

QUANTUM CHEMISTRY AND QSAR OF ANTIMALARIAL
ARTEMISININ AND ITS DERIVATIVES



Mr. Somsak Tonmunphean

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Chemistry

Department of Chemistry

Faculty of Science

Chulalongkorn University

Academic Year 2000

ISBN 974-13-0920-1

เคมีควอนตัมและความสัมพันธ์เชิงปริมาณระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพ

ของยาต้านมาลาเรียอาร์ติมิซินินและอนุพันธ์



นายสมศักดิ์ ตนหมั่นเพียร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาเคมี ภาควิชาเคมี

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2543

ISBN 974-13-0920-1

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title Quantum Chemistry and QSAR of Antimalarial
Artemisinin and Its Derivatives
By Mr. Somsak Tonmunphean
Field of Study Chemistry
Thesis Advisor Associate Professor Sirirat Kokpol, Ph.D.
Thesis Co-Advisor Associate Professor Vudhichai Parasuk, Ph.D.
Professor Karl Peter Wolschann, Ph.D.

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

..... Dean of Faculty of Science
(Associate Professor Wanchai Phothiphichitr, 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.)

..... Thesis Co-Advisor
(Professor Karl Peter Wolschann, Ph.D.)

..... Member
(Associate Professor Jerapan Krungkrai, Ph.D.)

..... Member
(Assistant Professor Supa Hannongbua, Dr.rer.nat.)

..... Member
(Assistant Professor Tirayut Vilaivan, D.Phil.)

สมศักดิ์ ตนหมั่นเพียร: เคมีควอนตัมและความสัมพันธ์เชิงปริมาณระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพของยาต้านมาลาเรียอาร์ติมิซินินและอนุพันธ์. (QUANTUM CHEMISTRY AND QSAR OF ANTIMALARIAL ARTEMISININ AND ITS DERIVATIVES) อ.ที่ปรึกษา: รศ. ดร.ศิริรัตน์ กักผล, อ.ที่ปรึกษาร่วม: รศ.ดร.วุฒิชัย พาราสุข, ศ.ดร.คาร์ล ปีเตอร์ ไวลชาน, 231 หน้า. ISBN 974-13-0920-1.

ได้ทำการศึกษาความสัมพันธ์เชิงปริมาณระหว่างโครงสร้างกับฤทธิ์ทางชีวภาพของยาต้านมาลาเรียอาร์ติมิซินินและอนุพันธ์จำนวน 104 สาร โดยปรับโครงสร้างของสารทั้งหมดด้วยวิธี HF/3-21G ใช้ฤทธิ์ทางชีวภาพของสารเหล่านี้ที่ทดสอบกับเชื้อมาลาเรีย 2 สายพันธุ์ที่แตกต่างกัน ได้แก่ D-6 และ W-2 จากวิธีการ QSAR แบบดั้งเดิม ได้แบบจำลองที่มีประสิทธิภาพในการทำนายที่ดีถึงดีมากสำหรับฤทธิ์ทางชีวภาพของทั้งสองสายพันธุ์ แบบจำลองเหล่านี้สามารถใช้ทำนายฤทธิ์ทางชีวภาพของสารในกลุ่มทดสอบได้ผลดี โดยให้ค่าที่ใกล้เคียงกับค่าจากการทดลอง อย่างไรก็ตามแบบจำลองที่สร้างจากวิธีการ CoMFA (3D-QSAR) มีประสิทธิภาพในการทำนายที่ระดับต่ำถึงปานกลาง สาเหตุเนื่องจากวิธีการนี้ไม่สามารถอธิบายแรงกระทำในสารประกอบเชิงซ้อนของฮีมกับอาร์ติมิซินิน ดังนั้นจึงได้ศึกษากลไกการออกฤทธิ์ของสารเหล่านี้โดยการคำนวณทางโมเลกุลาร์ดีอกกิ่งและการคำนวณทางเคมีควอนตัมแบบ IMOMO(B3LYP/6-31G**):HF/3-21G) ผลที่ได้บ่งชี้ว่า Fe^{2+} จะเข้าจับกับสารกลุ่มอาร์ติมิซินินที่อะตอม O_1 ได้ดีกว่าที่อะตอม O_2 นอกจากนี้ยังพบว่าสารที่มีฤทธิ์สูงจะมีแผนภาพพลังงานในปฏิกิริยาการออกฤทธิ์ที่แตกต่างไปจากสารที่มีฤทธิ์ต่ำอย่างมีนัยสำคัญ สุดท้ายได้ทำการเสนอแนวทางในการปรับปรุงโครงสร้างเพื่อให้มีฤทธิ์ที่สูงขึ้น

สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....เคมี.....	ลายมือชื่อนิสิต.....
สาขาวิชา.....เคมี.....	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2543.....	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....
	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

KEYWORD: QSAR / CoMFA / Molecular Docking / Mechanism of Action

SOMSAK TONMUNPHEAN: QUANTUM CHEMISTRY AND QSAR OF ANTIMALARIAL ARTEMISININ AND ITS DERIVATIVES. THESIS ADVISOR: ASSOC. PROF. SIRIRAT KOKPOL, Ph.D., THESIS CO-ADVISOR: ASSOC. PROF. VUDHICHAIR PARASUK, Ph.D., PROF. KARL PETER WOLSCHANN, Ph.D., 231 pp. ISBN 974-13-0920-1.

The quantitative structure-activity relationships (QSAR) of 104 antimalarial artemisinin and its derivatives were studied. All compounds were geometrically optimized at the HF/3-21G level. The activities measured against 2 different strains of malarial parasites, D-6 and W-2, were used. Models with good to excellent predictive ability were obtained from traditional QSAR method for both activities. The models were shown to predict activities of compounds in the test set very close to the experimental values. However, the derived CoMFA (3D-QSAR) models have only low to moderate predictive power since this method can not explain interactions in heme-artemisinin complex. Therefore, the mechanism of action of these compounds was investigated by means of molecular docking and quantum chemical calculations using the IMOMO (B3LYP/6-31G**: $\text{HF}/3\text{-}21\text{G}$) method. The obtained results reveal that Fe^{2+} approach artemisinin compounds at the O_1 atom more preferably than the O_2 atom. Moreover, we discovered that the high activity compounds have significantly different energy profiles in the reaction mechanism from the low activity compounds. Finally, information on how to enhance the activities were suggested.

Department...Chemistry.....

Field of study...Chemistry.....

Academic year...2000.....

Student's signature.....

Advisor's signature.....

Co-advisor's signature.....

Co-advisor's signature.....

ACKNOWLEDGEMENT

First, I would like to give my deep gratitude to my parents for their exceedingly understanding, encouragement, and supports.

I wish to express my sincere thanks to Associate Professor Dr. Sirirat Kokpol, Associate Professor Dr. Vudhichai Parasuk, Professor Dr. Karl Peter Wolschann, and Professor Dr. Bernd Michael Rode for useful guidance, kind helps, and valuable suggestions throughout my study.

I am grateful to Associate Professor Dr. Siri Varothai, Associate Professor Dr. Jerapan Krungkrai, Assistant Professor Dr. Supa Hannongbua, and Assistant Professor Dr. Tirayut Vilaivan for their advice as the thesis committee.

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 Unit Cell, Chulalongkorn University, Thailand, and the Computer Center of the University of Vienna, Austria for computer resources and other facilities.

Finally, I gratefully acknowledge the financial supports from the Austrian Federal Ministry for Foreign Affairs (ÖAD scholarship) and the ASEA-UNINET during my stay in Austria, and the Graduate School, Chulalongkorn University for my presentations in two international symposiums. In addition, the Research Assistantship from the Graduate School and the Research Division, Chulalongkorn University, the Professor Dr. Tab Nilanidhi Foundation scholarship, and the Rhone Poulenc-Professor Lehn scholarship are very appreciated.

CONTENTS

	Pages
ABSTRACT IN THAI.....	iv
ABSTRACT IN ENGLISH.....	v
ACKNOWLEDGEMENT.....	vi
LIST OF FIGURES.....	xii
LIST OF TABLES.....	xvii
LIST OF ABBREVIATIONS.....	xx
CHAPTER 1 INTRODUCTION.....	1
1.1 Malaria.....	2
1.1.1 Historical Outline.....	2
1.1.2 Life Cycle of Malarial Parasites.....	2
1.1.3 Malaria Symptoms.....	3
1.1.4 Malaria Control.....	4
1.1.4.1 Malaria Control in the Past.....	4
1.1.4.2 Malaria Control at the Present.....	6
1.1.4.3 Malaria Control in the Future.....	6
1.2 Antimalarial Drugs.....	8
1.2.1 Historical Outline.....	8
1.2.2 Classification of Antimalarial Drugs.....	11
1.3 Artemisinin and its Derivatives.....	12
1.3.1 Historical Outline.....	12
1.3.2 Chemical Structure.....	12
1.3.3 Antimalarial Activity.....	13
CHAPTER 2 STRUCTURE AND BIOLOGICAL DATA.....	16
2.1 Compounds Group 1.....	16
2.2 Compounds Group 2.....	19
2.3 Compounds Group 3.....	25
2.4 Structural Optimization.....	28

	Pages
2.5 Conformational Problem.....	31
CHAPTER 3 TRADITIONAL QSAR.....	35
3.1 Introduction.....	35
3.1.1 Historical Outlines of QSAR.....	37
3.1.2 Physicochemical Properties.....	39
3.1.2.1 Hydrophobicity Parameters.....	40
3.1.2.2 Polarizability Parameters.....	41
3.1.2.3 Electronic Parameters.....	41
3.1.2.4 Steric Parameters.....	42
3.1.3 Regression Analysis.....	43
3.1.4 Previous QSAR Studies.....	47
3.2 Calculations of Properties.....	51
3.3 Results and Discussion.....	52
3.3.1 All 104 Compounds.....	53
3.3.2 Compounds with C=O at C ₁₀ position (Group 1).....	55
3.3.3 Compounds with -OR or -NR at C ₁₀ position (Group 2).....	58
3.3.4 Compounds without a substituent at C ₁₀ position (Group 3)...	61
3.4 Summary.....	63
CHAPTER 4 COMPARATIVE MOLECULAR FIELDS ANALYSIS.....	65
4.1 Introduction.....	65
4.1.1 Methods of CoMFA Studies.....	65
4.1.1.1 Setup Step.....	65
4.1.1.2 Calculation Step.....	66
4.1.1.3 Statistical Analysis Step.....	68
4.1.1.4 Interpretation Step.....	69
4.1.2 Previous CoMFA Studies.....	69
4.2 Computational Methods.....	71
4.2.1 Structure and Charge Calculations.....	71
4.2.2 Alignment Rules.....	71

	Pages
4.2.3 CoMFA Calculations.....	72
4.2.4 Partial Least Squares Regression Analysis	73
4.3 Results and Discussions.....	73
4.3.1 Effect of Structure and Charge.....	73
4.3.2 Effect of Alignment Rule.....	74
4.3.3 Effect of Type and Charge of Probe Atom.....	75
4.3.4 Effect of Steric and Electrostatic Cut-off.....	76
4.3.5 Effect of Grid Position.....	77
4.3.6 q^2 -GRS approach.....	78
4.3.7 All 104 Compounds.....	79
4.3.8 Compounds with C=O at C ₁₀ position.....	85
4.3.9 Compounds without C=O at C ₁₀ position.....	88
4.4 Summary.....	91
 CHAPTER 5 MOLECULAR DOCKING.....	 93
5.1 Introduction.....	93
5.2 Computational Method.....	95
5.3 Determination of Suitable Docking Parameters.....	97
5.3.1 Force Field Parameters.....	97
5.3.2 Temperature Reduction Rate, Starting and Final Temperature.	98
5.3.3 Grid Spacing and Grid Dimension.....	100
5.3.4 Cluster Tolerance Value.....	102
5.3.5 Dielectric Constant.....	102
5.3.6 Atomic Charges Calculation Method.....	103
5.3.7 Optimization Method for Artemisinin Compounds	105
5.3.8 Heme Structures.....	106
5.3.9 Comparison with Experimental Data.....	119
5.4 Docking Calculations.....	121
5.4.1 Heme-hemin and artemisinin derivatives group 1.....	121
5.4.2 Heme-hemin and artemisinin derivatives group 2.....	123
5.4.3 Heme-hemin and artemisinin derivatives group 3.....	126

	Pages
5.5 Summary.....	128
CHAPTER 6 MECHANISM OF ACTION.....	131
6.1 Introduction.....	131
6.2 Computational Methods.....	135
6.2.1 Theoretical Background on the IMOMO Approach.....	135
6.2.2 Calculations Details.....	136
6.3 Results and Discussions.....	136
6.3.1 Comparison of Structural Data between the IMOMO Optimized Structure and the X-ray Structure.....	136
6.3.2 Comparison of all three pathways.....	138
6.3.3 Direct O-O bond breaking.....	138
6.3.4 Reaction Mechanism.....	140
6.3.4.1 Intermolecular 1,5-Hydrogen Shift (Pathway 1).....	142
6.3.4.2 Homolytic C-C Cleavage (Pathway 2).....	149
6.3.4.3 Comparison between the Two Pathways.....	155
6.3.4.4 Relationship to Biological Activity.....	158
6.4 Summary.....	161
CHAPTER 7 CONCLUSIONS.....	163
7.1 QSAR Models.....	163
7.2 Mechanism of Action.....	164
7.3 Effect of Substituent Group to Activity.....	165
7.4 Suggestions for New Compound.....	166
7.5 Suggestions for Further Works.....	166
REFERENCES.....	168
APPENDICES.....	181
Appendix A Publication I.....	182
Appendix B Publication II.....	196

	Pages
Appendix C Publication III.....	201
Appendix D Publication IV.....	211
CURRICULUM VITAE.....	228



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF FIGURES

Figures	Pages
1.1 The life cycle of <i>Plasmodium</i>	3
1.2 Structures of antimalarial drugs (a) quinine, (b) chloroquine, (c) primaquine, (d) pyrimethamine, (e) mefloquine.....	10
1.3 Stereochemistry and atomic numbering scheme of artemisinin.....	13
1.4 Structures of artemisinin derivatives (a) dihydroartemisinin, (b) artemether, (c) arteether, (d) artesunate.....	14
2.1 Structures of artemisinin derivatives number 1 to 31.....	17
2.2 Structures of artemisinin derivatives number 32 to 41.....	18
2.3 Structural difference between compound in group 1 (blue) and 2 (red).....	19
2.4 Structures of artemisinin derivatives number 42 to 60.....	20
2.5 Structures of artemisinin derivatives number 61 to 74.....	21
2.6 Structures of artemisinin derivatives number 75 to 78.....	23
2.7 Structures of artemisinin derivatives number 79 to 82.....	23
2.8 Structures of artemisinin derivatives number 83 to 87.....	24
2.9 Structures of artemisinin derivatives number 88 to 104.....	25
2.10 Superimposed structures of no search (blue), MM3 (black), and Tripos (red) conformations for compounds (a) 21, (b) 29, (c) 44, and (d) 61.....	32
3.1 Steps in drug discovery and development process.....	36
3.2 The common structure of artemisinin derivatives used in Ref. [84].....	48
3.3 Comparison between actual and predicted log(D-6) activities for 10 compounds in the testing set.....	54
3.4 Comparison between actual and predicted log(W-2) activities for 10 compounds in the testing set.....	54
3.5 Comparison between actual and predicted log(D-6) activities for 4 compounds in the testing set.....	55
3.6 Comparison between actual and predicted log(W-2) activities for 4 compounds in the testing set.....	56

Figures	Pages
3.7 Relation between actual and predicted log(D-6) activities for 36 compounds in group 2.....	59
3.8 Relation between actual and predicted log(W-2) activities for 36 compounds in group 2.....	59
3.9 Comparison between actual and predicted log(D-6) activities for 5 (model 3.45) and 4 (model 3.47) compounds in the testing set.....	60
3.10 Comparison between actual and predicted log(W-2) activities for 5 (model 3.46) and 4 (model 3.48) compounds in the testing set.....	60
3.11 Relation between actual and predicted log(D-6) activities for 14 compounds in group 3.	62
3.12 Relation between actual and predicted log(W-2) activity for 14 compounds in group 3.	62
4.1 The superimposition of all artemisinin compounds (hydrogen atoms are not shown).	66
4.2 Steps in the CoMFA calculations.....	67
4.3 Definition of five alignment rules used in CoMFA studies.....	72
4.4 Steric contour plot of the log(D-6) model for all compounds	81
4.5 Electrostatic contour plot of the log(D-6) model for all compounds	81
4.6 Steric contour plot of the log(W-2) model for all compounds	83
4.7 Electrostatic contour plot of the log(W-2) model for all compounds	83
4.8 Comparison between actual and predicted log(D-6) activities for 10 compounds in the testing set	84
4.9 Comparison between actual and predicted log(W-2) activities for 10 compounds in the testing set	84
4.10 Electrostatic contour plot of the log(D-6) model for compounds with C=O at the C ₁₀ position.....	86
4.11 Comparison between actual and predicted log(D-6) activities for 4 compounds with C=O at the C ₁₀ position in the testing set	87
4.12 Steric contour plot of the log(D-6) model for compounds without C=O at the C ₁₀ position.....	89

Figures	Pages
4.13 Steric contour plot of the log(W-2) model for compounds without C=O at the C ₁₀ position.....	90
4.14 Comparison between actual and predicted log(D-6) activities for 6 compounds without C=O at the C ₁₀ position in the testing set	91
4.15 Comparison between actual and predicted log(W-2) activities for 6 compounds without C=O at the C ₁₀ position in the testing set	91
5.1 Proposed mechanism of action of artemisinin compound.....	94
5.2 Grid base energy evaluation.....	96
5.3 Structures of 5 heme molecules (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, (e) heme-oxy.....	107
5.4 Superimposed docking configuration between the heme-pdb and 16 artemisinin derivatives (without hydrogen atoms).....	109
5.5 Superimposed docking configuration between the heme-model and 16 artemisinin derivatives (without hydrogen atoms).....	111
5.6 Superimposed docking configuration between the heme-hemin and 16 artemisinin derivatives (without hydrogen atoms).....	113
5.7 Superimposed docking configuration between the heme-deoxy and 16 artemisinin derivatives (without hydrogen atoms).....	115
5.8 Superimposed docking configuration between the heme-oxy and 16 artemisinin derivatives (without hydrogen atoms).....	116
5.9 Relationship between binding energies and activities for 16 artemisinin compounds using (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, and (e) heme-oxy.....	118
5.10 Structure of four artemisinin compounds used in Ref. [110].....	120
5.11 Relationship between binding energies and dissociation constants (K _d)....	120
5.12 Relationship between binding energies and activities for 40 artemisinin derivatives in group 2.....	126
5.13 Relationship between binding energies and activities for 16 artemisinin derivatives in group 3.....	128

Figures	Pages
6.1 Structure of heme molecule.....	132
6.2 The mechanism of action of artemisinin compound as proposed by Posner et al. (pathway 1) [115].....	132
6.3 The mechanism of action of artemisinin compound as proposed by Jefford et al. (pathway 2) [116].....	133
6.4 The mechanism of action of artemisinin compound as proposed by Haynes et al. (pathway 3) [123].....	134
6.5 The extrapolation scheme of IMOMO method.....	135
6.6 Labeled atoms that are used as the model part for IMOMO calculation....	136
6.7 Proposed mechanism of action of artemisinin compound.....	141
6.8 Optimized structures of (a) the O-centered radical 1A, (b) the transition state TS ₁ , and (c) the C-centered radical 1B for 15 artemisinin compounds.....	142
6.9 Optimized structures of (a) the O-centered radical 2A, (b) the transition state TS ₂ , and (c) the C-centered radical 2B for 15 artemisinin compounds.....	149
6.10 Energy scheme for the mechanism of action for artemisinin.....	156
6.11 Schematic energy scheme for the mechanism of action for 9 high activity artemisinin derivatives.....	156
6.12 Schematic energy scheme for the mechanism of action for 3 low activity artemisinin derivatives (except compound 104 that has 1A energy slightly lower than 2A energy).	157
6.13 Relationship between log(D-6) activity and EA ₁ for 12 artemisinin derivatives.	158
6.14 Relationship between log(W-2) activity and EA ₁ for 12 artemisinin derivatives.	159
6.15 Relationship between log(D-6) activity and ΔE ₁ for 12 artemisinin derivatives.	159
6.15 Relationship between log(W-2) activity and ΔE ₁ for 12 artemisinin derivatives.....	160

Figures	Pages
6.17 Relationship between $\log(D-6)$ activity and $\Delta E(1A-2A)$ for 12 artemisinin derivatives.....	160
6.18 Relationship between $\log(W-2)$ activity and $\Delta E(1A-2A)$ for 12 artemisinin derivatives.....	161



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF TABLES

Tables	Pages
2.1 Structures and biological data of compounds number 1-31 in group 1.....	17
2.2 Structures and biological data of compounds number 32-41 in group 1....	18
2.3 Structures and biological data of compounds number 42-60 in group 2....	20
2.4 Structures and biological data of compounds number 61-74 in group 2....	22
2.5 Structures and biological data of compounds number 75-78 in group 2....	23
2.6 Structures and biological data of compounds number 79-82 in group 2....	24
2.7 Structures and biological data of compounds number 83-87 in group 2....	24
2.8 Structures and biological data of compounds number 88-104 in group 3...	25
2.9 Effect of substituent at C ₉ position in compounds of group 1 and 3.....	27
2.10 Effect of substituent at C ₃ position in compounds of group 1 and 3.....	27
2.11 Comparison of important structural parameters of artemisinin between the X-ray structure, semiempirical, and <i>ab initio</i> optimized structures.....	29
2.12 Energy comparisons of different conformations of artemisinin compounds optimized at HF/3-21G level.....	32
3.1 Activity predictions of 3 compounds in the testing set of group 1.....	57
3.2 Activity predictions of 2 compounds in the testing set of group 3.....	62
4.1 Atoms selected for the definition of alignment rules.....	71
4.2 CoMFA results of different optimized structures and atomic charges.....	74
4.3 CoMFA results of different alignment rules.....	75
4.4 CoMFA results with different type and charge of probe atom.....	75
4.5 CoMFA results with different steric and electrostatic cutoffs.....	76
4.6 CoMFA results with different grid positions.....	77
4.7 CoMFA results with log(D-6) activity using q ² -GRS approach.....	79
4.8 CoMFA results with log(W-2) activity using q ² -GRS approach.....	79
4.9 CoMFA results of all 94 compounds with log(D-6) activity.....	80
4.10 CoMFA results of all 94 compounds with log(W-2) activity.....	82
4.11 CoMFA results of for log(D-6) activity of compounds with C=O at the C ₁₀ position.....	85

Tables	Pages
4.12 CoMFA results of for log(W-2) activity of compounds with C=O at the C ₁₀ position.....	87
4.13 CoMFA results for log(D-6) activity of compounds without C=O at the C ₁₀ position.....	88
4.14 CoMFA results for log(W-2) activity of compounds without C=O at the C ₁₀ position.....	89
5.1 Docking results of heme-pdb and artemisinin derivatives with AMBER and AMBER/MMFF force fields.....	98
5.2 Docking of artemisinin to heme-pdb with different starting and final temperatures.....	99
5.3 Docking of artemisinin to heme with different grid dimension and spacing.....	101
5.4 Docking of artemisinin to heme with different cluster tolerance value.....	102
5.5 Docking of artemisinin to heme with different dielectric constant value...	103
5.6 Docking results of heme-pdb and artemisinin HF/3-21G optimized structure with different atomic charges.....	104
5.7 Docking results of heme-pdb with different atomic charges and artemisinin HF/3-21G optimized structure with HF/3-21G charge.....	105
5.8 Docking results of heme-pdb and artemisinin optimized structures at various levels of theory.....	106
5.9 Docking results of heme-pdb and 16 artemisinin derivatives.....	109
5.10 Docking results of heme-model and 16 artemisinin derivatives.....	111
5.11 Docking results of heme-hemin and 16 artemisinin derivatives.....	113
5.12 Docking results of heme-deoxy and 16 artemisinin derivatives.....	115
5.13 Docking results of heme-oxy and 16 artemisinin derivatives.....	117
5.14 The dissociation constants and binding energies of 4 artemisinin derivatives.....	120
5.15 Docking results of the heme-hemin and artemisinin derivatives group 1..	122
5.16 Docking results of the heme-hemin and artemisinin derivatives group 2..	124
5.17 Docking results of the heme-hemin and artemisinin derivatives group 3..	127

Tables	Pages
6.1 Comparison of some important structural parameters between the X-ray structure and optimized structures of artemisinin.....	137
6.2 Comparison of energy difference (ΔE) of 3 intermediates in the mechanism of action of artemisinin.....	138
6.3 Energy of Singlet and Triplet states of Artemisinin derivatives optimized at the B3LYP/6-31G* level.....	139
6.4 Important structural parameters of radical 1A, transition state (TS ₁), and radical 1B in pathway 1, distances in Angstroms and angles in degrees....	144
6.5 Activation energy (EA ₁), energy difference between radical 1A and 1B (ΔE_1), and imaginary frequency (ν_1) of the transition state (TS ₁) in pathway 1.....	148
6.6 Important structural parameters of radical 2A, transition state (TS ₂), and radical 2B in pathway 2, distances in Angstroms and angles in degrees....	151
6.7 Activation energy (EA ₂), energy difference between radical 2A and 2B (ΔE_2), and imaginary frequency (ν_2) of the transition state (TS ₂) in pathway 2.....	154
6.8 Energy difference between radical 1A and 2A, $\Delta E(1A-2A)$, and between radical 1B and 2B, $\Delta E(1B-2B)$	155

LIST OF ABBREVIATIONS

2D	: Two Dimensional
3D	: Three Dimensional
AM1	: Austin Model 1
B3LYP	: Becke's three parameters and Lee-Yang-Parr for the correlation
CADD	: Computer-Aided Drug Discovery
CNDO	: Complete Neglect of Differential Overlap
CoMFA	: Comparative Molecular Field Analysis
DF	: Degree of Freedom
DFT	: Density Functional Theory
ESPFIT	: Electrostatic Potential Fit
HF	: Hartree-Fock Method
HOMO	: Highest Occupied Molecular Orbital
IMOMM	: Integrated Molecular Orbital and Molecular Mechanics
k	: Number of Variables
LUMO	: Lowest Unoccupied Molecular Orbital
MARS	: Multivariate Adaptive Regression Splines
MLR	: Multiple Linear Regression
MPA	: Mulliken Population Analysis
n	: Number of Compounds
NPA	: Natural Population Analysis
onc	: Optimal Number of Component
PLS	: Partial Least Squares
PRESS	: Predictive Residual Sum of Squares
q^2	: Cross-Validated r^2
q^2 -GRS	: Cross-Validated r^2 -Guided Region Selection
QSAR	: Quantitative Structure-Activity Relationships
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



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 1

Introduction

For pharmaceutical industry, the development of a new drug is a long and expensive process. Therefore, it is an ideal to have more effective drugs with low budget and in a short time. Over the years, many new technologies have been invented to assist the drug development. Computer-Aided Drug Discovery (CADD) is one of these technologies and it has been employed successfully in many pharmaceutical companies. The Quantitative Structure-Activity Relationship (QSAR), one of CADD methods, has been used for some decades to construct the model that can predict biological activity of new compounds without resorting to experimental works. Furthermore, the model gives information on how to modify chemical property to increase the biological activity. Another technique of CADD is the molecular docking which is used to investigate the interaction between drug and receptor. The knowledge obtained from the docking could be used for the structural modifications of drug to achieve the suitable interactions, hence aiding the design of more effective drugs.

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, it is cheaper and more convenient while gives very reliable values as compared to those of experiments. It is, therefore, no need to synthesize compounds. Furthermore, the power in terms of hardware and software is increasing while the costs of computing are steadily decreasing. Secondly, it is able to calculate some properties that are very hard 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, the QSAR techniques (both classical QSAR and 3D-QSAR methods), the molecular docking, and the quantum chemical calculations were applied to investigate and predict antimalarial activities of artemisinin compounds. Details on these techniques will be described in later chapters.

1.1 Malaria

Malaria is one of the most widespread and prevalent endemic diseases, which threatens approximately 40 percent of the world's population in more than 90 countries. This disease is estimated to cause approximately 300 to 500 million illnesses and 1.5 to 2.7 million deaths each year [1-4]. In Thailand, malaria is found mostly in the west region and approximately 100,000 people are infected with this disease and around 800 people die from the disease annually [5].

1.1.1 Historical Outline

Before the source of malaria was found, people believe that smelly pools are connected to malaria disease. Therefore, the Italian word *mal'aria* and the ancient Chinese term *zhangqi* which mean "bad air" were used to call this disease. The origin of the disease became clear at the end of the nineteenth century. In 1880, Alphonse Laveran, a French army surgeon working in Algeria, discovered that malaria was caused by parasitic protozoa in the genus *Plasmodium*. Sir Patrick Manson found evidence that mosquitoes were the vectors for this parasite and that people became infected when bitten. Ronald Ross, a doctor in Indian Medical Service, confirmed and announced these ideas in 1898 [6].

Further studies showed that there are some 100 species of these protozoa but only four are responsible for the disease in humans; *P. falciparum*, *P. malariae*, *P. vivax*, and *P. ovale*. Among these four species, *P. falciparum* is the most dangerous and life-threatening. It produces cerebral and severe disease, which is often fatal if left untreated in non-immune individuals [7]. The last three species produce the mild forms of malaria by destroying red blood cells in peripheral capillaries and thus causing anemia. Humans are affected by inoculation with infected blood or by infected female *Anopheles* mosquito.

1.1.2 Life Cycle of Malarial Parasites

The life cycle of malarial parasites (Figure 1.1) begins when the malarial sporozoites from the mosquito salivary gland (step 1) are injected into the human blood stream as the mosquito must inject the anticoagulant saliva to ensure an even flowing meal (step 2). These sporozoites rapidly penetrate the hepatocytes in the liver (step 3) and, over approximately one week, undergo asexual multiplication forming large tissue

schizonts that contain tens of thousands of merozoites (step 4). Tissue schizonts generally rupture after 5 to 20 days and release the merozoites into the blood stream (step 5), which then enter erythrocytic cells within minutes where they multiply rapidly. These erythrocytes disintegrate and release a new generation of merozoites. Then they invade new healthy erythrocytes. The merozoite multiplication continues for many cycles. Concurrently, some merozoites do not divide, but develop sexual forms of male and female gametocytes (step 6) [8]. These sexual forms are transferred to mosquito when it bites and feeds blood (step 7). The fertilization between male and female gametocytes occurs in the mosquito's gut (step 8). The resulting ookinete develops into an oocyst. Sporogony within the oocyst produces many sporozoites and, when the oocyst ruptures, the sporozoites migrate to the salivary gland (step 1), for injection into another host. At this point the life cycle is completed. More details about life cycle of malarial parasites can be found elsewhere [9-10].

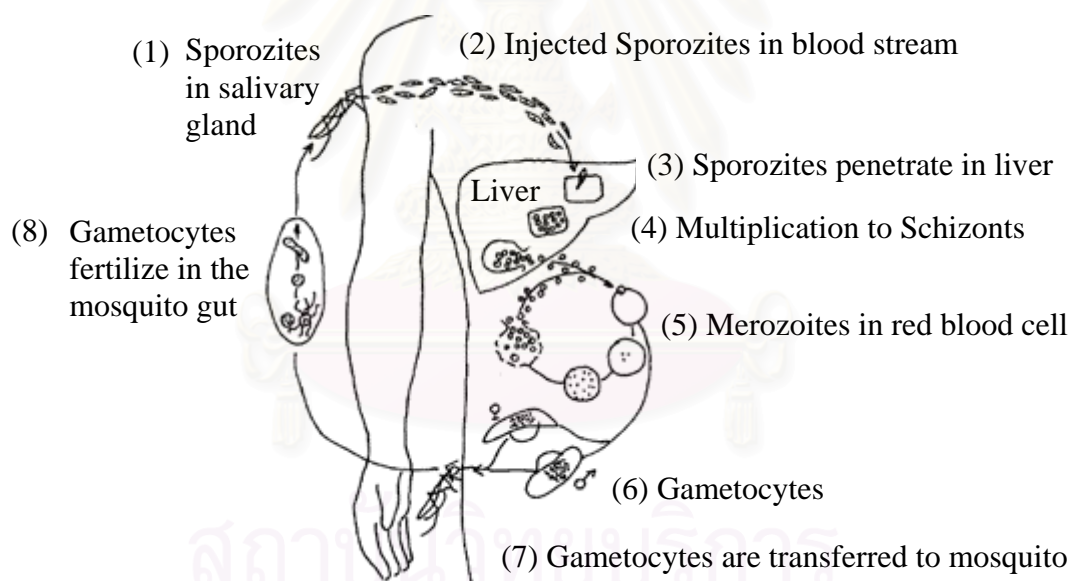


Figure 1.1 The life cycle of *Plasmodium*.

1.1.3 Malaria Symptoms

Malaria symptoms usually develop 10 to 35 days after a person was infected by mosquito's bite. Frequently, the first symptoms are a mild fever, headache, muscle aches, and chills, together with a general feeling of illness. Sometimes symptoms begin with shaking chills followed by fever. These symptoms last 2 or 3 days and are very similar to those of the flu. Subsequent symptoms and patterns of disease vary among

the four types of malaria. For vivax and ovale malaria, an attack may begin abruptly with a shaking chill, followed by sweating and a fever. Within a week, the typical pattern of intermittent attacks is established. A period of headache or of feeling ill may be followed by a shaking chill. The fever lasts 1 to 8 hours. After the fever subsides, the person feels well until the next chill. New attacks tend to occur every 48 hours in vivax malaria. For malariae malaria, an attack often begins abruptly. The attack is similar to that of vivax malaria but recurs every 72 hours. For falciparum malaria, an attack may begin as chills. The person's temperature rises gradually, then falls suddenly. The attack may last for 20 to 36 hours. The person may feel more ill than with vivax malaria and have a severe headache. Between attacks, during intervals that vary from 36 to 72 hours, the person usually feels miserable and has a mild fever.

In falciparum malaria, the parasite causes red blood cells to stick to the walls of arteries, which can result in severe anaemia, kidney failure, water in the lungs, shock or jaundice. It can also cause cerebral malaria with symptoms of a fever with at least 40 °C, severe headache, drowsiness, delirium, and confusion. Cerebral malaria can be fatal as a result of red blood cells sticking to the arteries of the brain and so cutting off blood supply. It most commonly occurs in infants, pregnant women, and travelers to high-risk areas.

Malaria can be diagnosed by the clinical symptoms and microscopic examination of the blood. It can normally be cured by antimalarial drugs. The symptoms quickly disappear once the parasites are killed. If the person is untreated, the symptoms of vivax, ovale, or malariae malaria subside spontaneously in 10 to 30 days but may recur at variable intervals. Untreated falciparum malaria is fatal in up to 20 percent of patients.

1.1.4 Malaria Control

1.1.4.1 Malaria Control in the Past

After the role of the mosquito as malarial vector was established, malarial control has been based on the eradication of this vector. In the 1930s and 1940s the development of low-cost antimalarial drugs and insecticides led experts to conclude that malaria had become an eradicable disease. Therefore, the World Health Organization (WHO) adopted a Global Malaria Eradication Campaign in 1955 [11]. The main

strategy for the campaign was based on the widespread regimented use of the insecticide DDT to kill mosquitoes, of larvicides for treating mosquito breeding sites, and of antimalarial drugs to treat infected individuals and to eliminate the parasite in human. These approaches had led to eradication of malaria in all endemic developed countries and to freedom or partially freedom from the risk of infection in large areas of subtropical Asia and Latin America by 1967. However, the eradication did not appear feasible in tropical Africa due to many problems. These included for examples, the insufficient financial funding, the lack of personnel trained in malaria and its eradication, logistic problems related to transport difficulties, the different habits of African populations that affect malaria transmission, and the resistance of vectors to DDT. In 1969, it became obvious that improvements in the malaria situation could not be maintained indefinitely without substantial national commitments and international assistance in view of the mounting operational, financial, and technical problems faced by the programs. The recognition of these constraints effectively ended the WHO campaign. Since then the financial support to antimalarial programs was considerably reduced. Furthermore, the capabilities of malaria-endemic countries to continue their antimalarial operations were reduced by the world economic crisis in the early 1970s that resulted in a dramatic rise in the prices of insecticides and drugs and of shipping costs. These make a gradual and sometimes even drastic resurgence of malaria through 1980s and 1990s.

With the concern in serious malaria situation, the WHO again adopted the World Declaration on the Control of Malaria and the Global Malaria Control Strategy in 1992. These were consequently confirmed by several organizations around the world. The objectives of the Global Malaria Control Strategy are to prevent mortality and reduce morbidity as well as social and economic loss due to disease through the progressive improvement and strengthening of local and national capabilities for malarial control. National programs were encouraged to focus on early diagnosis and prompt treatment, selective and sustainable prevention, early detection, containment and prevention of epidemics, and building local capacity to assess and manage the malaria situation.

1.1.4.2 Malaria Control at the Present

In 1998, the WHO launched the Roll Back Malaria project with an aim to halve the burden of malaria by 2010 [12]. The project will work through new tools for controlling malaria and also by strengthening health systems for sustainable health improvement. Roll Back Malaria's activities will be implemented through partnerships with international organizations, governments in endemic and non-endemic countries, academic institutions, private sectors, and non-governmental organizations. Partners will work together, at country level, towards common goals using agreed strategies and procedures. The Roll Back Malaria project will act as a pathfinder, helping to set the direction and strategy for more integrated actions in other priority areas.

At present, the malaria control is mainly based on a malaria prophylaxis. The use of insecticide-treated nets has recently been shown to bring about large reductions in mortality ranging from 14% to 63% in African trials [12-13]. However, current implementation in Africa remains limited and achieving high retreatment rates of nets has proved very difficult. Therefore, the current ways in controlling malaria are prevention of mosquito bite and prompt treatment for malaria patients with antimalarial drugs.

1.1.4.3 Malaria Control in the Future

A. Malaria Vaccine

An attractive weapon for preventing malaria is a vaccine [14-17]. Candidate vaccines are based on various antigens or combinations of antigens derived from different stages in the life cycle of the malarial parasite. Currently, three types of vaccines are focusing – asexual blood stage, pre-erythrocytic stage, and transmission-blocking vaccines.

Asexual blood stage vaccines prevent the parasites from entering or developing in red blood cells. These asexual stages (merozoites) of the parasite are responsible for the symptoms of malaria, and a vaccine affecting them would have a great impact on disease morbidity and death, although it would not necessarily prevent people from becoming infected. About a dozen promising asexual blood stage candidate vaccines for *P. falciparum* malaria, such as MSP-1, MSP-2, EBA-175, SERA, MAEBL, AMA-1,

and SPf66, are in various stages of research and development (mostly in clinical Phase II).

Pre-erythrocytic stage vaccines are designed to prevent the parasites's infective sporozoite stage from entering or developing within liver cells of an individual bitten by an infected mosquito. This type of vaccine would prevent infection in non-immune individuals, thus averting the severe, life-threatening consequences of malaria. Up to date, four major candidate vaccines, which are based on the circumsporozite (CS) protein, are in clinical testing.

Transmission-blocking vaccines, e.g., Pfs-25, are aimed to interrupt the development of the parasite in the mosquito host. It would help eliminate transmission of malaria in areas of low endemicity, but in areas of high transmission the vaccine would be used in combination with effective pre-erythrocytic and asexual blood stage antigens. The transmission-blocking vaccine would also contribute to controlling the emergence of drug resistant parasites and/or potential escape variants selected by partially effective pre-erythrocytic and asexual blood stage vaccines.

B. Malaria Genome Project

Malaria genome project is another way to help combating malaria. The main aim is to identify parasite genes. The differences arising from the comparison of parasite genes with host genes will reveal the complex interactions between host and parasite. Hence, this opens up ways of blocking or disrupting parasite development, e.g., interference with genes crucial for a particular stage of a parasite's life cycle. The identified parasite genes would contribute in the finding of new drugs, vaccines, and diagnostic test. Moreover, they also enable the understanding of the molecular basis in drug resistance, which is helpful in creating rational ways of dealing with malaria problem.

C. Transgenic Mosquito

Since the eradication of mosquitoes is notoriously difficult to control, a new approach for vector control has been proposed, i.e., by the manipulation of mosquito genes [18-19]. The ultimate aim is to replace the natural vectors of malaria in the wild with population of *Anopheles* mosquitoes that are incapable of transmitting malaria. Three main areas of research are under development. Based on the fact that the human

malarial parasite does not develop in all species of *Anopheles* mosquito, the first research area is to identify genes responsible for disrupting parasite development in the refractory mosquito. Then the responsible genes could be transferred to susceptible strains of mosquito. The second research area is the development of genetic and molecular tools to insert selected genes into the mosquito genome. And the last research area is the development of methods to spread selected genes in wild mosquito populations.

1.2 Antimalarial Drugs

1.2.1 Historical Outline

In the past, the bark of the cinchona tree was used in South America by the indigenous people to treat the chills and other symptoms associated with malaria. But the first record of the use of cinchona occurred in a religious book written in 1633 by Father Antonio de la Calancha, an Augustinian monk who had lived in Peru. It is alleged that the physician of the Viceroy of Peru was responsible for bringing the treatment back to Spain on his return from South America. From Spain, it was spread all over Europe by 1640 and was proved to be very successful in the treatment of malaria as mentioned in European medical literature in 1643. For almost 2 centuries the bark was employed for medicine as a powder, extract, or infusion. Many attempts were made to isolate the active principle from the bark. But it was until 1820 that Pierre Pelletier and Joseph Caventou succeeded in separating quinine from the bark. The quinine (Figure 1.2a) was found to be more palatable than the nauseating powder of cinchona bark [6]. Since then it became the main treatment for malaria. However, during the World War I (1914-1918) the normal supplies of quinine become unavailable. As a result, an extensive research program on synthetic antimalarials was settled in Germany.

From this extensive research program, pamaquine, a synthetic 8-aminoquinoline derivative, was synthesized in 1920s. The compound was found to be 60 times more effective than quinine in killing malarial parasites lodged in the liver [20]. A few years later, quinacrine or mepacrine was also developed and introduced for malarial therapy as a synthetic alternative to quinine. However, the quinine was still the chief antimalarial drug. During the World War II (1939-1945) quinine was again no longer

available to the Allies since the Japanese cut off the supply of cinchona bark from Java. Therefore, quinacrine became the official drug for the treatment of malaria. But its toxicity and inability to cure benign tertian malaria, *P. vivax* malaria, or to act as a true causal prophylactic stopped its use and made it as an obsolete antimalarial drug. In order to search for more active compounds, some extensive antimalarial research programs were established in many countries such as the United States and the United Kingdom.

From the investigation on a large number of 4-aminoquinoline derivatives since 1941 by the cooperative program of antimalarial research in the United States, chloroquine (Figure 1.2b) was found to be a very effective drug, which has fewer side effects and does not turn the patient yellow [21]. However, it was just recognized that the compound had been synthesized and studied as early as 1934 under the name Resochin by the Germans. Another successful drug, primaquine (Figure 1.2c), was found during the exhaustive search for more potent and less toxic compound than the pamaquine. The primaquine is particularly effective against *P. vivax*, the cause of benign tertian fever.

With the knowledge that pyrimidine compounds are of importance in the cell metabolism, the investigations on a large series of its derivatives were conducted under the antimalarial research program in the United Kingdom. This led to the discovery of proguanil or chloroguanide, a biguanidine compound, in 1945 [22]. The compound was proved to be an outstanding causal prophylactic agent for falciparum malaria and a satisfactory suppressive for vivax malaria. The investigations were further studied in the early 1950s by the joint research team of British and American. Pyrimethamine (Figure 1.2d) was then developed in 1951 [23]. Its antimalarial effects are identical to those of proguanil but, however, its potency is considerably greater. Undoubtedly, this is owing to the fact that it acts directly and its half-life is much longer than that of the active metabolite of proguanil. Therefore, pyrimethamine has been used widely for prophylaxis and suppression.

In the early 1960s the chloroquine-resistant strains of *P. falciparum* were reported in South America and South East Asia [24]. This prompted the U.S. Army to develop new effective antimalarial drugs. Resulting from the screening of about 300 of 4-quinolinemethanol derivatives since 1963, mefloquine (Figure 1.2e) displayed high

activity against the chloroquine- and the pyrimethamine-resistant strains [25]. Due to its highly effectiveness, mefloquine was widespreadly used since the late 1970s. However, the resistance to this drug since the early 1980s and its severe side effects [26] reduced its use.

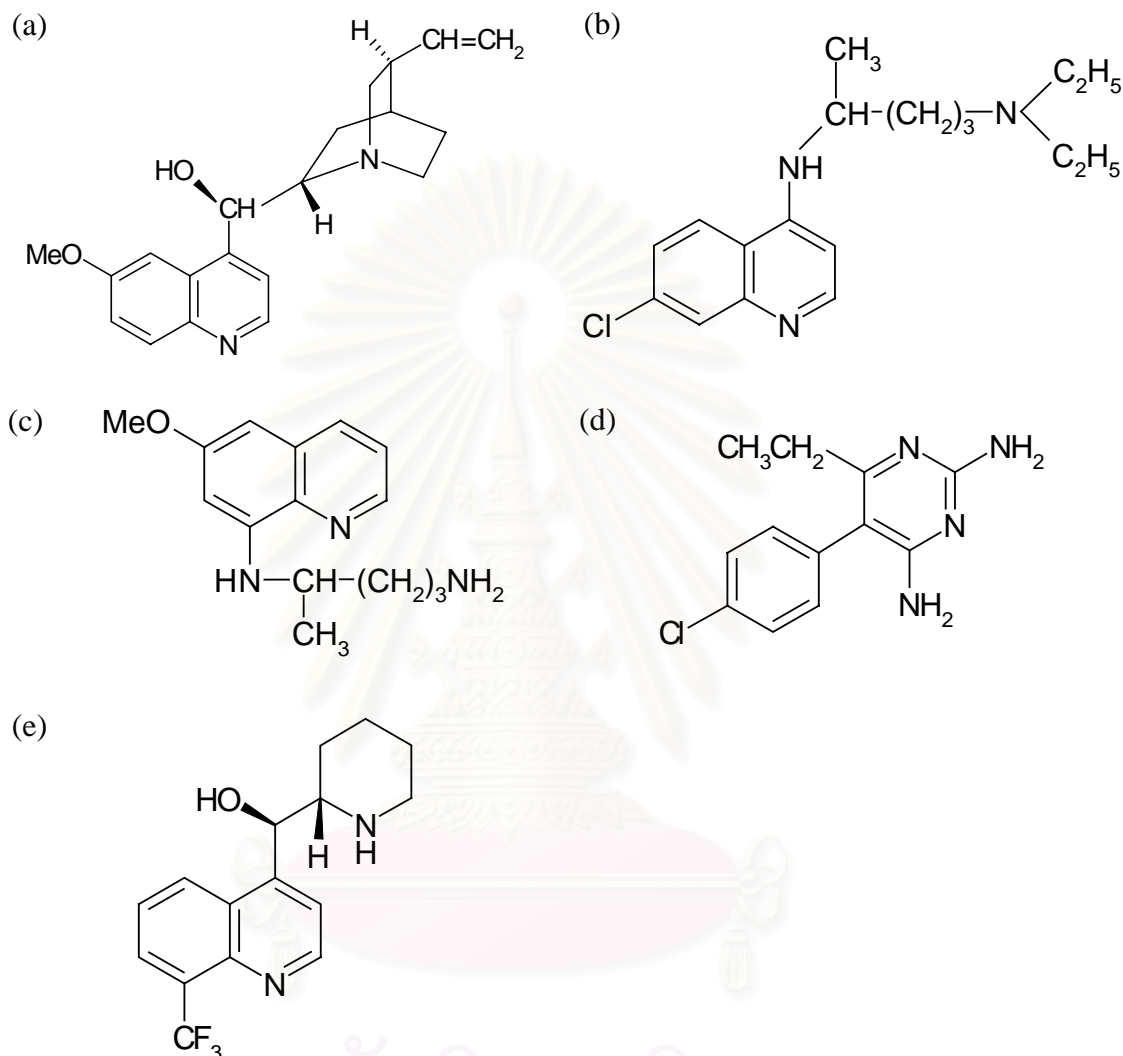


Figure 1.2 Structures of antimalarial drugs (a) quinine, (b) chloroquine, (c) primaquine, (d) pyrimethamine, (e) mefloquine

The developments of parasite resistant strains to most common chemotherapeutic agents, e.g., chloroquine, quinine, sulfa/pyrimethamine combination, and mefloquine, have been reported in many parts of the world [27-29]. As a result, the malaria situation becomes serious once again. The Chinese discovery of a new and potentially valuable artemisinin abates the situation. This drug and its derivatives are very effective against the drug-resistant strains.

1.2.2 Classification of Antimalarial Drugs

The classification of antimalarial drugs can be done in many ways depending on the criteria used, such as chemical structure, drug target, and drug action. However, the biological classification, based on the parasite stage in which drug mediates its action, is widely used [21]. According to this classification, 5 categories are defined as follows.

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

1.2.2.2 Secondary Tissue Schizontocides (radically curative drugs). The drugs, e.g., primaquine, eradicate the 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.

1.2.2.3 Schizontocides (blood schizontocides or schizontocidal drugs). The drugs act rapidly on the erythrocytic stages (schizon) of parasites in red blood cell. By interrupting the asexual reproduction of malarial parasite, the clinical attack is terminated. Continuing use of schizontocides for a longer period than the 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.

1.2.2.4 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, eliminates the reservoir from which mosquitoes are reinfected. They also act on the development stages of malarial parasites in the *Anopheles*, thus some of them form the next group of drugs.

1.2.2.5 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 Artemisinin and its derivatives

1.3.1 Historical Outline

The medicinal herb qinghao (*Artemisia annua* L., sweet wormwood, annual wormwood), a native plant in China, has been used as a remedy for fever in China since the ancient time. But the first record was found in the treatise “Fifty-two Prescriptions” discovered in the Mawangdui Han Dynasty Tomb of 168 B.C. It was also recorded in the “Shennong Bencao Jing” published in the 1-2 century A.D. However, its use as antimalarial agent was first mentioned in 341 A.D. by Ge Hong in the handbook of prescriptions for emergency treatments, “Zhouhou Bei Ji Fang”. The next evidence was discovered in the “Bencao Gangmu” (compendium of materia medica) written by Li Shizhen, the famous herbalist, in 1596 A.D. Also, qinghao decoction was noted in the “Wenbing Tiaobian” for malaria treatment in 1798 A.D. Qinghao was prescribed in many ways, such as soaked water, decoction, pill, and powder.

In 1967 the government of the People’s Republic of China launched a systematic examination of indigenous plants used in traditional remedies as sources of drugs. By 1971, the crude extraction of qinghao with ethyl ether was shown to be highly effective in mice and simian malaria. Further investigation led to the isolation of an effective antimalarial compound from the aerial portions of the plant in 1972 [30]. The compound was named “qinghaosu”, which means “active principle of qinghao”, and the more Western sounding name, “artemisinine”. But, the “-ine” suffix normally suggests the alkaloid or amine structures, may led to misunderstanding of its terpene structure. Therefore, the name “artemisinin” is preferred and used by the Chemical Abstracts.

1.3.2 Chemical Structure

Many experimental works have been done to elucidate the chemical structure of artemisinin. The compound, a colorless needle crystal, has a melting point of 156-157 °C. High resolution mass spectroscopy data ($m/e = 282.1742$, M^+) together with elemental analysis data (C = 63.72% and H = 7.86%) revealed the empirical formula of $C_{15}H_{22}O_5$, which suggested a sesquiterpene structure [31]. The absorption

peaks in the IR region at 1745 cm^{-1} (strong, delta-lactone) and at $722, 831, 881, 1115\text{ cm}^{-1}$ (peroxide) were observed [31]. The $^1\text{H-NMR}$ and $^{13}\text{C-NMR}$ spectra indicated the presence of three methyl groups (one tertiary and two secondary), an acetal function, and several kinds of aliphatic carbon atoms. Qualitative and quantitative reaction experiments verified the presence of the lactone and peroxy-group. Later, its absolute structure was investigated by the X-ray diffraction [30] and was reconfirmed again [32-33]. The exact chemical name was assigned as octahydro-3,6,9-trimethyl-3,12-epoxy-12H-pyrano-[4,3-j]-1,2-benzodioxepin-10(3H)-one. Its stereochemistry and atomic numbering scheme according to the IUPAC is shown in Figure 1.3.

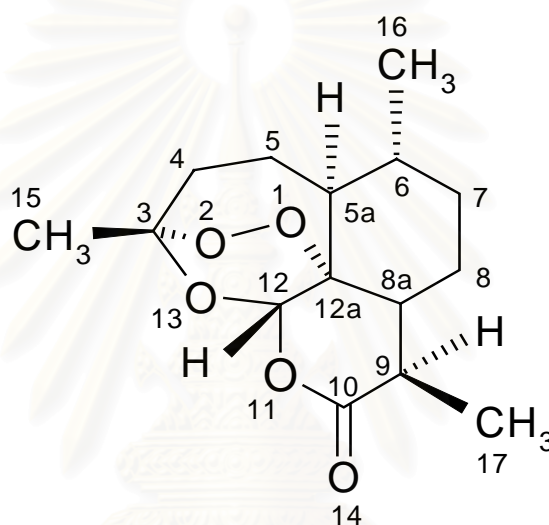


Figure 1.3 Stereochemistry and atomic numbering scheme of artemisinin.

1.3.3 Antimalarial Activity

From early *in vitro* experiments, artemisinin has potency comparable to chloroquine and mefloquine [34-35]. Moreover, it was shown to be effective against the chloroquine-resistant strains of *P. falciparum*. *In vivo* experiments in mice, chickens, and monkeys also shown that clearance of parasitemia could be accomplished by the administration of artemisinin. However, the high recrudescence rate of parasitemia (up to 70%) was found [30].

In human, the clinical cures of 1,511 patients with *P. vivax* and 588 patients with *P. falciparum* during 1973-1978 were reported by the qinghaosu antimalaria coordinating research group in 1979 [30]. In addition, 143 cases of chloroquine-resistant falciparum malaria and 141 cases of cerebral malaria were treated with good results. Artemisinin were administrated in four different dosage forms: tablets, in oil, as

an oil suspension, and as a water suspension. In *P. vivax* patients, the order of rapidity in parasite clearance is tablets > oil > oil suspension > water suspension. In *P. falciparum* patients, the oil form was the most rapid acting.

Artemisinin is a rapid acting drug as indicated by the parasite clearance time of about 40 hours (in tablet form) compared with 56 hours of the chloroquine in *P. vivax* malaria [30]. The drug was considered safe in normal patients and also in patients complicated by heart, liver, and renal diseases of pregnancy. Neither obvious adverse reactions nor noticeable side effects were seen during the treatment. Unfortunately, a high rate of recrudescence is its main problem [36]. In addition, the drawbacks are also contributed from its insolubility in both water and oil [37], its poor efficacy by oral administration [38], and short plasma half-life [36]. However, there exist solutions for all these problems.

In order to solve the problem of its poor solubility, some more soluble derivatives were synthesized [37], e.g., dihydroartemisinin (Figure 1.4a), artemether (Figure 1.4b), arteether (Figure 1.4c), and artesunate (Figure 1.4d). These derivatives were found to be more active than artemisinin and hence, they are now increasingly being used for malaria treatment against drug resistant strains of *P. falciparum* [39-41]. Combining these drugs with other antimalarials having a longer half-life, such as mefloquine, could solve the recrudescence and the short plasma half-life problem [39, 42-45].

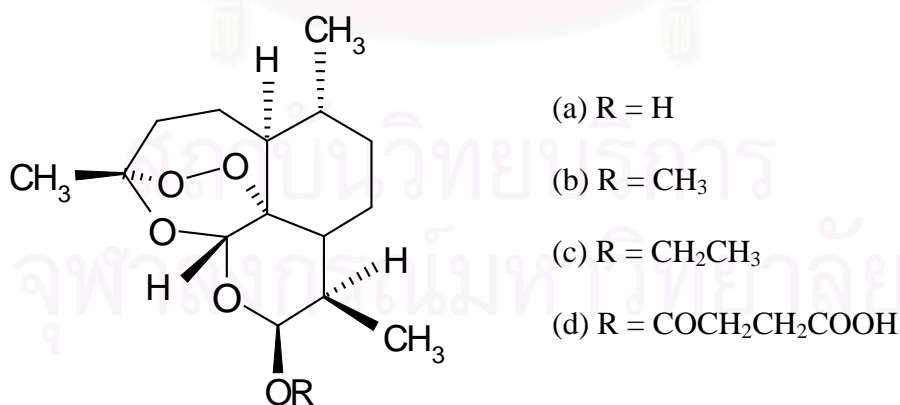


Figure 1.4 Structures of artemisinin derivatives (a) dihydroartemisinin, (b) artemether, (c) arteether, (d) artesunate

Despite the fact that artemisinin and its derivatives are active and are now being used worldwide, their long term uses possibly cause malarial parasite to give lower response to these compounds in a near future as happened to the other antimalarial drugs. Hence, new more effective derivatives are needed. Therefore, the QSAR techniques were applied on these compounds. Moreover, mechanism of action of these compounds, which will be discussed in chapter 6, was investigated by means of the molecular docking and the quantum chemical calculations. The understanding in mechanism of action could assist in designing more effective derivatives and also in developing new class of compounds, as well as in preventing the drug resistance.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 2

Structures and Biological Data

With the aim to explore all possible effects of structural differences in artemisinin compounds concerning their biological activities, totally 104 artemisinin derivatives with significantly different structures and biological activities [46-55] were used in this study. All compounds were categorized into 3 groups according to their structural similarity. The activities were measured as the IC₅₀ values, the inhibitory concentration of a compound required for 50% inhibition of the parasitemia, against the Sierra Leone (D-6) and the Indochina (W-2) clones of *P. falciparum*. The D-6 clone is mefloquine-resistant but chloroquine-sensitive while the W-2 clone is chloroquine-resistant but mefloquine-sensitive. Since the biological data arise from different sources, there might occur an inconsistency from individual experimental testing procedure. Therefore, the relative activity, the ratio of activity of artemisinin and the drug compound, was used. Moreover, in order to compare drug activities of different compounds of various molecular weights, it is necessary to convert the biological activities of the compounds (IC₅₀) in ng/mL unit to nM/mL unit according to the formula shown below. As a result, compounds with relative activities of higher than 1.00 are more active than artemisinin. On the other hand, compounds with relative activities of lower than 1.00 are less active than artemisinin. The structures of 104 compounds are given in Figure 2.1 to 2.8 and their corresponding biological data together with the reference sources are depicted in Table 2.1 to 2.8.

$$\text{activity} = (\text{IC}_{50} \text{ of artemisinin} / \text{IC}_{50} \text{ of analog}) \times (\text{MW of analog} / \text{MW of artemisinin})$$

2.1 Compounds Group 1

There are 41 compounds belonging to this group. All the structures have C=O group at the C₁₀ position. Compounds 2 to 31 are analogues of artemisinin (compound 1) with substituent groups at C₃ and C₉ positions (Figure 2.1). Compound 32 to 41, which contain a nitrogen atom at position 11 instead of oxygen atom, were included to study the effect of the N₁₁ atom (Figure 2.2). Moreover, for these compounds, the -CH₃ group at the C₉ position is disappeared and the substituents at the position 11 are varied.

From both D-6 and W-2 activities of these 10 compounds (Table 2.2), it appears that O₁₁ atom is not a necessary requirement for the antimalarial activity.

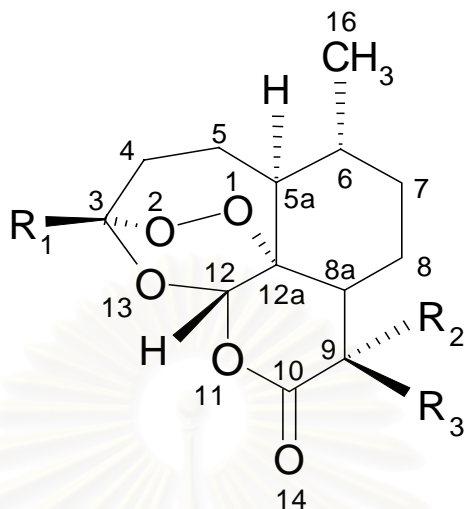


Figure 2.1 Structures of artemisinin derivatives number 1 to 31.

Table 2.1 Structures and biological data of compounds number 1-31 in group 1.

No	R ₁	R ₂	R ₃	D-6	W-2	Ref.
1	CH ₃	H	CH ₃	1.00	1.00	46
2	CH ₃		=CH ₂	0.35	0.13	47
3	CH ₃	H	H	1.71	6.18	46
4	CH ₃	CH ₃	H	0.14	0.67	46
5	CH ₃	CH ₃	CH ₃	0.02	0.006	46
6	CH ₃	H	C ₂ H ₅	6.74	12.87	46
7	CH ₃	H	<i>n</i> -C ₃ H ₇	6.05	13.47	46
8	CH ₃	H	<i>i</i> -C ₃ H ₇	2.23	0.92	46
9	CH ₃	H	<i>n</i> -C ₄ H ₈	1.10	1.47	46
10	CH ₃	H	<i>i</i> -C ₄ H ₈	0.23	0.28	46
11	CH ₃	H	<i>n</i> -C ₅ H ₁₁	4.40	10.39	46
12	CH ₃	H	<i>i</i> -C ₅ H ₁₁	1.88	1.17	46
13	CH ₃	H	<i>n</i> -C ₆ H ₁₃	5.83	6.87	46
14	CH ₃	H	<i>i</i> -C ₆ H ₁₃	0.39	0.92	46
15	CH ₃	H	(CH ₂) ₂ C ₆ H ₅	2.23	1.32	46
16	CH ₃	H	(CH ₂) ₃ C ₆ H ₅	8.36	6.09	46
17	CH ₃	H	(CH ₂) ₄ C ₆ H ₅	1.94	4.26	46

Table 2.1 (Continued)

No	R ₁	R ₂	R ₃	D-6	W-2	Ref.
18	CH ₃	H	CH ₂ COOH	0.002	0.003	46
19	CH ₃	H	CH ₂ CH=CH ₂	0.34	0.79	46
20	CH ₂ CH ₃	H	H	0.88	1.12	48
21	(CH ₂) ₂ CH ₃	H	H	21.02	6.73	48
22	(CH ₂) ₃ CH ₃	H	H	0.20	0.18	48
23	CH ₂ CH(CH ₃) ₂	H	H	0.53	0.45	48
24	(CH ₂) ₂ COOC ₂ H ₅	H	H	2.32	2.32	48
25	(CH ₂) ₂ C ₆ H ₅	H	H	0.03	0.01	48
26	<i>p</i> -ClC ₆ H ₄ (CH ₂) ₃	H	H	1.14	1.27	48
27	C ₆ H ₅ (CH ₂) ₄	H	H	2.20	2.81	48
28	CH ₂ CH ₃	H	(CH ₂) ₃ CH ₃	1.84	2.57	48
29	(CH ₂) ₄ C ₆ H ₅	H	(CH ₂) ₃ CH ₃	0.39	0.48	48
30	(CH ₂) ₃ CH ₃	H	(CH ₂) ₃ CH ₃	0.28	0.33	48
31	<i>p</i> -ClC ₆ H ₄ (CH ₂) ₃	H	(CH ₂) ₃ CH ₃	0.43	0.53	48

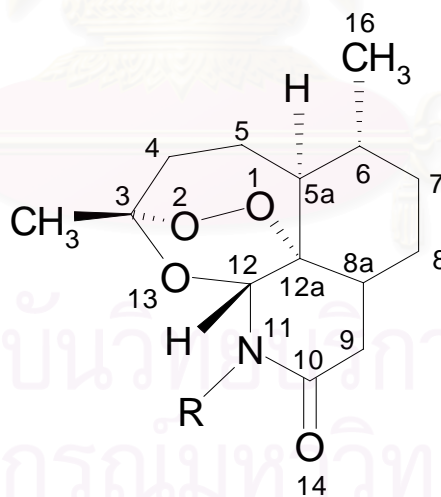


Figure 2.2 Structures of artemisinin derivatives number 32 to 41.

Table 2.2 Structures and biological data of compounds number 32-41 in group 1.

No	R	D-6	W-2	Ref.
32	CH ₃	2.13	5.00	49
33	<i>n</i> -C ₅ H ₁₁	1.10	0.63	49

Table 2.2 (Continued)

No	R	D-6	W-2	Ref.
34	(CH ₂) ₅ COOH	0.12	0.02	49
35	CH ₂ C ₆ H ₅	1.89	2.17	49
36	CH ₂ CH ₂ CH ₃	0.75	1.11	49
37	<i>i</i> -C ₄ H ₉	1.45	1.05	49
38	<i>i</i> -C ₅ H ₁₁	1.49	0.92	49
39	<i>p</i> -ClC ₆ H ₄ CH ₂	1.11	0.69	49
40	(CH ₂) ₂ C ₆ H ₅	1.97	1.43	49
41	(CH ₂) ₃ C ₆ H ₅	2.05	1.05	49

2.2 Compounds Group 2

Compounds in this category have either -OR or -NR functional group at the C₁₀ position. This makes their structures significantly different from those in the group 1, especially the lactone ring. The reason is clearly due to the difference in hybridization of the C₁₀ atom, i.e., sp² in the group 1 and sp³ in the group 2. This structural difference is demonstrated in Figure 2.3.

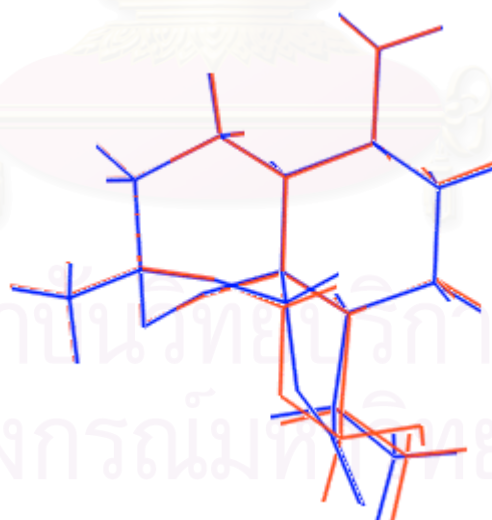


Figure 2.3 Structural difference between compound in group 1 (blue) and 2 (red).

Totally 46 compounds were chosen in this group. The structures of compounds 42 to 74 are very similar. They vary only the substituent groups at either the O₁₄ position (Figure 2.4) or at the C₁₈ position (Figure 2.5). And compound 79 to 82 (Figure 2.7) have substituent groups at both C₉ and O₁₄ positions.

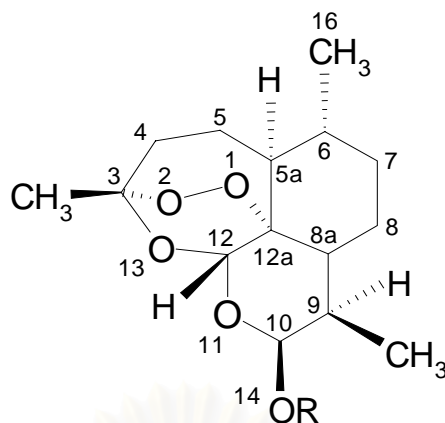
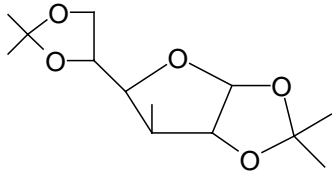
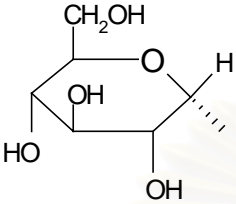
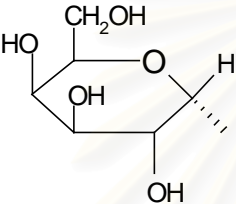
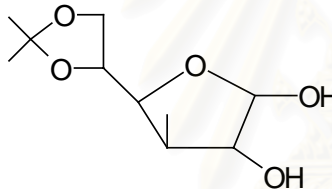


Figure 2.4 Structures of artemisinin derivatives number 42 to 60.

Table 2.3 Structures and biological data of compounds number 42-60 in group 2.

No	R	D-6	W-2	Ref.
42	H	7.20	0.96	50
43	CH ₃	2.80	3.57	51
44	CH ₂ COOCH ₂ CH ₃	6.41	3.33	50
45	(CH ₂) ₂ COOCH ₃	2.09	1.35	50
46	(CH ₂) ₃ COOCH ₃	1.30	0.95	50
47	CH ₂ C ₆ H ₄ COOCH ₃	5.83	2.73	50
48	CH ₂ COOK	0.07	0.03	50
49	(CH ₂) ₂ COOK	0.05	0.04	50
50	(CH ₂) ₃ COOK	0.06	0.09	50
51	CH ₂ C ₆ H ₄ COOK	2.72	1.16	50
52	(CH ₂) ₂ COOH	0.07	0.02	50
53	(CH ₂) ₃ COOH	0.21	0.11	50
54	CH ₂ C ₆ H ₄ COOH	1.07	0.71	50
55	Si(CH ₃) ₃	3.34	2.04	52
56		2.58	7.79	52

Table 2.3 (Continued)

No	R	D-6	W-2	Ref.
57		3.83	4.09	52
58		0.04	0.31	52
59		0.35	0.03	52
60		1.97	1.43	52

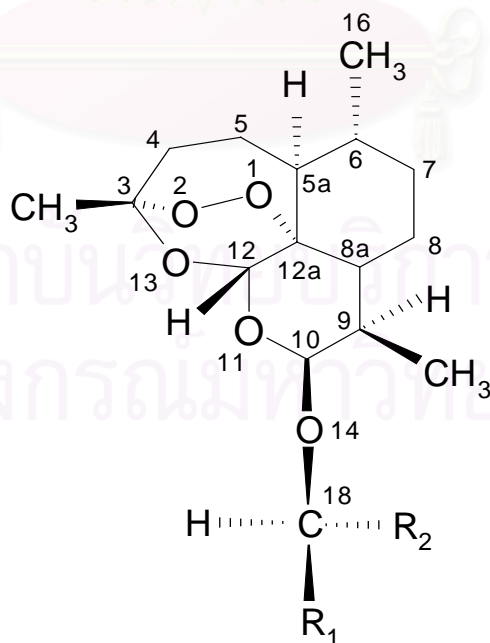
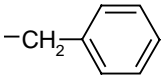
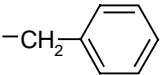


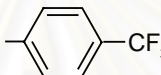


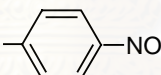
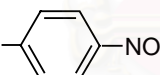
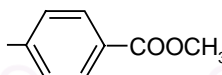
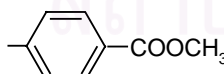
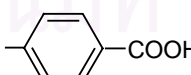
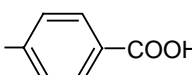
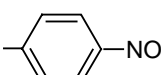


Figure 2.5 Structures of artemisinin derivatives number 61 to 74.

Table 2.4 Structures and biological data of compounds number 61-74 in group 2.

