Quantum chemical calculation and structure activity relationship of bioactive terpenoids

A Thesis submitted to the University of North Bengal

For the Award of

Doctor of Philosophy
in
Chemistry

By
Bhaskar Bagchi

GUIDE
Dr. Asim Kumar Bothra

CO-GUIDE
Dr. Pranab Ghosh

Department of Chemistry
University of North Bengal
July, 2016

This work is dedicated to my Parents-

Late Ajit Kumar Bagchi

and

Smt. Bani Bagchi

DECLARATION

I declare that the thesis entitled "Quantum chemical calculation and structure activity relationship of bioactive terpenoids" has been prepared by me under the guidance of Dr. Asim Kumar Bothra, Associate Professor of Department of Chemistry, Raiganj University as Guide and Dr. Pranab Ghosh, Professor of Department of Chemistry, University of North Bengal as co-Guide. No part of the thesis has formed the basis for the award of any degree or fellowship previously.

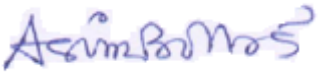
Bhaskar Bagchi

Bhaskar Bagchi
Department of Chemistry,
University of North Bengal,
Raja Rammohunpur
Siliguri, Dt. Darjeeling
PIN. 734013

DATE: 14.07.2016

CERTIFICATE

We certify that Mr. Bhaskar Bagchi has prepared the thesis entitled "Quantum chemical calculation and structure activity relationship of bioactive terpenoids", for the award of PhD degree of the University of North Bengal, under our guidance. He has carried out the work at the Department of Chemistry, Raiganj University.



[Dr. Asim Kumar Bothra]
Guide
Department of Chemistry
Raiganj University
Raiganj,
Dt. Uttar Dinajpur,
PIN.733134

DATE: 14.07.2016

[Dr. PranabGhosh]
Co-Guide
Department of Chemistry
University of North Bengal
Raja Rammohunpur
Siliguri, Dt. Darjeeling
PIN. 734013

DATE: 14.07.2016

ACKNOWLEDGEMENT

I have much pleasure in expressing my deepest sense of gratitude to my guide Dr. Asim Kumar Bothra, Associate professor, Department of Chemistry, Raiganj Universityand my co-guide Dr. Pranab Ghosh, Professor, Department of Chemistry, University of North Bengal for their inspiration, encouragement and guidance to execute the work. I would also like to thank all the faculty members and staff at Raiganj University for supporting me to complete the work. I am thankful to the Head of the Department and all the teachers and staff of the Department chemistry, University of North Bengal. I am also very grateful to Dr. Abul Hassan, Retired Associate professor of Malda College, for their invaluable help and scientific inspirations.

I am grateful to all my friends and Lab mates Uttam Mandal, Shyamal Sharma, Biswajit Das for their support and encouragement all through my PhD tenure. I express my heartiest thanks to my friend Mr. Tamal Goswami, Department of Chemistry, University of North Bengal, for helping me in various ways.

Finally, I express my gratitude and respect to my parents for their endless inspiration, affection and encouragement.

Bhaskar Bagchi

[Bhaskar Bagchi]

Department of Chemistry,

University of North Bengal,

DATE: 14.07.2016

ABSTRACT

The introductory chapter offers a brief description on terpenoids and the historical development of quantitative structure-activity relationship. A large number of terpenoids have a variety of biological activities such as anti-viral, anti-bacterial, anti-malarial, anti-inflammatory, anti-cancer activities and QSAR study may help to provide guidance in design and synthesis of very specific compounds that have high biological activities.

The second chapter presents the review of literature. This chapter is subdivided into two sections: i) terpenoids with different biological activity and ii) quantitative structure-activity relationships of bioactive terpenoids.

The third chapter is allotted for the methodology of research work. Different indices, using in this study, are incorporated. A brief account on regression analysis and statistical parameters are also provided.

The forth chapter is a QSAR study of a number of 23-hydroxybetulinic acid derivatives which were found to be potent glycogen phohphorylase a (GPa) inhibitors. Some derivatives of these triterpenes also exhibit anti-tumor activities against a variety of tumor cell lines. In this study, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa. The other set of equations predicts the antiproliferative activities against HeLa cells. We have also performed Docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa.

The fifth chapter presents the computational study on the redox reaction of puupehenone in aqueous solution by density functional theory. The redox potential of

puupehedienone/puupehenone couple was calculated at the DFT- B3LYP/6-311G(d,P) level of theory in conjugation with Polarizable Continuum Model (PCM). The influence of hydrogen-bond on the redox reaction was also investigated.

The sixth chapter is on the theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum. The molecular geometry of nine halogenated monoterpenoids in the ground state has been calculated by B3LYP/6-31G* and it was found that gap energy and stereochemical features of the compounds play an important role towards activity.

The docking and the QSAR study of oleanolic acid and its derivatives have been discussed in chapter seven. Oleanolic acid and its derivatives were found to be potent Protein-tyrosine phosphatase 1B inhibitors. The docking study shows that most of the ligands can form hydrogen bonds with ARG24 and/or ARG254. Two quantitative structure activity relationships models have been constructed using different descriptors and the significance of these models is judged on the basis of correlation, Fischer F test, and quality factor (Q).

In chapter eight, the molecular geometry of halomon in the ground state has been calculated by the DFT methods. The bond lengths, bond angles and dihedral angles derived from the quantum chemical method are compared with the experimental values and it is found that the calculated geometric parameters are close to the X-ray crystal structure.

A QSAR study of sesquiterpene lactones from Inula falconeri that possess diverse biological activities is discussed in chapter nine. Three statistically significant QSAR models are constructed which may be used to find out the activity of the designed compounds.

Finally chapter ten presents a general and comprehensive conclusion of all the chapters.

PREFACE

Structure-activity relationships (SAR) are unquestionably of great importance in medicinal chemistry which tries to modify the activity of bioactive chemical compounds by modifying their chemical structure. This method was later refined to build mathematical relationships between biological activity and chemical structure, known as quantitative structure-activity relationships (QSAR). The appropriate molecular descriptors are essential to obtain a significant QSAR model. The descriptors are theoretical, empirical, or derived from readily available experimental characteristics of the structures. When a correlation between structure and activity is established, any number of compounds, including those not yet synthesized, can be readily screened on the computer. Quantum chemistry provides an attractive source of molecular descriptors which express all of the electronic and geometric properties of molecules. QSAR currently are being applied in many disciplines like agricultural, biological, environmental, medicinal, and physical organic studies. It is an encouraging method of chemical researching all over the world today.

The dissertation embodies the results of research undertaken at the Department of Chemistry, Raiganj University over the period from November 2012 to June 2016. This work is an attempt to explore the structure activity relationship of various terpenoids by studing their quantum chemical and structural properties.

TABLE OF CONTENTS

Content	Page
Abstract	i-ii
Preface	iii
List of tables	viii - x
List of figures	xi - xiv
List of appendices	xv
Appendix A: List of publications	xvi
Appendix B: Software used in the present study	xvii
Appendix C: Web address used in present study	xviii
Appendix D: Abbreviations	xix
Chapter 1. Introduction	1 – 35
1.1. Medicinal natural products in history	1
1.2. Natural products and drug discovery	2
1.3. Co-evolution	2 – 3
1.4. Plant defense mechanism: production of secondary metabolites	3
1.5. Terpenoids: secondary metabolites	3
1.5.1. Structural features and classification of terpenoids	3 – 4
1.5.2. Biosynthesis of terpenoids	5 – 7
1.5.3. Extraction and isolation	7 – 8
1.5.4. Structural elucidation	8
1.5.5. Bioactivity of terpenoids	8 – 12
1.5.6. Therapeutically used terpenoids	12 – 13
1.5.7. Important receptors of terpenoids	14
1.5.7.1. Glycogen phosphorylase	14
1.5.7.2. Protein tyrosine phosphatase 1B	14 – 15
1.5.7.3. Lipoxygenase	15 – 16
1.6 Quantitative structure-activity relationship	16
1.6.1. Historical Development of QSAR	17 – 19
1.6.2. Development of QSAR Model	19 – 20

1.6.3. Conceptual DFT approach in QSAR analysis 21
1.7. Objective of the thesis 22
1.8. References 23 – 35
Chapter 2. Review of Literature 36 – 83
2.1. Terpenoids with different biological activity 36 – 49
2.2. Quantitative structure-activity relationships of bioactive terpenoids 49 – 66
2.3. References 67 – 83
Chapter 3. Methodology 84 – 110
3.1. Topological indexes 84 – 88
3.1.1. Wiener index 85
3.1.2. Harary index 85
3.1.3. Randić connectivity index 86
3.1.4. Information-theoretic topological indices 86 – 88
3.2. Molar Refraction 89
3.3. Molar Volume 90
3.4. Solvent Accessible Surface Area 90
3.5. Partition coefficient 91
3.6. Quantum Chemical Descriptors 91 – 94
3.7. Molar entropy 94 – 95
3.8. Molecular docking 95 – 97
3.9. Regression Analysis 97 – 105
3.10. The statistical features used for predicting the best model 105 – 106
3.10.1. Co-efficient of Correlation 105
3.10.2. Cross validated coefficient 105 – 106
3.10.3. F test 106
3.10.4. Quality factor 106
3.11. References 106 – 110

Chapter 4. QSAR study and Molecular docking of 23-hydroxybetulinic acid derivatives as RMGPa and HeLa cells inhibitors 111 – 148
4.1. Introduction 111 - 112
4.2. Materials and methods 113 - 119
4.2.1. Dataset and parameters 113 - 118
4.2.2. Statistical methods 119
4.2.3. Molecular docking 119
4.3. Results and discussion 120 - 143
4.4. Conclusion 144
4.5. References 144 - 148
Chapter 5. Computational study on redox reaction of puupehenone in aqueous solution by density functional theory 149 - 161
5.1. Introduction 149 - 150
5.2. Computational details 150 - 156
5.3. Results and discussion 157 - 159
5.4. Conclusion 159
5.5. References 159 - 161
Chapter 6. A theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum 162 - 169
6.1. Introduction 162
6.2. Computational studies 162 - 163
6.3. Results and discussion 164 - 168
6.4. Conclusion 169
6.5. References 169 - 170
Chapter 7. Molecular docking and DFT based QSAR study on oleanolic acid derivatives as Protein-tyrosine phosphatase 1B inhibitors 171 - 193
7.1. Introduction 171 - 172
7.2. Materials and methods 172 - 179
7.2.1. Molecular docking of the oleanolic acid derivatives to PTP1B enzyme 172
7.2.2. Descriptors and data set for QSAR 173 - 178

7.2.3. Statistical methods 178 - 179
7.3. Results and discussion 179 - 190
7.4. Conclusion 190 - 191
7.5. References 191 - 193
Chapter 8. Quantum chemical study of Halomon by the DFT methods 194 - 201
8.1. Introduction 194 - 195
8.2. Computational method 195
8.3. Results and discussion 195 - 199
8.4. Conclusions 200
8.5. Reference 200 - 201
Chapter 9. A QSAR study of sesquiterpene lactones from Inula falconeri as potent anti-inflammatory agents 202 - 210
9.1. Introduction 202
9.2. Materials and methods 202 - 206
9.3. Results and discussion 207 - 210
9.4. Conclusion 211
9.5. Reference 211 - 212
Chapter 10. Conclusion 213 - 215
Bibliography 216 - 262
Index 263 - 264

LIST OF TABLES

Table	Caption	Page
Table 1.1.	Classes of terpenoid with their isoprene units and molecular formula	4
Table 1.2.	Terpenoids approved for therapeutic use	12 -13
Table 3.1.	Molar refraction (R_M) at 589 nm, (cm^3mol^{-1})	89
Table 4.1.	Structures and activity of 47 compounds of 23-hydroxybetulinic acid derivatives	114 - 118
Table 4.2.	Correlation matrix of 15 training RMGPa inhibitors	120
Table 4.3.	SIC_1, CIC_1, quantum chemical descriptors, molar refractivity and molar volume of 15 training RMGPa inhibitors	121
Table 4.4.	List of experimental and predicted logIC_{50} of 15 training RMGPa inhibitors	123
Table 4.5.	Quantum chemical descriptors, SIC_1 and CIC_1 of 5 test RMGPa inhibitors	124
Table 4.6.	List of experimental and predicted logIC_{50} of 5 test RMGPa inhibitors	124
Table 4.7.	Correlation matrix of 20 training antiproliferative compounds against HeLa cells with quantum chemical, molar refractivity and molar volume parameters	126
Table 4.8.	Correlation matrix of 20 training antiproliferative compounds against HeLa cells with different topological indices	126
Table 4.9.	Quantum chemical descriptors, molar refractivity and molar volume of 20 training antiproliferative compounds against HeLa cells	127 - 128

Table 4.10.	Topological indices of 20 training antiproliferative compounds against HeLa cells	128 - 129
Table 4.11.	List of experimental and predicted logIC_{50} of 20 training antiproliferative compounds against HeLa cells	130 - 131
Table 4.12.	Quantum chemical descriptors, SIC_1 and CIC_1 of 8 test antiproliferative compounds against HeLa cells	131 - 132
Table 4.13.	List of experimental and predicted logIC_{50} of 8 test antiproliferative inhibitors against HeLa cells	132
Table 4.14.	Predicted Antiproliferative activity/ RMGPa inhibitory activity of studied compounds	134 - 135
Table 4.15.	The binding energies of 47 docked compounds	135 - 136
Table 5.1.	The Gibbs free energy of P and PH_2 in gas phase and solution, together with solvation free energies of species calculated at 6-31G* and 6-311G** basis set	157
Table 5.2.	Difference in the interaction energies (ΔE_{diff}), difference in solvation free energy (ΔG_{diff}), ΔG^0 (total) and absolute E^0 of puupehedienone (P) and puupehenone (PH_2) complexes with different water molecules	159
Table 6.1.	Minimal inhibitory concentration (MIC) and selected molecular electronic properties of the studied compounds	164
Table 6.2.	Correlation matrix of $MIC_{(SW480)}$ and the electronic descriptors for the studied compounds	168
Table 7.1.	Structural feature of oleanolic acid and its derivatives having PTP1B inhibitory activity	173 - 178
Table 7.2.	Binding energy (EB), Solvent accessible surface area (SASA), Molar refractivity (MR), Molar volume (MV), Partition coefficient (logP), HOMO energy (EH), LUMO energy (EL) and Dipole moment (μ) of 41 PTP1B inhibitors	184 - 185

