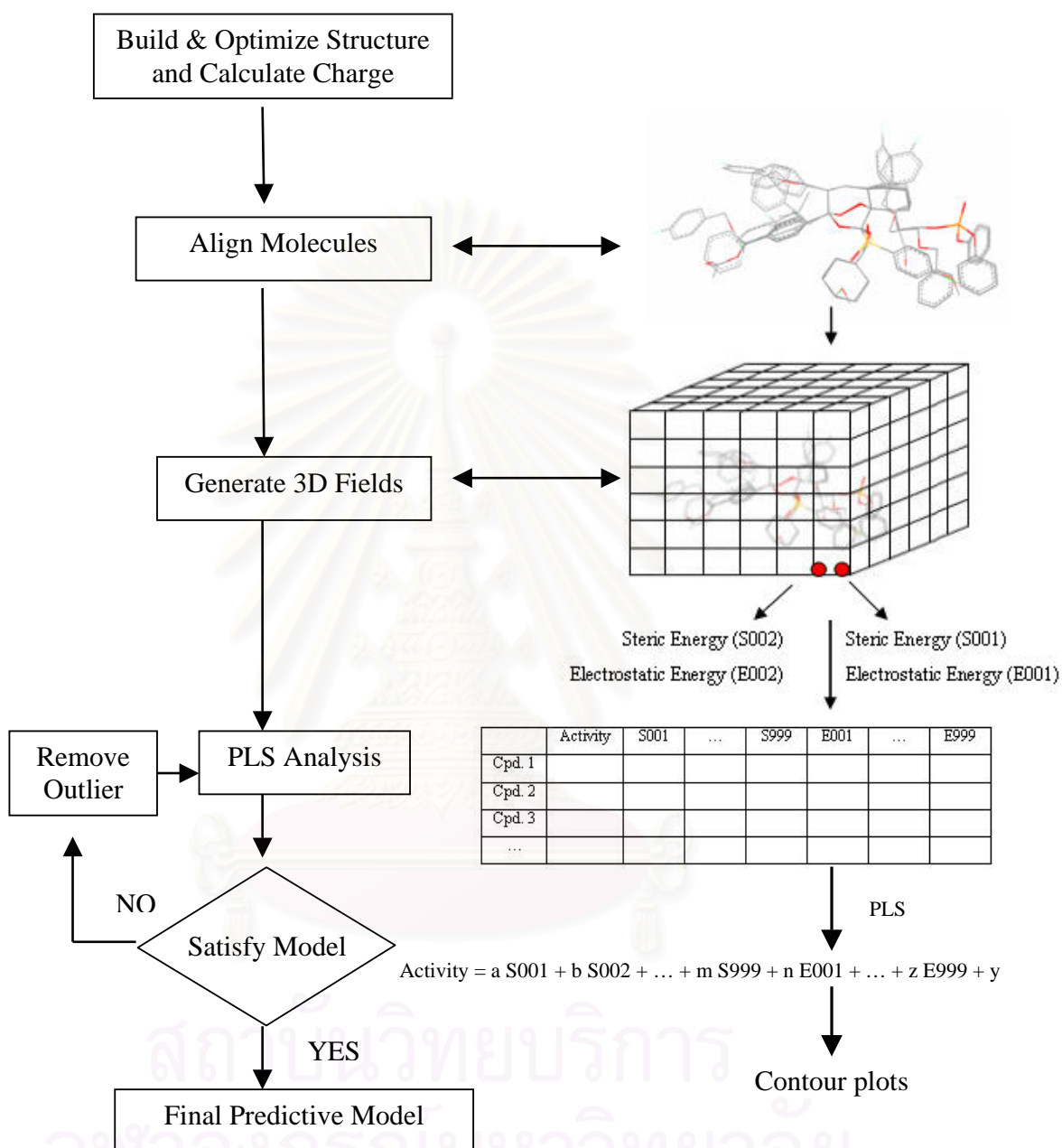


Figure 5.2 Flow chart of CoMFA methodology in this study.

5.2 Computational Methods

5.2.1 Conformation and Alignment Rule

The bioactive conformation of drug molecule, the conformation when it is bound to its target receptor, is usually used in the CoMFA study. However, this conformation is not known in our compounds. Therefore, two types of conformation were considered instead: 1) the conformation obtained from HF/3-21G optimization and 2) the conformation corresponding to the best docking configuration to heme (see chapter 6). For the first conformation type, the structures of all 32 compounds were optimized at HF/3-21G level and their atomic charges were subsequently assigned at the HF/6-31G(d) level using the Mulliken Population Analysis (MPA) method. The second conformation type was obtained by performing flexible docking calculations between each compound and heme (see details in Chapter 6). Fifteen alignment rules were used to study the influence of different alignments. For all alignments, the structures were adjusted using the “Fit Atom” option in the SYBYL⁷⁰ which minimizes the root-mean-square (RMS) differences of selected atoms to the ones of the reference molecule. The most active compound 31 was used as the template for the alignment (see Figure 5.3).

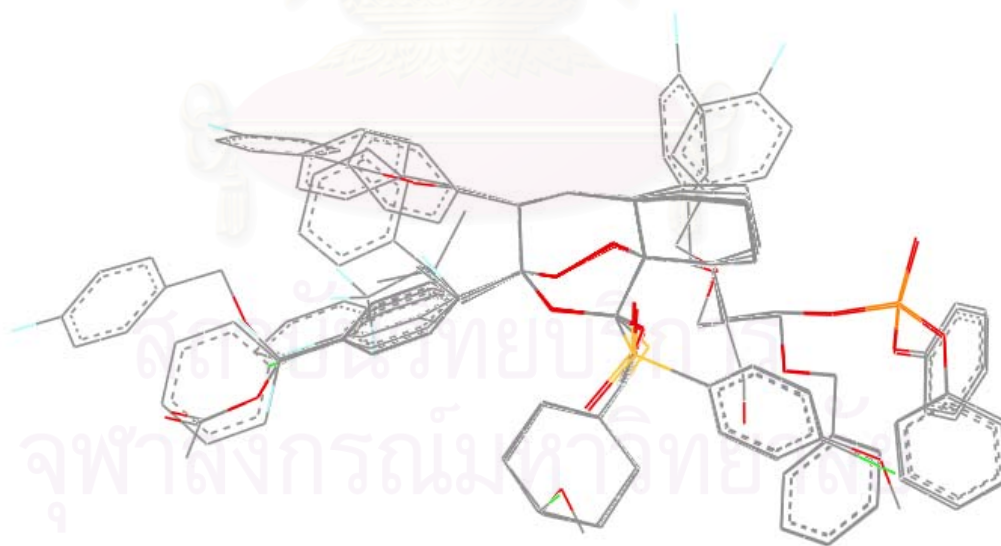
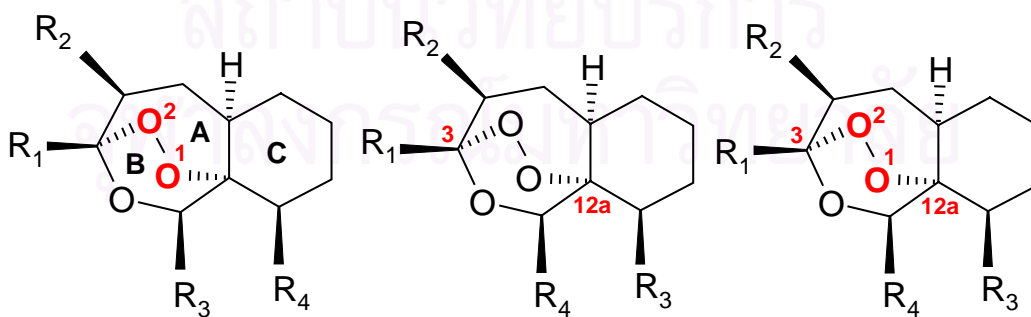


Figure 5.3. Superimposition of all tricyclic 1,2,4-trioxane compounds using the alignment rule 13 and compound 31 as template (hydrogen atoms were omitted for clarity).

The selected atoms for the definition of alignment rules are shown in Table 5.1 and Figure 5.4.

Table 5.1 Atoms selected for the definition of alignment rules.

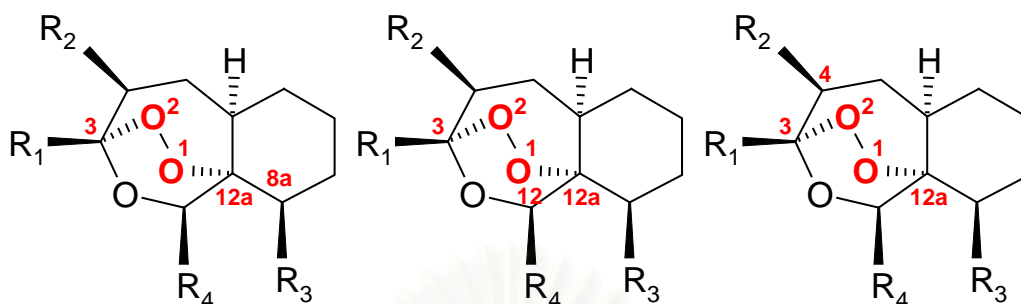
Alignment No.	Selected Atoms
1	O ₁ -O ₂
2	C _{12a} -C ₃
3	C _{12a} -O ₁ -O ₂ -C ₃
4	C _{8a} -C _{12a} -O ₁ -O ₂ -C ₃
5	C ₁₂ -C _{12a} -O ₁ -O ₂ -C ₃
6	C _{12a} -O ₁ -O ₂ -C ₃ -C ₄
7	C _{12a} -O ₁ -O ₂ -C ₃ -C ₄ -C ₅ -C _{5a}
8	C _{12a} -O ₁ -O ₂ -C ₃ -O ₁₃ -C ₁₂
9	O ₁ -O ₂ -C ₃ -C ₄ -C ₅ -C _{5a} -C _{12a} -C ₁₂ -O ₁₃
10	C _{5a} -C ₆ -C ₇ -C ₈ -C _{8a} -C _{12a}
11	O ₁ -O ₂ -C ₃ -C ₄ -C ₅ -C _{5a} -C ₆ -C ₇ -C ₈ -C _{8a} -C _{12a} -C ₁₂ -O ₁₃
12	C _{8a} -C _{12a} -C ₁₂ -O ₁₃ -C ₃ -C ₄
13	O ₁ -O ₂ -C ₃ -C ₄ -C _{8a} -C ₁₂
14	O ₁ -O ₂ -O ₁₃
15	O ₁ -O ₂ -C ₃ -C ₄ -C ₅ -C _{5a} -C _{12a} -C ₁₂ -O ₁₃ -C _{8a}



Alignment 1

Alignment 2

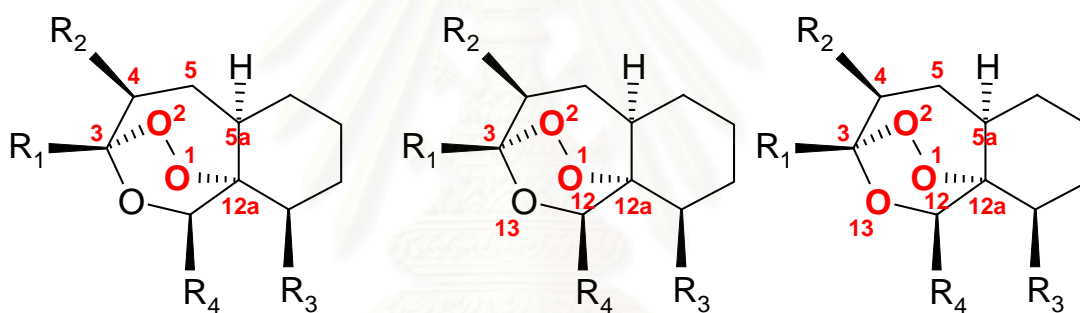
Alignment 3



Alignment 4

Alignment 5

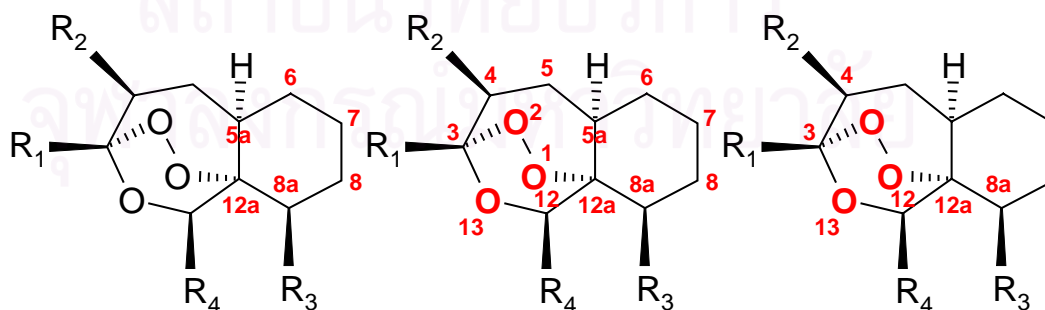
Alignment 6



Alignment 7

Alignment 8

Alignment 9



Alignment 10

Alignment 11

Alignment 12

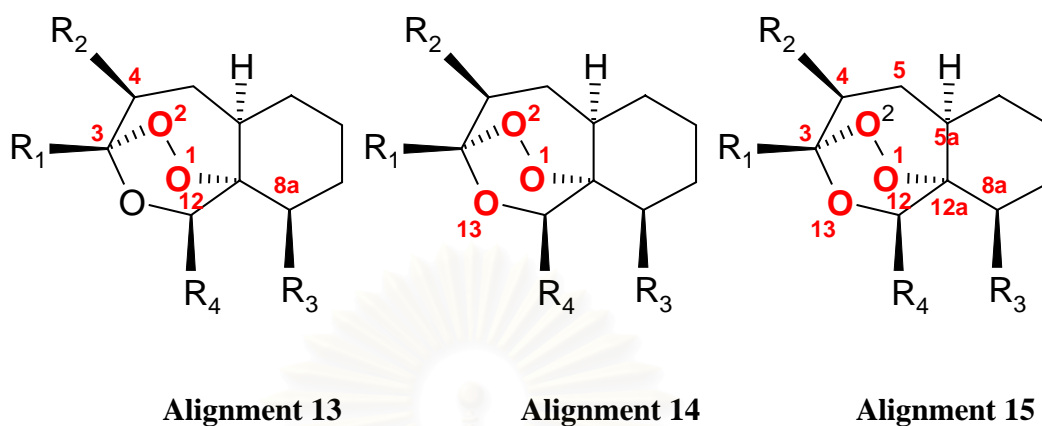


Figure 5.4 Definition of 15 alignment rules used in CoMFA studies.

5.2.2 CoMFA Calculations

A regular three-dimensional lattice with 2 Å spacing was created extending beyond molecular dimension of the largest molecule by 4.0 Å in all directions. Twelve types of probe atoms were used, +1.0 sp^3 carbon, +1.0 sp^2 carbon, +1.0 sp carbon, +1.0 ar carbon, -1.0 sp^3 oxygen, -1.0 sp^2 oxygen, +1.0 sp^3 nitrogen, +1.0 sp^2 nitrogen, +1.0 sp nitrogen, +1.0 ar nitrogen, +1.0 am nitrogen, and +1.0 hydrogen. The steric (Lennard-Jones 12,6 function) and electrostatic (Coulombic) interactions were calculated using the Tripos force field with a distance-dependent dielectric constant. The cut-off was set to 30 kcal/mol for both fields. The AutoCoMFA column, which uses all the default setting values, was additionally calculated.

5.2.3 Partial Least Squares Regression Analysis

All models were investigated using the full cross-validated partial least squares method (leave-one-out) with CoMFA standard options for scaling of variables. Minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve a signal-to-noise ratio by omitting those lattice points whose energy variation is below this threshold. To avoid an excessive number of components, the optimal number of components (onc) was selected as the one which results in an increase of the q^2 of more than 5% compared to

the model with fewer components. Subsequently, it was used for the PLS without cross-validation to derive the r^2 statistics.

5.3 CoMFA Results and Discussions

5.3.1 Effect of Conformation and Alignment Rule

The CoMFA results of different conformations and the alignment rules for tricyclic 1,2,4-trioxanes are given in Table 5.2. From these results, it is clear that the conformation from HF/3-21G optimization give better predictive models than the those from the docking calculations for all alignments. The structural orientation of all 32 compounds obtained from docking calculations are quite different from each other while the optimized structures have almost the same conformation. The poor superimpose to the template due to variation of orientations in definitely effect to the derived model and yield a bad statistical results. Therefore, we considered the effect of alignment rules only on the conformation from HF/3-21G optimization. All models have r^2 acceptable values, therefore only q^2 are discussed. Alignments 1 and 2 ($q^2 = 0.333$) give the same results. The alignment number 3 give better q^2 statistics (0.385) than alignments 1 and 2. Alignments 4, 5, 6 give lower q^2 values than alignment 3. Considering alignments 7-11, the better q^2 values of alignments 8, 10 compared to 7, 9, 11 indicate that rings B and C of tricyclic 1,2,4-trioxanes have more contribution to activity than ring A. For alignments 12-15, it was found that alignment 13 give higher q^2 value than other alignments. In addition, alignment 13 give the highest value among all 15 alignments used in this study. Therefore, alignment 13 was selected for further investigations.