No	R ₁	R ₂	D-6	W-2	Ref.
61	-CH ₂ CH ₂ CH ₃		1.15	0.81	51
62		-CH ₂ CH ₂ CH ₃	3.61	2.22	51
63		-COOCH ₂ CH ₃	12.68	17.11	51
64	-COOCH ₂ CH ₃		7.15	7.09	51
65	-CH ₃		2.55	1.43	51
66	-CH ₂ COOCH ₂ CH ₃		7.58	7.76	51
67		-CH ₂ COOCH ₂ CH ₃	8.18	11.52	51
68	-CH ₂ COOCH ₂ CH ₃		7.32	7.91	51
69		-CH ₂ COOCH ₂ CH ₃	16.74	37.32	51
70	-CH ₃		3.95	4.77	51
71		-CH ₃	5.26	4.79	51
72	-CH ₃		0.61	0.61	51
73		-CH ₃	1.20	1.26	51
74	-CH ₂ COOH		1.32	1.24	51

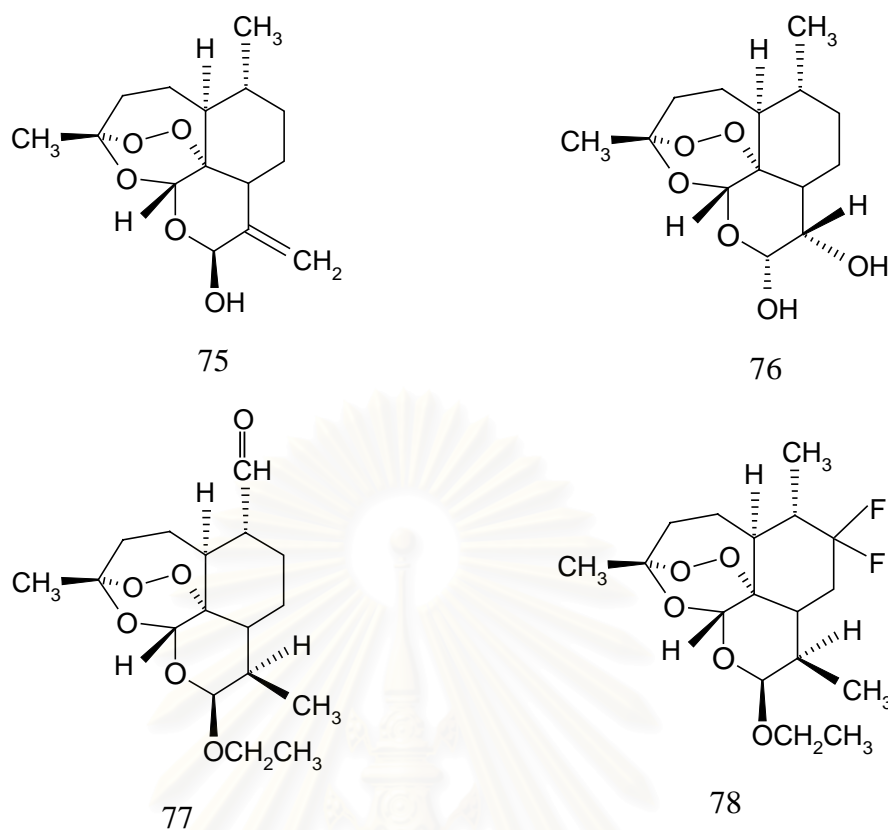


Figure 2.6 Structures of artemisinin derivatives number 75 to 78.

Table 2.5 Structures and biological data of compounds number 75-78 in group 2.

No.	D-6	W-2	Ref.
75	0.016	0.004	47
76	0.26	0.13	47
77	1.39	1.62	53
78	1.73	1.48	53

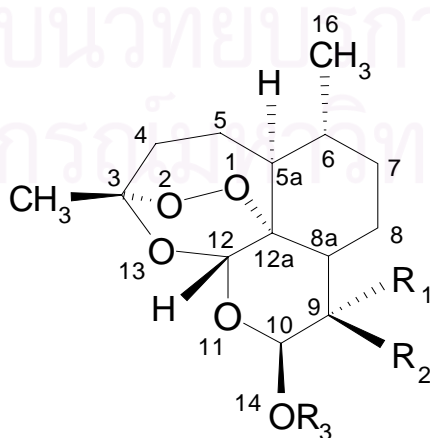


Figure 2.7 Structures of artemisinin derivatives number 79 to 82.

Table 2.6 Structures and biological data of compounds number 79-82 in group 2.

No	R ₁	R ₂	R ₃	D-6	W-2	Ref.
79	OH	CH ₃	CH ₂ CF ₃	2.44	2.03	53
80	CH ₃	OH	CH ₂ CF ₃	0.20	0.19	53
81	OH	CH ₃	CH ₂ CH ₃	0.52	0.35	53
82	CH ₃	OH	CH ₂ CH ₃	0.09	0.07	53

Unlike compounds 42 to 82, which have an oxygen atom at the position 14, the following five compounds have a nitrogen atom at the position 14. Therefore, the effect of substituent group at the C₁₀ position can be investigated. Note that all these five compounds have a bromine atom as substituent at the α -C₉ position.

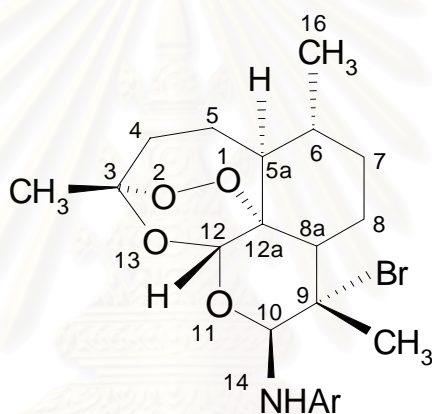
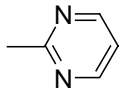


Figure 2.8 Structures of artemisinin derivatives number 83 to 87.

Table 2.7 Structures and biological data of compounds number 83-87 in group 2.

No	Ar	D-6	W-2	Ref.
83		0.97	4.58	54
84		5.66	6.16	54
85		0.25	1.50	54
86		0.11	0.80	54

Table 2.7 (Continued)

No	Ar	D-6	W-2	Ref.
87		0.03	0.17	54

2.3 Compounds Group 3

Unlike the previous two groups, compounds in this group have no substituent at the C₁₀ position (Figure 2.9). Most compounds in this group have higher activities than artemisinin itself. Hence, the O₁₄ atom seems to be not necessarily required for high antimalarial activity. Totally 17 compounds with different substituent groups at the C₃ and the β-C₉ position were selected.

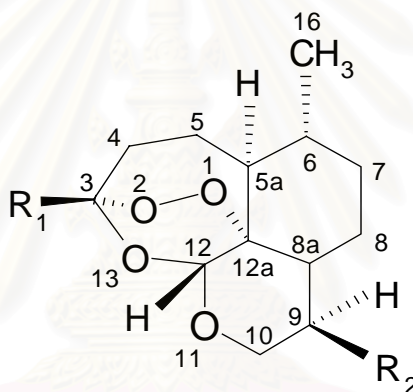


Figure 2.9 Structures of artemisinin derivatives number 88 to 104.

Table 2.8 Structures and biological data of compounds number 88-104 in group 3.

No	R ₁	R ₂	D-6	W-2	Ref.
88	CH ₃	CH ₃	6.59	5.67	55
89	CH ₃	H	2.37	1.90	55
90	CH ₃	CH ₂ CH ₃	9.14	4.66	55
91	CH ₃	(CH ₂) ₂ CH ₃	4.73	5.50	55
92	CH ₃	(CH ₂) ₃ CH ₃	58.26	20.90	55
93	CH ₃	(CH ₂) ₄ CH ₃	1.70	1.45	55
94	CH ₃	(CH ₂) ₃ C ₆ H ₅	50.73	25.06	55
95	CH ₃	<i>p</i> -ClC ₆ H ₄ (CH ₂) ₃	69.91	33.17	55
96	CH ₂ CH ₃	H	0.10	0.10	55

Table 2.8 (Continued)

No	R ₁	R ₂	D-6	W-2	Ref.
97	(CH ₂) ₂ CH ₃	H	7.22	6.85	55
98	(CH ₂) ₃ CH ₃	H	6.53	5.56	55
99	CH ₂ CH(CH ₃) ₂	H	1.83	2.50	55
100	(CH ₂) ₄ C ₆ H ₅	H	3.36	3.80	55
101	(CH ₂) ₂ C ₆ H ₅	H	0.06	0.02	55
102	<i>p</i> -ClC ₆ H ₄ (CH ₂) ₃	H	0.13	0.28	55
103	(CH ₂) ₂ COOC ₂ H ₅	H	4.22	5.06	55
104	(CH ₂) ₂ COOH	H	0.0009	0.0009	55

The main reason to distinguish this group from the two previous ones is the substituent at the C₁₀ position. All compounds in the group 2 are proposed to change rapidly to dihydroartemisinin upon entering the body [56]. But for compounds in the group 3, this process can not occur.

Comparing between some compounds in group 1 and group 3, the substituent at either C₉ (Table 2.9) or C₃ positions (Table 2.10) leads to different changes in activities for both groups. E.g., when the hydrogen atom at the β-C₉ position is changed to a methyl group, the activities increase for the compound 1 (group 1) whereas they decrease for the compound 88 (group 3). Note that the activities of each compound are reported as the relative values to those of the compound with R = -H for substituent at the C₉ position and with R = -CH₃ for substituent at the C₃ position.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 2.9 Effect of substituent at C₉ position in compounds of group 1 and 3.

R	Group 1			Group 3		
	No.	D-6	W-2	No.	D-6	W-2
H	3	1.00	1.00	89	1.00	1.00
CH ₃	1	0.58	0.16	88	2.78	2.98
CH ₂ CH ₃	6	3.94	2.08	90	3.86	2.45
(CH ₂) ₂ CH ₃	7	3.54	2.18	91	2.00	2.89
(CH ₂) ₃ CH ₃	9	0.64	0.24	92	24.58	11.00
(CH ₂) ₄ CH ₃	11	2.57	1.68	93	0.72	0.76
(CH ₂) ₃ C ₆ H ₅	16	4.89	0.99	94	21.41	13.19

Table 2.10 Effect of substituent at C₃ position in compounds of group 1 and 3.

R	Group 1			Group 3		
	No.	D-6	W-2	No.	D-6	W-2
CH ₃	3	1.00	1.00	89	1.00	1.00
CH ₂ CH ₃	20	0.51	0.18	96	0.04	0.05
(CH ₂) ₂ CH ₃	21	12.29	1.09	97	3.05	3.61
(CH ₂) ₃ CH ₃	22	0.12	0.03	98	2.76	2.93
CH ₂ CH(CH ₃) ₂	23	0.31	0.07	99	0.77	1.32

Table 2.10 (Continued)

R	No.	D-6	W-2	No.	D-6	W-2
<i>p</i> -ClC ₆ H ₄ (CH ₂) ₃	26	0.67	0.21	102	0.05	0.15
(CH ₂) ₄ C ₆ H ₅	27	1.29	0.45	100	1.42	2.00
(CH ₂) ₂ C ₆ H ₅	25	0.02	0.002	101	0.03	0.01
(CH ₂) ₂ COOC ₂ H ₅	24	1.36	0.38	103	1.78	2.66

2.4 Structural Optimization

Since the experimental structures are not available for all the compounds used in this study, quantum chemical calculations were used to determine the geometry of each compound. Therefore, it is necessary to establish a suitable level of accuracy for the geometry optimization. For this purpose, the artemisinin structure was optimized using CNDO, AM1, Hartree Fock (HF) with 3-21G and 6-31G** basis sets, and Density Functional Theory (DFT) level of theory with B3LYP functional and 6-31G** basis set. Subsequently, all optimized structures were compared with the X-ray structure [33]. The results indicated that the Hartree Fock method with 3-21G basis set (HF/3-21G) is the lowest level of theory that gives geometrical parameters within acceptable accuracy to the X-ray data (see Table 2.11). This applies especially to the bond length of the endoperoxide linkage which is believed to be responsible for the antimalarial activity [31]. Therefore, the HF/3-21G method was chosen as the optimization method for all 104 compounds. Artemisinin derivative structures with (group 1) and without (group 2 and 3) C=O at the C₁₀ position were built using the X-ray structure of artemisinin and dihydroartemisinin, respectively, as template. The structures were then geometrically optimized with the HF/3-21G method using the Gaussian 92 and 94 program [57-58].

For compound 48 to 51, the original structures are the salts with K⁺. But these compounds could not be calculated at the HF/3-21G level since the 3-21G basis set for potassium atom is not available (in Gaussian 92). Moreover, the active species for these compounds in solution are the anionic form, which do not contain potassium atom. Therefore, the anionic form was used instead.

Table 2.11 Comparison of important structural parameters of artemisinin between the X-ray structure, semiempirical, and *ab initio* optimized structures.

Parameter	X-ray	CNDO	AM1	HF/ 3-21G	HF/ 6-31G**	B3LYP/ 6-31G**
Bond Distance (Å)						
O ₁ -O ₂	1.475	1.229	1.289	1.462	1.390	1.460
O ₁ -C _{12a}	1.450	1.410	1.469	1.477	1.430	1.455
O ₂ -C ₃	1.417	1.394	1.447	1.441	1.396	1.414
C ₃ -C ₄	1.518	1.488	1.532	1.537	1.537	1.546
C _{5a} -C _{12a}	1.538	1.504	1.540	1.537	1.546	1.555
C ₅ -C _{5a}	1.537	1.486	1.526	1.551	1.543	1.547
C ₄ -C ₅	1.533	1.477	1.513	1.544	1.535	1.539
C ₃ -O ₁₃	1.448	1.390	1.427	1.436	1.409	1.441
C ₁₂ -O ₁₃	1.388	1.387	1.416	1.408	1.376	1.396
C _{5a} -C ₆	1.550	1.495	1.531	1.549	1.545	1.552
C _{8a} -C _{12a}	1.520	1.538	1.532	1.529	1.532	1.540
C ₁₂ -O ₁₁	1.455	1.401	1.421	1.428	1.408	1.439
C ₆ -C ₇	1.532	1.486	1.521	1.543	1.533	1.539
C ₈ -C _{8a}	1.533	1.490	1.517	1.535	1.533	1.538
C ₃ -C ₁₅	1.517	1.478	1.519	1.512	1.513	1.519
C ₉ -C ₁₆	1.530	1.474	1.516	1.542	1.532	1.536
C _{8a} -C ₉	1.540	1.504	1.524	1.539	1.537	1.543
C ₁₀ -O ₁₁	1.351	1.409	1.373	1.366	1.338	1.365
C ₁₀ =O ₁₄	1.201	1.340	1.231	1.197	1.183	1.207
C ₉ -C ₁₇	1.540	1.464	1.513	1.532	1.528	1.531
Bond Angle (°)						
C ₅ -C _{5a} -C _{12a}	112.6	112.1	111.2	111.8	112.0	112.1
C ₄ -C ₅ -C _{5a}	116.3	115.2	114.6	115.7	116.5	116.6
C ₃ -C ₄ -C ₅	114.5	115.2	113.9	113.0	114.1	114.1
O ₁₃ -C ₃ -C ₄	110.4	111.7	112.4	109.6	109.6	109.4
C ₁₂ -O ₁₃ -C ₃	113.5	115.0	115.5	115.7	115.3	114.1

Table 2.11 (Continued)

Parameter	X-ray	CNDO	AM1	HF/ 3-21G	HF/ 6-31G**	B3LYP/ 6-31G**
C ₆ -C _{5a} -C _{12a}	113.2	112.5	111.9	111.7	113.0	112.8
C _{8a} -C _{12a} -C _{5a}	112.8	111.1	112.9	113.3	112.8	112.9
O ₂ -C ₃ -C ₄	113.2	113.1	111.6	111.0	111.3	111.9
O ₁₁ -C ₁₂ -O ₁₃	105.9	105.5	101.0	108.8	107.7	107.5
C _{5a} -C ₆ -C ₇	111.6	108.6	111.7	111.0	111.1	111.4
C ₈ -C _{8a} -C _{12a}	111.0	118.0	110.6	110.3	111.3	111.2
C ₁₅ -C ₃ -C ₄	114.3	113.1	113.6	115.3	114.2	114.5
C ₁₆ -C ₆ -C _{5a}	110.8	113.8	110.8	111.3	112.1	112.1
C ₉ -C _{8a} -C _{12a}	109.2	107.3	108.8	108.5	109.0	109.1
C ₁₀ -O ₁₁ -C ₁₂	124.6	74.2	121.0	125.9	125.6	124.6
O ₁₄ -C ₁₀ -O ₁₁	117.2	112.1	111.6	119.2	118.9	118.3
C ₁₇ -C ₉ -C _{8a}	113.4	122.4	112.6	114.1	114.5	114.5
Torsion Angle (°)						
C ₄ -C ₅ -C _{5a} -C _{12a}	323.9	313.6	315.1	321.4	324.7	324.0
C ₃ -C ₄ -C ₅ -C _{5a}	56.2	57.7	61.5	60.4	55.1	56.4
O ₁₃ -C ₃ -C ₄ -C ₅	25.8	31.4	25.6	21.7	26.3	25.7
C ₁₂ -O ₁₃ -C ₃ -C ₄	272.7	283.9	281.4	271.7	269.7	270.6
O ₁ -C _{12a} -C _{5a} -C ₅	69.0	73.7	73.2	68.8	67.5	68.5
O ₂ -C ₃ -C ₄ -C ₅	265.8	273.3	269.7	263.3	267.2	265.4
O ₁₁ -C ₁₂ -O ₁₃ -C ₃	258.1	241.5	245.7	262.1	260.7	259.2
C ₇ -C ₆ -C _{5a} -C _{12a}	46.1	63.8	49.9	50.0	49.3	48.8
C ₈ -C _{8a} -C _{12a} -C _{5a}	55.5	33.0	53.8	56.4	53.1	53.1

2.5 Conformational Problem

Although the main structure of artemisinin compounds is rigid due to three fused-rings skeleton, some substituent groups on these rings are quite flexible. Therefore, the compounds could exist in several conformations. It is then crucial to determine which conformations should be used in the study. The most important conformation for the drug development is the “bioactive conformation”, the conformation when the compound binds to its target receptor. But there is a problem of how to find this conformation.

Generally, the conformation of the drug is adjusted after binding with the receptor. Therefore, the bioactive conformation is not necessary the same as that of the lowest energy of unbound state. The exact bioactive conformation can be identified only by experimental work from the complex compound between drug and receptor. But this experiment is very difficult and extremely time-consuming. As a result, two proposed solutions were considered. The first possibility, the more preferable way, is to assume the lowest energy conformation of unbound state obtained from the conformational analysis as the bioactive conformation. Another solution is not to establish the bioactive conformation but simply to choose the same conformation as that of the parent structure for all compounds.

To follow the first solution, the conformational analysis must be applied. Two force fields, the MM3 and the Tripos force field, were applied for this purpose. Compound 21, 29, 44, and 61, which have long chain substituent groups, were randomly selected in order to determine suitable force field for our systems. The lowest energy conformations obtained from both force fields, denoted as MM3 and Tripos, together with the non-analysis conformation (“no search”) of all 4 compounds were then optimized at HF/3-21G level. The optimized conformations for each compound were superimposed as illustrated in Figure 2.10. The energy comparisons of all three conformations are shown in Table 2.12.

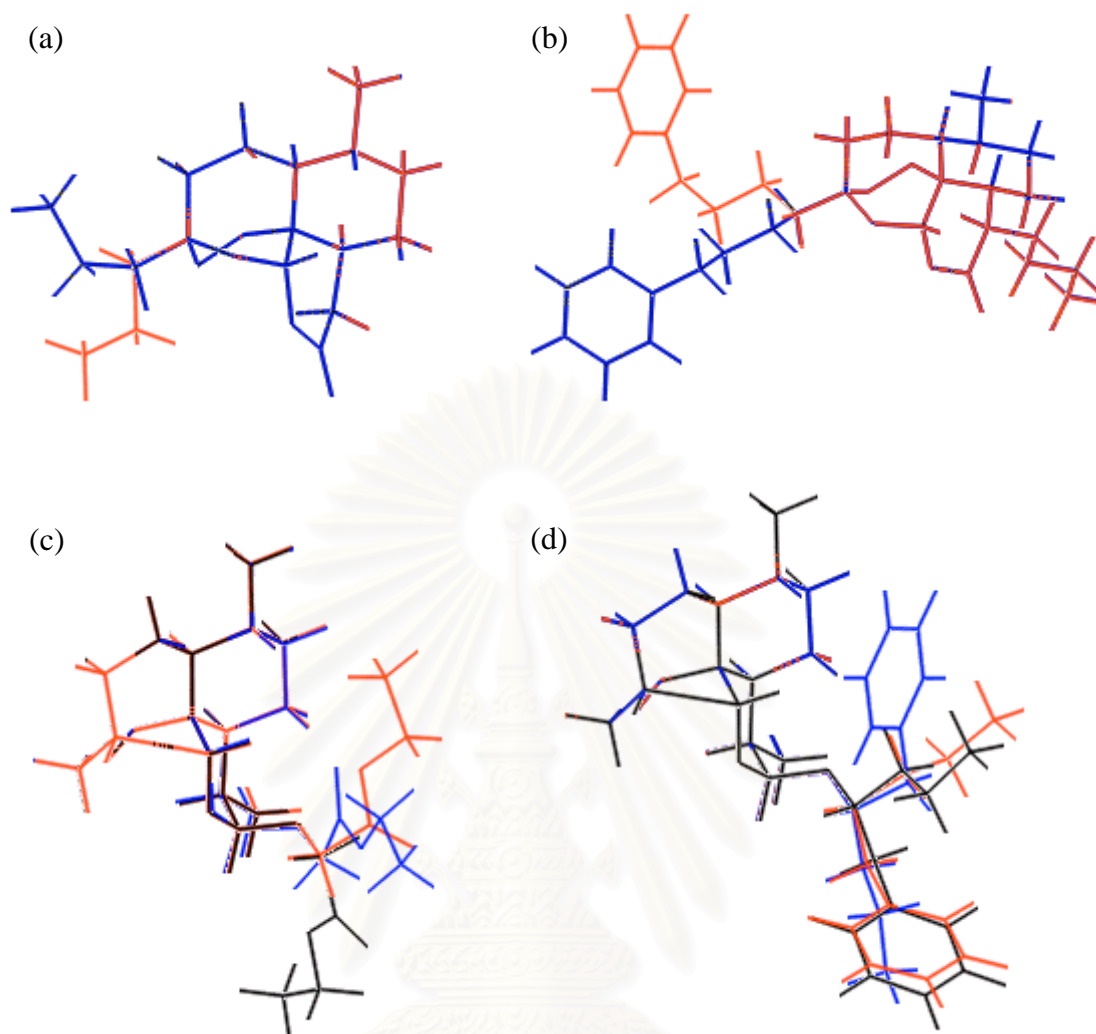


Figure 2.10 Superimposed structures of no search (blue), MM3 (black), and Tripos (red) conformations for compounds (a) 21, (b) 29, (c) 44, and (d) 61.

Table 2.12 Energy comparisons of different conformations of artemisinin compounds optimized at HF/3-21G level.

No	Conformational Search Method	Energy (a.u.)	ΔE (kcal/mol)
21	No search	-988.61697	0.00
	MM3	-988.61697	0.00
	Tripos	-988.62248	-3.46
29	No search	-1410.98601	0.00
	MM3	-1410.98601	0.00
	Tripos	-1410.98615	-0.09

Table 2.12 (Continued)

No	Conformational Search Method	Energy (a.u.)	ΔE (kcal/mol)
44	No search	-1253.97070	0.00
	MM3	-1253.97682	-3.84
	Tripos	-1253.96991	+0.50
61	No search	-1373.32246	0.00
	MM3	-1373.32523	-1.74
	Tripos	-1373.32561	-1.98

For compound 21, the MM3 conformation is the same as the “no search” conformation. Hence, their energies are equal. The Tripos conformation has lower energy than those of other two conformations ($\Delta E = -3.46$ kcal/mol). Compound 29 exhibits the same character as compound 21 but with only slightly lower energy of the Tripos conformation ($\Delta E = -0.09$ kcal/mol). In compound 44, the MM3 conformation has the lowest energy with ΔE of -3.84 kcal/mol while the Tripos conformation has the highest energy with ΔE of $+0.50$ kcal/mol. For compound 61, both MM3 and Tripos conformations have lower energy than that of the “no search” conformation. From these results, the Tripos force field seems to be the best choice for conformational analysis of our compounds. However, this force field does not guarantee the lowest energy conformation as in the case of compound 44. Moreover, this force field as well as the MM3 force field could not be applied to some compounds due to the lack of parameters. Furthermore, the conformational analysis was done in the gas phase not in the solution phase. The most stable conformation (lowest energy) of molecule, which is the most favorable orientation of dipoles under dielectric constant of the solvent, is depending on the medium used. But an information on the medium of the bioactive conformation is not available and the calculation that includes solvent effect in the system, e.g., molecular simulations, is time-consuming, therefore, it is not applicable to do such calculations for all compounds in this work.

In summary, from the above mentioned obstacles, it seems to be reasonable to choose the second solution for our research, i.e., simply choose the same conformation for all compounds by using the artemisinin and dihydroartemisinin X-ray structures as

the template to construct the structure of compounds with and without C=O at the C₁₀ position, respectively.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 3

Traditional QSAR

3.1 Introduction

A new drug does not just produce desired responses with minimum side effects. It, in addition, must be better than the existing therapies. Thus, the drug development process is supposed to be very long and expensive as schematically represented in Figure 3.1. In the first step, the discovery step, a lead compound is discovered and identified, which can be done in many ways, for example, by extraction from herbs or animals, by chemical synthesis in laboratory, and by modification from known drugs. In the next step, the compound undergoes testing *in vitro* to assess its biological activity. If the compound does not exhibit any acceptable potency, it is then disregarded and we have to return to the first step for a new compound. Until an active compound is found, we can then proceed to the next step, testing its biological activity in animal. This *in vivo* test is much more complicated than the *in vitro* test, and the active compound found in the *in vitro* test does not necessarily be active compound in the *in vivo* test. Again, if the compound does not show good activity, it is ignored and we have to go back to the discovery step again. But if the compound is considerably active, we can go further to the next step to test for toxicity and side effect. Since the compound may cause the side effect in long term, this step will require quite a long time, in average 3-5 years. During this time, the pharmaco-kinetic and pharmaco-dynamic of the compound are studied. After the compound was certainly proved to have low toxicity and no serious side effect, we can move to test the activity in human. As in animal, testing process for its toxicity and investigating the pharmaco-kinetic and pharmaco-dynamic should be carefully done. The time required for these steps are around 5-10 years. Moreover, in order to account for all races of human, the compound has to be tested on a large scale, the so-called clinical trial. Finally, it must undergo a review and approval process by the responsible government agency before it can be marketed. In summary, all the processes require in average 12 years and more than 200 million-dollar budget [59].

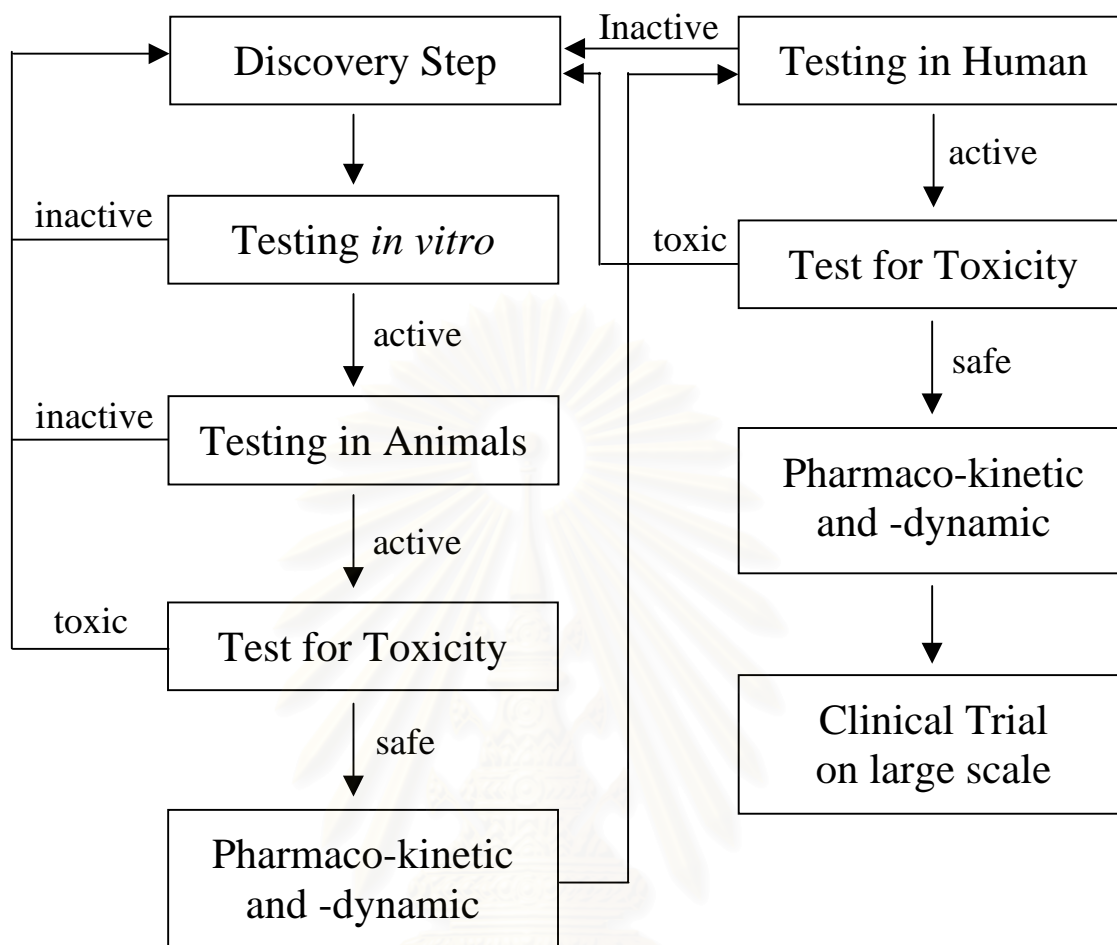


Figure 3.1 Steps in drug discovery and development process.

From the above paragraph, it is clear that a lot of money and time must be spent for a production of one new drug. Therefore, it would be great if there is a method that enables the prediction of biological activity of new compounds in advance based on knowledge of the chemical structure alone. Hence, it will cut down the number of analogues which have to be made. As a result, it greatly reduces the syntheses and testing biological activity efforts. Definitely, this will save a lot of time and money. Moreover, from this method, it can also help in deciding which features of a molecule give rise to its activity and how to make modification to enhance the activity of a compound. A Quantitative Structure-Activity Relationship (QSAR) is such a method. It relates molecular properties expressed in numerical values to the activity via a mathematical model.

3.1.1 Historical Outlines of QSAR

The history of QSAR started from the 19th century where some scientists noticed the relation between biological effect and chemical properties. For example, it was found that the toxicity of alcohols to mammals inversely follows their water solubility [60]. However, the relationship was not expressed in a quantitative way. The first general formulation of QSAR is considered to be an equation published by Crum-Brown and Fraser in 1868 [61]. In their studies of different alkaloids, the alkylation of the basic nitrogen atoms was recognized to produce significantly different biological effects. Therefore, they proposed that the "physiological activity", Φ , must be a function of the chemical structure C.

$$\Phi = f(C) \quad \text{.....(3.1)}$$

In 1893 Richet published the quantitative relationship between the toxicity of organic compounds and their water solubility [62] as shown in equation (3.2), where $\Delta\Phi$ are the differences in biological activity values and ΔC are changes in the properties.

$$\Delta\Phi = f(\Delta C) \quad \text{.....(3.2)}$$

Meyer at the University of Marburg in 1899 [63] and Overton at the University of Zürich in 1897 [64], working independently, observed linear relationships between potencies of narcotics and the partition coefficient of the compound between oil and water ($\log P$), see equation (3.3). The potencies of narcotics were expressed in terms of $\log (1/C)$ where C is a molar concentration producing iso-narcosis in tadpoles. For this study, 51 compounds were used for establishing the relation (3.3). The statistical fits were performed and the correlation coefficient (r) of 0.971 together with the standard deviation (s) of 0.280, which indicates the high degree of linear relationship, were obtained. (Details about these statistical values are given in the section 3.1.3.) From this linear relationship, Overton interpreted that the narcotic effect is due to physical changes caused by the dissolution of the drug in the lipid component of cells [65].

$$\log (1/C) = 0.94 \log P + 0.87 \quad \text{.....(3.3)}$$

$$r = 0.971, \quad n = 51, \quad s = 0.280$$

Later, Fühner found that within homologous series the narcotic activities increase in a geometric progression: 1, 3, 3², 3³, etc., as the number of carbon atoms increases [66]. This gave the first evidence of an additivity of group contributions to biological activity. From this time, there was little development of QSAR until the work of Hammett in 1937 [67], who correlated electronic properties of organic acids and bases with their equilibrium constants and reactivity. At first, he observed that adding substituents to the aromatic ring of benzoic acid had an orderly and quantitative effect on the dissociation constant, i.e., electron withdrawing group increases dissociation constant while electron donating group decreases that value. Later he also observed that substituents have a similar effect on the dissociation of other organic acids and bases. Therefore, the general correlation equation, the Hammett equation, was given as following.

$$\log (k/k') = \rho\sigma \quad \text{.....(3.4)}$$

where k = dissociation constant of the substituted compound
 k' = dissociation constant of the unsubstituted compound
 ρ = proportional constant relating the effect of substituent on that equilibrium to the effect of those substituent on the benzoic acid equilibrium
 σ = electronic parameter

The magnitude of the σ gives the relative strength of the electron-withdrawing or -donating properties of the substituents. σ is positive if the substituent is electron-withdrawing and negative if it is electron-donating. Many variables have been derived using the concept of the Hammett equation, such as lipophilicity parameter π [68], field (inductive) parameter, and resonance parameter [69]. And many QSAR works have been done using the Hammett parameters. However, the first (and for a long time the only one) real Hammett-type relationship was just published in 1962 by Hansen [70] which related the toxicities of substituted benzoic acids with the electronic σ constants of their substituents (equation 3.5). Noted that the F value indicates a level of statistical significance of the linear model (see section 3.1.3 for more details).

$$\log (1/C) = 1.454 \sigma + 1.787 \quad \text{.....(3.5)}$$

$r = 0.918$, $n = 13$, $s = 0.243$, $F = 58.91$

After the discovery of Hammett equation, the QSAR methodologies were gradually developed. In 1964, the time was ready for the modern QSAR methodology. Two new exciting methods were initially developed. One is the Hansch analysis (linear free energy-related approach, extrathermodynamic approach) [71], and another is the Free Wilson analysis (additivity model) [72]. The real breakthrough in QSAR resulted from the combination of different physicochemical parameters in a linear additive manner (equation 3.6) with the use of multiple linear regression method. Further contributions were the formulation of a parabolic equation for the quantitative description of nonlinear lipophilic-activity relationships (equation 3.7). Some improvements were done on the combination of Hansch equations with indicator variables [73], which may be considered as a mixed Hansch/Free Wilson model.

$$\log (1/C) = a (\log P) + b\sigma + \dots + \text{constant} \quad \dots\dots(3.6)$$

$$\log (1/C) = a (\log P)^2 + b (\log P) + c\sigma + \dots + \text{constant} \quad \dots\dots(3.7)$$

During the past four decades, many new parameters have been developed. Moreover, with the help of high efficiency computers, enormous numbers of various chemical properties and parameters could be calculated. In addition, a lot of powerful QSAR softwares are commonly available. This may lead to the case that the numbers of variables is larger than that of the objects, hence, the use of the regression analysis to derive the QSAR model is impossible. Therefore, many new statistical analysis techniques, such as the principle component analysis (PCA) and the partial least squares (PLS) [74], have been developed to handle such a case. This stimulates the growing of QSAR usage. To date, numerous QSAR applications have been published in many international journals.

3.1.2 Physicochemical Properties

Hundreds of different physical, structural, and chemical parameters have been used in the QSAR studies. In the early era (1970s), only a few parameters were recorded. The numbers of parameters used were gradually increased to more than 220 different parameters by 1980s. And they were exponentially increased to more than 17,000 different parameters by 1990s [75]. Although new parameters are steadily invented, there is still a lack of adequate parameters to describe some important interactions like the membrane partitioning of drugs, the strength of hydrogen bonds,

the influence of desolvation energies on drug-receptor affinity, and steric interactions with a binding site.

With the improvement of hardware and software, numbers of parameters are being developed and tried. Among these, the most commonly used are hydrophobic, polar, electronic, and steric properties. This is because they are easily to be quantified. Here, parameters are categorized into four types and are discussed below.

3.1.2.1 Hydrophobicity Parameters

No other physicochemical property has attracted as much attention in QSAR as hydrophobicity. The hydrophobicity determines how easy for a drug to cross cell membranes and is also important in assessing the drug-receptor interactions. A drug should have a suitable hydrophobicity. If it is too hydrophilic, it could not cross the hydrophobic cell membrane to the target site. But if it is too hydrophobic, it would quickly be extracted from an aqueous bloodstream and be stored in the fatty tissues of the body. Consequently, it does not reach the intended target site. This is the reason why parabolic relationships between hydrophobicity and activity are frequently seen. Changing substituents on a drug may well have significant effects on its hydrophobic character and hence its biological activity.

The hydrophobic character of a drug is widely presented by the partition coefficient (P), which is defined as the ratio of drug concentrations in the organic and aqueous phases of a two-compartment system under equilibrium conditions (equation 3.8). Although many organic/aqueous systems have been used since the past time, an *n*-octanol/water system is now accepted as the standard system for the experimental measurement of the P.

$$P = \frac{\text{Concentration of drug in organic phase}}{\text{Concentration of drug in aqueous phase}} \quad \dots\dots(3.8)$$

From the definition of P, it is obvious that hydrophobic compounds will have a high P value, whereas hydrophilic compounds will 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 would be much better if we could calculate P theoretically.

Since the scale of numbers involved in measuring P usually cover several factors of ten, the logarithm unit is used instead to allow more manageable numbers. The log P value can be calculated by using its additive property and there are a lot of molecular modeling softwares available for this purpose. In the software database, the hydrophobicity contribution of various substituents and the experimental log P values of as many compounds as possible are stored. By adding and/or subtracting these contributions to an experimental log P value of a parent compound, the log P value of the desired compound could be obtained. Both experimental and theoretical log P values play an important role in the QSAR studies as indicated by its contributions to many QSAR equations.

3.1.2.2 Polarizability Parameters

Since polarizability is the ability of electrons to change position in response to the presence of an outside electrical field, it is thus related to the electronic properties, size, and polarity of a compound. Therefore, it may have an influence on the drug-receptor interactions. Molar refractivity (MR), which is calculated by the equation (3.9), is the most utilized polarizability parameter.

$$MR = \frac{MW}{d} \cdot \frac{n^2 - 1}{n^2 + 2} \quad \dots\dots (3.9)$$

where MW = molecular weight

d = density

n = refractive index

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

3.1.2.3 Electronic Parameters

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. 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 dipole moments (μ), moment of inertia (I), hydrogen-bonding parameters, and parameters derived from quantum chemical calculation, e.g., orbital energies and partial atomic charges.

3.1.2.4 Steric Parameters

Generally, a drug has to approach and interact with its receptor to mediate its effects. The bulk, size, and shape of the drug must be appropriate for the binding. In some cases, a bulky substituent may act like a shield, which hinders the drug-receptor interaction. Alternatively, it may help a drug to orient properly for the maximum drug-receptor binding which then increases the activity. Therefore, it is worth to investigate a relationship between activity and steric character. Steric effects are sometimes difficult to describe since the 3D structures of the binding sites of drug are most often unknown, however, if the structure is known, there is still some problems to calculate the steric effects quantitatively. Such problems are, for example, different conformations of different ligands, small (or even large) differences in the binding modes of different analogs, and variations in the positions of side chains and even backbone atoms of the protein in different ligand-protein complexes.

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 theory [76] as the basis, are also widely used. Examples of these indices are the Wiener index [77], molecular connectivity indices (Chi) [78], valence-modified molecular connectivity indices (ChiV) [78], and molecular shape indices (Kappa) [78]. 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 [79].

3.1.3 Regression Analysis

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 (3.13). 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.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad \dots(3.13)$$

where y = dependent variable

X_1, X_2, X_3, \dots = independent variables

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

Since 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 3.14

and 3.15) of n chemical compounds, a model with only one X variable, the regression coefficients could be evaluated as following.

$$y_{\text{obs}} = \beta_0 + \beta_1 X + \varepsilon \quad \dots(3.14)$$

$$y_{\text{cal}} = \beta_0 + \beta_1 X \quad \dots(3.15)$$

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 $f = \Sigma (y_{\text{obs}} - \beta_0 - \beta_1 X)^2$ with respect to β_0 and β_1 are set to zero, i.e., $df/d\beta_0 = df/d\beta_1 = 0$ (equations 3.16 and 3.17).

$$df/d\beta_0 = 2 \cdot \Sigma (y - \beta_0 - \beta_1 X) \cdot (-1) = 0 \quad \dots(3.16)$$

$$df/d\beta_1 = 2 \cdot \Sigma (y - \beta_0 - \beta_1 X) \cdot (-X) = 0 \quad \dots(3.17)$$

Second, the so-called normal equations (3.18) and (3.19) are then resulted from equations (3.16) and (3.17).

$$\Sigma y = n\beta_0 + \beta_1 \Sigma X \quad \dots(3.18)$$

$$\Sigma (Xy) = \beta_0 \Sigma X + \beta_1 \Sigma (X)^2 \quad \dots(3.19)$$

Finally, the regression coefficients β_0 and β_1 (equations 3.20 and 3.21) could be obtained by mathematical solving of the equations (3.18) and (3.19).

$$\beta_1 = \frac{n\Sigma(Xy) - (\Sigma X) \cdot (\Sigma y)}{n\Sigma(X^2) - (\Sigma X)^2} \quad \dots(3.20)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{X} \quad \dots(3.21)$$

where \bar{y} = mean of y variable

\bar{X} = mean of x variable

For the multiple linear regression equation (equation 3.13), 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 3.22). 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 (3.23) to (3.25).

$$s = \text{SSE} / (n-k-1) \quad \dots(3.22)$$

$$\text{SSE} = \sum \varepsilon^2 = \sum (y_{\text{observe}} - y_{\text{calculate}})^2 \quad \dots(3.23)$$

$$\text{SSR} = \sum (y_{\text{calculate}} - y_{\text{mean}})^2 \quad \dots(3.24)$$

$$\text{SST} = \text{SSR} + \text{SSE} = \sum (y_{\text{observe}} - y_{\text{mean}})^2 = \sum y^2 - (\sum y)^2/n \quad \dots(3.25)$$

The second and most popular indicator used to measure the quality of the QSAR model is the Pearson correlation coefficient, r (equation 3.26). 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.

$$r = \frac{\sum(xy) - (\sum X)(\sum Y)/n}{\sqrt{\sum(X^2) - (\sum X)^2/n} \cdot \sqrt{\sum(Y^2) - (\sum Y)^2/n}} \quad \dots(3.26)$$

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 a regression model to the total variance (equation 3.27). 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, $\sum \varepsilon^2$) goes to zero, r^2 goes to 1. Then, the equation can explain all 100% of the variation in the biological activity.

$$\begin{aligned} r^2 &= \sum (y_{\text{calculate}} - y_{\text{mean}})^2 / \sum (y_{\text{observe}} - y_{\text{mean}})^2 \\ &= \text{SSR} / \text{SST} = 1 - \text{SSE} / \text{SST} \quad \dots(3.27) \end{aligned}$$

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 (3.28). 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 limits 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.

$$F = \frac{r^2 \cdot (n - k - 1)}{k \cdot (1 - r^2)} \quad \dots(3.28)$$

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 and 0.9 ($r^2 \geq 0.81$) for *in vitro* data [75]. 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 0.5 [80].

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 an information 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 [81], bootstrapping, random change of the values of the dependent variable, and dividing the original set into the training set and testing set. 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 3.29 to 3.31). 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 [82].

$$\text{PRESS} = \sum (y_{\text{observed}} - y_{\text{predicted}})^2 \quad \dots\dots(3.29)$$

$$q^2 = 1 - \text{PRESS}/\text{SST} \quad \dots\dots(3.30)$$

$$s_{cv} = (\text{PRESS}/n)^{1/2} \quad \dots\dots(3.31)$$

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, its real predictive power is obtained.

3.1.4 Previous QSAR Studies

There are only few QSAR publications on artemisinin and its derivatives. The first QSAR study on artemisinin compounds was published in 1982 by a Chinese group [83]. In that work, various ethers and ester derivatives of dihydroartemisinin were synthesized and their biological activities were measured in terms of SD_{90} (the dose required for 90% suppression of the parasitemia). The log P values of 49 compounds were determined from 1-octanol/water system by HPLC and were used to study the correlation with their antimalarial activities. A non-linear QSAR equation was derived by the regression analysis as following.

$$\log (1/C) = 1.120 - 0.198 (\log P)^2 + 1.155 (\log P) - 0.612 (I_{\alpha,\beta}) \quad \dots(3.33)$$

$$r = 0.911, \quad n = 49, \quad s = 0.153$$

where C = concentration

$I_{\alpha,\beta}$ = indicator variable which value is 1 for β -epimers and is 0 for α -epimers

In the equation (3.33), the relationship between activity and log P is in a parabolic form. Therefore, the suitable log P values were calculated to be between 2.60 and 2.90. However, it is not clear whether the biological activities of pure β - and α -epimers were correctly measured. Moreover, we could not confirm their results since the structures and antimalarial activities of some compounds were not given in the literature.

In 1996, the same Chinese group published another work on 21 ether and ester derivatives of 10-dihydroartemisinin (Figure 3.2) [84]. All structures were geometrically optimized by the semiempirical AM1 and PM3 methods. Five quantum chemical indices, C_{18} atomic charge (Q_{C18}) and four bond order parameters (P_{O2-C3} , $P_{O1-C12a}$, P_{O1-O2} , and $P_{O14-C18}$), were calculated at the same levels as geometry optimizations. These five chemical indices were supplied as input parameters in the neural network method for the statistical analysis. The QSAR models with quite good predictive capability were constructed for both AM1 and PM3 data. The statistical values for both models are as follows.

For AM1 model: $r^2 = 0.863$, $q^2 = 0.526$, $s^2 = 0.016$, $n = 21$

For PM3 model: $r^2 = 0.946$, $q^2 = 0.713$, $s^2 = 0.006$, $n = 21$

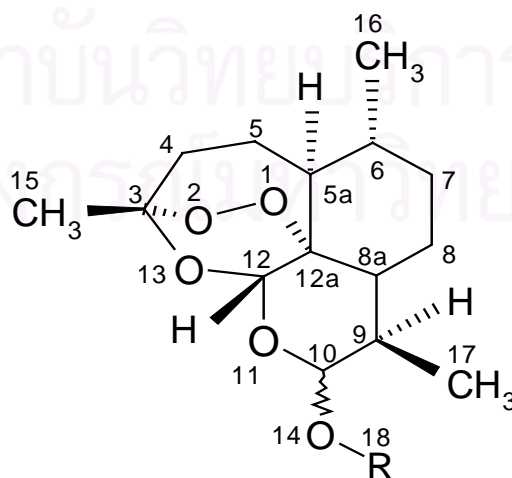


Figure 3.2 The common structure of artemisinin derivatives used in Ref. [84].

The relationships between the 5 indices and the antimalarial activity suggested that the stronger the O₁-O₂ bond and the weaker the O₂-C₃ and O₁-C_{12a} bonds the higher the antimalarial activity. Also, the activity increased as the net atomic charge on the C₁₈ and the bond order of O₁₄-C₁₈ increased. Based on those results, the proposal was made that artemisinin analogues should inactivate heme in red blood cells by oxidizing iron(II) ion to iron(III) ion. Then, if the bond O₁-O₂ is stronger and the bonds O₂-C₃, O₁-C_{12a} are weaker, the oxidative reaction occurs more easily. Furthermore, a highly positive C₁₈ and strong O₁₄-C₁₈ bond results in strong electrostatic interaction with the negative charge of the receptor.

In 1997, one more publication appeared by the same Chinese group [85]. All compounds and methodologies were the same as in their previous work [84] except the method for statistical analysis. In this publication, the Partial Least Squares (PLS) method was employed instead of the neural network method. The QSAR models were constructed by omitting 3 and 1 compounds for AM1 (equation 3.34) and PM3 (equation 3.35), respectively. The obtained models are as follows.

$$\begin{aligned}
 -\log C &= -217.745 + 0.434 Q_{C18} + 207.065 P_{O1-O2} - 35.698 P_{O2-C3} \quad \dots(3.34) \\
 &\quad + 47.029 P_{O1-C12a} + 4.536 P_{O14-C18} \\
 r^2 &= 0.832, \quad q^2 = 0.750, \quad n = 18, \quad s = 0.148, \quad F = 37.162
 \end{aligned}$$

$$\begin{aligned}
 -\log C &= 722.216 + 0.338 Q_{C18} - 973.297 P_{O1-O2} + 138.006 P_{O2-C3} \quad \dots(3.35) \\
 &\quad + 58.393 P_{O1-C12a} + 6.214 P_{O14-C18} \\
 r^2 &= 0.813, \quad q^2 = 0.712, \quad n = 20, \quad s = 0.152, \quad F = 23.148
 \end{aligned}$$

Due to the contradictions in the sign of regression coefficients between AM1 and PM3 models, the P_{O1-O2}, P_{O2-C3}, and P_{O1-C12a} were combined together to give a new parameter, P, i.e., P = P_{O1-O2} + P_{O2-C3} + P_{O1-C12a}. New QSAR models for AM1 (equation 3.36) and PM3 (equation 3.37) were evaluated with three parameters. Notably, the resulting r² and q² values were slightly decreased and one more compound was additionally omitted for the PM3 model.

$$\begin{aligned}
 -\log C &= -113.068 + 0.586 Q_{C18} + 4.870 P_{O14-C18} + 39.386 P \quad \dots(3.36) \\
 r^2 &= 0.799, \quad q^2 = 0.745, \quad n = 18, \quad s = 0.156, \quad F = 63.656
 \end{aligned}$$

$$-\log C = -129.959 + 0.558 Q_{C_{18}} + 4.655 P_{O_{14}-C_{18}} + 44.576 P \quad \dots(3.37)$$

$$r^2 = 0.807, \quad q^2 = 0.757, \quad n = 19, \quad s = 0.149, \quad F = 70.947$$

In that publication, the pharmacophore of artemisinin derivatives was also deduced to be a triangle formed with O₁, O₂, and C₁₈. But this pharmacophore model probably holds true only for the ether and ester derivatives of 10-dihydroartemisinin because many active compounds, such as artemisinin, lack the C₁₈ atom.

Another QSAR work was also done on the 10-dihydroartemisinin derivatives [86]. The QSAR models of 17 dihydroartemisinin derivatives were derived from atomic net charges by using Multivariate Adaptive Regression Splines (MARS), a nonlinear statistical analysis method. The biological activities were measured against D-6 and W-2 clones of *P. falciparum*. The structures were optimized using the semiempirical molecular orbital PM3 method. Subsequently, the Mulliken population analysis was used to calculate the net atomic charges. For the D-6 activity, the atomic charges of O₂, C_{5a}, C₈, and O₁₁ were presented in the model with r² and q² values of 0.949 and 0.896, respectively. The three-dimensional perspective plots were used to investigate the joint dependence of the activity on the atomic net charges of the model. The plots suggested that the antimalarial activity would increase if the atomic net charges at O₂ and O₁₁ decreased and those at C_{5a} and C₈ were kept at moderate positive values. For the W-2 activity, the O₁, C_{5a}, C₈, and O₁₁ were presented in the model with r² and q² values of 0.955 and 0.872, respectively. The investigations on graphical representations of the model revealed that in order to improve the antimalarial activity, the atomic net charge at O₁₁ should be decreased and those at O₁, C_{5a}, and C₈ should be kept at rather moderate values.

In a recent QSAR publication [87], a large data set of 202 artemisinin derivatives with the antimalarial activity against the W-2 clone of *P. falciparum* was used. All the structures were geometrically optimized with the empirical force field. Correlations between activity and 75 calculated chemical descriptors were investigated using a stepwise regression. The best correlation was found with 6 chemical descriptors: logP, molar refractivity, moment of inertia, Balaban Index, and 1 Å and 2 Å 3D autocorrelograms. However, this model is still statistically unimpressive with r² = 0.518, q² = 0.485, s = 0.738, and F = 28.6. With the attempt to find a better correlation,

the compounds were split into less structurally diverse data sets. Again, however, no good correlation was found.

3.2 Calculations of Properties

There are a lot of parameters that can be used in the field of QSAR. Totally, 102 physicochemical parameters were calculated using the TSAR [79] and Gaussian [57-58] softwares.

A. Hydrophobicity properties

For the hydrophobicity parameter, the log P was calculated. Moreover, the desolvation free energies for water (F_{h_2o}) and for octanol (F_{oct}), which indicated the ease for desolvating solute molecule (water or octanol) from a drug compound, were also included. This is important for the drug-receptor binding process since both drug and receptor must become at least partially desolvated before the binding.

B. Polarizability properties

For the polarizability parameter, the molar refractivity was calculated using the TSAR software.

C. Electronic properties

For the electronic parameters, atomic charges obtained from the Gaussian software were used. Since the Mulliken Population Analysis (MPA) method is known to be dependent on the basis set used, therefore, the Natural Population Analysis (NPA) method, which is much less sensitive to the basis set, was used instead for the atomic charge calculations. Atomic charges of 17 atoms namely O₁, O₂, C₃, C₄, C₅, C_{5a}, C₆, C₇, C₈, C_{8a}, C₉, C₁₀, O₁₁, C₁₂, C_{12a}, O₁₃, and O₁₄ were computed at the HF/3-21G level.

Twenty bond orders [88] were calculated at the HF/3-21G level which are denoted as follows – B(1-2), B(1-12a), B(2-3), B(3-4), B(3-13), B(4-5), B(5-5a), B(5a-6), B(5a-12a), B(6-7), B(7-8), B(8-8a), B(8a-9), B(8a-12a), B(9-10), B(10-11), B(10-14), B(11-12), B(12-12a), and B(12-13). The bond order parameter indicates types of bonding. The bond order values of 1.0, 2.0, and 3.0 correspond to the single, double, and triple bonds, respectively.

Dipole moment, HOMO energy, and LUMO energy were also calculated at the HF/3-21G level. Moreover, six moments of inertia were computed using the TSAR software. They are calculated in all three perpendicular axes, which pass through the center of mass of a molecule. And these results are reported in the project as Inertia Moment 1 Size, Inertia Moment 2 Size, Inertia Moment 3 Size, Inertia Moment 1 Length, Inertia Moment 2 Length, and Inertia Moment 3 Length. The moment of inertia indicates the ability for rotation, which may involve the binding.

D. Steric properties

Structural parameters, 12 bond lengths (R), 14 bond angles (A), and 16 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 artemisinin in Figure 1.2 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(1-12a), R(2-3), R(3-4), R(3-13), R(4-5), R(5-5a), R(5a-12a), R(10-11), R(11-12), R(12-12a), R(12-13), A(1-2-3), A(1-12a-5a), A(2-1-12a), A(2-3-4), A(2-3-13), A(3-4-5), A(3-13-12), A(4-3-13), A(4-5-5a), A(5-5a-12a), A(5a-12a-12), A(10-11-12), A(11-12-13), A(12a-12-13), T(1-2-3-4), T(1-12a-5a-5), T(1-12a-12-13), T(2-1-12a-5a), T(2-1-12a-12), T(2-3-4-5), T(2-3-13-12), T(3-2-1-12a), T(3-4-5-5a), T(3-13-12-12a), T(4-3-13-12), T(4-5-5a-12a), T(5-4-3-13), T(5-5a-12a-12), T(5a-12a-12-13), and T(10-11-12-13).

Topological index after the Balaban method [89] 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.

3.3 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.

3.3.1 All 104 Compounds

In order to access the real predictive ability, compounds were divided into the training set (90%) and testing set (10%). Therefore, ten compounds were purposively chosen for the testing set, i.e., compound number 10, 20, 30, 34, 50, 60, 70, 80, 90, and 100, which cover all groups of different structures. And the remaining 94 compounds, the training set, were used to derive the model. For the log(D-6) activity, six parameters were presented in the model as shown in equation (3.38) with the r^2 and q^2 of 0.404 and 0.368, respectively. For the log(W-2) activity, also 6 parameters were employed in the model, giving the r^2 and q^2 values of 0.393 and 0.278, equation (3.39). These two models have r^2 values of less than 0.81, hence not acceptable models. However, considering the q^2 value, the log(D-6) model is acceptable model while the log(W-2) model is not.

$$\begin{aligned} \log D-6 = & 0.586 * \text{Kappa } 2 + 292.877 * \text{B}(1-2) + 69.379 * \text{O}_{13} \\ & - 43.088 * \text{O}_1 + 0.0972 * \text{F}_{\text{h}_2\text{o}} - 0.507 * \log P - 228.469 \quad \dots(3.38) \\ r^2 = & 0.404, \quad q^2 = 0.368, \quad s = 0.702, \quad n = 94, \quad F = 9.843 \end{aligned}$$

$$\begin{aligned} \log W-2 = & 0.59 * \text{Kappa } 2 - 93.713 * \text{B}(3-4) + 107.166 * \text{O}_{13} \\ & - 0.772 * \text{C}_9 + 0.0939 * \text{F}_{\text{h}_2\text{o}} - 0.499 * \log P + 146.051 \quad \dots(3.39) \\ r^2 = & 0.393, \quad q^2 = 0.278, \quad s = 0.738, \quad n = 94, \quad F = 9.384 \end{aligned}$$

In both models, a plus sign of the Kappa 2 parameter indicates that compounds with higher Kappa 2 values are more potent. Similarly, a plus sign of the $F_{\text{h}_2\text{o}}$ parameter refers that compounds with higher $F_{\text{h}_2\text{o}}$ values (hard to desolvate water molecules) have higher activities. Therefore, water molecules may play some important roles in the mechanism of action. In addition, a minus sign of the log P points out that compounds with low log P (dissolve more favorably in water than in octanol) are more active.

The comparison between actual and predicted values of 10 compounds in the testing set for log(D-6) and log(W-2) activities was displayed in Figure 3.3 and 3.4, respectively.

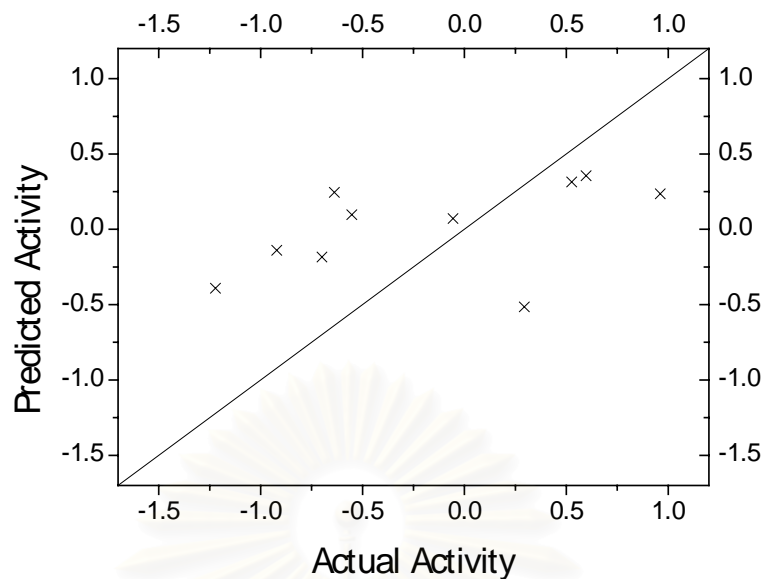


Figure 3.3 Comparison between actual and predicted log(D-6) activities for 10 compounds in the testing set.

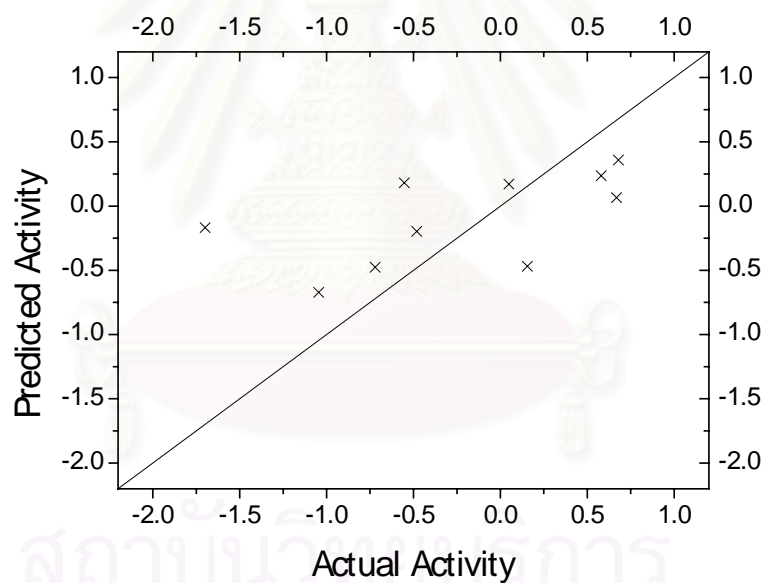


Figure 3.4 Comparison between actual and predicted log(W-2) activities for 10 compounds in the testing set.

From the models in equations (3.38) and (3.39) as well as the predicted activities in Figure 3.3 and 3.4, it can be seen that a good QSAR model for the whole set of compounds could not be derived from our calculated physicochemical properties. This may be contributed by the structural differences among the compounds (see Chapter 2). Therefore, the QSAR analysis was separately performed for each individual group of compounds with more closely related structures.

3.3.2 Compounds with C=O at C₁₀ position (Group 1)

Considering only the compounds having C=O group at the C₁₀ position (group 1 in Chapter 2), there are 41 compounds belonging to this group. Compound number 5, 9, 16, and 38 were randomly selected for the testing set. And the remaining 37 compounds were used for the training set. The results for log(D-6) and log(W-2) activities are shown in equations (3.40) and (3.41), respectively. The r^2 values in both models and the q^2 value of the log(D-6) model are not acceptable. But the q^2 value of the log(W-2) model is impressive.

$$\begin{aligned} \log D-6 = & 22.737*B(8a-9) + 164.793*O_{13} + 1045.67*C_6 \\ & + 0.0834*F_{h_{2o}} + 353.662 \end{aligned} \quad \dots(3.40)$$

$$r^2 = 0.501, \quad q^2 = 0.235, \quad s = 0.537, \quad n = 37, \quad F = 8.039$$

$$\begin{aligned} \log W-2 = & 68.86*B(2-3) + 33.71*B(10-11) + 196.844*O_{13} \\ & + 3.24*A(1-2-3) + 0.125*F_{h_{2o}} - 305.077 \end{aligned} \quad \dots(3.41)$$

$$r^2 = 0.604, \quad q^2 = 0.448, \quad s = 0.538, \quad n = 37, \quad F = 9.440$$

The comparison between actual and predicted values of 4 compounds in the testing set for log(D-6) and log(W-2) activities was displayed in Figure 3.5 and 3.6, respectively. The predictions are fairly good.

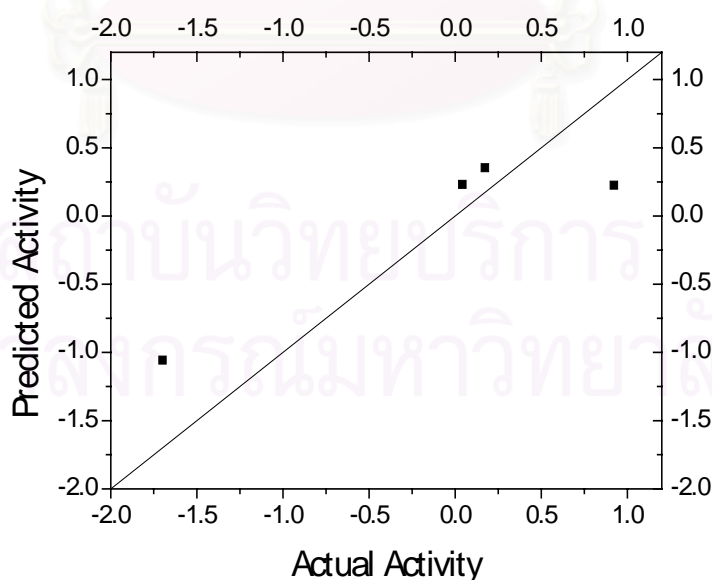


Figure 3.5 Comparison between actual and predicted log(D-6) activities for 4 compounds in the testing set.

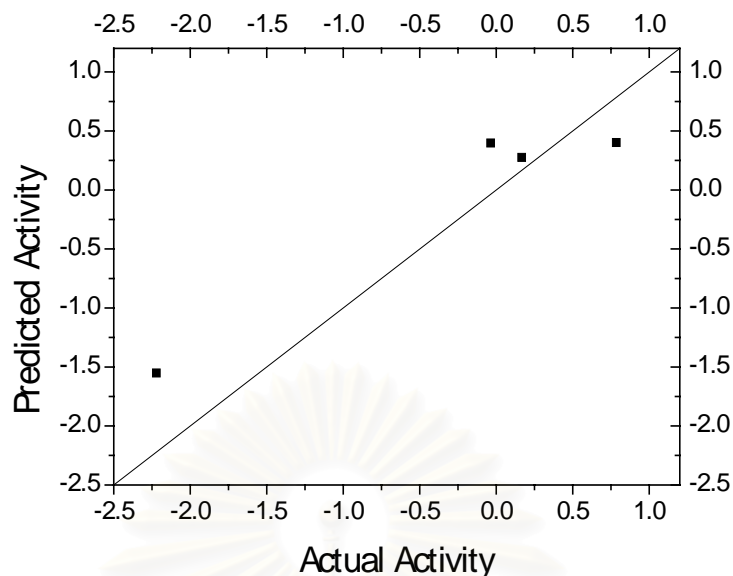


Figure 3.6 Comparison between actual and predicted $\log(W-2)$ activities for 4 compounds in the testing set.

In order to search for better predictive models, the compounds with very similar structures were considered. All 41 compounds in this group could be classified into 2 classes depending on the atom type at the position 11, i.e., oxygen atom or nitrogen atom. There are 31 compounds with an oxygen atom and 10 compounds with a nitrogen atom. Therefore, ten compounds with the N_{11} atom were excluded from both training and testing sets. The models derived from 28 compounds in the training set are shown in equation (3.42) and (3.43).

$$\begin{aligned} \log D-6 = & 172.22 * O_{13} - 562.97 * R(1-2) + 0.154 * F_{h_{2o}} \\ & - 0.575 * T(2-3-4-5) + 877.91 \end{aligned} \quad \text{.....(3.42)}$$

$$r^2 = 0.616, \quad q^2 = 0.457, \quad s = 0.535, \quad n = 28, \quad F = 9.207$$

$$\begin{aligned} \log W-2 = & 72.33 * B(3-13) + 1940.47 * C_6 - 485.67 * R(11-12) \\ & + 0.524 * T(10-11-12-13) + 1106.73 \end{aligned} \quad \text{.....(3.43)}$$

$$r^2 = 0.699, \quad q^2 = 0.623, \quad s = 0.510, \quad n = 28, \quad F = 10.233$$

The q^2 values are significantly improved from the previous models, i.e., 0.235 to 0.457 and 0.448 to 0.623 for $\log(D-6)$ and $\log(W-2)$ activities, respectively. Parameter with the highest correlation coefficient for $\log(D-6)$ model is $F_{h_{2o}}$ and for $\log(W-2)$ model is $T(10-11-12-13)$. An explanation for the importance of the $F_{h_{2o}}$ parameter in the model (3.42) is the same as described for the model (3.38) and (3.39) in the section

3.3.1. The T(10-11-12-13) parameter presented in the model (3.43) may involve in the structural change during reactions in the mechanism of action (see Chapter 6). Compound with larger T(10-11-12-13) has higher activity as indicated by a plus sign in the equation. For example, the angles in compound 2, 4 and 5, having substituent group at the α -C₉ position, are small, hence low activities. We suppose that compound must change its structure to have larger T(10-11-12-13) angle in the reaction mechanism. Therefore, compound with larger angle is easier to achieve this change.

From the investigation of residual value in the log(D-6) model, compound 21 shows the highest residual value. This is possibly due to its unusual high activity value as compared to other compounds with similar structures. For instance, the compound 20, 21, and 22 which have the substituents at the C₃ position as -CH₂CH₃, -(CH₂)₂CH₃, and -(CH₂)₃CH₃ respectively, possess the D-6 activity of 0.88, 21.02, and 0.20. This unusual high activity could not be explained by properties calculated in the section 3.2. Therefore, this compound was omitted and the quality of the model was clearly improved as shown in equation (3.44). The log(D-6) and log(W-2) activities of 3 compounds in the testing set were predicted using equation (3.44) and (3.43), respectively, and the results were shown in Table 3.1.

$$\begin{aligned} \log D-6 = & 203.54*O_{13} - 435.28*R(1-2) - 1.006*T(2-3-4-5) \\ & + 0.153*F_{h_{20}} + 668.995 \end{aligned} \quad \dots(3.44)$$

$$r^2 = 0.729, \quad q^2 = 0.625, \quad s = 0.430, \quad n = 27, \quad F = 14.798$$

Table 3.1 Activity predictions of 3 compounds in the testing set of group 1.

Compound No.	log(D-6) activity		log(W-2) activity	
	Actual	Predicted by (3.44)	Actual	Predicted by (3.43)
5	-1.699	-1.300	-2.222	-1.237
9	0.042	0.233	0.167	0.460
16	0.922	0.206	0.785	0.125

3.3.3 Compounds with -OR or -NR at C₁₀ position (Group 2)

Group 2 consists of 46 compounds with either -OR or -NR groups at the C₁₀ position. Compounds number 49, 55, 60, 73, and 86 were randomly selected for the testing set. The QSAR results for 41 compounds in the training set are given in equations (3.45) and (3.46). Although the r^2 values are slightly less than the acceptable values, the q^2 values of both models are quite impressive.

$$\begin{aligned} \log D-6 = & 0.00287 * \text{Inertia Moment 1 size} - 9.73 * B(11-12) - 58.09 * O_{11} \\ & + 0.487 * A(10-11-12) + 0.0415 * F_{h_{2o}} - 85.299 \quad \dots(3.45) \\ r^2 = & 0.710, \quad q^2 = 0.606, \quad s = 0.460, \quad n = 41, \quad F = 17.168 \end{aligned}$$

$$\begin{aligned} \log W-2 = & 0.00510 * \text{Inertia Moment 1 size} - 67.96 * B(2-3) - 2.74 * C_9 \\ & - 152.85 * R(12-12a) + 0.0331 * F_{h_{2o}} + 287.201 \quad \dots(3.46) \\ r^2 = & 0.711, \quad q^2 = 0.602, \quad s = 0.504, \quad n = 41, \quad F = 17.233 \end{aligned}$$

If only the compounds with -OR group at C₁₀ position were considered, statistically better models may be obtained. Therefore, 5 compounds with -NR group at the C₁₀ position were eliminated. For both activities, the same set of six parameters was presented in the models. By omitting the compound 52, which have the highest residual value, the obtained results are given in equations (3.47) and (3.48).

$$\begin{aligned} \log D-6 = & 0.0038 * \text{Inertia Moment 1 size} - 0.306 * \text{ChiV2 path} \\ & + 569.21 * B(1-2) + 0.918 * A(10-11-12) \\ & - 44.302 * B(9-10) + 0.0391 * F_{h_{2o}} - 568.506 \quad \dots(3.47) \\ r^2 = & 0.900, \quad q^2 = 0.801, \quad s = 0.267, \quad n = 36, \quad F = 43.498 \end{aligned}$$

$$\begin{aligned} \log W-2 = & 0.0051 * \text{Inertia Moment 1 size} - 0.274 * \text{ChiV2 path} \\ & + 565.27 * B(1-2) + 0.827 * A(10-11-12) \\ & - 45.605 * B(9-10) + 0.0358 * F_{h_{2o}} - 554.197 \quad \dots(3.48) \\ r^2 = & 0.823, \quad q^2 = 0.644, \quad s = 0.400, \quad n = 36, \quad F = 22.423 \end{aligned}$$

The r^2 and q^2 values for these two models are very impressive. In both models, the A(10-11-12) parameter has the highest correlation coefficient. This parameter may have a contribution to the structural change in the reaction mechanism. The moment of inertia parameter in the two models possibly represents the rotation of the substituent groups at the C₁₀ position. This is significant for their binding to the receptor molecule.