Table 7.3.	List of experimental and predicted pIC_{50} of 28 training compounds	186 - 187
Table 7.4.	List of experimental and predicted pIC_{50} of 7 test compounds	188
Table 8.1.	DFT calculated energies of the various conformers of halomon and the relative energies with respect to the most stable conformation	196
Table 8.2.	Optimized structural parameters of most stable conformer of halomon with experimental data	197 - 199
Table 9.1.	Structural feature of sesquiterpene lactones from Inula falconeri with anti-inflammatory activity	203 - 205
Table 9.2.	SIC_1, CIC_1, quantum chemical descriptors, entropy at 298 K and electronegativity of 20 inhibitors	207 - 208
Table 9.3.	List of experimental and predicted $logIC_{50}$ of 16 training compounds	209
Table 9.4.	List of experimental and predicted $logIC_{50}$ of 4 test compounds	210

LIST OF FIGURES

Figure	Caption	Page
Figure 1.1.	Illustration of isoprene rule of Myrcene (two isoprene units), p-Menthane (two isoprene units) and Cadinene (three isoprene units)	4
Figure 1.2.	Mevalonate pathway for isopentenyl diphosphate (IPP) biosynthesis	5
Figure 1.3.	Isomerization of isopentenyl diphosphate to dimethylallyl diphosphate	6
Figure 1.4.	General scheme of monoterpenoid biosynthesis	6
Figure 1.5.	General scheme for sesquiterpenoid and triterpenoid biosynthesis	6
Figure 1.6.	General scheme for diterpenoid, sesterterpenoid and polyterpenoid biosynthesis	7
Figure 1.7.	Structures of the some important monoterpenoids	10
Figure 1.8.	Structures of the some important sesquiterpenoids	10
Figure 1.9.	Structure of some important diterpenoids	11
Figure 1.10.	Structure of some important carotenoids	12
Figure 1.11.	General overview of developing QSAR process	20
Figure 3.1.	Hydrogen suppressed graph of isobutene	84
Figure 3.2.	Labeled graph of acetamide and sample calculation of IC_1, SIC_1 and CIC_1	88
Figure 4.1.	A plot between the predicted and the experimental activities for the training set of RMGPa inhibitors	125
Figure 4.2.	A plot between the predicted and the experimental activities for the test set of RMGPa inhibitors	125
Figure 4.3.	A plot between the predicted and the experimental activities for the training antiproliferative inhibitors against HeLa cells	133

Figure 4.4.	A plot between the predicted and the experimental activities for the test antiproliferative inhibitors against HeLa cells	133
Figure 4.5.	Phenogram using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) of 47 compounds of 23-hydroxy betulinic acid derivatives	137
Figure 4.6a.	Docked conformation of ligand 1 along with the important amino acid residues of RMGPa	139
Figure 4.6b.	Docked conformation of ligand 43 along with the important amino acid residues of RMGPa	140
Figure 4.6c.	Docked conformation of ligand 24 along with the important amino acid residues of RMGPa	140
Figure 4.6d.	Docked conformation of ligand 46 along with the important amino acid residues of RMGPa.	141
Figure 4.6e.	Docked conformation of ligand 2 along with the important amino acid residues of RMGPa.	141
Figure 4.6f.	Docked conformation of ligand 15 along with the important amino acid residues of RMGPa.	142
Figure 4.6g.	Docked conformation of ligand 16 along with the important amino acid residues of RMGPa.	142
Figure 4.6h.	Docked conformation of ligand 36 along with the important amino acid residues of RMGPa	143
Figure 4.6i.	Docked conformation of ligand 17 along with the important amino acid residues of RMGPa.	143
Figure 5.1.	The two protons, two electron redox process of puupehenone	150
Figure 5.2.	Thermodynamic cycle for obtaining ΔG^0 (total)	151
Figure 5.3.	Optimized geometries of PH_2 (a) and P (b) at B3LYP/6-311G(d,p) level in gas phase	152

Figure 5.4.	Optimized geometries of PH_2 (c) and P (d) at B3LYP/6-311G(d,p) level in water	153
Figure 5.5.	Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with one water molecule along with ΔE ($kJmol^{-1}$)	154
Figure 5.6.	Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with two water molecules at different configurations along with ΔE ($kJmol^{-1}$)	155
Figure 5.7.	Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with three water molecules at different configurations along with ΔE ($kJmol^{-1}$)	156
Figure 6.1.	Chemical structures of furoplocamioid C (1a), prefuroplocamioid (1b), pirene (1c), cyclohexanes (1d-1i), including mertensene (1g), and violacene (1h).	163
Figure 6.2.	Optimized structures of studied molecules obtained by B3LYP/6-31G* level.	165 - 167
Figure 7.1a.	Docked conformation of ligand 1 along with the important amino acid residues of PTP1B	180
Figure 7.1b.	Docked conformation of ligand 5 along with the important amino acid residues of PTP1B	181
Figure 7.1c.	Docked conformation of ligand 7 along with the important amino acid residues of PTP1B	181
Figure 7.1d.	Docked conformation of ligand 8 along with the important amino acid residues of PTP1B	182
Figure 7.1e.	Docked conformation of ligand 11 along with the important amino acid residues of PTP1B	182
Figure 7.1f.	Docked conformation of ligand 29 along with the important amino acid residues of PTP1B	183
Figure 7.2.	A plot between the predicted and the experimental pIC_{50} for the training set by model 7.1	188

Figure 7.3.	A plot between the predicted and the experimental pIC_{50} for the training set by model 7.2	189
Figure 7.4.	A plot between the predicted and the experimental pIC_{50} for the test set by model 7.1	189
Figure 7.5.	A plot between the predicted and the experimental pIC_{50} for the test set by model 7.2	190
Figure 8.1.	Molecular structure and atom labeling of halomon (hydrogen atoms are not numerated).	194
Figure 8.2.	Conformers of halomon	196-197
Figure 8.3.	IR spectra of the most stable conformer of halomon in gas phase	199
Figure 9.1.	A plot between the predicted and the experimental activities for the 16 training compounds using model 9.3	210

LIST OF APPENDICES

Appendix A: List of publications

Appendix B: Software used in the present study

Appendix C: Web address used in present study

Appendix D: List of Abbreviations

APPENDIX - A

List of Publications

[1] B. Bagchi, S. Sharma, A. Chatterjee, P. Ghosh, A.K. Bothra, QSAR Study and Molecular Docking of 23-hydroxybetulinic Acid Derivatives as RMGPa and HeLa Cells Inhibitors, Commun. Comput. Chem., 3 (2015) 75-102.

[2] B. Bagchi, T. Goswami, P. Ghosh, A.K. Bothra, Computational study on the redox reaction of puupehenone in aqueous solution by density functional theory, Asian J. Chem., 28 (2016) 2199-2203.

[3] B. Bagchi, A. Chatterjee , P. Ghosh, A.K. Bothra, A theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum, JOCPR, 4 (2012) 5076-5080.

[4] B. Bagchi, A.K. Bothra , Quantum chemical study of halomon by the DFT and MP2 Methods, IUP Journal of Chemistry, 4 (2011) 41-47.

[5] A QSAR study of sesquiterpene lactones from Inula falconeri as potent anti-inflammatory agents, B. Bagchi, P. Ghosh, A.K. Bothra, JOCPR, 7 (2015) 907-912.

[6] P. Ghosh, B. Bagchi, S. Sharma, A. Chatterjee, A.K. Bothra , Molecular docking and DFT based QSAR study on oleanolic acid derivatives as Protein–tyrosine phosphatase 1B inhibitors. (Manuscript under communication)

[7] S. Sharma, B. Bagchi, S. Mukhopadhyay, A.K. Bothra, Theoretical study of lysophosphatidic acid acyltransferase 2 inhibitors, JOCPR, 5 (2013) 348-355.

APPENDIX - B

Software used in the present study

Name	Executable	Description
GAMESS	Windows/Linux	AB initio quantum chemistry package.
Firefly	Linux	AB initio and DFT computational chemistry program
MOPAC	Windows/Linux	Semiempirical quantum chemistry program
AutoDock 4.0	Windows/Linux	Suite of automated docking tools
UCSF Chimera	Windows	Molecular visualization program
Molegro Molecular Viewer	Windows	Molecular visualization program
MATLAB	Windows	Numerical computing environment
Gabedit	Windows	Graphical user interface

APPENDIX - C

Web address used in present study

Name	Web address	Description
NCBI	http://www.ncbi.nlm.nih.gov/	For molecular chemistry information
PDB	www.rcsb.org/	Repository of 3D protein structures
Binding DB	www.bindingdb.org/	Focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules
DendroUPGMA	http://genomes.urv.cat/	A dendrogram construction utility
Cambridge Crystallographic Data Centre	www.ccdc.cam.ac.uk	A database of small molecule crystal structures

APPENDIX - D

List of Abbreviations

IPP: Isopentenyl diphosphate

GP: Glycogen phosphorylase

PTP1B: Protein tyrosine phosphatase 1B

QSAR: Quantitative structure-activity relationship

CoMFA: Comparative molecular field analysis

IC: Mean information content

SIC: Structural information content

CIC: Complementary information content

RMGPa: Rabbit muscle glycogen phosphorylase a

CHAPTER 1

Introduction

1.1. Medicinal natural products in history

The knowledge of natural product with medicinal property is as ancient as human civilization and nearly every civilization has accumulated experience of their use. The earliest record of natural products has come from ancient Mesopotamia, circa 2600 BC, and is written on clay tablets in cuneiform. Almost 1000 plants and plant derived substances, such as oils from *Cupressus sempervirens* (Cypress) and *Commiphora* species (myrrh) were documented which are still used today for treating coughs, colds and inflammation. The well known document of Egyptian pharmaceutical record is the Ebers Papyrus (1500 BC) which contains 700 plants based drugs and including formulas such as gargles, pills, and ointments. The Chinese are leaders for using natural products. The most primitive Chinese medical book, Wu Shi Er Bing Fang (1100 BC) contains 52 prescriptions followed by works such as the *Shennong* Herbal (~100 BC; 365 drugs), and the *Tang* Herbal (659 AD; 850 drugs). Documentation of traditional Indian medicine i.e. Ayurveda starts from about 1000 BC (Susruta and Charaka). The Greek scientist and philosopher, Theophrastus (~300 BC) dealt with the medicinal qualities of herbs while the collection, storage, and use of medicinal herbs were recorded by Greek physician Dioscorides (100 AD). Galen (130-200 AD) published at least 30 books on pharmacy and medicine in Rome. From the 5th to 12th centuries, the monasteries in England, Ireland, France, and Germany preserved this knowledge but Arabs were expanding their knowledge by including their own resources with Greco-Roman expertise together with Chinese and Indian herbs unknown to the Greco-Roman world [1].

1.2. Natural products and drug discovery

Various natural products contain different compounds and modern chemistry has provided the tools to purify various compounds and to determine their structures. In 1805, the German pharmacist Friedrich Wilhelm Sertürner isolated morphine from opium. This is the first naturally derived medicine commercialized by Merck in 1826. In fact, Western pharmaceutical companies quickly began to use purified natural products as ingredients for making drugs, rather than crude extracts. In addition, molecular structure determination of many compounds encouraged the chemistry to synthesize them rather than isolating them from natural sources which lowered drug production cost. Hence, a large number of natural compounds were identified, analyzed and synthesized: strychnine and brucine from *Strychnos nux-vomica* (strychnos), nicotine from *Nicotiana tabacum*, atropine from *Atropa belladonna* etc. In the 20th century, the discovery of penicillin from the mould *Penicillium notatum* gave physicians a powerful weapon in their battle against antibacterial diseases. Chemists have now modified the structure of natural compounds to suppress or enhance certain characteristics such as solubility, efficiency or stability in the human body [2]. Newman pointed out that approximately 60% of the drugs that are now available-including household names such as artemisinin, camptothecin, maytansine, paclitaxel, penicillin, reserpine and silibinin were either directly or indirectly derived from natural products [3].

1.3. Co-evolution

The important question is that why so many natural products are useful to human health and one explanation is that it is the result of long-term co-evolution within biological communities: simultaneous evolution of multiple interactive species. Hence, those natural compounds are able

to suppress the growth of bacteria; such molecules can also exert same effects in humans and are now used as antimicrobial drugs in medicine [4].

1.4. Plant defense mechanism: production of secondary metabolites

The production of secondary plant metabolites may be the most important strategy for plant defense against animal and microbial predators. Though the function of many plant secondary metabolites is not known, we can assume these chemicals may have general or specific activity against key target sites in bacteria, fungi, viruses, or neoplastic diseases [5].

1.5. Terpenoids: secondary metabolites

Terpenoids are the largest and most widespread class of secondary metabolites which occur mainly in plants and lower invertebrates. Traditionally, humans used plant-based terpenoids in the food, pharmaceutical, and chemical industries. A large number of terpenoids show a variety of biological activities such as anti-viral, anti-bacterial, anti-malarial, anti-inflammatory, inhibition of cholesterol synthesis and anti-cancer activities [6, 7].

1.5.1. Structural features and classification of terpenoids

The term terpene is generally used to represent only the hydrocarbons $(C_5H_8)_n$ while the term terpenoids represent the hydrocarbons as well as their oxygenated derivatives. The important structural feature of nearly all the terpenoids is their biosynthesis from one monomeric structural unit, isoprene (C_5H_8). Ingold suggested that the isoprene units in terpenoids were joined in head to tail linkages, may be referred to as the special isoprene rule (Figure 1.1). But this rule is not unique as many exceptions occur, e.g., the two halves of carotenoids are joined tail to tail fashion; cryptone, a natural terpenoid, have nine carbon atoms instead of ten and hence isoprene

rule cannot be applied [8, 9]. The C_5H_8 units polymerize and subsequently fix the number and position of double bond. The basic C_5 unit can be combined in many ways; it is not surprising to observe the amount and diversity of the structures [10].

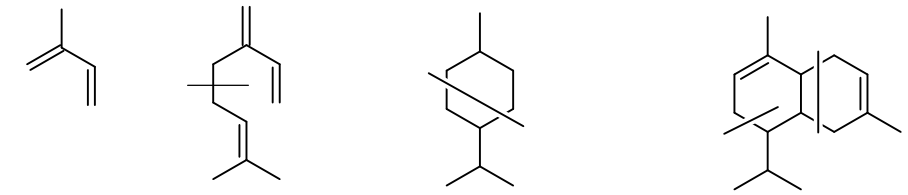


Figure 1.1. Illustration of isoprene rule of Myrcene (two isoprene units), p-Menthane (two isoprene units) and Cadinene (three isoprene units)

The classes of terpenoids found in nature are listed in Table 1.1.