Table 5.2 CoMFA results of different conformation and alignment rules.

Alignment No.	HF/3-21G Optimized Structure			Structure from Docking Calculations		
	q ²	onc	r ²	q ²	onc	r ²
1	0.333	5	0.975	0.188	5	0.945
2	0.333	5	0.975	0.188	5	0.945
3	0.385	5	0.942	0.197	5	0.947
4	0.362	5	0.940	0.195	5	0.947
5	0.362	5	0.940	0.195	5	0.947
6	0.376	5	0.940	0.218	6	0.976
7	0.370	5	0.939	0.199	5	0.946
8	0.394	5	0.940	0.203	5	0.947
9	0.378	5	0.941	0.200	5	0.946
10	0.395	5	0.963	0.225	6	0.986
11	0.363	5	0.939	0.204	5	0.968
12	0.386	5	0.907	0.188	5	0.945
13	0.401	5	0.931	0.286	5	0.966
14	0.388	5	0.940	0.233	5	0.948
15	0.356	5	0.940	0.212	5	0.947

5.3.2 Effect of Type of Probe Atom

Both steric and electrostatic fields are determined from the interaction energies between molecules and the selected probe atom. The type and charge of probe atom are significant to CoMFA results. Twelve types of probe atoms were selected, +1.0 sp³ carbon, +1.0 sp² carbon, +1.0 sp carbon, +1.0 ar (aromatic) carbon, -1.0 sp³ oxygen, -1.0 sp² oxygen, +1.0 sp³ nitrogen, +1.0 sp² nitrogen, +1.0 sp nitrogen, +1.0 ar (aromatic) nitrogen, +1.0 am (amine) nitrogen and +1.0 hydrogen. Comparing among these twelve probe atom types, there is no significant difference (see Table 5.3). Therefore, the default setting for type and charge of probe atom, C-sp³ with +1.0 charge were selected for further studies.

Table 5.3 Effect of Type of Probe Atom.

Type	Charge	q^2	onc	r^2
C-sp ³	+1.0	0.401	5	0.931
C-sp ²	+1.0	0.401	5	0.931
C-sp	+1.0	0.401	5	0.931
C-ar	+1.0	0.401	5	0.931
O-sp ³	-1.0	0.398	5	0.939
O-sp ²	-1.0	0.398	5	0.939
N-sp ³	+1.0	0.407	5	0.939
N-sp ²	+1.0	0.407	5	0.939
N-sp	+1.0	0.407	5	0.939
N-ar	+1.0	0.407	5	0.939
N-am	+1.0	0.407	5	0.939
H	+1.0	0.403	5	0.937

5.3.3 Effect of Steric and Electrostatic Cut-offs

The steric and electrostatic cut-offs are used to filter unimportant data before the statistical analysis step. Therefore, their effect on the CoMFA results was investigated by varying their values to 10 and 30 kcal/mol. The results are shown in Table 5.4., By setting the electrostatic cutoff to 30 kcal/mol and varying the steric cutoff to 10 and 30 kcal/mol, the resulting q^2 values are equal (0.401) but the r^2 values are slightly different (0.923 vs. 0.931). Therefore, the steric cut-off does not have significant effect. On the other hand, changing the electrostatic cut-off from 30 to 10 kcal/mol while keeping the steric cut-off constant at 30 kcal/mol; both q^2 and r^2 values were decreased. Therefore, steric and electrostatic cut-off values 30 kcal/mol were selected for further studies.

Table 5.4 CoMFA results with different steric and electrostatic cut-offs.

Steric Cut-off (kcal/mol)	Electrostatic Cut-off (kcal/mol)	q^2	onc	r^2
30	30	0.401	5	0.931
10	30	0.401	5	0.923
30	10	0.375	5	0.927
10	10	0.385	5	0.913

5.3.4 CoMFA results

In order to access the real predictive ability of the model, all 32 compounds were divided into 2 sets, i.e., training set and testing set. Compounds 15, 20, 23 and 28 were randomly selected for the testing set. The remaining 28 compounds were used as the training set. The CoMFA results of the training set are illustrated in Table 5.5.

Table 5.5 CoMFA results of all 28 compounds.

Model	q^2	onc	S_{PRESS}	r^2	S	F	Contribution Fraction	
							Steric	Elec.
1	0.490	6	0.342	0.972	0.080	121.506	0.591	0.409

From Table 5.5, although the CoMFA model has good statistical result (high r^2 value) but its predictive power ($q^2 = 0.490$) is lower than the acceptable value ($q^2 = 0.500$). Since the q^2 value is sensitive to a compound with high residual value, an improvement of the q^2 value could be attained by omitting such compounds, as usually done in all CoMFA studies. Therefore, compounds with high residual, compounds 6, 18 and 26 were omitted (see Table 5.6). Subsequently, the new model was evaluated and the results are given in Table 5.7. The predictive power (q^2) was improved to 0.524 with the onc of 4. A contribution ratio between steric and electrostatic fields of 3:2 indicate the importance of steric field.

Table 5.6 Predicted activities and residuals of 28 compounds in the training set.

Compound No.	Actual Activity	Predicted Activity	Residual
1	6.718	6.667	0.051
2	6.796	6.881	-0.085
3	6.959	6.959	0.000
4	6.998	6.982	0.016
5	7.090	7.094	-0.004
6	7.182	7.314	-0.132
7	7.198	7.302	-0.104
8	7.215	7.216	-0.001
9	7.279	7.243	0.036
10	7.342	7.293	0.049
11	7.391	7.429	-0.038
12	7.445	7.343	0.102
13	7.561	7.489	0.072
14	7.621	7.576	0.045
16	6.847	6.953	-0.106
17	6.939	6.931	0.008
18	7.140	7.028	0.112
19	7.370	7.428	-0.058
21	7.526	7.464	0.062
22	6.445	6.413	0.032
24	7.658	7.651	0.007
25	7.445	7.461	-0.016
26	7.538	7.678	-0.140
27	7.790	7.862	-0.072
29	7.895	7.815	0.080
30	8.050	8.026	0.024
31	8.097	8.005	0.092
32	8.081	8.113	-0.032

Table 5.7 CoMFA results of 25 compounds*.

Model	q ²	onc	S _{PRESS}	r ²	S	F	Contribution Fraction	
							Steric	Elec.
2	0.524	4	0.334	0.962	0.095	126.173	0.606	0.394

* compounds 6, 18, and 26 were omitted.

The predicted activities and residuals of all 25 compounds in the training set are given in Table 5.8. The plot of experimental and predicted activities of the non-cross-validated model is presented in Figure 5.6.

Table 5.8 Predicted activities and residuals of 25 compounds in the training set.

Compound No.	Actual Activity	Predicted Activity	Residual
1	6.718	6.711	0.007
2	6.796	6.855	-0.059
3	6.959	6.856	0.103
4	6.998	6.961	0.037
5	7.090	7.191	-0.101
7	7.198	7.312	-0.114
8	7.215	7.233	-0.018
9	7.279	7.208	0.071
10	7.342	7.281	0.061
11	7.391	7.360	0.031
12	7.445	7.319	0.126
13	7.561	7.557	0.004
14	7.621	7.523	0.098
16	6.847	6.956	-0.109
17	6.939	6.946	-0.007
19	7.370	7.376	-0.006
21	7.526	7.391	0.135
22	6.445	6.552	-0.107
24	7.658	7.672	-0.014
25	7.445	7.468	-0.023
27	7.790	7.854	-0.064
29	7.895	8.033	-0.138
30	8.050	7.986	0.064
31	8.097	7.957	0.140
32	8.081	8.206	-0.125

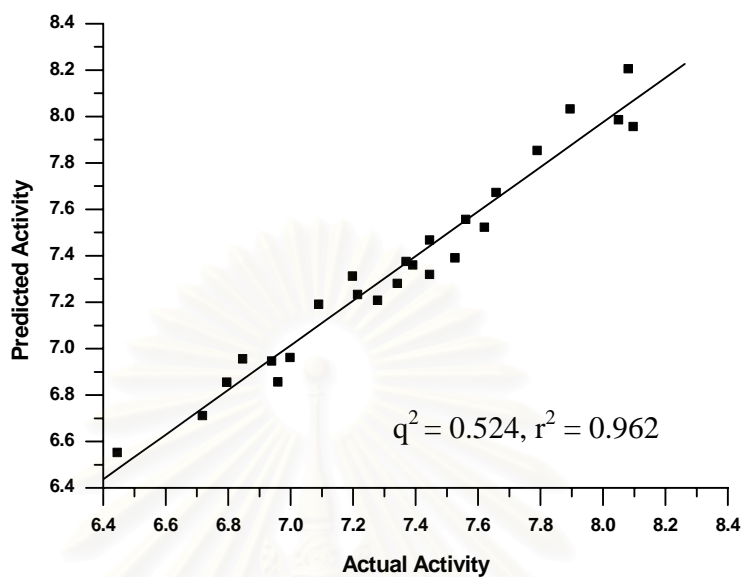


Figure 5.6 Comparison between actual and predicted activities for 25 compounds in the training set.

The predicted activities and residuals of 4 compounds in the testing set using the best CoMFA model is given in Table 5.9.

Table 5.9 Predicted activities and residuals of 4 compounds in the testing set.

Compound No.	Actual Activity	Predicted Activity	Residual
15	7.746	7.244	0.502
20	7.411	7.382	0.029
23	6.575	7.116	-0.541
28	7.822	7.452	0.370

The relation between actual and predicted activities of 4 compounds in the testing set is presented in Figure 5.7.

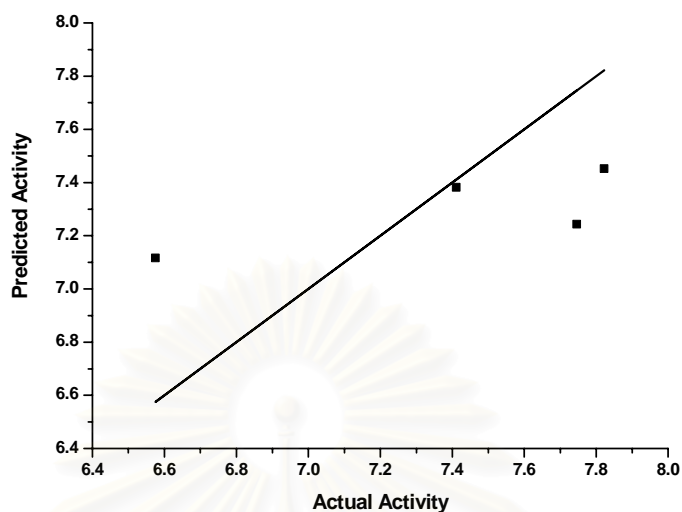


Figure 5.7 Comparison between actual and predicted activities for 4 compounds in the testing set.

Analysis of CoMFA contour maps

The CoMFA contour maps of the model with the best “predictive” r^2 (q^2), model 2, are discussed below. Contour maps were generated as scalar products of coefficients and standard deviation ($\text{stdev} \times \text{coeff}$). The CoMFA steric interactions are represented by green and yellow colored contours while electrostatic interactions are represented by red and blue colored contours. Bulky substituents are favored in green regions and disfavored in yellow regions. An increase in positive charge is favored in blue regions while an increase in negative charge is favored in red regions

The steric contour plots of our model are illustrated in Figures 5.8-5.11. A large yellow contour was located around substituent group at the C_3 position (R_1). It indicates that the steric at this position would diminish the activities. The alkyl substituent groups in compounds 2 and 3 are buried in this yellow region (see Figure 5.8A) and hence they have low activities. On the other hand, compounds 13-14 have only one carbon atom of phenyl substituent groups fall in this region (see Figure 5.8B). Thus, compounds 13-14 have higher activities than compounds 2 and 3.

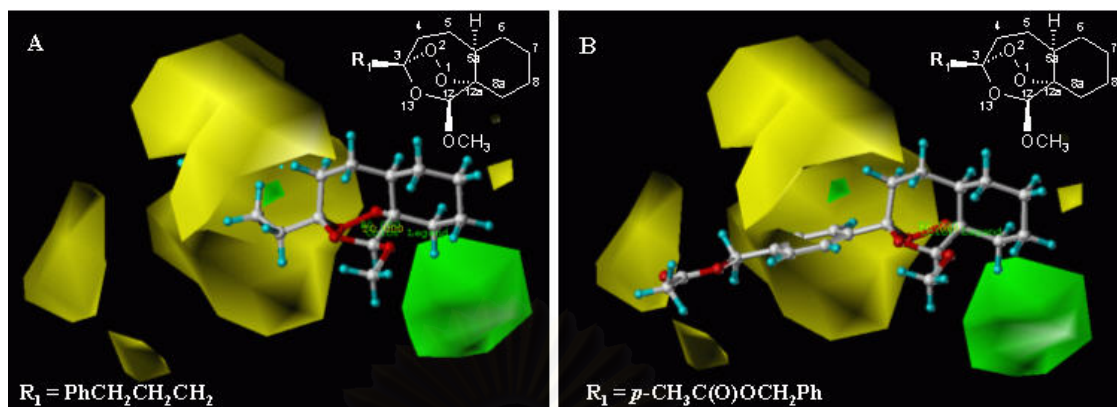


Figure 5.8 CoMFA S.D.*coeff. Steric contour maps, (A) compounds 3 and (B) compounds 14 which have substituents at C₃ position (R₁) are represented.

In Figure 5.9, there is no steric contour located on the substituent group at the C₄ position (R₂) of compounds 22-24. This is possibly because only two compounds having R₂ substituent were included in the training set of 25 compounds so their steric contribution to the model in the contour map is very small. Therefore, the CoMFA model does not represent their steric effect well.

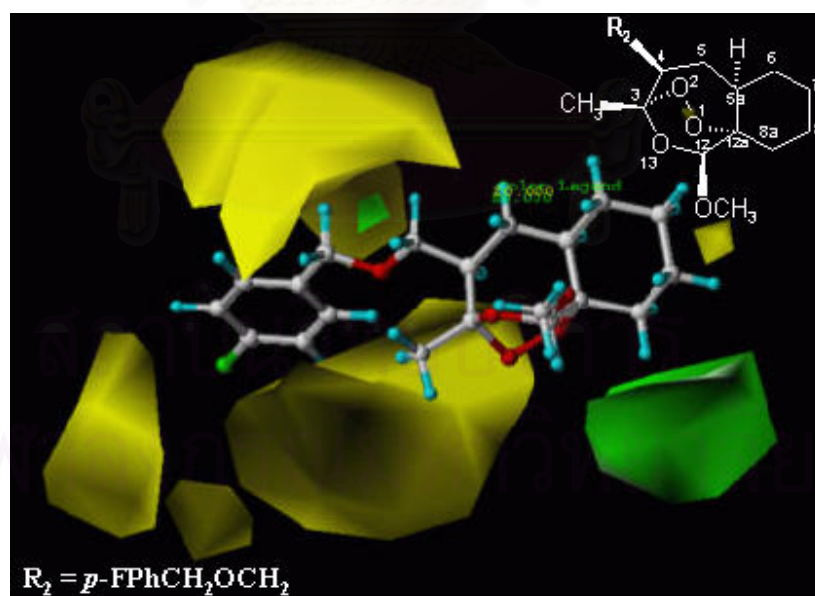


Figure 5.9 CoMFA S.D.*coeff. Steric contour maps, compound 24 which has substituent at C₄ position (R₂) is represented.

Green contour around OCH_2CH_2 group at the C_{8a} position (R_3) indicate the steric favor of substituent at this position (see Figure 5.10). This contour map suggest that the tricyclic 1,2,4-trioxanes compounds with OCH_2CH_2 substituents at the C_{8a} position would have high activities, e.g. compounds 25-32. In addition, a small green contour was found around substituent group at the C_4 position (R_2), which means that the methyl substitution at this position produced better activity than the hydrogen atom. For example, compounds with the methyl substituent at the C_4 position (R_2), compounds 30-32, showed higher activity than their unsubstituted compounds 25-27.