The relationships between actual and predicted values for log(D-6) and log(W-2) activities of 36 compounds in the training set using models (3.47) and (3.48) are shown in Figure 3.7 and 3.8, respectively.

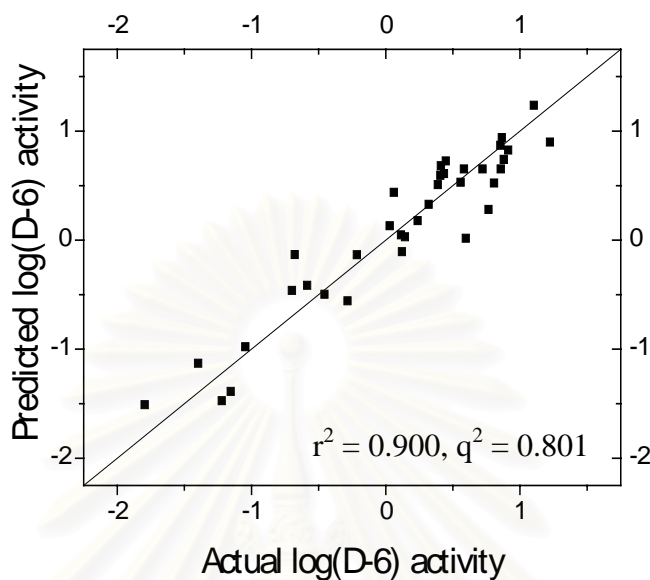


Figure 3.7 Relationship between actual and predicted log(D-6) activities for 36 compounds in group 2.

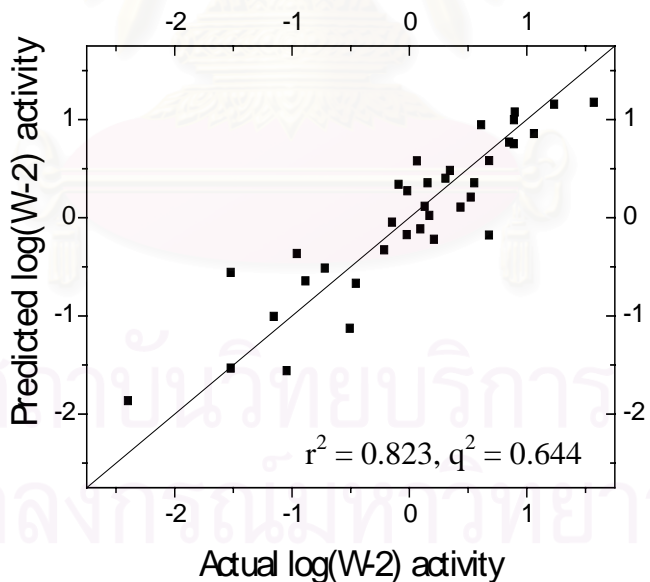


Figure 3.8 Relationship between actual and predicted log(W-2) activities for 36 compounds in group 2.

The comparison between actual and predicted values of 5 and 4 compounds in the testing set for log(D-6) and log(W-2) activities was displayed in Figure 3.9 and 3.10, respectively. The predictions are fairly good.

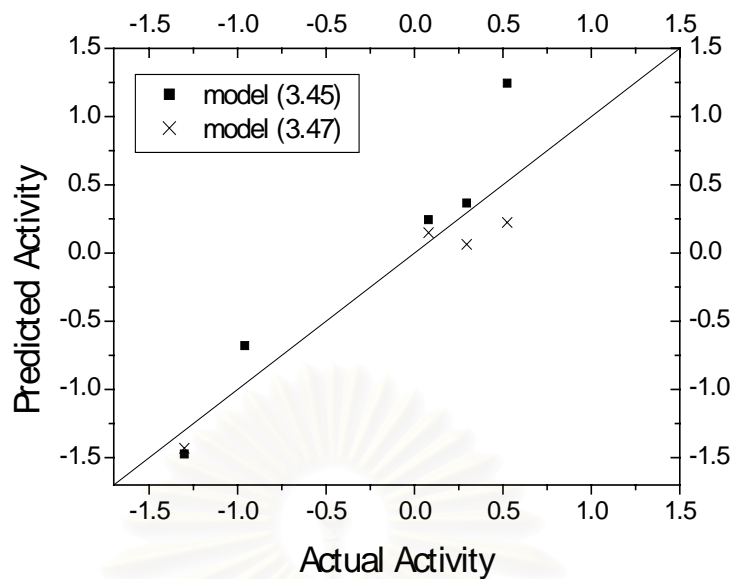


Figure 3.9 Comparison between actual and predicted log(D-6) activities for 5 (model 3.45) and 4 (model 3.47) compounds in the testing set.

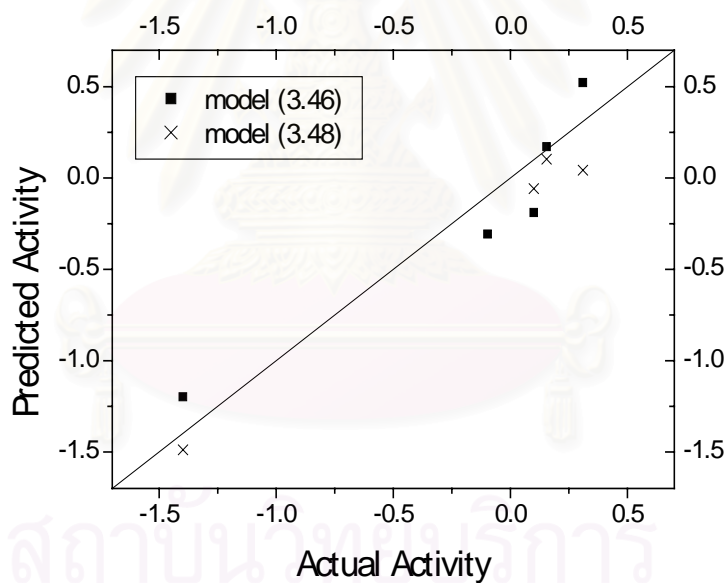


Figure 3.10 Comparison between actual and predicted log(W-2) activities for 5 (model 3.46) and 4 (model 3.48) compounds in the testing set.

3.3.4 Compounds without a substituent at C₁₀ position (Group 3)

Compounds in group 3 have no substituent group at the C₁₀ position. Totally 17 compounds are belonging to this group. Compounds number 89 and 102 were randomly selected for the testing set. The QSAR models were derived from 15 compounds in the training set. Two parameters, C₃ and O₁ atomic charges, were statistically selected for the models of both activities. The r² and q² values of 0.608 and 0.420 for log(D-6) activity and 0.593 and 0.218 for log(W-2) activity were obtained. Since these statistical values are not impressive, the residual value of each compound was examined. The highest residual value corresponds to compound 103. This is again possibly due to its unusually high activity. The compounds 103 and 104 have very similar structures. They are different only in the substituent at the C₃ position, i.e., -(CH₂)₂COOC₂H₅ for the compound 103 and -(CH₂)₂COOH for the compound 104, but the activities of these two compounds are very different (see Table 2.8). The change from COO-H to COO-C₂H₅ leads to the increment in activities of more than 4,688 times which is certainly not possible to be described by the calculated properties. Therefore, compound 103 was omitted and much better predictive QSAR models were obtained as shown in equations (3.49) and (3.50). The presence of C₃ and O₁ atomic charges in both models indicates that compounds with more negative O₁ charge and less C₃ charge would have higher activities. The O₁ charge has more contribution than the C₃ charge due to its statistically higher importance.

$$\log D-6 = 234.19 \cdot C_3 - 688.34 \cdot O_1 - 360.115 \quad \dots(3.49)$$

$$r^2 = 0.876, \quad q^2 = 0.846, \quad s = 0.510, \quad n = 14, \quad F = 38.921$$

$$\log W-2 = 273.23 \cdot C_3 - 705.18 \cdot O_1 - 388.642 \quad \dots(3.50)$$

$$r^2 = 0.917, \quad q^2 = 0.885, \quad s = 0.405, \quad n = 14, \quad F = 60.779$$

The r² and q² values of these two models are very impressive. The relationships between actual and predicted values for both log(D-6) and log(W-2) activities of 14 compounds in the training set are shown in Figure 3.11 and 3.12, respectively. These relationships together with the activity predictions of 2 compounds in the testing set (Table 3.2) indicate the high predictive ability of the models.

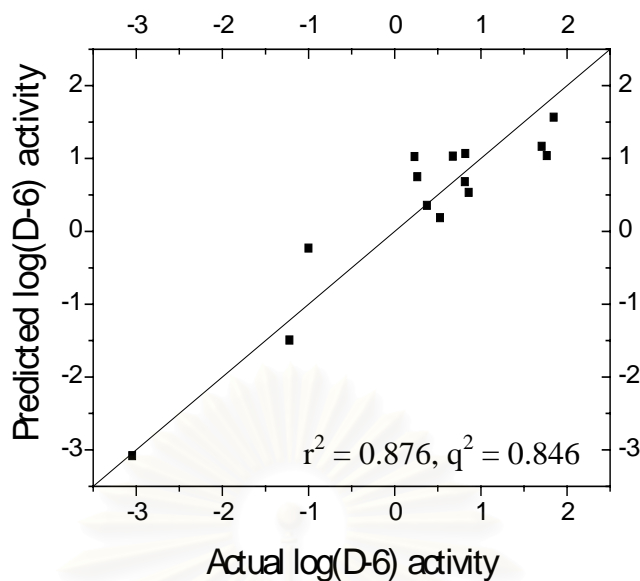


Figure 3.11 Relationship between actual and predicted log(D-6) activities for 14 compounds in group 3.

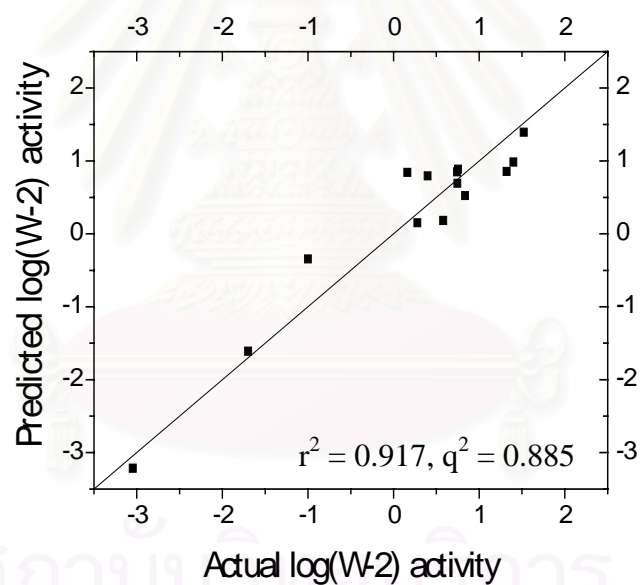


Figure 3.12 Relationship between actual and predicted log(W-2) activities for 14 compounds in group 3.

Table 3.2 Activity predictions of 2 compounds in the testing set of group 3.

Compound No.	log(D-6) activity		log(W-2) activity	
	Actual	Predicted by (3.49)	Actual	Predicted by (3.50)
89	0.375	0.343	0.279	0.105
102	-0.886	-1.048	-0.553	-1.068

3.4 Summary

From the statistical analysis of antimalarial activities against 2 strains of malarial parasite and chemical properties of artemisinin compounds, impressive relationships could not be derived for the group of all compounds. However, if we divided the compounds into 3 groups, impressive relationships with high predictive power were obtained. In case of the group of all compounds, both r^2 and q^2 values are below the acceptable value. For compounds with C=O at the C₁₀ position (group 1), models with moderate quality were obtained. Statistically better models could be attained if only compounds with the O₁₁ atom were used. This situation also holds true for the compounds with -OR or -NR at the C₁₀ position (group 2). Exclusion of compounds with the -NR group at the C₁₀ position could improve the quality of the models, i.e., r^2 from 0.710 to 0.900 and q^2 from 0.606 to 0.801 for the log(D-6) and r^2 from 0.711 to 0.823 and q^2 from 0.602 to 0.611 for the log(W-2). In case of compounds without substituent at the C₁₀ position (group 3), excellent predictive models were found with r^2 value of 0.876 and 0.917, and q^2 value of 0.846 and 0.885.

The $F_{h_{2o}}$ parameter was presented in almost all models. This implies the important effect of water molecules in reaction mechanism for antimalarial activities. However, this parameter appeared to have no significant role for the compounds in group 3 since it was not included in the models. Comparing the structures of the compounds in group 3 with the other two groups, water molecules possibly interact with the substituent group (C=O, -OR, and -NR) at the C₁₀ position. For the moment of inertia parameter, it existed only in the models for the compounds in group 2. With the consideration on their structural differences, this parameter appears to represent the rotation of the substituent groups (-OR and -NR) at the C₁₀ position, which perhaps have some significant effects on the binding to their receptor. Similarly, the A(10-11-12) parameter was presented only in the models for the compounds in group 2. The reason is that the compounds in group 1 and 3 have no variation in the substituent group at the C₁₀ position. Therefore, this angle was kept nearly unchanged and does not related to the activities. From all the results, the antimalarial activities of artemisinin compounds could be mainly described by their structures.

Finally, the real predictive ability of each model was judged from the 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. Furthermore, our models are statistically better than those of the previous works (section 3.1.4).



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 4

Comparative Molecular Field Analysis

4.1 Introduction

Although the traditional QSAR method is fast and simple and has been used successfully in the drug discovery and development process, it is sometimes not possible to account for the biological activity of some groups of compounds. This may be because in reality molecules have three dimensional (3D) structures but the traditional QSAR does not include the 3D structural information. Therefore, it usually can only be applied to a congeneric series. Moreover, 2D parameters may not be sufficient for describing the drug-receptor interactions. This is especially true with various steric descriptors. As a result, many 3D-QSAR techniques have been proposed as the solutions for this problem. Among these techniques, the Comparative Molecular Field Analysis (CoMFA) [90] is the most widely used method, which is available in the SYBYL software [91]. It is based on the assumption that drug-receptor interactions, which are responsible for biological activities, are usually non-covalent. In most of the molecular mechanics force fields, non-covalent interactions are usually treated as steric and electrostatic forces only. Therefore, the steric and electrostatic fields surrounding a set of drug molecules should have correlations with their biological activities. The CoMFA technique employs both interactive graphics and statistical techniques for correlating shapes and properties of molecules with their biological activity.

4.1.1 Methods of CoMFA Studies

Since CoMFA includes 3D structural information of molecules, it is much more complex than the traditional QSAR studies. Generally, there are 4 main steps for CoMFA study, i.e., setup, calculation, statistical analysis, and interpretation.

4.1.1.1 Setup Step

As in any QSAR study (Chapter 3), the first process in performing CoMFA is to collect structural and biological data of a set of interested compounds. Then, the 3D structures of all compounds are constructed. Usually, they are created by the molecular

modeling software and are geometrically optimized by a suitable quantum chemical method. Finally, each compound is superimposed and aligned together (Figure 4.1).

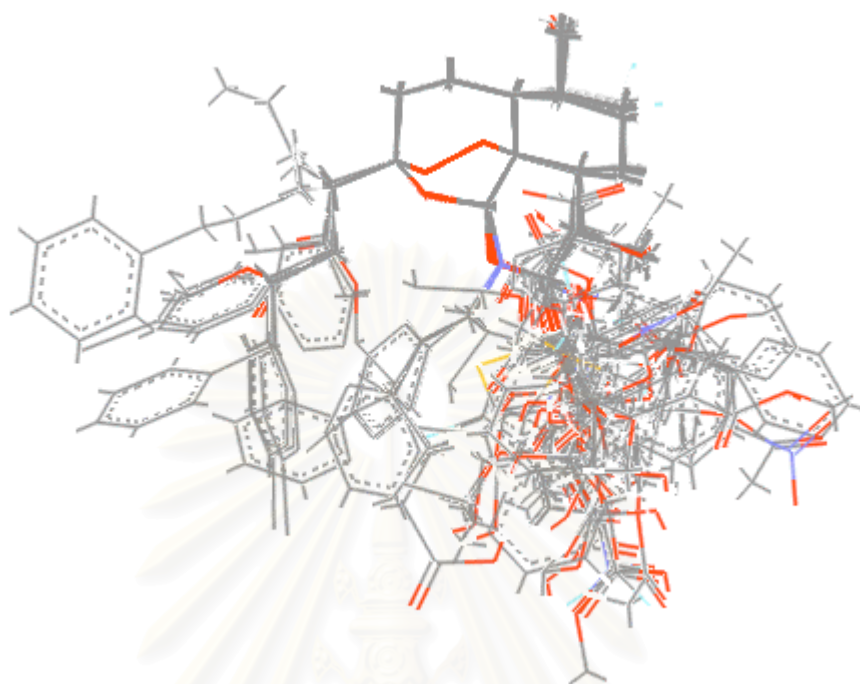


Figure 4.1 The superimposition of all artemisinin compounds (hydrogen atoms are not shown).

The superimposition of the molecules is a crucial step in CoMFA as the results of CoMFA analyses depend highly on the alignment of the molecules. Such an alignment can be done by several approaches. But most often, it is accomplished by an atom overlapping method which fits selected corresponding atom pairs of compounds in such a way that the root mean square (RMS) is minimized. The RMS value can be calculated by the following equation.

$$\text{RMS} = (\sum d^2/n)^{1/2} \quad \text{.....(4.1)}$$

where d = distance between selected atoms
 n = number of atom pairs

4.1.1.2 Calculation Step

In this step, a regular three-dimensional lattice with a selected grid spacing is created around the aligned molecules (Figure 4.2). The typical choice for grid spacing is 2 Å and the size of the grid box is about 3-4 Å larger than the union surface of the molecules. The grid spacing of 1 Å is seldom used because it requires much more

computational time and disk storage space. Moreover, it rarely improves the results. A selected probe is then placed at each lattice point. The probe may be a small molecule such as water, or chemical fragment such as a methyl group but normally atom such as sp^3 carbon is used. Note that many kinds of probe are usually employed in one CoMFA study since different probes may give different results. The steric and electrostatic interaction energies between the probe and each molecule are then calculated. For the steric fields, the Lennard-Jones 6-12 function is often used to model the van der Waals interactions whereas the Coulombic potential using atomic point charges of the molecule and a charged probe with no radius is applied for the calculation of the electrostatic interactions. Finally, all calculated energies are put in the data table.

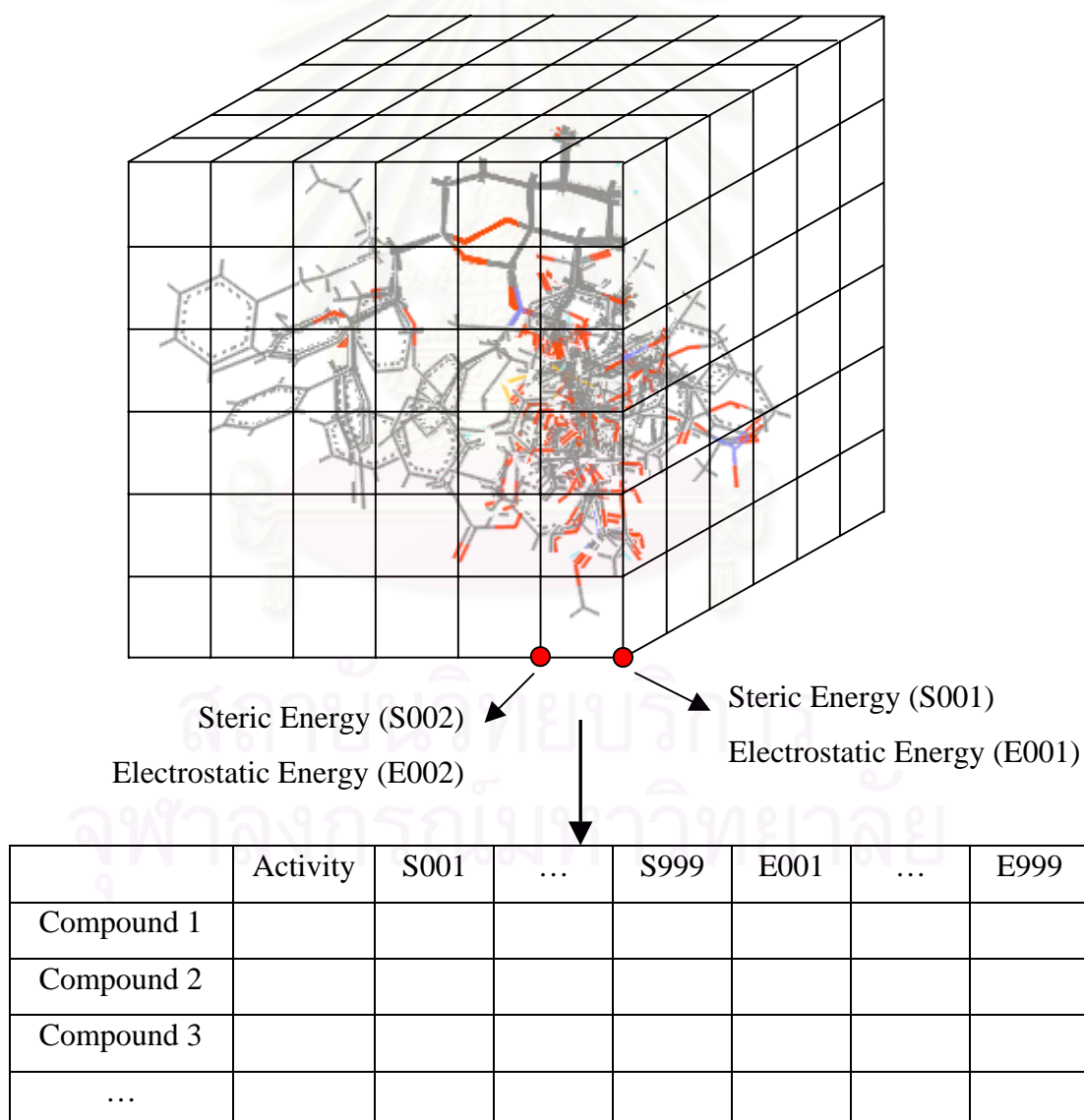


Figure 4.2 Steps in the CoMFA calculations.

4.1.1.3 Statistical Analysis Step

After the data table is constructed, a statistical method is employed to extract important features related to the biological activity. However, before performing the data analysis, the raw data are always pretreated for the reduction of the matrix size, which is usually based on the standard deviation cut-off and energy cut-off. In the standard deviation cut-off process, the energy columns with a low standard deviation are eliminated. This is because these columns do not significantly influence the CoMFA model while they require much longer computing time. The standard deviation cut-off value is called minimum sigma in the SYBYL and the default value is 2.0. The energy cut-off process is used to eliminate the skewness in energy values, i.e., large repulsive energies but small attractive forces. If the probe and the molecule are in close contact, the repulsive energies in a lattice point may become very large. Thus, the energy values usually vary from -10 kcal to minus infinite. This imbalance in energy values is usually handled by replacing all repulsive values larger than a specified cut-off value by this value. In order to include hydrogen bonding information, the cut-off value should be high enough to enable the probe atom to move closer to the molecule within the range of hydrogen bond. In the SYBYL, the default cut-off value for both steric and electrostatic energy is 30 kcal/mol.

In order to apply the classical Multiple Linear Regression (MLR) analysis, the following two requirements should be met. Firstly, every variable is independent and contains no error. And secondly, the number of compounds should be much higher than the number of descriptors. But a typical CoMFA data table usually contains hundreds or thousands of columns of interaction energy values while the number of compounds included in the study is relatively much smaller. Thus, the MLR is not applicable to CoMFA data. The Partial Least Squares (PLS) regression was developed to tackle such problems. It can accommodate hundreds or thousands of interrelated variables.

The PLS is an iterative procedure that produces solutions and summarizes the hundreds or thousands of descriptors to few orthogonal new variables called latent variables or scores. The PLS generates iteratively one component at a time by collectively maximizing the degree of commonality between all of the descriptors and the biological data. In contrast, MLR maximizes the commonality of an individual structural parameter column with the experimental data. The process stops when the

requested number of components is extracted. The number of significant PLS components (latent variables) is determined by the cross-validation test. The statistical values in CoMFA, r^2 , s , q^2 , and s_{cv} , are defined similarly as in MLR analysis (section 3.1.3). More details about the PLS can be found elsewhere [92].

4.1.1.4 Interpretation Step

Like all QSAR techniques, CoMFA results are an equation showing the contribution of energy fields at each lattice point. However, the thousands of terms would make a trouble for expressing them in a form of numerical equation. Since each lattice point has a coefficient (the coefficient implicitly being 0 for any lattice point that was omitted), it is much better to display the equation as 3D contour map showing the regions in space where specific molecular properties increase or decrease the potency. Typically there are two contour levels for each type of CoMFA energy field: the positive and the negative contours. The contours are colored in green and yellow for positive and negative steric effects, and in blue and red for positive and negative electrostatic effects. Positive steric contours illustrate the regions where substituents could increase the biological activity, whereas the negative steric contours indicate the area where substituents would decrease the activity. The positive electrostatic contours show the regions where positive charges could increase the activity, and the negative electrostatic contours display the regions where negative charges increase the activity.

4.1.2 Previous CoMFA Studies

There are at least three articles on artemisinin compounds with the CoMFA method [46, 93-94]. The first article [46] was published in 1993, in this work, 25 C-9 analogs of artemisinin compounds (group 1) and 17 inactive C-9 analogs of 1-deoxyartemisinin (group 2) were chosen. The activities were measured against the D-6 clone of *P. falciparum*. And the relative activity value, the IC_{50} of artemisinin divided by the IC_{50} of the analog, was used. The Gasteiger charges were taken for the electrostatic fields. The O_1 , O_2 , and O_{11} atoms were selected for the alignment. The CoMFA was first performed on 25 compounds of group 1. Omitting 4 compounds gave the model with r^2 of 0.985 and q^2 of 0.793 with 5 components. Although the statistical values of this model are impressive, there is only a weak correlation between the peroxy moiety and the steric and electrostatic fields. Therefore, compounds of group 2 were

then additionally included in the calculations so totally 42 compounds were resulted. The CoMFA calculations were performed again, however, without an omission of any compound. The model with r^2 of 0.975 and q^2 of 0.857 using 5 components was obtained. This time, the new model has a good correlation between electrostatic contours and the peroxide bridge. The steric contour maps revealed that the presence of steric bulk at C_3 and moderate steric group at C_9 position would enhance antimalarial potency. The electrostatic contour map showed the negative region surrounding the peroxide bridge.

In the second article [93], the CoMFA was performed on 22 ethers and ester derivatives of dihydroartemisinin. All structures were optimized by semiempirical AM1 and PM3 methods. And the atomic charges were calculated by AM1. The default values of SYBYL were used for setting steps. Four different alignments were selected. The q^2 values for 22 compounds are in the range of 0.186 to 0.343. By omitting 3 and 4 compounds, the q^2 were increased to around 0.713-0.774 using 3 to 5 components. The authors concluded from the best model (highest q^2 value) among four alignments that the chain of $-C_{12a}-O_1-O_2-C_3-O_{13}-C_{12}-O_{11}-C_{10}-O_{14}-$ and atom C_{18} were definitely important sub-structures for antimalarial activity of artemisinin analogs.

In the last article [94], the CoMFA was carried out on a large number of artemisinin derivatives. Totally, 171 compounds with the activity against the W-2 clone of *P. falciparum* were used in the study. All compounds were geometrically optimized with the Tripos force field available in the SYBYL software and their atomic charges were given after the Gasteiger-Hückel protocol. An sp^3 carbon +1 probe and 1 Å lattice spacing were employed. By gradually omitting compounds with high residual values until 154 compounds were remained (17 compounds were omitted), the satisfactory CoMFA model with q^2 of 0.840 was achieved. From the residual values, the predicted activities of the compounds in the data set are within 1.45 log unit of the observed value. And for 24 compounds in the testing set, the residual values ranged from 0 to ± 1.4 log units. In that work, the hydrophilicity field was introduced into CoMFA but even lower r^2 and q^2 values were resulted. This indicates a poor correlation between hydrophilicity/lipophilicity and activity. From the steric contour maps, adding steric bulk around C_7 , C_8 , and C_9 positions would increase the activity and steric hindrance at peroxide moiety would decrease the potency. As in the first article, the electrostatic

contour maps indicated the importance of negative charge around the peroxide moiety to the activity.

4.2 Computational Methods

4.2.1 Structure and Charge Calculations

From section 2.4, the HF/3-21G was shown to be the lowest level of theory that yields the geometrical parameters in good agreement to the X-ray data. However, this does not guarantee that the HF/3-21G optimized structure would also give good CoMFA results. Therefore, the comparison of CoMFA results among different optimized structures, i.e., AM1, HF/3-21G, and Tripos force field methods, were made. All optimized structures and their Mulliken population analysis (MPA) atomic charges obtained from quantum chemical methods were reimported into the SYBYL. In addition, the “Systematic Search” procedure in the SYBYL was used for the conformational search of all compounds. Subsequently, the compounds were geometry optimized with the Tripos force field. Moreover, since the electrostatic field was built from the atomic charges, the effect of different atomic charges should also be investigated. For this reason, the charges according to Gasteiger and Marsili, the electrostatic potential fit (ESPFIT) charges [95] at AM1 level, the MPA charges at PM3 level, and the natural population analysis (NPA) charges [96] at HF/3-21G level were additionally calculated.

4.2.2 Alignment Rules

Five 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 SYBYL which minimizes the root-mean-square (RMS) differences of selected atoms to the ones of the reference molecule. The selected atoms for the definition of alignment rules are shown in Table 4.1 and Figure 4.3. Note that alignment number 3 is the same as that of the reference [93].

Table 4.1 Atoms selected for the definition of alignment rules.

Alignment No.	Selected Atoms
1	O1-O2-C3-C4-C5-C5a-C12a-C12-O13
2	C5a-C6-C7-C8-C8a-C12a

Table 4.1 (Continued)

Alignment No.	Selected Atoms
3	O1-O2-O11
4	C12a-O1-O2-C3
5	C12a-O1-O2-C3-O13-C12-O11

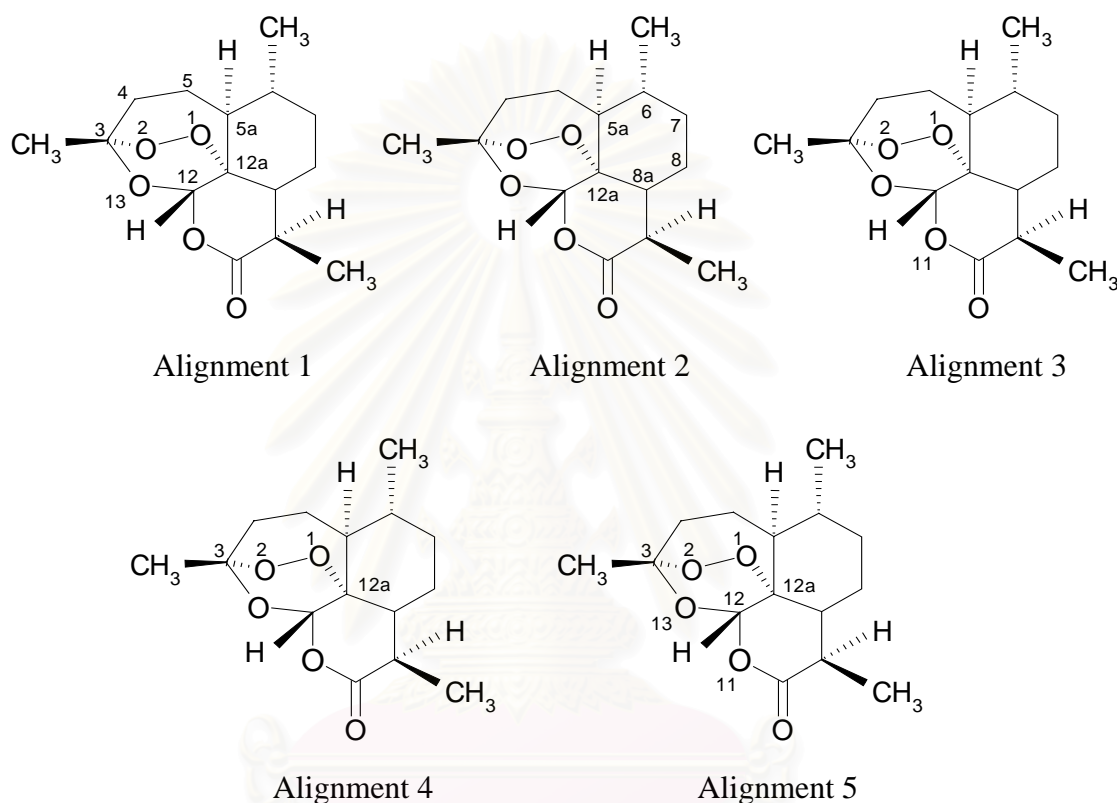


Figure 4.3 Definition of five alignment rules used in CoMFA studies.

4.2.3 CoMFA Calculations

A regular three-dimensional lattice with 2 Å spacing was created extending beyond the molecular dimensions of largest molecule by 4.0 Å in all directions. Two probe atoms, namely an sp^3 carbon with a charge of +1.0 and an sp^3 oxygen with a charge of -1.0, were employed. 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. An AutoCoMFA column was also created by using the automatically created region and a +1.0 C- sp^3 probe atom.

4.2.4 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 the 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.

4.3 Results and Discussions

In order to reduce computing time for the investigations of parameters affecting CoMFA results, a small set of compounds should be used instead of all 104 compounds. Therefore, 50 artemisinin compounds with activities against the D-6 clone of *P. falciparum* were selected. The full set of compounds was applied when all suitable parameters were justified.

4.3.1 Effect of Structure and Charge

The CoMFA calculations were performed on various optimized structures and atomic charge schemes to establish the suitable methods. Since the endoperoxide (O_1 and O_2) is crucial for the antimalarial activities and artemisinin compounds have variation of the substituent groups at the C_3 position, the alignment number 4, which has the smallest number of selected atoms that can cover the three atoms, was selected for this purpose. Using the AutoCoMFA column, the results are given in Table 4.2. From these results, it is clear that for all charge schemes the HF/3-21G optimized structures give the best predictive models and the atomic charges have a little effect on the q^2 value. Therefore, the HF/3-21G structures and HF/3-21G MPA charges, which are calculated simultaneously in the optimization, were selected for further study. Interestingly, the effect of atomic charges on the q^2 value is more pronounced in the other optimized structures. For example, using AM1 structure the q^2 was decreased from 0.319 with the AM1 MPA charge to 0.062 with Gasteiger-Marsili charge. All the above results are in good agreement with our preliminary investigations of 40 artemisinin derivatives using the AM1 and HF/3-21G optimized structures [97].

Notably, for the CoMFA using structures optimized by the Tripos force field, results of models with conformational search were comparable to those without conformational search.

Table 4.2 CoMFA results of different optimized structures and atomic charges.

Optimization Method	Atomic Charge	q^2	onc	PRESS
HF/3-21G	HF/3-21G MPA	0.480	5	0.548
	HF/3-21G NPA	0.478	5	0.549
	AM1 MPA	0.469	5	0.554
	AM1 ESPFIT	0.494	5	0.540
	PM3 MPA	0.473	5	0.552
	Gasteiger-Marsili	0.424	3	0.564
AM1	AM1 MPA	0.319	2	0.607
	Gasteiger-Marsili	0.062	2	0.713
PM3	PM3 MPA	0.363	3	0.593
Tripos force field	HF/3-21G MPA	0.404	5	0.586
	AM1 MPA	0.417	5	0.580
	Gasteiger-Marsili	0.235	2	0.644
Tripos force field with Conformational Search	Gasteiger-Marsili	0.264	5	0.652

4.3.2 Effect of Alignment Rule

In order to study an influence of different alignment rules, the AutoCoMFA column derived from the HF/3-21G optimized structures and MPA atomic charges were used. From the results in Table 4.3, the alignment number 4 and 5 give better statistics than the other three alignments. These two alignment rules contain the endoperoxide atoms which are essential for the antimalarial activity. However, the alignment number 4 has smaller number of selected atoms (i.e., more ease for superimposition step) and also slightly higher q^2 value. Therefore, it was selected for further investigations. Note that the alignment number 3, which is the same as that in reference [93], give the lowest q^2 value and the highest PRESS value.

Table 4.3 CoMFA results of different alignment rules.

Alignment No.	q^2	onc	PRESS
1	0.459	5	0.559
2	0.451	2	0.545
3	0.429	3	0.584
4	0.480	5	0.548
5	0.477	5	0.549

4.3.3 Effect of Type and Charge of Probe Atom

Both steric and electrostatic fields are determined from the interaction energies between molecules and the selected probe atom. Therefore, the type and charge of probe atom are significant to CoMFA results. Two types of probe atoms were selected, +1.0 sp^3 carbon and -1.0 sp^3 oxygen. Comparing between these two probe atom types, the oxygen sp^3 probe atom gave slightly better results (Table 4.4). Hence, both types were used. The influence of charge was also investigated by varying the charge of sp^3 oxygen probe atom from -0.1 to -2.0. The model with more negative probe charge gives better results than that with the lower value. However, the effect of charge on q^2 values is quite pronounced at low negative charges but less pronounced at high negative charges. Therefore, the probe charge of -1.0 or 1.0, which give the same statistics, seems to be appropriate for our study.

Table 4.4 CoMFA results with different type and charge of probe atom.

Type	Charge	q^2	onc	PRESS
C- sp^3	+1.0	0.480	5	0.548
O- sp^3	-1.0	0.494	5	0.540
O- sp^3	-0.1	0.350	3	0.599
O- sp^3	-0.2	0.287	2	0.656
O- sp^3	-0.3	0.317	5	0.628
O- sp^3	-0.4	0.454	5	0.561
O- sp^3	-0.5	0.465	5	0.555
O- sp^3	-0.6	0.452	5	0.562
O- sp^3	-0.7	0.461	5	0.557

Table 4.4 (Continued)

Type	Charge	q ²	onc	PRESS
O-sp ³	-0.8	0.476	5	0.549
O-sp ³	-0.9	0.484	5	0.545
O-sp ³	-1.1	0.499	5	0.537
O-sp ³	-1.5	0.508	5	0.532
O-sp ³	-2.0	0.515	5	0.528

4.3.4 Effect of Steric and Electrostatic Cut-off

As stated above in the section 4.1.1.3, 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. The results are shown in Table 4.5. By varying the steric cut-off values while keeping the electrostatic cut-off value at 30 kcal/mol, the lowest and highest q² values are 0.321 (model A1) and 0.560 (model A5), which were obtained from the steric cut-off values of 1 and 7 kcal/mol, respectively. The high difference indicates that the CoMFA results are sensitive to the steric cut-off. Comparing between different electrostatic cut-off values with the same steric cut-off values, e.g., model A1 and A10, model A3 and A11, the results revealed that the electrostatic cut-off has less effect on the q² value than the steric cut-off. With the default cut-off value (30 kcal/mol), the q² value of 0.494 (model A9) is not significantly different from the highest q² value (0.560). Therefore, the default cut-off values, which include hydrogen-bonding information, were selected for further studies.

Table 4.5 CoMFA results with different steric and electrostatic cut-offs.

Model	Steric Cut-off (kcal/mol)	Electrostatic Cut-off (kcal/mol)	q ²	onc	PRESS
A1	1	30	0.321	1	0.603
A2	4	30	0.381	4	0.598
A3	5	30	0.537	3	0.520
A4	6	30	0.554	3	0.512
A5	7	30	0.560	3	0.509
A6	8	30	0.548	3	0.515

Table 4.5 (Continued)

Model	Steric Cut-off (kcal/mol)	Electrostatic Cut-off (kcal/mol)	q^2	onc	PRESS
A7	9	30	0.532	3	0.523
A8	10	30	0.501	3	0.537
A9	30	30	0.494	3	0.540
A10	1	100	0.334	5	0.626
A11	5	100	0.536	3	0.521
A12	30	100	0.479	2	0.541
A13	50	50	0.500	5	0.551

4.3.5 Effect of Grid Position

In order to explore the effect of grid position, the changes in the X, Y, and Z directions were applied to the automatically generated grid from the AutoCoMFA column. The grid box was shifted by +0.5, -0.5, +1.0, and -1.0 Å in all 3 directions. The resulting q^2 values are in a range of 0.344 to 0.504 (Table 4.6) compared to 0.494 of the original position. Note that the location of grid box in Z direction significantly affects the statistics and the number of components. Therefore, the original grid position seems to be suitable for our studies.

Table 4.6 CoMFA results with different grid positions.

Model	Direction	Shift	q^2	onc	PRESS
Original	-	-	0.494	3	0.540
B1	X	-0.5	0.440	3	0.588
B2	X	+0.5	0.461	3	0.546
B3	X	-1.0	0.504	3	0.524
B4	X	+1.0	0.504	3	0.524
B5	Y	-0.5	0.444	2	0.549
B6	Y	+0.5	0.443	5	0.567
B7	Y	-1.0	0.441	3	0.556
B8	Y	+1.0	0.441	3	0.556
B9	Z	-0.5	0.402	5	0.588

Table 4.6 (Continued)

Model	Direction	Shift	q^2	onc	PRESS
B10	Z	+0.5	0.367	3	0.591
B11	Z	-1.0	0.344	3	0.602
B12	Z	+1.0	0.344	3	0.602

4.3.6 q^2 -GRS approach

One problem usually encountered in the CoMFA works is the reproducibility. Since the q^2 value is sensitive to the overall orientation of superimposed molecules, a problem in reproducing the q^2 values reported in the literatures may occur. Therefore, the cross-validated r^2 guided region selection approach (q^2 -GRS) was introduced [98]. The q^2 -GRS approach, which is a simple SYBYL script, was proposed to give the orientation-independent q^2 value, thus, solving a reproducibility problem. The routine in the q^2 -GRS is as follows. The rectangular lattice obtained initially with conventional CoMFA is divided into 125 ‘sublattices’. Subsequently, the CoMFA analysis is independently performed for each sublattice in turn using C-sp³ probe atom and both fields with the grid spacing of 1.0 Å. Only those sublattices which have the resulting q^2 above certain cutoff, by default $q^2 \geq 0.1$, are selected for further analysis. Finally, the selected sublattices are combined together to give a single region and the CoMFA calculation is repeated using this united region. Note that the q^2 -GRS approach does not investigate the effect of grid position as done in the section 4.3.5.

Since this approach was shown to give better q^2 value than the conventional method [94, 98], we applied this approach to our compounds (denoted as “all”) for both log(D-6) and log(W-2) activities. The results from the q^2 -GRS method were compared to the conventional ones as shown in Table 4.7 and 4.8. As in the QSAR study (Chapter 3), all compounds were divided into 3 groups with more structural similarity, i.e., 31 compounds with O₁₁ and C₁₀=O₁₄ atoms (set 1), 41 compounds with -OR at C₁₀ position (set 2), and 17 compounds without substituent at C₁₀ position (set 3).

Table 4.7 CoMFA results with log(D-6) activity using q^2 -GRS approach.

Compound	q^2 -GRS method			Conventional method		
	q^2 -GRS	onc	PRESS	q^2	onc	PRESS
all	0.169	7	0.822	0.124	3	0.823
set 1	0.021	5	0.919	-0.040	6	0.948
set 2	0.250	2	0.693	0.363	2	0.638
set 3	0.221	2	1.172	0.162	1	1.175

Table 4.8 CoMFA results with log(W-2) activity using q^2 -GRS approach.

Compound	q^2 -GRS method			Conventional method		
	q^2 -GRS	onc	PRESS	q^2	onc	PRESS
all	0.134	1	0.848	0.107	1	0.860
set 1	0.195	2	0.843	0.004	9	1.079
set 2	0.203	3	0.819	0.397	2	0.703
set 3	0.166	2	1.159	0.052	1	1.194

The q^2 -GRS values of models for both activities using all 104 compounds are only slightly higher than those of the conventional values. For the models of set 1 and 3, the improvements in q^2 values are moderate in log(W-2) activity while they are little in log(D-6) activity. But for the models of set 2 with both activities, the q^2 -GRS values are around 2 times lower than those of the conventional method. Therefore, the q^2 -GRS approach is not necessary to give better CoMFA results than the conventional method. And the effect of grid position should still be investigated. Moreover, the computational time for the q^2 -GRS is much longer than the conventional one. Hence, we decided not to apply this approach in further studies.

4.3.7 All 104 Compounds

The compounds were divided into 2 sets, i.e., training set and testing set, to access the real predictive ability of the model. Therefore, compounds 10, 20, 30, 34, 50, 60, 70, 80, 90, and 100 were intentionally selected for the testing set in order to cover all groups of different structures. The remaining 94 compounds were used as the training set. For log(D-6) activity, the CoMFA results have q^2 values of between 0.035

to 0.120 (Table 4.9). The C-sp³ probe atom gave higher q² values than the O-sp³. The models with only steric field have slightly higher q² values than those with both fields and only electrostatic field. Using both fields, the steric field has more contribution to the model than the electrostatic field, as shown by the “% Steric” of 62.0 % and 63.1 % for C-sp³ and O-sp³ probe atoms, respectively.

Table 4.9 CoMFA results of all 94 compounds with log(D-6) activity.

Probe	Fields	q ²	onc	PRESS	r ²	S.E.	F	% Steric
C-sp ³	Both	0.113	3	0.842	0.493	0.637	29.157	62.0
	Steric	0.120	3	0.839	0.401	0.692	20.065	100.0
	Elec.	0.035	3	0.878	0.517	0.621	32.154	0.0
O-sp ³	Both	0.095	2	0.846	0.378	0.701	27.643	63.1
	Steric	0.098	3	0.849	0.394	0.696	19.478	100.0
	Elec.	0.035	3	0.878	0.517	0.621	32.154	0.0
^a C-sp ³	Both	0.454	2	0.513	0.673	0.397	83.333	56.9

^a compounds 5, 18, 21, 22, 25, 27, 75, 96, 101, and 104 were omitted.

Since the q² value is sensitive to a compound with high residual value, the improvement of the q² value could be attained by omitting of such compounds, as usually done in almost all CoMFA studies [46, 93-94]. Therefore, compounds with residual values larger than 1.5 (compounds 5, 18, 21, 22, 25, 27, 75, 96, 101, and 104) were omitted from the model using C-sp³ probe atom and both fields. This improved the predictive power (q²) to 0.454 with the onc of 2. Other statistics are also improved.

The steric contour plot of the best predictive model for the log(D-6) activity is illustrated in Figure 4.4. A yellow contour was located in front of and below the β-C₁₀ position. The 6-membered ring substituent groups of compounds 48 and 49 fall in this region and hence they have low activities. In case that the steric substituent is placed in the right hand position to the C₁₀ atom, the activity could be enhanced as indicated by a green region around this area. Green contours around the β- and α-C₉ position indicate the steric favor of substituent group at this position. The very high activities of compounds 94 and 95, which have large substituents at this position, support this concept. Notably, no contour was observed around the C₃ position. This indicates that the steric of substituent at the C₃ position has no effect to the log(D-6) activity.

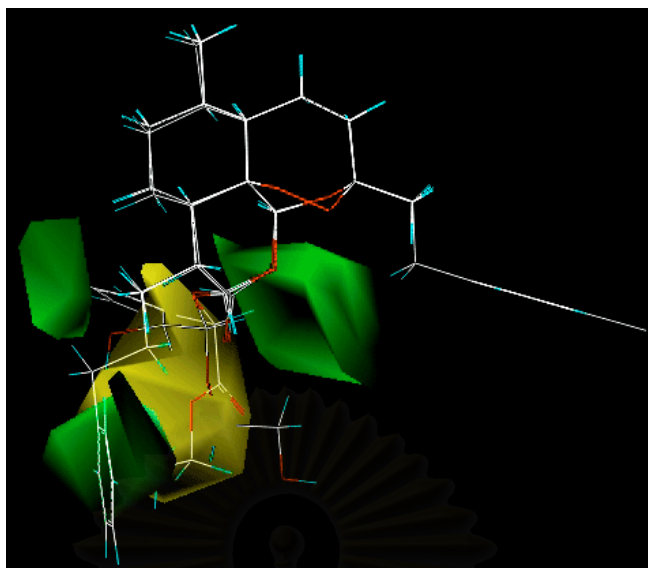


Figure 4.4 Steric contour plot of the log(D-6) model for all compounds.

The electrostatic contour plot of the best predictive model for the log(D-6) activity is illustrated in Figure 4.5. A large blue contour is located below the β -C₁₀ position indicating high antimalarial activity if atoms in this region have positive charge. A red contour is presented further away from the C₁₀ position. Therefore, atoms in this area should have negative charge to increase the activity.

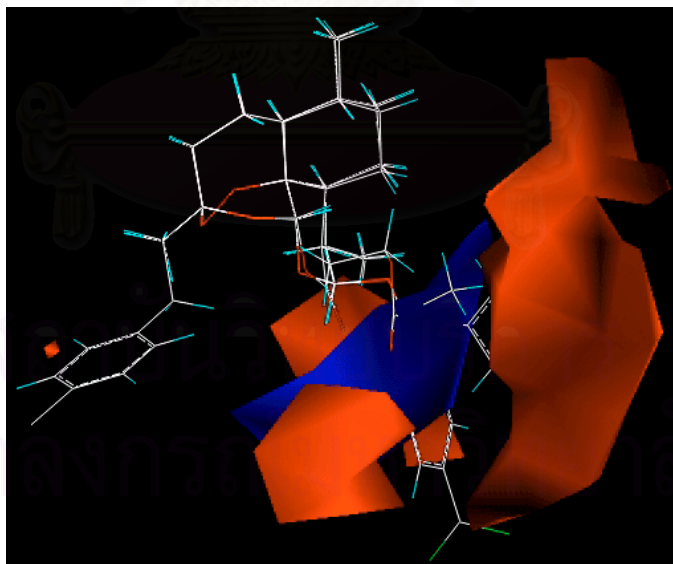


Figure 4.5 Electrostatic contour plot of the log(D-6) model for all compounds.

Similar to the log(D-6) activity, the CoMFA models of all 94 compounds with log(W-2) activity have quite low predictive ability, q^2 values of between 0.051 to 0.116 (Table 4.10). The models with C-sp³ probe atom have slightly higher q^2 values than those with O-sp³. Unlike the log(D-6) activity, the models with both fields have slightly higher q^2 values than those with only one field.

Table 4.10 CoMFA results of all 94 compounds with log(W-2) activity.

Probe	Fields	q^2	onc	PRESS	r^2	S.E.	F	% Steric
C-sp ³	Both	0.116	2	0.871	0.372	0.734	26.922	61.9
	Steric	0.111	3	0.879	0.379	0.734	18.312	100.0
	Elec.	0.051	3	0.908	0.497	0.661	29.690	0.0
O-sp ³	Both	0.112	2	0.873	0.365	0.738	26.157	60.9
	Steric	0.094	3	0.887	0.372	0.738	17.767	100.0
	Elec.	0.051	3	0.908	0.497	0.661	29.690	0.0
^a C-sp ³	Both	0.307	3	0.548	0.606	0.413	41.496	61.5

^a compounds 5, 18, 25, 52, 59, 75, 101, 103, and 104 were omitted.

After omitting compounds with residual values larger than 1.5 (compounds 5, 18, 25, 52, 59, 75, 101, 103, and 104) from the model using C-sp³ probe atom and both fields, the q^2 value was increased to 0.307. Other statistics were also increased but the steric contribution was nearly unchanged.

The steric contour plot of the best predictive model for the log(W-2) activity is illustrated in Figure 4.6. Similar to the log(D-6) model, a yellow contour was located in front of and below the β -C₁₀ position but this contour has a larger size. Therefore, compounds having large steric group in this region would have low activities. Moderate green contours (steric favor regions) were found in the right hand position to the C₁₀ atom and around the β -C₉ position. As in the log(D-6) model, no contour was observed around the C₃ position. Thus, the steric of substituent at the C₃ position also has no effect to the log(W-2) activity.

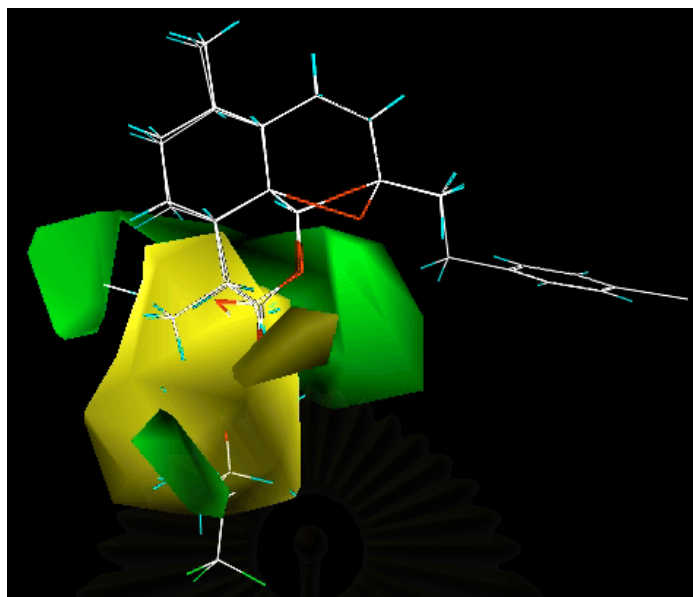


Figure 4.6 Steric contour plot of the log(W-2) model for all compounds.

The electrostatic contour plot of the best predictive model for the log(W-2) activity is illustrated in Figure 4.7. Blue contours are located in a region below the C₁₀ position, a region around substituent at the α -C₉ position, and a region in a right hand side of the C₁₀ position. Red contours are presented in a region around the carbon atom attached to the O₁₄ atom, a region further away from the O₁₄ position, and a region below the C₁₀ position.

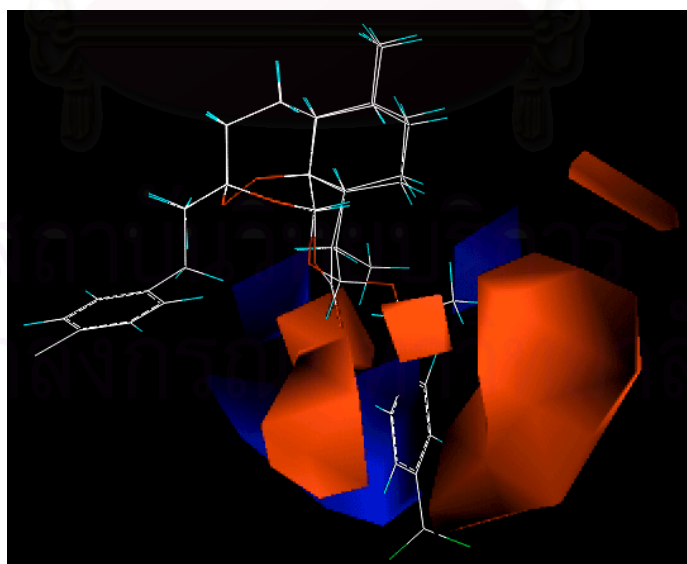


Figure 4.7 Electrostatic contour plot of the log(W-2) model for all compounds.

Both $\log(D-6)$ and $\log(W-2)$ activities of compounds in the testing set were predicted using the models that omitted compounds with a high residual value. And the results are shown in Figure 4.8 and 4.9.

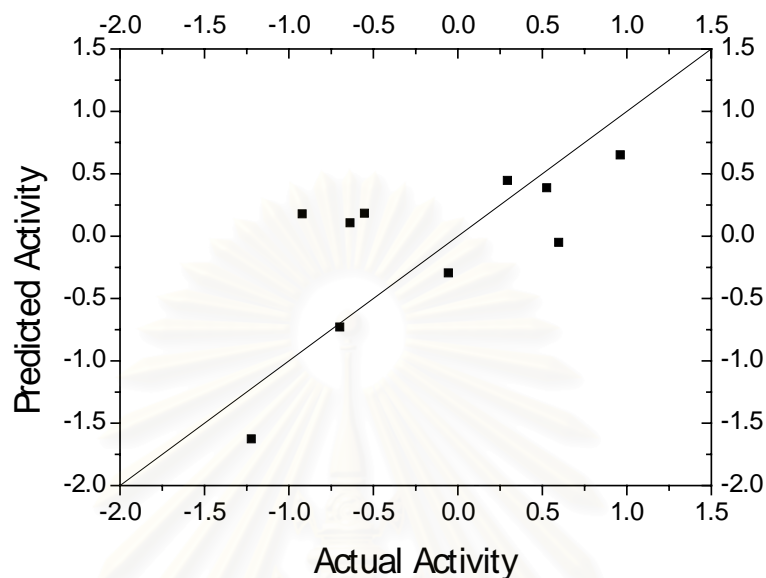


Figure 4.8 Comparison between actual and predicted $\log(D-6)$ activities for 10 compounds in the testing set.

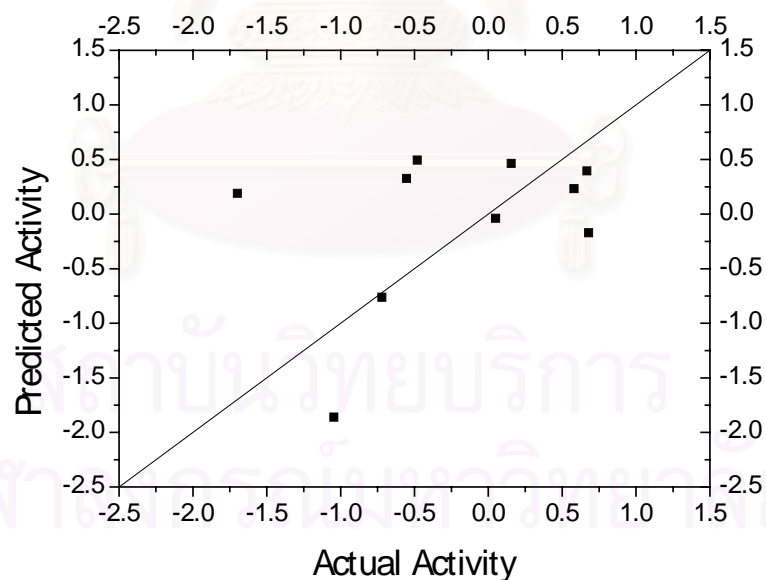


Figure 4.9 Comparison between actual and predicted $\log(W-2)$ activities for 10 compounds in the testing set.

From the statistical values of models for $\log(D-6)$ and $\log(W-2)$ activities and the predicted activities shown in Figure 4.8 and 4.9, it seems that a good CoMFA model can not successfully be derived for the whole set of our compounds containing various

types of structures. Actually, a better model could be obtained by further omitting more compounds until a satisfactory q^2 value was achieved as done in the reference [94]. However, this procedure appears to be not reasonable since the main goal in the QSAR study is not to get a model with the highest q^2 value but is to gain useful understanding of the properties affecting the activity that could be used for the enhancement of its activity.

The unsatisfactory CoMFA models are possible due to the structural differences among the compounds and the inconsistent effect of substituent groups to the activities of compounds with and without C=O at the C₁₀ position (see section 2.3). Therefore, all 104 compounds were divided into 2 groups, with or without C=O at the C₁₀ position. The CoMFA analysis was then performed for each individual group of compounds.

4.3.8 Compounds with C=O at the C₁₀ position

There are 41 compounds with C=O at the C₁₀ position. Compounds 5, 9, 16, and 38 were randomly selected for the testing set. And the remaining 37 compounds were used for the training set. The CoMFA results of compounds group 1 for both activities were illustrated in Table 4.11 and 4.12.

For log(D-6) activity, although the models have good relations (high r^2 values) but they do not have a predictive power as indicated by q^2 values of less than zero. Moreover, the one of these models are also high. Using only the electrostatic field gives substantially better results than both fields and only steric field.

Table 4.11 CoMFA results for log(D-6) activity of compounds with C=O at the C₁₀ position.

Probe	Field	q^2	onc	PRESS	r^2	S.E.	F	% Steric
C-sp ³	Both	-0.177	7	0.866	0.954	0.170	86.717	67.6
	Steric	-0.310	7	0.914	0.911	0.238	42.267	100.0
	Elec.	-0.047	4	0.778	0.846	0.298	43.993	0.0
O-sp ³	Both	-0.122	7	0.846	0.956	0.167	90.698	66.9
	Steric	-0.229	7	0.885	0.928	0.214	53.720	100.0
	Elec.	-0.047	4	0.778	0.846	0.298	43.993	0.0
^a O-sp ³	Elec.	0.382	5	0.556	0.950	0.157	107.342	0.0

^a compounds 21, 22, and 25 were omitted.

Improved model with moderate predictive power ($q^2 = 0.382$) could be obtained if 3 compounds with residual values larger than 1.5 (compounds 21, 22, and 25) were excluded while using O-sp³ probe atom and electrostatic field only. And its electrostatic contour plot is shown in Figure 4.10.

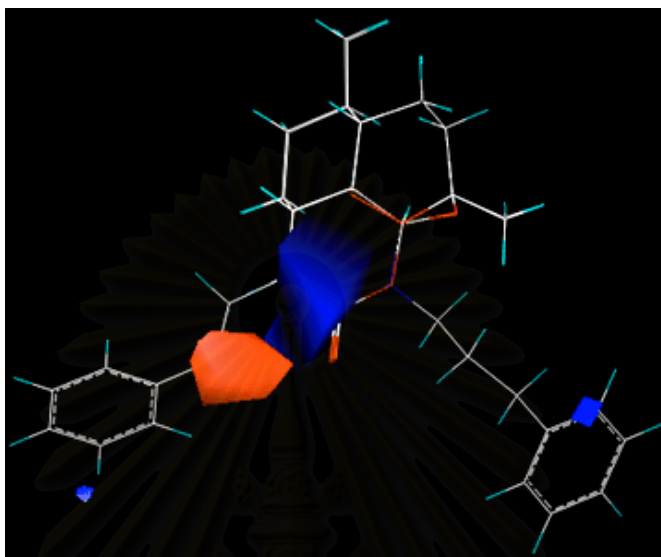


Figure 4.10 Electrostatic contour plot of the log(D-6) model for compounds with C=O at the C₁₀ position.

From the electrostatic contour plot, a large blue contour is located around the C₉ position indicating high antimalarial activity if atoms in this region have positive charge. A red contour is presented around the second carbon atom of substituent at the C₉ atom. Therefore, atoms in this area should have negative charge to increase the activity.

Our compounds were compared with that in the reference [46] where the CoMFA model for log(D-6) activity of only 22 compounds having C=O at the C₁₀ position with r^2 of 0.975 and q^2 of 0.857 values was presented. The lower q^2 statistics (0.382) of our model may be contributed by the fact that all 22 compounds in that article have very similar structures with variation only in the substituent group at the C₉ position while our compounds are different in the substituent group at both C₉ and C₃ positions. Note that the steric contribution in the model using both fields of that article (65.7 %) is comparable to ours.

For log(W-2) activity, both q^2 and r^2 values are not good. Unlike the log(D-6) activity, using only steric field give better model than using only electrostatic field and both fields. The suggested one for the models is only one. Omitting compounds with

high residual values from the model while using C-sp³ probe atom and only steric field could not improve the q² value as shown in Table 4.12. Since the models have no predictive power, information obtained from its contour plot would be unreliable. Thus, the contour was not displayed.

Table 4.12 CoMFA results for log(W-2) activity of compounds with C=O at the C₁₀ position.

Probe	Fields	q ²	onc	PRESS	r ²	S.E.	F	% Steric
C-sp ³	Both	-0.162	1	0.866	0.215	0.712	9.613	68.4
	Steric	-0.122	1	0.851	0.131	0.749	5.283	100.0
	Elec.	-0.208	1	0.883	0.250	0.696	11.694	0.0
O-sp ³	Both	-0.177	1	0.872	0.220	0.710	9.870	67.9
	Steric	-0.126	1	0.853	0.132	0.749	5.334	100.0
	Elec.	-0.208	1	0.883	0.250	0.696	11.694	0.0
^a C-sp ³	Steric	-0.217	1	0.551	0.159	0.458	5.868	100.0

^a compounds 6, 18, 25, and 34 were omitted.

The log(D-6) activity predictions of 4 compounds in the testing set are shown in Figure 4.11. The prediction is fairly good. The log(W-2) activity prediction was not performed because the CoMFA model for the log(W-2) activity has no predictive power.

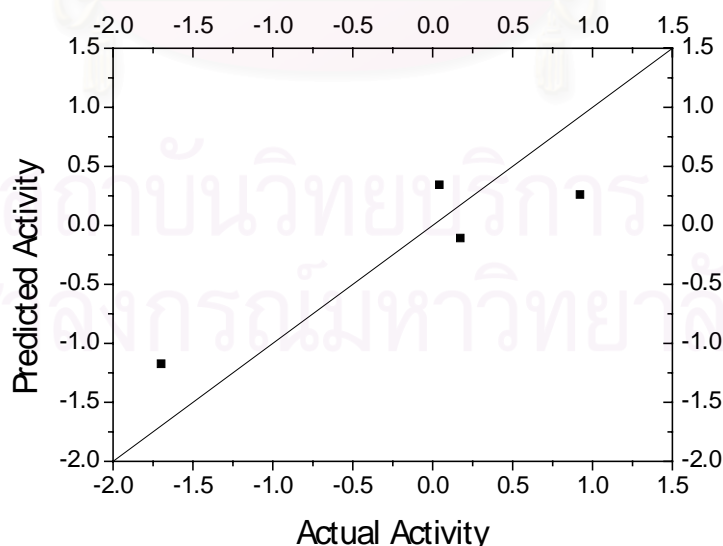


Figure 4.11 Comparison between actual and predicted log(D-6) activities for 4 compounds with C=O at the C₁₀ position in the testing set.

4.3.9 Compounds without C=O at the C₁₀ position

From all 63 compounds in this group, compounds number 48, 64, 71, 86, 91, and 102 are randomly chosen for the testing set. The CoMFA results of 57 compounds in the training set with log(D-6) activity are given in Table 4.13. These results have higher predictive ability than those of the previous group. Interestingly, the q^2 values of models derived from different types of probe atom and field are comparable. Compared to the preceding section, the steric contribution is lower for both C-sp³ and O-sp³ probe atoms. Thus, the combination of compounds with and without C=O at the C₁₀ position may not yield a good CoMFA model. And this is perhaps one of the reasons for the low statistics of models for all compounds (section 4.3.7.1). Exclusion of 2 compounds with residual values larger than 1.5, i.e., compounds 65 and 104, from the model while using C-sp³ probe atom and steric field only could lead to the better predictive model with the q^2 of 0.376. A reason for the high residual values of these two compounds possibly due to their unusual low activities. Therefore, the CoMFA could not explain their activities.

Table 4.13 CoMFA results for log(D-6) activity of compounds without C=O at the C₁₀ position.

Probe	Fields	q^2	onc	PRESS	r^2	S.E.	F	% Steric
C-sp ³	Both	0.227	2	0.842	0.546	0.646	32.413	58.6
	Steric	0.240	2	0.835	0.491	0.684	26.014	100.0
	Elec.	0.234	2	0.838	0.598	0.607	40.202	0.0
O-sp ³	Both	0.231	2	0.840	0.544	0.647	32.199	58.0
	Steric	0.221	2	0.846	0.483	0.689	25.238	100.0
	Elec.	0.234	2	0.838	0.598	0.607	40.202	0.0
^a C-sp ³	Steric	0.376	3	0.658	0.716	0.445	42.758	100.0

^a compounds 75 and 104 were omitted.

The steric contour plot of the best predictive model for the log(D-6) activity is illustrated in Figure 4.12. This plot is very similar to that of the model for all compounds (Figure 4.4). A large yellow contour is located in front of and below the O₁₄ atom. Green contours are placed in the right and left hand side of the yellow contour. As in the model for all compounds, no steric contour was observed around the C₃ position.

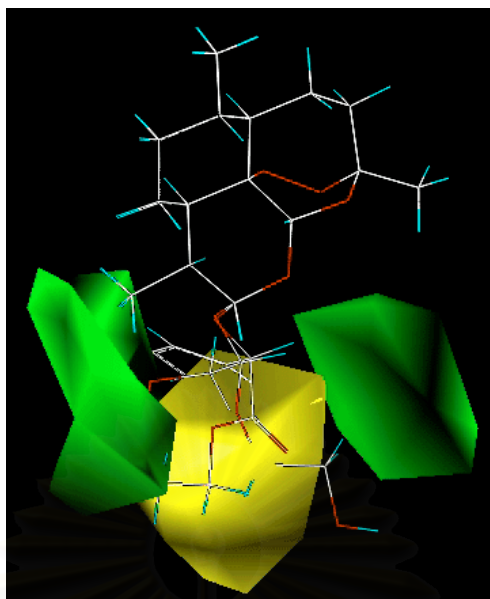


Figure 4.12 Steric contour plot of the log(D-6) model for compounds without C=O at the C₁₀ position.

For the log(W-2) activity, the CoMFA results are given in Table 4.14. The results also have higher predictive ability than those of the previous group, as indicated by higher q^2 value. As in the log(D-6) models, the types of probe atom and field have no significant effect to the predictive power. The analysis on residual values revealed that compounds 75, 101, 103, and 104 have residual values larger than 1.5. Therefore, these four compounds were excluded from the training set. Using the model with C-sp³ probe atom and steric field only, a large improvement in q^2 value from 0.192 to 0.492 was obtained.

Table 4.14 CoMFA results for log(W-2) activity of compounds with C=O at the C₁₀ position.

Probe	Fields	q^2	onc	PRESS	r^2	S.E.	F	% Steric
C-sp ³	Both	0.169	2	0.893	0.517	0.681	28.866	58.6
	Steric	0.192	3	0.889	0.577	0.643	24.093	100.0
	Elec.	0.165	2	0.895	0.542	0.663	31.986	0.0
O-sp ³	Both	0.164	2	0.895	0.512	0.684	28.363	56.3
	Steric	0.173	2	0.891	0.471	0.712	24.014	100.0
	Elec.	0.165	2	0.895	0.542	0.663	31.986	0.0
^a C-sp ³	Steric	0.492	3	0.575	0.747	0.406	48.245	100.0

^a compounds 75, 101, 103, and 104 were omitted.

From the best predictive model of the log(W-2) activity, the steric contour plot was drawn (Figure 4.13). A large steric favorable region (green color) was observed surrounding the space that is remote from the β -C₁₀ position. As in the log(D-6) model, a steric unfavorable region (yellow color) was located under the β -C₁₀ position. Moreover, a steric field around the C₃ position has no relation to the antimalarial activity since it did not appear in the contour plot.

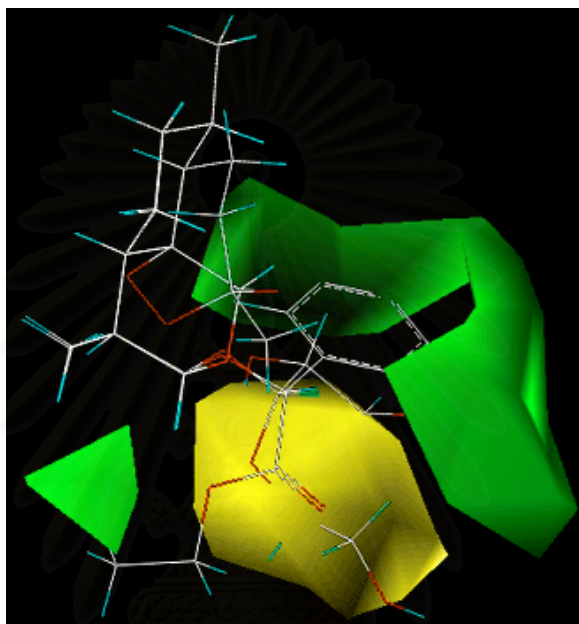


Figure 4.13 Steric contour plot of the log(W-2) model for compounds without C=O at the C₁₀ position.