Table 1.1. Classes of terpenoid with their isoprene units and molecular formula

Class	Number of Isoprene units (C_5H_8)	Molecular formula
Hemiterpenoids	1	C_5H_8
Monoterpenoids	2	$C_{10}H_{16}$
Sesquiterpenoids	3	$C_{15}H_{24}$
Diterpenoids	4	$C_{20}H_{32}$
Sesterterpenoids	5	$C_{25}H_{40}$
Triterpenoids	6	$C_{30}H_{48}$
Tetraterpenoids or Carotenoids	8	$C_{40}H_{64}$
Polyterpenoids	n	$(C_5H_8)_n$

Each class of terpenoid may be further subdivided on the basis of the number of rings present in the molecule i.e. monocyclic, bicyclic, tricyclic etc.

1.5.2. Biosynthesis of terpenoids

Plants have tremendous biosynthetic potentials and the fundamental units used in syntheses in the cell are water, carbon dioxide, formic acid (as active formate), and acetic acid (as active acetate). These active compounds form acyl derivatives of coenzyme A. This coenzyme is a complex thiol derivative and is represented as CoA-SH. The acetylcoenzyme A, usually written as CH_3CO-SCoA, may be formed by the condensation between acetic acid and coenzyme A. The conversion of acetyl-coenzyme A (5) into isopentenyl diphosphate (9, IPP) is believed to proceed via acetoacetyl-coenzyme A (6), 3-hydroxy-3-methylglutaryl-coenzyme A (7), and mevalonate (8) shown in Figure 1.2.

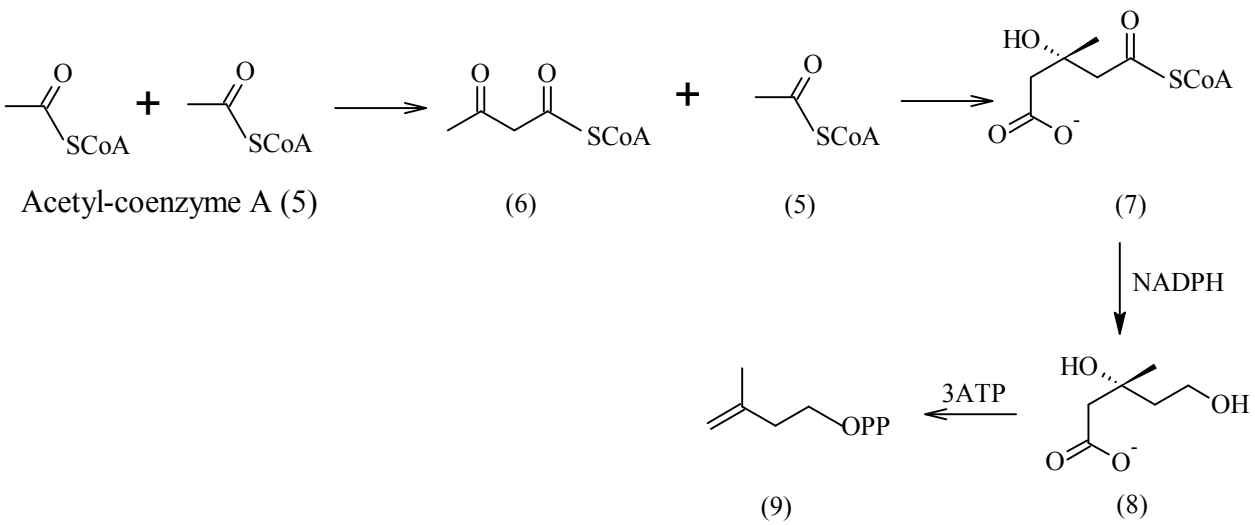


Figure 1.2. Mevalonate pathway for isopentenyl diphosphate (IPP) biosynthesis

This isopentenyl diphosphate is transformed into all different terpenoids found in nature. In the presence of appropriate enzyme, isopentenyl diphosphate (9, IPP) is isomerized to dimethylallyl diphosphate (10). Either of these can be converted into hemiterpenoids (Figure 1.3).

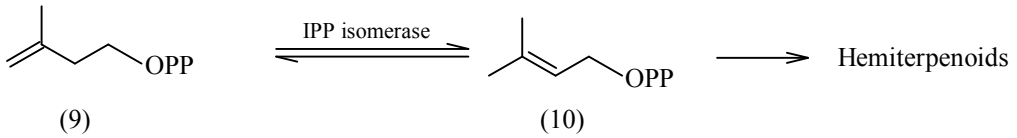


Figure 1.3. Isomerization of isopentenyl diphosphate to dimethylallyl diphosphate

The condensation of dimethylallyl diphosphate (10) with isopentenyl diphosphate (9) under the action of the enzyme prenyltransferase gives geranyl diphosphate (11) which can serve as the precursor for monoterpenoid biosynthesis (Figure 1.4).

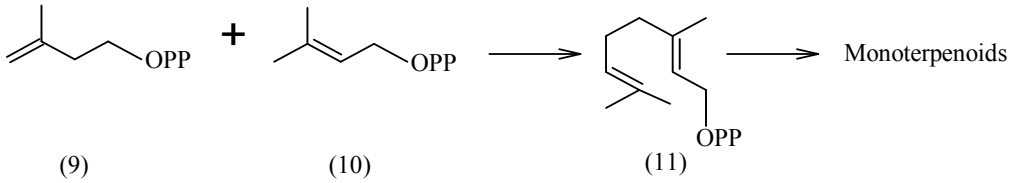


Figure 1.4. General scheme of monoterpenoid biosynthesis

The condensation of geranyl diphosphate (11) with another IPP molecule produces farnesyl diphosphate (12) which can be either channeled into sesquiterpenoid biosynthesis or undergo a tail-to-tail dimerization to form triterpenoids (13) (Figure 1.5).

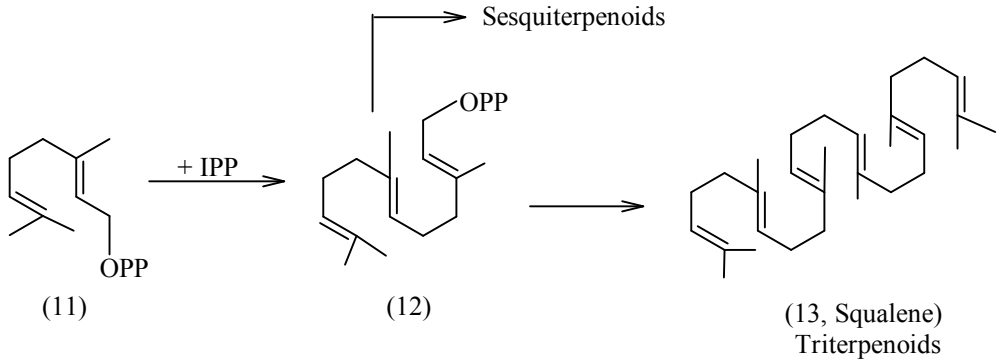


Figure 1.5. General scheme for sesquiterpenoid and triterpenoid biosynthesis

Farnesyl diphosphate (12) can undergo chain elongation with IPP to form geranylgeranyl diphosphate (14) which has three metabolic fates, to form diterpenoids, to undergo chain extension to form geranylfarnesyl diphosphate (15) and tail-to-tail dimerization to tetraterpenoids

(16). Geranylfarnesyl diphosphate (15) is transformed into sesterterpenoids or undergoes chain extension to produce polyterpenoids (Figure 1.6) [6, 11].

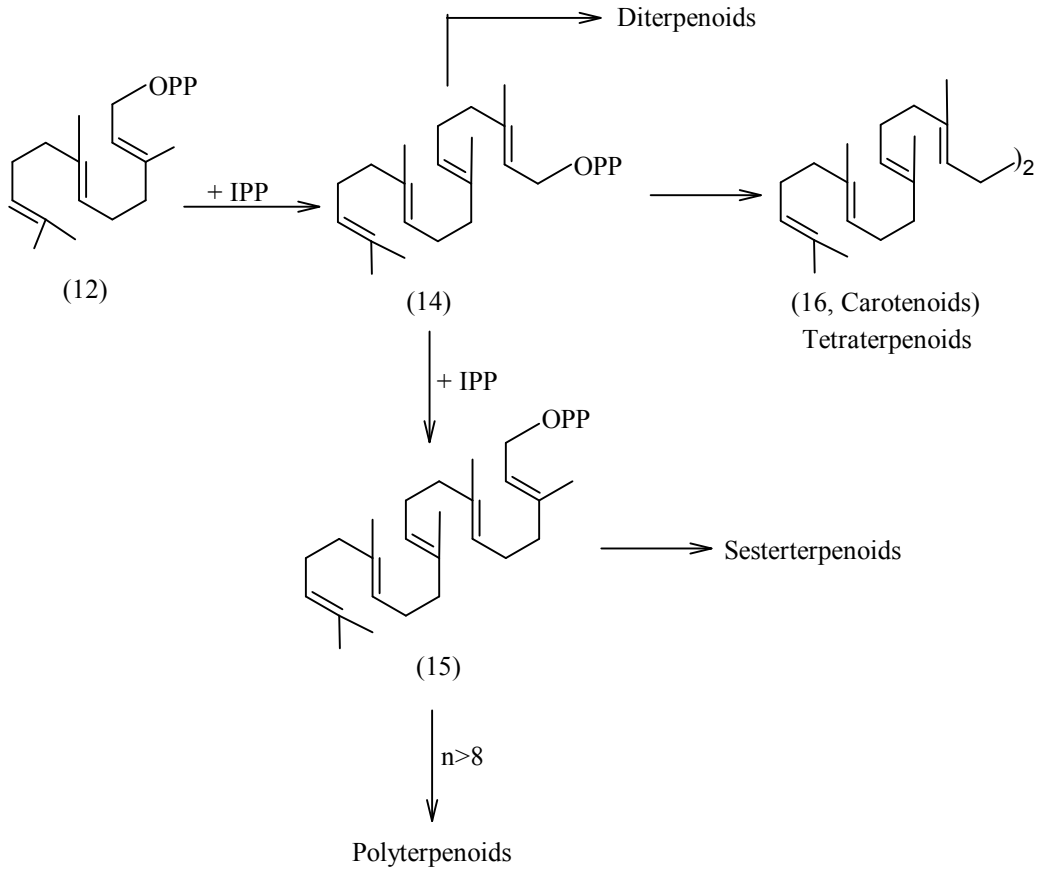


Figure 1.6. General scheme for diterpenoid, sesterterpenoid and polyterpenoid biosynthesis

1.5.3. Extraction and isolation

As the terpenoids are of wide occurrence, they can be isolated by a great variety of methods. The essential oils, which are mixture of lower terpenes (mono and sesquiterpenes), are extracted by subjecting plants to steam distillation. The individual terpenoids from the essential oils are then separated by successive fractional distillation. During fractional distillation, the terpenoid

hydrocarbons distill first followed by their oxygenated derivatives. The sesquiterpenoids are obtained under the distillation of the residue at reduced pressure. More recently, various forms of chromatography, such as column chromatography, preparative HPLC, radial chromatography, have been employed for the isolation and separation of terpenoids [12].

1.5.4. Structural elucidation

The structures of terpenoids are examined by a combination of chemical and spectroscopic methods. However, the recent trend is to determine the structures by the spectroscopic techniques which require small amount of the sample. Nuclear magnetic resonance spectroscopy (^{1}H, ^{13}C and 2D NMR) has been used to elucidate the structure of a new terpenoid including the relative stereochemistry of chiral centres. NMR has also been used to recognize substance comparing with a previously known compound, considering the huge literature data on these metabolites that have been gathered during the past decades. Moreover, the absolute configuration of a terpenoid can be further established by several methods like X-ray analysis, optical rotator dispersion, circular dichroism, exciton chirality, chemical correlation with a compound of known absolute configuration, resolution of racemates, as well as enzymatic reactions [13-15].

1.5.5. Bioactivity of terpenoids

A number of monoterpenoids possess many biological activities. Citral is an acyclic monoterpenoid hydrocarbon which occurs in lemon grass oil present at levels of 65-85%. Citral (3,7-dimethyl-2,6-octadienal) exists in two geometrical isomeric forms: trans citral (citral-a, geranial, 17) and cis citral (citral-b, neral, 18). Citral has good antimicrobial, antibacterial and antifungal activity [16, 17]. Citral also possesses anticancer activity against HeLa, ME-180, NB4

and MCF-7 cell lines in vitro [18-20]. Apoptosis or programmed cell death is suppressed in many diseases including cancer. It may be mentioned that citral induced apoptosis of NB4 cell takes place by decreasing mitochondrial membrane potential. Bcl-2 down-regulation, Bax up-regulation on mRNA level and NF-κB down-regulation on protein level may be involved in the mechanism of the apoptotic effect of citral on NB4 cells [20]. A very recent report on the anti-tumor activity of citral-isomers reveals that geranial is more potent in inhibiting cytotoxicity/tumor-growth in mammalian physiology [21]. The roles of limonene (19), perillyl alcohol (20), carvone (21), and carveol (22) have been investigated in human health due to their chemotherapeutic activities. They inhibit carcinogenesis both in the initiation and promotion/progression stages, and are effective in early and advanced cancers treatment [22-25]. Several monoterpenoids have been studied for their antioxidant activity and γ-terpinene (23) was shown to be an important antioxidant [26]. Monoterpene hydrocarbons such as menthol (24), limonene (19), linalool (25) etc. are used as flavouring agent in foods and beverages [27]. Figure 1.7 shows the structures of the some important monoterpenoids.

Sesquiterpenoids function as the pheromones and juvenile hormons in plants. Farnesol (26), a acyclic sesquiterpene, has anti tumor activity [28]. Because of its anti bacterial activity, it is also used as a deodorant [29]. The bicyclic sesquiterpene β-caryophyllene (27) was shown to be selective agonist of cannabinoid receptor type 2 which is the alternative therapeutic targets for the treatment of anxiety and depression [30]. Abscisic acid (28) is a plant hormone that accelerates the development process in many plants including seed development and dormancy [31]. Some important sesquiterpenoids are shown in Figure 1.8.