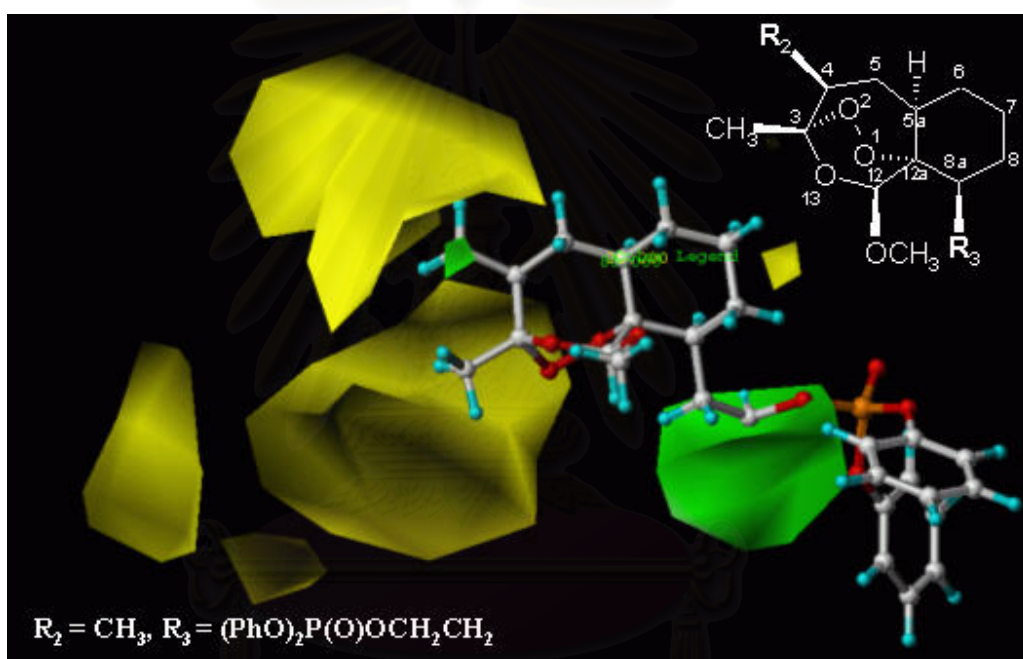


Figure 5.10 CoMFA S.D.*coeff. Steric contour maps, compound 31 which has substituent at C_{8a} position (R_3) is represented.

The green contour in the vicinity of R_4 is absent for compounds 16-21 (see Figure 5.11) although experimentally the steric occupancy with these bulky substituents increase the activities. Our CoMFA model would not describe well for compounds with sulfur atom in the substituent group.

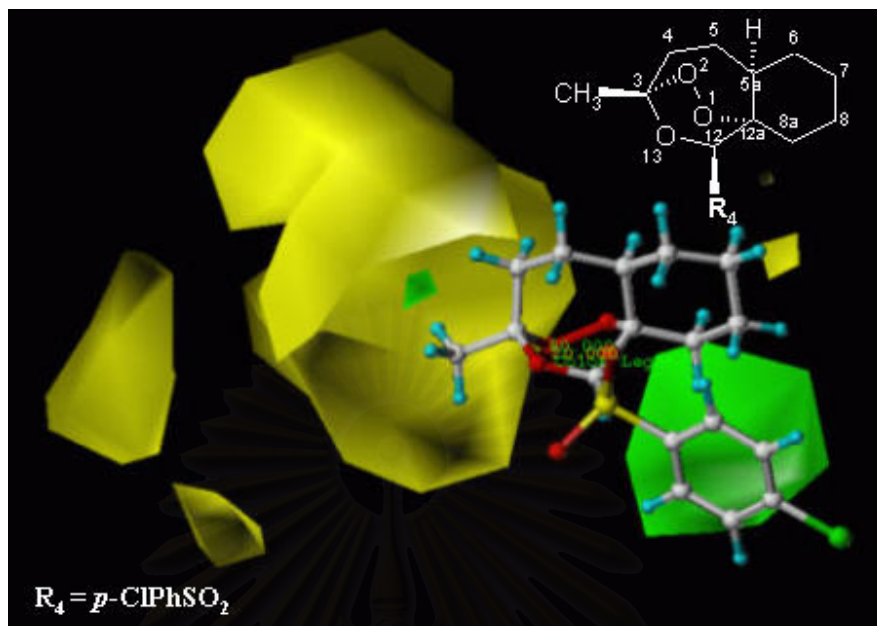


Figure 5.11 CoMFA S.D.*coeff. Steric contour maps, compound 21 which has substituent at C₁₂ position (R₄) is represented.

The electrostatic contour plots of our model are illustrated in Figures 5.12-5.15. A red contour is found around the carbon atom next to C₃ position (R₁) (see Figure 5.12), therefore, compound with negative charge at this position will have high activity. For example, compounds 2-4 have atomic charge on carbon atom next to C₃ position (R₁) of -0.338, -0.344 and -0.384, respectively so compound 4 (activity 6.998) more active than compound 3 (activity 6.959) and compound 2 (activity 6.796) accordingly.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

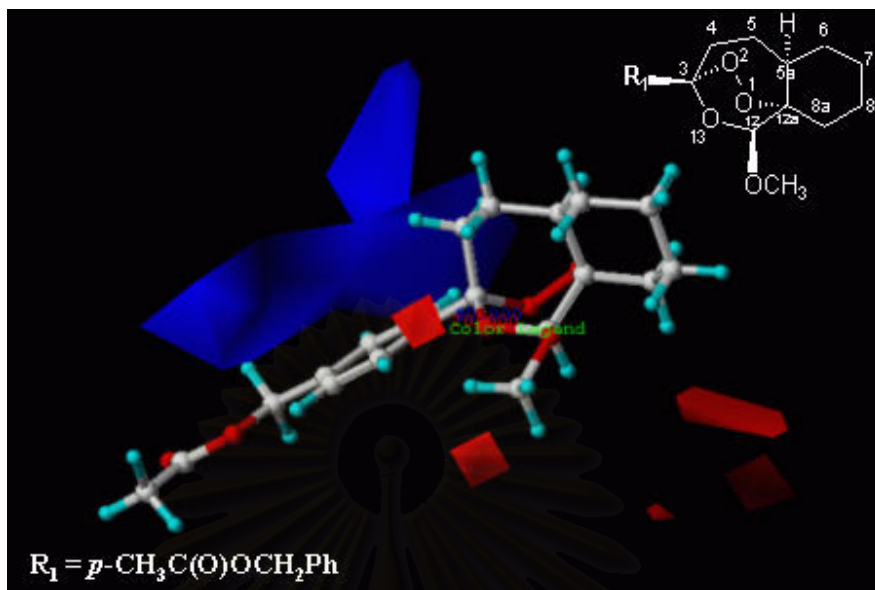


Figure 5.12 CoMFA S.D.*coeff. Electrostatic contour maps, compound 14 which have substituent at C₃ position (R₁) is represented.

A large blue contour is located around the first carbon atom of substituent group connected to the C₄ position (R₂) indicating that highly positive charge at this position is required to enhance binding affinity. As example, compounds 22-24 (see Figure 5.13), the first carbon atoms of substituent group at the C₄ position have atomic charges of -0.346, 0.000, 0.023, respectively, thus compound 24 display highly inhibition.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

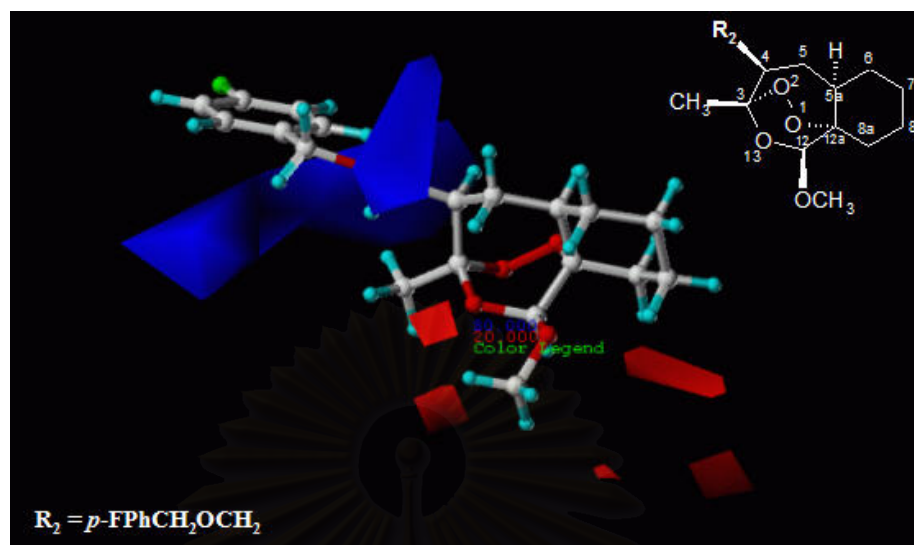


Figure 5.13 CoMFA S.D.*coeff. Electrostatic contour maps, compound 24 which has substituent at C₄ position (R₂) is represented.

A red contour is placed on carbon atom next to the C_{8a} position (R₃) (see Figure 5.14). The calculated atomic charge on carbon atom next to the C_{8a} position of compound 28 is -0.351 and of compound 31 is -0.366. The charge of compound 31 is more negative than that of compound 28 hence compound 31 shows higher activity than compound 28.

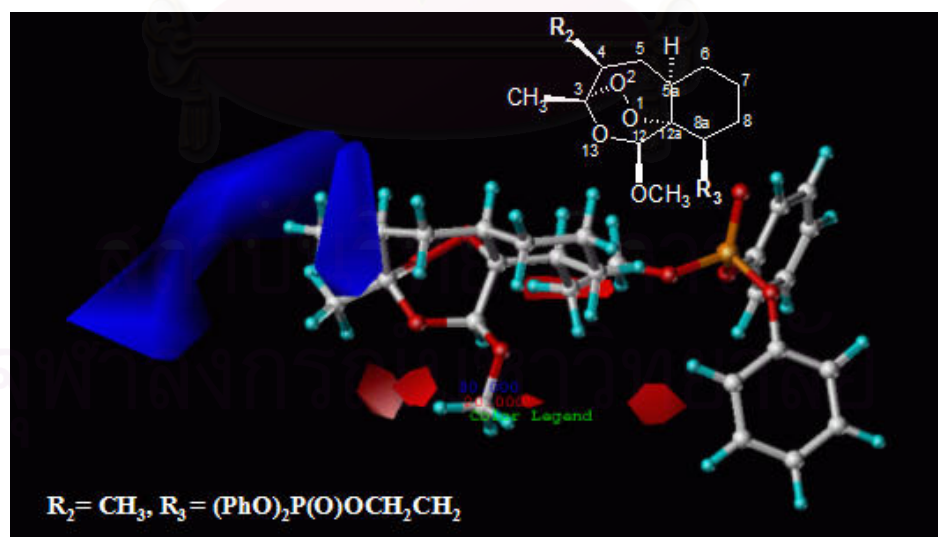


Figure 5.14 CoMFA S.D.*coeff. Electrostatic contour maps, compound 31 which has substituent at C_{8a} position (R₃) is represented.

Considering compounds 16-21 with substituent group at the C₁₂ position (R₄), no electrostatic contour was observed around the C₁₂ position (see Figure 5.15). This means that the electrostatic field contributed from substituent group of such compounds has no relationship with biological activity. It is also interesting to note that compounds 19-21 containing sulfone (SO₂) in the substituent groups (R₄) attached to the C₁₂ position display significantly malarial inhibition, about 2-6 times compared to compounds 16-18 containing sulfide (S) in the substituent groups (R₄). Our quantum chemical calculations show that the atomic charge of S in SO₂-R₄ substituent groups of compounds 19-21 are 15 times greater than that in S-R₄ substituent groups of compounds 16-18. This remark is interesting for further work.

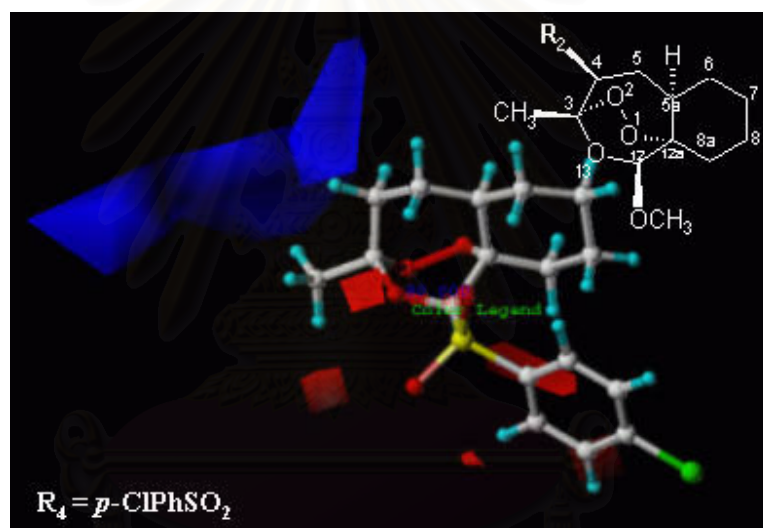


Figure 5.15 CoMFA S.D.*coeff. Electrostatic contour maps, compound 21 which has substituent at C₁₂ position (R₄) is represented.

5.4 CoMFA Summary

The effects of some adjustable parameters on CoMFA results of tricyclic 1,2,4-trioxane derivatives were studied. The conformation obtained from the HF/3-21G optimizations gave better results than those from the docking calculations. Fifteen alignment rules were used to study the influence of different alignments, the alignment 13 yield the best q^2 value, thus, alignment 13 was selected for CoMFA study. From the investigations, the type of probe atom does not have significant effect on the q^2 value.

Therefore, the default setting for probe atom type, C-sp³ with +1.0 charge was chosen. And the default setting for steric and electrostatic cut-off values 30 kcal/mol was shown to be the suitable values. These setting were subsequently used for the training set (28 compounds) to derive the CoMFA model. The predictive power ($q^2 = 0.490$) of the obtained model is lower than the acceptable value. Therefore 3 compounds with high residual values, i.e. compounds 6, 18 and 26 were excluded and the predictive power was significantly improved ($q^2 = 0.524$). The analysis of CoMFA contour maps in this study provides insight into the possible modification of the molecules for better activity.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 6

MOLECULAR DOCKING

6.1 Introduction

Computer aided rational drug design, which enable a large reduction of both budget and time, is explosively growth in the area of drug design and discovery over the past few years. One of the most popular methods for investigation of interactions between drug and receptor is a molecular docking method. It provides an estimate of a binding mode between drug and receptor. With an increasing number of known receptor structures available, interest in molecular docking is rapidly increased and much progresses have been made in recent years. An important feature of any docking method is an energy function that is capable of predicting binding modes.

A mechanism of action of any drug is very important in drug development. Generally, drug compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug-receptor interactions are required for high activity. To understand the nature of these interactions, theoretical calculations, in particular the molecular docking method, seem to be a proper tool for gaining such understanding. The obtained docking results will give information on how to modify chemical structure of the drug to achieve suitable interactions. Hence, this could bring about a development of new and more effective drugs.

Simplified tricyclic 1,2,4-trioxane analogues has similar structure to artemisinin. They contain peroxide linkage, rings A, B and C but lack of lactone ring D. Therefore, they are proposed to have similar mechanism as artemisinin^{71,72} as shown in Figure 6.1, It has been accepted that free iron from heme generated within the malarial parasite is responsible for activating the tricyclic 1,2,4-trioxane antimalarials to form cytotoxic radical intermediates. The Fe atom can make a reductive cleavage of the O-O bond by coordinating to either O₁ or O₂. In Pathway A, heme iron attacks tricyclic 1,2,4-trioxane at O₂ position and produces free radical at O₁ position (1A), which later rearranges to form C₄ free radical (2A). The compound 2A is changed to epoxide (3A) and then ultimately to hydroxylated product (4A). On the other hand, in Pathway B, heme iron

attacks tricyclic 1,2,4-trioxane at O₁ position and produces free radical at O₂ position (1B). Then, the C₃-C₄ bond is cleaved to give a carbon radical at C₄ (2B) after that ring closure is occurred to form ring-contracted product (3B).^{73,74,75}

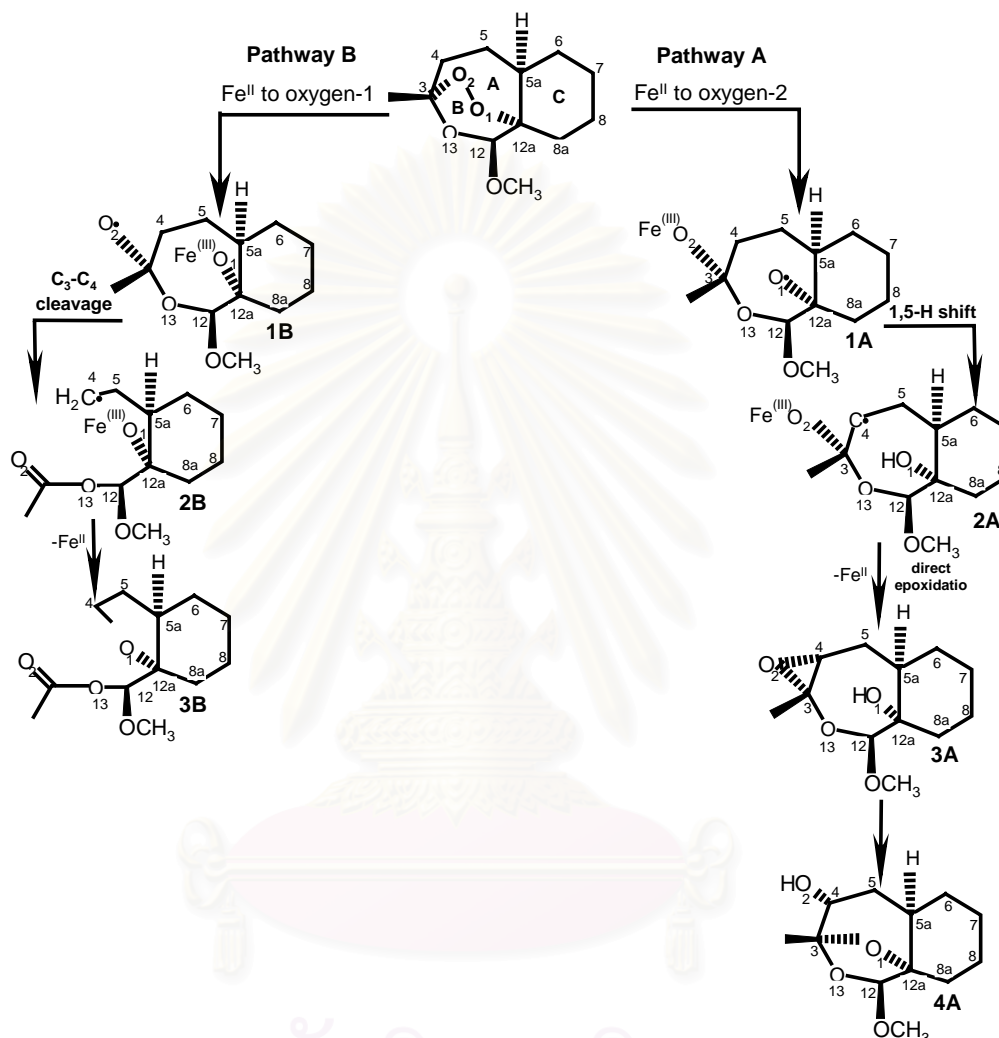


Figure 6.1 Proposed mechanisms for the Fe(II)-induced activation of simplified 1,2,4-trioxanes.