The predicted activities of compounds in the testing set using the best predictive models of log(D-6) and log(W-2) activities are given in Figure 4.14 and 4.15. The predictions are not so good quantitatively but they are fairly good in qualitative manner. If the predicted value is high, the real activity is also high, and if the predicted value is low, a compound is less active.

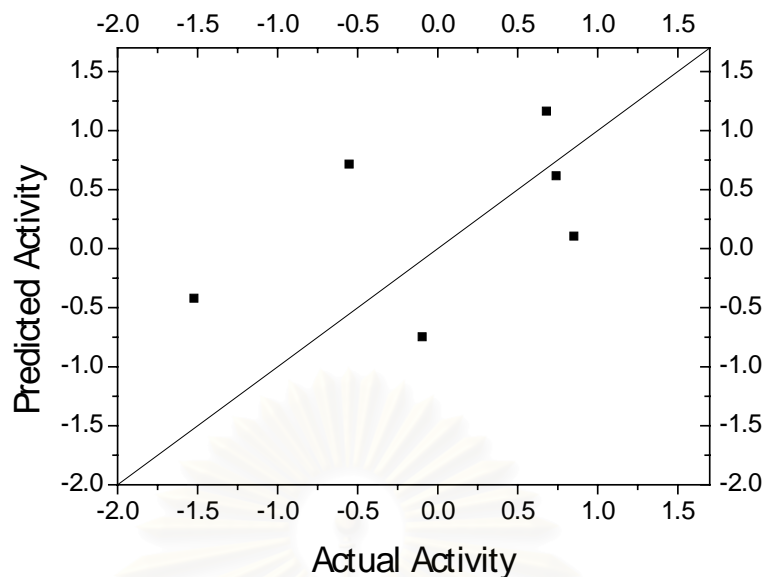


Figure 4.14 Comparison between actual and predicted log(D-6) activities for 6 compounds without C=O at the C₁₀ position in the testing set.

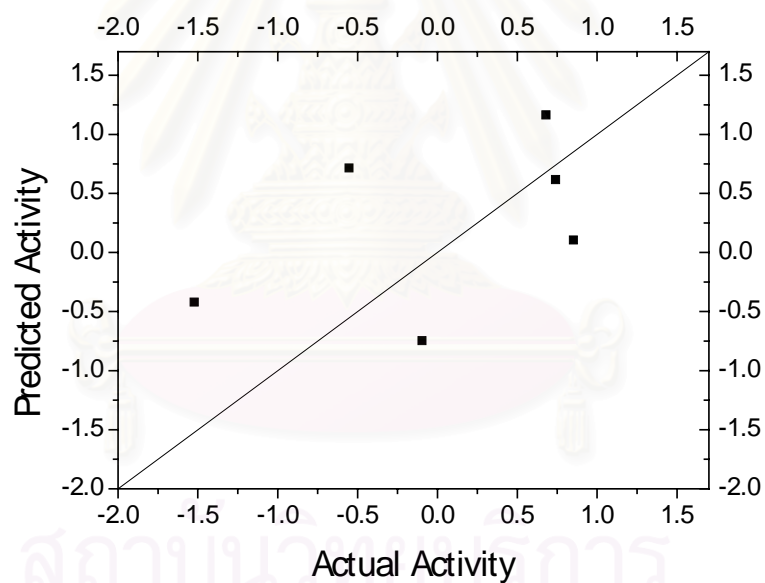


Figure 4.15 Comparison between actual and predicted log(W-2) activities for 6 compounds without C=O at the C₁₀ position in the testing set.

4.4 Summary

The effects of some adjustable parameters on CoMFA results of artemisinin derivatives were studied. The HF/3-21G method was justified as the suitable method for optimization and charge calculation. Interestingly, the geometry obtained from the conformational search using the Tripos force field gave worse results than the

HF/3-21G structure without the conformational search. Five different alignment rules gave slightly different q^2 values. The electrostatic cut-off value has less effect on the q^2 value than the steric cut-off does. And the default setting was shown to be the suitable values. From the investigations, the grid position has significant effect on the q^2 value. Although the q^2 -GRS method was proposed to give orientation-independent q^2 value, it does not guarantee the better q^2 value over the conventional method. Therefore, the application of q^2 -GRS method in the CoMFA study should be carefully considered.

After obtaining justified adjustable parameters, the CoMFA calculations were performed for all 104 artemisinin derivatives as well as two divided groups with more structural similarity, i.e., compounds with and without C=O at the C₁₀ position. The predictive power of all models is quite low. However, if compounds with high residual value were excluded, the predictive power was significantly improved. However, for log(W-2) activity of compounds with C=O at the C₁₀ position, no improvement could be obtained. The predictive power of models for divided groups is comparable to that for the set of all compounds. Therefore, this implies that the CoMFA is perhaps not a suitable method for explaining the activities of our types of compounds. The possible reason is that the steric and electrostatic fields only in the CoMFA can not be fully responsible for describing the activities of these compounds. Additional fields, such as hydrophilic field, are recommended. Although the hydrophilic field was reported to give a poor relationship with log(D-6) activities of 157 artemisinin analogues [94], the importance of F_{h2o} parameter to our compounds as indicated from the QSAR models in Chapter 3 shows the significance of the hydrophilic field for our compounds.

The interactions between heme iron and artemisinin compounds (see Chapter 6) could be another possibility for the unsatisfactory application of CoMFA on our compounds. Since molecular mechanics method is not suitable to describe the interactions between transition metal (iron) and oxygen atoms which involve free radicals, quantum chemical method that is more precise should be applied instead. The investigations on the mechanism of action of artemisinin compounds are reported in the Chapter 6.

CHAPTER 5

Molecular Docking

5.1 Introduction

From the pharmacological concept of drug-receptor interaction, good drugs must possess a specific structural change to yield a suitable geometry for the binding with their target receptor. In order to design a new effective drug, the knowledge of drug-receptor interaction is required. Hence, the structure of the complex between drug and receptor is needed. However, at the present time, there is relatively small experimental data for the drug-receptor complex due to experimental difficulties, e.g., the isolation of the complex without breaking it. For this reason, the binding between drug and receptor is advantageously studied theoretically using molecular modeling method, such as molecular docking.

The molecular docking method has been widely used for some years. Its ultimate goal is to obtain the precise docking structure, which corresponds to the energetically most stable configuration. In the docking process, the ligand and receptor are either constructed manually or obtained from databases, such as X-ray database and the Protein Data Bank. The docking algorithm could be performed either manually or automatically. The manual dock, which utilizes interactive computer graphics, could be very effective if good assumptions of the expected binding mode are provided, e.g., from the binding mode of a closely related ligand. However, there are evidences from crystallographic X-ray experiments that quite different binding modes could possibly be adopted even for very similar inhibitors. Thus, the manual dock must be used with cautions. On the other hand, automatic dock can be performed without a prior knowledge of the system since it employs the Monte Carlo simulation method, which allows many possibilities of docking configuration to be considered, to achieve its goal. Nevertheless, the automatic dock requires longer computational times.

In the mechanism of action of artemisinin compounds (see more details in Chapter 6), heme iron was proposed to attack the endoperoxide linkage of artemisinin either at O₁ position or O₂ position (Figure 5.1). Therefore, the docking calculations between artemisinin derivatives and heme were carried out to elucidate the interactions between heme iron and oxygen in the endoperoxide linkage.

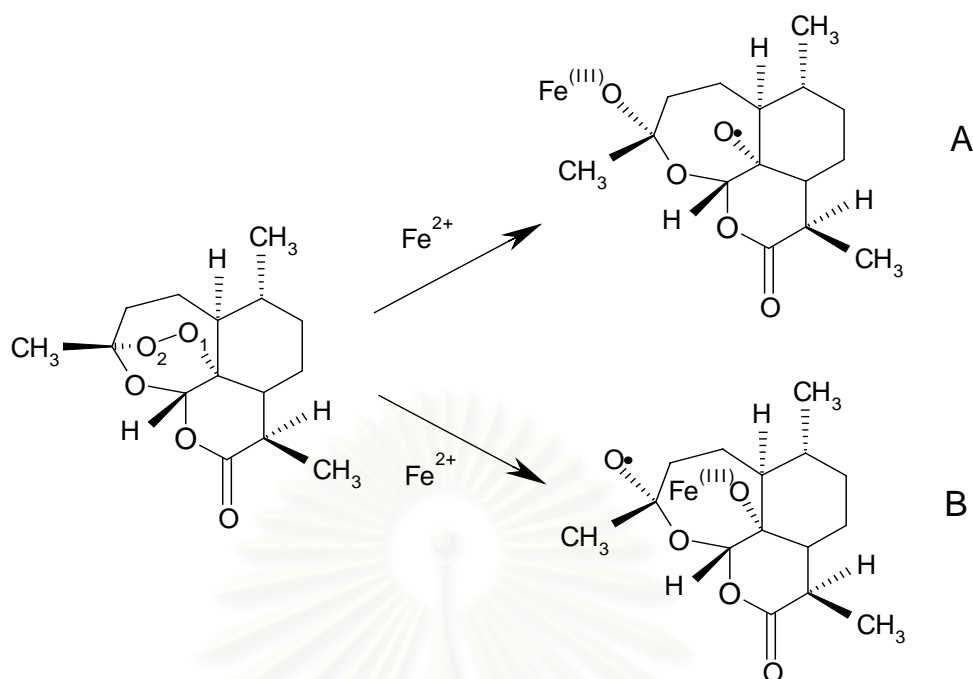


Figure 5.1 Proposed mechanism of action of artemisinin compounds.

Shukla and co-workers [99] studied the docking of artemisinin and deoxyartemisinin with hemin [Fe(II) and Fe(III)] using the Dock module in the SYBYL software, a direct docking algorithm. In their study, the artemisinin structure was built from the crystallographic X-ray structure of artemether. Although the study did not elaborate on how the structure of artemisinin was obtained from artemether, it is very likely that the geometry optimization was performed at either the molecular mechanics or semiempirical level since only these methods are available in the SYBYL. For their docking calculations, only 3 orientations of artemisinin around the heme molecule were considered. Furthermore, the Gasteiger method, an empirical method implemented in the SYBYL, was used for the atomic charge calculations. However, this empirical method has no parameters for iron; therefore, the charge of the heme iron was assigned under the assumption that the change in the charge distribution of the heme iron should be equal to that of the heme model where iron was replaced by aluminium. Moreover, the general parameters for metal were used in the docking calculations. Their employed docking scheme might influence the docking result in favor of one of the heme-artemisinin configurations and yield an inaccurate model for the complex. It is quite important to have an accurate model for the heme-artemisinin complex. Since this knowledge can be used to design better and more potent antimalarial drugs.

In this study, automated docking calculations were carried out to eliminate the bias in selecting preferred configurations (orientations). The X-ray crystallographic structure of artemisinin was used for artemisinin derivatives having C=O at the C₁₀ position instead of that of artemether which is quite different from the artemisinin structure especially at the lactone ring. And the crystallographic X-ray structure of dihydroartemisinin was used for artemisinin derivatives without C=O at the C₁₀ position. For the heme iron, accurate *ab initio* calculations were performed to obtain its atomic charge (as well as those of artemisinin) instead of using a crude approximation for the charge of iron, and specific parameters for iron were used in the docking calculations. Moreover, the effects of different heme structures were also considered. Furthermore, 104 artemisinin derivatives were studied to investigate the structural effect.

5.2 Computational Method

The docking calculations were performed using the automated docking program, AutoDock 2.4 software [100-101]. The AutoDock employs a simulated annealing Monte Carlo simulation in combination with a rapid grid-based energy evaluation method [102]. The rapid energy evaluation is achieved by precalculating atomic affinity potentials for each atom type present in the ligand molecule. For example, artemisinin 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 5.2).

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 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 Solmajer [103] is used to model solvent effect.

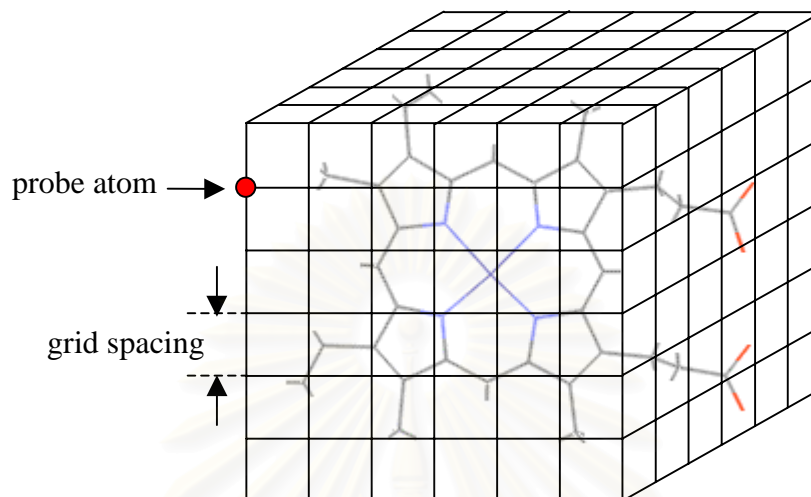


Figure 5.2 Grid base energy evaluation.

In one docking calculation, the simulations were performed for 100 annealing cycles. At the first cycle, the initial annealing temperature (RT) was set to 100 kcal/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 \AA and a maximum orientation step of 5° . The energy of the new configuration was then calculated. The selection of the new configuration was based on the Metropolis algorithm [101]. The cycle terminates if the ligand makes 30,000 accepted or 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 \AA were grouped together. In each group, the lowest energy configuration was selected as the representative of that group. The “% Occurrence” (it was sometimes denoted as “% Occur” due to the limited space in some tables) was used

to represent the number of members (configurations) in each group. Our attention was focused to the group with the highest % Occurrence or “the most occurring configuration”. And this configuration is most probably corresponding to the docked configuration in the real system.

5.3 Determination of Suitable Docking Parameters

Since there are some adjustable parameters in Autodock, series of docking calculations were performed in order to establish the suitable parameters for further studies. Unless stated otherwise, the following settings were applied as default parameters. A grid map has a dimension of 25 x 25 x 25 Å with a spacing of 0.5 Å. For a simulated annealing calculation, 100 docking runs with 100 annealing cycles per run were performed. A cycle is terminated if the ligand made 30,000 accepted or 30,000 rejected moves. The initial annealing temperature, RT, was 100 kcal/mol with the reduction factor of 0.90. The combined AMBER/MMFF [104-105] parameters for the Lennard-Jones 12,6 and Coulomb potentials, taken from the authors of the program [106], were used. The heme structure was taken from the Protein Data Bank [107] (denoted as heme-pdb).

5.3.1 Force Field Parameters

Since the AMBER force field has only general parameters for metal atom, the combined AMBER/MMFF force field which have specific parameters for iron were chosen. Nevertheless, the comparison between these two force fields was evaluated. For this purpose, the heme-pdb with HF/3-21G atomic charges was used. And for artemisinin derivative compounds, the HF/3-21G method was used for the geometrical optimization and atomic charge calculation. The docking results are shown in Table 5.1.

Interestingly, results of compound 2 and 69 are different for the two force fields. These differences can be explained by the comparison of the characteristics between these two force fields. The AMBER parameters have longer equilibrium internuclear separation and higher well depth than the combined AMBER/MMFF parameters. This indicates stronger repulsive forces of the AMBER parameters. Therefore, the AMBER force field may cause configurations with Fe away from the endoperoxide bridge becoming the most occurring configuration. However, the second most occurring configurations, which are slightly more stable, have Fe pointed toward to the bridge. As

indicated above that the AMBER force field has no specific parameter for Fe and it predicted the less stable configurations as the most occurring configurations. Therefore, the combined AMBER/MMFF force field was selected for the entire calculations.

Table 5.1 Docking results of heme-pdb and artemisinin derivatives with AMBER and AMBER/MMFF force fields.

Compound No.	Energy (kcal/mol)	% Occurrence	O ₁ -Fe Distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
AMBER						
1	-30.58	26	2.75	3.68	5.46	5.74
2	-29.20	17	6.09	6.46	4.52	4.76
	-29.29	16	2.84	3.72	5.37	5.86
42	-30.00	40	2.83	3.77	5.47	5.71
69	-30.58	24	5.11	5.34	6.71	4.83
	-30.71	22	2.81	3.84	5.50	5.65
AMBER/MMFF						
1	-30.27	22	2.72	3.62	5.41	5.73
2	-29.14	22	2.92	3.86	5.48	5.93
42	-29.85	28	2.88	3.91	5.51	5.75
69	-30.14	22	2.65	3.61	5.39	5.46

5.3.2 Temperature Reduction Rate, Starting and Final Temperature

For large compounds such as proteins and weakly bound complexes such as drug-receptor complexes, it is extremely difficult to locate the global minimum of their potential surfaces. The simulated annealing is hailed to be able to cope with this problem. Thus, it is employed in the Autodock program. 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. As the temperature decreases the molecules tend to line themselves up in a state of minimum energy. The temperature should be kept reducing until a suitable value or the best solution, i.e., arriving at a large single crystal, is finally achieved. Therefore, the temperature

reduction rate should be carefully controlled. Normally, this value should be above 0.90.

In this section, the effects of temperature reduction rate as well as the starting (T_s) and final temperatures (T_f) were investigated. Three different reduction rates, i.e., 0.90 (model A), 0.95 (model B), and 0.99 (model C), were considered. And various T_s and T_f were used. Fundamentally, the appropriate T_s and T_f are those above and below the melting point of the studied compound, respectively. The melting point for artemisinin is around 156-157 °C. Moreover, the T_s should be high enough to allow the molecular system to overcome the energy barriers of the local minimum. Therefore, the temperatures were varied from 503.3 to 50,327.1 K for the T_s and from 1.24 to 310.9 K for the T_f . The heme-pdb structure and artemisinin X-ray structure both with the HF/3-21G atomic charges were used. The results were shown in Table 5.2.

Table 5.2 Docking of artemisinin to heme-pdb with different starting and final temperatures.

Model	T_s (K)	T_f (K)	% Occur	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe Distance (Å)	O ₁₃ -Fe Distance (Å)	O ₁₁ -Fe distance (Å)
A1	50,327.1	1.3	22	-30.44	2.70	3.57	5.41	5.70
A2	48,867.6	310.9	9	-28.89	6.32	6.85	5.13	4.82
			6	-29.32	2.75	3.72	5.45	5.75
A3	25,163.6	1.3	21	-30.47	2.78	3.75	5.50	5.76
A4	12,581.8	1.3	14	-30.38	6.87	6.92	5.22	4.83
			13	-30.22	2.64	3.61	5.36	5.71
A5	5,032.7	1.4	23	-30.39	2.55	3.11	5.17	5.38
A6	5,032.7	155.5	10	-29.19	2.68	3.50	5.42	5.61
A7	2,516.4	1.3	21	-30.36	2.74	3.55	5.38	5.77
A8	1,258.2	1.3	24	-30.32	2.78	3.77	5.50	5.77
A9	503.3	1.2	23	-30.43	2.78	3.76	5.50	5.77
A10	503.3	142.1	6	-28.14	5.68	5.46	3.17	4.23
			5	-27.89	2.82	3.50	5.33	5.88
B1	12,581.8	74.5	14	-29.99	2.79	3.66	5.47	5.80
B2	5,032.7	29.8	16	-29.98	2.51	3.05	5.11	5.41

Table 5.2 (Continued)

Model	T _s (K)	T _f (K)	% Occur	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe Distance (Å)	O ₁₁ -Fe distance (Å)
B3	4,026.2	8.5	22	-30.26	2.77	3.74	5.50	5.75
B4	503.3	154.7	6	-29.42	2.56	3.19	5.23	5.45
C1	5,032.7	12.1	16	-30.86	2.50	3.09	5.16	5.37
C2	503.3	150.7	14	-29.77	6.33	6.79	5.04	4.64
			12	-30.20	2.78	3.83	5.51	5.75

From the results, if the T_f is still high (more than 100 K), the docked configurations were clustered into too many groups which consist of only a few members as observed in models A2, A6, A10, B4, and C2. Moreover, the most occurring configurations in the models A2, A4, A10, and C2 have heme irons away from the endoperoxide moiety. Comparing models A5 and A6 which have the same T_s, the model A5 with lower T_f has higher “% Occurrence” and a lower energy ($\Delta E = 1.2$ kcal/mol). Therefore, the low T_f is recommended. For the T_s value, models A1, A3, A4, A5, A7, A8, and A9 gave quite similar results except the model A4. The most occurring configuration in the model A4 has Fe away from artemisinin. However, the second most occurring configuration has Fe pointing toward O₁ atom as in the other model. It seems that the T_s, if high enough, does not have significant effect to the docking results. Therefore, the highest T_s in this study was used for further calculations. For the effect of temperature reduction rate, increasing the value from 0.90 to 0.95 and 0.99 requires more calculation time while nearly the same results were obtained. Therefore, the temperature reduction rate of 0.90 seems to be sufficient for these calculations.

5.3.3 Grid Spacing and Grid Dimension

In the Monte Carlo simulation, the energy of the molecular system has to be calculated for every movement. Nevertheless, it is impossible to calculate the energy by direct method for each movement because it is a very time-consuming process. Therefore, the Autodock uses a rapid grid-based energy evaluation for this purpose. The investigations on the effects of the grid spacing and grid dimension were carried out. Three grid boxes with dimension of 20 x 20 x 20 Å³ (model D), 25 x 25 x 25 Å³ (model

E), and $30 \times 30 \times 30 \text{ \AA}^3$ (model F) and various grid spacings were examined. The grid size is calculated from a side of grid box divided by grid spacing. The results were shown in Table 5.3.

Table 5.3 Docking of artemisinin to heme with different grid dimension and spacing.

Model	Grid Size	Grid Spacing	% Occur	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
D1	100	0.20	26	-32.99	2.53	3.43	5.26	5.52
D2	50	0.40	25	-31.61	2.57	3.45	5.28	5.56
D3	25	0.80	24	-27.89	6.22	6.70	4.95	4.68
			11	-27.19	2.77	3.72	5.53	5.73
D4	20	1.0	31	-25.20	5.65	5.26	3.02	3.60
			7	-25.10	3.55	4.85	6.31	6.04
E1	100	0.25	26	-32.58	2.59	3.48	5.30	5.59
E2	50	0.50	22	-30.44	2.70	3.57	5.41	5.70
E3	25	1.0	24	-24.97	5.65	5.25	2.98	3.69
			3	-24.72	3.60	4.91	6.35	6.10
F1	60	0.50	18	-30.39	2.53	3.02	5.12	5.38
F2	50	0.60	17	-29.59	2.81	3.76	5.54	5.79
F3	40	0.75	19	-27.74	3.00	4.07	5.71	5.97
F4	30	1.0	18	-25.54	6.49	7.10	5.39	5.49
			4	-24.98	3.40	4.66	6.18	6.07

Considering model D with the total grid dimension of $20 \times 20 \times 20 \text{ \AA}^3$, models with a grid spacing of 0.8 \AA (model D3) and 1.0 \AA (model D4) have Fe away from the endoperoxide group in the most occurring configuration. However, in the model D3 there is a configuration with Fe close toward the endoperoxide moiety. Models D1 and D2 with small grid spacing give comparable results having the O₁-Fe as the shortest distance. In model E with $25 \times 25 \times 25 \text{ \AA}^3$ grid dimension, the model E3 with the largest grid spacing has O₁₃-Fe as the shortest distance. However, configuration with shorter O₁-Fe is observed although the distance is quite far (3.60 \AA). Models E1 and E2 both have Fe pointed toward the O-O bond but the O₁-Fe distance in the model E1 which has smaller grid spacing is shorter than that in the model E2. The results for model F with

30 x 30 x 30 Å³ grid dimension display the same trend as those in the model E do. However, models F have lower “% Occurrence” in the most occurring configurations than models D and E. This is possibly due to its larger total grid dimension which has more space for artemisinin to be able to move around heme molecule. Hence, larger numbers of cluster groups were observed. From Table 5.3, it appears that consistent docking results could be obtained using small grid spacing and moderate grid dimension, i.e., not too large but big enough to cover the receptor molecule and free space for the movement of ligand. Therefore, the grid dimension of 25 x 25 x 25 Å³ with 0.5 Å grid spacing were chosen for further calculations.

5.3.4 Cluster Tolerance Value

The Autodock uses cluster method to analyze the docking results. The structures, which have root-mean-square-deviation (rmsd) less than a defined cluster tolerance value, are grouped together. Only the lowest energy representative from each cluster is presented in an output. Hence, a cluster tolerance is a crucial value. If a high value is selected, some detail information may be overlooked. Thus, the cluster tolerance values of 1, 2, 3, 4, and 5 Å were used to investigate its effect (Table 5.4). The cluster tolerance value of 1.0 Å was chosen for further docking calculations since detailed examination of results would be allowed.

Table 5.4 Docking of artemisinin to heme with different cluster tolerance value.

Cluster Tolerance (Å)	No. of Group	% Occur	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1.0	18	22	-30.44	2.70	3.57	5.41	5.70
2.0	4	42	-29.94	2.81	3.94	5.62	5.64
3.0	2	54	-29.85	2.70	3.60	5.39	5.72
4.0	2	55	-29.90	2.75	3.75	5.42	5.76
5.0	2	51	-29.92	2.72	3.73	5.46	5.69

5.3.5 Dielectric Constant

In docking calculations, the solvent effect is introduced in the simulation through the dielectric constant using a solvent screening model. Either the distance-dependent dielectric function or the constant dielectric function could be used. In order

to study the solvent effect, three values of difference dielectric constants for the constant dielectric function, i.e., 10.0, 50.0, and 100.0, and the distance-dependent dielectric function were used. The results were shown in Table 5.5.

Table 5.5 Docking of artemisinin to heme with different dielectric constant value.

Model	Dielectric Constant	% Occur	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
H1	dependent	22	-30.44	2.70	3.57	5.41	5.70
H2	10.0	26	-29.54	2.80	3.71	5.46	5.82
H3	50.0	28	-30.02	5.48	5.29	3.10	3.41
		24	-29.37	2.79	3.79	5.53	5.75
H4	100.0	33	-30.16	5.45	5.24	3.04	3.36
		19	-29.64	2.78	3.77	5.51	5.77

The high dielectric constants (models H3 and H4) made the configuration with Fe point away from O₁ as the most occurring configuration. However, the configurations with Fe point toward O₁ are still existed although with lower number of “% Occurrence”. In the model H2 with the dielectric constant of 10.0, the most occurring configuration is similar to that of the model H1 but with slightly higher energy. Therefore, the constant function with value of 10.0 and the distance-dependent function can be used. However, the distance-dependent function is selected as it is recommended in the AutoDock manual.

5.3.6 Atomic Charges Calculation Method

For docking calculations, the total energy is the summation of electrostatic and steric potentials. Since the electrostatic potential is built from atomic charges, the choice of atomic charges of both ligand and receptor molecules would have effects on the docking configurations. Therefore, the investigations on their effects were carried out.

A. Atomic charges of artemisinin molecule

In order to find the suitable atomic charges for artemisinin molecule, series of docking calculations were performed using atomic charges obtained from various methodologies. These included the Mulliken population analysis (MPA) at AM1,

HF/3-21G, HF/D95V, and HF/6-31G* levels, the electrostatic potential fit (ESPFIT), and the natural population analysis (NPA) at the HF/3-21G level. The docks of heme-pdb with the HF/3-21G charges to the HF/3-21G optimized structure of artemisinin were performed. The results given in Table 5.6 show no significant difference on the docking configurations. Therefore, it was justified to use the MPA atomic charges at HF/3-21G level for building electrostatic potential of artemisinin derivatives in docking calculations.

Table 5.6 Docking results of heme-pdb and artemisinin HF/3-21G optimized structure with different atomic charges.

Artemisinin Atomic charges	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
AM1 MPA	-30.04	2.79	3.73	5.46	5.81
HF/3-21G MPA	-30.27	2.72	3.62	5.41	5.73
HF/3-21G ESPFIT	-29.81	2.73	3.65	5.43	5.73
HF/3-21G NPA	-30.56	2.77	3.61	5.37	5.82
HF/D95V MPA	-29.97	2.81	3.80	5.50	5.80
HF/6-31G* MPA	-29.97	2.82	3.80	5.49	5.83

B. Atomic charges of heme molecule

Similarly, atomic charges of heme-pdb taken from ZINDO/S, HF/STO-3G, HF/3-21G, and HF/6-311G** calculations were used in docking calculations of heme-pdb and artemisinin (HF/3-21G optimized structure and charges) to determine the effect of charges of heme. The results were shown in Table 5.7. In this case, the docking results depended on the quality of heme-pdb atomic charges. The ZINDO/S atomic charges predicted a different configuration from the others. The docking configurations of heme-pdb with HF/STO-3G, HF/3-21G, and HF/6-311G** atomic charges have heme iron point toward the endoperoxide. Note that the HF/6-311G** results gave the smallest Fe-O distance (2.49 Å). This difference in Fe-O distance is quite crucial for the description of artemisinin-heme interaction. Thus, the HF/6-311G** charges were used for electrostatic potential of heme.

Table 5.7 Docking results of heme-pdb with different atomic charges and artemisinin HF/3-21G optimized structure with HF/3-21G charge.

Heme Atomic Charges	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
ZINDO/S	-30.56	5.66	5.65	3.51	3.71
STO-3G	-31.44	2.63	3.47	5.28	5.66
HF/3-21G	-30.27	2.72	3.62	5.41	5.73
HF/6-311G**	-31.40	2.49	3.12	5.14	5.40

5.3.7 Optimization Method for Artemisinin Compounds

From the section 2.4, the results indicated that HF/3-21G is the lowest level that can reproduce most of the X-ray structure, especially the bond length of the endoperoxide linkage. However, this does not mean that the docking results using HF/3-21G optimized structure would be the same as those with the X-ray structure. Therefore, we have to validate the use of optimized structures.

The docking calculations between AM1, HF/3-21G, and HF/6-31G* optimized structures of artemisinin and heme were carried out and compared with those of the artemisinin's X-ray structure where atomic charges were assigned from the single-point AM1, HF/3-21G, and HF/6-31G*. The heme-pdb structure with the HF/3-21G atomic charges was used for the structure of heme in the docking. A comparison of docking results between those using X-ray and theoretical optimized structures of artemisinin with corresponding atomic charges is shown in Table 5.8. Using the AM1 atomic charges, the AM1 optimized structure gave too short O-Fe distances as compared to those in the X-ray structure. The docking results of HF/3-21G and HF/6-31G* optimized structures and atomic charges have similar O-Fe distances to those of the X-ray structure with corresponding atomic charges. However, the use of HF/3-21G optimized structures would be preferred for economic reasons. Thus, the HF/3-21G level of theory was selected for the optimization of the artemisinin derivatives.

Table 5.8 Docking results of heme-pdb and artemisinin optimized structures at various levels of theory.

Structure	Atomic charges	Energy (kcal/mol)	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
X-ray	AM1	-30.17	2.73	3.58	5.40	5.75
	HF/3-21G	-30.44	2.70	3.57	5.41	5.70
	HF/6-31G*	-30.26	2.53	3.03	5.10	5.43
AM1	AM1	-30.13	2.57	3.09	5.18	5.58
HF/3-21G	HF/3-21G	-30.27	2.72	3.62	5.41	5.73
HF/6-31G*	HF/6-31G*	-30.27	2.54	3.10	5.14	5.38

5.3.8 Heme Structures

In order to investigate the effect of the heme structure to docking results, 5 heme structures were selected. These structures are all different owing to the source of heme and the oxidation state of iron. The first structure, heme-pdb, was taken from the Protein Data Bank (id 1CTJ). In this structure, the iron is positioned slightly above the porphyrin plane (Figure 5.3a). The second structure, heme-model, which was taken from the AMBER database [108], has the planar geometry (Figure 5.3b). The third structure, heme-hemin, was modified from the crystallographic X-ray structure of chlorohemin of the Cambridge Crystallography Data Bank [109]. This structure possesses a pyramidal shape with Fe on the top (Figure 5.3c).

In the hemoglobin degradation process by the malaria parasite, the proximal ligand (histidine) may possibly still be attached to the heme iron. As a result, the fourth and the fifth structures, heme-deoxy and heme-oxy, respectively, were obtained from the modifications of deoxy and oxy forms of hemoglobin containing histidine as the proximal which were taken from the Protein Data Bank (id 1A3N and 1HHO, respectively). In the heme-deoxy, the histidine pulls the Fe atom to lie below the porphyrin plane and gives it a basin-like structure (Figure 5.3d). In the heme-oxy, the oxygen molecule bound to heme iron was deleted before the docking calculations. As in the heme-deoxy, the porphyrin plane has a basin-like structure due to the attraction to the heme iron by histidine. However, the interaction with oxygen molecule causes the

Fe atom to be drawn up above the plane (Figure 5.3e) and thus results in the structure which is markedly different from the heme-deoxy.

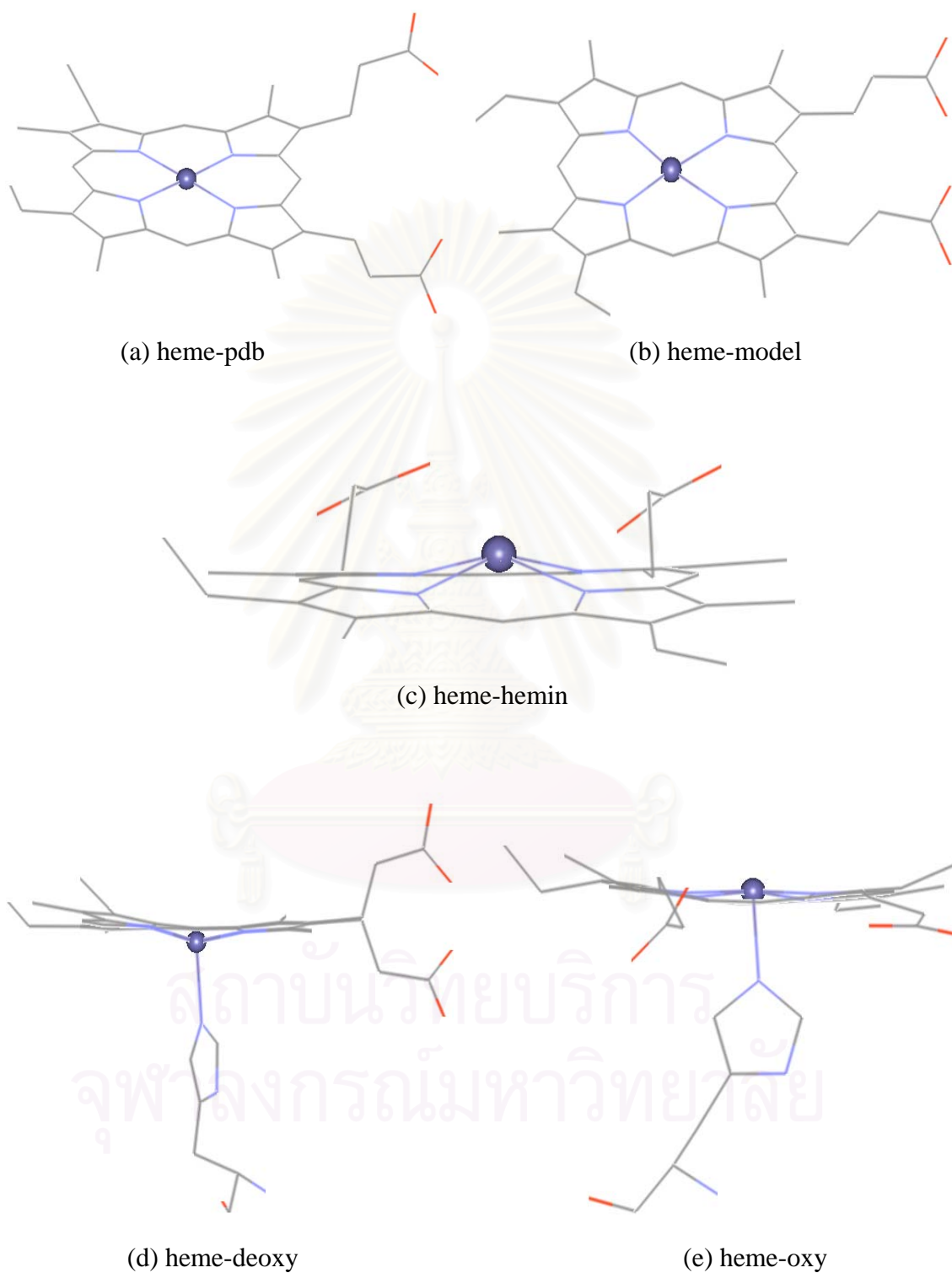


Figure 5.3 Structures of 5 heme molecules (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, (e) heme-oxy.

The atomic charges for all five heme molecules were assigned at the HF/6-311G** level. In order to investigate the effect of each heme structure on docked configuration, a moderate number of artemisinin derivatives are required. Therefore, 16 artemisinin compounds were selected. The HF/3-21G level was used for geometry optimization and atomic charge calculations. Results of docking between 16 artemisinin derivatives and five different heme structures were demonstrated in Table 5.9 to 5.13 and Figure 5.4 to 5.8. As mentioned earlier, the endoperoxide moiety is essential for antimalarial activity, therefore, the reports emphasized particularly on distances between the endoperoxide oxygens, O₁ and O₂, and the heme iron.

A. Heme-pdb

Almost all of docking configurations between the heme-pdb and artemisinin derivatives have the heme iron pointing towards the endoperoxide of artemisinin derivatives. And it is this heme iron that was located nearest to the artemisinin derivatives. The distances between Fe and O₁ and between Fe and O₂ in all compounds are in the range of 2.5 to 3.3 Å and 3.1 to 3.8 Å, respectively (Table 5.9). Thus, the binding of heme and artemisinin derivatives is mainly described by interactions between heme iron and endoperoxide oxygen(s). The docking configurations between the heme-pdb and artemisinin derivatives were illustrated in Figure 5.4. Compounds 85 and 87 have the longest O₁-Fe distances (3.27 and 3.22 Å) which could possibly be explained by the existence of Br group on the same side of the endoperoxide moiety, thus preventing the heme iron from binding with the endoperoxide oxygens. To prove this steric hypothesis, the analog of compound 85 without Br group (compound “85-no-br”) was built and docking calculation was then performed. The calculation showed the shortening of the O₁-Fe distance to 2.50 Å, which is comparable to those in other compounds. In exception to other compounds where the most occurring configurations showed O₁-Fe as the closest distances, the compound 2 had the most occurring configurations with O₁₃-Fe as the closest distance. This is possibly owing to the replacement of -CH by C=O at the C₉ position, which causes the structure of the lactone ring to change from a twisted-boat form to a nearly-envelop form. This structure facilitates the encroachment of Fe to O₁₃. However, the second most occurring configurations, which have the lowest energy, had shorter O₁-Fe distance. The compound 88 has two most occurring configurations. Notably, the one with shorter O₁-Fe distance had lower energy. Interestingly, deoxyartemisinin (compound “deoxy”), an

inactive analogous of artemisinin where the endoperoxide bridge was replaced by the epoxide bridge (single oxygen atom bridge), also bound with the heme iron with the binding energy (-29.04 kcal/mol) comparable to other compounds.

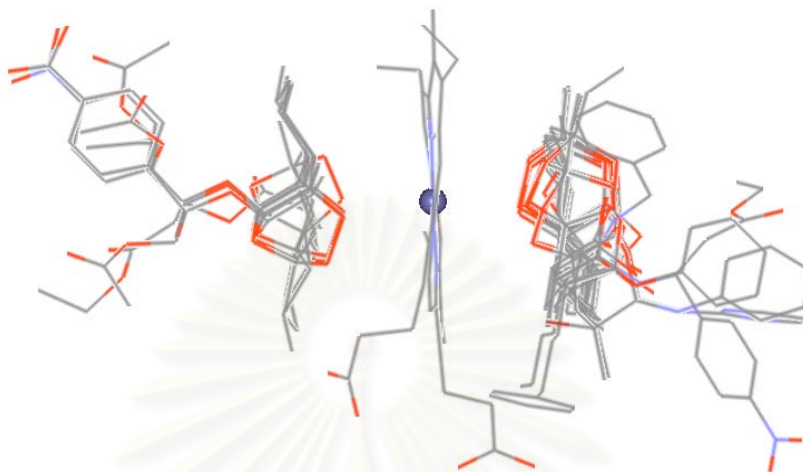


Figure 5.4 Superimposed docking configurations between the heme-pdb and 16 artemisinin derivatives (without hydrogen atoms).

Table 5.9 Docking results of heme-pdb and 16 artemisinin derivatives.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe Distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-31.40	25	2.49	3.12	5.14	5.40
2	-29.42	31	5.48	5.10	2.91	3.39
	-29.84	22	2.76	3.56	5.30	5.77
28	-35.67	37	2.78	3.67	5.40	5.83
35	-32.19	21	2.56	3.03	5.11	5.52
42	-30.90	30	2.51	3.15	5.16	5.27
43	-30.65	31	2.64	3.41	5.31	5.46
54	-30.53	38	2.64	3.56	5.38	5.47
56	-31.16	31	2.60	3.44	5.35	5.41
68	-31.44	31	2.52	3.05	5.09	5.30
69	-31.22	43	2.65	3.54	5.36	5.48
75	-29.47	35	2.73	3.53	5.30	5.62
85	-29.61	28	3.27	3.73	5.26	6.12
87	-29.33	19	3.22	3.67	5.27	6.08

Table 5.9 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe Distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
88	-30.52	25	2.53	3.17	5.23	5.23
	-30.26	25	2.67	3.53	5.36	5.53
94	-35.66	18	2.86	3.79	5.47	5.76
Deoxy	-29.04	28	-	3.24	5.41	6.19
85-no-br	-31.21	33	2.50	3.12	5.12	5.25

B. Heme-model

Most of the docking configurations between the heme-model and artemisinin derivatives have Fe closest to the endoperoxide bond, with O₁-Fe and O₂-Fe distances of 2.5 to 5.1 Å and 3.2 to 6.4 Å, respectively (see Table 5.10). Thus, the interaction between heme iron and the endoperoxide group appears to be an essential requirement. Similar to heme-pdb's results, compounds 85 and 87 have longer O₁-Fe distances than other compounds. However, these O₁-Fe distances are longer than in those of the heme-pdb. Since the heme-model has a planar structure, compounds 85 and 87 could not approach as close to heme iron as in the heme-pdb. The dock calculation on compound "85-no-br" revealed the O₁-Fe distance of 2.71 Å, which is in the same range as those in other compounds. This again confirms the steric hypothesis of the Br group. Like in the heme-pdb, the most occurring configuration of compound 2 places the endoperoxide group away from the heme iron. The most occurring configuration of the compound 35, which have Fe closer to O₁₁ than the endoperoxide oxygens, differs from other compounds. This might be owing to two factors. Firstly, the replacement of the oxygen atom by a nitrogen atom at the position 11 caused the change in the structure of the lactone ring from a chair to a twisted-boat form. Secondly, the heme-model has a planar structure. Both factors facilitated the approach of Fe to the lactone ring of the compound 35. However, the second most occurring configuration has O₁-Fe as the shortest distance between artemisinin and heme.

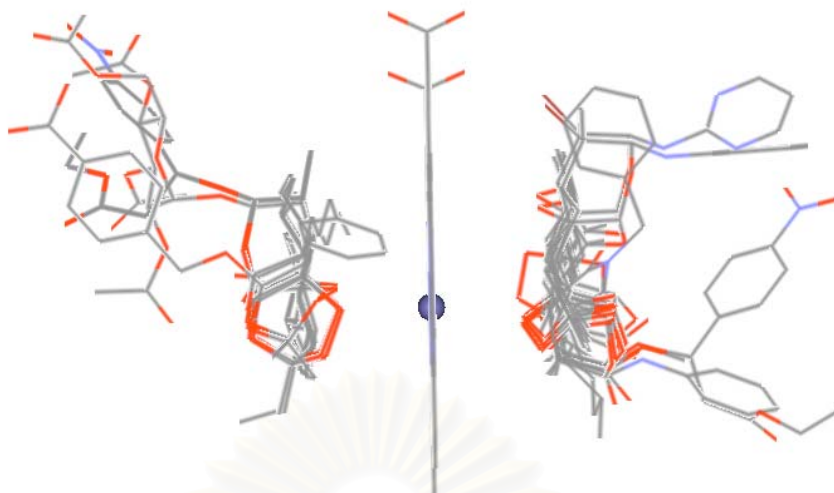


Figure 5.5 Superimposed docking configurations between the heme-model and 16 artemisinin derivatives (without hydrogen atoms).

Table 5.10 Docking results of heme-model and 16 artemisinin derivatives.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-29.92	22	2.75	3.66	5.38	5.79
2	-29.19	30	5.51	5.03	2.97	3.20
	-28.63	6	2.87	3.69	5.42	5.89
28	-33.16	25	2.69	3.49	5.36	5.69
35	-31.49	9	7.12	6.56	4.57	4.51
	-30.61	8	2.54	3.24	5.27	5.45
42	-29.62	25	2.72	3.64	5.43	5.57
43	-29.50	28	2.69	3.54	5.38	5.54
54	-29.45	19	2.64	3.43	5.28	5.50
56	-29.39	24	2.82	3.74	5.40	5.75
68	-29.36	25	2.82	3.67	5.40	5.74
69	-29.64	21	2.72	3.61	5.39	5.59
75	-28.28	16	2.81	3.44	5.26	5.69
85	-27.87	16	4.80	6.00	6.59	7.27
87	-27.74	13	5.12	6.35	6.74	7.42
88	-29.22	31	2.69	3.58	5.34	5.59
94	-33.20	19	2.71	3.53	5.40	5.57

Table 5.10 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
Deoxy	-28.71	16	-	3.17	5.30	6.17
85-no-br	-29.55	23	2.71	3.57	5.35	5.57

C. Heme-hemin

It was observed that the heme iron in nearly all docking configurations was located near the endoperoxide oxygens of artemisinin derivatives with O₁-Fe and O₂-Fe distances of between 1.9 to 2.7 Å and 2.4 to 3.6 Å, respectively (see Table 5.11). The heme-hemin has a pyramidal-like structure with Fe atom on the top, the structure which facilitates the approach of Fe to the endoperoxide moiety. As a result, the distances between Fe and the endoperoxide oxygens are very short (as compared to other models of heme). The O₁-Fe distances in most compounds are comparable to the experimental bond length in the oxy hemoglobin A (1.86 Å). Interestingly, the most occurring configurations of compound 85 and 87, where O₁ is placed close to Fe with obtained O₁-Fe distances of 2.60 Å and 2.62 Å, respectively, are completely different from those of the heme-pdb and the heme-model. This is possibly due to the pyramidal-like structure of the heme-hemin that reduces the steric interaction between Br and porphyrin plane. Therefore, the O-O group can be placed very close to the Fe atom. However, we decided to continue performing docking calculation of the compound “85-no-br” and the much shorter O₁-Fe distance of 2.06 Å was found. This once again confirms our steric hypothesis of the Br group. Similar to the heme-pdb and the heme-model, compound 2 still has the most occurring configuration with the much longer O₁-Fe distances of 4.87 Å.

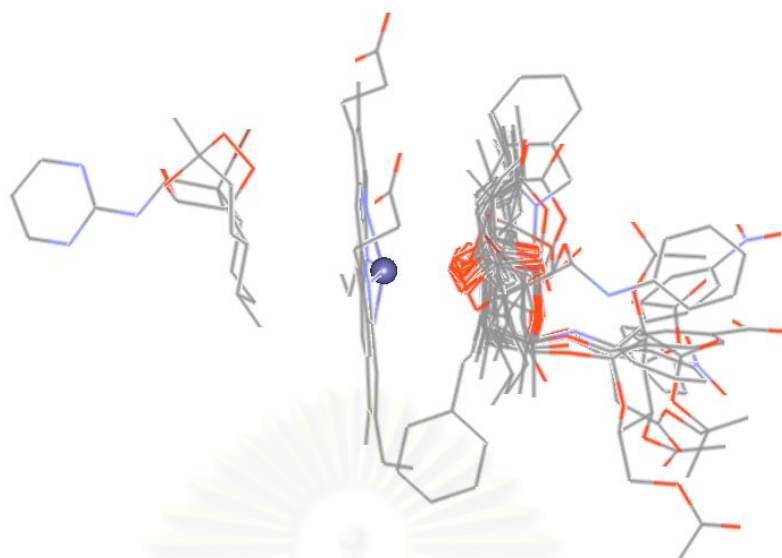


Figure 5.6 Superimposed docking configurations between the heme-hemin and 16 artemisinin derivatives (without hydrogen atoms).

Table 5.11 Docking results of heme-hemin and 16 artemisinin derivatives.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-33.13	24	2.00	2.65	4.67	4.90
2	-32.51	27	4.87	4.33	2.08	3.22
	-31.21	4	2.16	2.88	4.77	5.12
28	-36.96	32	2.05	2.95	4.72	5.09
35	-35.91	30	2.01	2.44	4.54	4.93
42	-33.27	26	1.98	2.66	4.67	4.69
43	-33.18	27	1.95	2.54	4.59	4.66
54	-33.22	31	1.97	2.66	4.66	4.71
56	-33.27	18	2.00	2.69	4.67	4.79
68	-33.33	17	2.00	2.78	4.69	4.83
69	-33.28	36	2.00	2.66	4.66	4.76
75	-30.81	21	2.17	3.02	4.85	5.02
85	-28.43	14	2.60	3.57	5.40	5.28
87	-30.22	12	6.53	6.51	4.41	6.13
	-28.09	8	2.62	3.47	5.42	5.22
88	-32.74	24	1.98	2.67	4.66	4.75

Table 5.11 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
94	-37.57	29	2.10	2.68	4.76	4.77
Deoxy	-29.64	13	-	5.17	3.71	2.16
	-30.58	11	-	2.69	4.89	5.54
85-no-br	-33.26	39	2.06	2.86	4.75	4.86

D. Heme-deoxy

Since the proximal ligand is still attached to the heme molecule, it obliges artemisinin to approach heme molecule from only one direction. Moreover, this proximal ligand pulls the Fe atom under the porphyrin plane (Figure 5.3d), thus, the most occurring configurations with longer O₁-Fe distances were resulted as shown in Table 5.12 and Figure 5.7. Because of the basin-like structure of heme-deoxy, the O₁₃-Fe or O₁₁-Fe attractions can become more preferable than the O₁-Fe and O₂-Fe attractions for some derivatives as indicated by their respective most occurring configurations. For instance, the most occurring configurations of the compound 1, 2, 85, and 88 have O₁₃-Fe as the shortest heme-artemisinin distances while the most occurring configurations of the compound 54, 56, 69, and deoxy have O₁₁-Fe as the shortest ones. Interestingly, the docking results of most compounds always contained configurations with short O₁-Fe distances. Only exception is the compound 69 where all dock configurations have long O₁-Fe distances.

From the results, the heme-deoxy is very unlikely being the target for artemisinin since their O₁-Fe and O₂-Fe distances are too long to enable the reaction of oxygen free radical formation. If, however, the heme-deoxy is the real source of iron for parasites in human body, the hemoglobin degradation must change its structural conformation or the proximal ligand must be eliminated so that artemisinin could become effective for this target.

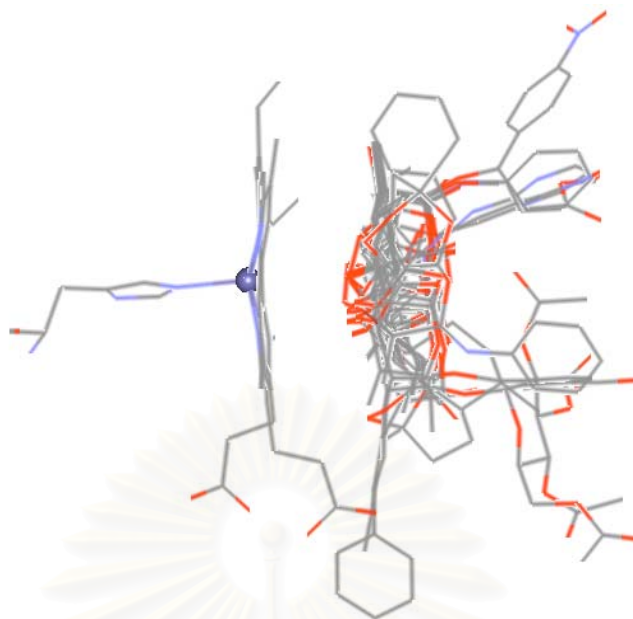


Figure 5.7 Superimposed docking configurations between the heme-deoxy and 16 artemisinin compounds (without hydrogen atoms).

Table 5.12 Docking results of heme-deoxy and 16 artemisinin derivatives.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-31.03	39	5.95	5.53	3.26	4.03
	-30.18	13	3.19	4.18	5.85	6.19
2	-30.18	30	5.84	5.33	3.22	3.59
	-29.54	3	3.28	3.77	5.61	6.26
28	-33.54	27	3.34	3.95	5.58	6.41
35	-32.34	38	3.14	3.14	5.40	5.81
42	-29.85	25	3.19	4.13	5.84	6.06
43	-30.04	33	3.14	4.08	5.82	5.99
54	-29.75	26	6.32	5.82	5.93	4.01
	-29.06	16	3.73	5.00	6.34	6.37
56	-31.86	15	7.95	7.30	7.86	6.13
	-30.71	15	3.22	3.68	5.54	6.11
68	-30.04	17	3.15	4.07	5.82	6.03
69	-28.92	21	5.90	5.95	7.10	5.07
75	-28.96	26	3.66	4.30	5.72	6.55

Table 5.12 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
85	-28.22	21	6.61	6.75	5.01	6.89
	-27.17	8	4.92	6.28	6.98	7.22
87	-28.28	17	4.83	6.18	6.90	7.17
88	-30.18	24	5.95	5.54	3.24	4.25
	-29.56	8	3.14	3.83	5.64	6.06
94	-35.05	45	3.62	4.68	6.08	6.50
Deoxy	-28.25	17	-	5.66	3.67	3.31
	-28.34	8	-	4.04	6.09	6.85
85-no-br	-30.07	27	3.23	4.16	5.83	6.09

E. Heme-oxy

The docking results between the heme-oxy and artemisinin derivatives are shown in Table 5.13 and Figure 5.8. For most artemisinin derivatives, the docks gave the most occurring configurations with O₁-Fe as the shortest heme-artemisinin distances. Similar to other heme structures, the compound 85 and 87 also have the most occurring configurations with longer O₁-Fe distances due to the steric effect from the Br group. And the distance can be shortened by removing this Br group as shown in the compound “85-no-br”.



Figure 5.8 Superimposed docking configurations between the heme-oxy and 16 artemisinin derivatives (without hydrogen atoms).

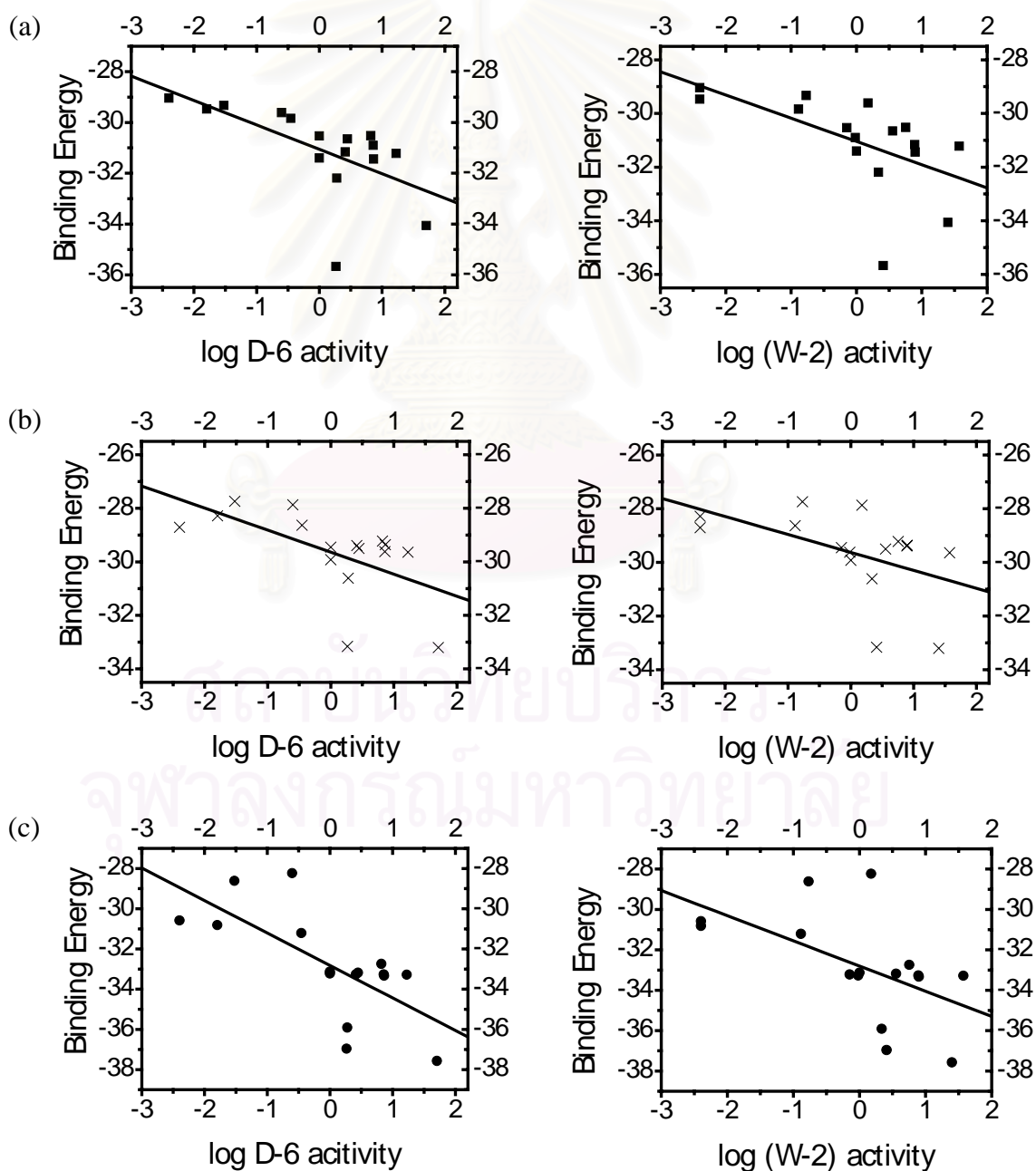
Table 5.13 Docking results of heme-oxy and 16 artemisinin derivatives.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-32.32	51	2.52	3.32	5.12	5.57
2	-31.02	37	2.76	3.56	5.16	5.79
28	-36.47	64	2.54	3.18	5.13	5.53
35	-33.14	38	2.54	2.88	5.08	5.24
42	-31.84	49	2.57	3.34	5.21	5.42
43	-31.89	50	2.56	3.36	5.19	5.44
54	-31.48	43	2.61	3.49	5.23	5.51
56	-32.14	27	2.52	3.32	5.18	5.41
68	-32.24	51	2.56	3.42	5.20	5.46
69	-29.30	26	2.62	3.37	5.17	5.52
75	-30.40	40	2.68	3.41	5.19	5.58
85	-30.04	31	4.27	5.64	6.36	6.60
	-29.99	15	3.17	4.07	5.45	6.07
87	-29.63	20	4.17	5.54	6.34	6.53
	-29.54	15	3.31	4.21	5.48	6.19
88	-31.66	47	2.53	3.34	5.22	5.37
94	-36.65	37	2.81	3.82	5.50	5.67
Deoxy	-30.23	32	-	3.15	5.11	6.21
85-no-br	-32.01	51	2.53	3.29	5.13	5.37

From all docking calculations performed (section A to E), it can be seen that the heme iron prefers to bind with artemisinin compounds at the O₁ than the O₂ of the endoperoxide linkage as indicated by the shorter O₁-Fe distances. This preference possibly comes from the more negative charge of O₁ and the steric hindrance at the position O₂.

Since the endoperoxide group is very crucial for the binding of artemisinin and heme, we supposed that there would be some relationships between antimalarial activities and 3 parameters taken from docking information, i.e., O₁-Fe distance, O₂-Fe distance, and the binding energy. Moreover, the effect of different heme structures to

the activity relationships was investigated. Using both activities in logarithm units, the linear correlation coefficients (r) were evaluated. Among the 3 parameters, the binding energy showed the best correlation with the activities having the r values in a range of -0.52 to -0.70 for the D-6 activity and -0.41 to -0.58 for the W-2 activity. The relationships were depicted in Figure 5.9. This figure clearly demonstrated that compounds that bind firmly with heme (low binding energy) have high activity. From these relationships as well as the docked O-Fe distances compared to the experimental data, the heme-hemin seems to be the best candidate for the receptor structure in the docking study. Therefore, it was selected for the docking calculations for all artemisinin derivatives.



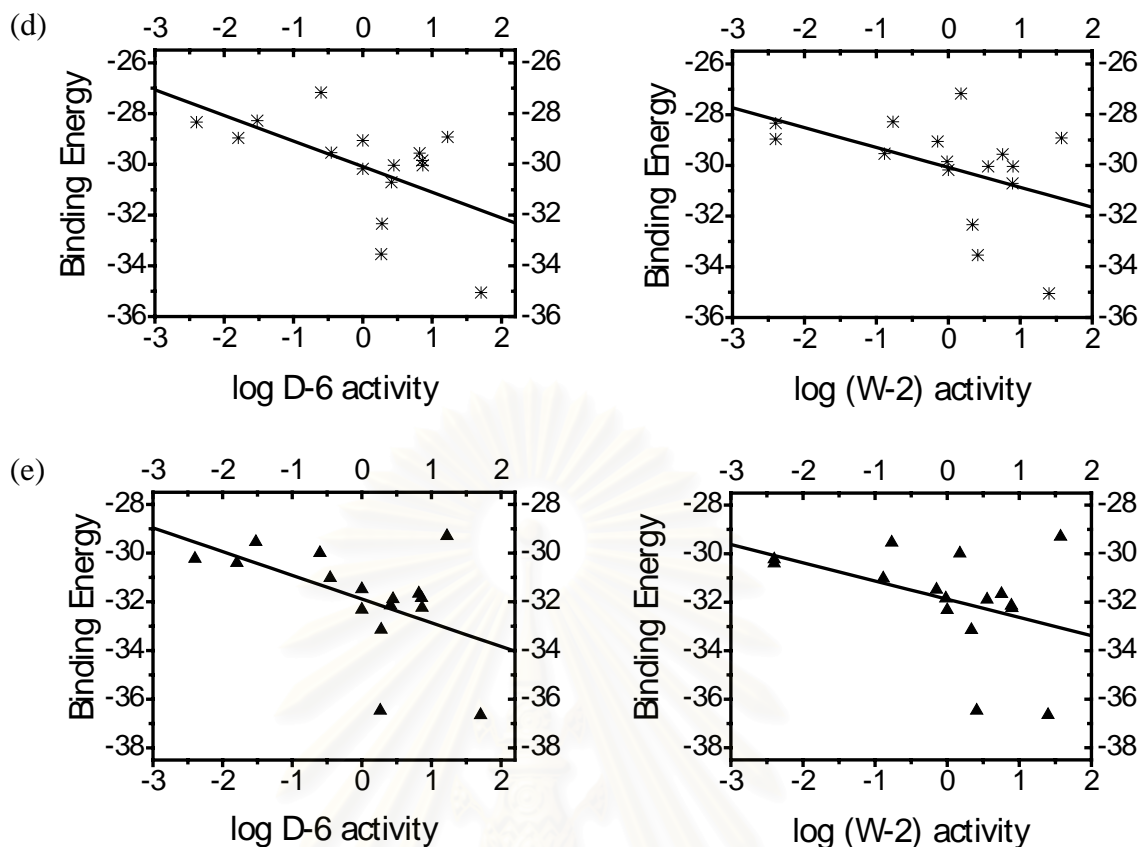


Figure 5.9 Relationships between binding energies and activities for 16 artemisinin compounds using (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, and (e) heme-oxy.

5.3.9 Comparison with Experimental Data

From the experimental data of the spectral change due to the interactions between artemisinin derivatives and ferriprotoporphyrin IX (heme), Paitayatat et al. [110] could calculate the dissociation constants (K_d) for their binding. In order to compare the binding energy from docking results with the K_d from the experimental data, we performed docking calculations between the heme-hemin and 3 artemisinin derivatives (Figure 5.10) which were used in that literature. All of these 3 compounds are well-known drugs, i.e., dihydroartemisinin (**1**), artemether (**2**), and arteether (**3**). The $\log K_d$ and binding energies were given in Table 5.14. A relationship between $\log K_d$ and binding energy was shown in Figure 5.11. The relationship implies that our docking results are consistent with the experimental data.

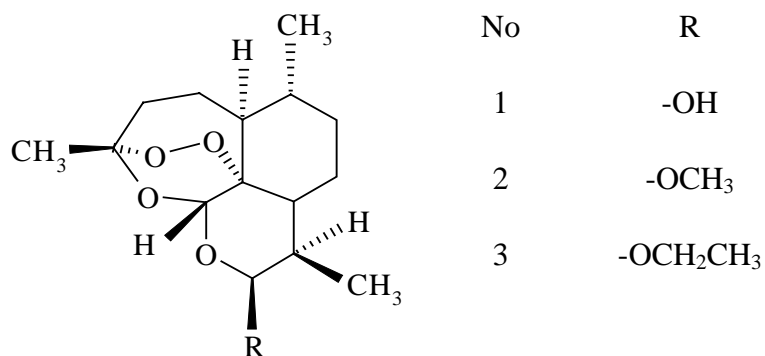


Figure 5.10 Structures of four artemisinin compounds used in Ref. [110].

Table 5.14 The dissociation constants and binding energies of 4 artemisinin derivatives.

Compound No	$\log K_d (M^{-1})$	Binding Energy (kcal/mol)
1	-4.6778	-33.27
2	-4.5686	-33.18
3	-4.4949	-33.13

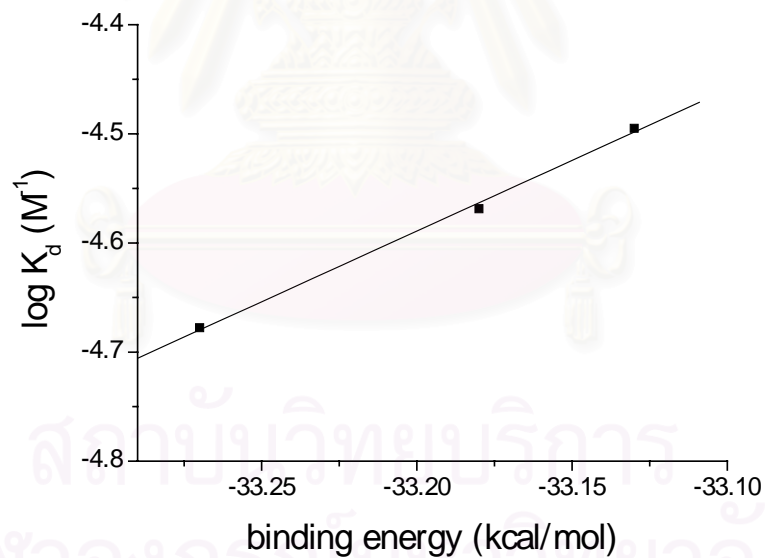


Figure 5.11 Relationship between binding energies and dissociation constants (K_d).

5.4 Docking Calculations

All suitable parameters determined from the previous section were employed for the docking calculations of all artemisinin derivatives. Since there are no specific parameters for silicon atom, compound 55 was omitted. For the comparison purpose, the inactive compound, deoxyartemisinin, was additionally included. Therefore, totally 104 compounds were studied. All artemisinin derivatives were optimized at HF/3-21G and the atomic charges were assigned by the same method. The atomic charges of the heme-hemin was calculated at the HF/6-311G** level. For expedient reason, the docking results between 104 artemisinin derivatives and heme-hemin were separated into 3 groups according to the structural similarity of artemisinin derivatives (see Chapter 2).

5.4.1 Heme-hemin and artemisinin derivatives group 1

The docking results of the heme-hemin and 41 artemisinin derivatives in group 1 as well as deoxyartemisinin were given in Table 5.15. The most occurring configurations in most artemisinin compounds have O₁ pointing toward the heme iron. In exception, compounds 2, 4, 5, 8, and 10 have the most occurring configurations with O₁₃ pointing toward the heme iron. The replacement of -CH by C=O at the C₉ position in compound 2 and the -CH₃ substituent groups at the α -C₉ position in compound 4 and 5 cause the structure of the lactone ring to change from a twisted-chair form to a nearly-envelop form. This structure facilitates the encroachment of Fe to O₁₃. However, there is a configuration with shorter O₁-Fe distance for compound 2. The iso-alkyl groups at the β -C₉ position in compounds 8 and 10 cause the steric hindrance at the endoperoxide side. Therefore, the heme iron prefers to approach at the O₁₃ side. However, there are some configurations with O₁ pointing toward heme iron but these configurations have quite long O₁-Fe distances. The replacement of the oxygen atom by a nitrogen atom at the position 11 in compounds 32 to 41 does not have significant effect to the docking results. Docking information have no relationship with antimalarial activities. For the inactive deoxyartemisinin compound, its most occurring configurations have O₁₁-Fe as the shortest distance. However, the second most occurring configurations have heme iron closed to O₂ atom. These configurations even have lower interaction energies than the most occurring ones (-30.58 kcal/mol compared to -29.64 kcal/mol).

Table 5.15 Docking results of the heme-hemin and artemisinin derivatives group 1.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
1	-33.13	24	2.00	2.65	4.67	4.90
2	-32.51	27	4.87	4.33	2.08	3.22
	-31.21	4	2.16	2.88	4.77	5.12
3	-32.03	26	2.01	2.76	4.72	4.93
4	-32.06	24	4.89	4.32	2.12	3.20
	-27.05	2	4.92	6.21	6.94	7.34
5	-33.24	38	4.85	4.31	2.08	3.22
6	-34.08	29	2.00	2.86	4.67	5.03
7	-35.21	31	2.06	2.97	4.73	5.10
8	-33.25	44	4.98	4.36	2.16	3.36
	-29.54	2	3.50	4.20	5.83	6.58
9	-36.15	35	2.01	2.89	4.68	5.04
10	-33.31	41	4.96	4.31	2.13	3.33
	-30.59	6	3.63	4.32	5.81	6.69
11	-36.60	29	1.99	2.79	4.62	5.03
12	-36.50	35	1.97	2.69	4.59	4.98
13	-36.49	24	2.20	3.21	4.99	5.13
14	-37.02	28	2.05	2.86	4.71	5.07
15	-37.85	19	2.05	2.87	4.73	5.07
16	-37.94	30	2.05	2.98	4.75	5.08
17	-37.20	33	2.16	3.22	4.92	5.12
18	-35.52	27	2.08	2.94	4.79	5.08
19	-34.25	35	2.05	2.85	4.67	5.08
20	-33.66	44	2.17	2.99	4.84	5.18
21	-34.80	43	2.10	3.06	4.88	5.03
22	-35.20	31	2.11	3.10	4.90	5.02
23	-33.78	28	1.99	2.69	4.68	4.87
24	-36.12	32	1.97	2.59	4.57	4.96

Table 5.15 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
25	-37.68	25	1.96	2.57	4.63	4.86
26	-36.81	26	2.01	2.45	4.60	4.70
27	-36.72	15	1.97	2.37	4.48	4.81
28	-36.96	32	2.05	2.95	4.72	5.09
29	-37.66	41	1.99	2.77	4.61	5.02
30	-38.44	51	1.95	2.63	4.56	4.94
31	-39.77	38	1.91	2.57	4.48	4.91
32	-32.42	42	2.00	2.61	4.64	4.99
33	-32.75	42	1.98	2.52	4.59	5.01
34	-32.88	34	1.96	2.58	4.62	4.99
35	-35.91	30	2.01	2.44	4.54	4.93
36	-32.91	40	2.00	2.51	4.61	4.97
37	-32.50	40	1.99	2.54	4.58	4.97
38	-32.81	35	1.96	2.47	4.55	4.98
39	-36.47	32	1.96	2.41	4.50	4.88
40	-32.96	40	1.95	2.47	4.55	4.96
41	-33.36	22	1.94	2.53	4.56	5.01
Deoxy	-29.64	13	-	5.17	3.71	2.16
	-30.58	11	-	2.69	4.89	5.54

5.4.2 Heme-hemin and artemisinin derivatives group 2

The docking results of the heme-hemin and 45 artemisinin derivatives in group 2 were given in Table 5.16. For compound with -OR group at the C₁₀ position (compounds 42 to 82), the most occurring configurations in almost compounds have O₁ pointing toward the heme iron. The O₁-Fe distances are in the range of 1.93 to 2.60 Å. The only one exception is compound 76. This compound has heme iron closest to the O₁₃ atom. The reason may be due to its structure. The -OH groups at both α -C₉ and α -C₁₀ positions hinder the heme iron to approach at the endoperoxide side but facilitate

the approach of the heme iron to the O₁₃ side. However, there are some configurations with shorter O₁-Fe distance. For compounds with -NR groups at the C₁₀ position (compounds 83 to 87), the Br group at the α -C₉ position causes the steric hindrance for heme iron in approaching the endoperoxide moiety. Therefore, longer O₁-Fe distances in these 5 compounds were observed. Moreover, the most occurring configurations in compounds 86 and 87 have O₁₃-Fe as the shortest distances. Nevertheless, there still be some configurations with shorter O₁-Fe distances. In order to prove the steric hindrance hypothesis once again, docking calculations of compounds “85-no-br” and “86-no-br” were performed. The most occurring configurations with very short O₁-Fe distance were obtained.