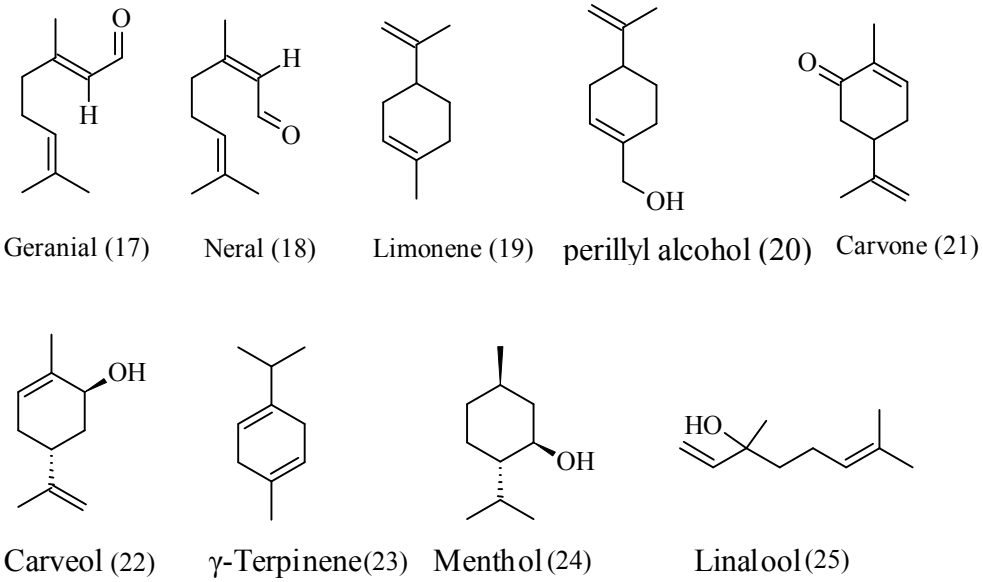


Figure 1.7. Structures of the some important monoterpenoids

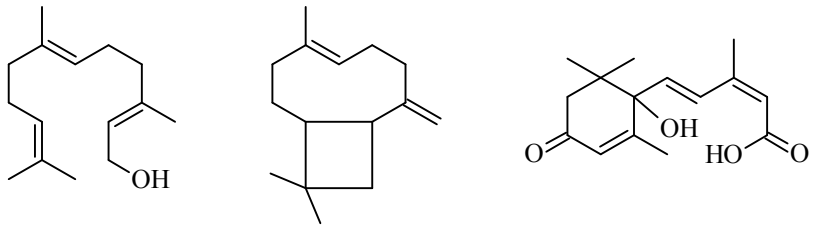
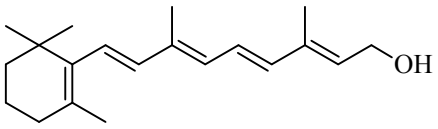


Figure 1.8. Structures of the some important sesquiterpenoids

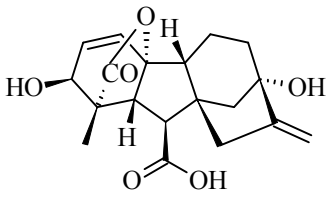
The important diterpenoids associated with vision are those with vitamin A activity. Besides this, retinoids (e.g.retinol, 29) influence the growth and activate the tumor suppressor genes [25, 32]. Gibberellic acid (30), a plant hormone controlling plant growth and development is a tetracyclic diterpene acid [33]. Carnosic acid (31) and carnosol (32) are found in the herb rosemary (Rosmarinus officinalis). Carnosol (32) is an ortho-diphenolic diterpenoid possessing anti oxidant, anti inflammatory and anti cancer activity [34]. Carnosic acid (31) shows

antimicrobial and antioxidative properties and it is also used as preservatives [35]. Structures of some important diterpenoids are given in Figure 1.9.

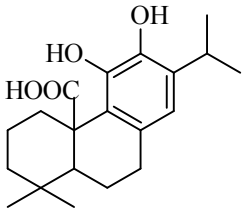
The main group of tetraterpenoids is the carotenoids which are found in plants as well as in humans. Photoprotection, light harvesting in photosynthesis and pigmentation are the main functions of carotenoids in plants [36]. For man, their main function is associated with vision. Carotenoids are good anti-oxidizing agents. They are involved in scavenging of singlet oxygen and peroxide radical and thereby protect human from various disease like cancer [37]. Some important carotenoids are lycopene (33), β-carotene (34), lutein (35) etc are shown in Figure 1.10.



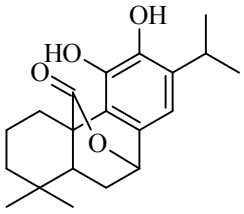
Retinol (29)



Gibberellic acid (30)

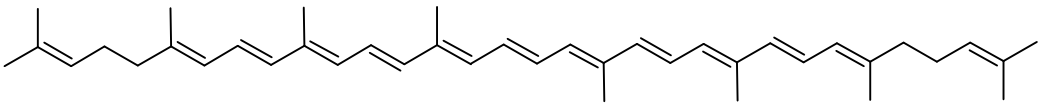


Carnosic acid (31)

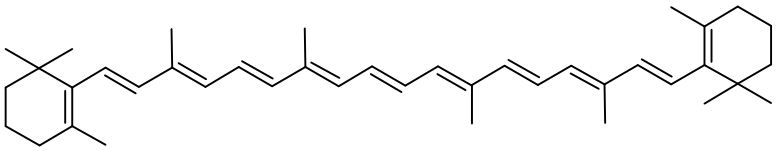


Carnosol (32)

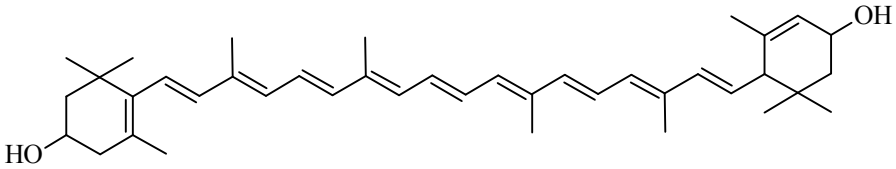
Figure 1.9. Structure of some important diterpenoids



Lycopene (33)



β-Carotene (34)



Lutein (35)

Figure 1.10. Structure of some important carotenoids

1.5.6. Therapeutically used terpenoids

Terpenoids are used for the treatment of various diseases [38]. Table 1.2 depicted a few of them.

Table 1.2. Terpenoids approved for therapeutic use

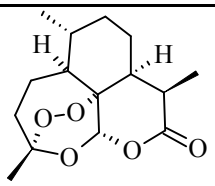
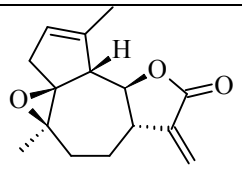
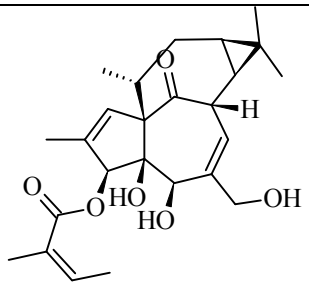
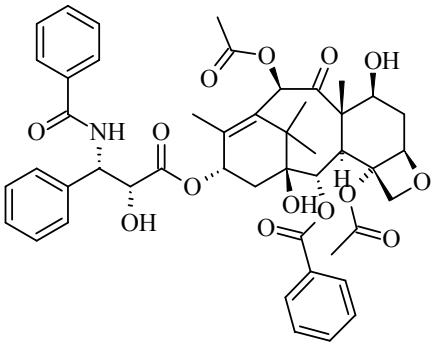
Generic name (trade name)	Chemical structure	Plant species	Type of terpenoids	Indication and mechanism of action
Artemisinin (Artemisin)		*Artemisia annua L.*	Sesquiterpene lactone endoperoxide	Malaria treatment (radical formation)

Table 1.2 (continued)

Generic name (trade name)	Chemical structure	Plant species	Type of terpenoids	Indication and mechanism of action
Arglabin (Arglabin)		*Artemisia obtusiloba var. glabra Ledeb.*	Sesquiterpene lactone	Cancer chemotherapy (inhibition of farnesyl transferase)
Ingenol mebutate (Picato)		*E. peplus L.*	Diterpenoid	Actinic keratosis (inducer of cell death)
Paclitaxel (Nanoxel)		*T. brevifolia Nutt.*	Diterpenoid core	Cancer chemotherapy (antimitotic agent)

1.5.7. Important receptors of terpenoids

1.5.7.1. Glycogen phosphorylase

Glycogen is the principal reserve of carbohydrate in animals and its breakdown into glucose-1-phosphate is mediated by glycogen phosphorylase (GP) with the help of a debranching enzyme [39]. The brain, liver, and muscle isoforms of GP share approximately 80% sequence homology and each has an important role in glycogen metabolism. The brain isoform supplies glucose during the period of insufficient glucose level in brain. The muscle isoform gives energy for muscle contraction while the liver isoform provides glucose to the rest of the body and hence it is an attractive target for treatment of type 2 diabetes [40-43].

GP exists in two forms: GPa and GPb and they are interconvertible. GPa has high activity and remains predominantly in the active R state (Relaxed state) while GPb has low activity and resides mainly in the T state (Tense state). During physical activity or hormone simulation, the enzyme is converted from T state b form (GPb) to the R state a form (GPa) through the phosphorylation of Ser14 which is catalyzed by phosphorylase kinase [44, 45]. The regulatory sites of GP are: glucose analogues at the catalytic site, azasugar inhibitors, lactones at the allosteric site (AMP), caffeine at the purine inhibitor site, indole-2-carboxamide at the indole binding site and cyclodextrins at the glycogen storage site [46, 47]. The X-ray analysis suggests that pentacyclic triterpenes bind at the allosteric site [48].

1.5.7.2. Protein tyrosine phosphatase 1B

Reversible Protein tyrosine phosphorylation is important for nearly all cellular processes such as growth, differentiation, migration, survival, and apoptosis. Protein tyrosine kinases (PTKs) are enzymes that catalyze protein tyrosine phosphorylation where as protein-tyrosine phosphatases

(PTPs) catalyze the dephosphorylation process [49]. In men, more than one hundred PTPs are present and they function as negative or positive modulators of biological pathways [50]. Among the member of PTPs, protein tyrosine phosphatase 1B (PTP1B) negatively regulates insulin signaling pathway and is an excellent drug targets for type 2 diabetes and insulin resistance [51, 52].

A number of natural compounds show PTP1B inhibitory activity. Thareja et al. reviewed approximately 50 natural PTP1B inhibitors [53]. The first reported sesterterpenoid which shows PTP1B inhibitory activity is sulfircin 236 isolated as its N,N-dimethylguanidinium salt from a deep-water sponge Ircinia (unknown species) collected from Andros and Bahamas [54]. In 2002, five natural flavonoids were reported as PTP1B inhibitors [55]. Since then, approximately 300 known or new natural compounds have been identified which possess PTP1B inhibitory activity.

PTP1B consists of 435 amino acids with active site (site A) of approximately 8 to 9 Å depth and is defined by residues His-Cys-Ser-Ala-Gly-Ile-Gly-Arg (214 to 221). The important amino acid residues of second binding site (site B) are Arg24, Arg254, and Glu262. Other amino acid residues in site B are Tyr46, Asp48, Val49, Ile219, and Met258. Molecular docking studies indicate that triterpenes bind in the site B, not in the active site [56, 57].

1.5.7.3. Lipoxygenase

Lipoxygenases are a class of non-heme iron containing oxidative enzymes, occurring in a number of plants and animals [58, 59]. These enzymes catalyze the incorporation of molecular oxygen in the naturally occurring poly-unsaturated fatty acids (PUFAs) such as arachidonic acid and linoleic acid [60]. More importantly, lipoxygenases are involved in the regulation of inflammatory responses that can promote human disease. For example, human 5-lipoxygenase

(5-HLO), human 12-lipoxygenase (12-HLO) and human 15- lipoxygenase (15-HLO) are involved in several diseases like asthma, arthritis, allergy, psoriasis, atherosclerosis and tumorigenesis [61-66].

Several studies have suggested that diets containing high fat, particularly omega-6 polyunsaturated fatty acids, are linked with pancreatic cancer development and growth [67, 68]. The pathways associated in the conversion of unsaturated fatty acid arachidonic acid and linoleic acid to their lipid metabolites such as prostaglandins and leukotrienes which are involved in development and growth of multiple human cancers [69, 70]. Hence, the blockade of lipoxygenase pathway inhibits pancreatic cancer cell proliferation and induces apoptosis through the mitochondria-mediated pathway, with cytochrome c release and caspase activation [71].

The lipoxygenase inhibitors are classified into two groups, redox and nonredox inhibition. The redox active compound reduces lipoxygenase from ferric oxidation state to its inactive ferrous form where as allosteric inhibition can occur in nonredox mechanism [72, 73].

1.6. Quantitative structure-activity relationship

A number of natural products were isolated and attempts were made to correlate their chemical structures with their physiological activities. Though the physiological activity of a compound is associated with a particular structural unit or group, there are no hard and fast rules connecting chemical structure and physiological activity. Quantitative structure-activity relationship (QSAR) makes it feasible to predict the activities of a given compound as a function of its molecular substituent. New compounds with similar structural features can also be incorporated in the development of QSAR models assuming that they possess similar activities. Hence QSAR methods have become major tools in modelling and designing novel compounds.