In this thesis works, the molecular docking framework was applied to investigate and predict antimalarial activities of tricyclic 1,2,4-trioxane derivatives. The relationships between biological activity and properties obtained from dock results such as binding energy, O₁-Fe distance, O₂-Fe distance, and O₁₃-Fe distance were studied. It is also expected that results might be able to give some information about the mode of action.

6.2 Docking Theory

The Binding Energy in AutoDock2.4 program can be calculated by van der waals potential and electrostatic potential terms.

Van der Waals Potential Energy

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

$$V(r) = \frac{Ae^{-br}}{r} - \frac{C_6}{r^6} \quad (6.1)$$

Graphically, if r_{eqm} is the equilibrium internuclear separation and ϵ is the well depth at r_{eqm} , then:

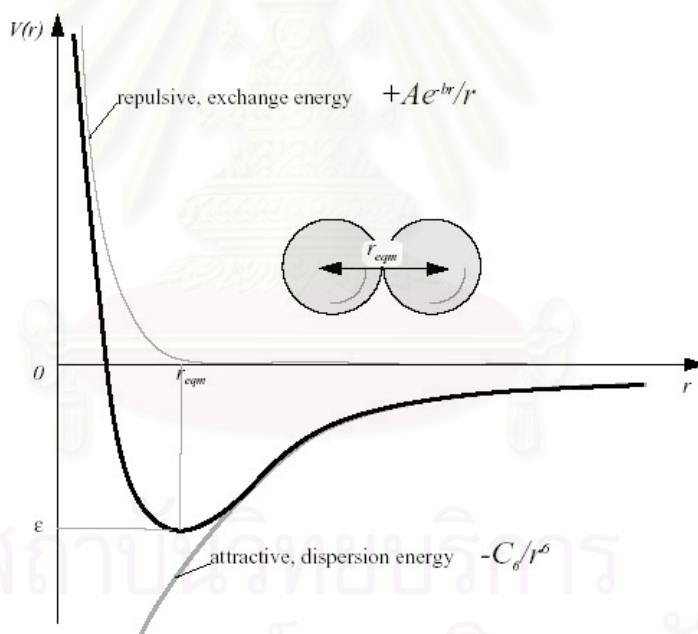


Figure 6.2 Graph showed relationship between internuclear distance (r) and well-depth (ϵ)

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

$$\frac{Ae^{-br}}{r} \approx \frac{C_{12}}{r^{12}} \quad (6.2)$$

Hence pairwise-atomic interaction energies can be approximated using the following general equation,

$$V(r) \approx \frac{C_n}{r^n} - \frac{C_m}{r^m} = C_n r^{-n} - C_m r^{-m} \quad (6.3)$$

where m and n are integers.

We can derive a general relationship between the coefficients, equilibrium separation and well depth as follows. At the equilibrium separation, r_{eqm} , the potential energy is a minimum and equal to the well depth: in other words, $V(r_{eqm}) = -\varepsilon$. The derivative of the potential with respect to separation will be zero at the minimum potential:

$$\frac{dV}{dr} = -\frac{nC_n}{r^{n+1}} + \frac{mC_m}{r^{m+1}} = 0$$

so:

$$\frac{nC_n}{r^{n+1}} = \frac{mC_m}{r^{m+1}}$$

therefore:

$$C_m = \frac{nC_n r^{m+1}}{m r^{n+1}} = \frac{n}{m} C_n r^{(m-n)} \quad (6.4)$$

Substituting C_m into the original equation for $V(r)$, then at equilibrium we obtain,

$$-\varepsilon = \frac{C_n}{r_{eqm}^n} - \frac{nC_n r_{eqm}^{(m-n)}}{m r_{eqm}^m}$$

$$C_n \left(\frac{m r_{eqm}^m - n r_{eqm}^n r_{eqm}^{(m-n)}}{m r_{eqm}^n r_{eqm}^m} \right) = -\varepsilon \quad (6.5)$$

Rearranging, equation (6.5), substituting C_m into the original equation for $V(r)$, then at equilibrium we obtain:

$$C_n = \frac{m}{n-m} \epsilon r_{eqm}^n \quad (6.6)$$

Therefore, the coefficient C_n can be expressed in terms of n , m , ϵ and r_{eqm} , then substituent C_n into equation (6.4) for C_m

$$C_m = \frac{n}{n-m} \epsilon r_{eqm}^m \quad (6.7)$$

then we obtain the general equation for $V(r)$ at any n , m :

$$V(r) \approx \frac{\frac{m}{n-m} \epsilon r_{eqm}^n}{r^n} - \frac{\frac{n}{n-m} \epsilon r_{eqm}^m}{r^m} \quad (6.8)$$

C_n and C_m are constants whose values depend on the depth of the energy well and the equilibrium separation of the two atoms nuclei. They can be calculated equation (6.9) and (6.10) by:

$$\epsilon_{XY} = \sqrt{\epsilon_{XX} \epsilon_{YY}}$$

when ϵ are well-depth

$$r_{eqm,XY} = \frac{1}{2} (r_{eqm,XX} + r_{eqm,YY}) \quad (6.9)$$

r_{eqm} are Van der Waals radius of a given atom for all pairwise distances.

Then a derivation for the Lennard-Jones potential, we obtain:

$$r_{eqm,XY} = 2^{\frac{1}{6}} \sigma \quad (6.10)$$

and the Lennard-Jones 12-6 potential becomes:

$$V_{12-6}(r) = 4\epsilon_{XY} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (6.11)$$

Hence, the coefficients C_{12} and C_6 are given by:

$$C_{12} = \epsilon_{XY} r_{eqm,XY}^{12}$$

$$C_6 = 2\epsilon_{XY} r_{eqm,XY}^6$$

Typically the 12-6 Lennard-Jones parameters ($n=12$, $m=6$) are used to model the Van der Waals' forces experienced between two instantaneous dipoles.

Electrostatic Potential Grid Maps

In addition to the atomic affinity grid maps, AutoDock requires an electrostatic potential grid map. Partial atomic charges must be assigned to the macromolecule receptor. AutoGrid calculates Coulombic interactions between the macromolecule and a probe of charge e , $+1.60219 \times 10^{-19}$ C; there is no distance cutoff used for electrostatic interactions. A sigmoidal distance-dependent dielectric function is used to model solvent screening, based on the work of Mehler and Solmajer.⁷⁶

$$\epsilon(r) = \frac{A + B}{1 + ke^{-\lambda Br}}$$

where parameters are:

$$B = \epsilon_0 - A$$

$$\epsilon_0 = \text{the dielectric constant of bulk water at } 25^\circ\text{C} = 78.4$$

$$A = -8.5525$$

$$\lambda = 0.003627$$

$$k = 7.7839$$

$$r = \text{distance}$$

6.2.1 Automated Docking

AutoDock is a program for docking small flexible ligands to a rigid protein or rigid macromolecule. It combines a fast energy evaluation through precalculated grids of affinity potentials with a Monte Carlo-simulated annealing search algorithm. It was developed by the A. J. Olson's group.⁷⁷ AutoDock uses an atomic representation of the ligand. Flexible torsions in the ligand may be defined with the utility AutoTors. The

macromolecule is treated as rigid and represented by a set of affinity grids. The grids are generated with AutoGrid on the basis of typical force field terms for van der Waals and Coulombic interactions. At every grid points, the interaction energy between a probe atom/probe charge and the whole macromolecule is calculated, supplying a “map” of affinity potentials for each defined atom type, as well as a map for the electrostatic potential. These maps serve as look-up tables for the calculation of the interaction energy during the docking process. The actual docking is performed with the main AutoDock program. To search for suitable sites of interaction, the ligand is moved through the receptor near space by small random displacements along translational, rotational, and torsional degrees of freedom. Evaluation of the interaction energy at every steps is followed by application of the Metropolis algorithm to decide on the acceptance of the new position and thus on the point from which the search will proceed. Since this is coupled to a process of simulated annealing, a wide region of conformational space can be searched, while immediately finishing in local minima next to the starting position is avoided.

The individual components of the AutoDock program, whose basics will be explained below, are: AutoTors, AutoGrid, and AutoDock.

AutoTors is the simplest of the components-it defines which bonds in the ligand are rotatable, affecting the degrees of freedom (DF) of the ligand, and thus the complexity of the computations.⁷⁷ Each rotatable torsional angle adds an extra DF, so large ligands with many torsional angles quickly become too complex to compute.

AutoGrid pre-calculates a three-dimensional grid of interaction energies based on the macromolecular target using the AMBER force fields. Since the structure of the receptor is rigid and known, interaction energies between the probe and surrounding atoms can be calculated at each point in the grid and stored in a table. Additional tables are made for each atom type in the ligand, taking into account dispersion/repulsion and hydrogen bonding energies. A second grid is made to allow for electrostatic effects, using a probe with a single positive charge. After the grid has been completed, AutoDock can begin the simulation. First, the ligand moves randomly in any one of six degrees of freedom (either translation or rotation) and the energy of the new ligand “state” is calculated. If the energy of the new state is higher than the old state, the new one is

automatically accepted as the next step in docking. However, if it is higher, then the step is accepted by the following probability function:

$$P(\Delta E) = e^{\frac{-\Delta E}{k_B T}}$$

where k_B = Boltzmann's constant (1.38×10^{-23} J/K), ΔE = energy, T = temperature (K)

The system starts at a high temperature T in order to accept most initial steps. The steps are cycled, and at the beginning of each new cycle, the temperature is reduced, making it more progressively more difficult for the docking to precede to a new step. A final low energy bound conformation is returned.

6.3 Computational Methods

The docking calculations were performed using the automated docking program, AutoDock 2.4 software.^{78,79} The AutoDock employs a simulated annealing Monte Carlo simulation in combination with a rapid grid-based energy evaluation method.⁸⁰ The rapid energy evaluation is achieved by precalculating atomic affinity potentials for each atom type present in the ligand molecule. For example, tricyclic 1,2,4-trioxane has only three atom types in the molecule (carbon, oxygen, and hydrogen); therefore, three atomic affinity potentials, i.e., receptor-carbon, receptor-oxygen, and receptor-hydrogen interaction energies, are required. To create these potentials, a grid map, which is a regular three-dimensional lattice with a selected grid spacing, is placed covering the active site of the receptor (Figure 6.3).

Considering the C atomic affinity potential, a probe atom, which is the same atom type used to create the atomic affinity potential (in this case is carbon), is placed at the edge of every lattice points. For each lattice point, the interaction energy between the probe atom and receptor atoms within a non-bonded cutoff radius of 8 Å is calculated using the Lennard-Jones 12-6 potential and is assigned to that lattice point. The O and H atomic affinity potentials are calculated in the same manner as that of the carbon.

In addition to the atomic affinity grid maps, an electrostatic potential grid map of the receptor molecule is created. The electrostatic interaction energy between the receptor and a probe of charge e , $+1.60219 \times 10^{-19}$ C is calculated using a Coulomb potential. A

sigmoidal distance-dependent dielectric function based on the work of Mehler and Solmajer⁷⁶ is used to model solvent effect.

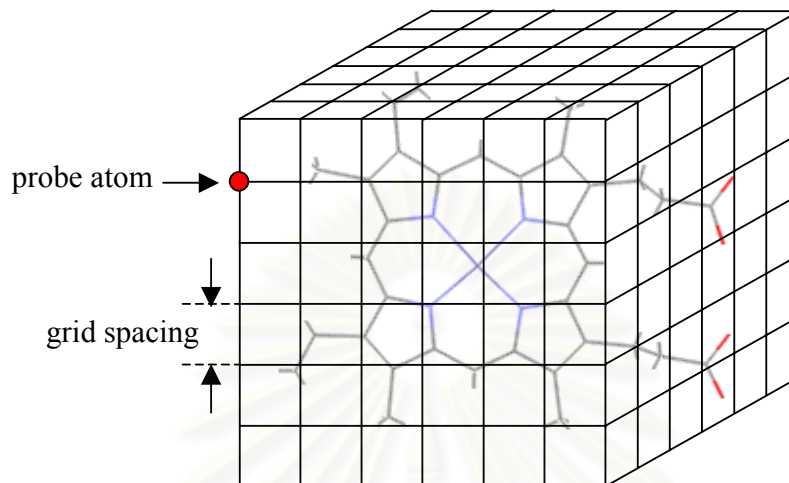


Figure 6.3 Grid base energy evaluation.

In one docking calculation, the simulations were performed for 100 annealing cycles. At the first cycle, the initial annealing temperature (RT) was set 1390.9 cal/mol and then the temperature was reduced at the rate of 0.90 per cycle. During each cycle, the ligand was gradually moved around the receptor molecule by a random displacement with a maximum translation step of 0.2 Å and a maximum orientation step of 50. The energy of new configuration was then calculated. The selection of the new configuration was based on the Metropolis algorithm.⁷⁹ The cycle terminates if the ligand makes 30,000 rejected moves. Then the simulation moves to the next cycle.

Since the Monte Carlo simulation is based on random movements, the final docked configuration depends on the starting configuration. In order to avoid any bias and to generate as many final docked configurations as possible, the starting configuration was assigned in a random manner for each docking calculation and 100 docking calculations were performed. A cluster analysis was used to categorize all 100 docked configurations into groups. Configurations with root-mean-square-deviation (rmsd) values of less than 1 Å were grouped together. In each group, the lowest energy configuration was selected as the representative of that group. The “% Frequency” was used to represent the number of members (configurations) in each group. Our attention

was focused to the group with the highest % Frequency or “the most frequency configuration”. And this configuration is most probably corresponding to the docked configuration in the real system.

6.4 Determination of Suitable Docking Parameters

Because a core structure of trioxane compounds is similar to that of artemisinin compounds, dock parameters of artemisinins taken from our previous works⁸⁰ were used as the initial setting as following. A grid map of dimension 25x25x25 Å³ with a grid spacing of 0.50 Å was employed. For the simulated annealing calculations, 100 docking runs with 100 annealing cycles per run were performed. A cycle terminated if the ligand makes either 30,000 accepted or 30,000 rejected moves. An initial annealing temperature, RT, was given as 100 kcalmol⁻¹ with a reduction factor of 0.90. The combined AMBER/MMFF parameters for the Lennard-Jones 12,6 potentials and the Coulomb potentials were taken from authors of the program.⁸¹ The docking results obtained by this initial set of parameters seem to be not suitable for trioxane compounds. Therefore, investigations of important parameters grid dimension, grid spacing, starting temperature (T_s), final temperature (T_f), and temperature reduction rate factor were carried out to establish appropriate values. In order to reduce computing time for the investigations of parameters affecting docking results, only four compounds, namely 4, 8, 10, and 15, were used.

6.4.1 Grid Dimension

The AutoDock program uses the Monte Carlo simulation for searching docking conformation. As energy of a molecular system is calculated for every movements in the simulation, it is very time consuming process to evaluate the energy by a direct method. In order to avoid such situation, the AutoDock employs a rapid grid-based energy evaluation method. Therefore, it is necessary to investigate effects of the grid dimension and the grid spacing on docking results. Three different grid boxes with dimensions of 30 x 30 x 30 Å³, 25 x 25 x 25 Å³ and 20 x 20 x 20 Å³ were used. The starting temperature of 50,327.1 K, the final temperature of 1.3 K, the temperature reduction rate of 0.90, and

the grid spacing of 0.50 Å were used for all calculations. The results are given in Table 6.1

Table 6.1 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with three different grid dimensions.