Table 5.16 Docking results of the heme-hemin and artemisinin derivatives group 2.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
42	-33.27	26	1.98	2.66	4.67	4.69
43	-33.18	27	1.95	2.54	4.59	4.66
44	-34.90	38	2.01	2.65	4.64	4.80
45	-33.15	41	1.98	2.69	4.67	4.73
46	-33.26	30	1.98	2.83	4.68	4.83
47	-33.07	29	1.99	2.78	4.67	4.84
48	-32.43	31	1.93	2.59	4.62	4.60
49	-32.20	26	2.06	2.77	4.75	4.80
50	-32.32	29	1.99	2.64	4.68	4.65
51	-32.10	29	1.99	2.61	4.65	4.73
52	-33.28	36	2.03	2.80	4.76	4.78
53	-33.33	30	2.03	2.89	4.75	4.86
54	-33.22	31	1.97	2.66	4.66	4.71
56	-33.27	18	2.00	2.69	4.67	4.79
57	-34.73	29	1.98	2.66	4.68	4.69
58	-33.17	27	1.96	2.67	4.68	4.64
59	-33.04	34	1.95	2.70	4.68	4.68
60	-35.87	26	2.05	2.79	4.70	4.89

Table 5.16 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
61	-33.13	28	2.01	2.70	4.71	4.72
62	-33.27	30	1.99	2.72	4.67	4.79
63	-34.64	26	2.03	2.55	4.64	4.69
64	-33.30	22	2.17	2.83	4.74	5.03
65	-33.55	28	2.01	2.69	4.68	4.79
66	-33.83	36	2.04	2.71	4.68	4.84
67	-33.55	28	2.08	2.83	4.76	4.89
68	-33.33	17	2.00	2.78	4.69	4.83
69	-33.28	36	2.00	2.66	4.66	4.76
70	-33.76	24	2.12	2.69	4.70	4.90
71	-33.20	31	1.95	2.67	4.66	4.67
72	-33.83	26	1.99	2.60	4.68	4.56
73	-33.76	21	1.98	2.56	4.62	4.68
74	-34.07	29	2.07	2.75	4.68	4.91
75	-30.81	21	2.17	3.02	4.85	5.02
76	-34.04	44	4.89	4.34	2.09	3.38
	-31.73	3	2.28	3.04	4.83	5.12
77	-34.09	23	2.08	2.77	4.73	4.87
78	-34.01	26	2.06	2.64	4.66	4.83
79	-32.29	22	2.31	3.07	4.77	5.19
80	-30.90	37	2.20	3.05	4.87	5.10
81	-32.49	20	2.34	3.08	4.78	5.20
82	-30.72	28	2.27	3.14	4.97	5.14
83	-28.63	31	4.97	6.33	7.03	7.33
	-28.97	7	2.63	3.73	5.28	5.48
84	-28.47	16	2.41	3.23	5.04	5.27
85	-28.43	14	2.60	3.57	5.40	5.28
85-no-br	-33.26	39	2.06	2.86	4.75	4.86

Table 5.16 (Continued)

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
86	-31.64	12	5.21	4.64	2.40	3.67
	-28.45	9	2.73	3.76	5.47	5.55
86-no-br	-33.38	35	2.04	2.74	4.71	4.80
87	-30.22	12	6.53	6.51	4.41	6.13
	-28.09	8	2.62	3.47	5.42	5.22

The correlations between antimalarial activities and properties from docking calculations of compounds 42 to 82 were investigated. Compound 83 to 87 were excluded from the correlation study because of their distinct docking properties. The binding energy has significant relation with both D-6 and W-2 activities in logarithm unit as indicated by the *r* values of -0.580 and -0.579, respectively (Figure 5.12).

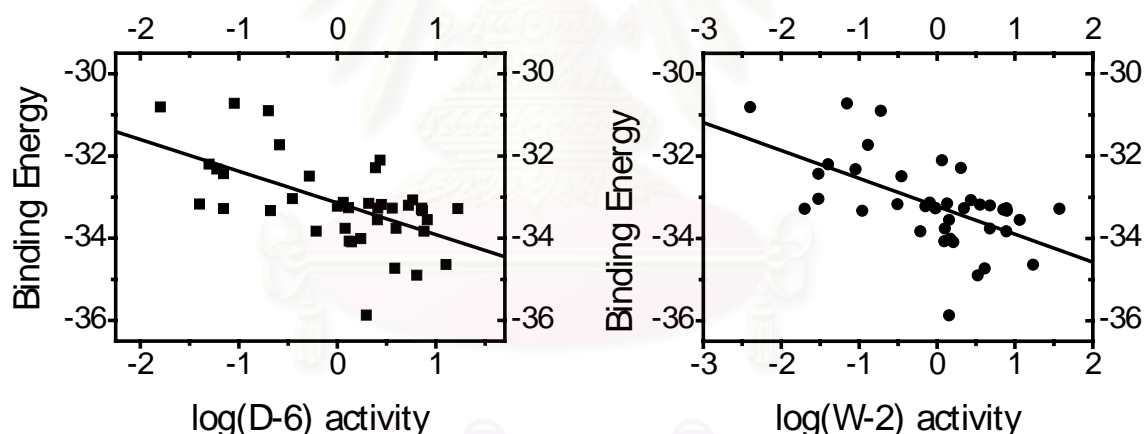


Figure 5.12 Relationship between binding energies and activities for 40 artemisinin derivatives in group 2.

5.4.3 Heme-hemin and artemisinin derivatives group 3

Compounds in this group have no substituent at the C₁₀ position but the docking results are very similar to the other two groups (Table 5.17). Therefore, the substituent groups at the C₁₀ position tend to have no effect on the binding characteristic to its receptor. The evidence of high antimalarial activities of some compounds in this group support the hypothesis. The most occurring configurations in all 17 compounds have O₁

pointing toward the heme iron with the O₁-Fe distances of between 1.89 to 3.05 Å. Only the compound 102 exhibited different results from the others by having longer O₁-Fe distance. This is possible because of the steric hindrance of *p*-ClC₆H₄(CH₂)₃ substituent at the C₃ position.

Table 5.17 Docking results of heme-hemin and artemisinin derivatives group 3.

Compound No.	Energy (kcal/mol)	% Occur	O ₁ -Fe distance (Å)	O ₂ -Fe distance (Å)	O ₁₃ -Fe distance (Å)	O ₁₁ -Fe distance (Å)
88	-32.74	24	1.98	2.67	4.66	4.75
89	-32.27	38	1.89	2.42	4.50	4.58
90	-33.98	22	1.95	2.68	4.58	4.82
91	-35.35	37	2.02	2.81	4.69	4.88
92	-36.01	31	1.98	2.71	4.65	4.81
93	-36.56	29	2.01	2.70	4.68	4.80
94	-37.57	29	2.10	2.68	4.76	4.77
95	-37.85	32	2.11	2.76	4.82	4.78
96	-33.34	19	1.90	2.48	4.52	4.65
	-32.79	19	1.91	2.67	4.66	4.62
97	-34.30	42	1.95	2.50	4.57	4.64
98	-34.90	45	1.93	2.41	4.51	4.60
99	-35.75	41	1.94	2.44	4.52	4.66
100	-35.84	16	1.95	2.52	4.60	4.61
101	-36.41	28	1.98	2.36	4.54	4.40
102	-32.12	16	3.05	3.30	5.52	5.60
103	-35.30	26	1.95	2.41	4.51	4.63
104	-35.02	35	1.95	2.47	4.55	4.64

As in the other two groups, the relationships between docking parameters and antimalarial activities were investigated. When all 17 compounds were used, no significant relationship was found. However, if the compound 102 was omitted, relationships between antimalarial activities (both D-6 and W-2) and binding energies (Figure 5.13) were observed.

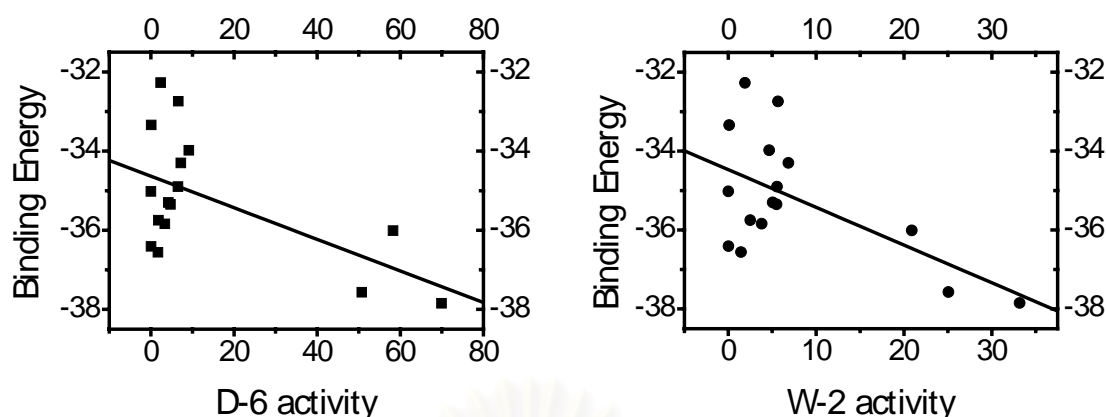


Figure 5.13 Relationship between binding energies and activities for 16 artemisinin derivatives in group 3.

5.5 Summary

The automated molecular docking method using the simulated annealing Monte Carlo simulations was employed to investigate the binding between heme and 104 artemisinin compounds and finally their relation to the antimalarial activities. The parameters affecting the docking results were also investigated. For the simulated annealing calculations, the high initial temperature, the very low final temperature (below 100 K), and the temperature reduction rate of 0.90 were suggested. The small grid spacing and the moderate grid size are recommended for the grid-based energy evaluations. The tolerance value of 1.0 Å should be employed for the cluster analysis to allow detail information of configurations. The distance-dependent dielectric function should be used for applying the solvent effect in the simulation.

We also observed that the docking results depend on the structures and atomic charges of both artemisinin and heme. From our study, it appears that the HF/3-21G level of theory is suitable for the geometry optimization of artemisinin and its derivatives and the docking configurations were significantly affected by the atomic charges of heme and, to a much lesser extent, of artemisinin. The high quality atomic charges, HF/6-311G**, are recommended for the electrostatic potential of heme. The heme structures with no or little steric hindrance at the Fe position, i.e., heme-pdb, heme-hemin, and heme-oxy, facilitates the binding of heme and endoperoxide oxygens, and they are recommended for using in docking calculation. Comparing between the docking results of heme-deoxy and heme-oxy, the heme-oxy structure, whose structure is very close to the receptor structure in the bound state, gave docking results that are in

agreement with those of other heme structures. And from the relationships between docking properties and activities of 16 artemisinin compounds, the heme-hemin was selected for the docking study of all artemisinin derivatives.

From the automated docking results of all 104 artemisinin derivatives, artemisinin compounds interact with heme molecule mainly at the endoperoxide linkage. The heme iron approaches O_1 more preferably than O_2 . Artemisinin compounds that have large substituent group on the same side of endoperoxide group, i.e., at the α - C_9 position, will hinder the approach of heme iron to the endoperoxide moiety such as in compounds 8, 10, 76, and 83 to 87. If such a substituent group is removed, the heme iron can be placed close to the endoperoxide as seen in compound "85-no-br". The steric hindrance was also observed in artemisinin compounds with substituent groups at the C_9 position which cause the conformation change in the lactone ring, such as compound 2, 4, and 5. The distances of each 4 oxygen atoms to heme iron from docking configuration and the binding energy were used to find the correlation with the antimalarial activities. Significant relationships between antimalarial activities and binding energies were found for compounds in group 2 and 3. From our docking results, we would propose that the binding between artemisinin compounds and heme is mainly controlled by the steric effect. Therefore, in order to design a more effective antimalarial drug, the compound should have no steric hindrance at the same side of the endoperoxide moiety, e.g., at the α - C_9 position. Moreover, a substituent group at the C_{10} position is not required because it does not affect the binding characteristic. Since only the dock information is not sufficient to explain the activities of all compounds, additional information from other methods should be considered together (see Chapter 7).

Our calculations show somewhat different docking data from the previous work [99]. In their results, the O_2 -Fe distance is shorter than the O_1 -Fe distance and the O_{14} plays an important role in the binding. In our calculations, the O_{14} is far away from the heme iron (more than 6 Å); hence it could not play any role for the binding. The evidence of compounds in group 3, which have no O_{14} atom but have higher antimalarial activities than artemisinin, supports our predictions. Also, the previous results of deoxyartemisinin are in contrast to ours. They reported that iron attaches preferably O_{11} and O_{13} whereas the preferable position in our calculations is O_2 . Moreover, the energy of their docking configuration when O_2 is close to iron is higher

than when O_1 is close to iron. The main reasons for these differences could possibly be the lack of specific parameters for iron, the low quality of atomic charges, the low accuracy level for geometry optimization, and the manual docking method in the previous docking calculations.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 6

Mechanism of Action

6.1 Introduction

The mechanism of action for artemisinin compounds still remains uncertain. However, it is generally accepted that the endoperoxide group in artemisinin compounds plays an important role in its biological activity. The evidence is that deoxyartemisinin analogs, which lack the endoperoxide moiety, are devoid of antimalarial activity [31, 46]. The endoperoxide moiety is believed to produce free radicals which are necessary for mediating the effects. One evidence that supports the importance of free radicals is the enhancement of antimalarial activity by oxidant agents (free-radical-generating compounds) and the retard of antimalarial activity by antioxidants (free-radical scavengers) [56]. In addition to the endoperoxide group, some amount of iron is required as suggested from *in vitro* experiments [111-113]. The mechanism of action was proposed to involve two sequential steps, i.e., activation and alkylation steps [56].

A. Activation Step

In the activation step, the endoperoxide linkage is attacked by ferrous ion, Fe (II), to produce reactive species, such as oxygen free radicals and hydroperoxide compound [114-116]. The specific source of iron for this activation step is still uncertain. Malarial parasites, which live in the red blood cell, an extremely Fe rich source, digest up to 80% of hemoglobin contained in the host cell to give heme (Figure 6.1) and globin as products. Globin is then hydrolyzed to provide amino acids as their sources for protein synthesis. Heme portion is normally discarded but owing to its toxicity, it is mostly detoxified by polymerization process to hemozoin. However, some free heme may be transiently present. Therefore, it is very probable that the free iron may be released from hemoglobin heme [117]. It was proposed that the reaction between artemisinin and free heme could intercede the hemoglobin degradation and the heme detoxification processes which may cause the parasites to die [118]. This idea is supported by the observation that the chloroquine-resistant strains of *P. berghei* which lack hemozoin are extremely resistant to artemisinin [119].

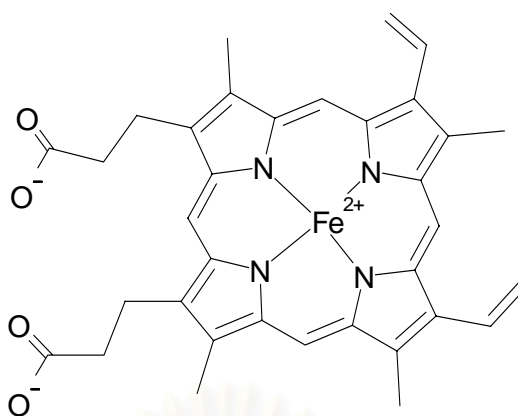


Figure 6.1 The structure of heme.

There are three proposed mechanisms for the approaching of heme iron to the endoperoxide linkage of artemisinin compounds. Posner and co-workers [115] proposed that the iron attacks the compound at the O₂ position and produces a free radical at the O₁ position (1A) (see Figure 6.2). It is then rearranged to form a C₄ free radical (1B) via the intermolecular 1,5-hydrogen shift process. This radical (1B) was suggested to be an important substance for antimalarial activity [120]. The supporting evidence is that trioxanes compound having substituent group at the α-C₄ position, which makes it difficult or even impossible to produce such a radical, is virtually devoid of activity [121].

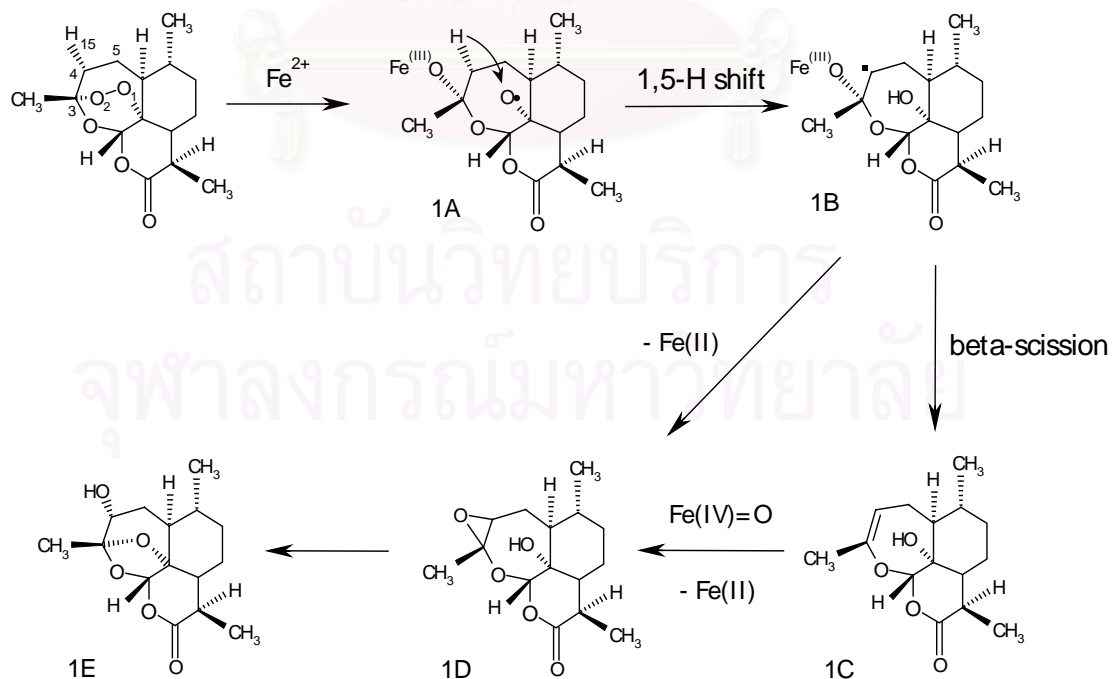


Figure 6.2 The mechanism of action of artemisinin compound as proposed by Posner et al. (pathway 1) [115].

The radical 1B is then changed to the vinyl ether 1C by a beta-scission reaction, which also generates the Fe(IV)=O as another product. Subsequently, the intermolecular reaction between compound 1C and Fe(IV)=O leads to an epoxide compound 1D. Alternatively, the epoxide 1D could be derived directly from the radical 1B by a direct expulsion of iron. Finally, a C₄-hydroxylated product 1E is formed. The compound 1D is able to alkylate the specific proteins of the malarial parasites and possibly causes damage to the parasites [122].

On the other hand, Jefford and co-workers [116] believed that the iron attacks the compound at the O₁ position and produces a free radical at the O₂ position (2A) (Figure 6.3). After that the homolytic C₃-C₄ bond cleavage is occurred giving a carbon radical at the C₄ (2B). This radical (2B) could also be very harmful to the parasites [116] in the similar way to the 1D. In the last step, expulsion of iron leads to tetrahydrofuran 2C.

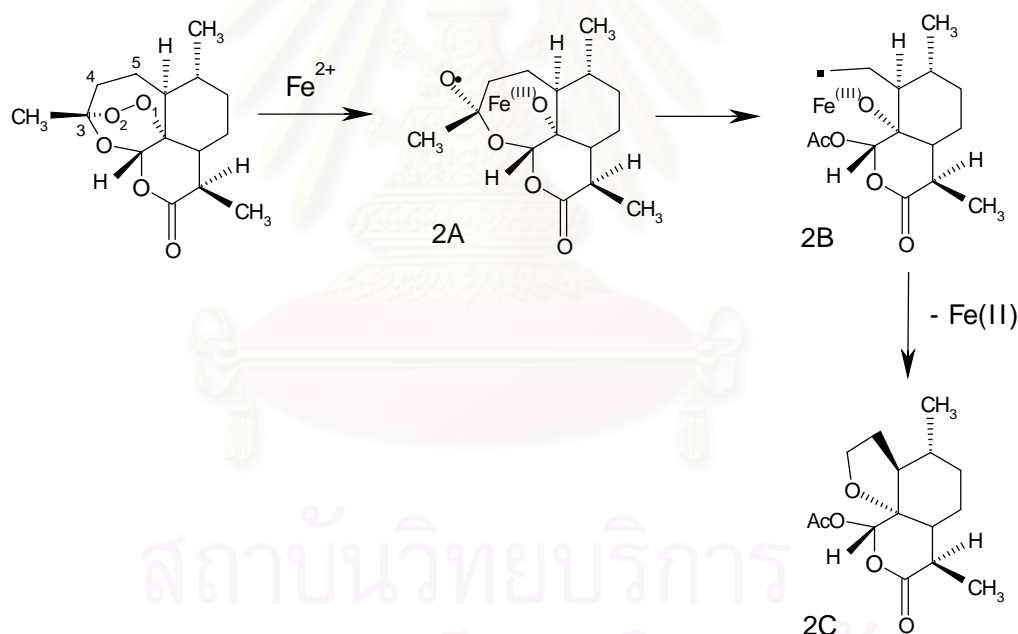


Figure 6.3 The mechanism of action of artemisinin compound as proposed by Jefford et al. (pathway 2) [116].

Haynes and Vonwiller presented a different hypothesis [123]. Instead of the O-O bond breaking, a heterolytic bond cleavage between O₂-C₃ by iron yielding the iron-oxo olefin (3A) was proposed as the first step in the mechanism of action (Figure 6.4). The loss of iron via protolysis reaction takes place producing the hydroperoxide 3B. The lactone 3C and diketo acid 3D are then resulted as the final products. Since

hydrogen peroxide is known to have antimalarial activity, the compound 3B was postulated to be the active species. However, this mechanism is still in doubt. Because the formation of free radical compound was strongly evident [124-125] but no free radical was proposed in this mechanism.

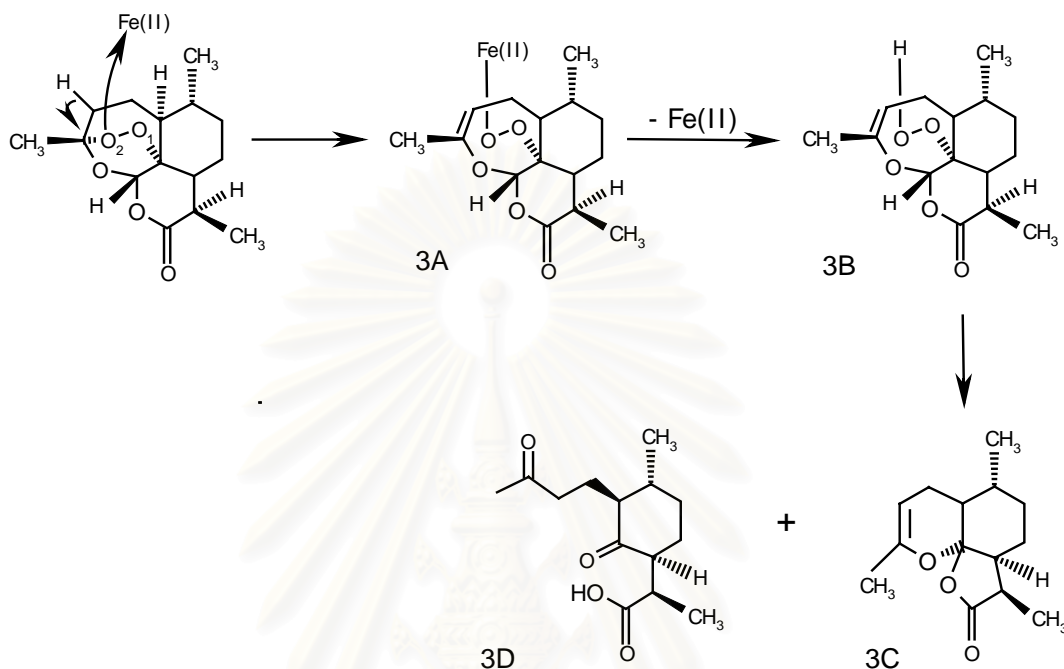


Figure 6.4 The mechanism of action of artemisinin compound as proposed by Haynes et al. (pathway 3) [123].

B. Alkylation Step

In the alkylation step, radicals and reactive intermediates formed from the activation step would rapidly react with nearby molecules due to their high reactivity. Alkylations of proteins [122, 126-127] and heme [128-129] but not of DNA [127] have been reported. The alkylation reaction was observed with many proteins, including human serum albumin, glycoprotein, hemoproteins, catalase, and cytochrome *c*. But most importantly, artemisinin and its derivatives were found to alkylate specific malaria proteins which are 25, 32, 42, 50, 65, and > 200 kDa in size via covalent linkage [122]. It seems that the heme alkylation does not play any significant role in the mode of action since the performed adduct(s) of artemisinin and heme have almost no antimalarial activity [128].

The understanding in the mechanism of action of drugs is definitely very crucial for the design of new more effective compounds. Therefore, investigations on the mechanism of action of artemisinin derivatives were carried out.

6.2 Computational Methods

According to the size of the artemisinin molecules, the Hartree-Fock (HF) method seems to be an appropriate choice for the study. However, from our study [130], we found that the HF with 3-21G basis set was not sufficient to predict energy of artemisinin compound with free radical. Instead, Density Functional Theory (DFT) with B3LYP functional was recommended. However, the use of DFT method on a large set of artemisinin compounds is not applicable due to the expensive computational costs. Therefore, the Integrated Molecular Orbital and Molecular Orbital (IMOMO) approach [131] was selected.

6.2.1 Theoretical Background on the IMOMO Approach

The IMOMO approach can markedly reduce the computational costs. The strategy is to partition the molecular system into 2 parts, “model” and “real” parts, in which different levels of molecular orbital theory are applied. The “model” part represents important section of the system and thus it is treated with more accurate method (high level). Whereas the “real” part, which is the rest of system, is treated with less accurate method (low level).

In order to calculate the energy of the entire system (denoted as “all” part), the extrapolation method as shown in Figure 6.5 is employed. It is assumed that the energy difference (ΔE) between high and low levels in “all” and in “model” parts is equal, i.e., $\Delta E(\text{all}) = \Delta E(\text{model})$. Therefore, the energy of the compound at high level, $E_{\text{high}}(\text{all})$, could be obtained by using equation (6.1).

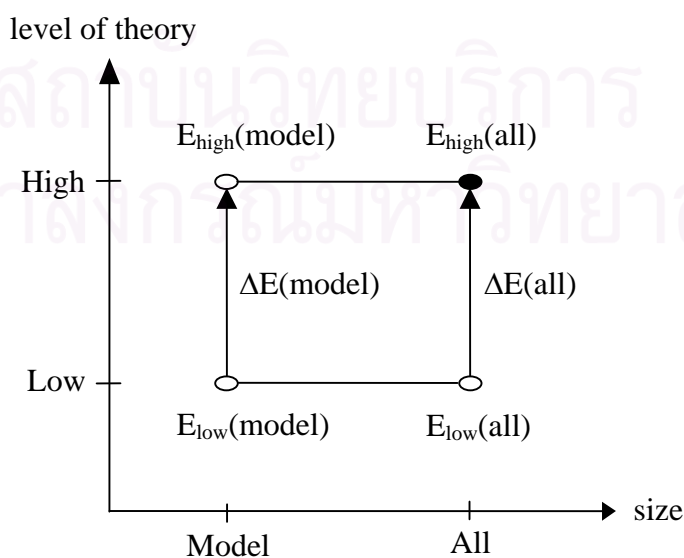


Figure 6.5 The extrapolation scheme of IMOMO method.

$$\begin{aligned}
 E_{\text{high}}(\text{all}) &= E_{\text{low}}(\text{all}) + \Delta E(\text{all}) \\
 &= E_{\text{low}}(\text{all}) + \Delta E(\text{model}) \\
 &= E_{\text{low}}(\text{all}) + [E_{\text{high}}(\text{model}) - E_{\text{low}}(\text{model})] \quad \dots\dots(6.1)
 \end{aligned}$$

where $E_{\text{low}}(\text{all})$ is the energy of the entire system calculated at the low level of theory.

$E_{\text{high}}(\text{model})$ is the energy of the model part calculated at the high level of theory.

$E_{\text{low}}(\text{model})$ is the energy of the model part calculated at the low level of theory.

6.2.2 Calculations Details

All atoms in the endoperoxide ring together with the O₁₁ atom (as labeled in Figure 6.6), which involve in the free radical formation, were assigned as the model part and they were treated at the highly accurate B3LYP/6-31G** level. While the real part was treated at HF/3-21G, which is sufficient to reproduce most of structural parameters of artemisinin (see section 2.4). Therefore, the method was denoted as the IMOMO(B3LYP/6-31G**:^{HF}/3-21G). All quantum chemical calculations were performed using the Gaussian 98 [132].

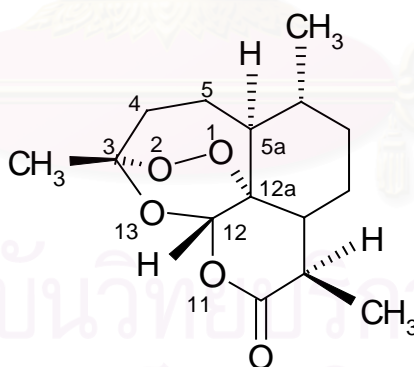


Figure 6.6 Labeled atoms that are used as the model part for IMOMO calculation.

6.3 Results and Discussions

6.3.1 Comparison of Structural Data between the IMOMO Optimized Structure and the X-ray Structure

For the purpose of elucidating the quality of the IMOMO method, artemisinin structure was fully optimized at HF/3-21G, B3LYP/6-31** and the IMOMO(B3LYP/6-

31G**: $\text{HF}/3\text{-}21\text{G}$). Subsequently, they were compared to the X-ray structure [33]. Important structural parameters which involve in the reaction mechanism of artemisinin are given in Table 6.1. It can be seen that the geometries obtained from the IMOMO method are in better agreement with the X-ray data than those from $\text{HF}/3\text{-}21\text{G}$ and gave the same accuracy as those from $\text{B3LYP}/6\text{-}31\text{G}^{**}$. For example, the $\text{O}_2\text{-C}_3$ bond length of 1.411 Å was obtained from the IMOMO while of 1.441 Å from $\text{HF}/3\text{-}21\text{G}$ and of 1.414 Å from $\text{B3LYP}/6\text{-}31\text{G}^{**}$ compared to 1.418 Å from the X-ray data. Moreover, some of the IMOMO structural parameters are in better agreement with the X-ray data than those of $\text{B3LYP}/6\text{-}31\text{G}^{**}$, e.g., $\text{O}_2\text{-C}_3\text{-C}_4$ bond angle (113.1° for X-ray, 112.3° for IMOMO, and 111.9° for $\text{B3LYP}/6\text{-}31\text{G}^{**}$). Therefore, the IMOMO method seems to be a suitable choice for this study.

Table 6.1 Comparison of some important structural parameters between the X-ray structure and optimized structures of artemisinin.

Parameters	X-ray	$\text{HF}/3\text{-}21\text{G}$	$\text{B3LYP}/6\text{-}31\text{G}^{**}$	IMOMO
Bond Distance (Å)				
$\text{O}_1\text{-O}_2$	1.474	1.462	1.460	1.461
$\text{O}_2\text{-C}_3$	1.418	1.441	1.414	1.411
$\text{O}_1\text{-C}_{12a}$	1.450	1.477	1.455	1.451
$\text{C}_3\text{-C}_4$	1.517	1.537	1.546	1.543
Bond Angle ($^\circ$)				
$\text{O}_1\text{-O}_2\text{-C}_3$	107.7	107.1	108.3	108.4
$\text{O}_2\text{-O}_1\text{-C}_{12a}$	111.6	111.3	111.6	112.1
$\text{O}_2\text{-C}_3\text{-C}_4$	113.1	111.0	111.9	112.3
$\text{O}_1\text{-C}_{12a}\text{-C}_{5a}$	106.4	105.1	105.8	106.7
$\text{C}_3\text{-C}_4\text{-C}_5$	114.6	112.9	114.1	114.2
Torsion Angle ($^\circ$)				
$\text{C}_3\text{-O}_2\text{-O}_1\text{-C}_{12a}$	47.7	50.3	47.9	45.3
$\text{O}_1\text{-O}_2\text{-C}_3\text{-C}_4$	46.3	45.0	46.8	48.4
$\text{O}_2\text{-O}_1\text{-C}_{12a}\text{-C}_{5a}$	251.7	249.1	250.8	251.9
$\text{O}_2\text{-C}_3\text{-C}_4\text{-C}_5$	265.8	263.3	265.4	265.7
$\text{O}_1\text{-C}_{12a}\text{-C}_{5a}\text{-C}_5$	69.0	68.8	68.5	69.6

6.3.2 Comparison of All Three Pathways

From the three proposed mechanisms, pathways 1 (Figure 6.2) and 2 (Figure 6.3) involve O-O bond breaking while pathway 3 (Figure 6.4) involves C-O bond breaking. Generally, the energy for the C-O bond breaking is higher than that for the O-O bond breaking. Therefore, the pathway 3 would have low probability to occur. In order to verify this hypothesis, the difficulty of bond breaking process in three pathways of artemisinin compound was measured from the energy difference (ΔE) between artemisinin and first intermediates in each pathway (1A, 2A, and 3A), i.e., $\Delta E = E_{\text{intermediate}} - E_{\text{artemisinin}}$. The energies were calculated using the IMOMO(B3LYP/6-31G**: $\text{HF}/3\text{-}21\text{G}$) method. Since the compound 3A has one electron less than the compounds 1A and 2A, the energy of one electron (-0.5 a.u.) was added into the absolute energy of 3A for the calculation of ΔE . From the results in Table 6.2, the intermediate 3A has much higher absolute energy than the other two intermediates and hence lower ΔE . Therefore, the pathway 3 appears to be an inferior process compared to the other two pathways and it was disregarded in our further studies.

Table 6.2 Comparison of energy difference (ΔE) of 3 intermediates in the mechanism of action of artemisinin.

Compound	Energy (a.u.)	ΔE (kcal/mol)
Artemisinin	-955.83019	0.00
1A	-956.44883	-388.20
2A	-956.45297	-390.79
3A	-955.79208	-289.84

6.3.3 Direct O-O Bond Breaking

From the above section, it is clear that the O-O bond breaking process is the crucial step in the mechanism of action. And this process was believed to occur via a complexation with Fe^{2+} . However, in many organic synthesis reactions, bicyclic endoperoxide compounds usually follow the homolysis of O-O bond to form a diradical as an intermediate [133-136]. Therefore, it may be possible that the endoperoxide bond in artemisinin compounds is broken to form the diradical compound before attacking Fe^{2+} (direct bond breaking). As a result, it is quite interesting to investigate the possibility for this alternative route.

The electronic state of this diradical intermediate can be described by either an open-shell singlet or triplet wave function. But the triplet state would normally be energetically favorable, and therefore is a good choice for representing the electronic state of the diradical intermediate. As a result, the O-O bond breaking process could be possibly explained by the energy gap between the triplet state and the singlet ground state (closed-shell singlet). And this may also be capable in describing the biological activities of artemisinin compounds.

At the time we performed these investigations on the singlet-triplet state of artemisinin compounds, the IMOMO method is not available for us. Therefore, a pure B3LYP with 6-31G* basis set was used for the calculations using the Gaussian 94 program packages [58]. Totally 11 artemisinin derivatives with high and low activities were randomly selected.

The singlet energy (E_S), triplet energy (E_T), and singlet-triplet energy gap (ΔE_{T-S}), which was calculated by $E_T - E_S$, together with D-6 and W-2 activities of all 11 compounds are shown in Table 6.3. For all compounds, the triplet state (diradical compound) has much higher energy than the singlet state (reactant compound) with the energy gap of between 29.48 to 31.24 kcal/mol. This implies that the diradical formation is rather difficult and would require some facilitation. Therefore, the complexation with Fe^{2+} appears to be a dominant route for the O-O bond breaking. The calculated energy gaps were not related to their antimalarial activities.

Table 6.3 Energy of Singlet and Triplet states of Artemisinin derivatives optimized at the B3LYP/6-31G* level.

No.	D-6 Activity	W-2 Activity	Singlet Energy (a.u.)	Triplet Energy (a.u.)	ΔE_{T-S} (kcal/mol)
1	1.0	1.0	-960.88937	-960.83959	31.24
3	1.71	6.18	-921.57597	-921.52680	30.85
20	0.88	1.12	-960.88887	-960.84033	30.46
32	7.2	0.96	-962.07610	-962.02907	29.51
33	2.80	3.57	-1001.38404	-1001.33679	29.51
65	0.016	0.004	-960.84125	-960.79432	29.45
78	2.13	5.00	-941.02545	-940.97603	31.01

Table 6.3 (Continued)

No.	D-6 Activity	W-2 Activity	Singlet Energy (a.u.)	Triplet Energy (a.u.)	ΔE_{T-S} (kcal/mol)
88	6.59	5.67	-886.85468	-886.80735	29.70
92	58.26	20.90	-1004.79418	-1004.74717	29.49
96	0.10	0.10	-886.85529	-886.80831	29.48
104	0.0009	0.0009	-1075.42056	-1075.37282	29.96

6.3.4 Reaction Mechanism

In pathways 1 and 2, the C_4 free radicals are believed to be the active species responsible for antimalarial activity. The understanding of the thermodynamics and kinetics of the reactions leading to the formation of the C_4 free radicals should be very helpful in the design of more potent antimalarial drugs. Therefore, the investigations on these two pathways were carried out. In human body the C_4 free radical formation reaction takes place in the proton rich environment (aqueous solution). Thus, the iron (III) ion which is attached to the oxygen atom could possibly be replaced by the proton in the solution. Intuitively, it is inspired to use this replacement as the model for the free radicals formed during the activation step in our calculations as in other theoretical studies [137-138].

It is also of interest to identify whether pathway 1 or 2 is more preferable (Figure 6.7). Moreover, there is still a debate regarding the intramolecular 1,5-H shift process in pathway 1 [116]. For example, the predicted atomic distance between the transferred H atom (H_{15}) and the O_1 radical is longer than the critical distance of 2.1 Å, which was proposed as the maximum allowable distance for the occurring of hydrogen transfer process. Also, the energy barrier in this process is too high. In an attempt to answer these questions, Gu et al. [137-138] had performed theoretical calculations on 6,7,8-trioxybicyclo[3,2,2]nonane, a model system for artemisinin. Based on B3LYP/6-31G* calculations [137], the carbon free radical analogous to 2B was predicted to be more stable than those analogous to 1B radical. The energy differences between the oxygen and carbon free radicals for pathway 1 and 2 were reported to be 2.7 and 12.1 kcal/mol, respectively. Hence, the homolytic C-C bond cleavage (pathway 2) is more

preferable from the energetic point of view. Since the transition state was not included in that study, the information about the activation energy was not available.

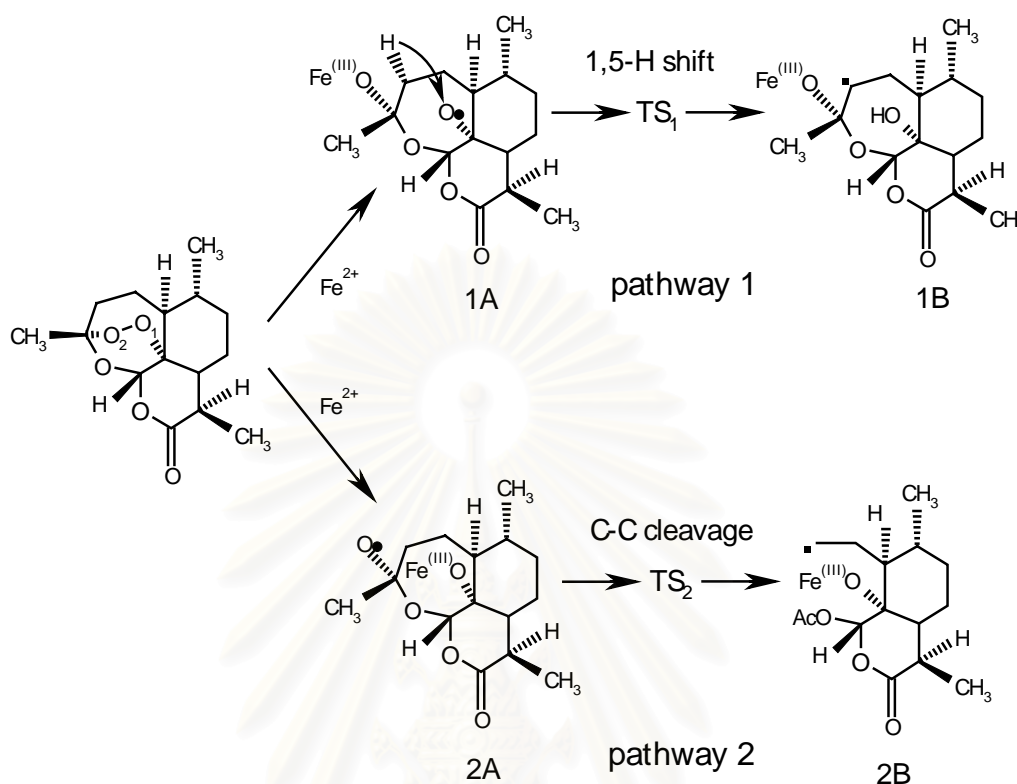


Figure 6.7 Proposed mechanism of action of artemisinin compound.

Recently, B3LYP/6-31G** calculations on the transition state of the 1,5-H shift process for the 6,7,8-trioxybicyclo[3,2,2]nonane were reported [138]. A low value of activation energy for the intramolecular 1,5-H shift process (pathway 1) was obtained (6.4 kcal/mol). Also, the critical distance between the transferred hydrogen and the receptor oxygen atom was predicted to be 2.34 Å, indicating that this distance could be larger than 2.1 Å. However, there is a certain difference between structures of 6,7,8-trioxybicyclo[3,2,2]nonane and artemisinin. The latter has more ring strain due to two additional rings attached which could affect the stability of the (oxygen and carbon) free radicals and the transition states. Thus, the previous prediction could be reversed. Therefore, it is our intention to perform theoretical calculations on the C_4 free radical formation pathways for the real system (artemisinin) instead of the model system (6,7,8-trioxybicyclo[3,2,2]nonane). Furthermore, to establish whether there is any relation between the antimalarial activity of artemisinin and the ease (or the difficulty) of forming the C_4 free radical, theoretical calculations were performed for 15 artemisinin derivatives. These compounds were classified into 2 group according to

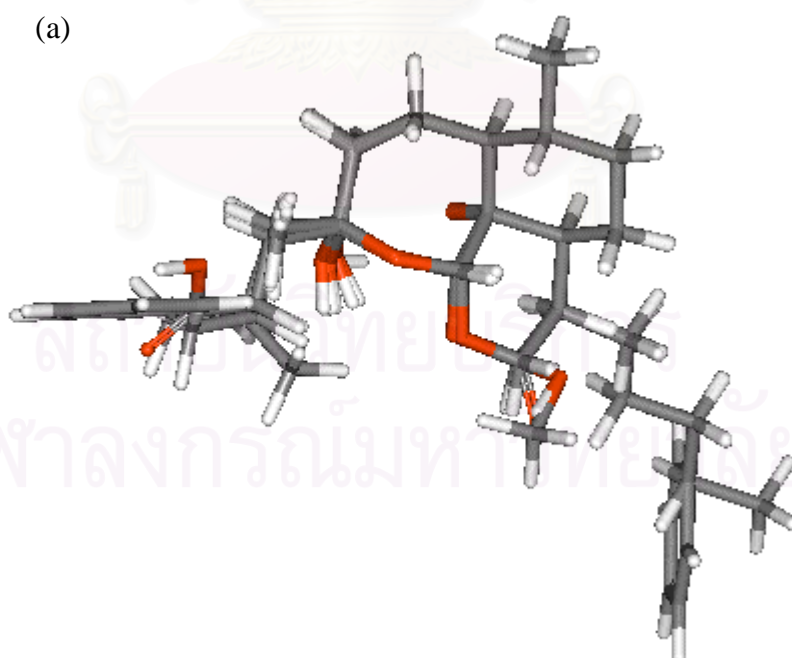
their antimalarial activities. Compounds which are more potent than artemisinin, i.e., $\log(\text{activity}) > 0.0$, were denoted as “high activity compounds” whereas compounds which are less active than artemisinin, i.e., $\log(\text{activity}) < 0.0$, were denoted as “low activity compounds”.

Structures of 15 artemisinin derivatives, their corresponding free radicals (1A, 1B, 2A, and 2B), and their transition states (TS1 and TS2) were fully optimized using the IMOMO(B3LYP/6-31G**):HF/3-21G) method. Furthermore, to confirm the optimized transition state structures, additional frequency calculations were carried out.

6.3.4.1 Intramolecular 1,5-Hydrogen Shift (Pathway 1)

A. Structural Information

The optimized structures of the O-centered (1A), the C-centered (1B) radicals, and the transition state (TS₁) for all 15 artemisinin compounds are illustrated in Figure 6.8. Some important structural parameters in radical 1A, transition state TS₁, and radical 1B of 15 artemisinin compounds and the model system are given in Table 6.4. Since some structural data of the model system were not reported in reference [138], this compound was recalculated and its structural data were used in this work.



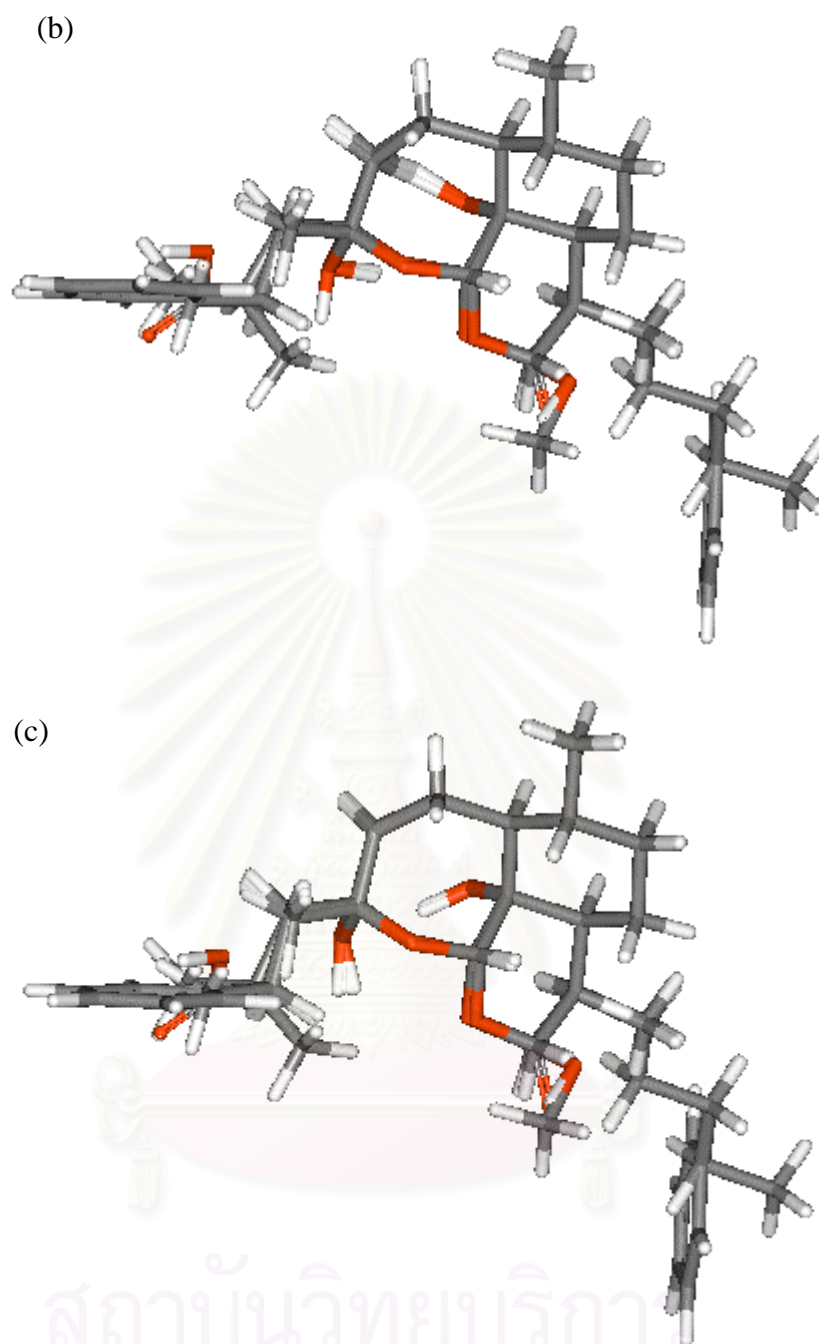


Figure 6.8 Optimized structures of (a) the O-centered radical 1A, (b) the transition state TS_1 , and (c) the C-centered radical 1B for 15 artemisinin compounds.

Table 6.4 Important structural parameters of radical 1A, transition state (TS₁), and radical 1B in pathway 1, distances in Angstroms and angles in degrees.

		1	42	43	88	89	90
O ₁ -H ₁₅	1A	2.392	2.264	2.256	2.381	2.320	2.318
	TS ₁	1.298	1.286	1.286	1.299	1.285	1.288
	1B	0.978	0.975	0.975	0.976	0.976	0.976
O ₁ -C ₄	1A	2.893	2.817	2.815	2.842	2.842	2.836
	TS ₁	2.435	2.410	2.411	2.422	2.409	2.410
	1B	2.937	2.800	2.802	2.792	2.793	2.789
C ₄ -H ₁₅	1A	1.093	1.094	1.094	1.092	1.093	1.093
	TS ₁	1.249	1.250	1.250	1.239	1.250	1.247
	1B	2.474	2.517	2.510	2.484	2.532	2.522
O ₂ -H ₁₅	1A	2.331	2.367	2.367	2.353	2.362	2.363
	TS ₁	2.595	2.487	2.487	2.411	2.425	2.426
	1B	1.827	1.829	1.828	1.810	1.826	1.822
O ₁ -H ₁₅ -C ₄	1A	106.3	109.0	109.4	103.5	107.2	106.9
	TS ₁	145.9	143.7	143.9	145.1	143.7	143.8
	1B	108.7	96.5	97.0	98.0	95.1	95.4

Table 6.4 (Continued)

		91	92	93	94	96	97
O ₁ -H ₁₅	1A	2.340	2.307	2.303	2.340	2.168	2.452
	TS ₁	1.288	1.288	1.288	1.288	1.290	1.288
	1B	0.976	0.976	0.976	0.976	0.976	0.976
O ₁ -C ₄	1A	2.844	2.832	2.831	2.845	2.719	2.870
	TS ₁	2.411	2.410	2.410	2.410	2.412	2.411
	1B	2.785	2.787	2.787	2.789	2.800	2.798
C ₄ -H ₁₅	1A	1.093	1.093	1.093	1.093	1.098	1.092
	TS ₁	1.248	1.248	1.248	1.248	1.247	1.248
	1B	2.524	2.523	2.523	2.518	2.498	2.507
O ₂ -H ₁₅	1A	2.361	2.365	2.365	2.362	2.490	2.363
	TS ₁	2.487	2.487	2.487	2.489	2.477	2.479
	1B	1.823	1.825	1.826	1.818	1.817	1.823
O ₁ -H ₁₅ -C ₄	1A	106.1	107.4	107.5	106.1	108.2	101.1
	TS ₁	143.8	143.8	143.8	143.8	143.9	143.9
	1B	95.0	95.3	95.2	95.7	97.7	96.9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 6.4 (Continued)

		99	101	104	model
O ₁ -H ₁₅	1A	2.508	2.172	2.171	2.479
	TS ₁	1.293	1.288	1.281	1.245
	1B	0.976	0.976	0.975	0.972
O ₁ -C ₄	1A	2.908	2.720	2.718	3.123
	TS ₁	2.414	2.411	2.409	2.424
	1B	2.831	2.795	2.797	3.001
C ₄ -H ₁₅	1A	1.093	1.098	1.097	1.097
	TS ₁	1.246	1.249	1.253	1.285
	1B	2.502	2.490	2.477	2.593
O ₂ -H ₁₅	1A	2.360	2.492	2.494	2.491
	TS ₁	2.457	2.480	2.482	2.571
	1B	1.792	1.825	1.828	1.904
O ₁ -H ₁₅ -C ₄	1A	100.2	108.0	108.0	116.3
	TS ₁	143.9	143.7	143.9	146.7
	1B	99.5	97.8	98.8	105.4

In order to examine the effect of substituent group at the C₁₀ position, the results of four compounds with different substituents at this position, i.e., compound 1 with C=O group, 42 with -OH group, 43 with -OCH₃ group, and 88 with -H group, were compared. Some structural parameters of compound 1 and its corresponding radicals are different from those of the other three compounds, e.g., O₁-C₄, O₂-H₁₅, and O₁-H₁₅-C₄. The reason may possibly due to the difference in conformation of the lactone ring. In the other 3 compounds, this ring is in the chair form whereas the C=O group at the C₁₀ position in compound 1 causes the ring to become a distort-chair form. However, from the activities of these four compounds, the substituent at the C₁₀ position appears to be not necessary for the high activity. Therefore, compounds without substituent group at the C₁₀ position were selected for further investigations.

For 12 compounds without substituent group at the C₁₀ position (compound 88 to 104), some interesting points were found. For the radical 1A, the distance between H₁₅ (transferred hydrogen atom) and the radical site O₁ is ranging from 2.168 to 2.508 Å compared to 2.479 Å for the model system. Our values are also longer than the

suggested critical value of 2.1 Å [116]. The O₁-H₁₅-C₄ angle of artemisinin compounds (100.2° to 108.2°) is around 8-16° smaller than that of the model system (116.3°). And the O₁-C₄ distance in the model system (3.123 Å) is too long compared to those of artemisinin compounds (2.718 to 2.908 Å). The differences in the O₁-H₁₅-C₄ angle and the O₁-C₄ distance between artemisinin compounds and the model system arise from the additional ring strained of the real system. Interestingly, differences in the O₁-H₁₅ distance and the O₁-C₄ distance between high activity compounds (compounds 88-94, 97, and 99) and low activity compounds (compounds 96, 101, and 104) were observed. The values of these two distances for low activity compounds are around 0.1 to 0.3 Å shorter than those for the high activity compounds. In the transition state TS₁, the H₁₅ atom is moved close to the O₁ atom, hence the O₁-H₁₅ distance becomes shortening to the value between 1.281 to 1.299 Å. The O₁-H₁₅-C₄ angle of artemisinin compounds is around 144 to 145°. This reconfirms that the collinear transition state is not necessary [116, 139]. For the radical 1B, the H₁₅ atom is placed close to the O₂ atom giving the distance of 1.792 to 1.828 Å, which is in the range of the hydrogen bond.

B. Energy Information

The activation energy of pathway 1 (EA₁), the energy difference between radical 1A and 1B (ΔE₁) as well as the imaginary frequency (ν₁) of the transition state for all 15 compounds are given in Table 6.5. The results of compounds 1, 42, 43, and 88 were again compared to investigate the effect of substituent at the C₁₀ position on the energy profile of pathway 1. The EA₁ of compounds 1, 42, 43, and 88 are 8.16, 4.05, 3.98, and 4.67 kcal/mol, respectively. The value of compound 1 significantly differs from the others which can again reasoned by the difference in conformation of the lactone ring. It seems that the substituent at the C₁₀ position has no significant effect on the activation energy of pathway 1 unless it causes the change in the conformation of the lactone ring. Comparing ΔE₁, we did not observe any significant difference in the values.

For 12 compounds without substituent at the C₁₀ position, significant differences in energy values between high and low activity compounds were discovered. The energy values of EA₁ and ΔE₁ for all high activity compounds (compounds 88-94, 97, and 99) are in the same range, i.e., 4.59 to 4.78 kcal/mol and -9.12 to -9.80 kcal/mol, respectively. For the low activity compounds (compounds 96, 101, and 104), the EA₁

and ΔE_1 are between 7.29 to 7.37 and -5.13 to -6.70 kcal/mol. Remarkably, compounds with low EA_1 and ΔE_1 (high activity) have substituent groups at the β -C₉ position while compounds with high EA_1 and ΔE_1 (low activity) have substituents at the C₃ position. Therefore, the position of substituent appears to have significant effect on EA_1 and ΔE_1 and hence, the activities. The model system in reference [138] has EA_1 and ΔE_1 energies of 7.22 and -5.16 kcal/mol, which is very close to those of the low activity compounds. This indicates that the model system exhibits the same characteristics as low activity compounds and, therefore, it tends to have low activity if exists.

Table 6.5 Activation energy (EA_1), energy difference between radical 1A and 1B (ΔE_1), and imaginary frequency (ν_1) of the transition state (TS_1) in pathway 1.

No	log (D-6)	log (W-2)	EA_1 (kcal/mol)	ΔE_1 (kcal/mol)	ν_1 (cm ⁻¹)
1	0.000	0.000	8.16	-9.65	1506.22 <i>i</i>
42	0.857	-0.018	4.05	-9.56	1411.18 <i>i</i>
43	0.447	0.553	3.98	-9.56	1391.05 <i>i</i>
88	0.819	0.754	4.67	-9.75	1404.24 <i>i</i>
89	0.375	0.279	4.71	-9.68	1408.47 <i>i</i>
90	0.961	0.668	4.65	-9.72	1389.31 <i>i</i>
91	0.675	0.740	4.66	-9.73	1394.04 <i>i</i>
92	1.765	1.320	4.66	-9.71	1397.06 <i>i</i>
93	0.230	0.161	4.66	-9.71	1396.00 <i>i</i>
94	1.705	1.399	4.64	-9.80	1404.01 <i>i</i>
96	-1.000	-1.000	7.29	-6.70	1399.77 <i>i</i>
97	0.859	0.836	4.78	-9.12	1407.45 <i>i</i>
99	0.262	0.398	4.59	-9.38	1371.70 <i>i</i>
101	-1.222	-1.699	7.35	-5.86	1415.59 <i>i</i>
104	-3.046	-3.046	7.37	-5.13	1385.63 <i>i</i>
ref. [138]	-	-	7.22	-5.16	1426 <i>i</i>

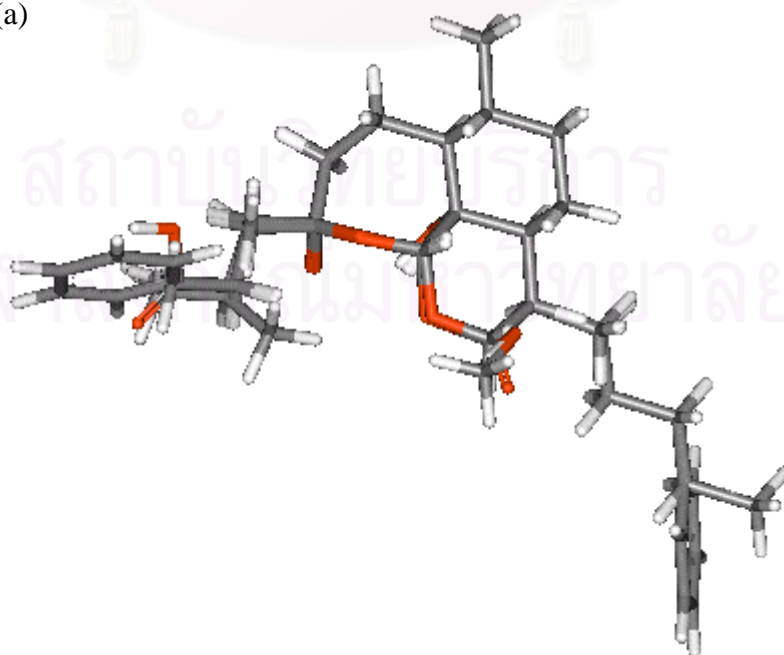
6.3.4.2 Homolytic C-C Cleavage (Pathway 2)

A. Structural Information

For all 15 compounds, some important structural parameters of the radical 2A, the radical 2B, and the transition state TS_2 are summarized in Table 6.6 and their optimized structures are given in Figure 6.9. In this pathway the features involved with C_3 and C_4 atoms along the reaction are the most relevant. Therefore, C_3 - C_4 distance, C_4 - C_5 - C_{5a} - C_{12a} and C_3 - O_{13} - C_{12} - C_{12a} dihedral angles were monitored. As in the pathway 1, compound 1, 42, 43, and 88 were compared to study the effect of substituent group at the C_{10} position. The radical intermediates of compound 1 have some structural parameters quite different from those of the other three compounds, especially in the radical 2B (see Table 6.6 and Figure 6.9c). Again, these disagreements may come from the difference in conformations of the lactone ring.

Considering 12 compounds without substituents at the C_{10} position, all the structural parameters of each compound are similar. No significant difference is observed. The C_3 - C_4 distance is obviously increasing as the reaction proceeds, i.e., from around 1.6 Å in 2A to around 2.0 Å in TS_2 and to 3.3-3.4 Å in 2B. The decrease in the C_4 - C_5 - C_{5a} - C_{12a} dihedral angle of around 21.9-23.0° and the increase in the C_3 - O_{13} - C_{12} - C_{12a} dihedral angle of around 31.0-32.7° from radicals 2A to 2B indicate that the increasing of C_3 - C_4 distance is also facilitated by the rotation of these two angles.

(a)



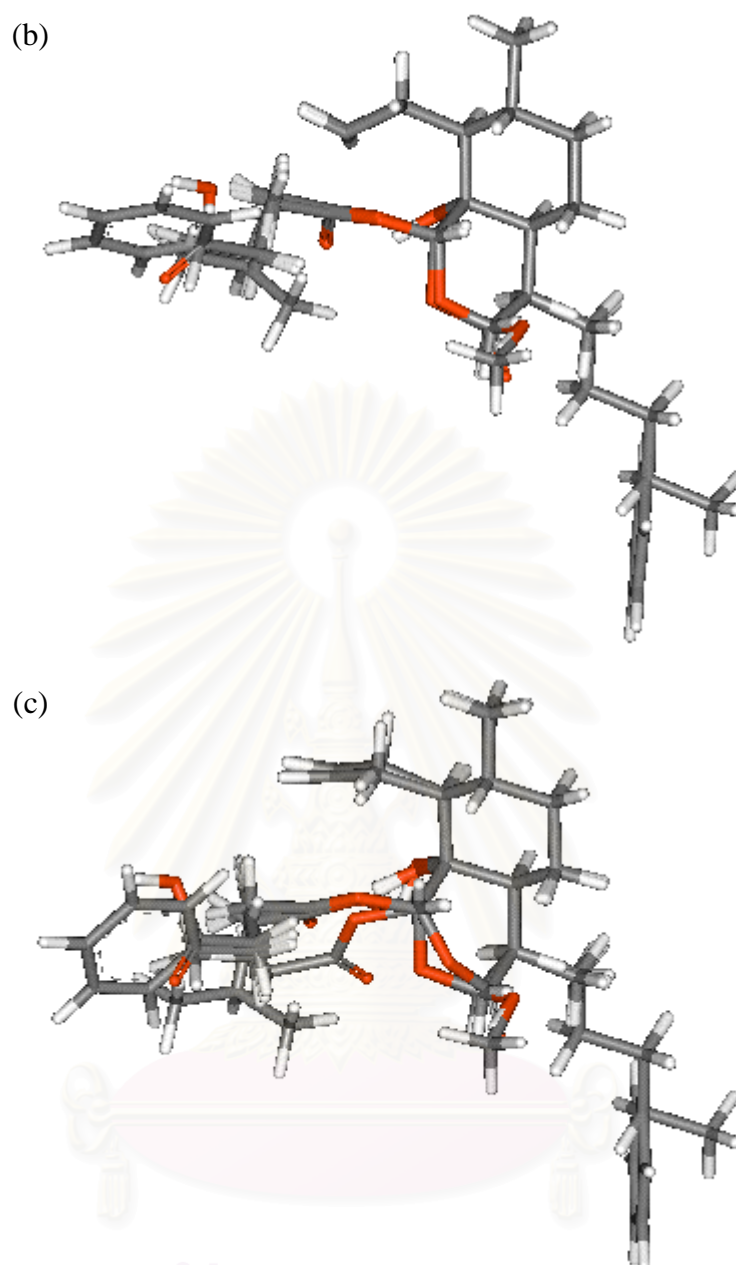


Figure 6.9 Optimized structures of (a) the O-centered radical 2A, (b) the transition state TS_2 , and (c) the C-centered radical 2B for 15 artemisinin compounds.

For the model system in reference [138], two different processes were found in the pathway 2. The C_3 - C_4 cleavage breaks the ring system to give an open chain compound that has more flexibility. In this case the second C-centered radical is allowed. But for artemisinin compounds, the ring strain caused by the existence of two remaining cyclic rings does not permit such a radical. Therefore, the model compound would not be a suitable representative for artemisinin systems in the study of mechanism of action.

Table 6.6 Important structural parameters of radical 2A, transition state (TS₂), and radical 2B in pathway 2, distances in Angstroms and angles in degrees.

		1	42	43	88	89	90
C ₃ -C ₄	2A	1.617	1.601	1.601	1.600	1.604	1.599
	TS ₂	1.979	1.967	1.968	1.968	1.968	1.967
	2B	3.508	3.332	3.345	3.324	3.293	3.302
O ₂ -C ₃	2A	1.332	1.336	1.336	1.335	1.334	1.336
	TS ₂	1.256	1.256	1.256	1.256	1.256	1.257
	2B	1.212	1.213	1.213	1.213	1.213	1.213
O ₂ -C ₃ -O ₁₃	2A	117.1	116.3	116.3	116.4	116.5	116.4
	TS ₂	122.8	122.6	122.7	122.7	122.7	122.7
	2B	125.2	125.7	125.8	125.8	125.8	125.8
O ₂ -C ₃ -O ₁₃ -C ₁₂	2A	41.2	49.6	49.5	48.8	48.9	49.3
	TS ₂	24.4	28.6	28.6	27.8	27.9	28.1
	2B	-19.7	-12.5	-12.9	-14.2	-13.4	-14.0
C ₄ -C ₅ -C _{5a} -C _{12a}	2A	325.3	327.8	327.9	328.1	327.9	328.0
	TS ₂	321.5	323.5	323.3	323.7	323.4	323.6
	2B	300.8	304.4	304.7	305.1	305.6	305.5
C ₃ -O ₁₃ -C ₁₂ -C _{12a}	2A	46.4	42.0	42.2	43.6	43.6	43.3
	TS ₂	53.9	52.7	52.8	53.8	53.7	53.5
	2B	69.8	73.4	73.3	75.2	75.1	75.1

Table 6.6 (Continued)

		91	92	93	94	96	97
C ₃ -C ₄	2A	1.599	1.599	1.599	1.600	1.599	1.599
	TS ₂	1.965	1.966	1.966	1.967	1.968	1.967
	2B	3.297	3.302	3.305	3.295	3.314	3.296
O ₂ -C ₃	2A	1.336	1.336	1.336	1.335	1.335	1.335
	TS ₂	1.257	1.257	1.257	1.257	1.256	1.256
	2B	1.213	1.213	1.213	1.213	1.214	1.214
O ₂ -C ₃ -O ₁₃	2A	116.4	116.4	116.4	116.4	116.1	116.1
	TS ₂	122.7	122.7	122.7	122.7	122.5	122.5
	2B	125.8	125.8	125.8	125.8	125.4	125.4
O ₂ -C ₃ -O ₁₃ -C ₁₂	2A	49.1	49.4	49.1	49.0	48.3	48.5
	TS ₂	28.0	28.0	28.0	28.0	27.9	27.8
	2B	-13.8	-14.1	-14.1	-13.6	-14.8	-14.3
C ₄ -C ₅ -C _{5a} -C _{12a}	2A	327.9	327.7	327.7	327.9	327.6	327.5
	TS ₂	323.6	323.6	323.6	323.7	323.4	323.3
	2B	305.5	305.8	305.7	305.6	305.4	305.4
C ₃ -O ₁₃ -C ₁₂ -C _{12a}	2A	43.2	43.0	43.0	43.2	43.3	43.3
	TS ₂	53.4	53.6	53.6	53.5	53.5	53.4
	2B	75.2	75.4	75.4	75.1	75.4	75.3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 6.6 (Continued)

Parameter		99	101	104	model
C ₃ -C ₄	2A	1.605	1.598	1.596	1.640
	TS ₂	1.949	1.971	1.986	1.947
	2B	3.369	3.308	3.305	3.843
O ₂ -C ₃	2A	1.335	1.335	1.335	1.326
	TS ₂	1.259	1.255	1.254	1.262
	2B	1.215	1.213	1.212	1.212
O ₂ -C ₃ -O ₁₃	2A	116.5	116.0	116.0	118.3
	TS ₂	122.5	122.5	122.8	123.0
	2B	125.3	125.4	125.7	126.4
O ₂ -C ₃ -O ₁₃ -C ₁₂	2A	44.5	49.3	49.7	40.3
	TS ₂	25.1	28.0	27.9	26.4
	2B	-18.4	-14.3	-14.4	-2.18
C ₄ -C ₅ -C _{5a} -C _{12a}	2A	327.1	327.5	327.3	325.9
	TS ₂	323.9	323.4	323.1	321.2
	2B	304.5	305.2	305.4	295.4
C ₃ -O ₁₃ -C ₁₂ -C _{12a}	2A	44.0	42.9	42.6	51.3
	TS ₂	54.1	53.5	53.7	58.6
	2B	75.0	75.1	75.3	85.4

B. Energy Information

The activation energy of pathway 2 (EA₂), the energy difference between radical 2A and 2B (ΔE_2), and the imaginary frequency (ν_2) of the transition state for all 15 artemisinin compounds are shown in Table 6.7. As in pathway 1, the results of compound 1, 42, 43, and 88 were compared to study the influence of the substituent at the C₁₀ position to the energy profile of pathway 2. The values of EA₂ and ΔE_2 for these compounds are 4.46, 4.95, 4.96, 5.07 kcal/mol and -11.29, -8.23, -8.32, -8.52 kcal/mol, respectively. This reveals that the substituent at the C₁₀ position which changes the conformation of lactone ring significantly affects only the energy of radical 2B but not TS₂. This effect can be explained from the structures shown in Figure 6.9. In the TS₂ intermediate, the C₃-C₄ distance is only around 0.3-0.4 Å longer than that in 2A. The conformation of compound still remains unchanged. Therefore, no effect from the

lactone ring was observed. But in the 2B intermediate of compound 1, the distort-chair conformation of the lactone ring causes the C₃ atom to move remotely apart from the C₄ atom. The resulted longer C₃- C₄ distance leads to a lower energy (less repulsion) and hence lower ΔE_2 . Similarly to the pathway 1, although substituent at the C₁₀ position has the effect on the intermediate 2B but it appears not to be a necessarily requirement for high activity.