1.6.1. Historical development of QSAR

More than a century ago, in 1863, Cros showed a relationship between the toxicity of primary aliphatic alcohols and their water solubility [74]. In 1868, Crum-Brown and Fraser formulated a suggestion that the physiological action of a substance is a function of its chemical composition and constitution [75]. Shortly after, Richet (1893), Meyer (1899) and Overton (1901) separately concluded that narcotic (depressant) activity is dependent on the lipophilicity of the molecules [76-78]. In 1935 Hammett proposed a relationship which deals with the electron donating and withdrawing substituents m- and p- to the reaction site in monosubstituted benzene [79, 80]. The mathematical expression of Hammett equation is:

$$\log (K/K_0) = \sigma\rho \quad \ldots\ldots\ldots \quad (1.1)$$

Where K and K_0 are the equilibrium constant of substituted and unsubstituted phenyl compounds of a given reaction. The σ is a substituent constant and ρ is a reaction constant. Taft (1952, 1953) has examined aliphatic reactions and incorporates a steric substituent parameter, Es [81]. The contributions from Hammett and Taft set forth the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita (1964). The mathematical combination of Hammett σ constants and log P values (P is the octanol-water partition coefficient of the unionized molecule) gives the linear Hansch equation and its many extended forms [82, 83].

$$\log (1/C) = k_1\pi + k_2\sigma + k_3 \quad \ldots\ldots\ldots \quad (1.2)$$

Where C is the molar concentration, the constants k_1, k_2, k_3 are obtained via least square methods and $\pi = \log Px - \log P_H$. P_H and Px are the partition coefficient of the parent molecule and its derivative. This equation is better than Hammett equation. However in case of highly hydrophilic molecules, Hansch postulated parabolic equation:

$$\log (1/C) = -k_1(\log P)^2 + k_2(\log P) + k_3\sigma + k_4 \ldots\ldots\ldots (1.3)$$

Free and Wilson constructed a model of substituent contribution to the overall biological activity [84]. It is given by the following equation:

$$BA = \Sigma a_i x_i + u \ldots\ldots\ldots (1.4)$$

where BA is the biological activity, u is the average contribution of the parent molecule, and a_i is the contribution of each structural feature; x_i denotes the presence $x_i = 1$ or absence $x_i = 0$ for a particular structural fragment. In 1971, Fujita and Ban simplified the Free and Wilson equation by using logarithm of activity which brought the activity parameter in line with other free energy-related terms [85]. In the 1970s, quantum chemical indices appear in QSAR modeling [86]. Topological descriptors are actually structural descriptors as they are only based on the two-dimensional chemical structure. These descriptors, which were originally proposed by Balaban [87], Randić [88], Kier and Hall [89], provided significant developments of the QSAR approaches in the mid-1980s. However simple descriptors like counts of atoms/fragments or topological descriptors have only a limited utility due to the lack of consideration of the 3D structure of the molecules, leading to the development of the 3D-QSAR. A large variety of geometrical descriptors like shadow indexes [90], charged partial surface area descriptors [91], WHIM descriptors [92], gravitational indexes [93], EVA descriptors [94], 3D-MoRSE descriptors [95], GETAWAY descriptors [96] etc which were derived from the 3D geometrical structure of a molecule are now used in QSAR modeling. In 1988, Cramer et al. proposed a powerful 3D-QSAR methodology, Comparative Molecular Field Analysis (CoMFA) [97]. In 3D-QSAR method the biological activity of a ligand can be predicted from the molecular interaction field properties in 3D space, such as steric demand, lipophilicity, and electrostatic interactions.

Finally, the scientific community has shown interest in virtual screening and moleculer design, for which several similarity/diversity approaches, cell-based methods, and scoring functions have been developed on the basis of the substructural descriptors such as molecular fingerprints [98, 99].

1.6.2. Development of QSAR model

The construction of QSAR model consists of two main steps: (i) selection of molecules (data set) followed by determination of molecular descriptors. Low quality descriptors are excluded as they will lower the quality of the model. Hence, it is better to use the most informative descriptors in the QSAR modeling. However in many cases too much information in the independent variables (the descriptors) with respect to the response is often seen as noise in the model, thus giving instable or unpredictive models. (ii) Modeling methods like multiple linear regression, logistic regression and machine learning methods etc are used for correlating molecular descriptors with observed activities. During the development of the QSAR model the data set are divided into two groups: training set and test set. The descriptors of the training set are used to build QSAR models that describe the empirical relationship between the structure and activity. The finalized QSAR models are verified by the statistical evaluation and with a testing set. Figure 1.11 illustrates general overview of developing QSAR process.

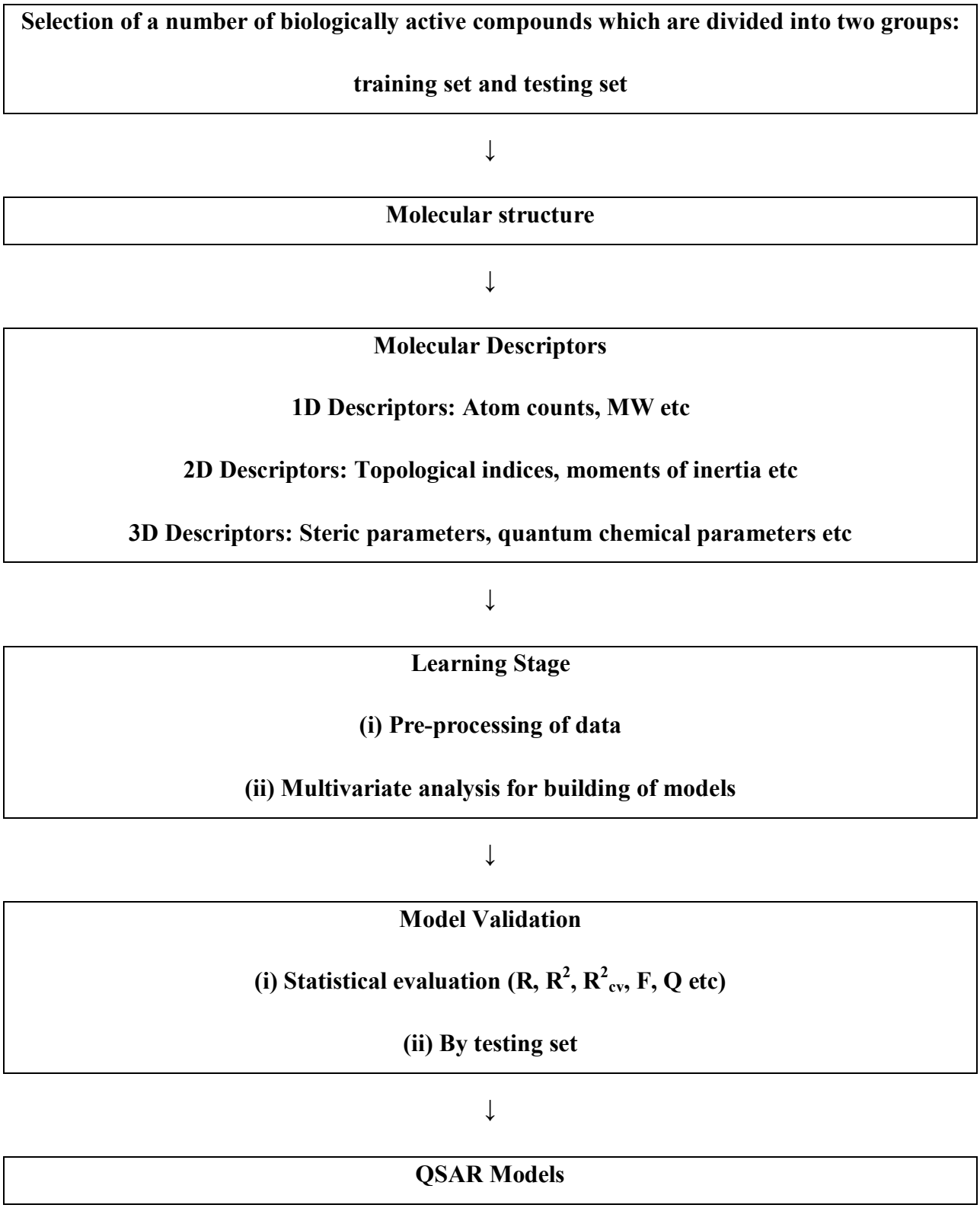


Figure 1.11. General overview of developing QSAR process

1.6.3. Conceptual DFT approach in QSAR analysis

In recent years, QSAR studies have been carried out using conceptual density functional theory (DFT) based descriptors. These descriptors are able to provide accurate quantitative description of the molecular structures and chemical properties. The important DFT descriptors which have found immense usefulness in QSAR studies are molecular orbital energies, chemical hardness, chemical softness, dipole moments etc.

The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies are important quantum chemical descriptors. These orbitals play a major role in many chemical reactions by determining electronic band gaps in solids and are also responsible for the formation of many charge transfer complexes [100, 101]. According to the frontier molecular orbital theory, the formation of a transition state is due to an interaction between HOMO and LUMO of the reacting species [102]. The HOMO and LUMO energies are also important in radical reactions [103, 104]. Ionization energy measures the energy of removing an electron from the HOMO and characterizes the susceptibility of the molecule toward attack by electrophiles while electron affinity represents the energy liberated when an electron is brought from infinity to the LUMO and characterizes the susceptibility of the molecule toward attack by nucleophiles. The hard and soft acid base concept of nucleophiles and electrophiles are directly related to the relative energy of the HOMO/LUMO orbitals [105]. The HOMO/LUMO gap energy is an important stability index. A large HOMO/LUMO separation implies high stability for the molecule in chemical reactions. The HOMO/LUMO gap energy also estimates the lowest excitation energy of the molecule [106]. The often used polarity quantity of the molecule is the dipole moment which reflects the global polarity of a molecule [107].

1.7. Objectives of the thesis

i) Construction of QSAR equations between molecular properties and the biological activities of a number of terpenoids and the quality of the model can be assessed by several statistical parameters.

ii) Docking study was performed to explore the binding mode of different bioactive terpenoids.

iii) To study the redox potential of bioactive terpenoids.

iv) To study the structure of bioactive terpenoids by quantum chemical methods and compared it with the experimental value.

Every chapter in this thesis is complete by itself; that is, it contains its own introduction, complete list of references, figures, tables, and interim conclusions etc.

1.8. References

[1] G.M. Cragg, D.J. Newman, Biodiversity: A continuing source of novel drug leads, Pure Appl Chem. 77 (2005) 7-24.

[2] H.F. Ji, X.J. Li, H.Y. Zhang, Natural products and drug discovery, EMBO rep. 10 (2009) 194-200.

[3] D.J. Newman, Natural products as leads to potential drugs: an old process or the new hope for drug discovery?, J Med Chem. 51 (2008) 2589-2599.

[4] W.J. Zhang, G.H. Liu, Coevolution: A synergy in biology and ecology, Selforganizology. 2 (2015) 35-38.

[5] S.J. Cutler, H.G. Cutler, Biologically Active Natural Products: Pharmaceuticals, CRC Press, New York, 2000.

[6] B. delasHeras, B. Rodríguez, L. Boscá, A.M. Villar, Terpenoids: sources, structure elucidation and therapeutic potential in inflammation, Curr Top Med Chem. 3 (2003) 171-185

[7] D. Tholl, Biosynthesis and biological functions of terpenoids in plants, Adv Biochem Eng Biotechnol. 148 (2015) 63-106.

[8] Z.L. Ruzicka, The isoprene rule and the biogenesis of terpenic compounds, Experientia. 9 (1953**)** 357-367.

[9] J. Gershenzon, W. Kreis, Biosynthesis of monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins, Biochemistry of plant secondary metabolites,

Annual plant reviews, in: M. Wink (Ed), Sheffield Academic Press: Sheffield, UK, Vol. 2, 1999.

[10] J. Gershenzon, N. Dudareva, The function of terpene natural products in the natural world, Nat Chem Biol. 3 (2007) 408-414.

[11] M. Rohmer, The Discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, Nat Prod Rep. 16 (1999) 565-574.

[12] I.L Finar, Stereochemistry and the chemistry of natural products, Pearson Education, Singapore, Vol. 2, 2004.

[13] P. Crabbé, Ord and Cd in chemistry and biochemistry, Academic Press, London, 1972.

[14] N. Harada, K. Nakanishi, Circular dichroic spectroscopy, Exciton coupling in organic stereochemistry, University Science Books, Mill Valley, CA, 1983.

[15] E.L. Eliel, S.H. Wilen, Stereochemistry of organic compounds, Wiley-Interscience, New York, 1994.

[16] A.A. Saddiq, S.A. Khayyat, Chemical and antimicrobial studies of monoterpene: Citral, Pestic Biochem Physiol. 98 (2010) 89-93.

[17] C. de Bona da Silva, S.S. Guterres, V. Weisheimer, E.E.S. Schapoval, Antifungal activity of the lemongrass Oil and citral against Candida spp., Braz J Infect Dis. 12 (2008) 63-66.

[18] K. Ghosh, Anticancer effect of lemongrass oil and citral on cervical cancer cell lines, Phcog Commn. 3 (2013) 41-48.

[19] W. Chaouki, D.Y. Leger, B. Liagre, J.L. Beneytout, M. Hmamouchi, Citral inhibits cell proliferation and induces apoptosis and cell cycle arrest in MCF-7 cells, Fundam Clin Pharmacol. 23 (2009) 549-556.

[20] H. Xia, W. Liang, Q. Song, X. Chen, X. Chen, J. Hong, The in vitro study of apoptosis in NB4 cell induced by citral, Cytotechnology 65 (2013) 49-57.

[21] S. Zeng, A. Kapur, M.S. Patankar, M.P. Xiong, Formulation, characterization, and antitumor properties of trans- and cis-citral in the 4T1 breast cancer xenograft mouse model, Pharm Res. 32 (2015) 2548-2558.

[22] P.L. Crowell, Monoterpenes in breast cancer chemoprevention, Breast Cancer Res Treat. 46 (1997) 191-197.

[23] P.L. Crowell, Prevention and therapy of cancer by dietary monoterpenes, J Nutr. 129 (1999) 775S-778S.

[24] M.N. Gould, Cancer chemoprevention and therapy by monoterpenes, Environ Health Perspect. 105 (1997) 977-979.

[25] K.H. Wagner, I. Elmadfa, Biological relevance of terpenoids-overview focusing on mono, di and tetraterpenes, Ann Nutr Metab. 47 (2003) 95-106.

[26] H.S. Choi, H.S. Song, H. Ukeda, M. Sawamura, Radical scavenging activities of citrus essential oils and their components: detection using 1,1-diphenyl-2-picrylhydrazyl, J Agric Food Chem. 48 (2000) 4156-4161.

[27] L. Caputi, E. Aprea, Use of terpenoids as natural flavouring compounds in food industry, Recent Pat Food Nutr Agric. 3 (2011) 9-16.

[28] J.H. Joo, A.M. Jetten, Molecular mechanisms involved in farnesol-induced apoptosis, Cancer Lett. 287 (2010) 123-135.

[29] L. Kromidas, E. Perrier, J. Flanagan, R. Rivero, I. Bonnet, Release of antimicrobial actives from microcapsules by the action of axillary bacteria, Int J Cosmet Sci. 28 (2006) 103-108.

[30] A. Bahi, S.A Mansouri, E.A Memari, M.A Ameri, S.M Nurulain, S. Ojha, β-Caryophyllene, a CB_2 receptor agonist produces multiple change relevant toanxiety and depression in mice, Physiol Behav. 135 (2014) 119-124.