Compound No.	Binding Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
A. Grid dimension 30 x 30 x 30 Å ³					
4	-30.69	43	1.99	2.96	4.75
8	-31.52	22	2.11	2.98	4.95
10	-30.37	21	2.14	2.97	4.86
15	-28.69	20	5.56	5.13	4.78
	-31.57	18	2.17	3.05	4.93
B. Grid dimension 25 x 25 x 25 Å ³					
4	-30.65	39	1.96	2.96	4.78
8	-31.53	30	2.09	2.97	4.90
10	-30.28	35	2.00	2.91	4.80
15	-28.71	21	5.55	5.13	4.77
	-31.27	17	2.14	2.93	4.93
C. Grid dimension 20 x 20 x 20 Å ³					
4	-30.84	55	1.98	2.94	4.81
8	-31.91	37	2.10	3.06	4.94
10	-30.50	22	2.01	2.96	4.83
15	-32.01	29	2.09	2.92	4.87

The results from grid dimensions of 25 x 25 x 25 Å³ and of 30 x 30 x 30 Å³ are similar. In compounds 4, 8, and 10 the distance between heme iron and O₁ is shorter than that of O₂ and O₁₃ which indicates that heme iron interacts with O₁ more preferably than O₂ and O₁₃. On the other hand, compound 15 has the O₁₃-Fe as the shortest distance. This distance is too large to account for any meaningful interactions. The compound 15 docks underneath the porphyrin plane instead of over the plane like in other compounds.

However, its second highest frequent configuration which has the lowest energy, has O₁-Fe distance as the shortest distance. Unlike the first two cases, results from the grid dimension of 20 x 20 x 20 Å³ indicated that all 4 compounds have O₁-Fe as the shortest distance. It can be seen that heme iron prefers to bind with the endoperoxide linkage of 1,2,4-trioxane compounds at the O₁ position.

Comparing these three grid dimensions, the 20 x 20 x 20 Å³ size give lower energy and larger number of member in the most frequency cluster group. In addition, the most occurring configuration in compound 15 has O₁-Fe shortest distance which is in the same range as the other compounds. This is possibly due to its smaller grid dimension that does not allow 1,2,4-trioxane to move away from heme molecule. Therefore, the grid dimension of 20 x 20 x 20 Å³ was chosen for further calculations.

6.4.2 Grid Spacing

The investigations on the effects of grid spacing were preformed by using grid spacing 0.40 and 0.50 Å. The results are shown in Table 6.2.

Table 6.2 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with two different grid spacings.

Compound No.	Binding Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
A. grid spacing 0.50 Å					
4	-30.84	55	1.98	2.94	4.81
8	-31.91	37	2.10	3.06	4.94
10	-30.50	22	2.01	2.96	4.83
15	-32.01	29	2.09	2.92	4.87
B. grid spacing 0.40 Å					
4	-31.71	40	1.98	2.94	4.81
8	-32.63	33	2.02	2.82	4.80
10	-31.32	24	1.94	2.84	4.75
15	-32.89	37	1.99	2.91	4.80

The dock results using grid spacing of 0.40 Å have shorter O₁-Fe than those with grid spacing of 0.50 Å except compound 4, which are equal. Moreover, binding energies obtained with grid spacing of 0.40 Å are lower than those with grid spacing of 0.50 Å. Thus, the grid spacing of 0.40 Å was used for further calculations.

6.4.3 Starting Temperature (T_s), Final Temperature (T_f) and Temperature Reduction Rate

The AutoDock program employs a simulated annealing method to find a global energy minimum of drug-receptor complex. In this method, the annealing is defined as a process where the temperature of a molten substance is gradually reduced until the material crystallizes. Hence, the rate of decreasing temperature is very important. The temperature should be slowly lowering so that there is enough time for the substance to attain thermal equilibrium within each stage. The temperature should be kept reducing until a suitable value or the best solution is obtained. Therefore, the temperature reduction rate should be carefully controlled. In this study, the effects of the starting temperature, the final temperature, and the temperature reduction rate were investigated. The results are shown in Tables 6.3, 6.4 and 6.5, respectively.

Table 6.3 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different starting temperature (T_s).

Compound No.	Binding Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
A. $T_s = 50,327.1$ K					
4	-31.71	40	1.98	2.94	4.81
8	-32.63	33	2.02	2.82	4.80
10	-31.32	24	1.94	2.84	4.75
15	-32.89	37	1.99	2.91	4.80
B. $T_s = 2,516.4$ K					
4	-31.89	42	1.98	2.94	4.80
8	-32.62	59	2.03	2.82	4.82
10	-31.97	45	1.98	2.88	4.77
15	-32.78	52	1.96	2.76	4.73
C. $T_s = 700$ K					
4	-31.91	45	1.98	2.94	4.81
8	-32.63	48	2.00	2.82	4.80
10	-31.34	54	1.96	2.89	4.81
15	-32.72	52	1.96	2.87	4.77

The effects of starting temperature were investigated. Three starting temperatures 50,327.1 K, 2,516.4 K and 700 K were selected. The grid dimension of 20 x 20 x 20 Å³ with the grid spacing of 0.40 Å and the final temperature of 1.3 K and the temperature reduction rate of 0.90 were used. The results are shown in Table 6.3. The results from three different starting temperature are similar. The distance between heme iron and O₁ is shorter than that of O₂ and O₁₃, which indicates that heme iron interacts with O₁ more preferably than O₂ and O₁₃. It is found that starting temperature 700 K and 2,516.7 K give comparable results. Consider the lesser CPU time the temperature 700 K is therefore chosen for the calculation.

Table 6.4 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different final Temperature (T_f).

Com pound No.	Binding Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
A. $T_f = 1.3$ K					
4	-31.91	45	1.98	2.94	4.81
8	-32.63	48	2.00	2.82	4.80
10	-31.34	54	1.96	2.89	4.81
15	-32.72	52	1.96	2.87	4.77
B. $T_f = 129.7$ K					
4	-28.89	17	2.37	3.13	5.04
	-30.00	16	1.95	2.89	4.77
8	-30.16	13	1.97	2.83	4.76
10	-29.55	12	1.94	2.89	4.77
15	-27.95	14	5.05	4.91	4.98
	-29.74	12	2.04	2.91	4.74
C. $T_f = 0.02$ K					
4	-32.79	42	1.94	2.82	4.72
8	-34.35	49	1.96	2.76	4.74
10	-32.97	54	2.00	2.89	4.81
15	-34.34	49	1.96	2.77	4.77

The effects of final temperature were investigated. Three final temperatures 1.3 K, 129.7 K and 0.02 K were selected. The grid dimension of $20 \times 20 \times 20 \text{ \AA}^3$ with the grid spacing of 0.40 \AA and the starting temperature of 700 K and the temperature reduction rate of 0.90 were used. The results are shown in Table 6.4. Considering the results with final temperature 1.3 K of compound 4, 8, 10 and 15 the distance between heme iron and O₁ is shorter than that of O₂ and O₁₃, which indicates that heme iron interacts with O₁ more than preferably than O₂ and O₁₃. Considering the results obtained with final

temperature of 129.7 K, the docked configurations were clustered into too many groups which consist of only a few members. The results from final temperature 0.02 K are similar to those with final temperature 1.3 K but the energy is lower. Therefore, the final temperature 0.02 K was chosen for further calculations.

Table 6.5 Docking results of heme and 4 derivatives of 1,2,4-trioxane compounds with different Temperature Reduction Rate.

Compound No.	Binding Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
A. Temperature Reduction Rate 0.90					
4	-32.79	42	1.94	2.82	4.72
8	-34.35	49	1.96	2.76	4.74
10	-32.97	54	2.00	2.89	4.81
15	-34.34	49	1.96	2.77	4.77
B. Temperature Reduction Rate 0.95					
4	-32.95	52	2.02	2.90	4.81
8	-34.37	49	1.95	2.77	4.73
10	-24.01	47	4.02	4.78	5.96
	-29.65	45	1.88	2.66	4.62
15	-34.36	44	1.96	2.76	4.74
C. Temperature Reduction Rate 0.99					
4	-32.95	54	2.01	2.89	4.81
8	-34.39	50	1.97	2.76	4.74
10	-24.09	48	4.00	4.77	5.95
	-29.67	45	1.85	2.66	4.62
15	-34.40	48	1.95	2.78	4.73

The effects of temperature reduction rate were investigated. Three temperature reduction rates 0.90, 0.95 and 0.99 were selected. The grid dimension of 20 x 20 x 20 Å³ with the grid spacing of 0.40 Å and the starting temperature of 700 K and the final

temperature of 0.02 K were used. The results are shown in Table 6.5. Considering results from temperature reduction rate 0.90, the distance between heme iron and O₁ is shorter than that of O₂ and O₁₃. The temperature reduction rate 0.95 indicated that all four compounds having shorter O₁-Fe distance are the most probable docking configuration. However, compound 10 has very long O₁-Fe distance, 4.02 Å, which is too large to account for any intently interactions. Nevertheless, its second highest frequent configuration which has the lowest energy, has O₁-Fe distance with in the same range as the others compounds. For temperature reduction rate 0.99, the obtained results is similar to those of temperature reduction rate 0.95. As the temperature reduction rate of 0.90 gives a better results and requires less calculation time, it seem to be an appropriate choice for further studies.

6.5 Docking of All 32 compounds

All suitable parameters determined from the previous section were employed for the docking calculations of all 32 compounds. These derivatives were optimized at HF/3-21G and the atomic charges were calculated at the HF/6-31G(d) level. The atomic charges of heme were assigned at the HF/6-311G(d,p) level. The docking results of 32 derivatives and heme are given Table 6.6. The distances from Fe²⁺ to O₁, O₂ and O₁₃ are in the range of 1.84 to 2.27 Å, 2.38 to 3.11 Å, and 4.51 to 4.94 Å, respectively. The most occurring configurations in most tricyclic 1,2,4-trioxane compounds have O₁ pointing toward the heme iron with the compound are over the porphyrin plane(see Figure 6.5) except compounds 16-21 which have O₂ pointing toward the heme iron(see Figure 6.6). The replacement of the oxygen atom by sulfur atom at the C₁₂ position in compounds 16 to 21 has significant effect on the docking results. This structure facilitates the encroachment of Fe²⁺ to O₂ more than O₁ due to steric effect is more pronounce than electrostatic effect. Consider the electrostatic effect, the partial charges of O₁ and O₂ are not different, therefore the preference for heme iron to bind with O₁ and O₂ is equal. For steric effect, the bulky substituent R₄ is closer to O₁ than O₂ will hinder heme to approach O₁ and thus interacts with O₂. The substituent groups at C₃ position (R₁) of compounds 1-15 cause the steric hindrance at these positions, despite partial charges of O₂ are more negative than O₁, heme iron did not dock at the O₂, heme iron prefers to approach at the

O₁ position. It is revealed that the steric effect at C₃ position (R₁) in compounds 1-15 is dominant than electrostatic contribution. Compounds 22-24 with substituent groups at the C₄ position (R₂), the most frequency configurations in those compounds have O₁ pointing toward the heme iron. The reason may be due to its structure. The substituent groups at the C₄ position (R₂) hinder the heme iron to approach at the O₂ side but facilitate the approach of heme iron to the O₁ side. Moreover, the partial charges of compounds 22-24 at the O₁ position are more negative than O₂ position. Therefore, both steric and electrostatic effect are promoted for docking of these compounds. Compounds 25-32 have substituent groups at the C₄ and/or C_{8a} positions (R₂ and/or R₃), the highest frequency configurations in those compounds have O₁ pointing toward the heme iron. This suggested that electrostatic effect contributions to activity more than steric effect because the substituent groups at the C_{8a} position (R₃) is far from the endoperoxide linkage, steric effect should not play dominate role on the binding characteristic to heme. The partial charges of compounds 25-32 at the O₁ position are more negative charge than O₂ position, these support the hypothesis of only electrostatic effect is the main features for activity. The superimposed docking configurations between heme and 32 derivative of 1,2,4-trioxane compounds are illustrated in Figure 6.3.

The Correlations between antimalarial activities and properties from docking calculations, i.e., binding energy, O₁-Fe distance, O₂-Fe distance and O₁₃-Fe distance of 32 derivatives of 1,2,4-trioxane compounds were investigated. No significant relationship was found as indicated by r² values of 0.009, 0.010, 0.003 and 0.011, respectively.

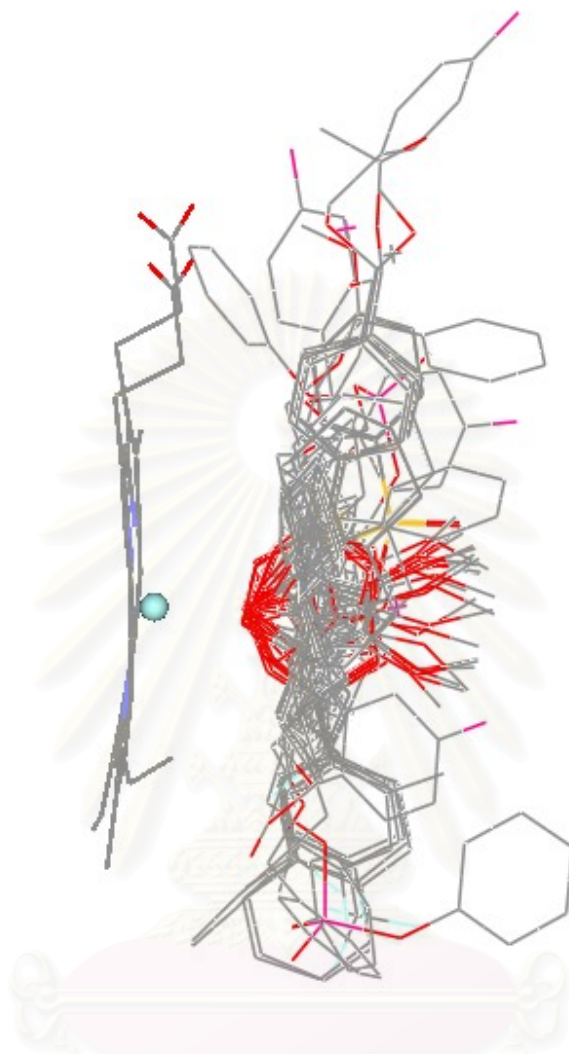


Figure 6.4 Superimposed docking configurations between heme and 32 derivative of 1,2,4-trioxane compounds (without hydrogen atoms).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

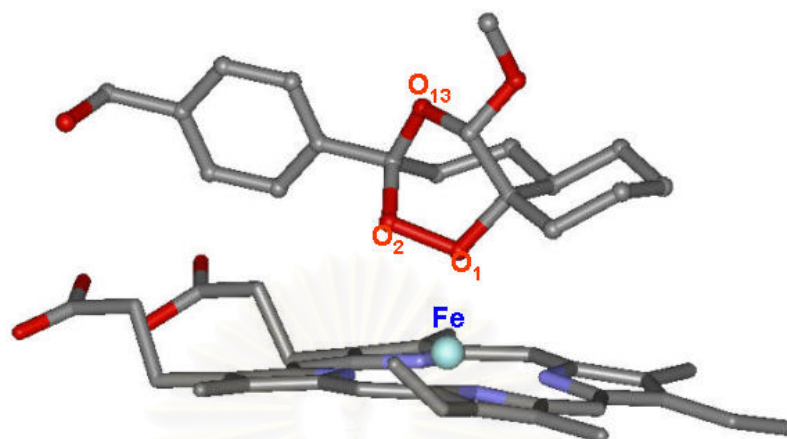


Figure 6.5. Docking configuration between heme and compound 15 (without hydrogen atoms).

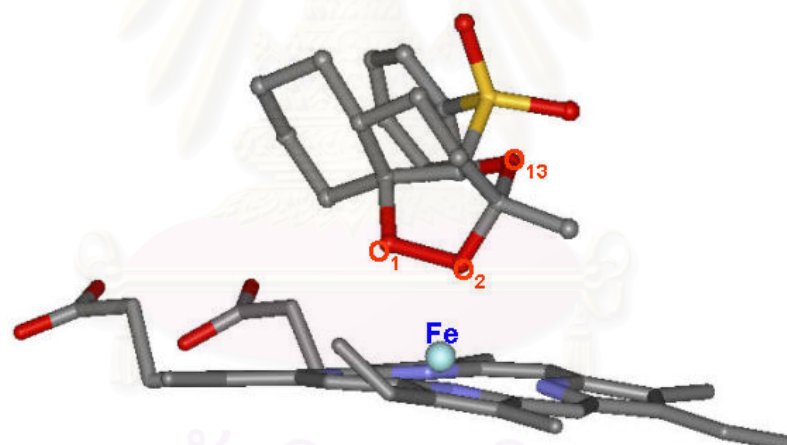


Figure 6.6. Docking configuration between heme and compound 19 (without hydrogen atoms).