Table 6.7 Activation energy (EA₂), energy difference between radical 2A and 2B (ΔE_2), and imaginary frequency (ν_2) of the transition state (TS₂) in pathway 2.

No	log (D-6)	log (W-2)	EA ₂ (kcal/mol)	ΔE_2 (kcal/mol)	ν_2 (cm ⁻¹)
1	0.000	0.000	4.46	-11.29	473.31 <i>i</i>
42	0.857	-0.018	4.95	-8.23	477.65 <i>i</i>
43	0.447	0.553	4.96	-8.32	482.98 <i>i</i>
88	0.819	0.754	5.07	-8.52	484.91 <i>i</i>
89	0.375	0.279	5.99	-7.54	483.36 <i>i</i>
90	0.961	0.668	5.11	-8.55	488.85 <i>i</i>
91	0.675	0.740	5.10	-8.54	485.76 <i>i</i>
92	1.765	1.320	5.08	-8.53	483.04 <i>i</i>
93	0.230	0.161	5.08	-8.53	483.46 <i>i</i>
94	1.705	1.399	5.06	-8.56	485.46 <i>i</i>
96	-1.000	-1.000	5.19	-8.32	479.02 <i>i</i>
97	0.859	0.836	5.13	-8.51	474.45 <i>i</i>
99	0.262	0.398	4.36	-10.20	483.51 <i>i</i>
101	-1.222	-1.699	5.23	-8.06	464.15 <i>i</i>
104	-3.046	-3.046	5.63	-7.41	447.21 <i>i</i>
ref. [138]	-	-	7.76	-10.07	470 <i>i</i>

Considering 12 compounds without substituents at the C₁₀ position, unlike the pathway 1, no significant difference in energy values among high and low activity compounds was detected. The energy values of EA₂ and ΔE_2 for all compounds are in the range of 4.36 to 5.99 kcal/mol and -7.41 to -10.20 kcal/mol, respectively. Therefore, the substituent groups at either C₃ or C₉ positions do not change the pattern of both EA₂ and ΔE_2 energies. Comparing the model system with artemisinin compounds, its ΔE_2

value (-10.07 kcal/mol) is in our range while its EA_2 of 7.76 kcal/mol is too high. Therefore, the model compound may have more difficulty to proceed along this pathway.

6.3.4.3 Comparison between the Two Pathways

As described in the introduction that the two pathways are competitive processes, it is interesting to estimate the ease of reactions by comparing energies of radicals in pathway 1 to those in pathway 2. Therefore, the energy differences between radical 1A and 2A, $\Delta E(1A-2A)$, and between radical 1B and 2B, $\Delta E(1B-2B)$, for all compounds are calculated (Table 6.8). And the energy schemes for both pathways of artemisinin, high and low activity compounds are summarized in Figure 6.10, 6.11, and 6.12, respectively.

Table 6.8 Energy difference between radical 1A and 2A, $\Delta E(1A-2A)$, and between radical 1B and 2B, $\Delta E(1B-2B)$.

Compound	$\log(D-6)$	$\log(W-2)$	$\Delta E(1A-2A)$ (kcal/mol)	$\Delta E(1B-2B)$ (kcal/mol)
1	0.000	0.000	+2.59	+4.23
42	0.857	-0.018	+3.71	+2.38
43	0.447	0.553	+3.68	+2.45
88	0.819	0.754	+3.68	+2.45
89	0.375	0.279	+4.57	+2.42
90	0.961	0.668	+3.70	+2.54
91	0.675	0.740	+3.71	+2.52
92	1.765	1.320	+3.68	+2.49
93	0.230	0.161	+3.67	+2.50
94	1.705	1.399	+3.72	+2.48
96	-1.000	-1.000	+0.80	+2.42
97	0.859	0.836	+3.33	+2.72
99	0.262	0.398	+4.02	+4.84
101	-1.222	-1.699	+0.26	+2.46
104	-3.046	-3.046	-0.328	+1.95
ref. [138]	-	-	+1.51	+6.42

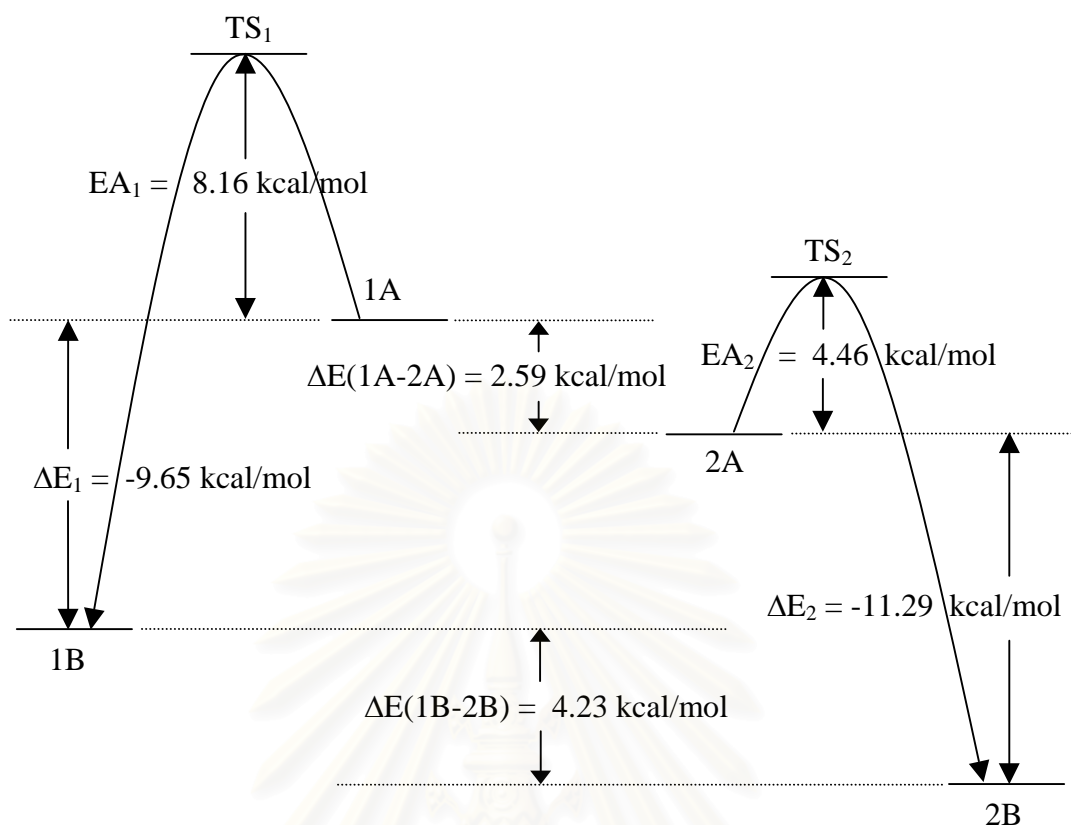


Figure 6.10 Energy scheme for the mechanism of action for artemisinin.

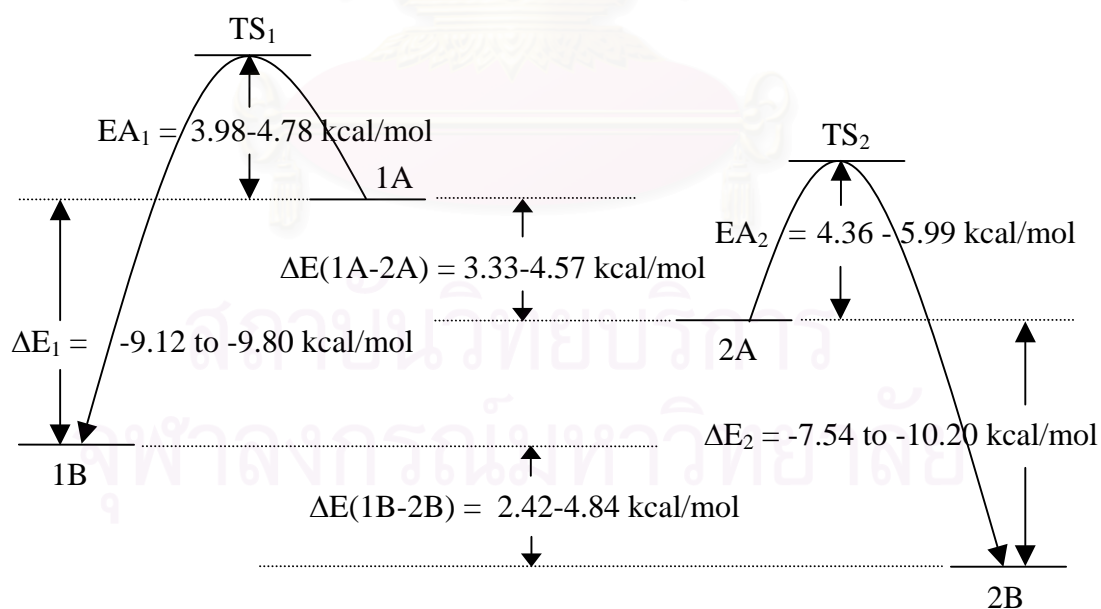


Figure 6.11 Schematic energy scheme for the mechanism of action for 9 high activity artemisinin derivatives.

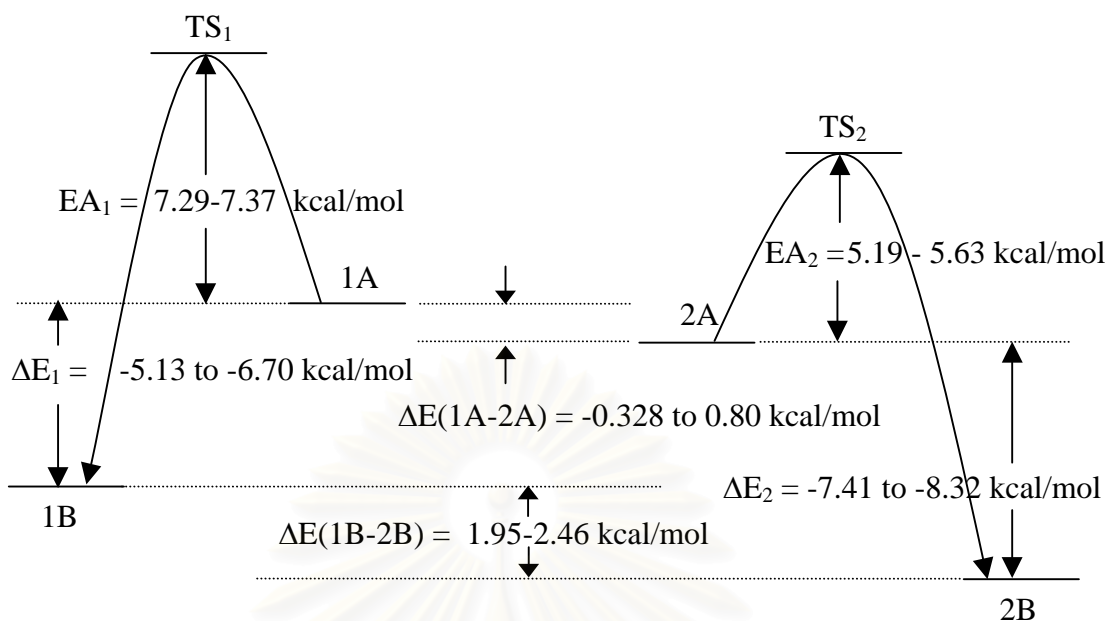


Figure 6.12 Schematic energy scheme for the mechanism of action for 3 low activity artemisinin derivatives (except compound 104 that has 1A energy slightly lower than 2A energy).

For artemisinin (Figure 6.10), the radical 2A has lower energy than the radical 1A. And the activation energy for pathway 2 (EA_2) is also lower than that for pathway 1. Moreover, the ΔE_2 value is less than the ΔE_1 value. Therefore, from the energetic point of view, the pathway 2 should occur more preferable than the pathway 1. For other compounds without C=O at the C₁₀ position, it is clearly seen that the high activity compounds always have higher $\Delta E(1A-2A)$, 3.33 to 4.57 kcal/mol (Figure 6.11), while the low activity compounds have lower values, -0.328 to 0.80 kcal/mol (Figure 6.12). This implies that the stability of the radical 2A is very possibly a requirement for high antimalarial activity and also emphasizes the importance of the pathway 2 over the pathway 1. Considering the C-centered radical, the energy of radical 2B in all compounds is also lower than that of radical 1B, $\Delta E(1B-2B)$ of 1.95 to 4.84 kcal/mol.

6.3.4.4 Relationship to Biological Activity

As one of our objectives for this study is to investigate relationships between antimalarial activities and properties implicated in the mechanism of action, all energy data (EA_1 , ΔE_1 , EA_2 , ΔE_2 , $\Delta E(1A-2A)$, and $\Delta E(1B-2B)$) and all structural parameters presented in Table 6.4 and 6.6 were considered. The relationships were identified by plotting all the properties against activities. Since the substituent group at the C_{10} position seems to have no significant effect on the biological activities, only 12 compounds without substituent at this position were considered.

For the structural parameters, no relationship with activities was found. However, three energy parameters, i.e., EA_1 , ΔE_1 , and $\Delta E(1A-2A)$, have relations with antimalarial activities. And their relationships were displayed in Figure 6.14 to 6.19. The graphically illustrations clearly demonstrate that these energies can be used to distinguish between high and low activity compounds.

Considering the activation energy for pathway 1 (EA_1), compounds with low EA_1 have high activities (see Figure 6.13 and 6.14). Since the EA_1 refers to the ease in proceeding along reaction of pathway 1, the radical 1B seems to be an important species for high activities.

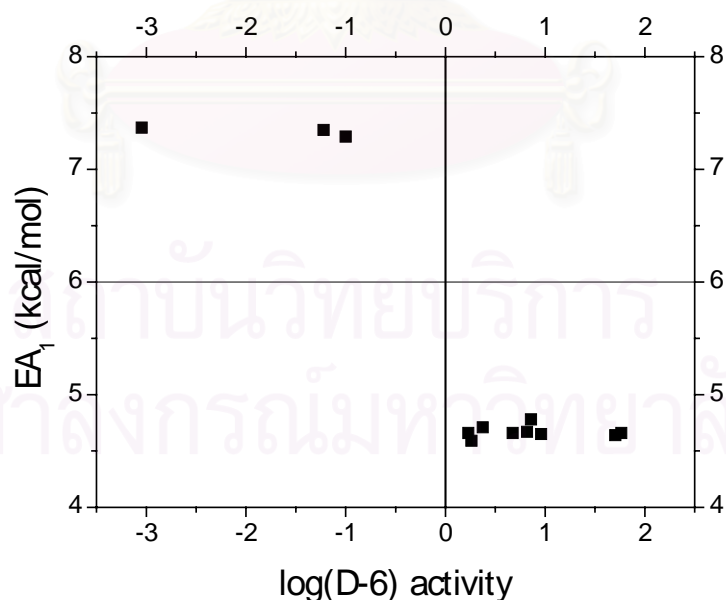


Figure 6.13 Relationship between log(D-6) activities and EA_1 for 12 artemisinin derivatives.

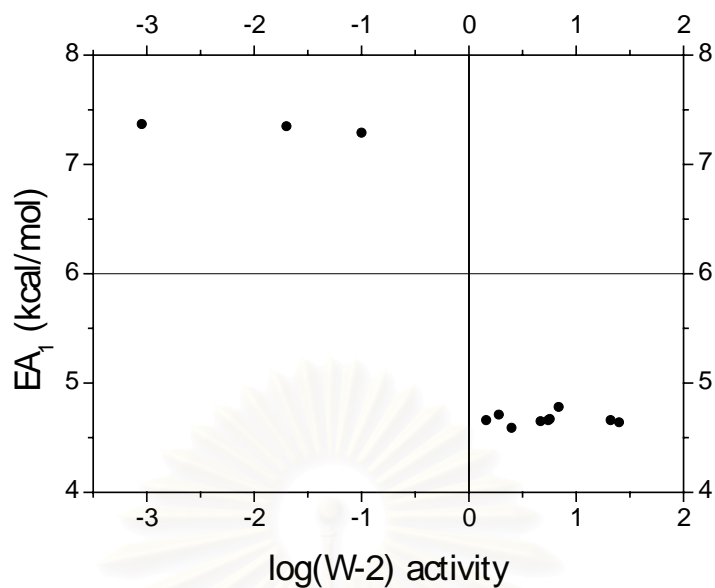


Figure 6.14 Relationship between log(W-2) activities and EA₁ for 12 artemisinin derivatives.

For the ΔE_1 parameter, compounds with lower ΔE_1 have higher activities (see Figure 6.15 and 6.16). This points out once again that the radical 1B is essential for high activities because the ΔE_1 refers to the ease of radical 1B formation.

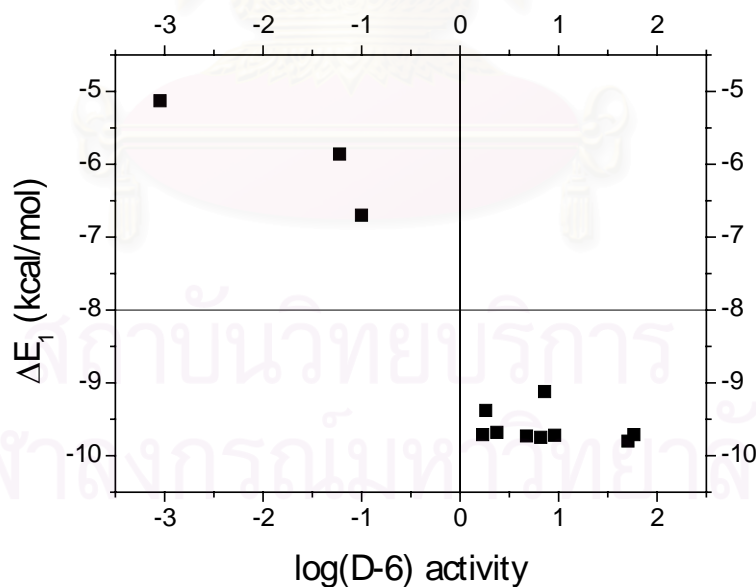


Figure 6.15 Relationship between log(D-6) activities and ΔE_1 for 12 artemisinin derivatives.

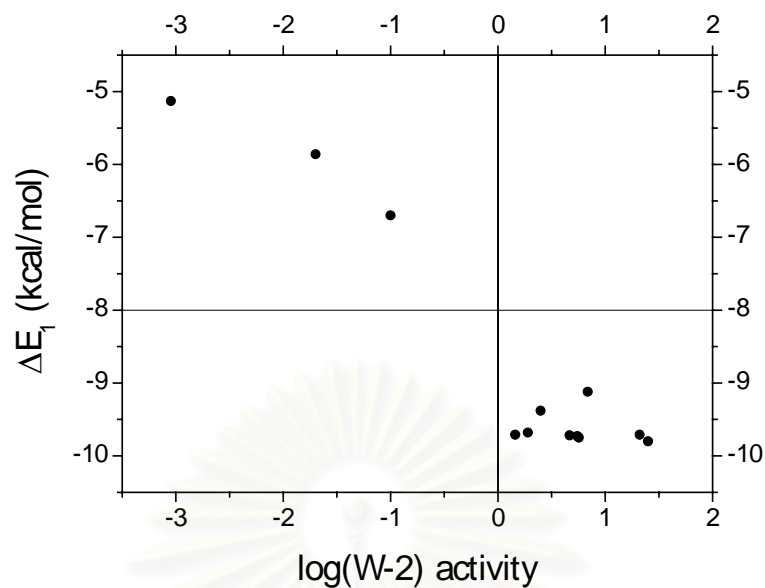


Figure 6.16 Relationship between $\log(W-2)$ activities and ΔE_1 for 12 artemisinin derivatives.

In the case of $\Delta E(1A-2A)$ parameter (Figure 6.17 and 6.18), high activity compounds have high $\Delta E(1A-2A)$ values, i.e., radical 2A is more stable than radical 1A. This implies that the energetically more preferable pathway 2 is also importance for high activities.

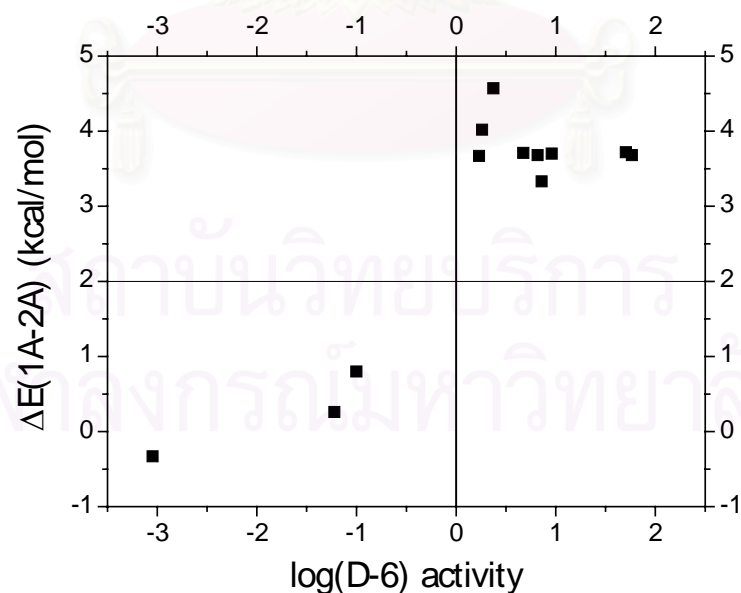


Figure 6.17 Relationship between $\log(D-6)$ activities and $\Delta E(1A-2A)$ for 12 artemisinin derivatives.

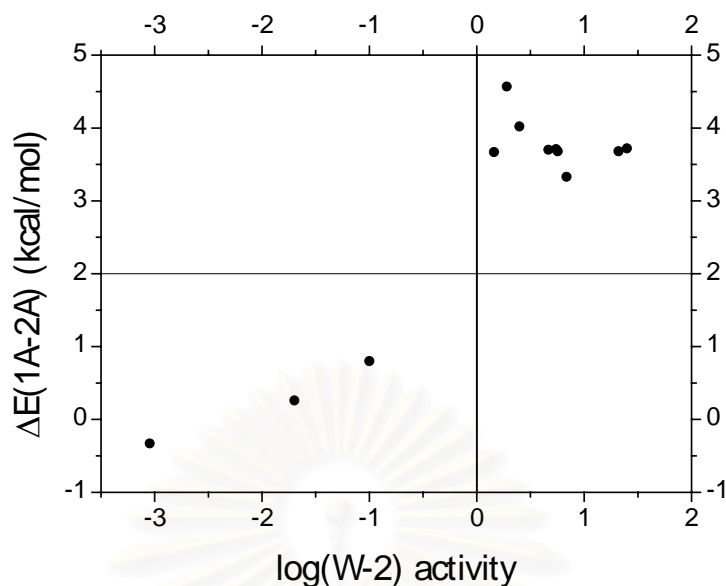


Figure 6.18 Relationship between $\log(W-2)$ activities and $\Delta E(1A-2A)$ for 12 artemisinin derivatives.

6.4 Summary

From the three proposed mechanisms, two of them (pathway 1 and 2) involve the O-O bond breaking while the other one involves the C-O bond breaking (pathway 3). Pathway 3 was shown to be an inferior pathway as indicated by much higher energy of intermediate 3A. For the pathways 1 and 2, the homolysis of O-O bond to form the diradical before attacking the Fe^{2+} was proposed. The energy gap between close-shell singlet ground state (reactant compound) and triplet state (diradical compound) of 11 artemisinin compounds was evaluated. Since the HF/3-21G was shown to be not sufficient for the prediction of singlet-triplet splitting, the B3LYP/6-31G* level was used instead. The calculated energy gap value of between -31.24 to -29.48 kcal/mol showed that the diradical formation is rather difficult and would require some facilitation.

Furthermore, the reactions leading to the formation of C_4 free radicals in pathways 1 and 2 were investigated. The quantum chemical calculations of oxygen radical, transition state, and carbon radical along the two competitive reaction mechanisms of 15 artemisinin compounds using the IMOMO method were carried out. The IMOMO(B3LYP/6-31G**):HF/3-21G yielded better structural parameters than the HF/3-21G method and in some cases better than the B3LYP/6-31G** level. The effect of substituents at the C_{10} position on the reaction mechanisms was studied. Only a

substituent group which changes the conformation of the lactone ring (such as C=O group) affects the EA_1 value due to the strain of this ring. However, the substituent group appears to have no influence on the activities. Therefore, only 12 compounds without substituent group at the C_{10} positions were considered.

Good relationships between antimalarial activities and three energy parameters, EA_1 , ΔE_1 , and $\Delta E(1A-2A)$, were found. The graphical illustrations of these relationships show that the three energies could be used to distinguish high activity compounds from low activity compounds. Since compounds with low EA_1 energy and high ΔE_1 energy, a case which indicates the ease of intramolecular 1,5-hydrogen shift reaction in the pathway 1, have high antimalarial activities, the radical 1B tends to be an important species.

Considering the $\Delta E(1A-2A)$, for the high activity compounds, the homolytic C-C cleavage reaction (pathway 2) is energetically more preferable than the intramolecular 1,5-hydrogen shift process (pathway 1). This is in well agreement with the docking results between artemisinin compounds and heme [140-141], which illustrate the preference of heme iron to approach the O_1 atom at the endoperoxide linkage of artemisinin compounds. On the other hand, the energy of the oxygen free radical in the two pathways for the low activity compounds is comparable, with a maximum difference of only 0.80 kcal/mol. Therefore, the pathway 2 is also important for high activities.

Our calculations also demonstrated that the use of 6,7,8-trioxobicyclo[3,2,2]nonane as a representative compound for artemisinin systems in reference [137-138] is inappropriate for the studies of mechanism of action. Although the calculations on this molecule gave some of structural parameters in agreement with artemisinin compounds, it, however, exhibits the same energetic profiles as those of the low activity artemisinin compound.

Finally, the obtained data help us to understand the mechanism of action of artemisinin compounds in more details which could aid in the development of new more effective antimalarial agents.

CHAPTER 7

Conclusions

7.1 QSAR Models

The quantitative structure-activity relationships of antimalarial artemisinin compounds were investigated by means of traditional QSAR, CoMFA (3D-QSAR), automated molecular docking using the simulated annealing Monte Carlo simulations, and quantum chemical calculations using the IMOMM(B3LYP/6-31G**):HF/3-21G method. The activities against two strains of malarial parasite, i.e., D-6 and W-2, were used. The HF/3-21G method was justified to be a suitable method for geometry optimization.

In the traditional QSAR, it seems that the analysis for the whole set of compounds could not be succeeded as indicated by the low statistical values. Therefore, compounds were classified into 3 groups by the structural similarity. Investigations on each individual group, impressive models with high predictive ability could be derived. The obtained models could predict activities of compounds in the test set quite close to the experimental values, thus indicating their high predictive power.

As in the traditional QSAR, the CoMFA analysis was performed on both the whole set and each individual set of compounds, in which moderate and low predictive models were obtained. The CoMFA, which considers steric and electrostatic fields only, is not able to describe the activities of our compounds. Therefore, other fields, such as hydrophilic field, should be additionally used to achieve a better model. Moreover, the force fields used in the CoMFA are not suitable to account for the interactions involving free radical. As a result, the quantum chemical method must be applied instead.

From the docking calculations, significant relationships between activities and binding energies were found for compounds without C=O group at the C₁₀ position. Furthermore, EA₁, ΔE₁, and ΔE(1A-2A) involving the mechanism of action for 12 10-deoxoartemisinin derivatives also show high correlations to the activities as indicated by the r² values of between 0.694 to 0.897.

7.2 Mechanism of Action

The results obtained from the automated molecular docking and the quantum chemical calculation facilitate the better understanding of the mechanism of action. The mechanism proposed by Haynes (pathway 3) which involves the C-O bond breaking appears to be not practicable since its intermediate has much higher energy (around 100 kcal/mol) than other two proposals from Posner and Jefford (pathway 1 and 2) which involve the O-O bond breaking. There are two possible reactions for the O-O bond breaking, either via the direct bond breaking (diradical formation) or via the complexation with Fe^{2+} . The singlet-triplet gap for the diradical formation calculated at the B3LYP/6-31G* level is quite high (around 30 kcal/mol). Therefore, from the energetic point of view, the O-O bond breaking should occur via the complexation with Fe^{2+} rather than the direct bond breaking.

From the docking calculations, the Fe^{2+} approaches artemisinin compounds at the O_1 more preferably than at the O_2 . This is possibly due to the more negative charge of O_1 over O_2 . More interestingly, it is observed that the compound with more negative O_1 charge becomes more active as indicated by the QSAR models (i.e., models 3.51 and 3.52). The docking results are in agreement with the IMOMO(B3LYP/6-31G**: $\text{HF}/3\text{-}21\text{G}$) calculations, where the radical 2A is energetically more favorable than the radical 1A.

From the above, it could be concluded that the reaction moves towards the production of radical 2B more than 1B. Although the radical 2B (primary carbon radical) is more efficient in the alkylation step than the radical 1B (secondary carbon radical), the radical 1B seems to have some significant contributions to the activities. It is evident from the fact that compounds with high EA_1 and low ΔE_1 (difficult to produce the radical 1B) are inactive. The importance of the radical 2B is implicitly seen from the data of inactive compounds having low $\Delta\text{E}(1\text{A}-2\text{A})$, which means the formation of the radical 2A (and hence 2B) is reduced. Therefore, both 1B and 2B radicals play important roles in the mechanism of action.

7.3 Effect of Substituent Group to Activity

The primary goal of this study is finding the relationships between structure and activity in a quantitative manner, but the relationships in a qualitative manner should not be neglected because they are also very helpful in designing a new potent drug. Therefore, the effects of the substituents at the C₁₀, C₃, and C₉ positions were considered.

- a) Substituent group at the C₁₀ position, particularly the C=O group, that modifies conformation of the lactone ring make a compound different from other compounds without this substituent. Although the substituent does not change the binding characteristics in the docking, it leads to the different character in the mechanism of action from other compounds. This would be the main reason for poor CoMFA models since the CoMFA method can be applied to only compounds having the same mode of action. Moreover, such a substituent also causes the substituent groups at C₃ and C₉ position to manifest their effects differently from those in other compounds (see Chapter 2). This is also possibly one of the reasons for the poor predictive models in QSAR and CoMFA methods when compounds of this group were considered together with other compounds.
- b) Substituents at the C₃ position have no steric effect on the activities as indicated from the CoMFA model. But it has significant influence, possibly from electronic effect, on the mechanism of action by raising the EA₁ and lowering both ΔE_1 and $\Delta E(1A-2A)$, thus leading to inactive compound.
- c) Substituent groups at the α -C₉ position, such as -Br and -OH, have an effect on the binding characteristics in the docking by hindering the approach of heme iron to the endoperoxide moiety as seen in compounds 8, 10, 76, and 83 to 87. In addition, substituents that cause conformational change in the lactone ring, such as C=O group in compound 2 and 5, also produces the steric hindrance and hence low activities.
- d) From the CoMFA results, the steric favorable contour was found around the substituent group at the β -C₉ position. Therefore, large steric groups could increase the activities.

7.4 Suggestions for New Compound

Using the information obtained from all investigations in this study, we could afford some suggestions on the structural modification of artemisinin compounds to enhance the activities as follows. Firstly, it is not necessary to include a substituent group at the C₁₀ position. Secondly, the new compound must have moderate solubility in both water and oil. Thirdly, it should contain steric substituent group at the C₉ position without causing the change in the conformation of the lactone ring. Finally, it is advised to include substituent at either C₃ or C₉ positions that causes the atomic charge of O₁ atom to be more negative as compared to that of the O₂.

7.5 Suggestions for Further Works

It is very interesting to further study the QSAR and reaction mechanism of artemisinin compounds. The following points are suggested as a guideline for further works.

- a) In order to find a better QSAR model, a non-linear statistical method is worth to apply. This is because the relationship between molecular properties and biological activity is not necessary a linear form.
- b) From the traditional QSAR, the hydrophilicity is important to antimalarial activities. Hence, hydrophilic field is recommended for further CoMFA study.
- c) In the reaction mechanism of artemisinin compounds, the main interactions are between heme iron and endoperoxide linkage of artemisinin. Therefore, the use of Fe²⁺ as probe atom would be of interest for further CoMFA study.
- d) From the quantum chemical studies of reaction mechanism, artemisinin was found to have different energy profile from other compounds. This may suggests a different mode of action between compounds with (group 1) and without (group 2 and 3) C=O at the C₁₀ position. Therefore, the IMOMO calculations of more compounds in the group 1 are very informative and should be performed.
- e) The investigations on reaction mechanism with Fe atom are very interesting. However, the calculations are really difficult and expensive process. Thus, it is not possible for us to tackle this problem since we do not have proper tools (both

software and hardware). Nevertheless, these calculations should be carried out when efficacious software and hardware are available.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

REFERENCES

1. Malaria Unit, World Health Organization. Global malaria control. Bull. WHO 71 (1993): 281-284.
2. World Health Organization. The World Health Report 1996.
3. Butter, D.; Maurice, J.; and O'Brien, C. Time to put malaria control on the global agenda. Nature 386 (1997): 535-536.
4. Maclean, M.; Anderson, J.; and Davies, C. Making malaria research bite. Nature Med. 3 (1997): 14-16.
5. กรองทอง ทิมาสาร. ความสูญเสียจากโรคมาลาเรีย. วารสารมาลาเรีย 33 (1998): 217-218.
6. Butler, A.R.; and Wu, Y.-L. Artemisinin (Qinghaosu): A New Type of Antimalarial Drug. Chem. Soc. Rev. (1992): 85-90.
7. Wyler D.J. Malaria: overview and update. Clin. Infect. Dis. 16 (1993): 449-458.
8. Barradell, L.B.; and Fitton, A. Artesunate: A Review of its Pharmacology and Therapeutic Efficacy in the Treatment of Malaria. Drugs 50 (1995): 714-741.
9. Hoffman, S.L. Diagnosis, treatment, and prevention of malaria. Med. Clin. North Am. 76 (1992): 1327-1355.
10. Bray, R.S.; and Garnham, P.C.C. The Life-Cycle of Primate Malaria Parasites. Br. Med. Bull. 38 (1982): 117-122.
11. Trigg, P.I.; and Kondrachine, A.V. Commentary: Malaria control in the 1990s. Bull. WHO 76(1998): 11-16.
12. World Health Organization. The World Health Report 1999: Make a difference.
13. Butter, D.; Maurice, J.; and O'Brien, C. Vaccines: a roller-coaster of hopes. Nature 386 (1997): 537-538.
14. Godson, G.N. Molecular Approaches to Malaria Vaccines. Sci. Am. 252 (1985): 32-39.
15. World Health Organization. Product Research and Development. In World Health Organization, Tropical Disease Research Progress 1997-98. Fourteenth Programme Report, pp. 33-37. 1998.
16. Marshall, E. Malaria Research – What Next? Science 247 (1990): 399-402.
17. Anders, R.F.; and Sual, A. Malaria Vaccines. Parasitol. Today 16 (2000): 444-447.

18. Butter, D.; Maurice, J.; and O'Brien, C. Transgenic mosquitoes: a new solution? Nature 386 (1997): 538.
19. World Health Organization. Strategic Research. In World Health Organization, Tropical Disease Research Progress 1997-98. Fourteenth Programme Report, pp. 43-52. 1998.
20. Muehlens, P. Die Behandlung der natürlichen menschlichen Malaria-Infektion mit Plasmochin. Naturwissenschaften 14 (1926): 1162-1166.
21. Bruce-Chwatt, L.J. (ed.), Chemotherapy of Malaria. Revised second edition, England: World Health Organization, 1986.
22. Curd, F.H.S.; Davey, D.G.; and Rose, F.L. Studies on synthetic antimalarial drugs. X. Some biguanide derivatives as new types of antimalarial substances with both therapeutic and causal prophylactic activity. Ann. Trop. Med. Parasitol. 39 (1945): 208-216.
23. Falco, E.A.; Goodwin, L.G.; Hitchings, G.H.; Rollo, I.M.; and Russell, P.B. 2,4-Diaminopyrimidines – a new series of antimalarials. Br. J. Pharmac. Chemother. 6 (1951): 185-200.
24. Moore, D.V.; and Lanier, J.E. Observations on two *Plasmodium falciparum* infections with an abnormal response to chloroquine. Am. J. Trop. Med. Hyg. 10 (1961): 5-9.
25. Canfield, C.J.; and Heiffer, M.H. The U.S. Army Drug Development Program. In M. Adolphe (ed.), Advances in Pharmacology and Therapeutics, Proceedings of the 7th International Congress of Pharmacology, Paris 1978, Volume 10 Chemotherapy, pp. 99-108. Paris, 1978.
26. Berrett, P.J.; Emmins, P.D.; Clarke, P.D.; and Bradley, D.J. Comparison of adverse events associated with use of mefloquine and combination of chloroquine and proguanil as antimalarial prophylaxis: Postal and telephone survey of travellers. Br. Med. J. 313 (1996): 525-528.
27. Olliaro, P.L.; and Trigg, P.I. Status of antimalarial drugs under development. Bull. WHO 73 (1995): 565-571.
28. Mockenhaupt, F.P. Mefloquine resistance in *Plasmodium falciparum*. Parasitol. Today 11 (1995): 248.

29. Boudreau, E.F.; Webster, H.K.; Pavanand, K.; and Thosingha, L. Type II mefloquine resistance in Thailand. Correspondence. Lancet 2 (1982): 1335.
30. Qinghaosu Antimalaria Coordinating Research Group. Antimalaria Studies on Qinghaosu. Chin. Med. J. 92 (1979): 811-816.
31. China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials. Chemical studies on qinghaosu (artemisinin), J. Tradit. Chin. Med. 2 (1982): 3-8.
32. Luo, X.-D.; Yeh, H.J.C.; Brossi, A.; Flippen-Anderson, J.L.; and Gilardi, R. The Chemistry of Drugs: Configurations of Antimalarials Derived from Qinghaosu: Dihydroqinghaosu, Artemether, and Artesunic Acid. Helv. Chim. Acta 67 (1984): 1515-1522.
33. Leban, I.; and Golic, L. Crystal and molecular structure of Qinghaosu: a redetermination. Acta Pharm. Jugosl. 38 (1988): 71-77.
34. Klayman, D.L.; Lin, A.J.; Acton, N.; Scovill, J.P.; Hock, J.M.; Milhous, W.K.; Theoharides, A.D.; and Dobek, A.S. Isolation of artemisinin (qinghaosu) from *Artemisia annua* growing in the United States. J. Nat. Prod. 47 (1984): 715-717.
35. Jiang, J.-B.; Li, G.-Q.; Guo, X.-B.; Kong, Y.C.; and Arnold, K. Antimalarial Activity of Mefloquine and Qinghaosu. Lancet 8 (1982): 285-288.
36. Klayman, D.L. Qinghaosu (Artemisinin): An Antimalarial Drug from China. Science 228 (1985): 1049-1055.
37. China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials. The Chemistry and Synthesis of Qinghaosu Derivatives, J. Tradit. Chin. Med. 2 (1982): 9-16.
38. China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials. Clinical studies on the treatment of malaria with quighaosu and its derivatives. J. Tradit. Chin. Med. 2 (1982): 45-50.
39. Looareesuwan, S.; Viravan, C.; Vanijanonta, S.; Wilairatana, P.; Suntharasamai, P.; Charoenlarp, P.; Arnold, K.; Kyle, D.; Canfield, C.; and Webster, H.K. A randomised trial of mefloquine, artesunate, and artesunate followed by mefloquine in acute uncomplicated falciparum malaria. Lancet 339 (1992): 821-824.

40. White, N.J.; Waller, D.; Crawley, J.; Nosten, F.; Chapman, D.; Brewster, D.; and Greenwood, B.M. Comparison of artemether and chloroquine for severe malaria in Gambian children. Lancet 339 (1992): 317-321.
41. White, N.J. Artemisinin: current status. Trans. Royal Soc. Trop. Med. Hyg. 88 (Supplement 1, 1994): 3-4.
42. Win, K.; Than, M.; and Thwe, Y. Comparison of combinations of parenteral artemisinin derivatives plus oral mefloquine with intravenous quinine plus oral tetracycline for treating cerebral malaria. Bull. WHO 70 (1992): 777-782.
43. Shwe, T.; and Hla, K.K. The effect of artemether plus mefloquine on Myanmar patients with complicated falciparum malaria. Southeast Asia J. Trop. Med. Pub. Hlth. 23 (Supplement 4, 1992): 117-121.
44. Nosten, F.; Luxemburger, C.; ter Kuile, F.O.; Woodrow, C., Pa Eh, J.; Chongsuphajaisiddhi, T.; and White, N.J. Treatment of Multidrug-Resistant *Plasmodium falciparum* Malaria with 3-Day Artesunate-Mefloquine Combination. J. Infect. Dis. 171 (1994): 971-977.
45. White, N.J. Preventing antimalarial drug resistance through combinations. Drug Resistance Updates 1 (1998): 3-9.
46. Avery, M.A.; Gao, F.; Chong, W.K.M.; Mehrotra, S.; and Milhous, W.K. Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 1. Synthesis and Comparative Molecular Field Analysis of C-9 Analogs of Artemisinin and 10-Deoxoartemisinin. J. Med. Chem. 36 (1993): 4264-4275.
47. Acton, N.; Karle, J.M.; and Miller, R.E. Synthesis and Antimalarial Activity of Some 9-Substituted Artemisinin Derivatives. J. Med. Chem. 36 (1993): 2552-2557.
48. Avery, M.A.; Mehrotra, S.; Bonk, J.D.; Vroman, J.A.; Goins, D.K.; and Miller, R. Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 4. Effect of Substitution at C-3. J. Med. Chem. 39 (1996): 2900-2906.
49. Avery, M.A.; Bonk, J.D.; Chong, W.K.M.; Mehrotra, S.; Miller, R.; Milhous, W.; Goins, D.K.; Venkatesan, S.; Wyandt, C.; Khan, I.; and Avery, B.A. Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 2. Effect of Heteroatom Substitution at O-11: Synthesis and Bioassay of *N*-Alkyl-11-aza-9-desmethylartemisinins. J. Med. Chem. 38 (1995): 5038-5044.

50. Lin, A.J.; Klayman, D.L.; and Milhous, W.K. Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. J. Med. Chem. 30 (1987): 2147-2150.
51. Lin, A.J.; and Miller, R.E. Antimalarial Activity of New Dihydroartemisinin Derivatives. 6. α -Alkylbenzylic Ethers. J. Med. Chem. 38 (1995): 764-770.
52. Lin, A.J.; Li, L.-Q.; Andersen, S.L.; and Klayman, D.L. Antimalarial Activity of New Dihydroartemisinin Derivatives. 5. Sugar Analogues. J. Med. Chem. 35 (1992): 1639-1642.
53. Pu, Y.M.; Torok, D.S.; Ziffer, H.; Pan, X.-Q.; and Meshnick, S.R. Synthesis and Antimalarial Activities of Several Fluorinated Artemisinin Derivatives. J. Med. Chem. 38 (1995): 4120-4124.
54. Lin, A.J.; Lee, M.; and Klayman, D.L. Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. 3. Aromatic Amine Analogues. J. Med. Chem. 33 (1990): 2610-2614.
55. Avery, M.A.; Mehrotra, S.; Johnson, T.L.; Bonk, J.D.; Vroman, J.A.; and Miller, R. Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 5. Analogs of 10-Deoxyartemisinin Substituted at C-3 and C-9. J. Med. Chem. 39 (1996): 4149-4155.
56. Meshnick, S.R.; Taylor, T.E.; and Kamchonwongpaisan, S. Artemisinin and the Antimalarial Endoperoxides: from Herbal Remedy to Targeted Chemotherapy. Microbiol. Rev. 60 (1996): 301-315.
57. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Gill, P.M.W.; Johnson, B.G.; Wong, M.W.; Foresman, J.B.; Robb, M.A.; Head-Gordon, M.; Replogle, E.S.; Gomperts, R.; Andres, J.L.; Raghavachari, K.; Binkley, J.S.; Gonzalez, C.; Martin, R.L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; and Pople, J. A. Gaussian 92/DFT, Revision F.4 [Computer Software]. Pittsburgh PA: Gaussian Inc., 1993.
58. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Gill, P.M.W.; Johnson, B.G.; Robb, M.A.; Cheeseman, J.R.; Keith, T.; Petersson, G.A.; Montgomery, J.A.; Raghavachari, K.; Al-Laham, M.A.; Zakrzewski, V.G.; Ortiz, J.V.; Foresman, J.B.; Peng, C.Y.; Ayala, P.Y.; Chen, W.; Wong, M.W.; Andres, J.L.; Replogle, E.S.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Binkley, J.S.; Defrees, D.J.; Baker, J.; Stewart, J.P.; Head-Gordon, M.; Gonzalez, C.; and

- Pople, J. A. Gaussian 94, Revision B.3 [Computer Software]. Pittsburgh PA: Gaussian Inc., 1995.
59. Kleinberg, M. L.; and Wanke, L. A. New approaches and technologies in drug design and discovery. Am. J. Health-Syst. Pharm. 52 (1995): 1323-1336.
 60. Borman, S. New QSAR Techniques Eyed for Environmental Assessments. Chem. Eng. News 68 (1990): 20-23.
 61. Crum-Brown, A.; and Fraser, T. R. On the connection between chemical constitution and physiological action I. On the physiological action of the salts of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia and nicotia. Trans. Roy. Soc. Edinburgh 25 (1868): 151-203.
 62. Richet, C. On the relationship between the toxicity and the physical properties of substances. Compt. Rend. Soc. Biol. 45 (1893): 775-776.
 63. Meyer, H. H. On the theory of alcohol narcosis I. Which property of anesthetics gives them their narcotic activity? Arch. Exper. Pathol. Pharmacol. 42 (1899): 109-118.
 64. Overton, E. Ueber die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie. Vierteljahressch. Naturforsch. Ges. Zürich 44 (1899): 88.
 65. Overton, E. Studien ueber die narkose, Fischer, Jena, Germany 1901; English translation by Lipnick, R.L. (ed.), Studies on Narcosis, Charles Ernest Overton, London: Chapman and Hall, 1991.
 66. Hansch, C. Quantitative Structure-Activity Relationships in Drug Design. In E.J. Ariëns (ed.), Drug Design, Volume I, pp. 271-342. New York: Academic Press, 1971.
 67. Hammett, L. P. Physical Organic Chemistry. 1st edition, New York: McGraw-Hill, 1940.
 68. 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.
 69. Swain, C. G.; and Lupton Jr., E. C. Field and Resonance Components of Substituent Effects. J. Am. Chem. Soc. 90 (1968): 4328-4337.

70. Dunn III, W.J. Quantitative structure-activity relationships. In C.R. Clark and W.H. Moos, *Drug Discovery Technologies*, pp. 22-43. England: Ellis Horwood, 1990.
71. Hansch, C. A Quantitative Approach to Biochemical Structure-Activity Relationships. Acc. Chem. Res. 2 (1969): 232-239.
72. Free Jr., S. M.; and Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. J. Med. Chem. 7 (1964): 395-399.
73. Hansch, C.; and Yoshimoto, M. Structure-Activity Relationships in Immunochemistry 2. Inhibition of Complement by Benzamidines. J. Med. Chem. 17 (1974): 1160-1167.
74. Wold, S.; Ruhe, A.; Wold, H.; and Dunn III, W. J. The collinearity problem in linear and nonlinear regression. The partial least squares (PLS) approach to general inverses. SIAM J. Sci. Stat. Comput. 5 (1984): 735-743.
75. Kubinyi, H. *Methods and Principles in Medicinal Chemistry. Vol. 1 QSAR: Hansch Analysis and Related Approaches.* New York: VCH Publishers, 1993.
76. Trinajstić, N. *Chemical Graph Theory. Vols. I, II.* Florida: CRC Press, 1983.
77. Wiener, H. Correlation of heats of isomerization, and difference in heats of vaporization of isomers, among the paraffin hydrocarbons. J. Am. Chem. Soc. 69 (1947): 2636.
78. 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.
79. Oxford Molecular Groups. TSAR 3.2 [Computer Software]. Oregon: Oxford Molecular Groups Inc., 1998.
80. 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.
81. Stone, M. Cross-validated choice and assessment of statistical prediction. J. Royal Stat. Soc. B 36 (1974): 111-133.

82. 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.
83. Wu, J.-A.; and Ji, R.-Y. A quantitative structure-activity study on artemisinin analogues. *Acta Pharmacol. Sinica* 3 (1982): 55-60.
84. Tang, Y.; Jiang, H. L.; Chen, K. X.; and Ji, R. Y. QSAR study of artemisinin (Qinghaosu) derivatives using neural network method. *Indian J. Chem.* 35B (1996): 325-332.
85. Jiang, H. L.; Chen, K. X.; Tang, Y.; Chen, J. Z.; Li, Y.; Wang, Q. M.; Ji, R. Y.; and Zhuang, Q. K. Theoretical and cyclic voltammetry studies on antimalarial mechanism of artemisinin (Qinghaosu) derivatives. *Indian J. Chem.* 36B (1997): 154-160.
86. Nguyen-Cong, Y.; Dang, G. V.; and Rode, B. M. Using multivariate adaptive regression splines to QSAR studies of dihydroartemisinin derivatives. *Eur. J. Med. Chem.* 31(1996): 797-803.
87. Avery, M. A.; Alvim-Gaston, M.; and Woolfrey, J. R. Synthesis and Structure-Activity Relationships of Peroxidic Antimalarials based on Artemisinin. *Adv. Med. Chem.* 4 (1999): 125-217.
88. Mayer, I. Charge, Bond Order and Valence in the Ab Initio SCF Theory. *Chem. Phys. Lett.* 97 (1983): 270-274.
89. Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* 89 (1982): 399-404.
90. Cramer III, R.D.; Petterson, 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.
91. Tripos. *Sybyl 6.5* [Computer Software]. Missouri: Tripos Inc., 1998.
92. 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.
93. Liang, J.H.; Xian, C.K.; Wu, W.H.; Yun, T.; Zhong, C.J.; and Yun, J.R. 3D-QSAR study on ether and ester analogs of artemisinin with comparative molecular field analysis. *Acta Pharmacol. Sinica* 15 (1994): 481-487.

94. Woolfrey, J.R.; Avery, M.A.; and Doweyko, A.M. Comparison of 3D quantitative structure-activity relationship methods: Analysis of the in vitro antimalarial activity of 154 artemisinin analogues by hypothetical active-site lattice and comparative molecular field analysis. J. Comput.-Aided Mol. Design 12 (1998): 165-181.
95. Singh, U.C.; and Kollman, P.A. An approach to computing electrostatic charges for molecules. J. Comput. Chem. 5 (1984) 129-145..
96. Reed, A.E.; Curtiss, L.A.; and Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor-Acceptor Viewpoint. Chem. Rev. 88 (1988): 899-926.
97. Tonmunphean, S.; Kokpol, S.; Parasuk, V.; Wolschann, P.; Winger, R.H.; Liedl, K.R.; and Rode, B.M. Comparative molecular field analysis of artemisinin derivatives: Ab initio versus semiempirical optimized structures. J. Comput.-Aided Mol. Design 12 (1998): 397-409.
98. Cho, S.J.; and Tropsha, A. Cross-Validated R^2 -Guided Region Selection for Comparative Molecular Field Analysis: A Simple Method to Achieve Consistent Results. J. Med. Chem. 38 (1995): 1060-1066.
99. Shukla, K.L.; Gund, T.M.; and Meshnick, S.R. Molecular modeling studies of the artemisinin (qinghaosu)-heim interaction: docking between the antimalarial agent and its putative receptor. J. Molec. Graphics 13 (1995): 215-222.
100. Morris, G.M.; Goodsell, D.S.; Huey, R.; and Olson, A.J. AutoDock (version 2.4) [Computer Software]. California, U.S.A, 1996.
101. Morris, G.M.; Goodsell, D.S.; Huey, R.; and Olson, A.J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. J. Comput.-Aided Mol. Des. 10 (1996): 293-304.
102. 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.
103. Mehler, E.L.; and Slomajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Eng. 4 (1991): 903-910.
104. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; and Case, D.A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. J. Comput. Chem. 7 (1986): 230-252.

105. Halgren, T.A. Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. J. Am. Chem. Soc. 114 (1992): 7827-7843.
106. <http://www.scripps.edu/pub/olson-web/doc/autoflex/parameters.html>
107. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; and Bourne, P.E. The Protein Data Bank. Nucleic Acids Research 28 (2000): 235-242.
108. Case, D.A.; Pearlman, D.A.; Caldwell, J.W.; Cheatham III, T.E.; Ross, W.S.; Simmerling, C.L.; Darden, T.A.; Merz, K.M.; Stanton, R.V.; Cheng, A.L.; Vincent, J.J.; Crowley, M.; Ferguson, D.M.; Radmer, R.J.; Seibel, G.L.; Singh, U.C.; Weiner, P.K.; and Kollman, P.A. AMBER 5 [Computer Software]. San Francisco: University of California, 1997.
109. <http://www.ccdc.cam.ac.uk>
110. Paitayatat, S.; Tarnchompoo, B.; Thebtaranonth, Y.; and Yuthavong, Y. Correlation of Antimalarial Activity of Artemisinin Derivatives with Binding Affinity with Ferroprotoporphyrin IX. J. Med. Chem. 40 (1997): 633-638.
111. Meshnick, S.R.; Thomas, A.; Ranz, A.; Xu, C.-M.; and Pan, H.-Z. Artemisinin (qinghaosu): the role of intracellular heme in its mechanism of antimalarial action. Mol. Biochem. Parasitol. 49 (1991): 181-190.
112. Meshnick, S.R.; Yang, Y.-Z.; Lima, V.; Kuypers, F.; Kamchonwongpaisan, S.; and Yuthavong, Y. Iron-Dependent Free Radical Generation from the Antimalarial Agent Artemisinin (Qinghaosu). Antimicrob. Agents Chemother. 37 (1993): 1108-1114.
113. Padmanaban, G.; and Rangarajan, P.N. Heme Metabolism of Plasmodium Is a Major Antimalarial Target. Biochem. Biophys. Res. Comm. 268 (2000): 665-668.
114. Meshnick, S.R.; Tsang, T.W.; Lin F.-B.; Pan, H.-Z.; Chang, C.-N.; Kuypers, F.; Chiu, D.; and Lubin, B. Activated oxygen mediates the antimalarial activity of qinghaosu. Prog. Clin. Biol. Res. 313 (1989): 95-104.
115. Posner, G.H.; Cummings, 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.

116. Jefford, C.W., Vicente, M.G.H., Jacquier, Y., Favarger, F., Mareda, J., MillassonSchmidt, 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.
117. Gaby, T.; Krugliak, M.; Shalmiev, G.; and Ginsburg, H. Inhibition by antimalarial drugs of haemoglobin denaturation and iron release in acidified red blood cells lysates-a possible mechanism of their antimalarial effect? Parasitology 108 (1994): 371-381.
118. Pandey, A.V.; Tekwani, B.L.; Singh, R.L.; and Chauhan, V.S. Artemisinin, an Endoperoxide Antimalarial, Disrupts the Hemoglobin Catabolism and Heme Detoxification Systems in Malarial Parasite. J. Biol. Chem. 274 (1999): 19383-19388.
119. Peter, W.; Li, Z.L.; Robinson, B.L.; and Warhurst, D.C. The chemotherapy of rodent malaria, XL. The action of artemisinin and related sesquiterpenes. Ann. Trop. Med. Parasitol. 80 (1986): 483-489.
120. Posner, G.H.; Oh, C.H.; Wang, D.; Gerena, L.; Milhous, W.K.; Meshnick, S.R.; and Asawamahasakda, 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.
121. Posner, G.H.; Wang, D.; Cumming, J.N.; Oh, C.H.; French, A.N.; Bodley, A.L.; and Shapiro, T.A. Further Evidence Supporting the Importance of and the Restrictions on a Carbon-Centered Radical for High Antimalarial Activity of 1,2,4-Trioxanes Like Artemisinin. J. Med. Chem. 38 (1995): 2273-2275.
122. Asawamahasakda, W.; Ittarat, I.; Pu, Y.-M.; Ziffer, H.; and Meshnick, S.R. Reactions of antimalarial endoperoxides with specific parasite proteins. Antimicrob. Agents. Chemother. 38 (1994): 1854-1858.
123. Haynes, R.K.; and Vonwiller, S.C. The Behavior of Qinghaosu (Artemisinin) in the Presence of Non-heme Iron(II) and (III). Tetrahedron Lett. 37 (1996): 257-260.
124. Posner, G.H.; and Oh, C.H. A Regiospecifically Oxygen-18 Labeled 1,2,4-Trioxane: A Simple Chemical Model System to Probe the Mechanism(s) for

- the Antimalarial Activity of Artemisinin (Qinghaosu). J. Am. Chem. Soc. 114 (1992): 8328-8329.
125. Butler, A.R.; Gilbert, B.C.; Hulme, P.; Irvine, L.R.; Renton, L.; and Whitwood, A.C. EPR Evidence for the Involvement of Free Radicals in the Iron-Catalysed Decomposition of Qinghaosu (Artemisinin) and Some Derivatives; Antimalarial Action of Some Polycyclic Endoperoxides. Free Rad. Res. 28 (1998): 471-476.
126. Yang, Y.-Z.; Asawamahasakda, W.; and Meshnick, S.R. Alkylation of human albumin by the antimalarial artemisinin. Biochem. Pharmacol. 46 (1993): 336-339.
127. Yang, Y.-Z.; Little, B.; and Meshnick, S.R. Alkylation of proteins by artemisinin. Effects of heme, pH, and drug structure. Biochem. Pharmacol. 48 (1994): 569-573.
128. Meshnick, S.R.; Thomas, A.; Ranz, A.; Xu, C.-M.; and Pan, H.-Z. Artemisinin (qinghaosu): the role of intracellular heme in its mechanism of antimalarial action. Mol. Biochem. Parasitol. 49 (1991): 181-189.
129. Hong, Y.-L.; Yang, Y.-Z.; and Meshnick, S.R. The interaction of artemisinin with malarial hemozoin. Mol. Biochem. Parasitol. 63 (1994): 121-128.
130. Tonmunpuean, S.; Irle, S.; Kokpol, S.; Parasuk, V.; and Wolschann, P. *Ab initio* and density functional study on singlet and triplet states of artemisinin. J. Mol. Struct. (Theochem) 454 (1998): 87-90.
131. Humbel, S.; Sieber, S.; and Morokuma, L. The IMOMO Method: Integration of Different Levels of Molecular Orbital Approximations for Geometry Optimization of Large Systems. Test for n-butane Conformation and SN2 Reaction: $\text{RCl}^+ \text{Cl}^-$. J. Chem. Phys. 105 (1996): 1959-1967.
132. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Zakrzewski, V.G.; Montgomery, Jr., J.A.; 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.; Baboul, A.G.; 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.; Chen, W.; Wong, M.W.; Andres, J.L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E.S.; and Pople, J. A. Gaussian 98, Revision A.7 [Computer Software]. Pennsylvania: Gaussian Inc., 1998.
133. Rigaudy, J. Photorearrangements of Endoperoxides. In W.M. Horspool and P.-S. Song (eds.), CRC Handbook of Photochemistry and Photobiology, pp. 325-332. Florida: CRC Press, 1995.
134. Kamata, M.; Nishikata, Y.; and Kato, M. Novel photo-rearrangements of 3,3,5-triaryl-1,2-dioxolanes: evidence for 1,5-dioxyl diradical intermediates. Chem. Commun. (1996): 2407-2408.
135. Benassi, R.; and Taddei, F. Homolytic Bond-Dissociation in Peroxides, Peroxyacids, Peroxyesters and Related Radicals: *ab-initio* MO Calculations. Tetrahedron 50 (1994): 4795-4810.
136. Adam, W. Oxygen Diradicals Derived from Cyclic Peroxides. Acc. Chem. Res. 12 (1979): 390-396.
137. Gu, J.; Chen, K.; Jiang, H.; and Leszczynski, J. A model molecule study of the O-centered and the C-centered free radical intermediates of artemisinin. J. Mol. Struct. (Theochem) 491 (1999): 57-66.
138. Gu, J.; Chen, K.; Jiang, H.; and Leszczynski, J. The Radical Transformation in Artemisinin: A DFT Study. J. Phys. Chem. A 103 (1999): 9364-9369.
139. Wu, W.-M.; Wu, Y.; Wu, Y.-L.; Yao, Z.-J.; Zhou, C.-M.; Li, Y.; and Shan, F. Unified Mechanistic Framework for the Fe(II)-Induced Cleavage of Qinghaosu and Derivatives/Analogues. The First Spin-Trapping Evidence for the Previously Postulated Secondary C-4 Radical. J. Am. Chem. Soc. 120 (1998): 3316-3325.
140. Tonmunphean, S.; Parasuk, V.; and Kokpol, S. QSAR Study of Antimalarial Activities and Artemisinin-Heme Binding Properties Obtained from Docking Calculations. Quant. Struct.-Act. Rel. 19 (2000): 475-483.
141. Tonmunphean, S.; Parasuk, V.; and Kokpol, S. Automated docking calculations of artemisinin to heme. J. Mol. Model. (2001): in press.



APPENDICES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix A

Comparative molecular field analysis of artemisinin derivatives: *Ab initio* versus semiempirical optimized structures

Somsak Tonmunphean, Sirirat Kokpol,
Vudhichai Parasuk, Peter Wolschann,
Rudolf H. Winger, Klaus R. Liedl, Bernd M. Rode

J. Comput.-Aided Mol. Des. 12 (1998): 397-409.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Comparative molecular field analysis of artemisinin derivatives: Ab initio versus semiempirical optimized structures

Somsak Tonmuphean^a, Sirirat Kokpol^a, Vudhichai Parasuk^a, Peter Wolschann^b, Rudolf H. Winger^{c,*}, Klaus R. Liedl^c & Bernd M. Rode^c

^aDepartment of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand; ^bInstitute of Theoretical Chemistry and Radiation Chemistry, University of Vienna, Waehringer Strasse 17, A-1090 Wien, Austria; ^cDepartment of Theoretical Chemistry, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a, A-6020 Innsbruck, Austria

Received 13 August 1997; Accepted 22 January 1998

Key words: AM1 method, antimalarial drug, CoMFA, Hartree-Fock (HF), QSAR

Abstract

Based on the belief that structural optimization methods, producing structures more closely to the experimental ones, should give better, i.e. more relevant, steric fields and hence more predictive CoMFA models, comparative molecular field analyses of artemisinin derivatives were performed based on semiempirical AM1 and HF/3-21G optimized geometries. Using these optimized geometries, the CoMFA results derived from the HF/3-21G method are found to be usually but not drastically better than those from AM1. Additional calculations were performed to investigate the electrostatic field difference using the Gasteiger and Marsili charges, the electrostatic potential fit charges at the AM1 level, and the natural population analysis charges at the HF/3-21G level of theory. For the HF/3-21G optimized structures no difference in predictability was observed, whereas for AM1 optimized structures such differences were found. Interestingly, if ionic compounds are omitted, differences between the various HF/3-21G optimized structure models using these electrostatic fields were found.

Introduction

Malaria is one of the most widespread and prevalent endemic diseases, a risk for 40% of the world's population, which causes around 300 to 500 million illnesses and around 1.5 to 2.7 million deaths each year [1,2]. In many parts of the world, malarial parasites have developed resistance to most of the common chemotherapeutic agents, such as chloroquine, quinine, sulfa/pyrimethamine combination, and mefloquine [3-5]. This led to the introduction of artemisinin, a sesquiterpene endoperoxide compound isolated from a Chinese herb in 1972 [6]. However, its poor solubility stimulated scientists to synthesize more soluble and active derivatives such as dihydroartemisinin, artesunate, and artemether [7]. These derivatives are now being increasingly used for

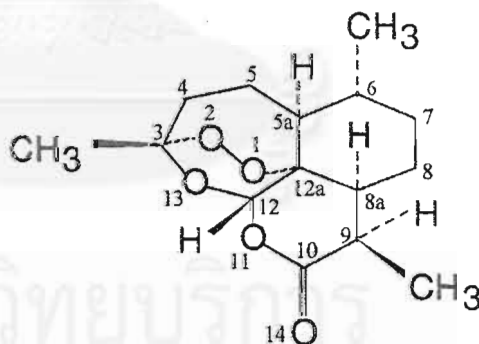


Figure 1. Stereochemistry and atomic numbering scheme of artemisinin.

malaria treatment due to their low toxicity and high potency against drug-resistant strains of *Plasmodium falciparum* [8-10]. Nevertheless, there is still a need for new and more potent antimalarial drugs.

*To whom correspondence should be addressed.

Table 1. Comparison of the most important structural parameters of artemisinin between X-ray structure, semiempirical, and ab initio optimized structures

Structural parameter	X-ray	CNDO	AM1	HF/STO-3G	HF/3-21G	HF/6-31G(d,p)
<i>Bond length (Å)</i>						
O1-O2	1.475	1.229	1.289	1.397	1.462	1.390
O1-C12a	1.450	1.410	1.468	1.467	1.477	1.430
O2-C3	1.417	1.394	1.447	1.450	1.441	1.396
O11-C12	1.455	1.401	1.421	1.442	1.428	1.408
C10-O14	1.201	1.340	1.231	1.214	1.197	1.183
<i>Bond angle (deg)</i>						
O1-C12a-C5a	106.4	107.4	106.0	105.9	105.1	105.6
O2-C3-C4	113.2	113.1	111.6	111.6	111.0	111.3
O11-C12-O13	105.9	105.5	101.0	105.8	108.8	107.7
O14-C10-O11	117.2	112.1	111.6	119.2	119.2	118.9
<i>Torsional angle (deg)</i>						
C12-O13-C3-C4	272.7	283.9	281.4	266.5	271.7	269.7
O1-C12a-C5a-C5	69.0	73.7	73.2	67.8	68.8	67.5
O2-C3-C4-C5	265.8	273.3	269.7	264.8	263.3	267.2
O11-C12-O13-C3	258.1	241.5	245.7	267.7	262.1	260.7

For drug discovery and development, quantitative structure-activity relationships (QSARs) are often employed because the bioactivity of new compounds can be predicted with strongly reduced efforts, and thus time and money can be saved [11]. Comparative Molecular Field Analysis (CoMFA), a three-dimensional quantitative structure-activity relationship (3D-QSAR) method [12], is based on the assumption that drug-receptor interactions, which are responsible for biological activities, are usually non-covalent. Therefore, the steric and electrostatic fields (noncovalent interactions) surrounding the analogues of drug molecules should correlate with their biological activities. From geometry optimization of artemisinin structures (Figure 1) with semiempirical (CNDO and AM1 methods) and ab initio molecular orbital methods (Hartree-Fock with STO-3G, 3-21G, and 6-31G(d,p) basis sets) we found that the ab initio level, using Hartree-Fock theory with the 3-21G basis set (HF/3-21G), reproduced most of the structural parameters very reliably in comparison to the X-ray structures [13] (see Table 1). This applies especially to the bond length of the endoperoxide linkage, which is held responsible for the antimalarial activity [7-9] and which is believed to be attacked by ferrous ions (Fe^{2+}) and then to be broken to form an oxy-radical in the key step of the mechanism of action [14]. The 1,2,4-trioxane compounds show that bulky groups at the C_4

position, which block the ferrous ions from attacking the peroxide linkage, significantly decrease the biological activity [15]. Therefore, structural optimization methods that give structures more closely to the experimental ones should give better, i.e. more relevant, steric fields and hence a better, i.e. more predictive, CoMFA model. This suggested the choice of the HF/3-21G level of theory for the optimization. In most of the CoMFA studies, molecules have been geometry optimized with empirical or semiempirical quantum mechanical methods [12,16-21]. For this reason, we additionally used the AM1 method to optimize the structures in order to compare these results with those obtained at the HF/3-21G level of theory.

Biological data

The structures and biological data are listed in Figure 2 and Table 2. Because the biological data arise from different sources [22-26], the relative activity, IC_{50} of artemisinin over IC_{50} of the compounds, was used to reduce inconsistencies due to individual experimental environments. The activities were measured against two strains of *Plasmodium falciparum*, i.e. the W-2 clone and the D-6 clone. The W-2 clone is chloroquine-resistant but mefloquine-sensitive, while

Table 2. Experimental biological activities and values predicted from the HF/3-21G optimized structures

No.	log(relative activity)					
	D-6 clone			W-2 clone		
	Expt.	Cal. ^a	Residual	Expt.	Cal. ^a	Residual
1	0.854	0.737	0.117	-0.019	0.364	-0.383
2	0.689	0.666	0.023	0.405	0.494	-0.089
3	0.202	0.221	-0.019	0.013	0.021	-0.008
4	0.580	0.531	0.049	0.251	0.203	0.048
5	-1.264	-1.266	0.002	-1.599	-1.599	0.000
6	-1.463	-1.353	-0.110	-1.547	-1.610	0.063
7	-1.411	-1.451	0.040	-1.202	-1.238	0.035
8	0.226	0.238	-0.012	-0.144	-0.154	0.010
9	-0.786	-0.656	-0.130	-1.086	-0.972	-0.114
10	-0.786	-0.937	0.151	-0.014	-0.101	0.087
11	-1.139	-1.121	-0.018	-0.290	-0.329	0.039
12	-1.666	-1.701	0.035	-0.960	-0.981	0.021
13	0.423	0.252	0.171	0.529	0.115	0.414
14	-0.122	-0.100	-0.022	-0.276	-0.393	0.117
15	0.375	0.411	-0.036	0.163	0.145	0.018
16	0.904	0.907	-0.003	1.034	0.926	0.108
17	0.655	0.785	-0.130	0.651	0.786	-0.134
18	0.199	-0.119	0.318	-0.053	-0.254	0.201
19	0.667	0.691	-0.024	0.677	0.679	-0.002
20	0.700	0.722	-0.022	0.849	0.888	-0.039
21	0.612	0.685	-0.073	0.645	0.677	-0.032
22	0.971	1.001	-0.030	1.319	1.240	0.079
23	0.522	0.541	-0.019	0.481	0.445	0.036
24	-0.399	-0.277	-0.122	-0.398	-0.423	0.025
25	-0.105	-0.121	-0.016	-0.084	-0.037	-0.467
26	-0.094	0.078	-0.172	-0.123	-0.068	-0.055
27	0.079	0.098	-0.019	0.146	0.037	0.109
28	0.146	0.202	-0.056	0.079	0.100	-0.021
29	0.255	0.354	-0.099	0.176	0.199	-0.023
30	-0.824	-0.985	0.161	-0.854	-1.058	0.204
31	-0.347	-0.245	-0.102	-0.523	-0.230	-0.293
32	-1.097	-0.826	-0.271	-1.222	-0.852	-0.370
33	0.423	0.210	0.213	0.209	0.136	0.074
34	0.411	0.460	-0.049	0.782	0.599	0.182
35	-0.018	-0.153	0.135	-0.260	-0.296	0.036
36	-1.052	-1.062	0.010	-1.830	-1.754	-0.076
37	0.174	0.115	0.059	0.234	0.367	-0.133
38	1.302	1.191	0.111	0.807	0.968	-0.161
39	0.186	0.184	0.002	0.331	0.339	-0.008
40	0.841	0.917	-0.076	0.776	0.697	0.078

^a Calculated from alignment 4 with O.sp³ (-1.0) as the probe atom using the HF/3-21G optimized structures (Table 7).

the D-6 clone is mefloquine-resistant but chloroquine-sensitive.