[31] M. Seo, T. Koshiba, Complex regulation of ABA biosynthesis in plants, trends in plant sci. 7 (2002) 41-48.

[32] T.R. Evans, S.B. Kaye, Retinoids: present role and future potential, Br J Cancer. 80 (1999) 1-8.

[33] R. Gupta, S.K. Chakrabarty, Gibberellic acid in plant Still a mystery unresolved, Plant Signal Behav. 8 (2013) e25504.

[34] J.J. Johnson, Carnosol: a promising anti-cancer and anti-inflammatory agent, Cancer Lett. 305 (2011) 1-7.

[35] S. Birtić, P. Dussort, F.X. Pierre, A.C. Bily, M. Roller, Carnosic acid, Phytochemistry. 115 (2015) 9-19.

[36] G.E. Bartley, P.A. Scolnik, Plant carotenoids: pigments for photoprotection, visual attraction, and human health, Plant Cell. 7 (1995) 1027-1038.

[37] J. Grassmann, Terpenoids as plant antioxidants, Vitam Horm. 72 (2005) 505-535.

[38] A.G. Atanasov, B. Waltenberger, E.M.P. Wenzig , T. Linder , C. Wawrosch, P. Uhrin, V. Temml, L. Wang, S. Schwaiger, E. Heiss, J.M. Rollinger, D. Schuster, J.M. Breuss, V. Bochkov, M.D. Mihovilovic, B. Kopp, R. Bauer, V.M. Dirsch, H. Stuppner, Discovery and resupply of pharmacologically active plant-derived natural products: a review, Biotechnol Adv. 33 (2015) 1582-1614.

[39] M. Bollen, S. Keppens, W. Stalmans, Specific features of glycogen metabolism in the liver, Biochem J. 336 (1998) 19-31.

[40] H.G. Hers, The control of glycogen metabolism in the liver, Annu Rev Biochem. 45 (1976) 167-190.

[41] W. Stalmans, M. Laloux, H.G. Hers, The interaction of liver phosphorylase a with glucose and AMP, Eur J Biochem. 49 (1974) 415-427.

[42] W. Pimenta, N. Nurjhan, P.A. Jansson, M. Stumvoll, J. Gerich, M. Korytkowski, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia. 37 (1994) 697-702.

[43] M.K. Hellerstein, R.A. Neese, P. Linfoot, M. Christiannsen, S. Turner, A. Letscher, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest. 100 (1997) 1305-1319.

[44] N.G. Oikonomakos, E.D. Chrysina, M.N. Kosmopoulou, D.D. Leonidas, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 Å resolution, Biochim Biophys Acta. 1647 (2003) 325-332.

[45] V.L. Rath, M. Ammirati, D.E. Danley, J.L. Ekstrom, E.M. Gibbs, T.R. Hynes, A.M. Mathiowetz, R.K. McPherson, T.V. Olson, J.L. Treadway, D.J. Hoover, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol. 7 (2000) 677-682.

[46] R. Kurukulasuriya, J.T. Link, D.J. Madar, Z. Pei, S.J. Richards, J.J. Rohde, A.J. Souers, B.G. Szczepankiewicz, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem. 10 (2003) 123-153.

[47] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[48] X. Wen, H. Sun, J. Liu, K. Cheng, P. Zhang, L. Zhang, J. Hao, L. Zhang, P. Ni, S.E. Zographos, D.D. Leonidas, K.M. Alexacou, T. Gimisis, J.M. Hayes, N.G. Oikonomakos, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem. 51 (2008) 3540-3554.

[49] T. Hunter, Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling, Cell. 80 (1995) 225-236.

[50] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, Protein tyrosine phosphatases in the human genome, Cell. 117 (2004) 699-711.

[51] Z.Y. Zhang, S.Y. Lee, PTP1B inhibitors as potential therapeutics in the treatment of type 2 diabetes and obesity, Expert Opin Investig Drugs. 12 (2003) 223-233.

[52] A.P. Combs, Recent advances in the discovery of competitive protein tyrosine phosphatase 1B inhibitors for the treatment of diabetes, obesity, and cancer, J Med Chem. 53 (2010) 2333-2344.

[53] S. Thareja, S. Aggarwal, T.R. Bhardwaj, M. Kumar, Protein tyrosine phosphatase 1B inhibitors: a molecular level legitimate approach for the management of diabetes mellitus, Med Res Rev. 32 (2012) 459-517.

[54] R.E. Cebula, J.L. Blanchard, M.D. Boisclair, K. Pal, N.J. Bockovich, Synthesis and phosphatase inhibitory activity of analogs of sulfircin, Bioorg Med Chem Lett. 7 (1997) 2015-2020.

[55] R.M. Chen, L.H. Hu, T.Y. An, J. Li, Q. Shen, Natural PTP1B inhibitors from Broussonetia papyrifera, Bioorg Med Chem Lett. 12 (2002) 3387-3390.

[56] C.S. Jiang, L.F. Liang, Y.W. Guo, Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades, Acta Pharmacologica Sinica. 33 (2012) 1217-1245.

[57] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[58] E.I. Solomon, J. Zhou, F. Neese, E.G. Pavel, New insights from spectroscopy into the structure/function relationships of lipoxygenases, Chem Biol. 4 (1997) 795-808.

[59] H.W. Gardner, Biological roles and biochemistry of the lipoxygenase pathway, Hortscience. 30 (1995) 197-205.

[60] R. Wisastra, F.J. Dekker, Inflammation, cancer and oxidative lipoxygenase activity are intimately linked, Cancers. 6 (2014) 1500-1521.

[61] L.A. Dailey, P. Imming, 12-Lipoxygenase: classification, possible therapeutic benefits from inhibition, and inhibitors, Curr Med Chem. 6 (1999) 389-398.

[62] V.E. Steele, C.A. Holmes, E.T. Hawk, L. Kopelovich, R.A. Lubet, J.A. Crowell, C.C. Sigman, G.J. Kelloff, Lipoxygenase inhibitors as potential cancer chemopreventives, Cancer Epidem Biomar. 8 (1999) 467-483.

[63] B. Samuelsson, S.E. Dahlen, J.A. Lindgren, C.A. Rouzer, C.N. Serhan, Leukotrienes and lipoxins: structures, biosynthesis, and biological effects, Science. 237 (1987) 1171-1176.

[64] X.Z. Ding, W.G. Tong, T.E. Adrian, 12-Lipoxygenase metabolite 12(S)-HETE stimulates human pancreatic cancer cell proliferation via protein tyrosine phosphorylation and ERK activation, Int J Cancer. 94 (2001) 630-636.

[65] J.A. Cornicelli, B.K. Trivedi, 15-Lipoxygenase and its inhibition: a novel therapeutic target for vascular disease, Curr Pharm Des. 5 (1999) 11-20.

[66] U.P. Kelavkar, J.B. Nixon, C. Cohen, D. Dillehay, T.E. Eling, K.F. Badr, Overexpression of 15-lipoxygenase-1 in PC-3 human prostate cancer cells increases tumorigenesis, Carcinogenesis. 22 (2001) 1765-1773.

[67] B.D. Roebuck, D.S. Longnecker, K.J. Baumgartner, C.D. Thron, Carcinogen-induced lesions in the rat pancreas: effects of varying levels of essential fatty acid, Cancer Res. 45 (1985) 5252-5256.

[68] J. Zhang, V.L. Go, High fat diet, lipid peroxidation, and pancreatic carcinogenesis, Adv Exp Med Biol. 399 (1996) 165-172.

[69] D.P. Rose, Dietary fatty acids and cancer, Am J Clin Nutr. 66 (1997) 998S-1003S.

[70] P.L. Zock, M.B. Katan, Linoleic acid intake and cancer risk: a review and meta-analysis, Am J Clin Nutr. 68 (1998) 142-153.

[71] W.G. Tong, X.Z. Ding, R.C. Witt, T.E. Adrian, Lipoxygenase inhibitors attenuate growth of human pancreatic cancer xenografts and induce apoptosis through the mitochondrial pathway, Mol Cancer Therapeutics. 1 (2002) 929-935.

[72] C. Kemal, P. Louis-Flamberg, R. Krupinski-Olsen, A.L. Shorter, Reductive inactivation soybean lipoxygenase-1 by catechols: a possible mechanism for regulation of lipoxygenase activity, Biochemistry. 26 (1987) 7064-7072.

[73] R. Mogul, E. Johansen, T.R. Holman, Oleyl sulfate reveals allosteric inhibition of soybean lipoxygenase-1 and human 15-lipoxygenase, Biochemistry. 39 (2000) 4801-4807.

[74] A.F.A. Cros, Ph.D. dissertation thesis: Action de l'alcohol amylique sur l'organisme, University of Strasbourg, Strasbourg, 1863.

[75] A. Crum-Brown, T.R. Fraser, On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived

from strychia, brucia, thebaia, codeia, morphia and nicotia, Trans R Soc Edinburgh. 25 (1868) 151-203.

[76] C. Richet, On the relationship between the toxicity and the physical properties of substances, Compt Rendus Seances Soc Biol. 9 (1893) 775-776.

[77] E. Overton, Osmotic properties of cells in the bearing on toxicology and pharmacology, Z Physik Chem. 22 (1897) 189-209.

[78] H. Meyer, On the theory of alcohol narcosis I. Which property of anesthetics gives them their narcotic activity?, Arch Exp Pathol Pharmakol. 42 (1899) 109-118.

[79] L.P. Hammett, Some relations between reaction rates and equilibrium constants, Chem Rev. 17 (1935) 125-136.

[80] L.P. Hammett, The effect of structure upon the reactions of organic compounds. benzene derivatives, J Am Chem Soc. 59 (1937) 96-103.

[81] R.W. Taft, Polar and steric substituent constants for aliphatic and o- Benzoate groups from rates of esterification and hydrolysis of esters1, J Am Chem Soc. 74 (1952) 3120-3128.

[82] C. Hansch, T. Fujita, p-σ-π Analysis. A method for the correlation of biological activity and chemical structure, J Am Chem Soc. 86 (1964) 1616-1626.

[83] C. Hansch, Quantitative approach to biochemical structure-activity relationships, Acc Chem Res. 2 (1969) 232-239.

[84] S.M. Free Jr, J.W. Wilson, A Mathematical contribution to structure-activity studies, J Med Chem. 7 (1964) 395-399.

[85] T. Fujita, T. Ban, Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters, J Med Chem. 14 (1971) 148-152.

[86] L.B. Kier, Molecular orbital theory in drug research, Academic Press, New York, 1971.

[87] A.T. Balaban, F. Harary, The characteristic polynomial does not uniquely determine the topology of a molecule, J Chem Doc. 11 (1971) 258-259.

[88] M. Randić, On the recognition of identical graphs representing molecular topology, J Chem Phys. 60 (1974) 3920-3928.

[89] L.B. Kier, L.H. Hall, W.J. Murray, M. Randic, Molecular connectivity I: relationship to nonspecific local anesthesia, J Pharm Sci. 64 (1975) 1971-1974.

[90] R.H. Rohrbaugh, P.C. Jurs, Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships, Anal Chim Acta. 199 (1987) 99-109.

[91] D.T. Stanton, P.C. Jurs, Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies, Anal Chem. 62 (1990) 2323-2329.

[92] R. Todeschini, M. Lasagni, E. Marengo, New molecular descriptors for 2D- and 3D-structures, Theory J Chemom. 8 (1994) 263-273.

[93] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, J Phys Chem. 100 (1996) 10400-10407.

[94] A.M. Ferguson, T.W. Heritage, P. Jonathon, S.E. Pack, L. Phillips, J. Rogan, P.J. Snaith, EVA: a new theoretically based molecular descriptor for use in QSAR/QSPR analysis, J Comput Aided Mol Des. 11 (1997) 143-152.

[95] J. Schuur, P. Selzer, J. Gasteiger, The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity, J Chem Inf Comput Sci. 36 (1996) 334-344.

[96] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors, J Chem Inf Comput Sci. 42 (2002) 682-692.

[97] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J Am Chem Soc. 110 (1988) 5959-5967.

[98] J. Gasteiger, Handbook of chemoinformatics. From data to knowledge in 4 volumes, Wiley-VCH, Weinheim, 2003.

[99] T.I. Oprea, 3D QSAR modeling in drug design, in: P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere (Eds.), Computational medicinal chemistry for drug discovery, Marcel Dekker, New York, 2004.

[100] R. Franke, Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.

[101] K. Osmialowski, J. Halkiewicz, A. Radecki, R. Kaliszan, Quantum chemical parameters in correlation analysis of gas-liquid chromatographic retention indices of amines, J Chromatogr. 346 (1985) 53-60.

[102] K. Fukui, Theory of orientation and stereoselection, Springer- Verlag, New York, 1975.

[103] H. Sklenar, J. Jäger, Molecular structure-biological activity relationships on the basis of quantum-chemical calculations, Int J Quant Chem. 16 (1979) 467-484.

[104] K. Tuppurainen, S. Lötjönen, R. Laatikainen, T. Vartiainen, U. Maran, M. Strandberg, T. Tamm, About the mutagenicity of chlorine-substituted furanones and halopropenals. A QSAR study using molecular orbital indices, Mutat Res Fund Mol Mech Mut. 247 (1991) 97-102.

[105] I. Fleming, Frontier orbitals and organic chemical reactions, John Wiley & Sons, New York, 1976.

[106] D.F.V. Lewis, C. Ioannides, D.V. Parke, Interaction of a series of nitriles with the alcohol inducible isoform of P450- computer analysis of structure-activity relationships, Xenobiotica. 24 (1994) 401-408.

[107] M. Karelson, V.S. Lobanov, Quantum-Chemical Descriptors in QSAR/QSPR Studies, Chem Rev. 96 (1996) 1027-1043.

CHAPTER 2

Review of Literature

Terpenoids, also called isoprenoids, are substances that are made up from isoprene (2-methylbuta- 1, 3-diene) units. Most of the terpenoids are plant origin and more than 40,000 individual compounds have been accounted. In plants, some terpenoids are involved in cellular function and maintenance as primary metabolites. However, thousands of terpenoids produced by plants have been characterized as secondary metabolites and they show significant pharmacological applications such as anti-viral, anti-bacterial, anti-malarial, anti-inflammatory, inhibition of cholesterol synthesis and anti-cancer activities [1]. So many researchers have paid attentions on naturally obtained terpenoids or their synthetic modifications and developed quantitative structure activity relationships (QSARs) of a number of terpenoids with different biological activities in order to predict the activity of the compounds that have not yet been synthesized or experimentally tested.