Table 6.6. The docking results of heme and 32 derivatives of 1,2,4-trioxane compounds.

Compound No.	Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
1	-30.01	58	<u>1.86</u>	2.69	4.63
2	-32.03	91	<u>1.95</u>	2.81	4.67
3	-35.98	60	<u>1.91</u>	2.40	4.51
4	-32.79	42	<u>1.94</u>	2.82	4.72
5	-35.34	52	<u>1.97</u>	2.75	4.73
6	-33.34	50	<u>1.98</u>	2.86	4.79
7	-33.70	45	<u>1.99</u>	2.87	4.79
8	-34.35	49	<u>1.96</u>	2.76	4.74
9	-33.91	52	<u>1.95</u>	2.88	4.76
10	-32.97	54	<u>2.00</u>	2.89	4.81
11	-33.41	41	<u>1.91</u>	2.79	4.69
12	-32.97	43	<u>2.01</u>	2.90	4.81
13	-36.05	45	<u>1.97</u>	2.76	4.74
14	-34.42	27	<u>1.94</u>	2.79	4.73
15	-34.34	49	<u>1.96</u>	2.77	4.77
16	-31.48	51	2.90	<u>2.03</u>	3.65
17	-31.41	36	2.90	<u>2.03</u>	3.65
18	-31.50	40	2.84	<u>2.00</u>	3.70
19	-33.57	25	2.38	<u>1.86</u>	3.98
20	-34.42	35	2.38	<u>1.85</u>	3.97
21	-34.04	77	2.38	<u>1.84</u>	3.98
22	-34.59	64	<u>2.00</u>	2.65	4.59
23	-33.50	46	<u>1.90</u>	2.68	4.64
24	-35.91	45	<u>1.84</u>	2.49	4.55
25	-35.43	61	<u>2.21</u>	2.81	4.81
26	-30.23	42	<u>2.22</u>	2.78	4.84

Table 6.6 (Continued)

Compound No.	Energy (kcal/mol)	% Frequency	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)
27	-36.95	41	<u>1.92</u>	2.60	4.62
28	-38.20	52	<u>2.19</u>	2.84	4.79
29	-34.48	46	<u>2.09</u>	2.92	4.78
30	-34.25	55	<u>2.08</u>	2.66	4.61
31	-40.18	53	<u>2.27</u>	2.38	4.53
32	-32.94	68	<u>2.11</u>	3.11	4.94

6.6 Docking Summary

The molecular docking method using the simulated annealing Monte Carlo simulations was employed to investigate the binding between heme and 32 derivatives of tricyclic 1,2,4-trioxane compounds. The parameters affecting the docking results were also investigated. For the simulated annealing calculations, the moderate initial temperature (700K), the very low final temperature (0.02K), and the temperature reduction rate of 0.90 were suggested. The small grid spacing (0.40Å) and the moderate grid size (20x20x20Å³) are recommended for the grid-based energy evaluations. From the docking results of all 32 derivatives of tricyclic 1,2,4-trioxane compounds interact with heme molecule mainly at the endoperoxide linkage as in artemisinin.⁷⁷ Therefore they confirm the same mechanism of action as artemisinin. The most occurring configurations in most tricyclic 1,2,4-trioxane compounds have O₁ pointing toward the heme iron. In exception, compounds 16-21 have the most occurring configurations with O₂ pointing toward the heme iron. From the results, we can conclude that the binding between tricyclic 1,2,4-trioxane compounds and heme is controlled by the steric effect at C₃ position(R₁) for compounds 1-15, by steric effect at C₁₂ position(R₄) for compounds 16-21, by both steric effect at C₄ position(R₂) and electrostatic effect for compounds 22-24 and by only electrostatic effect for compounds 25-32. Moreover, the dock parameters such as binding energy, O₁-Fe distance, O₂-Fe distance and O₁₃-Fe distance of 32 derivatives of tricyclic 1,2,4-trioxane compounds were used to find the correlation

with the antimalarial activities, however, no significant relationship was found. Therefore, it seems that, the dock information is possibly not sufficient to explain the activities of these compounds and additional information from other methods should be considered together.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 7

CONCLUSIONS

Quantitative Structure Activity Relationship (QSAR) approaches have been applied to investigate the structure requirements of antimalarial activity of tricyclic 1,2,4-trioxane compounds by means of classical QSAR (2D-QSAR), CoMFA (3D-QSAR), automated molecular docking using the simulated annealing monte carlo simulations. The optimized geometries based on the *ab initio* method at the HF/3-21G level of theory has been utilized to calculate electronic and molecular properties of compounds.

In the classical QSAR and CoMFA techniques, the whole set of compounds were used to derive the model because the number of compounds in each group is small and not sufficient to divide into training set and testing set for establishing the model for each group and predicting the compounds. The obtained models provide a good relationships between activities and molecular properties with rather high predictive power.

The information obtained from both classical QSAR and CoMFA provides some suggestions on the structural modification of tricyclic 1,2,4-trioxane compounds to increase the antimalarial effects as the following;

- (1) The substituent group at the C₃ position (R₁) especially alkyl group is not preferred. However when it is replaced by phenyl substituent group, the activity is enhanced.
- (2) The small steric substitution at the C₄ position (R₂) and the highly positive charge at the first carbon atom of substituent group connected to the C₄ position (R₂) are required to improve the binding potency.
- (3) The substituent group at C_{8a} (R₃) and C₁₂ (R₄) should cause C_{8a} and C₁₂ more positive charges, or another word the electron withdrawing substituent group at those positions are conducive to the activity.

Besides the 2D-QSAR and CoMFA results that giving information on relationship between structures of ligands and biological activities, docking experiments were performed to predict the interaction parameters of tricyclic 1,2,4-trioxane with the inhibition binding site of heme. Docking of the compounds revealed a consistent set of

recurring binding modes i.e. the O₁ of compounds pointing toward heme iron with the O₁-Fe distance in the range of 1.84-2.27 Å which is comparable to artemisinin derivatives, except compounds 16-21 containing substituent at the C₁₂ position (R₄) have O₂ approach heme iron with the O₂-Fe distance in the range of 1.84-2.03 Å. Thus, the binding of heme and tricyclic 1,2,4-trioxane could be mainly described by interaction between heme iron and peroxide oxygen(s) as in artemisinin, hence, the same mechanism of action. The different approach to oxygen atom of peroxide linkage of compounds 16-21 may result from the S and SO₂ in the substituent group compare to the other compounds that contain only carbon and hydrogen atoms. The reasons possible due to the different mechanism of action or parameters for sulfur(S) from Amber force field obtained from amino acid of protein may not suitable for this docking calculation. In order to support the QSAR studies, docking parameter e.g. binding energy, O₁-Fe distance, O₂-Fe distance were used to correlate with antimalarial activities, however the good relationship could not be achieved. One explanation might arise from the diverse substituents of the compounds.

The combination of ligand-based (classical QSAR and CoMFA) and structural-based design (molecular docking) results give better insights into the structural features requirements of tricyclic 1,2,4-trioxane compounds for NF54 strain and the importance of peroxide bridge in the mode of action of these derivatives. These results could provide a basis guideline to design and develop new compounds with higher activities.

Further Work

Several improvement are possible:

- 1) If number of compounds is large enough, it is recommended to classify all compounds into groups according to their structural similarity.
- 2) Other techniques such as Molecular Dynamics simulation (MD) may show how inhibitor interact with receptor in solution phase.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

REFERENCES

1. Robert, A.; Dechy-Cabaret, O.; Cazelles, J.; Benoit-Vical, F.; Meunier, B. From Mechanistic Studies on Artemisinin Derivatives to New Modular Antimalarial Drugs. Acc. Chem. Res. 35 (2002): 167-174.
2. World Health Organization. The World Health Reported, 2003.
3. Balter, M.; Marshall, E.; Vogel, G.; Taubes, G.; Pennisi, E.; and Enserink, M. MALARIA: A Renewed Assault on an Old and Deadly Foe. The Case for Deemphasizing Genomics in Malaria Control. Science, 290 (2000): 428-430.
4. กรองทอง ทิมาสาร. ความสูญเสียจากโรคมาลาเรีย. วารสารมาลาเรีย 33 (1998): 217-218.
5. Wilairatana, P; and Looareesuwan, S. Malaria in Thailand in former times. J. Roy. Inst. Thai. 27 (2002): 434-440.
6. Looareesuwan, S; Wilairatana, P; Chalermrut, K; Rattanapong, Y; Canfield, C; and Hutchinson, DBA. Efficacy and safety of atovaquone/proguanil compared with mefloquine for treatment of acute Plasmodium falciparum malaria in Thailand. Am. J. Trop. Med. Hyg. 60 (1999): 526-532.
7. Looareesuwan, S; Wilairatana, P; and Glanarongran, R. Atovaquone and proguanil hydrochloride followed by primaquine for treatment of Plasmodium vivax malaria in Thailand. Trans. R. Soc. Trop. Med. Hyg. 93 (1999): 637-640.
8. World Health Organization. Severe falciparum malaria. Trans. R. Soc. Trop. Med. Hyg. 94(2000): 1-90.
9. Looareesuwan, S; Vanijanonta, S; and Viravan, C.; et al. Randomised trial of mefloquine tetracycline, and quinine-tetracycline for acute uncomplicated falciparum malaria. Acta. Tropica. 57 (1994): 47-53.
10. Looareesuwan, S.; Viravan, C; and Vanijanonta, S.; et al. Randomised trial of mefloquinedoxycycline, and artesunate-doxycycline for treatment of acute uncomplicated falciparum malaria. Am. J. Trop. Med. Hyg. 50 (1994): 784-789.
11. Looareesuwan, S; Wilairatana, P; Chokejindachai, W; et al. A randomised, double-blind, comparative trial of a new oral combination of artemether and benflumetol

- (CGP 56697) with mefloquine in the treatment of acute *Plasmodium falciparum* malaria in Thailand. Am. J. Trop. Med. Hyg. 60 (1999): 238-243.
12. Looareesuwan, S; Wilairatana, P; Vanijanonta, S; Vugt, M; Wilairatana, P.; and Gemperli, B.; et al. Efficacy of six doses of artemetherlumefantrine (benflumetol) in multidrug resistant *Plasmodium falciparum* malaria. Am. J. Trop. Med. Hyg. 60 (1999): 936-942.
 13. Bruce-Chwatt, L.J. (ed.), Chemotherapy of Malaria. Revised second edition, England: White, N. J.; Fosten, F.; Looareesuwan, S.; Watkins, W. M.; Marsh, K.; Snow, R. W.; Kokwaro, G.; Ouma, J.; Hien, T. T.; Molyneux, M. E.; Taylor, T. E.; Newbold, C. I.; Ruebush, T.K.; Danis, M.; Greenwood, B. M.; Anderson, R. M.; and Olliaro, P. Averting a Malaria Disaster. The Lancet. 353 (1999): 1965-1967.
 14. Price, R. N.; Nosten, F.; Luxemburger, C.; Van Vught, M.; Phaipun, L.; Chongsuphajaisiddhi, T.; and White, N. J. Artesunate/mefloquine treatment of multi-drug resistant *falciparum* malaria. Trans. R. Soc. Trop. Med. Hyg. 91 (1997): 574-577.
 15. Brockman, A.; Price, R. N.; van Vught, M.; Heppner, D. G.; Walsh, D.; Sookto, P.; Wimonwattrawatee, T.; Looareesuwan, S.; White, N. J.; Nosten, F. *Plasmodium falciparum* antimalarial drug susceptibility on the northwestern border of Thailand during five years of extensive artesunate-mefloquine use. Trans. R. Soc. Trop. Med. Hyg. 94 (2000): 537-544.
 16. Anderson, R. F.; and Saul, A. Malaria Vaccines. Parasitol. Today. 16(2000): 444-447.
 17. Daubersies, P.; Thomas, A. W.; Millet, P.; Brahimi, K.; Langermans, J. A. M.; Ollomo, B.; BenMohamed, L.; Slierendregt, B.; Eling, W.; Van Belkum, A.; Dubreuil, G.; Meis, J. F. G. M.; Guérin-Marchand, C.; Cayphas, S.; Cohen, J.; Gras-Masse, H.; and Druilhe, P. Protection against *Plasmodium falciparum* malaria in chimpanzees by immunization with the conserved pre-erythrocytic liver-stage antigen 3. Nat. Med. 6 (2000): 1258-1263.
 18. Curtis, C. F. and Hoffman, S. L. The Case for Deemphasizing Genomics in Malaria Control. Science. 290 (2000): 1508-1509.

19. Calas, M.; Ancelin, M. L.; Cordina, G.; Portefaix, P.; Piquet, G.; Vidal-Sailhan, V.; Vial, H. Antimalarial Activity of Compounds Interfering with Plasmodium falciparum Phospholipid Metabolism: Comparison between Mono- and Bisquaternary Ammonium Salts. J. Med. Chem. 43 (2000): 505-516.
20. Ridley, R. G.; Hofheinz, W.; Matile, H.; Jaquet, C.; Dorn, A.; Masciadri, R.; Jolidon, S.; Rich ter, W. F.; Guenzi, A.; Girometta, M.-A.; Urwyler, H.; Huber, W.; Thaitong, S.; Pe ters, W. 4-aminoquinoline analogs of chloroquine with shortened side chains retain activity against chloroquine-resistant Plasmodium falciparum. Antimicrob. Agents Chemother. 40 (1996): 1846-1854.
21. Girault, S.; Grellier, P.; Berecibar, A.; Maes, L.; Lemièrre, P.; Mouray, E.; Davioud-Charvet, E.; Sergheraert, C. Antiplasmodial Activity and Cytotoxicity of Bis-, Tris-, and Tetraquinolines with Linear or Cyclic Amino Linkers. J. Med. Chem. 44 (2001): 1658-1665.
22. Biot, C.; Glorian, G.; Maciejewski, L. A.; Brocard, J. S. Synthesis and Antimalarial Activity in Vitro and in Vivo of a New Ferrocene-Chloroquine Analogue. J. Med. Chem. 40 (1997): 3715-3718.
23. Duc, D. D.; de Vries, P. J.; Nguyen, X. K.; Le Nguyen, B.; Kager, P. A.; van Boxtel, C. J. The pharmacokinetics of a single dose of artemisinin in healthy Vietnamese subjects. Am. J. Trop. Med. Hyg. 51 (1994): 785-790.
24. Posner G. H.; Jeon, H. B.; Parker, M. H.; Krasavin, M.; Paik, I-H.; Shapiro, T. A. Antimalarial Simplified 3-Aryltrioxanes: Synthesis and Preclinical Efficacy/Toxicity Testing in Rodents J. Med. Chem. 44 (2001): 3054-3058.
25. Hien, T.T. and White, N. Qinghaosu. The Lancet. 341 (1993): 603-608.
26. Brossi, A.; Venugopalan, B.; Gerpe, L.D.; Yeh, H.J.C.; Flippen-Anderson, J.L.; Buchs, P.; Luo, X.D.; Milhous, W.; and Peters, W. Arteether, a new antimalarial drug: synthesis and antimalarial properties. J. Med. Chem. 31 (1988): 645-650.
27. Jefford, C. W.; McGoran, E. C.; Boukouvalas, J.; Richardson, G.; Robinson, B. L. and Peter, W. Synthesis of new 1,2,4-trioxanes and their antimalarial activity. Helv. Chim. Acta. 71 (1988):1805-1812.

28. Jefford, C. W.; Velarde, J. A.; Bernardinelli, G.; Bray, D. H.; Warhurst, C.; Milhous, W. K. Synthesis Structure and Antimalarial Activity of Tricyclic 1,2,4-Trioxanes Related to Artemisinin. Helv. Chim. Acta. 76 (1993): 2775-2787.
29. Posner, G. H.; Oh, C. H.; and Milhous, W. K. Olefin oxidative cleavage and dioxetane formation using triethylsilyl hydrotrioxide: Applications to preparation of potent antimalarial 1,2,4-trioxanes. Tetrahedron Lett. 32 (1991): 4235-4238.
30. Posner, G. H.; Oh, C. H.; Gerena, L.; Milhous, W. K. Extraordinarily Potent Antimalarial Compounds: New, Structurally Simple, Easily Synthesized, Tricyclic 1,2,4-Trioxanes. J. Med. Chem. 35 (1992): 2459-2467.
31. Posner, G. H.; Wang, D.; Cumming, J. N.; Oh, C. H.; French, A. N.; Bodley, A. L.; Shapiro, T. A. Further Evidence Supporting the Importance of and Restrictions on a Carbon-Centered Radical for High Antimalarial Activity of 1,2,4-Trioxanes Like Artemisinin. J. Med. Chem. 38 (1995): 2273-2275.
32. Grigorov, M.; Weber, J.; Tronchet, J.M.J.; Jefford, C.W.; Milhous, W.K.; and Maric, D. A QSAR Study of the Antimalarial Activity of Some Synthetic 1,2,4-Trioxanes. J. Chem. Inf. Comput. Sci. 37 (1997): 124-130.
34. Born, M.; and Oppenheimer, J. R. Zur quantumthorie der molekeln. Ann. Physik. 84 (1927): 457.
35. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. Rev. Mod. Phys. 23 (1951): 69.
36. Hall, G. G. The molecular orbital theory of chemical valence. VIII - A Method of calculating ionization potentials. Proc. Roy. Soc. (London) A205, (1951): 541.
37. Slater, J.C. Atomic Shielding Constant. Phys. Rev. 36 (1930): 57.
38. Boys, S. F. Electronic Wave Functions. I. A. general Method of calculation for the Stationary States of Any Molecular System. Proc. Roy. Soc. (London) A200, (1950): 542.
39. Mulliken, R. S. Charge Population. J. Chem. Phys. 23 (1955): 1833, 1841, 2338, 2343.
40. Jefford, C. W.; Velarde, J. A.; Bernardinelli, G.; Bray, D. H.; Warhurst, C.; Milhous, W. K. Synthesis Structure and Antimalarial Activity of Tricyclic 1,2,4-Trioxanes Related to Artemisinin. Helv. Chim. Acta. 76 (1993): 2775-2787.