Table 3. Atoms selected for the definition of alignment rules

Alignment	Selected atoms
1	O1-O2-C3-C4-C5-C5a-C12a-C12-O13
2	C5a-C6-C7-C8-C8a-C12a
3	O1-O2-O11
4	C12a-O1-O2-C3
5	C12a-O1-O2-C3-O13-C12-O11

Computational methods

Structural optimization

All structures were built using SYBYL 6.3 [27] and optimized at the HF/3-21G and AM1 level of theory using GAUSSIAN 94 [28]. The quantum mechanically optimized structures with their Mulliken population analysis (MPA) atomic charges were reimported into SYBYL. The charges according to Gasteiger and Marsili [29], the electrostatic potential fit (ESPFIT) charges [30] at the AM1 level, and the natural population analysis (NPA) charges [31] at the HF/3-21G level were calculated using SYBYL, MOPAC 6.0 [32], and GAUSSIAN 94, respectively.

Alignment rules

Five alignment rules were selected for this work in order to study the influence of different alignments. We also tested the alignment of Ref. 17 (denoted as alignment 3 in the following). For all alignments, the structures were adjusted using the 'Fit Atom' option in SYBYL which minimizes the root-mean-square (rms) differences between selected atoms of the structure in question and dihydroartemisinin. Dihydroartemisinin was chosen as the reference molecule due to its most common structure within the analogues (see Table 3). Data (structures used) are available upon request at the following e-mail address: Rudolf.Winger@uibk.ac.at.

CoMFA calculations

A regular 3D lattice with 2 Å spacing was created extending beyond the molecular dimensions of molecule number 22 by 4.0 Å in all directions. Two probe atoms, namely an sp³ carbon with a charge of +1.0 and an sp³ oxygen with a charge of -1.0, were used. Steric (Lennard-Jones 6-12 function) and electrostatic (Coulombic) interactions were calculated using the Tripos force field [33] with a distance-dependent dielectric constant. The cutoff was set to 30 kcal/mol. An

Table 4. CoMFA results for the 32 dihydroartemisinin analogues optimized with AM1

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.675	3	0.932	0.657	3	0.921	0.678	3	0.929
	2	0.665	3	0.917	0.667	4	0.947	0.627	3	0.899
	3	0.648	5	0.964	0.638	5	0.978	0.653	5	0.977
	4	0.619	4	0.940	0.609	5	0.977	0.652	5	0.976
	5	0.624	3	0.913	0.709 ^b	5	0.977	0.604	3	0.899
W-2	1	0.580	2	0.768	0.599	2	0.786	0.580	2	0.774
	2	0.587	2	0.762	0.583	2	0.773	0.594	2	0.775
	3	0.586	2	0.755	0.567	2	0.764	0.609 ^b	5	0.960
	4	0.576	2	0.750	0.563	2	0.762	0.564	2	0.754
	5	0.598	2	0.777	0.586	2	0.772	0.581	2	0.764

^aNumber of components.^bThe best model.

Table 5. CoMFA results for the 32 dihydroartemisinin analogues optimized with HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.698	4	0.926	0.703	3	0.932	0.679	3	0.932
	2	0.700	3	0.926	0.683	3	0.926	0.679	3	0.930
	3	0.647	3	0.923	0.706	3	0.933	0.692	3	0.932
	4	0.649	3	0.921	0.730 ^b	3	0.940	0.681	3	0.934
	5	0.699	4	0.965	0.711	3	0.935	0.683	3	0.931
W-2	1	0.606	2	0.812	0.646	2	0.836	0.623	2	0.825
	2	0.585	2	0.803	0.654	2	0.840	0.632	2	0.824
	3	0.591	2	0.808	0.648	2	0.838	0.630	2	0.825
	4	0.599	2	0.809	0.686 ^b	3	0.927	0.637	2	0.831
	5	0.607	2	0.810	0.649	2	0.837	0.633	2	0.827

^{a,b}As in Table 4

Table 6. CoMFA results for the 40 artemisinin analogues optimized with AM1

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.663 ^b	4	0.948	0.646	4	0.944	0.661	4	0.946
	2	0.656	4	0.928	0.626	4	0.932	0.631	4	0.929
	3	0.573	4	0.922	0.559	4	0.933	0.608	5	0.966
	4	0.577	4	0.923	0.545	4	0.935	0.598	5	0.962
	5	0.563	4	0.925	0.592	4	0.926	0.576	4	0.918
W-2	1	0.480	4	0.908	0.515	4	0.920	0.502	4	0.921
	2	0.548 ^b	4	0.912	0.487	4	0.911	0.497	4	0.907
	3	0.490	4	0.899	0.377	4	0.897	0.468	5	0.940
	4	0.477	4	0.902	0.381	4	0.901	0.403	4	0.888
	5	0.493	4	0.909	0.465	4	0.906	0.477	4	0.907

^{a,b}As in Table 4

Table 7. CoMFA results for the 40 artemisinin analogues optimized with HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.610	4	0.953	0.708	5	0.977	0.662	5	0.978
	2	0.640	3	0.925	0.679	5	0.975	0.630	3	0.922
	3	0.569	4	0.944	0.709	5	0.979	0.668	5	0.976
	4	0.592	4	0.954	0.723 ^b	5	0.979	0.673	5	0.980
	5	0.602	4	0.946	0.714	5	0.980	0.658	5	0.976
W-2	1	0.528	6	0.972	0.578	5	0.963	0.434	3	0.889
	2	0.494	4	0.935	0.551	5	0.970	0.458	3	0.884
	3	0.436	4	0.921	0.583 ^b	5	0.961	0.479	4	0.930
	4	0.476	4	0.931	0.581	5	0.963	0.515	5	0.960
	5	0.556	6	0.970	0.572	5	0.961	0.485	4	0.934

a,b As in Table 4

Table 8. CoMFA results for the 32 dihydroartemisinin analogues optimized with AM1 and adjusted to be in the same orientation as those resulting from HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.637	4	0.944	0.718 ^b	5	0.980	0.639	4	0.946
	2	0.696	5	0.974	0.702	5	0.976	0.640	3	0.904
	3	0.631	4	0.945	0.713	5	0.980	0.608	3	0.905
	4	0.650	4	0.945	0.710	5	0.979	0.642	4	0.947
	5	0.682	4	0.954	0.707	5	0.980	0.610	3	0.907
W-2	1	0.587	2	0.763	0.605	2	0.780	0.600	2	0.775
	2	0.559	2	0.761	0.613 ^b	2	0.791	0.608	2	0.777
	3	0.572	2	0.769	0.598	2	0.782	0.602	2	0.776
	4	0.592	2	0.766	0.609	2	0.785	0.599	2	0.775
	5	0.580	2	0.773	0.600	2	0.780	0.601	2	0.776

a,b As in Table 4

Table 9. CoMFA results for the 40 artemisinin analogues optimized with AM1 and adjusted to be in the same orientation as those resulting from HF/3-21G

Activity	Alignment	AutoCoMFA			O.sp ³ (-1.0)			C.sp ³ (+1.0)		
		q ²	noc ^a	r ²	q ²	noc	r ²	q ²	noc	r ²
D-6	1	0.570	4	0.925	0.596	4	0.929	0.593	4	0.926
	2	0.566	5	0.953	0.583	6	0.974	0.583	5	0.941
	3	0.497	4	0.892	0.535	4	0.900	0.529	4	0.902
	4	0.567	4	0.925	0.589	4	0.929	0.642 ^b	4	0.947
	5	0.585	4	0.929	0.613	4	0.929	0.618	4	0.932
W-2	1	0.497	4	0.911	0.509	4	0.909	0.491	4	0.906
	2	0.376	4	0.866	0.445	4	0.853	0.442	4	0.849
	3	0.341	4	0.872	0.397	4	0.876	0.382	4	0.877
	4	0.472	4	0.908	0.506	4	0.909	0.488	4	0.906
	5	0.490	4	0.910	0.521 ^b	4	0.909	0.505	4	0.910

a,b As in Table 4

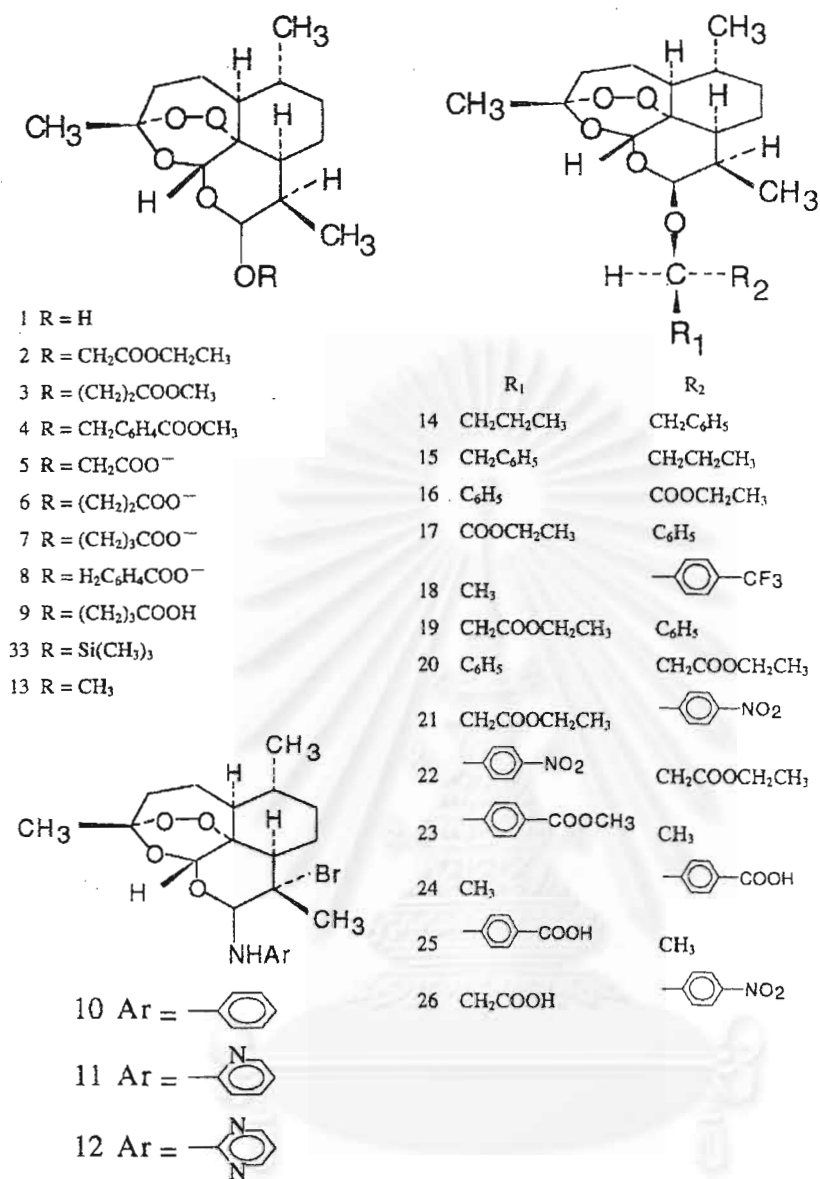
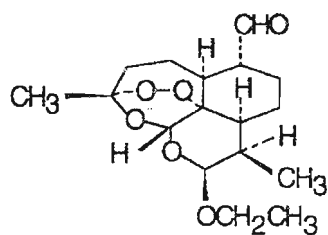


Figure 2. Structures of artemisinin derivatives used for the CoMFA studies.

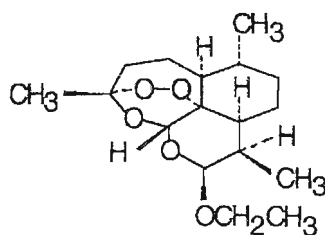
additional 'AutoCoMFA' column was created by using the automatically created region, which extends the region beyond the molecular dimension of the largest compound by 4.0 Å in all directions and uses C.sp³ with a charge of +1.0 as the probe atom, in order to see the influence of grid box position and dimension.

Partial least-squares regression analysis

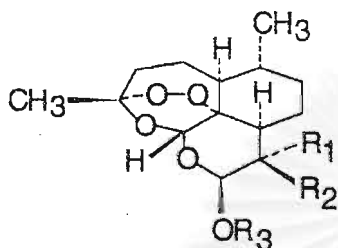
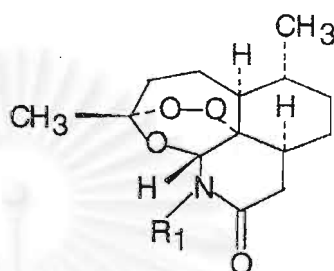
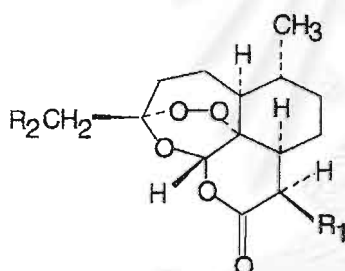
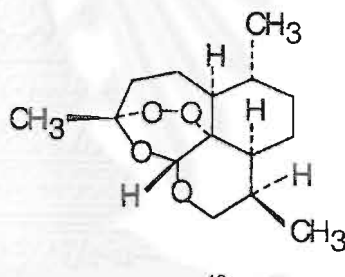
All models were investigated using the full cross-validated partial least-squares (PLS) method [34,35] (leave-one-out) with CoMFA standard options for scaling of variables. Minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation is below this threshold. To avoid an excessive number of components, the optimal number



27



28

29 $R_1 = \text{OH}$, $R_2 = \text{CH}_3$, $R_3 = \text{CH}_2\text{CF}_3$ 30 $R_1 = \text{CH}_3$, $R_2 = \text{OH}$, $R_3 = \text{CH}_2\text{CF}_3$ 31 $R_1 = \text{OH}$, $R_2 = \text{CH}_3$, $R_3 = \text{CH}_2\text{CH}_3$ 32 $R_1 = \text{CH}_3$, $R_2 = \text{OH}$, $R_3 = \text{CH}_2\text{CH}_3$ 34 $R = \text{CH}_3$ 35 $R = n\text{-C}_5\text{H}_{11}$ 36 $R = (\text{CH}_2)_5\text{COOH}$ 37 $R = \text{CH}_2\text{C}_6\text{H}_5$ 38 $R_1 = \text{H}$, $R_2 = \text{CH}_2\text{CH}_3$ 39 $R_1 = (\text{CH}_2)_3\text{CH}_3$, $R_2 = \text{CH}_3$ 

40

Figure 2. Continued.

of components (noc) was selected as the one which results in an increase of the cross-validated r^2 (also called r_{cv}^2 or q^2), which indicates the predictive ability of the CoMFA model, of more than 5% [36] compared to the model with fewer components. Subsequently, it was used for the PLS without cross-validation and with scaling of variables.

Two sets of compounds were chosen: the dihydroartemisinin analogues (compounds 1–32 that have an -OR or -NR group at the C10 position) and the

artemisinin analogues (compounds 1–40 that have either an -OR or -NR or =O group at the C10 position).

Results and discussion

The CoMFA results obtained using $O.sp^3$ (-1.0) and $C.sp^3$ (+1.0) as probe atoms and the five alignment rules for the dihydroartemisinin analogues (32 compounds) optimized at the AM1 and HF/3-21G level of theory are listed in Tables 4 and 5, respectively.

Table 10. CoMFA results for the 32 dihydroartemisinin and 40 artemisinin analogues optimized with AM1 using O.sp³ with charge -1.0 as the probe atom and alignment 4 and translated grid position along the x, y, and z axes showing the q² value and the optimum number of components in parentheses

Activity	Set ^a	x-axis				y-axis				z-axis			
		+0.5	-0.5	+1.0	-1.0	+0.5	-0.5	+1.0	-1.0	+0.5	-0.5	+1.0	-1.0
D-6	1	0.585(3)	0.610(5)	0.670(4)	0.674(5)	0.531(5)	0.687(5)	0.458(2)	0.458(2)	0.676(4)	0.607(3)	0.618(5)	0.618(5)
	2	0.579(4)	0.557(4)	0.658(4)	0.645(4)	0.462(4)	0.592(4)	0.421(4)	0.421(4)	0.641(4)	0.603(4)	0.603(5)	0.603(5)
W-2	1	0.614(2)	0.592(2)	0.594(2)	0.580(2)	0.594(2)	0.630(4)	0.570(2)	0.570(2)	0.601(2)	0.627(3)	0.660(5)	0.660(5)
	2	0.536(4)	0.497(4)	0.585(4)	0.577(4)	0.399(4)	0.504(4)	0.501(6)	0.501(6)	0.545(5)	0.521(4)	0.611(5)	0.611(5)

^a 1: dihydroartemisinin analogues (32 compounds);

2: artemisinin analogues (40 compounds).

Table 11. CoMFA results for the 32 dihydroartemisinin and 40 artemisinin analogues optimized with HF/3-21G using O.sp³ with charge -1.0 as the probe atom and alignment 4 and translated grid position along the x, y, and z axes showing the q² value and the optimum number of components in parentheses

Activity	Set ^a	x-axis				y-axis				z-axis			
		+0.5	-0.5	+1.0	-1.0	+0.5	-0.5	+1.0	-1.0	+0.5	-0.5	+1.0	-1.0
D-6	1	0.723(3)	0.682(2)	0.714(3)	0.721(3)	0.674(3)	0.700(3)	0.610(3)	0.610(3)	0.706(3)	0.625(3)	0.655(3)	0.655(3)
	2	0.655(3)	0.628(5)	0.646(4)	0.656(4)	0.554(3)	0.677(5)	0.521(4)	0.524(4)	0.628(3)	0.612(5)	0.632(5)	0.632(5)
W-2	1	0.612(2)	0.654(2)	0.610(2)	0.616(2)	0.616(2)	0.659(3)	0.584(2)	0.584(2)	0.670(3)	0.613(2)	0.647(3)	0.647(3)
	2	0.533(5)	0.389(2)	0.477(4)	0.460(3)	0.396(3)	0.598(7)	0.418(4)	0.424(4)	0.526(3)	0.517(5)	0.494(5)	0.494(5)

^a 1: dihydroartemisinin analogues (32 compounds);

2: artemisinin analogues (40 compounds).

With AM1, there are slight differences in q² when comparing the five alignment rules. The reason may be that the backbone structures in all compounds are nearly the same. Alignment 3, which is the one from Ref. 17, also gave acceptable results, although all the structures in Ref. 17 have =O at position C10 while our compounds are dihydroartemisinin analogues with -OR at C10. With HF/3-21G, there is also a slight difference in q² when comparing the five alignment rules. Comparing AM1 and HF/3-21G optimized structures, some significant differences can be observed: for example, for D-6 activity using alignment 4 with O.sp³ as the probe atom, HF/3-21G gives q² = 0.730 (noc = 3) compared to 0.609 (noc = 5) from AM1, and for W-2 activity using the same alignment and probe atom, HF/3-21G gives q² = 0.686 (noc = 3) compared to 0.563 (noc = 2) resulting from AM1. However, the comparison of the highest q² resulting from AM1 for D-6 (0.709, noc = 5) and W-2 (0.609, noc = 5) and from HF/3-21G for D-6 (0.730, noc = 3) and W-2 (0.686, noc = 3) shows that HF/3-21G yielded only slightly better results.

The CoMFA results obtained using O.sp³ (-1.0) and C.sp³ (+1.0) as probe atoms and the five align-

ment rules for the artemisinin analogues (40 compounds) optimized at the AM1 and HF/3-21G level of theory are listed in Tables 6 and 7, respectively. Comparing the five alignment rules for both AM1 and HF/3-21G results, larger differences in q² are observed than for the dihydroartemisinin analogues. For W-2 activity using AM1 and O.sp³ as the probe atom, alignment 3 gives q² = 0.377 (noc = 4) compared to 0.515 (noc = 4) of alignment 1 and for D-6 activity using AutoCoMFA, alignment 5 gives q² = 0.563 (noc = 4) compared to 0.663 (noc = 4) of alignment 1. The reason may be that the geometry of the lactone ring for the artemisinin analogues, with =O at C10, and for the dihydroartemisinin analogues, with -OR at that position, is quite different. As in the case of the dihydroartemisinin analogues, HF/3-21G optimizations usually give better results than AM1 optimizations. For D-6 activity, using alignment 4 with O.sp³ as the probe atom, HF/3-21G gives q² = 0.723 (noc = 5) in comparison to 0.545 (noc = 4) for AM1, and for W-2 activity, using the same alignment and probe atom, HF/3-21G gives q² = 0.581 (noc = 5) in comparison to 0.381 (noc = 4) of AM1. Comparing the highest q² resulting from HF/3-21G and AM1, the difference

Table 12. Predicted and measured activities of five additional dihydroartemisinin analogues from HF/3-21G (MPA, NPA, AM1, and Gasteiger charges) and AM1 optimized structures using the best corresponding alignment rule for each model

Activity	Set ^a	Steric Field	Electrostatic Field	Compound no.				
				41	42	43	44	45
D-6	Measured	—	—	0.0739	0.3130	-1.6162	-0.6499	0.0584
	1	HF/3-21G	HF/3-21G MPA	0.8850	0.4107	-0.5078	-0.2603	-0.0377
		HF/3-21G	HF/3-21G NPA	0.9136	0.4190	-0.4264	-0.2048	-0.0302
		HF/3-21G	AM1	0.8091	0.5036	-0.3964	-0.2149	0.0379
		HF/3-21G	Gasteiger	0.6944	0.4686	-0.3914	-0.2006	0.1019
		AM1	AM1	1.1251	0.9170	-0.2673	-0.0794	0.5192
	2	HF/3-21G	HF/3-21G MPA	1.0587	0.3936	-0.6243	-0.4481	-0.0733
		HF/3-21G	HF/3-21G NPA	1.0413	0.3423	-0.5477	-0.3925	-0.0650
		HF/3-21G	AM1	0.9060	0.4347	-0.5043	-0.4411	-0.0231
		HF/3-21G	Gasteiger	1.0840	0.6225	-0.3795	-0.2970	0.1416
AM1		AM1	1.1563	0.7775	-0.3223	-0.2719	0.4504	
W-2	Measured	—	—	0.5538	0.3411	-1.6999	-1.7189	-0.0813
	1	HF/3-21G	HF/3-21G MPA	0.8500	0.5505	-0.4057	-0.0978	0.2128
		HF/3-21G	HF/3-21G NPA	0.8180	0.3954	-0.1821	0.0947	0.0817
		HF/3-21G	AM1	0.7807	0.5820	-0.2582	-0.0117	0.1816
		HF/3-21G	Gasteiger	0.7232	0.6093	-0.0609	0.1257	0.2119
		AM1	AM1	0.8700	0.5964	0.0415	0.1199	0.2538
	2	HF/3-21G	HF/3-21G MPA	1.1120	0.5649	-0.6100	-0.3382	0.3549
		HF/3-21G	HF/3-21G NPA	1.1489	0.5582	-0.5607	-0.2716	0.2524
		HF/3-21G	AM1	1.0705	0.7389	-0.4692	-0.3184	0.2250
		HF/3-21G	Gasteiger	1.1359	0.7177	-0.3815	-0.2012	0.2633
AM1		AM1	1.0071	0.8233	-0.2255	-0.1491	0.3351	

^aAs in Table 10.

between D-6 activity, 0.723 (HF, noc = 5) compared to 0.663 (AM1, noc = 4), and W-2 activity, 0.583 (HF, noc = 5) compared to 0.548 (AM1, noc = 4), is nearly the same. The comparison between dihydroartemisinin analogues and artemisinin analogues reveals that the latter give slightly worse results.

It is well known that the orientation of the compounds within the predefined grid significantly affects the CoMFA results [12,37]. Therefore, we adjusted the AM1 structures to be in the same orientation as those resulting from HF/3-21G and recalculated the CoMFA. The CoMFA results using O.sp³ (-1.0) and C.sp³ (+1.0) as probe atoms and the five alignment rules of dihydroartemisinin and artemisinin analogues which were optimized at the AM1 level of theory and then oriented in the same way as the HF/3-21G optimized structures are shown in Tables 8 and 9, respectively. Interestingly, some results for the dihydroartemisinin analogues are remarkably improved,

especially those using O.sp³ as the probe atom. For D-6 activity using O.sp³ as the probe atom in alignment 4, q² rose from 0.609 (noc = 5) to 0.710 (noc = 5) and for W-2 activity with the same probe atom and alignment q² rose from 0.563 (noc = 2) to 0.609 (noc = 2). In contrast to the dihydroartemisinin analogues, the results for artemisinin analogues are often worse than the original ones. For W-2 activity with AutoCoMFA in alignment 2, q² decreased from 0.548 (noc = 4) to 0.376 (noc=4); on the other hand in some models the results were improved, e.g. for W-2 activity using alignment 4 with O.sp³ as the probe atom, where q² rose from 0.381 (noc = 4) to 0.506 (noc = 4).

We also translated the grid in the AM1 models by ±0.5 and ±1.0 Å in the x, y, and z directions. The CoMFA results using O.sp³ as the probe atom and alignment 4 of both the dihydroartemisinin and artemisinin analogues, which were optimized with AM1 and HF/3-21G, are shown in Tables 10 and 11,

Table 13. Comparison of different electrostatic fields using HF/3-21G and AM1 optimized structures with alignment 4 of 40 artemisinin analogues showing the q^2 value, the optimum number of components in parentheses and the steric contribution

Optimization method			HF/3-21G				AM1			
Charges method			HF/MPA	HF/NPA	AM1/MPA	GM ^a	AM1/MPA	GM ^a		
Activity	Type	Field								
D-6	Auto	Both	0.592 (4)	0.577 (4)	0.586 (4)	0.645 (5)	0.567 (4)	0.255 (4)		
		Steric	0.495 (4)	0.495 (4)	0.495 (4)	0.495 (4)	0.239 (3)	0.239 (3)		
		Electrostatic	0.596 (4)	0.579 (4)	0.547 (4)	0.417 (4)	0.604 (2)	0.060 (1)		
	O.sp ³	sc ^b		0.604	0.597	0.602	0.621	0.538	0.605	
			Both	0.723 (5)	0.718 (5)	0.725 (5)	0.702 (5)	0.589 (4)	0.345 (5)	
		Steric		0.509 (2)	0.509 (2)	0.509 (2)	0.509 (2)	0.302 (3)	0.302 (3)	
			Electrostatic	0.236 (4)	0.263 (4)	0.566 (4)	0.568 (6)	0.613 (3)	0.072 (1)	
		C.sp ³	sc ^b		0.562	0.565	0.568	0.603	0.529	0.591
			Both	0.673 (5)	0.619 (4)	0.659 (5)	0.644 (5)	0.590 (4)	0.344 (5)	
	Steric		0.452 (2)	0.452 (2)	0.452 (2)	0.452 (2)	0.352 (4)	0.352 (4)		
	W-2	Auto	Electrostatic	0.236 (4)	0.263 (4)	0.566 (4)	0.568 (6)	0.613 (3)	0.072 (1)	
			sc ^b		0.579	0.582	0.574	0.596	0.518	0.610
Both			0.476 (4)	0.503 (7)	0.463 (6)	0.443 (5)	0.427 (4)	0.088 (4)		
O.sp ³		Steric	0.314 (4)	0.314 (4)	0.314 (4)	0.314 (4)	0.057 (1)	0.057 (1)		
		Electrostatic	0.340 (2)	0.352 (2)	0.354 (1)	0.398 (5)	0.480 (2)	-0.030 (4)		
		sc ^b		0.561	0.540	0.566	0.551	0.491	0.575	
C.sp ³	Both	0.581 (5)	0.550 (5)	0.521 (4)	0.487 (5)	0.506 (4)	0.224 (5)			
	Steric	0.290 (4)	0.290 (4)	0.290 (4)	0.290 (4)	0.092 (2)	0.092 (2)			
	Electrostatic	0.347 (1)	0.328 (1)	0.416 (2)	0.433 (4)	0.477 (2)	0.064 (4)			
C.sp ³	sc ^b		0.520	0.515	0.523	0.539	0.500	0.573		
		Both	0.515 (5)	0.512 (6)	0.517 (6)	0.453 (5)	0.488 (4)	0.216 (5)		
	Steric		0.283 (4)	0.283 (4)	0.283 (4)	0.283 (4)	0.088 (1)	0.088 (1)		
		Electrostatic	0.347 (1)	0.328 (1)	0.416 (2)	0.433 (4)	0.477 (2)	0.064 (4)		
sc ^b		0.536	0.536	0.525	0.544	0.498	0.575			

^a Gasteiger and Marsili charges.

^b Steric contribution.

respectively. Although the results show a strong influence of the CoMFA study on the grid position, HF/3-21G usually gives better results than AM1.

To check whether the addition of a smaller gridded box in a specific region significantly improves the results, we performed a CoMFA calculation of both HF/3-21G and AM1 optimized structures with an O.sp³ probe atom using an additional small grid of dimensions 10 × 9 × 4 Å and a grid spacing of 0.6 Å. Although there is slight improvement of the HF model, the enormous cost does not justify the use of this smaller grid.

To test the validity of the CoMFA models, we predicted the bioactivity of five additional compounds [38] using the best alignment rules of the dihydroartemisinin and artemisinin analogues. The structures, logarithmic relative biological activities and

results are shown in Figure 3 and Table 12, and they confirm that HF/3-21G models give better results than those based on AM1.

To study the influence of atomic charges, we performed CoMFA studies using either only the steric or the electrostatic field interaction with alignment 4 (data not shown). Interestingly, the AM1 method using only electrostatic field interaction gives slightly better results than the use of both fields. This is in agreement with some previous reports [21]. The HF/3-21G method using both fields gives, however, better results than the use of only one field.

In order to investigate further which field is responsible for the difference between the HF/3-21G and AM1 results, we performed single point calculations to obtain AM1 charges using HF/3-21G optimized structures and then did the CoMFA using alignment 4

Table 14. Comparison of different electrostatic fields using HF/3-21G optimized structures with alignment 4 of 36 artemisinin analogues showing the q^2 value, the optimum number of components in parentheses and the steric contribution

Activity	Type	Field	HF/MPA	HF/NPA	AM1/MPA	AM1/ESPFIT	GM ^a
D-6	Auto	Both	0.580 (5)	0.602 (5)	0.509 (5)	0.489 (5)	0.594 (4)
		Steric	0.609 (4)	0.609 (4)	0.609 (4)	0.609 (4)	0.609 (4)
		Electrostatic	0.633 (3)	0.631 (3)	0.516 (3)	0.414 (2)	0.439 (4)
		sc ^b	0.562	0.550	0.510	0.516	0.641
	O.sp ³	Both	0.659 (5)	0.674 (5)	0.623 (5)	0.589 (5)	0.710 (4)
		Steric	0.685 (4)	0.685 (4)	0.685 (4)	0.685 (4)	0.685 (4)
		Electrostatic	0.339 (5)	0.345 (5)	0.493 (3)	0.557 (5)	0.644 (4)
		sc ^b	0.545	0.553	0.516	0.528	0.612
	C.sp ³	Both	0.663 (5)	0.670 (5)	0.609 (5)	0.531 (5)	0.669 (4)
		Steric	0.616 (3)	0.616 (3)	0.616 (3)	0.616 (3)	0.616 (3)
		Electrostatic	0.339 (5)	0.345 (5)	0.493 (3)	0.557 (5)	0.644 (4)
		sc ^b	0.545	0.573	0.503	0.516	0.618
W-2	Auto	Both	0.514 (5)	0.550 (5)	0.428 (6)	0.477 (6)	0.631 (5)
		Steric	0.408 (4)	0.408 (4)	0.408 (4)	0.408 (4)	0.408 (4)
		Electrostatic	0.544 (5)	0.560 (5)	0.400 (5)	0.254 (2)	0.436 (4)
		sc ^b	0.533	0.527	0.510	0.515	0.600
	O.sp ³	Both	0.546 (5)	0.559 (5)	0.524 (5)	0.521 (5)	0.629 (4)
		Steric	0.465 (4)	0.465 (4)	0.465 (4)	0.465 (4)	0.465 (4)
		Electrostatic	0.320 (3)	0.313 (3)	0.397 (3)	0.374 (3)	0.557 (4)
		sc ^b	0.514	0.513	0.481	0.501	0.568
	C.sp ³	Both	0.530 (5)	0.563 (5)	0.493 (5)	0.498 (5)	0.646 (5)
		Steric	0.456 (4)	0.456 (4)	0.456 (4)	0.456 (4)	0.456 (4)
		Electrostatic	0.320 (3)	0.313 (3)	0.397 (3)	0.374 (3)	0.557 (4)
		sc ^b	0.509	0.536	0.482	0.500	0.588

^{a,b}As in Table 13.

with the artemisinin analogues. The results are shown in Table 13. We found no significant difference between electrostatic fields built with different charges, as long as the HF/3-21G optimized structures were maintained. We also compared the AM1 charge model with HF/3-21G and AM1 optimized structures and found that almost all the models from HF/3-21G optimized structures give better results than the AM1 models.

In a previous study [39], a strong influence of different electrostatic field descriptors on CoMFA results was found. We, therefore, additionally used NPA charges at the HF/3-21G level. CoMFA was performed using alignment 4, and the results are shown in Table 13. For the HF/3-21G optimized structures, for the D-6 clone there is no significant difference among the electrostatic fields, whereas for the W-2 clone the Gasteiger and Marsili charge model is substantially worse than the others. Using O.sp³ as the probe atom with both fields, the HF/3-21G/MPA charge model

gives $q^2 = 0.581$ (noc = 5) compared to 0.487 (noc = 5) from the Gasteiger and Marsili charge model. Using the AM1 optimized structures, the Gasteiger and Marsili charges give much worse results than AM1/MPA. The AM1/MPA charge models with HF/3-21G optimized structures using both fields give better results than the use of only one field, while with AM1 optimized structures the use of only the electrostatic field gives better results. Among all methodical variations, HF/3-21G optimized structures give the best results in almost every model. We also calculated the ESPFIT charges at the AM1 level of the HF/3-21G optimized structures without compounds 7–10 in order to avoid problems arising from anionic compounds and then performed the CoMFA calculations using alignment 4. The results are shown in Table 14. Interestingly, different electrostatic fields significantly affect the CoMFA results. The Gasteiger and Marsili charge model yields substantially better results than the others, and the AM1/ESPFIT charge model usually

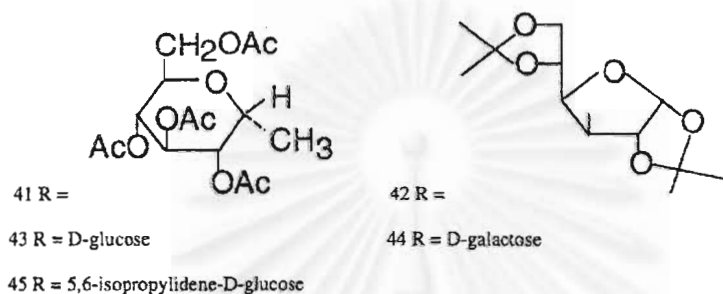
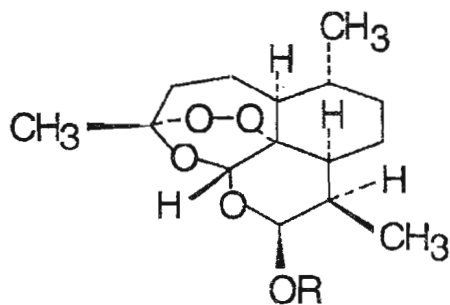


Figure 3. Structures of the five compounds in the test set.

yields worse results than the others: e.g. for D-6 activity with C.sp³ as the probe atom and using both fields, the AM1/ESPFIT charge model gives $q^2 = 0.531$ (noc = 5) compared to 0.663 (noc = 5) resulting from HF/3-21G/MPA.

Conclusions

In CoMFA studies, there are many adjustable parameters which sometimes significantly affect the results [40]. Therefore, a systematic variation of some parameters can significantly increase their quality. In this study, we have examined the influence of the steric field, the electrostatic field, and the orientation of the molecules. For the steric field, HF/3-21G geometry optimization usually gives better results than AM1, albeit the computational demand is much higher. For the electrostatic field, different charge calculation methods sometimes significantly affect the CoMFA results, depending both on the level of theory used for optimization and on the data set.

Acknowledgements

The authors would like to thank the Computer Center of the University of Vienna, Austria for providing computing resources at their workstation clusters. V.P. would like to thank Prof. C.J. Cramer, Department of Chemistry, University of Minnesota, for allowing some calculations to be carried out at his facilities. S.T. gratefully acknowledges a scholarship for his visit to the University of Innsbruck from the Austrian Federal Ministry for Foreign Affairs.

References

1. World Health Organization, Bull. WHO, 71 (1993) 281.
2. Butter, D., Maurice, J. and O'Brien, C., Nature, 386 (1997) 535.
3. Olliaro, P.L. and Trigg, P.I., Bull. WHO, 73 (1995) 565.
4. World Health Organization, WHO Drug Inf., 7 (1993) 7.
5. Mockenhaupt, F.P., Parasitol. Today, 11 (1995) 248.
6. Qinghaosu Antimalaria Coordinating Research Group, Chin. Med. J., 92 (1979) 811.
7. China Cooperative Research Group on Qinghaosu and its Derivatives as Antimalarials, J. Tradit. Chin. Med., 2 (1982) 9.
8. Klayman, D.L., Science, 228 (1985) 1049.
9. Luo, X.D. and Shen, C.-C., Med. Res. Rev., 7 (1987) 29.
10. Meshnick, S.R., Taylor, T.E. and Kamchonwongpaisan, S., Microbiol. Rev., 60 (1996) 301.

11. Hansch, C. and Fujita, T., *J. Am. Chem. Soc.*, 86 (1964) 1616.
12. Cramer, R.D., Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
13. Leban, I. and Golic, L., *Acta Pharm. Jugosl.*, 38 (1988) 71.
14. Posner, G.H. and Oh, C.H., *J. Am. Chem. Soc.*, 114 (1992) 8328.
15. Posner, G.H., Wang, D., Cumming, J.N., Oh, C.H., French, A.N., Bodley, A.L. and Shapiro, T.A., *J. Med. Chem.*, 38 (1995) 2273.
16. Jiang, H.-L., Chen, K.-X., Wang, H.-W., Tang, Y., Chen, J.-Z. and Ji, R.-Y., *Acta Pharmacol. Sin.*, 15 (1994) 481.
17. Avery, M.A., Gao, F., Chong, W.K.M., Mehrotra, S. and Milhous, W.K., *J. Med. Chem.*, 36 (1993) 4264.
18. Kim, K.H., In Dean, P.M. (Ed.) *Molecular Similarity in Drug Design*, London, 1995, pp. 291-331.
19. Kroemer, R.T. and Hecht, P., *J. Comput.-Aided Mol. Design*, 9 (1995) 396.
20. Horwitz, J.P., Massova, I., Wiesc, T.E., Besler, B.H. and Corbett, T.H., *J. Med. Chem.*, 37 (1994) 781.
21. Hannongbua, S., Lawtrakul, L., Sottriffer, C.A. and Rode, B.M., *Quant. Struct.-Act. Relatsh.*, 15 (1996) 1.
22. Lin, A.J., Klayman, D.L. and Milhous, W.K., *J. Med. Chem.*, 30 (1987) 2147.
23. Lin, A.J., Li, L.Q., Klayman, D.L., George, C.F. and Anderson, J.L.F., *J. Med. Chem.*, 33 (1990) 2610.
24. Lin, A.J. and Miller, R.E., *J. Med. Chem.*, 38 (1995) 764.
25. Pu, Y.M., Torok, D.S., Ziffer, H., Pan, X.-Q. and Meshnick, S.R., *J. Med. Chem.*, 38 (1995) 4120.
26. Avery, M.A., Mehrotra, S., Johnson, T.L., Bonk, J.D., Vromn, J.A. and Miller, R., *J. Med. Chem.*, 39 (1996) 4149.
27. SYBYL Molecular Modeling Software, v. 6.3, Tripos Associates Inc., St. Louis, MO, U.S.A., 1996.
28. GAUSSIAN 94, Gaussian Inc., Pittsburgh, PA, U.S.A., 1995.
29. Gasteiger, J. and Marsili, M., *Tetrahedron*, 36 (1980) 3219.
30. Singh, U.C. and Kollman, P.A., *J. Comput. Chem.*, 5 (1984) 129.
31. Reed, A.E. and Weinhold, F.J., *J. Am. Chem. Soc.*, 108 (1986) 3586.
32. Stewart, J.J.P., MOPAC 6.0, Quantum Chemical Program Exchange 455, 1990.
33. Clark, M., Cramer, R.D. and VanOpdenbosch, N., *J. Comput. Chem.*, 10 (1989) 982.
34. Dunn, W.J., Wold, S., Edlund, U., Hellberg, S. and Gasteiger, J., *Quant. Struct.-Act. Relatsh.*, 3 (1984) 131.
35. Geladi, P., *J. Chemometrics.*, 2 (1988) 231.
36. Kubinyi, H. and Abraham, U., In Kubinyi, H. (Ed.) *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993, pp. 712-728.
37. Cramer, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.) *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993, pp. 443-485.
38. Lin, A.J., Li, L.-Q., Andersen, S.L. and Klayman, D.L., *J. Med. Chem.*, 35 (1992) 1639.
39. Kroemer, R.T., Hecht, P. and Liedl, K.R., *J. Comput. Chem.*, 17 (1996) 1296.
40. Folkers, G., Merz, A. and Rognan, D., In Kubinyi, H. (Ed.) *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, 1993, pp. 583-618.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix B

Ab initio and density functional study on singlet
and triplet states of artemisinin

Somsak Tonmunphean, Stephan Irle, Sirirat Kokpol,
Vudhichai Parasuk, Peter Wolschann

J. Mol. Struct. (THEOCHEM) 454 (1998): 87-90.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ELSEVIER

Journal of Molecular Structure (Theochem) 454 (1998) 87–90

THEO
CHEM

Ab initio and density functional study on singlet and triplet states of artemisinin

Somsak Tonmunpuean^a, Stephan Irle^b, Sirirat Kokpol^a, Vudhichai Parasuk^a, P. Wolschann^{b,*}^aDepartment of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand^bInstitute of Theoretical Chemistry and Radiation Chemistry, University of Vienna, Währinger Strasse 17, A-1090 Vienna, Austria

Received 17 February 1998; received in revised form 24 April 1998; accepted 30 April 1998

Abstract

Calculations on the singlet (S_0) and triplet (T_1) states of artemisinin were carried out to study its singlet–triplet excitations using several levels of theory. Geometries of the singlet and the triplet states were at first optimized at Hartree–Fock (HF) level using 3-21g basis. An additional calculation was performed for the triplet state using the ground-state geometry. The adiabatic transition (ΔE_{ad}) of -4.62 kcal/mol was obtained from the energy difference of S_0 and T_1 at their optimized geometries. The vertical transition (ΔE_v) of 90.96 kcal/mol was computed from the energy difference of S_0 and T_1 at the singlet optimized geometry. The result suggests that artemisinin would readily become diradical under normal condition which is in contrast to the experiments. Thus, B3LYP density functional theory calculations using 6-31g* basis were performed to confirm the HF results. In contrast to the HF calculations, the ΔE_{ad} and ΔE_v of 37.42 and 96.53 kcal/mol were obtained, respectively. Geometry optimizations at B3LYP level using 6-31g* basis were also performed on both singlet and triplet. Using B3LYP's geometries, the Møller–Plesset second order perturbation theory (MP2) were applied to the two states and in respective order ΔE_{ad} and ΔE_v of 50.81 and 117.48 kcal/mol were reported. In summary, HF theory is not sufficient to estimate the adiabatic transition of this endoperoxide compound, and correlated calculations would be required. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Artemisinin; Ab initio calculation; Density functional calculation; Singlet–triplet splitting

Artemisinin, a compound isolated from chinese herb [1], has a unique structure with an endoperoxide linkage. This compound and its derivatives were used as antimalarial drugs. One of the proposed mechanism for its antimalarial activity [2–4] suggested that ferrous ions (Fe^{2+}) contained in the malaria parasites reduce the endoperoxide linkage of artemisinin to form an oxy-radical and in a further step to produce an intermediate radical center at C_4 (see Fig. 1).

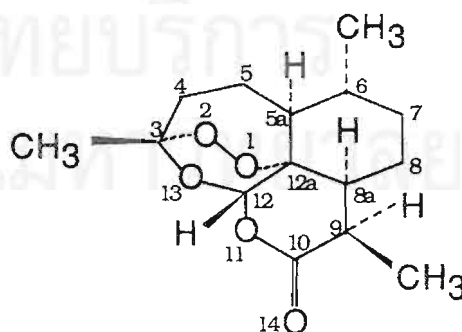


Fig. 1. Stereochemistry and atomic numbering scheme of artemisinin.

* Corresponding author. Tel: +43 1 40480603; fax: +43 1 40480647; e-mail: Karl.Peter.Wolschann@univie.ac.at

Evidence for the importance of the C₄ position can be seen from the fact that derivatives with β -substituents at C₄ are at least 12–200 times more active than the corresponding C₄ α -substituted derivatives [5].

Without the present of ferrous ions, the oxy-radical intermediate of artemisinin has radical centers at the position O₁ and O₂. The electronic state of this oxy-radical intermediate can either be described by an open-shell singlet or triplet wavefunction. The triplet state would normally be energetically favorable and, therefore, is a good choice for representing the electronic state of the oxy-radical intermediate. Thus, the energetic profile of the electronic transition between the inactive form, the singlet ground state (closed-shell singlet), and the oxy-radical intermediate of artemisinin can be understood from electronic structures of the corresponding singlet and triplet states of artemisinin. It is possible that this singlet–triplet energy gap can be used for predicting bioactivity of artemisinin derivatives. According to the size of the molecule, the Hartree–Fock (HF) method seems to be an appropriate choice for the study. However, singlet–triplet splittings from the HF theory are usually too small by about 20 kcal/mol, which corresponds to the electron correlation energy of the broken electron pair involved in the dioxo-bond [6,7], e.g. high-level correlated methods were necessary to reproduce accurately the energy difference between ³ Σ_g^- and ¹ Δ_g electronic states of O₂ [8]. These methods are computationally very demanding and their use is, therefore, prohibited for investigations on molecules of the size of artemisinin. An interesting alternative offers density functional theory (DFT) which has been shown to reproduce energy differences between singlet and triplet states of systems involving first transition row elements within an accuracy comparable to Møller–Plesset second order perturbation theory (MP2) [9]. It is also the aim of this work to investigate the reliability of DFT in the context of singlet–triplet splittings of artemisinin.

All quantum chemical calculations were carried out using the Gaussian 92/DFT suite of programs [10]. For the density functional we selected the B3LYP hybrid parameterization [11]. Open-shell MP2 calculations were based on both the spin-restricted HF (ROHF) and the UHF method. Throughout all unrestricted calculations on the triplet state, spin

Table 1

Vertical and adiabatic singlet–triplet splittings of artemisinin using the HF, B3LYP and MP2 methods

Method	Basis set	Vertical transition (kcal/mol)	Adiabatic transition (kcal/mol)
HF/3-21G optimized geometries			
RHF/UHF	3-21G	90.96	– 4.62
RHF/ROHF	3-21G	—	– 0.76
RHF/UHF	6-31G*	91.53	– 5.88
B3LYP	6-31G*	96.53	37.42
B3LYP/6-31G* optimized geometries			
B3LYP	6-31G*	96.31	31.24
RMP2/UMP2	6-31G*	117.63	50.73
RMP2/ROMP2	6-31G*	117.48	50.81

contamination was found to be very low with a deviation of $\langle S^2 \rangle$ of at most 0.03 from the theoretical value of 2.

In a first step, we performed geometry optimization at the HF level on the lowest singlet and triplet electronic states of artemisinin, employing the 3-21G basis set. Surprisingly, it was found that at this level of theory the triplet state is lower in energy by 4.6 kcal/mol as compared with the singlet state (see Table 1).

The optimized geometries (compare with Fig. 2) show that the O₁–O₂ bond of the endoperoxide bridge is indeed opened in going from the singlet to the triplet state showing a large separation of both oxygen atoms of about 2.62 Å. Spin densities of the triplet state confirm the existence of two radical electrons at these oxygen atoms. Whereas the C–O bonds of the endoperoxide bridge are shortened considerably by about 0.1 Å, especially the C₃–C₄ bond becomes stretched by more than 0.05 Å with a bond length of 1.60 Å upon excitation into the triplet state. Bond length elongations are visible for the C₁₂–C_{12a} bond as well as for the C₃–C₄. Bond length shortenings of about 0.01 Å can be observed for C₁₂–O₁₃ and C₁₂–O₁₁ distances. The distances in the cyclohexane ring remain unchanged upon excitation. Changes of the bond angles in triplet state can be observed around O₂ of the dioxo bond. Interestingly, the position of O₁ is kept constant (O₁–C_{12a}–C₁₂). Broad changes of the torsional angles show the rearrangement of the whole molecular skeleton owing to the opening of the strained O–O bond.

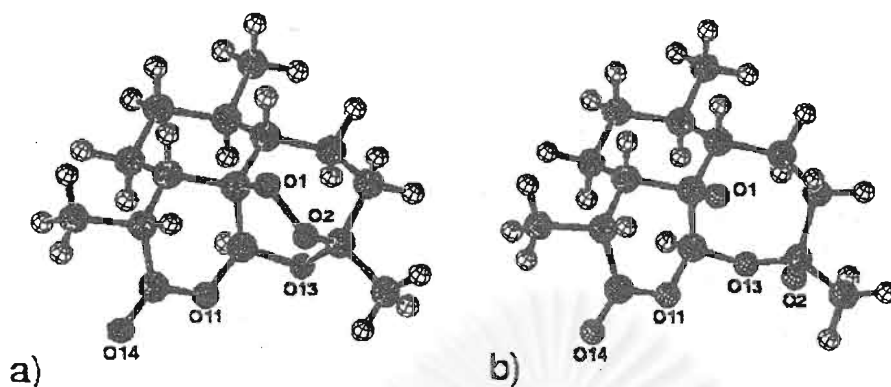


Fig. 2. B3LYP/6-31G* optimized geometries of artemisinin in its: (a) singlet; and (b) triplet state.

In a further step we included electron correlation effects using B3LYP/6-31G* single point calculations at the HF/3-21G optimized singlet and triplet geometries. In these calculations, the triplet state was found to be 37.4 kcal/mol higher in energy than the singlet state (Table 1). The unusually large difference of 42.0 kcal/mol in the singlet–triplet splittings obtained at the HF/3-21G and B3LYP/6-31G* levels cannot be attributed to the use of different basis sets as demonstrated by HF/6-31G* single point calculations which confirm the HF/3-21G results (see Table 1). Obviously, DFT lowers the energy of the singlet ground state in comparison to HF by roughly twice the correlation energy of a single electron pair. In contrast, HF and B3LYP vertical singlet–triplet excitation energies at the singlet geometry are in remarkable agreement with values of 91.5 and 96.5 kcal/mol,

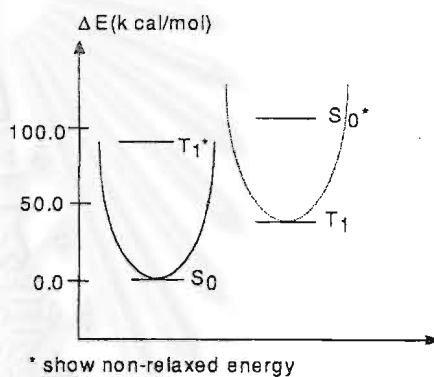


Fig. 3. Energy diagram of singlet and triplet state excitation.

respectively (see Table 1 and Fig. 3). Obviously, relaxation of ring strain plays an important role in the energetics of the opening of the endoperoxide bridge which influences significantly the singlet–triplet

Table 2

Most important structural parameters for the singlet and triplet states of artemisinin at the B3LYP/6-31G* optimized geometries (distances in Å and angles in degrees)

Bond length	Bond angle		Torsional angle					
	Singlet	Triplet	Singlet	Triplet				
C ₃ –C ₄	1.547	1.605	C ₄ –C ₃ –O ₁₃	109.4	110.4	C ₄ –C ₃ –O ₁₃ –C ₁₂	– 89.5	– 73.4
C ₃ –C ₁₃	1.436	1.452	C ₃ –O ₁₃ –C ₁₂	114.1	121.7	C ₃ –O ₁₃ –C ₁₂ –C _{12a}	27.3	41.8
C ₁₂ –C _{12a}	1.529	1.562	O ₁ –C _{12a} –C ₁₂	111.3	111.6	O ₂ –O ₁ –C ₁₂ –C _{12a}	11.7	29.3
O ₁ –C _{12a}	1.477	1.360	O ₂ –O ₁ –C _{12a}	111.6	102.7	C ₃ –O ₂ –O ₁ –C _{12a}	47.9	31.1
O ₁ –O ₂	1.462	2.625	O ₁ –O ₂ –C ₃	108.3	88.1	O ₁ –O ₂ –C ₃ –C ₄	46.9	57.8
O ₂ –C ₃	1.414	1.330	O ₂ –C ₃ –C ₄	111.9	105.7	O ₁ –C _{12a} –C ₁₂ –O ₁₁	73.1	54.4
C ₁₂ –O ₁₃	1.396	1.386	O ₁₁ –C ₁₂ –O ₁₃	107.5	108.8	O ₁ –C _{12a} –C _{8a} –C ₈	167.9	171.5
O ₁₁ –C ₁₂	1.439	1.430	C ₁₀ –O ₁₁ –C ₁₂	124.7	122.4	C ₃ –O ₁₃ –C ₁₂ –O ₁₁	– 100.9	– 86.3
C ₁₀ –O ₁₁	1.365	1.362	O ₁₁ –C ₁₀ –O ₁₄	118.3	118.6	C ₁₂ –O ₁₁ –C ₁₀ –O ₁₄	165.5	161.6
C ₁₀ –O ₁₄	1.207	1.207	C _{5a} –C _{12a} –C _{8a}	112.9	112.1	C ₁₀ –C ₉ –C _{8a} –C _{12a}	– 51.7	– 44.6

splittings in artemisinin and makes them difficult to predict from estimates based on HF calculations.

To check the reliability of the HF/3-21G geometries, the geometries of both singlet and triplet state were reoptimized at the B3LYP/6-31G* level. The most important structure parameters obtained from these calculations are collected in Table 2.

These geometries were found to be very similar to the HF/3-21G geometries, and owing to the similarity of the HF/3-21G and B3LYP/6-31G* geometries the singlet–triplet splittings obtained at these levels should not differ much. At this stage we performed MP2/6-31G* single point calculations as a check of the accuracy of B3LYP. The singlet ground state is confirmed by the MP2 with the stabilization energy of around 50 kcal/mol, a number which is even larger in comparison to that of B3LYP. It is known that certain functionals could not describe lone pair interactions very well and some functionals require basis set of sufficient flexibility to represent interactions properly. It is probably that B3LYP quantitatively underestimates the singlet–triplet splitting. Unlike HF, B3LYP still predicts correct ground state. It is concluded that the B3LYP calculations should give the more reliable results and is in preference over the qualitatively incorrect HF method.

Acknowledgements

The authors thank the Austrian–Thai Center for Computer Assisted Chemical Education and Research (ATC) and Computational Chemistry Unit Cell,

Department of Chemistry, Faculty of Science, Chulalongkorn University, Thailand, and the Computer Center of the University of Vienna, Austria for providing ample computing resources at their DEC Alpha 2100 4/275 workstation cluster. S.I. gratefully acknowledges support during his visit to Chulalongkorn University from the program ASEA Uninet which is sponsored by the Austrian Government and Chulalongkorn University.

References

- [1] Qinghaosu Antimalaria Coordination Research Group, *Chin. Med. J.* 92 (1979) 811.
- [2] G.H. Posner, J.N. Cumming, P. Ploypradith, C.H. Oh, *J. Am. Chem. Soc.* 117 (1995) 5885.
- [3] G.H. Posner, D.J. McGarvey, C.H. Oh, N. Kumar, S.R. Meshnick, W. Asawamahsakda, *J. Med. Chem.* 38 (1995) 607.
- [4] G.H. Posner, S.B. Park, M.D. Bachi, *J. Am. Chem. Soc.* 118 (1996) 3537.
- [5] G.H. Posner, C.H. Oh, L. Gerena, W.K. Milhous, *Heteroatom Chem.* 6 (1995) 105.
- [6] E.R. Davidson, *J. Chem. Phys.* 41 (1964) 656.
- [7] E.R. Davidson, *J. Chem. Phys.* 42 (1964) 4199.
- [8] V. Staemmler, R. Jaquet, *Theor. Chim. Acta* 59 (1981) 501.
- [9] A. Ricca, C.W. Bauschlicher Jr, *Theor. Chim. Acta* 92 (1996) 123.
- [10] M.J. Frisch, G.W. Trucks, M. Head-Gordon, P.M.W. Gill, M.W. Wong, J.B. Foresman, B.G. Johnson, H.B. Schlegel, M.A. Robb, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, J.S. Binkley, C. Gonzales, R.L. Martin, D.J. Fox, D.J. Defrees, J. Baker, J.J.P. Stewart, J.A. Pople, *Gaussian 92 Rev. B*, Gaussian Inc., Pittsburgh, PA, 1992.
- [11] A.D. Becke, *J. Chem. Phys.* 98 (1993) 5648.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix C

QSAR Study of Antimalarial Activities and
Artemisinin-Heme Binding Properties Obtained
from Docking Calculations

Somsak Tonmunphean,
Vudhichai Parasuk, Sirirat Kokpol

Quant. Struct.-Act. Rel. 19 (2000): 475-483.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

QSAR Study of Antimalarial Activities and Artemisinin-Heme Binding Properties Obtained from Docking Calculations

Somsak Tonmunpheap, Vudhichai Parasuk* and Sirirat Kokpol

Department of Chemistry, Faculty of Science, Chulalongkorn University, Patumwan, Bangkok, 10330 Thailand

Abstract

The quantitative structure-activity relationships (QSAR) between antimalarial activities and artemisinin-heme binding properties were studied by means of docking calculations. Automated molecular dockings of 30 artemisinin derivatives to heme revealed a significant relationship between biological activity and binding energy ($r = -0.93$) and less significantly with the O1-Fe distance

($r = -0.55$). The QSAR models were constructed and the predicted biological activities were in good agreement with the corresponding experimental values. The docking also showed that artemisinin compounds approach heme by pointing O1 at the endoperoxide linkage toward the iron center, a mechanism controlled by the steric hindrance.

1 Introduction

Currently, the worldwide malaria situation has become serious once again since malarial parasites have developed resistance to the most common drugs [1–3]. Artemisinin (Figure 1), a sesquiterpene endoperoxide isolated from a Chinese medicinal herb [4], was found to be a potent antimalarial drug against the resistant strains of *Plasmodium falciparum* [5–6]. Since then numerous artemisinin derivatives have been synthesized and their biological activities have been tested. Some of them showed higher activity than the original compound. This high potency against the resistant strains is possibly due to their unusual structures, with a mode of action different from those of other antimalarial drugs. However, the mechanism for antimalarial activity of these compounds is still inconclusive, although researchers agree on the significance of the endoperoxide group in artemisinin compounds to the biological activity. This is evidence by the impotency of deoxyartemisinin analogs, which lack the endoperoxide moiety [7]. *In vitro* experiments suggested that some amount of iron must be added for artemisinin to show antimalarial activity [8–10]. In the human body, the heme compound (Figure 2), which is a product from the digestive process, is believed to be the source of iron for artemisinin [11].

Malarial parasites in humans degrade hemoglobin from the red blood cells into heme and globin. Subsequently, globin is hydrolyzed to provide amino acids as the sources for protein synthesis. The toxic heme portion will be mostly detoxified by the polymerization process to hemozoin.

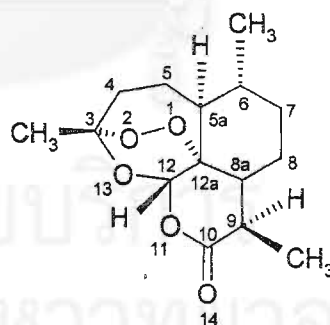


Figure 1. Structure of artemisinin compound with atom numbering.

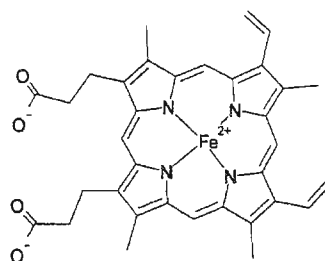


Figure 2. Structure of the heme molecule.

* To receive all correspondence. e-mail: parasuk@atc.atccu.chula.ac.th

Key words: QSAR, docking, antimalarial drug, endoperoxide, heme

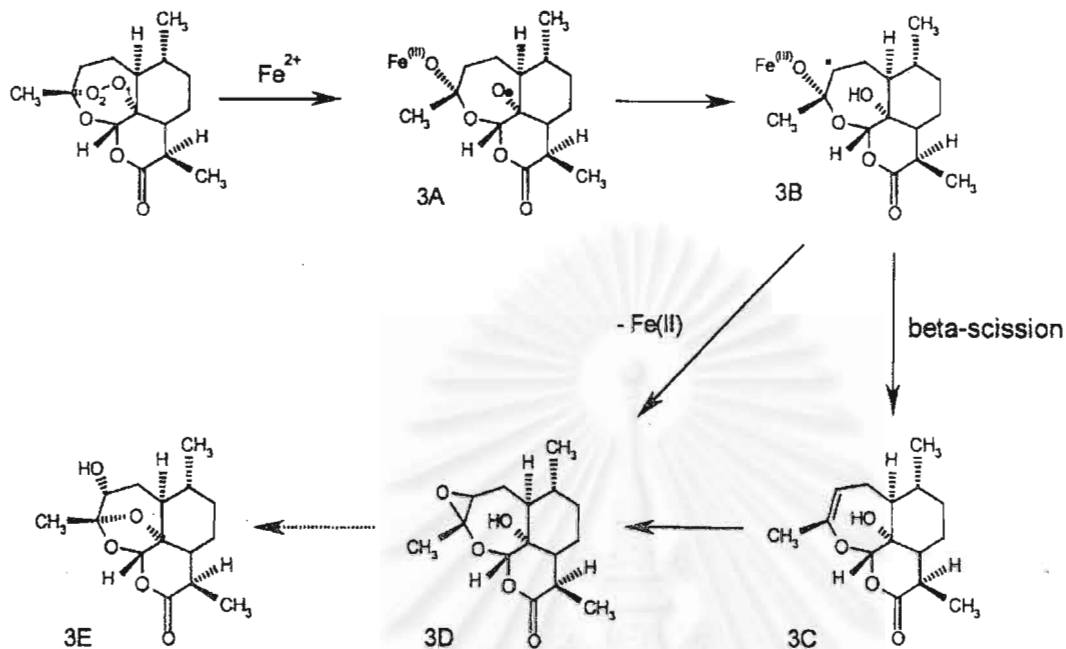


Figure 3. Mechanism of action of artemisinin proposed by Posner [12].

There are two possible ways for heme iron to approach the endoperoxide linkage of artemisinin compounds. Posner and co-workers [12] proposed that the iron attacks the compound at the O2 position and produces a free radical at the O1 position which later is rearranged to form a C4 free radical (Figure 3). This radical (3B) was suggested to be an important substance for antimalarial activity [10]. The 3B radical is changed to the 3C compound by a beta-scission reaction. Subsequently, the 3C compound is rearranged to form an epoxide compound (3D). Alternatively, a direct intramolecular formation of the 3B radical gives the 3D

compound. This compound is able to alkylate the specific proteins of the malarial parasites and possibly causes damage to the parasites [13]. On the other hand, Jefford and co-workers [14] believed that the iron attacks the compound at the O1 position and produces a free radical at the O2 position (Figure 4). After that the C3-C4 bond is cleaved to give a carbon radical at C4 (4B). This radical could also be very harmful to the parasites.

In drug discovery and development, knowledge of the mechanism of action could assist the development of new

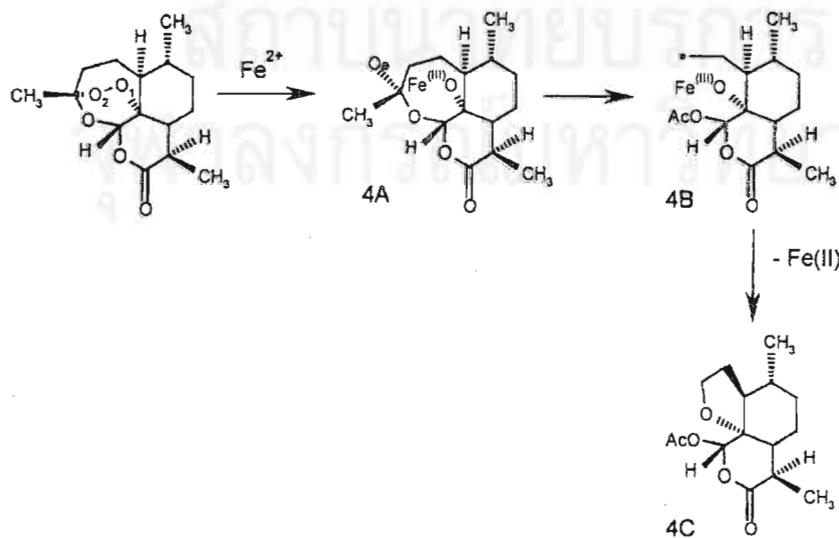


Figure 4. Mechanism of action of artemisinin proposed by Jefford [14].

and more effective drugs. Generally, the drug mediates its effects at a specific target, a receptor. In order to show high activity, the drug has to bind firmly with the receptor, like a "hand and glove" model. If the drug-receptor complex structure is known, the drug-receptor interactions could be investigated in more detail. Hence, some more effective drugs could be designed by structural modifications to achieve the suitable interactions. Molecular docking is one method for investigating the drug-receptor interactions.

Shukla and co-workers [15] studied the docking of artemisinin and deoxyartemisinin (an inactive analog) with hemin using the Dock module in SYBYL software, a direct docking algorithm. In the most stable dock configurations, the closest distance between artemisinin and hemin was found to be between O2 and Fe (ranging from 2.63 to 2.81 Å). The situation was different for deoxyartemisinin where hemin iron approached very close to the O13 atom. Therefore, one might propose that the difference in docking configurations should be responsible for the difference in antimalarial activity of these 2 compounds. However, there is no clear evidence to support this proposal. Therefore, it is our goal to investigate the relation between the docking configuration and antimalarial activity. In this study, docking calculations were carried out for 30 artemisinin derivatives. The docking results were then used to perform a QSAR study on antimalarial activities and binding energies of heme-artemisinin derivatives complexes.

2 Compounds and Methods

2.1 Docking Calculations

The docking calculations were carried out using the automated docking program, the AutoDock 2.4 software [16], on the personal computer (Pentium III 550 MHz) running on the Linux platform. The program uses a simulated annealing Monte Carlo simulation in combination with a rapid grid-based energy evaluation method where a grid map of a dimension of $25 \times 25 \times 25$ Å with a 0.5 Å spacing was used. Since the AMBER force fields have no parameters for iron, the combined AMBER/MMFF [17–18] parameters for the Lennard-Jones 12, 6 potentials and Coulomb potentials were obtained from the authors of the program [19].

For one docking calculation, the simulations were performed for 100 annealing cycles of decreasing temperature with a reduction fraction of 0.90. The initial annealing temperature, RT, was set to 100 kcal/mol. During each cycle, the ligand was gradually moved by a random displacement and the energy of the new configuration was then calculated. The selection of the new configuration was based on the Metropolis algorithm [20]. The cycle terminated if the ligand made 30 000 accepted or 30 000 rejected moves. Then the simulation moved to the next cycle.

As stated above, the Monte Carlo simulation is based on random movement, hence, the final docked configuration may differ if we start from a different configuration. In order to account for this, 100 docking calculations with different starting configurations were performed for each compound. Then, a cluster analysis was used to categorize all 100 docked configurations into groups. The configurations which had root-mean-squared deviation (rmsd) values less than 1 Å were grouped together. In each group, the lowest energy configuration was selected as the representative of that group. The "% Occurrence" was used to represent the number of members (configurations) in each group. Our attention was focused to the group with the highest % Occurrence or "the most occurring configuration". This configuration represents the highest probability to occur in the real system.

2.2 Structures of Artemisinin Derivatives and Heme

Quantum mechanical calculations were used to determine the geometry of all artemisinin compounds. To justify this, the suitable level of accuracy for the geometry optimization must be established. Optimized geometries of artemisinin obtained from various levels of accuracy, i.e., CNDO, AM1, HF/STO-3G, HF/3-21G, HF/6-31G**, were compared with its X-ray structure [21]. HF/3-21G seemed to be the lowest level of theory that can reproduce most of the geometrical parameters as compared with the X-ray structure. Therefore, HF/3-21G was selected for the geometry optimization of all artemisinin derivatives in this work [22]. The quantum chemical calculations were carried out using the Gaussian 94 program [23].

For the receptor molecule, the heme structure was taken from the modification of the X-ray structure of chlorohemin of the Cambridge Crystallography Data Bank [24]. Chlorohemin has a pyramidal structure with Fe on the top as shown in Figure 5. Our preliminary calculations suggested that this heme structure gave the highest drug-receptor binding energy, compared with other heme structures investigated [25]. A high level of accuracy, HF/6-311G**, was used for the calculation of atomic charges of heme. These charges were used in further docking calculations.

2.3 Biological Data

With the aim of exploring all possible effects of structural differences of artemisinin compounds on the biological

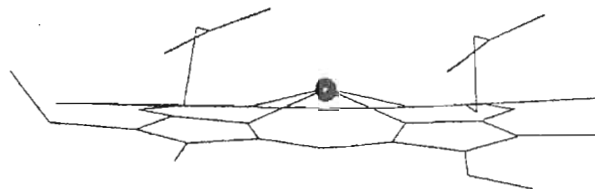


Figure 5. Three-dimensional structure of the heme molecule, showing its pyramid-like shape.

activities, a total of 30 compounds with significantly different structures and biological activities were selected [26–34]. To reduce inconsistency from individual experimental environments, the relative activity, or the ratio of the activity of a compound to that of artemisinin, was used. The structures of 30 compounds and their corresponding biological data are listed in Figure 6 and Table 1. These activities were measured *in vitro* against the D-6 and the W-2 clone of *Plasmodium falciparum* using the Milhous-Desjardins assay method. The D-6 clone is mefloquine-resistant but chloroquine-sensitive while the W-2 clone is chloroquine-resistant but mefloquine-sensitive. Since the biological activities of all compounds

were reported as IC_{50} in the ng/mL unit, it is necessary to convert the values to nM/mL unit so that these activities can be compared directly.

3 Results and Discussion

3.1 Docking Calculations

Results of docking between artemisinin derivatives and heme are demonstrated in Table 2 and Figure 7. The averaged CPU time required for docking a molecule is approximately 7 h.