This review work has been briefly divided into two subgroups which are as follows:

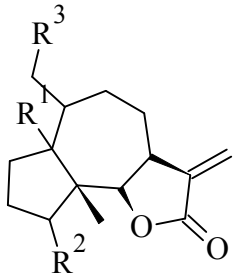
2.1. Terpenoids with different biological activity

2.2. Quantitative structure-activity relationships of bioactive terpenoids

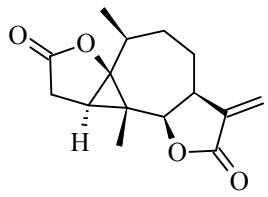
2.1. Terpenoids with different biological activity

Terpenoids are the most abundant and diverse class of natural products. Many biological activities of terpenoids have made them a widely used resource for traditional and modern human utilization. The sesquiterpene lactones are known for their wide variety of biological activities. Some medicinal plants containing sesquiterpene lactones show anti-inflammatory

properties. Recio et al. reported the in vivo anti-inflammatory activities of seven pseudoguaianolide type of sesquiterpene lactones: 4-α-Oacetyl- pseudoguaian-6β-olide (1), hymenin (2), ambrosanolide (3), tetraneurin (4), parthenin (5), hysterin (6) and confertdiolide (7) isolated from several species of Parthenium where confertdiolide (7) was the most active compound [2-6].

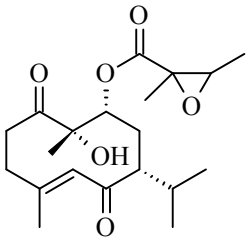


(1): R^1=α-H, R^2=α-OAc, R^3=H; (2): R^1=α-OH, R^2=O, R^3=H; (3): R^1=α-OH, R^2=β-OAc, R^3=H; (4): R^1=α-OH, R^2=O, R^3=OAc; (5): R^1=β-OH, R^2=O, R^3=H; (6): R^1=α-H, R^2=α-OAc, R^3=OH



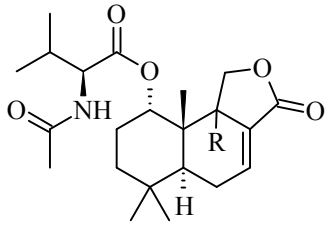
(7)

Ishibashi et al. isolated two new sesquiterpenoid esters with nine known flavonoids from *Blumea balsamifera*, a tropical Compositae plant. Compound (8) proved to be weakly cytotoxic against Jurkat human T-cell leukemia cells [7].



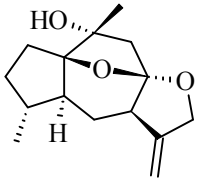
(8)

Stierle et al. isolated two new drimane sesquiterpene lactones and one new tricarboxylic acid derivative from the Berkeley Pit extremophilic fungus *Penicillium solitum*. The structures of the new compounds were elucidated by spectroscopic analyses and it was found that berkedrimanes A (9) and berkedrimanes B (10) inhibited the signal transduction enzymes caspase-1 and caspase-3 and mitigated the production of interleukin 1-β in the induced THP-1 (pro-monocytic leukemia cell line) assay [8].

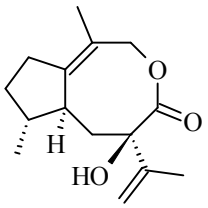


(9): R=H; (10): R=OH

Zhao et al. isolated two new sesquiterpenoids with guaiane skeletons holosericin A (11) and holosericin B (12) from the medicinal plant *Daphne holosericea* (Diels) Hamawa (Thymelaeceae). Both the compounds were evaluated for inhibitory activities against acetylcholinesterase and compound (12) exhibited a moderate activity with 31% inhibition [9].

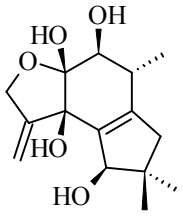


(11)

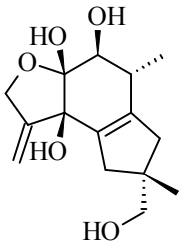


(12)

Liu et al. have isolated some sesquiterpenoids, clitocybulol derivatives from the solid culture of the edible fungus *Pleurotus cystidiosus*. Compounds (13) and (14) exhibited moderate inhibitory activity against protein tyrosine phosphatase-1B (PTP1B) with IC_{50} values of 49.5 μM and 38.1μM respectively [10].

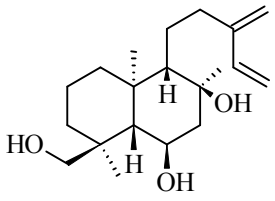


(13)

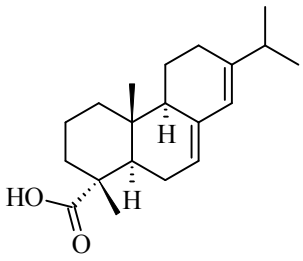


(14)

A large number of diterpenoids have been reported for several biological actions including antibacterial, antifungal, antiinflammatory, antileishmanial, antialgal, cytotoxic and antitumour activities [11-15]. Andalusol (15) and abietic acid (16) exhibited anti-inflammatory activities in vivo and in vitro models of inflammation [16, 17].

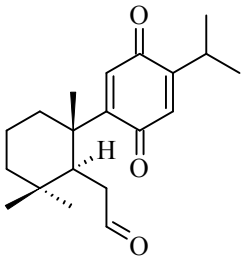


(15)

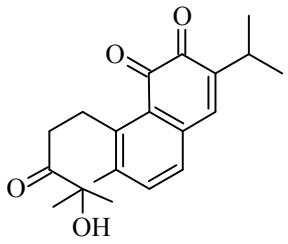


(16)

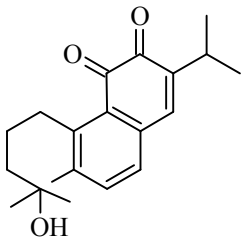
Zhang et al. isolated a new 7,8-seco-abietane diterpene derivative, 7,8-seco-para-ferruginone (17), and two new 4,5-seco-5,10-friedo-abietane diterpenoids, 4-hydroxysaprorthoquinone (18) and 3-keto-4- hydroxysaprorthoquinone (19) together with two new compounds from the roots of *Salvia prionitis*. Compound 17 showed antimicrobial activities against Staphylococcus aureus and Micrococcus luteus. Compound 18 exhibited significant inhibitions against topoisomerase I. Compound 19 displayed cytotoxic activities against HL-60 human leukemia and the SGC-7901 and MKN-28 stomach cancer cell lines [18].



(17)



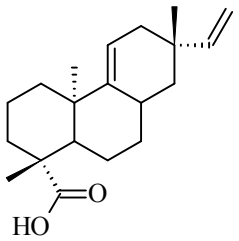
(18)



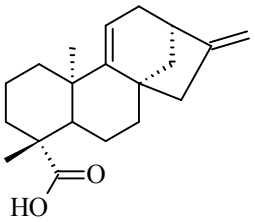
(19)

Ahn et al. isolated eight diterpenoids from CH_2Cl_2-soluble extract of the roots of *Acanthopanax koreanum* (Araliaceae) and were evaluated for their inhibitory effect on protein tyrosine phosphatase 1B (PTP1B) which has been proposed as a therapeutic target for the treatment of type 2 diabetes and obesity. A kaurane-type diterpene, 16αH,17-isovaleryloxy-*ent*-kauran-19-oic

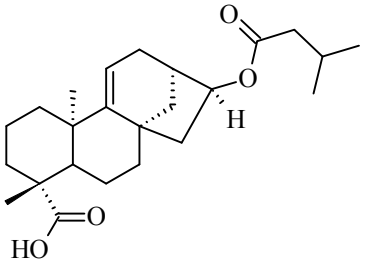
acid (20), inhibited PTP1B in a non-competitive manner where as acanthoic acid (21) and *ent*-kaur-16-en-19-oic acid (22) inhibited PTP1B in dose-dependent manners [19].



(20)

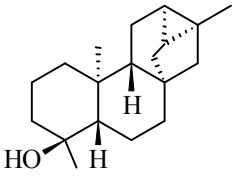


(21)

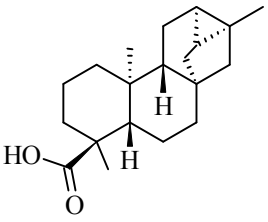


(22)

Litaudon et al. isolated a new diterpenoid ent-trachyloban-4β-ol (23), and five known ent-trachylobane or ent-atisane compounds from *Xylopia caudata*. All the compounds exhibited cytotoxicity against KB and HCT-116 cell lines and only ent-trachyloban-18-oic acid (24) exhibited weak binding activity to antiapoptotic protein Bcl-xL [20].

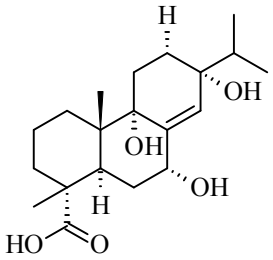


(23)



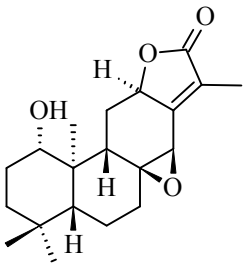
(24)

Guo et al. extracted ten new abietane diterpenoids, aquilarabietic acids A–J, and a new podocarpane diterpenoid, aquilarabietic acid K from the petroleum ether and ethanol extracts of Chinese eaglewood. Aquilarabietic acids A (25) exhibited remarkable antidepressant activity in vitro by inhibiting norepinephrine reuptake in rat brain synaptosomes by 81.4% [21].



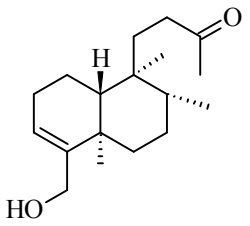
(25)

Yue et al. reported that some isolated diterpenoids, named eurifoloids A–R, from *Euphorbia neriifolia* exhibited anti-HIV activities. It was stated that eurifoloids F (26) showed significant anti-HIV activities with EC_{50} values of 7.40 ± 0.94 μM [22].



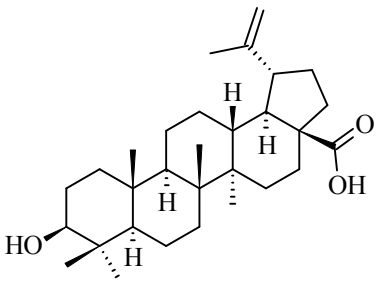
(26)

Luo et al. isolated five new iridoid glucoside derivatives, three new diterpenoids and 11 known compounds from the aqueous EtOH extract of *Caryopteris glutinosa*. Cell-based estrogen biosynthesis assays indicated that caryopterisoid B (27), a diterpenoid, promote the biosynthesis of estrogen E2, with EC_{50} values 8.0 μM, in human ovarian granulosa-like KGN cells via upregulating the expression of aromatase [23].

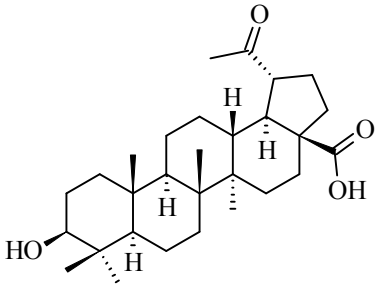


(27)

Triterpenoids often exhibit a variety of biological activities such as anti-HIV, anti-inflammatory, ichthyotoxic, anti-tumour-promotin and antimycobacterial activities [24-30]. Lee et al. isolated betulinic acid (28) and platanic acid (29) from the leaves of *Syzigium claviflorum*. These compounds were found to be inhibitors of HIV replication in H9 lymphocyte cells [31].

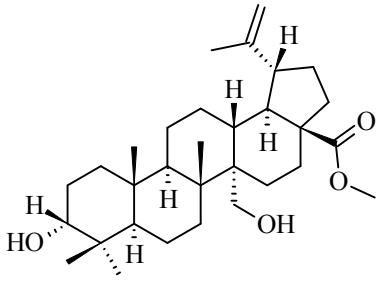


(28)



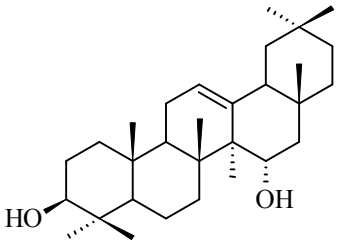
(29)

Nomura et al. isolated three new triterpenoids, 3α, 27-dihydroxylup-20(29)-en-28-oic acid methyl ester, 3α-acetoxy-27-hydroxylup-20(29)-en-28-oic acid methyl ester, and 3α-acetoxyolean-12-ene-27,28-dioic acid 28-methyl ester along with four known lupene-type triterpenoids from the roots of *Peganum nigellastrum*. They elucidated their structures by means of NMR techniques and only 3α, 27-dihydroxylup-20(29)-en-28-oic acid methyl ester (30) is a DNA topoisomerase II inhibitor that plays a crucial role in DNA metabolism [32].

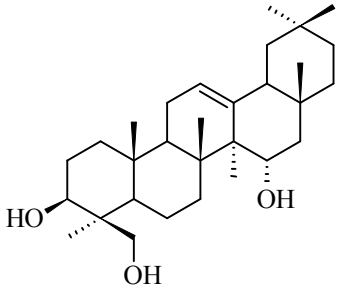


(30)

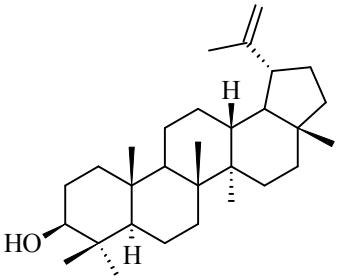
Wada et al. screened naturally occurring lupane- and oleanane-type triterpenoids for human DNA topoisomerases I and II inhibitory activities and it was found that Olean-12-en-3β,15α-diol (31), olean-12-en-3β,15α,24-triol (32), lupeol (33), and betulin (34) showed significant activity to DNA topoisomerases II [33].