41. Posner, G. H.; Park, S. B.; Gonza'lez, L.; Wang, D.; Cumming, J. N.; Klinedinst, D.; Shapiro, T. A.; Bachi, M. D. Evidence for the Importance of High-Valent Fe=O and of a Diketone in the Molecular Mechanism of Action of Antimalarial Trioxane Analogs of Artemisinin. J. Am. Chem. Soc. 118 (1996): 3537-3538.
42. Posner, G. H.; Coming, J. N.; Woo, S. H.; Ploypradith, P. Xie, S. J. Shairo, T. A. Orally Active Antimalarial 3-Substituted Trioxanes: New Synthetic Methodology and Biological Evaluation. J. Med. Chem. 41 (1998): 940-951.
43. Posner, G. H.; O'sDowd, H.; Caferro, T.; Cumming, J. N.; Ploypradith, P.; Xie, S.; Shapiro, T. A.; Antimalarial Sulfone Trioxanes. Tetrahedron Lett. 39 (1998): 2273-2276.
44. Cumming, J. N.; Wang, D.; Park, S. B.; Shapiro, T. A.; Posner, G. H. Design Synthesis Derivatization and Structure-Activity Relationships of Simplified Tricyclic 1,2,4-Trioxane Alcohol Analogues of the Antimalarial Artemisinin. J. Med. Chem. 41 (1998): 952-964.
45. Tonmunphean, S.; Kokpol, S.; Parasuk, V.; Wolschann, P., Winger, R.H.; Liedl, K.R.; Rode, B.M. Comparative Molecular Field Analysis of Artemisinin Derivatives: Ab initio versus Semiempirical Optimized Structures, J. Comput-Aided Mol. Des. 12 (1998): 397-409.
46. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, revision A.11; [Computer Software]. Pennsylvania: Gaussian, 1998.

47. Richet, C. On the relationship between the toxicity and the physical properties of substances. Compt. Rend. Soc. Biol. 45 (1893): 775-776.
48. Hansch C. A. Quantitative Approach to Biochemical Structure-Activity Relationships. Acc. Chem. Res. 2 (1969):232-239
49. Free Jr., S. M.; and Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. J. Med. Chem. 7 (1964): 395-399.
50. Hansch, C.; and Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J. Am. Chem. Soc. 86 (1964): 1616-1626.
51. Kansy, M. Molecular properties. In: Structure-property correlations in drug research, Han van de Waterbeemd. New York: Academic Press; 1996. p. 19-26.
52. Trinajstić, N. Chemical Graph Theory. Vols. I, II. Florida: CRC Press, 1983.
53. Wiener, H. Correlation of Heats of Isomerization, and Differences in Heats of Vaporization of Isomers, Among the Paraffin Hydrocarbons. J. Am. Chem. Soc. 69 (1947): 2636-2638.
54. Hall, L. H.; and Kier, L.B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In K. B. Lipkowitz and D. B. Boyd (eds), Reviews in Computational Chemistry, Volume II, pp. 367-422. New York: VCH Publishers, 1991.
55. Oxford Molecular Groups. TSAR 3.2 [Computer Software]. Oregon: Oxford Molecular Groups, 1998.
56. Taylor, J.B.; Kennewell, P.D., editors. Modern medicinal chemistry. 1st ed. New York: Ellis Horwood; 1993. pp. 131-134.
57. Höltje, H.D., Folkers, G. Molecular modeling: Basic and principle and applications. In: Mannhold, R.; Kubinyi, H; Timmerman, H editors Methods and principles in medicinal chemistry. 5th ed. Weinheim: VCH; 1996. pp.. 1-39.
58. Leach A.R. editor. Molecular modeling: Principles and applications: London: Addison Wesley Longman; 1996. pp. 25-130.
59. Levine, I. N. Ab initio treatments of polyatomic molecules. In: Forsyth, N.; Petraities, P.; editors. Quantum chemistry. 4th ed. New Jersey(NJ): Prentice-Hall; 1991. pp. 455-544.

60. Lien, E. J.; Guo, Z. R.; Li, R. L. Use of dipole moments as a parameter in drug-receptor interactions and quantitative structure-activity relationship studies. J. Pharm. Sci. 71 (1982): 641-655.
61. Peter, C. J.; Steven, L.D.; Leanne, M. E. Molecular concepts. In: Han van de Waterbeemd, editor. 2nd ed. Chemometric methods in molecular design. Weinheim: VCH; 1995. pp. 31-32.
62. Kubinyi, H. Methods and Principles in Medicinal Chemistry. Vol. 1 QSAR: Hansch Analysis and Related Approaches. New York: VCH publishers, 1993.
63. Van de Waterbeemd, H. Quantitative Approaches to Structure-Activity Relationships. In C. G. Wermuth (ed.), The Practice of Medicinal Chemistry, pp. 367-389. London: Academic Press, 1996.
64. Stone, M. Cross-validators choice and assessment of statistical prediction. J. Royal Stat. Soc. B 36 (1974): 111-133.
65. Clementi, S.; and Wold, S. How to choose the proper statistical method. In H. Van der Waterbeemd (ed.), Chemometric Methods in Molecular Design, pp. 319-338. New York: VCH publishers, 1995.
66. Balaban, A.T. Highly discriminating distance-based topological index. Chem. Phys. Lett. 89 (1982): 399-404
67. Cramer III, R.D.; Patterson, D.E.; and Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. J. Am.Chem. Soc. 110 (1988): 5959-5967.
68. Wold, S.; Johansson, E.; and Cocchi, M. PLS - Partial Least-Squares Projections to Latent Structures. In H. Kubinyi (ed.), 3D-QSAR in Drug Design: Theory, Methods and Applications, pp. 523-550. The Netherlands: ESCOM, 1993.
69. Tripos. Sybyl6.8 [Computer Software]. Missouri: Tripos, 2000.
70. Posner, G. H.; Cumming, J.N.; Ploypradith, P.; and Oh. C.O. Evidence for Fe(IV)=O in the Molecular Mechanism of Action of the Trioxane Antimalarial Artemisinin. J. Am. Chem. Soc. 117 (1995): 5885-5886.
71. Jefford, C. W.; Vicente, M.G.H., Jacquier, Y., Favarger, F., Mareda, J., MillassonSchidt, P., Brunner, G., Burger, U. The Deoxygenation and

- Isomerization of Artemisinin and Artemether and Their Relevance to Antimalarial Action. Helv. Chim. Acta. 79 (1996): 1475-1487.
72. Cumming, J. N.; Ploypradith, P.; Posner, G. H. Antimalarial Activity of Artemisinin (Qinghaosu) and Related Trioxanes: Mechanism of Action. Adv. Pharmacol. 37 (1997): 2253-2297.
73. Posner, G. H.; Wang, D.; Cumming, J. N.; Oh, C. H.; French, A.N.; Bodley, A. L.; Shapiro, T. A. Further Evidence Supporting the Importance of and Restrictions on a Carbon-Centered Radical for High Antimalarial Activity of 1,2,4-Trioxanes Like Artemisinin. J. Med. Chem. 38 (1995): 2273-2275.
74. Posner, G. H.; Oh, C.H.; Wang, D.; Gerena, L.; Milhous, W. K.; Meshnick, S.R.; and Asawamahasadka, W. 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 (1994): 1256-1258.
75. Mehler, E.L. and Solmajer, T. Electrostatic Effects in Proteins: Comparison of Dielectric and Charge Models. Protein Eng. 4 (1991): 903-910.
76. Goodsell, D., Morris, G., Olson, A. "Automated Docking of Flexible Ligands: Applications of AutoDock." J. Mol. Rec. 9 (1996): 1-5.
77. Morris, G. M.; Goodsell, D.S.; Huey, R.; and Olson, A.J. AutoDock (version2.4) [Computer Software]. California, U.S.A, 1996.
78. Morris, G. M.; Goodsell, D.S.; Huey, R.; and Olson, A.J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock2.4. J. Comput.-Aided Mol. Des. 10 (1996): 293-304.
79. Goodsell, D.S.; and Olson, A.J. Automated Docking of Substrates to Proteins by Simulated Annealing. Proteins: Str. Func. and Genet. 8 (1990): 195-202.
80. Somsak Tonmunphean. Quantum Chemistry and QSAR of antimalarial artemisinin and its derivatives. Doctoral dissertation, Chemistry Department of Chemistry, Faculty of Science, Chulalongkorn University, 2000.
81. Morris, G. M.; Goodsell, D.S.; Huey, R.; and Olson, A.J. [Online]. 2000. Available from: <http://www.scripps.edu/pub/olson-web/doc/autoflex/parameters.html> [2001, June 5]



APPENDICES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table A1. Atomic net charges obtained from HF/6-31G(d) calculations and molecular properties of 32 derivative of tricyclic 1,2,4-trioxane compounds.

Compound	O ₁	O ₂	O ₁₃	C ₃	C ₄	C ₅	C _{5a}	C ₆
1	-0.392	-0.393	-0.745	0.608	-0.356	-0.330	-0.149	-0.318
2	-0.400	-0.396	-0.744	0.660	-0.359	-0.321	-0.149	-0.318
3	-0.399	-0.398	-0.744	0.663	-0.359	-0.321	-0.149	-0.318
4	-0.393	-0.402	-0.744	0.671	-0.362	-0.322	-0.150	-0.319
5	-0.397	-0.401	-0.756	0.711	-0.346	-0.323	-0.148	-0.319
6	-0.394	-0.403	-0.756	0.712	-0.347	-0.323	-0.148	-0.319
7	-0.393	-0.403	-0.756	0.713	-0.348	-0.324	-0.148	-0.319
8	-0.398	-0.401	-0.755	0.710	-0.345	-0.323	-0.148	-0.318
9	-0.396	-0.407	-0.753	0.719	-0.346	-0.323	-0.148	-0.319
10	-0.397	-0.401	-0.755	0.708	-0.346	-0.323	-0.148	-0.319
11	-0.391	-0.422	-0.747	0.718	-0.354	-0.325	-0.148	-0.319
12	-0.395	-0.402	-0.757	0.713	-0.346	-0.323	-0.148	-0.319
13	-0.397	-0.400	-0.756	0.711	-0.345	-0.323	-0.148	-0.319
14	-0.397	-0.401	-0.756	0.711	-0.346	-0.323	-0.148	-0.319
15	-0.398	-0.401	-0.755	0.710	-0.345	-0.323	-0.148	-0.318
16	-0.392	-0.390	-0.696	0.639	-0.355	-0.321	-0.149	-0.336
17	-0.394	-0.391	-0.693	0.639	-0.355	-0.321	-0.148	-0.336
18	-0.393	-0.390	-0.693	0.639	-0.355	-0.320	-0.148	-0.336
19	-0.388	-0.385	-0.674	0.635	-0.357	-0.338	-0.151	-0.354
20	-0.389	-0.385	-0.675	0.636	-0.357	-0.339	-0.150	-0.354
21	-0.387	-0.384	-0.675	0.634	-0.357	-0.338	-0.152	-0.354
22	-0.400	-0.396	-0.743	0.634	-0.210	-0.315	-0.150	-0.319
23	-0.402	-0.390	-0.738	0.626	-0.197	-0.332	-0.148	-0.318
24	-0.402	-0.390	-0.737	0.626	-0.203	-0.332	-0.148	-0.318
25	-0.408	-0.401	-0.736	0.641	-0.352	-0.321	-0.149	-0.317
26	-0.404	-0.388	-0.738	0.638	-0.354	-0.321	-0.147	-0.317
27	-0.406	-0.390	-0.736	0.638	-0.355	-0.321	-0.148	-0.317
28	-0.410	-0.400	-0.739	0.640	-0.191	-0.313	-0.149	-0.317
29	-0.403	-0.394	-0.740	0.638	-0.192	-0.312	-0.148	-0.318
30	-0.403	-0.393	-0.740	0.638	-0.192	-0.312	-0.147	-0.317
31	-0.405	-0.396	-0.739	0.638	-0.193	-0.313	-0.148	-0.317
32	-0.402	-0.394	-0.746	0.641	-0.203	-0.314	-0.149	-0.318

Table A1. (Continued)

Compound	C ₇	C ₈	C _{8a}	C ₁₂	C _{12a}	log P	(log P) ²
1	-0.320	-0.327	-0.319	0.576	0.195	2.724	7.420
2	-0.320	-0.327	-0.318	0.576	0.202	4.299	18.481
3	-0.320	-0.327	-0.318	0.576	0.201	5.185	26.884
4	-0.320	-0.328	-0.318	0.574	0.199	3.833	14.692
5	-0.320	-0.327	-0.318	0.575	0.201	6.025	36.301
6	-0.320	-0.327	-0.318	0.575	0.199	4.859	23.610
7	-0.320	-0.327	-0.318	0.575	0.199	5.224	27.290
8	-0.320	-0.327	-0.318	0.576	0.202	4.084	16.679
9	-0.320	-0.327	-0.318	0.575	0.201	5.343	28.548
10	-0.320	-0.327	-0.318	0.575	0.201	4.341	18.844
11	-0.320	-0.327	-0.318	0.576	0.201	4.947	24.473
12	-0.320	-0.327	-0.318	0.575	0.200	4.480	20.070
13	-0.320	-0.327	-0.318	0.575	0.201	6.000	36.000
14	-0.320	-0.327	-0.318	0.575	0.201	3.805	14.478
15	-0.320	-0.327	-0.318	0.576	0.202	3.805	14.478
16	-0.316	-0.344	-0.319	0.025	0.222	4.723	22.307
17	-0.316	-0.343	-0.318	0.025	0.223	3.855	14.861
18	-0.316	-0.343	-0.318	0.026	0.223	4.635	21.483
19	-0.317	-0.337	-0.334	-0.059	0.238	3.786	14.334
20	-0.317	-0.337	-0.332	-0.057	0.237	3.533	12.482
21	-0.318	-0.336	-0.335	-0.057	0.238	4.304	18.524
22	-0.320	-0.326	-0.318	0.580	0.202	4.327	18.723
23	-0.320	-0.327	-0.318	0.576	0.205	2.202	4.849
24	-0.320	-0.327	-0.318	0.575	0.204	4.397	19.334
25	-0.320	-0.318	-0.177	0.580	0.204	4.454	19.838
26	-0.319	-0.320	-0.179	0.579	0.193	4.314	18.611
27	-0.324	-0.330	-0.164	0.583	0.185	6.954	48.358
28	-0.320	-0.318	-0.175	0.579	0.208	4.956	24.562
29	-0.319	-0.320	-0.179	0.580	0.195	4.817	23.203
30	-0.319	-0.319	-0.178	0.579	0.195	2.762	7.629
31	-0.324	-0.329	-0.164	0.584	0.187	7.456	55.592
32	-0.318	-0.328	-0.172	0.579	0.196	4.375	19.141

จุฬาลงกรณ์มหาวิทยาลัย

Table A1. (Continued)

Compound	Dipole Moment	HOMO Energy	LUMO Energy	MR
1	2.947	0.21046	-0.41384	56.749
2	1.540	0.21299	-0.40543	75.283
3	1.795	0.15355	-0.31463	90.829
4	3.723	0.20562	-0.41456	66.917
5	1.717	0.11685	-0.30175	101.898
6	2.394	0.12610	-0.33770	81.567
7	4.296	0.11081	-0.35266	82.736
8	1.194	0.14316	-0.31775	88.329
9	1.580	0.10261	-0.28708	93.212
10	1.674	0.14183	-0.32900	76.762
11	2.893	0.13627	-0.32860	82.019
12	2.592	0.13194	-0.33399	76.978
13	2.597	0.13007	-0.32616	113.158
14	0.916	0.13476	-0.32650	83.578
15	0.916	0.14267	-0.31812	83.578
16	0.824	0.16903	-0.34739	88.971
17	1.696	0.18024	-0.33855	90.671
18	1.449	0.17767	-0.34161	84.251
19	4.197	0.08553	-0.36617	83.864
20	4.726	0.09800	-0.34671	90.327
21	3.217	0.07267	-0.37840	88.668
22	1.698	0.14308	-0.32281	81.575
23	3.006	0.21407	-0.40072	63.048
24	4.004	0.13490	-0.32669	92.628
25	3.714	0.14404	-0.31652	97.645
26	1.846	0.15227	-0.31657	97.428
27	2.045	0.13511	-0.32413	128.292
28	3.657	0.14502	-0.31787	102.061
29	1.848	0.15230	-0.31653	101.845
30	1.741	0.21614	-0.39984	72.481
31	2.174	0.13522	-0.32405	132.709
32	2.835	0.14479	-0.32130	97.177