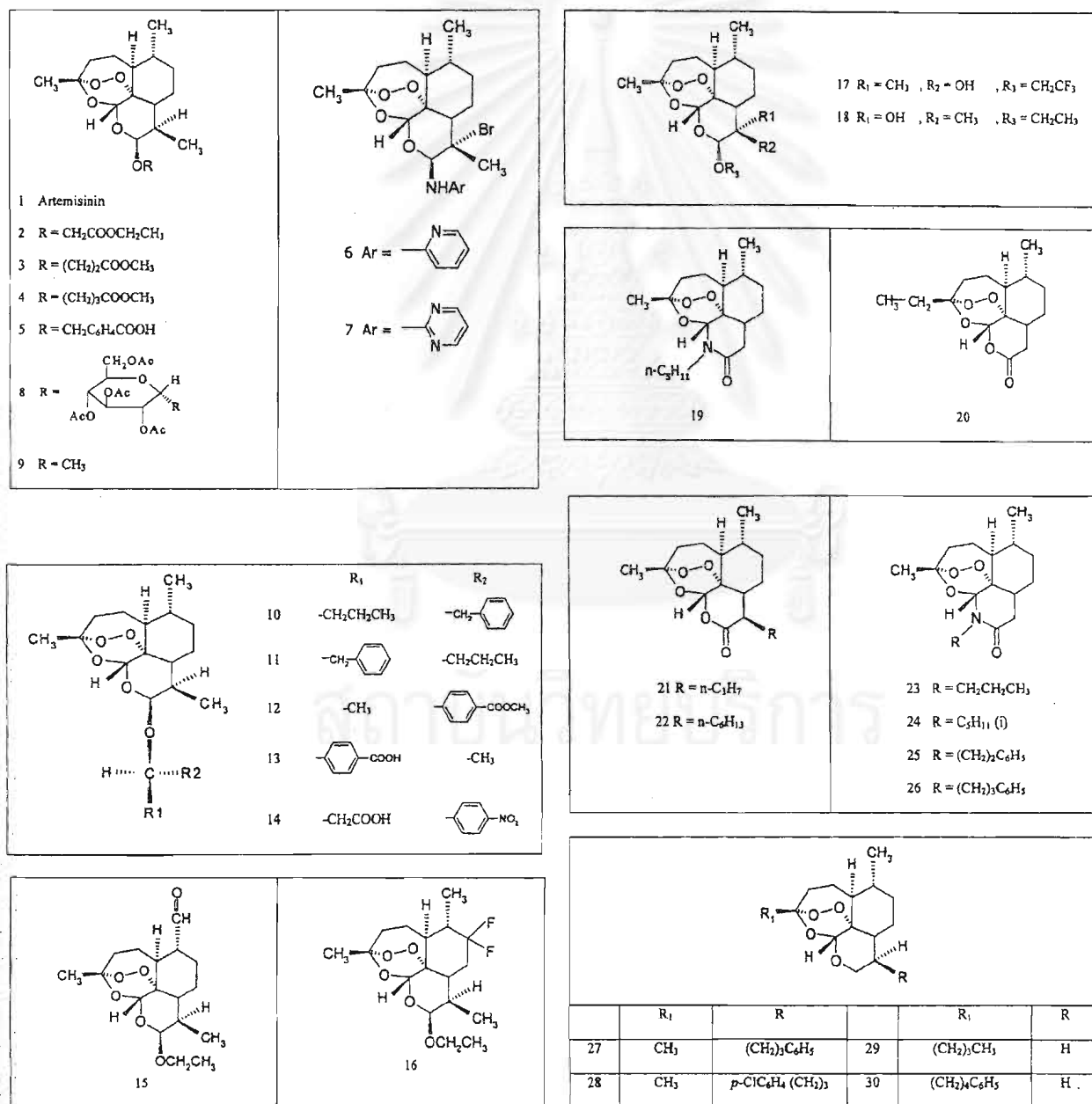


Figure 6. Structures of artemisinin and its derivatives.

Table 1. Biological activity of artemisinin and its derivatives

No.	D-6 activity ^(a)	W-2 activity ^(a)	No.	D-6 activity	W-2 activity
1	1.00	1.00	16	1.73	1.48
2	6.41	3.33	17	0.20	0.19
3	2.09	1.35	18	0.52	0.35
4	1.30	0.95	19	1.10	0.63
5	1.07	0.71	20	0.88	1.12
6	0.11	0.80	21	6.05	13.47
7	0.03	0.17	22	5.83	6.87
8	2.58	7.79	23	0.75	1.11
9	2.80	3.57	24	1.49	0.92
10	1.15	0.81	25	1.97	1.43
11	3.61	2.22	26	2.05	1.05
12	3.95	4.77	27	50.73	25.06
13	1.20	1.26	28	69.91	33.17
14	1.32	1.24	29	6.53	5.56
15	1.39	1.62	30	3.36	3.80

^(a)Activity = (IC₅₀ of artemisinin/IC₅₀ of analog) × (M.W. of analog/M.W. of artemisinin).

As mentioned earlier, the endoperoxide moiety is essential for antimalarial activity, therefore, our report emphasized particularly on binding energies and distances between the endoperoxide oxygens, O1 and O2, and the heme iron.

Almost all docking configurations between heme and artemisinin derivatives showed that the Fe atom of heme positions itself at the proximity of the endoperoxide bond, with O1-Fe and O2-Fe distances between 1.93 to 2.73 Å and 2.41 to 3.76 Å, respectively. This emphasizes the importance of the interaction between heme iron and the endoperoxide group. The docked configurations between heme and artemisinin derivatives are illustrated in Figure 7.

The obtained O1-Fe distances are comparable to the experimental bond length between heme iron and an oxygen atom in oxy hemoglobin A (1.86 Å), taken from the Protein Data Bank (id 1HHO). Inconsistent with other compounds, the most frequently occurring configurations of compounds 6 and 7 have an O13 atom pointing toward the heme iron instead of the endoperoxide oxygens. However, the second most frequently occurring configurations have an O1 atom pointing toward the heme iron, though with a slightly longer O1-Fe distance (2.73 and 2.62 Å). This is possibly because of the existence of the Br group on the same side of the endoperoxide moiety which prevents the heme iron from binding with the endoperoxide oxygens. To prove this steric hypothesis, a docking calculation for the analog of compound

Table 2. Docking results for heme and artemisinin derivatives

Compound no.	Energy (kcal/mol)	% Occurrence	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
1	-33.13	24	2.00	2.65	4.67	4.90
2	-34.90	38	2.01	2.65	4.64	4.80
3	-33.15	41	1.98	2.69	4.67	4.73
4	-33.26	30	1.98	2.83	4.68	4.83
5	-33.22	31	1.97	2.66	4.66	4.71
6	-31.64	12	5.21	4.64	2.40	3.67
	-28.45	9	2.73	3.76	5.47	5.55
7	-30.22	12	6.53	6.51	4.41	6.13
	-28.09	8	2.62	3.47	5.42	5.22
8	-33.27	18	2.00	2.69	4.67	4.79
9	-33.18	27	1.95	2.54	4.59	4.66
10	-33.13	28	2.01	2.70	4.71	4.72
11	-33.27	30	1.99	2.72	4.67	4.79
12	-33.76	24	2.12	2.69	4.70	4.90
13	-33.76	21	1.98	2.56	4.62	4.68
14	-34.07	29	2.07	2.75	4.68	4.91
15	-34.09	23	2.08	2.77	4.73	4.87
16	-34.01	26	2.06	2.64	4.66	4.83
17	-30.90	37	2.20	3.05	4.87	5.10
18	-32.49	20	2.34	3.08	4.78	5.20
19	-32.75	42	1.98	2.52	4.59	5.01
20	-33.66	44	2.17	2.99	4.84	5.18
21	-35.21	31	2.06	2.97	4.73	5.10
22	-36.49	24	2.20	3.21	4.99	5.13
23	-32.91	40	2.00	2.51	4.61	4.97
24	-32.92	35	2.01	2.49	4.57	5.12
25	-32.96	40	1.95	2.47	4.55	4.96
26	-33.36	22	1.94	2.53	4.56	5.01
27	-37.57	29	2.10	2.68	4.76	4.77
28	-37.85	32	2.11	2.76	4.82	4.78
29	-34.90	45	1.93	2.41	4.51	4.60
30	-35.84	16	1.95	2.52	4.60	4.61
31	-33.38	35	2.04	2.74	4.71	4.80

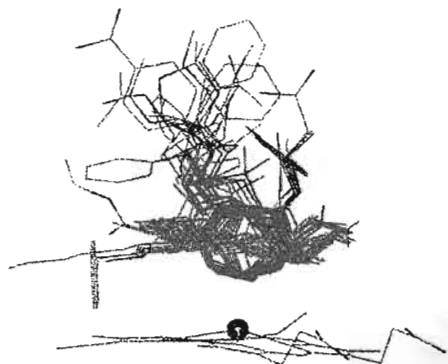


Figure 7. Docking configuration between heme and artemisinin derivatives (hydrogen atoms not shown).

6 without the Br group (compound 31) was performed. The shorter O1-Fe distance of 2.04 Å, which is comparable to those in other compounds, was obtained.

Considering the O1-Fe and O2-Fe distances in docked configurations of all 30 artemisinin compounds, it is obvious that the heme iron approaches the endoperoxide moiety at the O1 position more preferably than the O2 position. This result differs somewhat from that of previous work [15] which reported the preference of configurations with O2-Fe as the shortest heme-artemisinin distance and the important role of O14 in the binding. Interestingly, our calculations showed that O14 is further away from the iron (more than 6 Å) and should not play any role for the binding and hence activity. The evidence of compounds 27 to 30, which have no O14 in the structure but still have biological activity higher than artemisinin, supports our predictions. Moreover, in previous calculations [15] the binding energy of the configurations where O2 attaches to iron is larger than where O1 attaches to iron. This result was contradicted in our study. The contradiction may be ascribed to the following reasons. In their study, the artemisinin structure was modified from the X-ray structure of artemether, not from the artemisinin itself as in our study. They started with only 3 orientations of artemisinin around the heme molecule in contrast to our calculations where all possible configurations are considered. In addition, previously an empirical method was used to calculate the atomic charges and the charge of the heme iron was assigned using an approximate method since there is no parameter for

Table 3. Correlation coefficient (*r*) between activity and docking parameters of 30 artemisinin derivatives

Variable	log(D-6)	log(W-2)
Binding Energy	-0.933*	-0.814*
O1-Fe distance	-0.555*	-0.311
O2-Fe distance	-0.510*	-0.263
O13-Fe distance	-0.514*	-0.243
O11-Fe distance	-0.559*	-0.403

*Correlation is significant at the 0.01 level (2-tailed).

iron in SYBYL. But in our study the atomic charges of all atoms, including Fe, were calculated at a high level of accuracy.

3.2 Quantitative Structure-Activity Relationship

Since the endoperoxide group is very crucial for the binding of artemisinin and heme, relationships between biological activity and 5 parameters, i.e., the O1-Fe distance, O2-Fe distance, O13-Fe distance, O11-Fe distance, and the binding energy, were investigated. Using activities in logarithm units, values of the linear correlation coefficient (*r*), which measures the degree of linear relationship, were listed in Table 3. The value of *r* is always between -1 and 1 or $-1 \leq r \leq 1$, 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. Together with *r*, the level of significance, which determines the reliability of the

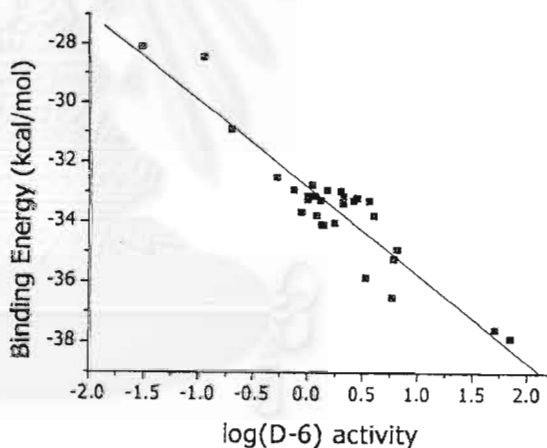


Figure 8. Relationship between each binding energy and antimalarial activity of D-6.

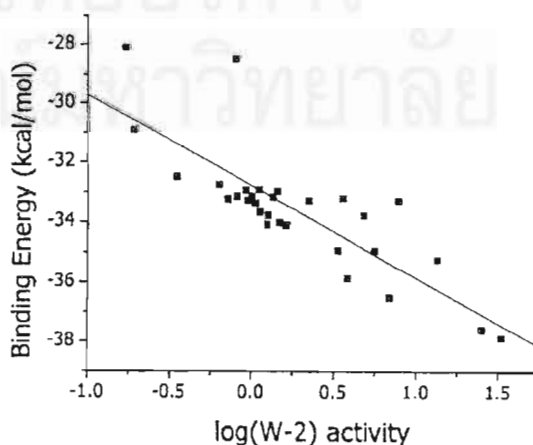


Figure 9. Relationship between each binding energy and antimalarial activity of W-2.

Table 4. Docking results of heme and 5 additional artemisinin derivatives

Compound no.	Energy (kcal/mol)	% Occurrence	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe Distance (Å)
32	-34.73	29	1.98	2.66	4.68	4.69
33	-33.55	28	2.01	2.69	4.68	4.79
34	-32.89	38	1.99	2.49	4.57	5.09
35	-35.35	37	2.02	2.81	4.69	4.88
36	-35.30	26	1.95	2.41	4.51	4.63

Table 5. Actual and predicted activities of 5 additional artemisinin derivatives

No.	log(D-6) activity		log(W-2) activity	
	Actual	Predict	Actual	Predict
32	0.5832	0.6005	0.6117	0.5194
33	0.1614	0.0522	0.0212	0.1201
34	0.6749	0.7853	0.7404	0.6540
35	0.6253	0.7704	0.7042	0.6431
36	0.4065	0.2489	0.1553	0.2634

statistics, was also reported. Among the 5 parameters, binding energies showed the best correlation with activity, with *r* values of -0.933 for log activity of D-6 and -0.814 for log

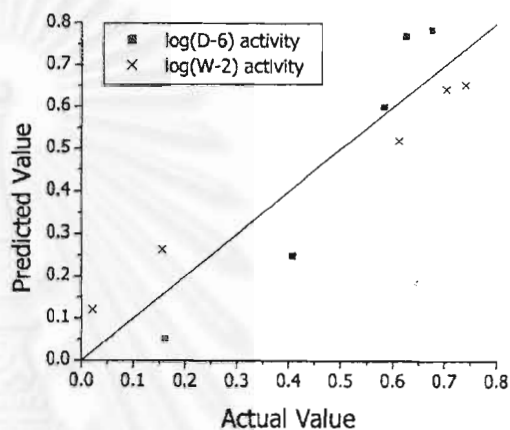


Figure 11. Relationship between actual and predicted activities.

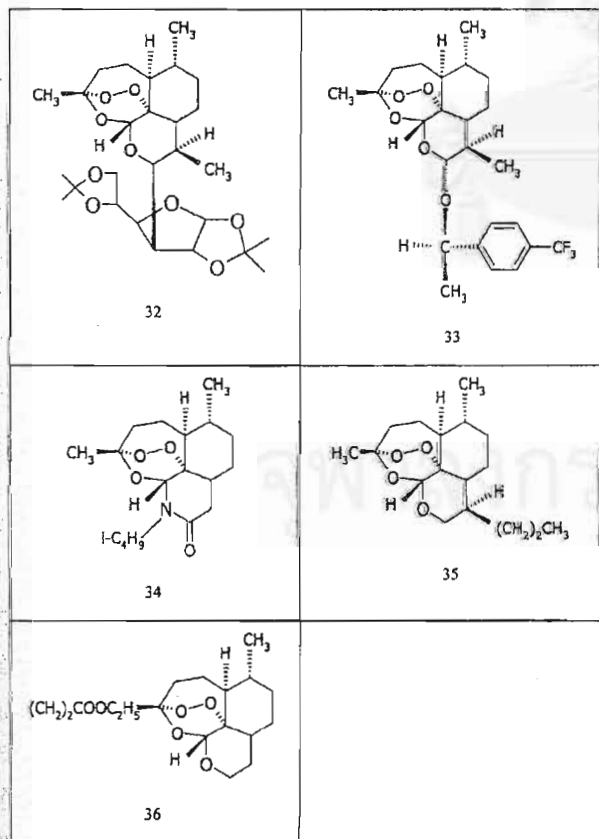


Figure 10. Structures of 5 additional artemisinin derivatives.

activity of W-2, both at the 0.01 level of significance. The minus sign of *r* indicates that a compound which binds more firmly with heme (lower binding energy) has a higher activity (also see Figures 8 and 9). All four O-Fe distances have only moderate relationships with the log activity of D-6 (*r* values of -0.510 to -0.559) and almost no relationship with the log activity of W-2. Also this suggests that a compound with a shorter O-Fe distance has a higher activity. These docking results imply the importance of Fe in the mechanism of action.

To investigate the predictivity of the binding energy for antimalarial activity, the QSAR models were constructed for both the log activities of D-6 and W-2 using the multiple linear regression method. The equations as shown below were obtained.

$$\log(\text{D-6 activity}) = -9.749 - 0.298 * (\text{binding energy}) \quad (1)$$

$$\log(\text{W-2 activity}) = -7.017 - 0.217 * (\text{binding energy}) \quad (2)$$

In order to test the quality of these two models, activities of 5 additional artemisinin derivatives were predicted. Structures of these compounds are schematically displayed in Figure 10. The docking calculations were performed with the same conditions as those of the 30 original artemisinin derivatives.

The docking results are shown in Table 4. The predicted and actual values of both log activities of D-6 and W-2 are given in Table 5. A plot of actual activity against predicted activity is drawn in Figure 11. The plot shows the good agreement between predicted and actual values, which indicates the good quality of our QSAR models and once again confirms the proposed mechanism of action [12, 14].

Acknowledgements

The authors would like to thank the Computer Center of the University of Vienna, Austria for providing computing resources at their workstation clusters and the Austrian-Thai Center (ATC) for Computer Assisted Chemical Education and Research, Department of Chemistry, Faculty of Science, Chulalongkorn University, Thailand for computer resources and other facilities. Tonmunpheap S. would like to thank Prof. Dr. Peter Wolschann, Institute for Theoretical Chemistry and Molecular Structural Biology, University of Vienna, Austria for helpful discussions and comments.

References

[1] Moore, D.V., and Lanier, J.E., Observations on Two *Plasmodium falciparum* Infections with an Abnormal Response to Chloroquine, *Am. J. Trop. Med. Hyg.* 10, 5-9 (1961).

[2] Mockenhaupt, F.P., Mefloquine Resistance in *Plasmodium falciparum*, *Parasitol. Today* 11, 248-253 (1995).

[3] Olliaro, P.L., and Trigg, P.I., Status of Antimalarial Drugs Under Development, *Bull. WHO* 73, 565-571 (1995).

[4] Qinghaosu Antimalaria Coordinating Research Group, Antimalaria Studies on Qinghaosu, *Chin. Med. J.* 92, 811-816 (1979).

[5] Klayman, D.L., Qinghaosu (Artemisinin): An Antimalarial Drug from China, *Science* 228, 1049-1055 (1985).

[6] Luo, X.D., and Shen, C.-C., The Chemistry, Pharmacology, and Clinical Applications of Qinghaosu (Artemisinin) and Its Derivatives, *Med. Res. Rev.* 7, 29-52 (1987).

[7] China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials, Chemical Studies on Qinghaosu (Artemisinin), *J. Trad. Chin. Med.* 2, 3-8 (1982).

[8] Meshnick, S.R., Thomas, A., Ranz, A., Xu, C.-M., and Pan, H.-Z., Artemisinin (Qinghaosu): The Role of Intracellular Hemin in its Mechanism of Antimalarial Action, *Mol. Biochem. Parasitol.* 49, 181-190 (1991).

[9] Meshnick, S.R., Yang, Y.-Z., Lima, V., Kuypers, F., Kamchonwongpaisan, S., and Yuthavong, Y., Iron-Dependent Free Radical Generation from the Antimalarial Agent Artemisinin (Qinghaosu), *Antimicrob. Agents Chemother.* 37, 1108-1114 (1993).

[10] Posner, G.H., Oh, C.H., Wang, D., Gerena, L., Milhous, W.K., Meshnick, S.R., and Asawamahsakda, 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, 1256-1258 (1994).

[11] Hong, Y.-L., Yang, Y.Z., and Meshnick, S.R., The interaction of artemisinin with malarial hemozoin, *Mol. Biochem. Parasitol.* 63, 121-128 (1994).

[12] Posner, G.H., Cummings, 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, 5885-5886 (1995).

[13] Asawamahsakda, W., Ittarat, I., Pu, Y.-M., Ziffer, H., and Meshnick, S.R., Reaction of Antimalarial Endoperoxides with Specific Parasite Proteins, *Antimicrob. Agents Chemother.* 38, 1854-1858 (1994).

[14] Jefford, C.W., Vicente, M.G.H., Jacquier, Y., Favarger, F., Mareda, J., Millasson-Schmidt, P., Brunner, and G., Burger, U., The Deoxygenation and Isomerization of Artemisinin and Artemether and Their Relevance to Antimalarial Action, *Helv. Chim. Acta* 79, 1475-1487 (1996).

[15] Shukla, K.L., Gund, T.M., and Meshnick, S.R., Molecular Modeling Studies of the Artemisinin (Qinghaosu)-Hemin Interaction: Docking between the Antimalarial Agent and its Putative Receptor, *J. Mol. Graph.* 13, 215-222 (1995).

[16] Morris, G.M., Goodsell, D.S., Huey, R., and Olson, A.J., AutoDock Version 2.4, The Scripps Research Institute, Department of Molecular Biology, MB-5, La Jolla, California, USA.

[17] Weiner, S.J., Kollman, P.A., Nguyen, D.T., and Case, D.A., An All Atom Force Field for Simulations of Proteins and Nucleic Acids, *J. Comput. Chem.* 7, 230-252 (1986).

[18] Halgren, T.A., Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters, *J. Am. Chem. Soc.* 114, 7827-7843 (1992).

[19] <http://www.scripps.edu/pub/olson-web/doc/autoflex/parameters.html>

[20] Morris, G.M., Goodsell, D.S., Huey, R., and Olson, A.J., Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4, *J. Comput.-Aided Mol. Des.* 10, 293-304 (1996).

[21] Tonmunpheap, S., Kokpol, S., Parasuk, V., Wolschann, P., Winger, R.H., Liedl, K.R., and Rode, B.M., Comparative Molecular Field Analysis of Artemisinin Derivatives: *Ab initio* versus Semiempirical Optimized Structures, *J. Comput.-Aided Mol. Des.* 12, 397-409 (1998).

[22] The comparison and geometries for all derivatives can be obtained through the private communication with the authors.

[23] Gaussian 94, Revision B.3, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C., and Pople, J.A., Gaussian, Inc., Pittsburgh PA, 1995.

[24] <http://www.ccdc.cam.ac.uk/>

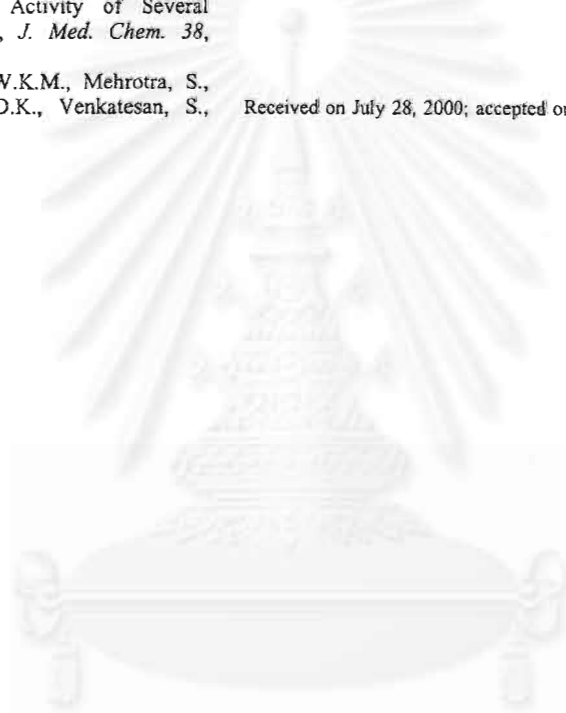
[25] Tonmunpheap, S., Parasuk, V., and Kokpol, S., Automated Molecular Docking of Artemisinin Derivatives to Heme, The Fourth Annual National Symposium on Computational Science and Engineering, 27-29 March 2000, Faculty of Engineering, Kasetsart University, Bangkok, Thailand, pp. 181-189

[26] Lin, A.J., Klayman, D.L., and Milhous, W.K., Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives, *J. Med. Chem.* 30, 2147-2150 (1987).

[27] Lin, A.J., Li, L.-Q., Klayman, D.L., George, C.F., and Flippen-Anderson, J.L., Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. 3. Aromatic Amine Analogues, *J. Med. Chem.* 33, 2610-2614 (1990).

- [28] Lin, A.J., Li, L.-Q., Andersen, S.L., and Klayman, D.L., Antimalarial Activity of New Dihydroartemisinin Derivatives. 5. Sugar Analogues, *J. Med. Chem.* 35, 1639–1642 (1992).
- [29] Avery, M.A., Gao, F., Chong, W.K.M., Mehrotra, S., and Milhous, W.K., Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 1. Synthesis and Comparative Molecular Field Analysis of C-9 Analogs of Artemisinin and 10-Deoxyartemisinin, *J. Med. Chem.* 36, 4264–4275 (1993).
- [30] Lin, A.J., and Miller, R.E., Antimalarial Activity of New Dihydroartemisinin Derivatives. 6. α -Alkylbenzylic Ethers, *J. Med. Chem.* 38, 764–770 (1995).
- [31] Pu, Y.M., Torok, D.S., Ziffer, H., Pan, X.-Q., and Meshnick, S.R., Synthesis and Antimalarial Activity of Several Fluorinated Artemisinin Derivatives, *J. Med. Chem.* 38, 4120–4124 (1995).
- [32] Avery, M.A., Bonk, J.D., Chong, W.K.M., Mehrotra, S., Miller, R., Milhous, W., Goins, D.K., Venkatesan, S., Wyandt, C., Khan, I., and Avery, B.A., Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 2. Effect of Heteroatom Substitution at O-11: Synthesis and Bioassay of *N*-Alkyl-11-aza-9-desmethylartemisinins, *J. Med. Chem.* 38, 5038–5044 (1995).
- [33] Avery, M.A., Mehrotra, S., Bonk, J.D., Vroman, J.A., Goins, D.K., and Miller, R., Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 4. Effect of Substitution at C-3, *J. Med. Chem.* 39, 2900–2906 (1996).
- [34] Avery, M.A., Mehrotra, S., Johnson, T.L., Bonk, J.D., Vroman, J.A., and Miller, R., Structure-Activity Relationships of the Antimalarial Agent Artemisinin. 5. Analogs of 10-Deoxyartemisinin Substituted at C-3 and C-9, *J. Med. Chem.* 39, 4149–4155 (1996).

Received on July 28, 2000; accepted on October 16, 2000



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Appendix D

Automated calculation of docking of artemisinin to heme

Somsak Tonmunphean,
Vudhichai Parasuk, Sirirat Kokpol

J. Mol. Model. 7 (2001): in press.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Original Paper

Automated calculation of docking of artemisinin to heme

Somsak Tonmunphean · Vudhichai Parasuk(✉) · Sirirat Kokpol

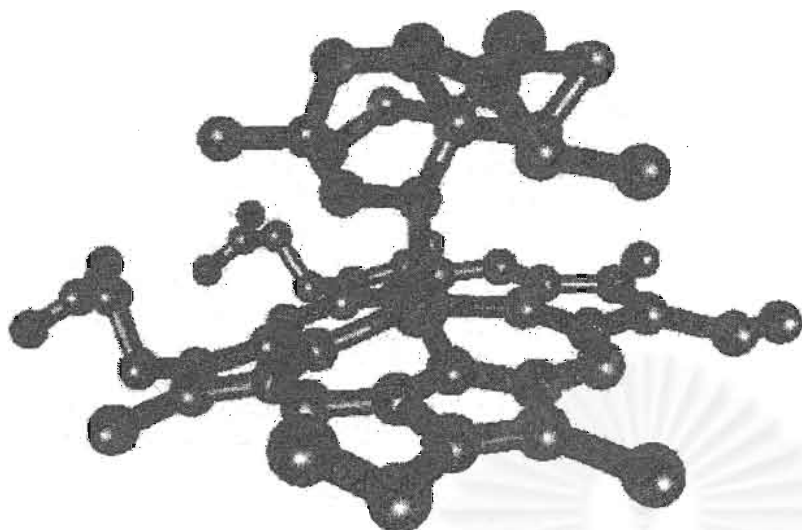
S. Tonmunphean · V. Parasuk · S. Kokpol
Department of Chemistry, Faculty of Science, Chulalongkorn University, Patumwan, Bangkok, 10330
Thailand

✉ E-mail: parasuk@atc.atccu.chula.ac.th
Phone: 662 218 5221
Fax: 662 252 1730

Received: 28 September 2000 / Accepted: 1 February 2001

Abstract. We report automated molecular docking of artemisinin to heme. The effects of atomic charges, and ligand and heme structures on the docking results were investigated. Several charge schemes for both artemisinin and heme, artemisinin structures taken from various optimization methods, X-ray data, and five heme models, were employed for this purpose. The docking showed that artemisinin approaches heme by pointing O1 at the endoperoxide linkage toward the iron center, a mechanism that is controlled by steric hindrance. This result differs from that reported by Shukla et al. which suggested that heme binds with artemisinin at the O2 position. The docking results also depended on the structures of both artemisinin and heme. Moreover, the atomic charges of heme have a significant effect on the docking configurations.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



Docking configuration between artemisinin and heme-hemin

Keywords. Docking - Antimalarial drug - Endoperoxide - Mechanism of Action - Heme

Introduction

Malaria is one of the most widespread and prevalent endemic diseases; it threatens approximately 40 percent of the world's population in more than 90 countries. This disease is estimated to cause approximately 300 to 500 million illnesses and up to 3 million deaths each year [1]. This tremendous prevalence might be partly because of the resistance of malaria parasites to most antimalarial agents, e.g. chloroquine, quinine, and mefloquine [2, 3]. Artemisinin (Fig. 1), a sesquiterpene endoperoxide isolated from a Chinese medicinal herb [4], is, however, a potent antimalarial drug against the resistant strains of *Plasmodium falciparum* [5, 6]. Its unusual structure might be indicative of a different mode of action from those of the other antimalarial drugs, and hence the high potency against the resistant strains. Although the mechanism of its antimalarial activity is still in doubt, there is general agreement on the significance of the endoperoxide group of artemisinin to the antimalarial activity. This is evident from the inactivity of the deoxyartemisinin compound that lacks the endoperoxide moiety [7]. In addition, in-vitro experiments revealed that iron is required for artemisinin to have antimalarial activity [8, 9, 10].

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

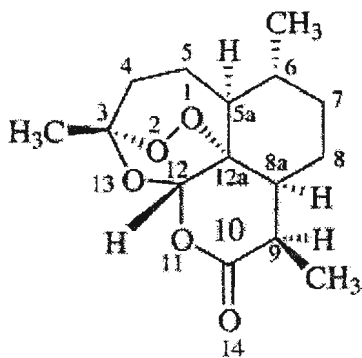


Fig. 1. The structure of artemisinin, with atom numbering

In humans, malarial parasites digest more than 70% of the hemoglobin within the infected red blood cell [11], giving globin and heme as the products. The globin is hydrolyzed to give amino acids, which are used in protein synthesis by the parasite. The toxic heme (Fig. 2) is mostly detoxified by a specific mechanism of heme polymerization into hemozoin. The heme polymerization is a target for some antimalarials, such as chloroquine, that inhibit this process [12]. A recent study reported that artemisinin also inhibits heme polymerization [13]. The chloroquine-resistant strain of *Plasmodium berghei* that lacks hemozoin, possibly because heme polymerization does not occur, is also resistant to artemisinin [14]. This supports the view that inhibition of heme polymerization is the mode of action of artemisinin. It is very possible that artemisinin interacts with heme and hence inhibits the polymerization process.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

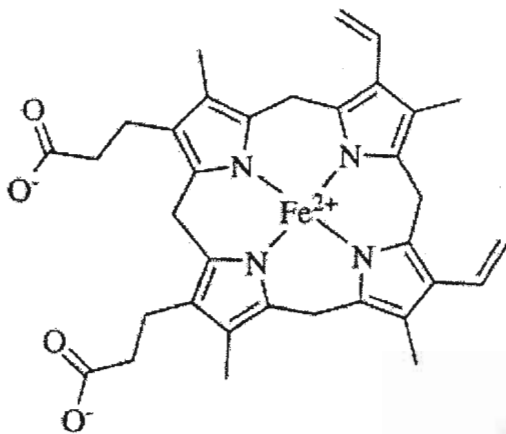


Fig. 2. The structure of heme

It has been proposed that heme iron attacks the endoperoxide linkage of artemisinin either at the O1 [15] or O2 position [16] (Fig. 3). In pathway A, heme iron attacks the compound at the O2 position and produces a free radical at the O1 position. Later it rearranges to form the C4 free radical. In pathway B, heme iron attacks the compound at the O1 position and produces a free radical at the O2 position. After that the C3-C4 bond is cleaved to give a carbon radical at C4. It has been suggested that the C4 free radical in both pathways is an important substance in antimalarial activity [10].

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

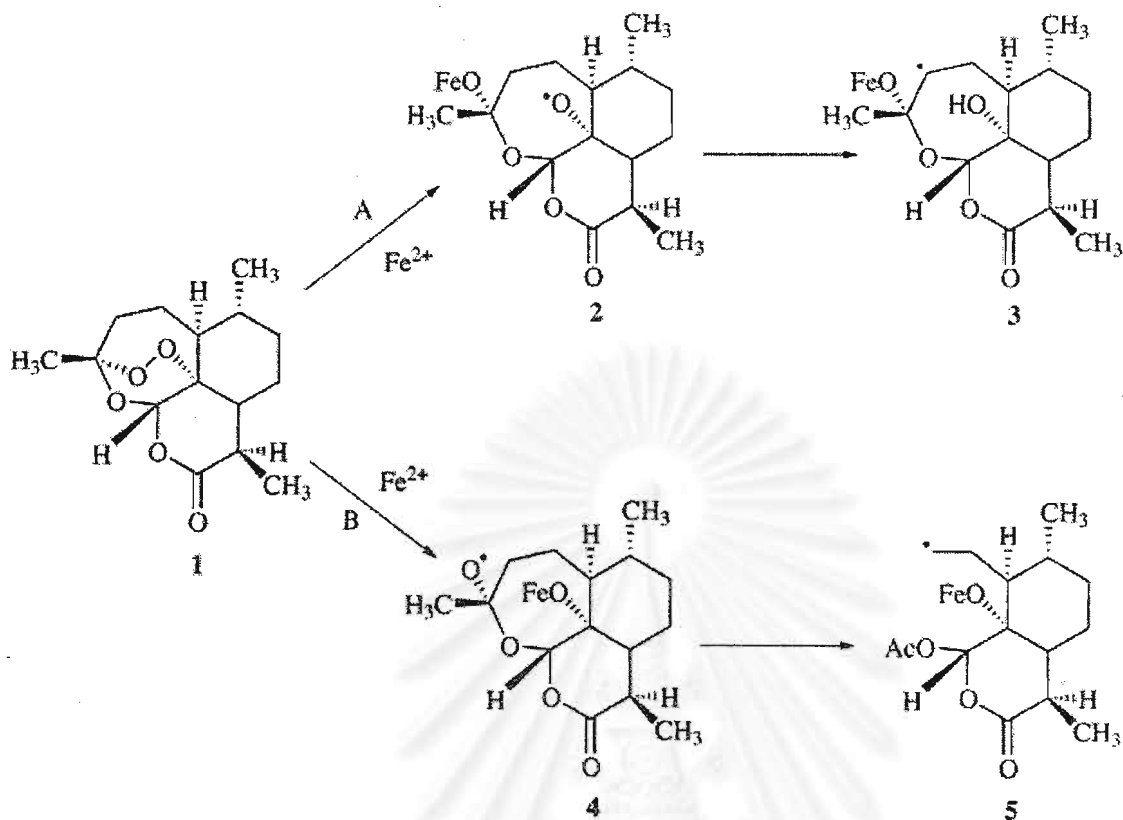


Fig. 3. Proposed mechanism of action of artemisinin

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

For this reason, Shukla and co-workers [17] studied the docking of artemisinin and deoxyartemisinin with hemin [Fe(II) and Fe(III)] using the Dock module in SYBYL software, a direct docking algorithm. In their study the artemisinin structure was built from the crystallographic X-ray structure of artemether. Although the study did not elaborate on how the structure of artemisinin was obtained from artemether, it is very likely that the geometry optimization was performed at either the molecular mechanics or semi-empirical level, because only these methods are available in SYBYL. For their docking calculations, only three orientations of artemisinin around the hemin molecule were considered. Furthermore, the Gasteiger method, an empirical method implemented in the SYBYL, was used for the atomic charge calculations. Because this empirical method has no parameters for iron, however, the charge of the heme iron was assigned under the assumption that the change in the charge distribution of the heme iron should be equal to that of the heme model where iron was replaced by aluminum. Moreover, the general parameters for metals were used in the docking calculations. The docking scheme they employed might influence the docking result in favor of one of the heme-artemisinin configurations and yield an inaccurate model for the complex. It is quite important

to have an accurate model for the heme-artemisinin complex, because this knowledge can be used to design better and more potent antimalarial drugs.

In this study, automated docking calculations were performed to eliminate the bias in selecting preferred configurations (orientations). Thus, all possible configurations between heme and artemisinin were explored. The crystallographic X-ray structure of artemisinin was used for artemisinin instead of that of artemether, which is quite different from the artemisinin structure, especially at the lactone ring. In addition, because few crystallographic X-ray structures of artemisinin derivatives are available, it is worth establishing a suitable geometry optimization scheme to determine structures of artemisinin derivatives for further investigations [18]. For the heme iron, accurate ab initio calculations were performed to obtain its atomic charge (and those of artemisinin) instead of using a crude approximation for the charge of iron, and specific parameters for iron were used in the docking calculations. The effects of different heme structures were also considered. Thus, five heme structures taken from the literature were studied.

The knowledge obtained from this study has been used as a guide for series of docking experiments between heme and artemisinin derivatives and we found a very pronounced relationship between their binding energies and antimalarial activity [18].

Computational details

Docking calculations

AutoDock 2.4 [19], an automated docking program, was used for the docking calculations. The automated docking is performed using a simulated annealing Monte Carlo simulation in combination with a rapid grid-based energy-evaluation method. A grid map of dimensions $25 \times 25 \times 25 \text{ \AA}^3$ with

a 0.5 \AA spacing was selected. The combined AMBER/MMFF parameters [20, 21] were chosen for the Lennard-Jones 12,6 potentials and Coulomb potentials to calculate the interaction energy, instead of using the AMBER force field that contains no parameters for iron. These parameters were taken from the authors of the program [22].

In one docking calculation, the simulations were performed for 100 annealing cycles. At the first cycle, the initial annealing temperature (RT) was set to $100 \text{ kcal mol}^{-1}$ and then the temperature was reduced at the rate of 0.90 per cycle. During each cycle, the ligand was gradually moved by a random displacement with a maximum translation step of 0.2 \AA and a maximum orientation step of 5° . The energy of the new configuration was then calculated. The selection of the new configuration was based on the Metropolis algorithm [23]. The cycle terminates if the ligand makes 30,000 accepted or 30,000 rejected moves. Then the simulation moves to the next cycle.

Because the Monte Carlo simulation is based on random movements, the final docked configuration depends on the starting configuration. To avoid any bias and to generate as many final docked configurations as possible, the starting configuration was assigned randomly 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 \AA were grouped together. The lowest energy configuration was selected as a representative for each group. Our attention was focused on the group with the highest number of members, referred to as "the most occurring configuration". Thus, it is most probable that this configuration represents the real system.

Ligand and receptor structures

In addition to the crystallographic X-ray structure, the docking of heme and the optimized geometries of artemisinin obtained at AM1, HF/3-21G, and HF/6-31G* levels of theory were investigated (these structures were taken from Ref. [24]). For the receptor molecule, five heme structures, i.e., heme-pdb, heme-model, heme-hemin, heme-deoxy, and heme-oxy, were considered. These structures are all different owing to the source of heme and the oxidation state of iron. The first structure, heme-pdb, was taken from the Protein Data Bank (id 1CTJ) [25]. In this structure, Fe positions itself slightly above the porphyrin plane (Fig. 4a). The second structure, heme-model, which was taken from the AMBER database [26] has the planar geometry (Fig. 4b). The third structure, heme-hemin, was modified from the crystallographic X-ray structure of chlorohemin of the Cambridge Crystallography Data Bank [27]. This structure has a pyramidal shape with Fe on the top (Fig. 4c).



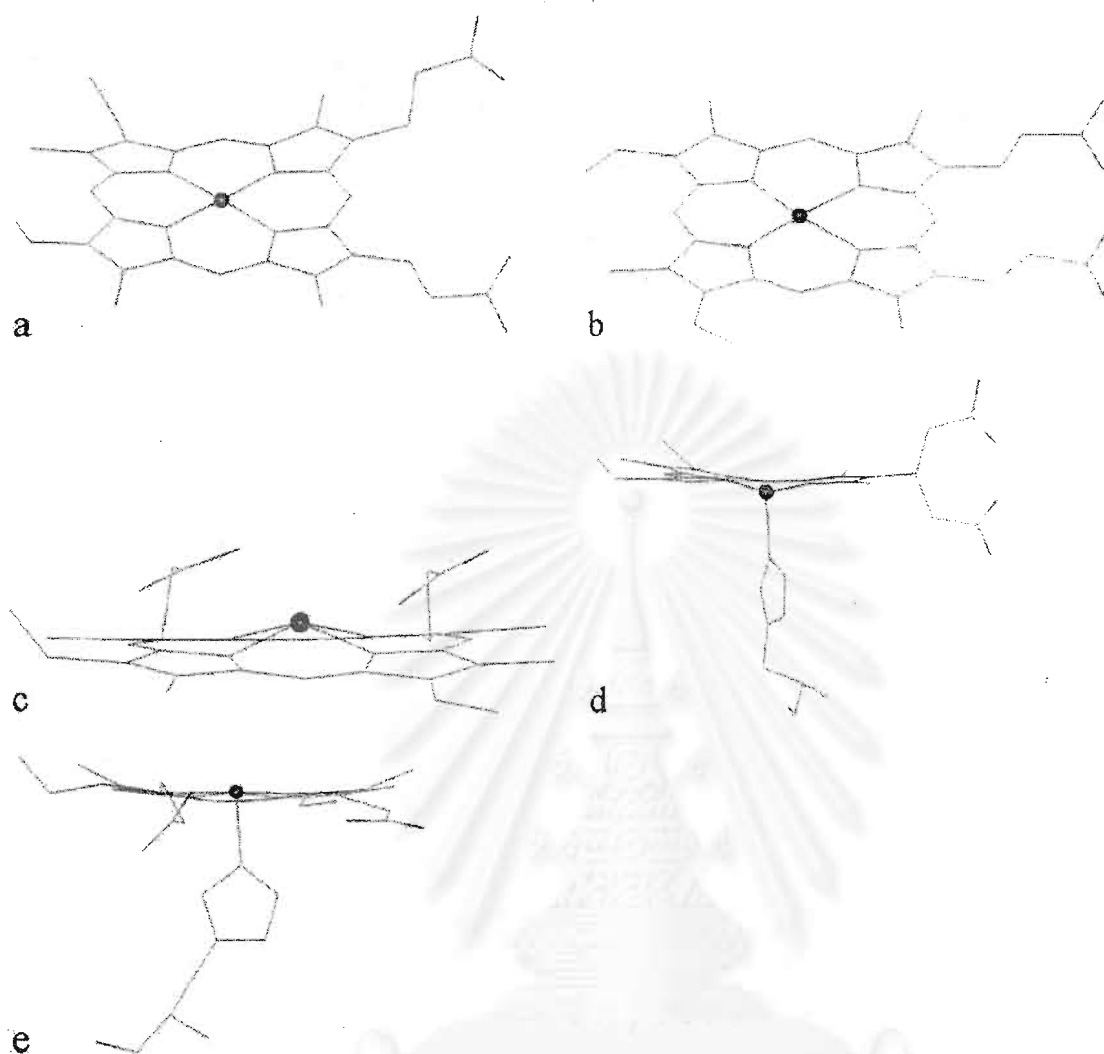


Fig. 4. The structures of the five heme compounds: (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, (e) heme-oxy

In the process of hemoglobin degradation by the malaria parasite, the proximal ligand may possibly still be attached to the heme iron and, therefore, it is very possible that the histidine remains with the heme structure. As a result, the fourth and the fifth structures, heme-deoxy and heme-oxy, respectively, were obtained from the modifications of deoxy and oxy forms of hemoglobin which contain histidine as the proximal. Both deoxy and oxy forms of hemoglobin were taken from the Protein Data Bank (id 1A3 N and 1HHO, respectively). In the heme-deoxy, the histidine pulls the Fe atom to lie below the protoporphyrin plane and gives it a basin-like structure (Fig. 4d). In the oxy hemoglobin structure, there are six coordinations for heme iron, i.e. with four N atoms in the protoporphyrin ring, with the proximal ligand (histidine), and with O₂. Thus, for docking purposes, the O₂ coordination was deleted while maintaining the coordinates of the rest; this modified structure was taken as the receptor structure. As in heme-deoxy, the protoporphyrin plane has a basin-like structure, because of the attraction to the heme iron by histidine. Interaction with O₂ causes the Fe

atom to be drawn up above the plane (Fig. 4e), however, and thus results in a structure which is markedly different from the heme-deoxy.

Atomic charge calculations

To investigate the effect of the atomic charge on docked configurations, atomic charges of both artemisinin and heme obtained at various levels of theory were used. For heme, the ZINDO/S, STO-3G, HF/3-21G, and HF/6-311G** atomic charges were calculated. For artemisinin, atomic charge calculations were performed at AM1, PM3, HF/3-21G, HF/D95, HF/6-31G*, and HF/6-311G**. All quantum chemical calculations were carried out using the Gaussian 94 program [28].

Results and Discussion

Effect of atomic charges

In docking calculations, the electrostatic potential is built from atomic charges. Therefore, the choices for atomic charges of both the ligand and receptor would have an effect on the docking results. Using charges obtained from ZINDO/S, HF/STO-3G, HF/3-21G, and HF/6-311G** levels of theory for heme-pdb, the docking to the artemisinin X-ray structure with HF/3-21G charges was performed. The results in Table 1 showed that the docking configurations depend on the heme-pdb atomic charges and especially the charge of Fe. With the exception of ZINDO/S charges, all docking calculations agree that the heme iron binds with endoperoxide oxygens, where the O1-Fe distance is the shortest. Among these calculations, docking with HF/6-311G** charges yielded the shortest O1-Fe distance of 2.51 Å. This O1-Fe distance is markedly much shorter than those predicted using HF/STO-3G (2.71 Å) and HF/3-21G (2.70 Å) charges. For the binding energy, the docking with HF/STO-3G charges gave the lowest energy while that with HF/6-311G** charges gave the second lowest. Thus, the employed charge scheme for heme does have a profound effect on the docking result. It is, however, quite difficult to judge which charge scheme leads to the most accurate result, because there is no supporting experimental evidence. Theoretically, HF/6-311G** is the most accurate level of theory employed. It is, therefore, reasonable to choose atomic charges from HF/6-311G** for heme in further docking calculations. To study the effect of atomic charges of artemisinin, the docking calculations using various charge schemes, i.e., AM1, PM3, HF/3-21G, HF/D95, HF/6-31G*, and HF/6-311G** for the artemisinin X-ray structure and HF/6-311G** charges for heme-pdb structure were performed. The docking results are given in Table 2 and the atomic charges of four oxygen atoms in artemisinin for each charge scheme are listed in Table 3. From Table 2, the dockings with *ab initio* charges (HF/3-21G, HF/D95, HF/6-31G*, and HF/6-311G**) gave similar results, whereas those with semi-empirical charges (AM1 and PM3) gave longer O-Fe distances. Thus, for the sake of saving CPU times, the HF/3-21G charges were chosen for artemisinin.

Table 1. Results for docking of heme-pdb with different atomic charges and the artemisinin X-ray structure with HF/3-21G charge

Heme-pdb charge	Fe charge	Energy (kcal mol ⁻¹)	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
ZINDO/S	0.127	-30.70	5.50	5.34	<u>3.10</u> ^a	3.64
STO-3G	0.780	-31.57	<u>2.71</u>	3.69	5.45	5.69
HF/3-21G	1.371	-30.44	<u>2.70</u>	3.57	5.41	5.70
HF/6-311G**	1.589	-31.55	<u>2.51</u>	3.09	5.16	5.37

^aThe underlined values are the shortest O-Fe distances

Table 2. Results for docking of heme-pdb with HF/6-311G** charge and the artemisinin X-ray structure with different atomic charges

Artemisinin atomic charges	Energy (kcal mol ⁻¹)	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
AM1	-30.70	<u>2.76</u> ^a	3.61	5.41	5.79
PM3	-30.58	<u>2.73</u>	3.59	5.41	5.75
HF/3-21G	-31.55	<u>2.51</u>	3.09	5.16	5.37
HF/D95	-31.71	<u>2.56</u>	3.11	5.19	5.42
HF/6-31G*	-31.40	<u>2.50</u>	3.10	5.16	5.37
HF/6-311G**	-30.58	<u>2.53</u>	3.03	5.10	5.42

^aThe underlined values are the shortest O-Fe distances

Table 3. Atomic charges of four oxygen atoms in artemisinin for all charge schemes

Artemisinin atomic charges	O1 charge	O2 charge	O13 charge	O11 charge
AM1	-0.155	-0.142	-0.289	-0.336
PM3	-0.132	-0.127	-0.252	-0.281
HF/3-21G	-0.374	-0.359	-0.669	-0.710
HF/D95	-0.322	-0.259	-0.489	-0.478
HF/6-31G*	-0.405	-0.368	-0.689	-0.666
HF/6-311G**	-0.333	-0.282	-0.510	-0.461

Effect of artemisinin structure

In our previous study [24], artemisinin was geometry-optimized at various levels of accuracy, ranging from the semi-empirical CNDO and AM1 to ab initio HF/STO-3G, HF/3-21G, and HF/6-31G**. Comparison of these optimized geometries with the crystallographic X-ray structure [29] showed that HF/3-21G gave geometry parameters in good agreement with those of crystallographic X-ray data, especially for the bond length of the endoperoxide linkage, whereas AM1 and HF/6-31G* yielded an O-O bond distance that was too short. This shorter O-O bond length for AM1 and HF/6-31G* is not

only found in artemisinin but also in other peroxide systems [30]. The HF/3-21G method is, therefore, recommended for the optimization of artemisinin derivatives. This recommendation is, however, based on geometrical criteria only, which does not necessarily guarantee good docking results.

To validate the use of this optimized artemisinin structure, the docking calculations between heme-pdb with HF/6-311G** atomic charges and the AM1, HF/3-21G, and HF/6-31G* optimized structures of artemisinin were performed. The results were compared with those obtained using the artemisinin crystallographic X-ray structure. For the optimized structures, atomic charges of artemisinin were taken according to the optimization methods, i.e. AM1 charges for the AM1 structure, etc. For the X-ray structure, three docking calculations using AM1, HF/3-21G, and HF/6-31G* charges for artemisinin were performed. The docking results are given in Table 4. Comparison of the configurations which occur most often reveals good agreement between the docking using the X-ray structure and HF/3-21G structure for artemisinin. The largest deviation is 0.03 Å (O11-Fe distance). Comparing the AM1 and the X-ray structures, the optimized structure yielded an O1-Fe distance that was short by 0.2 Å, with the largest deviation 0.55 Å (O2-Fe distance). Although much better for docking than the AM1 structure, when comparing the HF/6-31G* and X-ray structures, the optimized structure gave an O1-Fe distance that was too long by 0.08 Å, with the largest deviation of 0.12 Å (O11-Fe distance). The discrepancy between the docking results obtained from the AM1 and the HF/6-31G* structures and the X-ray structure is clearly rooted in the deficiency of the methods, which yielded O-O distances that were too short. Thus, the method which gives a good structure (compared with the X-ray structure) will also give good docking results. HF/3-21G is, therefore, the recommended method for geometry optimization of artemisinin derivatives in further study although it has a lower level of accuracy than HF/6-31G*. It can be argued that for artemisinin derivatives it is possible that the good agreement between the HF/3-21G and the X-ray structures no longer exists, so it would be wiser to employ the more accurate method, HF/6-31G*. From previous calculations on artemisinin, however, and the current docking results the difference between the structures obtained from the two methods is not pronounced. Thus, the HF/3-21G method is still preferred, because of its faster computation time.

Table 4. Results for docking of heme-pdb with HF/6-311G** charge and artemisinin optimized structures at various levels of theory

Artemisinin structure	Artemisinin atomic charges	Energy (kcal mol ⁻¹)	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
	AM1	-30.70	<u>2.76</u> ^a	3.61	5.41	5.79
X-Ray	HF/3-21G	-31.55	<u>2.51</u>	3.09	5.16	5.37
	HF/6-31G*	-31.40	<u>2.50</u>	3.10	5.16	5.37
AM1	AM1	-30.60	<u>2.56</u>	3.06	5.16	5.55
HF/3-21G	HF/3-21G	-31.40	<u>2.49</u>	3.12	5.14	5.40
HF/6-31G*	HF/6-31G*	-31.38	<u>2.58</u>	3.13	5.14	5.49

^aThe underlined values are the shortest O-Fe distances

Effect of heme structure

To investigate the effect of the heme structure, five heme structures were selected as described in the section on computational details. The atomic charges were assigned as HF/6-311G** charges for all five heme molecules. For artemisinin compounds, the HF/3-21G optimized structure and atomic charges were used. The results are shown in Table 5 and Fig. 5.

Table 5. Results for docking of different heme structures and artemisinin HF/3-21G optimized structure

Heme	Energy (kcal mol ⁻¹)	Frequency (%)	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
heme-pdb	-31.40	25	<u>2.49</u> ^a	3.12	5.14	5.40
heme-model	-29.92	22	<u>2.75</u>	3.66	5.38	5.79
heme-hemin	-33.13	24	<u>2.00</u>	2.65	4.67	4.90
heme-deoxy	-31.03	39	5.95	5.53	<u>3.26</u>	4.03
	-30.18	13	<u>3.19</u>	4.18	5.85	6.19
heme-oxy	-32.32	51	<u>2.52</u>	3.32	5.12	5.57

^aThe underlined values are the shortest O-Fe distances

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

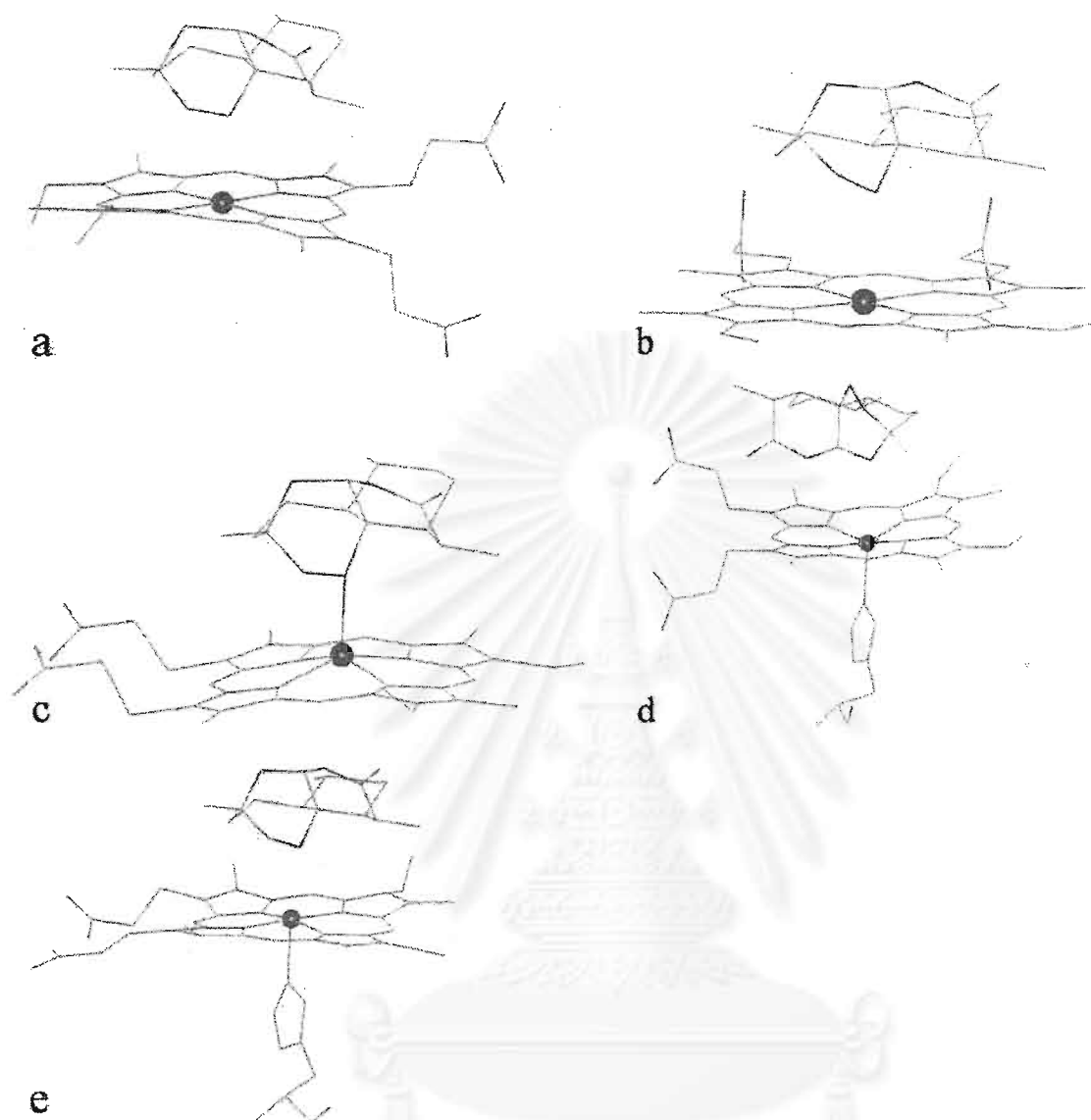


Fig. 5. Docking configuration between artemisinin and (a) heme-pdb, (b) heme-model, (c) heme-hemin, (d) heme-deoxy, (e) heme-oxy

The heme structure chosen does have an effect on the docking results. Although we could not observe agreement on O-Fe distances, all docking calculations with different heme structures (except heme-deoxy) suggested that artemisinin prefers to dock at endoperoxide oxygens (O1 and O2). Using heme-pdb for the heme structure, the docking results showed that artemisinin pointed its endoperoxide moiety toward the heme iron for the most occurring configuration. The O1-Fe and O2-Fe distances were measured and found to be 2.49 Å and 3.12 Å, respectively (Fig. 5a); the binding energy obtained was $-31.40 \text{ kcal mol}^{-1}$. Owing to the planar structure of the heme-model, the repulsion between artemisinin and the protoporphyrin ring of heme prevents artemisinin from approaching the heme iron as closely as for heme-pdb. Thus, the O1-Fe and O2-Fe distances of 2.75 Å and 3.66 Å (Fig. 5b) were obtained, with a binding energy of $-29.92 \text{ kcal mol}^{-1}$, the weakest among the heme structures

investigated. Unlike the first two models, the distances between the endoperoxide oxygens and Fe for heme-hemin are very short; 2.00 Å and 2.65 Å for O1-Fe and O2-Fe (Fig. 5c), with a binding energy of -33.13 kcal mol⁻¹ (the lowest). This is probably because of the pyramidal-like structure of heme-hemin which facilitates the approach of Fe to the endoperoxide moiety. The O1-Fe distance of 2.00 Å is comparable with the experimental bond length between the heme iron and oxygen atom in oxyhemoglobin A (1.86 Å), taken from the Protein Data Bank (id 1HHO).

For the heme-deoxy, because of its basin-like structure (see Fig. 4d), the binding with the endoperoxide moiety of artemisinin is less favorable and a stronger O13-Fe attraction is resulted (binding energy -31.03 kcal mol⁻¹). This could be observed from the most occurring configuration, which has the shorter O13-Fe distance of 3.26 Å, compared with 5.95 and 5.53 Å for O1-Fe and O2-Fe (Fig. 5d). Interestingly, the second most occurring configuration has shorter O1-Fe and O2-Fe distances. Still, this distance is longer than those obtained from the docking with other heme structures. For heme-oxy, the most occurring configuration has O1-Fe as the shortest heme-artemisinin distance with the binding energy of -32.32 kcal mol⁻¹ (Fig. 5e). The O1-Fe and O2-Fe distances of 2.52 Å and 3.32 Å are comparable with those of heme-pdb. Note that heme-oxy and heme-pdb have similar structures.

From the results from the five heme structures, it can be concluded that the structure of the heme molecule has a significant effect on the docking configurations. The steric hindrance at the Fe position plays an important role in the binding. The proximal ligand that increases the steric hindrance at the Fe position would significantly affect the docking results, as in heme-deoxy. If, however, the proximal ligand does not increase the steric hindrance, results similar to those without the proximal ligand, i.e. for heme-oxy and heme-pdb, would be obtained. Therefore, the heme structures which facilitate binding between Fe and endoperoxide oxygens, such as heme-pdb, heme-hemin, and heme-oxy, are recommended for further investigation of the heme-artemisinin system.

All docking calculations similarly reported O1-Fe as the shortest heme-artemisinin distance and O2-Fe as the second shortest. It could then be concluded that iron in heme interacts with O1 more preferably than O2, a preference which might arise from the more negative charge at O1 and the steric hindrance at O2. This observation is in agreement with the proposal of Posner et al. [16] (pathway B). From their docking results, however, Shukla et al. [17] reported O2-Fe as the shortest heme-artemisinin distance. This disagreement is possibly the result of using poor atomic charges, *ab initio* rather than empirical models, and a poor geometry for artemisinin.

Conclusions

The docking results for five heme structures all agreed that the heme iron approaches the endoperoxide moiety at the O1 position in preference to the O2 position. The docking configuration depends on the structures and atomic charges of both artemisinin and heme. The HF/3-21G level of theory is suitable for the geometry optimization of artemisinin and its derivatives. The docking configurations were significantly affected by the atomic charges of heme and to a much lesser extent by the atomic charges of artemisinin. The high quality atomic charges, 6-311G**, are recommended for the electrostatic potential of heme. Heme structures with no or little steric hindrance at the Fe position facilitate binding of heme and endoperoxide oxygens as in heme-pdb, heme-hemin, and heme-oxy, and they are recommended for use in docking calculations. Comparison of docking results for heme-deoxy and heme-oxy, the heme-oxy structure, whose structure is very close to the receptor structure in the bound state, gave docking results that are in agreement with those of other heme structures.

Acknowledgements. The authors would like to thank the Computer Center of the University of Vienna, Austria, for providing computing resources at their workstation clusters and the Austrian-Thai Center (ATC) for Computer-Assisted Chemical Education and Research, Department of Chemistry, Faculty of Science, Chulalongkorn University, Thailand, for computer resources and other facilities. Tonmunphean S. would like to thank Prof. Dr. Peter Wolschann, Institute of Theoretical Chemistry and Molecular Biology, University of Vienna, Austria for helpful discussion and comments.

References

1. World Health Organization. *The World Health Report 1999*.
2. Moore, D. V.; Lanier, J. E. *Am. J. Trop. Med. Hyg.* 1961, 10, 5.
3. Mockenhaupt, F. P. *Parasitol. Today* 1995, 11, 248.
4. Qinghaosu Antimalaria Coordinating Research Group *Chin. Med. J.* 1979, 92, 811.
5. Klayman, D. L. *Science* 1985, 228, 1049.
6. Luo, X. D.; Shen, C.-C. *Med. Res. Rev.* 1987, 7, 29.
7. China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials *J. Trad. Chin. Med.* 1982, 2, 3.
8. Meshnick, S. R.; Thomas, A.; Ranz, A.; Xu, C.-M.; Pan, H.-Z. *Mol. Biochem. Parasitol.* 1991, 49, 181.
9. Meshnick, S. R.; Yang, Y.-Z.; Lima, V.; Kuypers, F.; Kamchonwongpaisan, S.; Yuthavong, Y. *Antimicrob. Agents Chemother.* 1993, 37, 1108.
10. Posner, G. H.; Oh, C. H.; Wang, D.; Gerena, L.; Milhous, W. K.; Meshnick, S. R.; Asawamahsakda, W. *J. Med. Chem.* 1994, 37, 1256.
11. Francis, S. E.; Sullivan, D. J.; Goldberg, D. E. *Annu. Rev. Microbiol.* 1997, 51, 97.
12. Slater, A. F. *Pharmacol. Ther.* 1993, 57, 203.
13. Pandey, A. V.; Tekwani, B. L.; Singh, R. L.; Chauhan, V. S. *J. Biol. Chem.* 1999, 274, 19383.
14. Peters, W.; Li, Z. L.; Robinson, B. L.; Warhurst, D. C. *Ann. Trop. Med. Parasitol.* 1986, 80, 483.
15. Jefford, C. W.; Vicente, M. G. H.; Jacquier, Y.; Favarger, F.; Mareda, J.; Millasson Schmidt, P.; Brunner, G.; Burger, U. *Helv. Chim. Acta* 1996, 79, 1475.
16. Posner, G. H.; Cummings, J. N.; Ploypradith, P.; Oh, C. H. *J. Am. Chem. Soc.* 1995, 117, 5885.
17. Shukla, K. L.; Gund, T. M.; Meshnick, S. R. *J. Mol. Graph.* 1995, 13, 215.
18. Tonmunphean, S.; Parasuk, V.; Kokpol S. *Quant. Struct-Act. Rel.* 2000, *in press*.
19. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., AutoDock version 2.4, The Scripps Research Institute, Department of Molecular Biology, MB-5, La Jolla, California, U.S.A.

20. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* 1986, 7, 230.
21. Halgren, T. A. *J. Am. Chem. Soc.* 1992, 114, 7827.
22. <http://www.scripps.edu/pub/olson-web/doc/autoflex/parameters.html>
23. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J. Comput-Aided Mol. Des.* 1996, 10, 293.
24. Tonmunpuean, S.; Kokpol, S.; Parasuk, V.; Wolschann, P.; Winger, R. H.; Liedl, K. R.; Rode, B. M. *J. Comput-Aided Mol. Des.* 1998, 12, 397.
25. <http://www.rcsb.org/pdb/>
26. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E. III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. (1997), AMBER 5, University of California, San Francisco.
27. <http://www.ccdc.cam.ac.uk/>
28. Gaussian 94, Revision B.3, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
29. Leban, I.; Golic, L. *Acta. Pharm. Jugosl.* 1988, 38, 71.
30. Bernardinelli, G.; Jefford, C. W.; Maric, D.; Thomson, C.; Weber, J. *Int. J. Quant. Chem. Bio. Sym.* 1994, 21, 117.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CURRICULUM VITAE

Somsak Tonmunphean

Born September 4th, 1973 in Bangkok, Thailand

Education

- 1980-1986 Primary School
Bangkok Christian College, Bangkok, Thailand
- 1986-1991 Secondary School
Bangkok Christian College, Bangkok, Thailand
- 1991-1995 Bachelor of Science (First Class Honors) in Chemistry
Department of Chemistry, Faculty of Science,
Chulalongkorn University, Bangkok, Thailand
- 1995-2001 Doctor of Philosophy degree student in Chemistry
at Department of Chemistry, Faculty of Science,
Chulalongkorn University, Bangkok, Thailand



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

PUBLICATIONS

A. Articles

1. Somsak Tonmunphean, Sirirat Kokpol, Vudhichai Parasuk, Peter Wolschann, Rudolf H. Winger, Klaus R. Liedl, and Bernd M. Rode, "Comparative molecular field analysis of artemisinin derivatives: *Ab initio* versus semiempirical optimized structures", Journal of Computer-Aided Molecular Design 12 (1998): 397-409.
2. Somsak Tonmunphean, Stephan Irle, Sirirat Kokpol, Vudhichai Parasuk, and P. Wolschann, "Ab initio and density functional study on singlet and triplet states of artemisinin", Journal of Molecular Structure (Theochem) 454 (1998): 87-90.
3. Somsak Tonmunphean, Vudhichai Parasuk, and Sirirat Kokpol, "QSAR Study of Antimalarial Activities and Artemisinin-Heme Binding Properties Obtained from Docking Calculations", Quantitative Structure-Activity Relationship 19 (2000): 475-483.
4. Somsak Tonmunphean, Vudhichai Parasuk, and Sirirat Kokpol, "Automated docking calculations of artemisinin to heme", Journal of Molecular Modeling *in press*.

B. Proceedings

1. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "QSAR Study of Artemisinin Derivatives: Docking between Artemisinin Derivatives and Heme", The Third Annual National Symposium on Computational Science and Engineering, 24-26 March 1999, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, pp. 196-205.
2. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Automated molecular docking of artemisinin derivatives to heme", The Fourth Annual National Symposium on Computational Science and Engineering, 27-29 March 2000, Faculty of Engineering, Kasetsart University, Bangkok, Thailand, pp. 181-189.

C. International Conferences

1. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Molecular Docking between Artemisinin Derivatives and Heme", International Medicinal Chemistry Symposium, September 12-15, 1999, Beijing, China, OP16, p.91.

2. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Automated Molecular Docking of Artemisinin derivatives to heme", Post Eurasia Symposium, 4 March 2000, P-06, Conference Room, Main Auditorium, Chulalongkorn University, Bangkok, Thailand.
3. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Relationships between the biological activity of antimalarial artemisinins and properties from the automated molecular docking to heme", 13th European Symposium on Quantitative Structure-Activity Relationships: Rational Approaches to Drug Design, 27 August – 1 September 2000, P.120, Heinrich-Heine-University, Duesseldorf, Germany.

D. National Conferences

1. Somsak Tonmunphean, Sirirat Kokpol, Sopa Hwanseng, Ria Armunanto, "QSAR of artemisinin and its derivatives", 22nd Congress on Science and Technology of Thailand, 16-18 October 1996, A-53, pp. 210-211, Bangkok Convention Centre, The Central Plaza Ladprao, Bangkok, Thailand.
2. Somsak Tonmunphean, "Comparative Molecular Field Analysis (CoMFA) of artemisinin derivatives using different optimized structures", The 7th Annual Scientific Meeting of Faculty of Science, Chulalongkorn University, 22-23 March 1999, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, P21.
3. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Molecular docking between artemisinin derivatives and heme", 25th Congress on Science and Technology of Thailand, 20-22 October 1999, A-17, pp. 122-123, Amarin Lagoon Hotel, Pitsanuloke, Thailand.
4. Somsak Tonmunphean, Vudhichai Parasuk, Sirirat Kokpol, "Quantum Chemical Calculations on Mechanism of Action of Antimalarial Artemisinin Derivatives", The 9th Annual Scientific Meeting of Faculty of Science, Chulalongkorn University, 20-21 March 2001, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, C31.

SCHOLARSHIPS AND AWARDS

- 1995 Research Assistantship from the Graduate School and the Research Division, Chulalongkorn University, Thailand (5 years)
- 1996 Professor Dr. Tab Nilanidhi Foundation scholarship, Chulalongkorn University, Thailand (1 year)
- 1996 Rhone Poulenc-Professor Lehn scholarship, France (1 year)
- 1997 ÖAD scholarship from the Austrian Federal Ministry for Foreign Affairs, Austria (1 year)
- 1998 Rhone Poulenc-Professor Lehn scholarship, France (1 year)
- 1999 Best Oral Presentation Award in the 7th Annual Scientific Meeting of Faculty of Science, Chulalongkorn University, 22-23 March 1999, Faculty of Science, Chulalongkorn University, Thailand
- 2000 ASEA-UNINET scholarship, Austria (2 months)

RESEARCH STAYS

- Mar. 1997 – Aug. 1997 Department of Theoretical Chemistry, Institute of General, Inorganic, and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria.
- Sep. 1997 – Feb. 1998 Institute for Theoretical Chemistry and Molecular Structural Biology, University of Vienna, Vienna, Austria.
- September 2000 Department of Theoretical Chemistry, Institute of General, Inorganic, and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria.
- October 2000 Institute for Theoretical Chemistry and Molecular Structural Biology, University of Vienna, Vienna, Austria.