จุฬาลงกรณ์มหาวิทยาลัย

Table A1. (Continued)

Compound	R(1-2)	R(2-3)	R(3-4)	R(4-5)	R(3-13)	R(12-13)
1	1.464	1.437	1.528	1.536	1.427	1.438
2	1.465	1.441	1.536	1.538	1.438	1.434
3	1.464	1.442	1.536	1.538	1.436	1.434
4	1.463	1.441	1.536	1.538	1.429	1.438
5	1.464	1.442	1.542	1.539	1.429	1.438
6	1.463	1.441	1.542	1.539	1.428	1.440
7	1.463	1.440	1.542	1.539	1.427	1.441
8	1.464	1.443	1.542	1.539	1.429	1.438
9	1.463	1.441	1.541	1.538	1.430	1.438
10	1.464	1.442	1.542	1.541	1.429	1.438
11	1.461	1.445	1.542	1.539	1.430	1.439
12	1.463	1.442	1.542	1.539	1.428	1.440
13	1.464	1.442	1.542	1.539	1.429	1.439
14	1.463	1.442	1.542	1.539	1.429	1.438
15	1.464	1.443	1.542	1.539	1.429	1.438
16	1.464	1.439	1.535	1.538	1.439	1.440
17	1.464	1.439	1.535	1.538	1.437	1.440
18	1.464	1.439	1.535	1.538	1.438	1.440
19	1.327	1.440	1.559	1.552	1.439	1.444
20	1.464	1.437	1.533	1.538	1.440	1.422
21	1.464	1.437	1.532	1.538	1.442	1.421
22	1.463	1.449	1.549	1.543	1.436	1.438
23	1.463	1.446	1.549	1.544	1.438	1.433
24	1.464	1.445	1.549	1.544	1.436	1.433
25	1.465	1.446	1.534	1.537	1.431	1.436
26	1.464	1.443	1.535	1.537	1.432	1.436
27	1.464	1.444	1.535	1.537	1.432	1.436
28	1.462	1.450	1.546	1.543	1.433	1.436
29	1.463	1.447	1.547	1.543	1.434	1.436
30	1.463	1.447	1.547	1.543	1.434	1.436
31	1.463	1.448	1.547	1.543	1.434	1.436
32	1.463	1.448	1.549	1.546	1.438	1.435

จุฬาลงกรณ์มหาวิทยาลัย

Table A1. (Continued)

Compound	R(8a-12a)	R(8-8a)	R(12a-12)	A(4-5-5a)	A(3-4-5)
1	1.532	1.538	1.530	114.1	111.6
2	1.532	1.538	1.529	114.2	112.3
3	1.532	1.538	1.529	114.2	112.3
4	1.532	1.538	1.529	114.1	112.1
5	1.532	1.538	1.530	114.6	112.4
6	1.532	1.538	1.530	114.5	112.3
7	1.532	1.538	1.530	114.5	112.2
8	1.532	1.538	1.530	114.6	112.4
9	1.532	1.538	1.530	114.6	112.5
10	1.532	1.538	1.530	114.6	112.4
11	1.532	1.538	1.530	114.5	112.6
12	1.532	1.538	1.530	114.6	112.4
13	1.532	1.538	1.530	114.6	112.4
14	1.532	1.538	1.530	114.6	112.4
15	1.532	1.538	1.530	114.6	112.4
16	1.536	1.540	1.534	113.8	112.3
17	1.536	1.540	1.534	113.8	112.2
18	1.536	1.540	1.534	113.8	112.2
19	1.566	1.550	1.575	113.0	112.4
20	1.538	1.538	1.537	113.2	112.5
21	1.538	1.539	1.538	113.2	112.5
22	1.532	1.538	1.529	115.4	111.7
23	1.531	1.538	1.527	115.6	110.7
24	1.531	1.538	1.527	115.4	110.6
25	1.545	1.544	1.532	114.9	112.0
26	1.543	1.541	1.532	114.4	111.9
27	1.544	1.540	1.533	114.4	112.0
28	1.543	1.545	1.531	116.7	110.3
29	1.542	1.541	1.531	116.2	110.4
30	1.542	1.541	1.531	116.2	110.4
31	1.543	1.540	1.532	116.2	110.4
32	1.543	1.542	1.528	116.2	109.8

จุฬาลงกรณ์มหาวิทยาลัย

Table A1. (Continued)

Compound	A(13-3-4)	A(12-13-3)	A(8a-12a-5a)	A(2-3-4)	A(8-8a-12a)
1	112.1	114.1	111.9	112.5	113.2
2	109.9	114.4	112.0	112.4	113.2
3	110.1	114.5	112.0	111.8	113.1
4	110.6	114.2	112.0	111.8	113.1
5	109.7	113.9	112.0	111.7	113.1
6	109.9	113.8	112.0	111.8	113.1
7	110.0	113.8	112.0	111.9	113.1
8	109.7	113.9	112.0	111.0	113.1
9	109.7	113.8	112.0	111.8	113.1
10	109.7	113.9	112.0	111.7	113.1
11	110.2	114.1	112.0	111.2	113.1
12	109.8	113.9	112.0	111.7	113.1
13	109.7	113.9	112.0	111.7	113.1
14	109.7	113.9	112.0	111.7	113.1
15	109.7	113.9	112.0	111.6	113.1
16	110.0	113.5	111.8	112.1	114.3
17	109.9	113.6	111.8	112.1	114.3
18	109.9	113.6	111.8	112.1	114.4
19	109.7	111.9	111.4	112.5	115.3
20	109.6	114.7	111.9	113.1	115.3
21	109.6	114.5	112.0	113.2	115.2
22	110.8	114.6	112.0	111.0	113.2
23	110.0	114.5	112.0	111.0	113.1
24	110.0	114.5	112.0	110.9	113.1
25	110.4	114.4	111.5	111.3	111.4
26	110.4	114.3	111.2	111.6	111.2
27	110.5	114.5	111.0	111.6	111.5
28	110.2	114.5	111.6	111.2	111.3
29	110.2	114.4	111.4	111.3	111.4
30	110.2	114.5	111.4	111.3	111.4
31	110.3	114.6	111.1	111.4	111.5
32	109.7	114.2	110.9	111.1	112.0

จุฬาลงกรณ์มหาวิทยาลัย

Table A1. (Continued)

Compound	A(12a-12-13)	A(13-3-2)	A(3-2-1)	T(12a-5a-5-4)	T(3-4-5-5a)
1	111.8	106.9	108.7	44.4	62.3
2	111.8	106.1	109.0	44.5	63.0
3	111.8	106.3	109.1	44.3	63.1
4	111.8	106.9	109.0	44.3	62.9
5	111.7	107.1	108.7	43.3	62.2
6	111.7	107.2	108.7	43.4	62.2
7	111.6	107.3	108.7	43.5	62.2
8	111.7	107.0	108.7	43.2	62.2
9	111.7	107.0	108.7	43.3	62.1
10	111.7	107.0	108.7	43.3	62.2
11	112.0	106.1	109.0	43.1	62.8
12	111.7	107.2	108.7	43.4	62.2
13	111.7	107.1	108.7	43.3	62.2
14	111.7	107.1	108.7	43.3	62.2
15	111.7	107.0	108.7	43.2	62.2
16	111.4	106.9	109.0	46.2	63.2
17	111.3	107.0	108.9	46.2	63.2
18	111.3	107.0	108.9	46.2	63.2
19	111.1	110.0	113.2	48.5	57.0
20	111.9	106.0	109.6	50.3	64.1
21	112.0	105.9	109.6	50.2	64.0
22	112.0	106.2	110.1	43.5	61.6
23	111.6	106.6	109.7	43.9	62.5
24	111.5	106.7	109.6	44.0	62.7
25	112.1	106.6	108.2	43.2	62.4
26	111.5	106.7	108.5	43.2	62.6
27	111.6	106.6	108.4	43.3	62.8
28	112.0	106.6	109.0	42.9	61.2
29	111.4	106.7	109.3	43.0	61.5
30	111.5	106.7	109.3	43.1	61.5
31	111.5	106.6	109.3	43.1	61.7
32	111.0	106.7	109.7	42.9	61.6

Table A1. (Continued)

Compound	T(13-3-4-5)	T(4-3-13-12)	T(5-4-3-2)	T(8-8a-12a-5a)
1	22.6	90.4	97.9	51.3
2	21.5	91.0	96.4	51.3
3	21.2	90.8	96.7	51.3
4	21.8	90.9	97.2	51.3
5	22.1	92.0	96.5	51.4
6	22.3	92.0	96.6	51.3
7	22.4	92.0	96.7	51.4
8	22.0	92.0	96.5	51.4
9	22.2	91.9	96.4	51.4
10	22.0	92.0	96.5	51.4
11	21.0	90.7	96.3	51.4
12	22.2	92.0	96.5	51.3
13	22.0	92.0	96.5	51.3
14	22.0	92.0	96.5	51.3
15	22.0	92.0	96.5	51.4
16	22.1	93.3	96.6	48.5
17	22.1	93.4	96.7	48.5
18	22.0	93.4	96.7	48.5
19	32.5	97.5	90.2	43.4
20	22.3	93.0	95.7	45.2
21	22.3	93.2	95.6	45.2
22	22.2	90.9	95.6	51.3
23	21.5	92.7	96.3	51.3
24	21.2	92.7	96.6	51.3
25	21.4	91.6	96.9	51.6
26	21.1	91.7	97.4	53.6
27	21.0	91.1	97.5	54.5
28	22.6	93.6	95.4	51.6
29	22.3	93.6	95.9	53.5
30	22.3	93.5	95.9	53.4
31	22.2	93.1	96.0	54.5
32	21.5	95.5	96.2	52.7

Table A1. (Continued)

Compound	T(1-2-3-13)	T(3-2-1-12a)	T(13-12-12a-1)	T(12a-12-13-3)
1	71.2	44.0	58.7	31.3
2	73.8	46.4	58.9	33.6
3	73.4	45.9	58.9	33.6
4	73.5	45.2	58.8	33.2
5	73.5	46.7	59.2	35.0
6	73.6	46.4	59.1	34.7
7	73.7	46.4	59.1	34.6
8	73.4	46.7	59.3	35.1
9	73.7	46.8	59.1	34.9
10	73.5	46.7	59.2	35.0
11	74.4	47.8	58.9	34.6
12	73.6	46.6	59.1	34.9
13	73.5	46.7	59.2	35.0
14	73.5	46.7	59.2	35.0
15	73.4	46.7	59.3	35.1
16	72.9	45.4	61.8	37.3
17	72.8	45.3	61.8	37.3
18	72.8	45.4	61.8	37.4
19	68.5	38.3	59.8	32.7
20	70.8	41.9	63.0	37.2
21	70.9	42.1	63.3	37.6
22	72.3	44.8	59.3	33.7
23	71.8	45.9	59.8	35.9
24	71.8	46.0	59.8	36.1
25	74.1	48.7	57.4	34.6
26	73.5	46.8	58.8	34.5
27	73.6	46.6	58.4	33.8
28	72.8	48.8	58.2	36.5
29	72.4	47.0	59.5	36.3
30	72.3	46.9	59.4	36.2
31	72.4	46.8	59.2	35.8
32	71.4	47.4	61.3	39.4

Table A1. (Continued)

Compound	T(12-13-3-2)	T(4-3-2-1)
1	33.3	49.2
2	30.7	46.3
3	30.6	46.8
4	31.1	47.6
5	29.4	46.6
6	29.7	46.9
7	29.9	47.0
8	29.3	46.6
9	29.6	46.4
10	29.4	46.6
11	29.8	45.4
12	29.5	46.7
13	29.4	46.6
14	29.4	46.6
15	29.3	46.6
16	28.6	47.6
17	28.5	47.7
18	28.4	47.7
19	26.7	54.1
20	29.4	49.3
21	29.2	49.1
22	29.7	48.2
23	27.8	48.0
24	27.7	48.0
25	29.5	46.4
26	29.8	47.2
27	30.4	47.2
28	27.2	47.3
29	27.5	48.0
30	27.5	48.0
31	28.0	48.0
32	24.9	48.1

Table A2. Atomic charge of heme were obtained at HF/6-311G(d,p) level.

Number	Atom	Charge	Number	Atom	Charge
1	C	0.358	43	H	0.110
2	C	-0.284	44	H	0.110
3	C	0.403	45	H	0.072
4	C	-0.099	46	H	0.125
5	C	-0.131	47	H	0.114
6	C	0.354	48	H	0.050
7	C	-0.278	49	H	0.118
8	C	0.321	50	H	0.093
9	C	-0.034	51	H	0.040
10	C	-0.151	52	H	0.086
11	C	0.335	53	H	0.104
12	C	-0.100	54	N	-0.912
13	C	-0.276	55	N	-0.967
14	C	-0.123	56	N	-0.962
15	C	-0.187	57	N	-0.896
16	C	-0.188	58	O	-0.632
17	C	-0.138	59	O	-0.631
18	C	-0.204	60	O	-0.588
19	C	-0.118	61	O	-0.639
20	C	0.390	62	H	0.090
21	C	-0.203	63	H	0.106
22	C	-0.120	64	H	0.085
23	C	-0.140	65	H	0.104
24	C	0.431	66	H	0.108
25	C	-0.154	67	H	0.098
26	C	-0.145	68	H	0.094
27	C	-0.199	69	H	0.125
28	C	-0.183	70	H	0.095
29	C	0.343	71	H	0.100
30	C	-0.287	72	H	0.102
31	C	0.368	73	H	0.077
32	C	-0.150			
33	C	-0.105			
34	C	0.393			
35	Fe	1.539			
36	H	0.099			
37	H	0.201			
38	H	0.110			
39	H	0.092			
40	H	0.123			
41	H	0.109			
42	H	0.045			



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CURRICULUM VITAE

Koonwadee Ratanasak

- Born** June 5th, 1979 in Ranong, Thailand
- Education**
- 1991-1993 Primary School
Rat Borikanukroh, Ratchaburi, Thailand
- 1994-1996 Secondary School
Rat Borikanukroh, Ratchaburi, Thailand
- 1997-2000 Bachelor of General Science, Minor of Chemistry,
Department of Science, Faculty of Liberal Arts & Science,
Kasetsart University, Nakornprathom, Thailand
- 2001-2003 Master of degree student in Chemistry
Department of Chemistry, Faculty of Science,
Chulalongkorn University, Bangkok, Thailand



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Proceedings and Conferences

1. Koonwadee Ratanasak, Vudhichai Parasuk, Sirirat Kokpol, "Automated Docking Calculations of 1,2,4-Trioxane Derivatives to Heme", The 10th Annual Scientific Symposium of Faculty of Science, Chulalongkorn University, March, 20-22, 2002, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.
2. Koonwadee Ratanasak, Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "The Study of Interaction Between 1,2,4-Trioxane Derivatives and Heme by Docking Calculations", 28th Congress on Science and Technology of Thailand (STT28), October, 24-26, 2002, Queen Sirikit National Convention Center or QSNCC, Bangkok, Thailand. 02-21-P.
3. Koonwadee Ratanasak, Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Automated Docking Calculation of 1,2,4-Trioxane Derivatives to Heme", International Conference & Exhibition on Pure and Applied Chemistry 2002 (PACCON 2002), 29-31 May 2002, Amari Watergate Hotel, Bangkok, Thailand, 31-BCDD-11, p.121.
4. Koonwadee Ratanasak, Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol. "Automated Docking Calculations of 1,2,4-Trioxane Derivatives to Heme", The Sixth Annual National Symposium on Computational Science and Engineering (ANSCSE6), April, 3-5, 2002, Walailuk University, Nakhon Sri Thammarat, Thailand.
5. Koonwadee Ratanasak, Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol. "Binding of 1,2,4-trioxane compounds to heme and its relation with antimalarial activity", The Seventh Annual National Symposium on Computational Science and Engineering (ANSCSE7), March, 24-26, 2003, Chulalongkorn University, Bangkok, Thailand.
6. Koonwadee Ratanasak, Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol. "Three-Dimensional Quantitative Structure-Activity Relationship (3D-QSAR) Studies of Antimalarial Tricyclic 1,2,4-trioxanes", 29th Congress on Science and Technology of Thailand (STT29), October, 20-22, 2003, Golden Jubilee Convention Hall, Khon Kaen University, Khon Kaen.