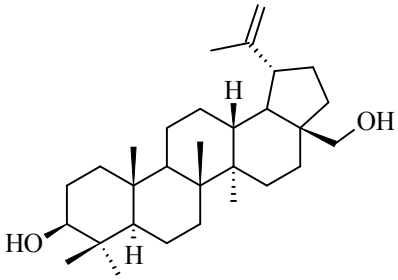
(31)



(32)

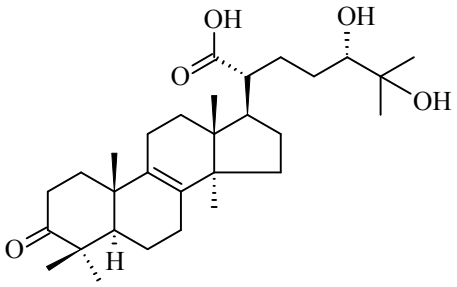


(33)



(34)

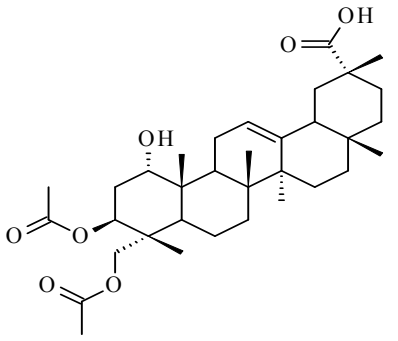
Yoshikawa et al. isolated two new lanostane triterpenoids and ten new lanostane triterpene glycosides from the fruit bodies of *Fomitopsis pinicola*. Their biological activity was investigated against COX-1 and COX-2 and compound (35) showed significant activities corresponding to indomethacin against COX-2 [34].



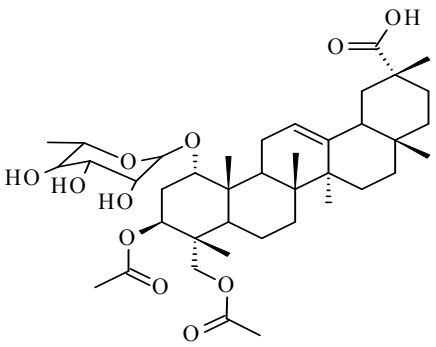
(35)

Oh et al. extracted two new ursane-type triterpenoids, 3α,19α-dihydroxyurs-12,20(30)-dien-24,28-dioic acid and 3α,19α-dihydroxyurs-12-en-24,28-dioic acid, together with 12 known ursane- and oleanane-type triterpenoids from the leaves of persimmon (*Diospyros kaki*). Triterpenoids with a 3α-hydroxy moiety were not active but with a 3β-hydroxy group were found to inhibit protein tyrosine phosphatase 1B (PTP1B) [35].

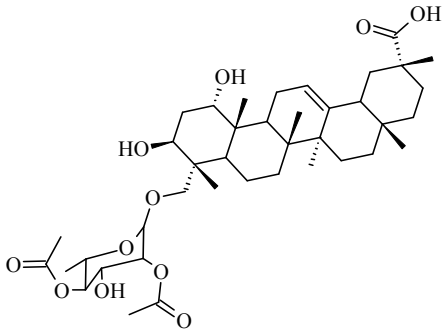
Litaudon et al. isolated 15 pentacyclic triterpenoids possessing olean-12-en-28-oic acid and olean-12-en-29-oic acid aglycons from the ethyl acetate extracts of the leaves and flowers of *Combretum sundaicum* and the leaves of *Lantana camara*. Only compounds (36-40) have binding affinity of the antiapoptotic protein Bcl-xL, capable of disrupting the Bcl-xL/Bak association [36].



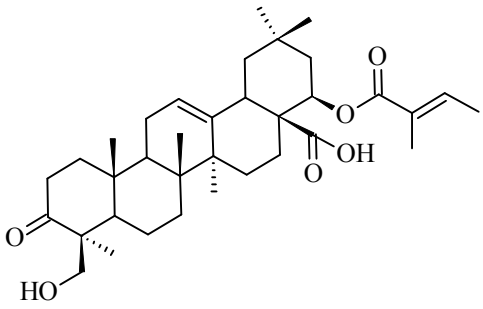
(36)



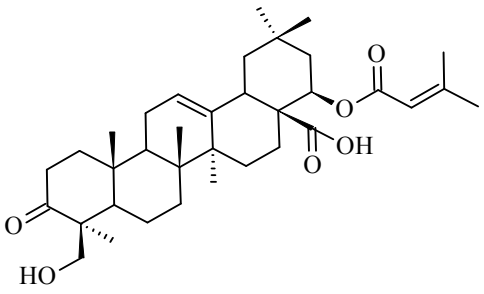
(37)



(38)



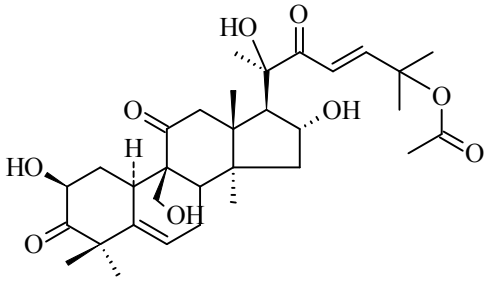
(39)



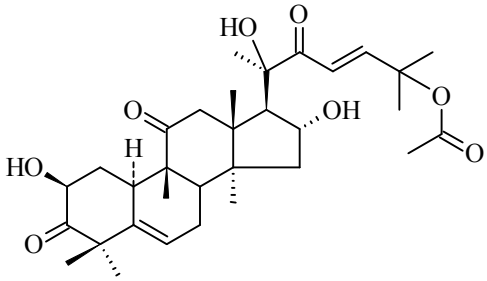
(40)

Zhao et al. isolated 21 cucurbitane-type triterpenoids from the stems of *Cucumis melo*. Their structures were elucidated on the basis of spectroscopic studies, chemical methods, and comparison with spectroscopic data in the literature. Two compounds, cucurbitacin A (41) and

cucurbitacin B (42), exhibited significant cytotoxic activity against the proliferation of A549/ATCC and BEL7402 cells in vitro [37].

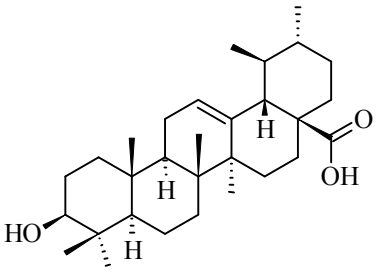


(41)

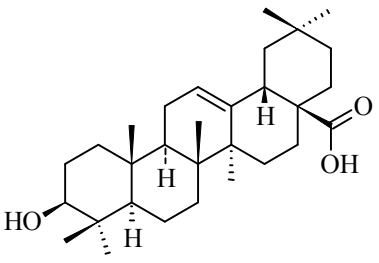


(42)

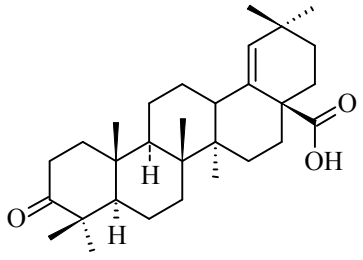
Li et al. investigated cytotoxicity and inhibition on human DNA topoisomerase I and II of 74 plant-originated triterpenoids and triterpenoid glycosides [38]. Soto et al. investigated the oral antidiabetic activity of four structurally-related triterpenic acids: ursolic (43), oleanolic (44), moronic (45) and morolic (46) acids and all compounds showed significant antidiabetic activity [39].



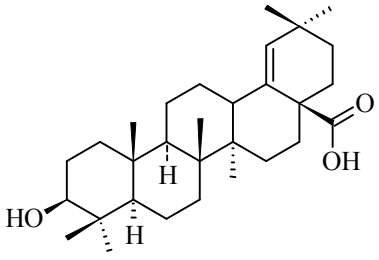
(43)



(44)

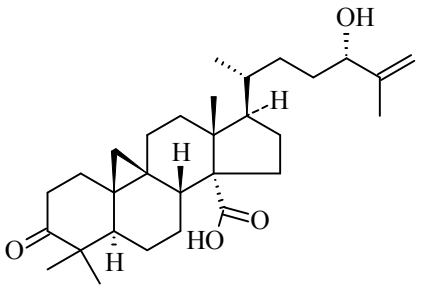


(45)

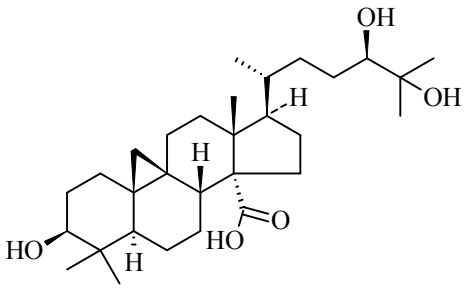


(46)

Hao et al. isolated five new triterpenoids, caloncobic acids A and B, caloncobalactones A and B, and glaucalactone, along with the known compounds 3β,21β-dihydroxy-30-nor-(D:A)-friedo-olean-20(29)-en-27-oic acid and acetyltrichadenic acid B from the leaves of *Caloncoba glauca*. Caloncobic acids A and B (47 and 48) exhibited strong inhibitory activities against mouse and human 11β-hydroxysteroid dehydrogenase type 1 which can influence factors affecting metabolic syndrome such as insulin resistance and dyslipidemia [40].



(47)

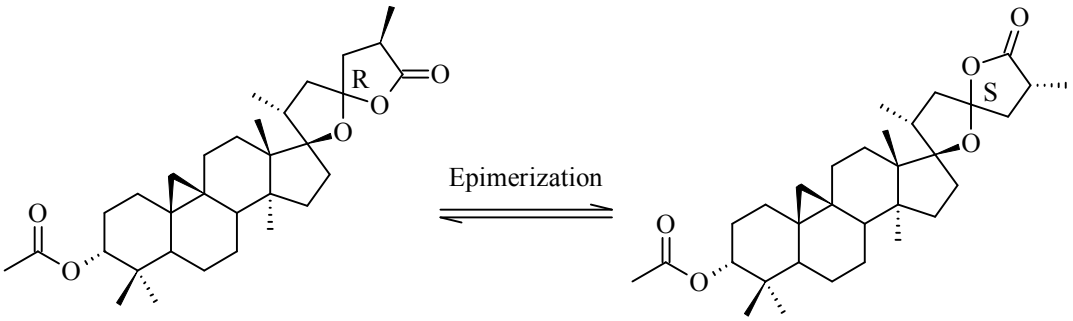


(48)

Shi et al. isolated five triterpenoids with a new 25-norfern carbon skeleton , a lupane triterpenoid, and four 20-hydroxyprogesterone acyl esters, together with 23 known compounds from the stem (with skin removed) of *Sinocalamus affinis*. These five triterpenoids with a new 25-norfern carbon skeleton exhibited inhibitory activity against protein tyrosine phosphatase 1B [41].

Yu et al. have reported that some of isolated novel iridal-type triterpenoids from the ethanol extract of the rhizomes of *Iris tectorum* exhibited neuroprotective activities against serum-deprivation-induced PC12 cell damage [42].

Eight pairs of epimeric triterpenoids were isolated from *Abies faxoniana* as inseparable mixtures of C-23 epimers in a specific proportion. It was stated that Compound 49 showed cytotoxicity against three hepatoma cell lines, namely, HepG2, Huh7, and SMMC7721 but exerted low cytotoxicity on normal QSG7701 hepatic cells, indicating its selective cytotoxicity for hepatoma cells [43].



49a 49b

2.2. Quantitative structure-activity relationships of bioactive terpenoids

The structure and reactivity relationship on a quantitative basis was introduced by Hammett (1935, 1937) where substituent effects on reaction mechanisms are explained by using two parameters namely substituent constant and reaction constant [44, 45]. In the 1950s, Taft proposed a way for separating polar, steric, and resonance effects of substituent in aliphatic compounds and introduced the first steric parameter, E_S [46]. The contributions from Hammett and Taft set forth the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita. (1964). Fujita and Hansch then constructed a linear equation (called Hansch linear

equation) by combining hydrophobic constants with Hammett's electronic constants [47, 48]. The failure of linear equations led to the development of the parabolic Hansch equation in cases with extended hydrophobicity ranges [49]. At the same time, Free and Wilson formulated an additive model of substituent contributions to biological activities, giving a push to the QSAR development [50].The quantitative structure-activity relationships have been further developed by Kier and Hall on their studies of connectivity indices based on hydrogen suppressed molecular structures [51-53]. Other popular QSAR approaches are HQSAR, Inverse QSAR, and Binary QSAR [54-57]. In 1988, Cramer et al. proposed 3D-QSAR methodology, Comparative Molecular Field Analysis (CoMFA) [58]. Other 3D-QSAR approaches have been developed, such as Comparative Molecular Similarity Indices Analysis (CoMSIA) [59, 60] or Self Organizing Molecular Field Analysis (SomFA) [61].

Jiang et al. used ether and ester analogs of artemisinin for comparative molecular field analysis (CoMFA) to correlate between the physicochemical properties and the in vitro activities. Four alignment models were used in this study and correlation study suggested that all to have good predictive values [62].

A three-dimensional quantitative structure-activity relationship paradigm was used by Chen et al. to correlate between the physicochemical properties and the in vitro bioactivities of ginkgolide analogues. They designed compounds on the basis of CoMFA analysis and it was found that three of these new designed compounds are more potent than that of ginkgolides [63].

Woolfrey et al. compared two 3D-QSAR methods, comparative molecular field analysis (CoMFA) and hypothetical active site lattice (HASL), with respect to the analysis of a training set of 154 artemisinin analogues. Five models including a complete HASL and two trimmed

versions, as well as two CoMFA models were created. Although the differences between CoMFA contour maps and the HASL output, each of the four predictive models exhibited a similar ability to predict the activity of a test set of 23 artemisinin analogues [64].

Heravi et al. carried out a quantitative structure-property relationship study based on multiple linear regressions (MLR) and artificial neural network (ANN) techniques to investigate the retention behavior of some terpenes on the polar stationary phase. A collection of 53 noncyclic and monocyclic terpenes were divided into two groups, a training set of 41 molecules and a test set of 12 molecules. A total of six descriptors containing one electronic, two geometric, two topological and one physicochemical descriptors were appeared in the MLR model and it has been found that electronic, topological and physicochemical descriptors have a pronounced effect on the retention behavior of the terpenes [65].

Grodnitzky et al. developed quantitative structure-activity relationships (QSARs) to predict insect toxicity of monoterpenoids and its derivatives. They found a linear relationship between house fly toxicity and Mulliken populations in aromatic monoterpenoids. Multiple linear regression study of an E-State descriptor and a GETAWAY (GEometry, Topology and Atomic Weights AssemblY) descriptor established a relationship with house fly toxicity for a number of monoterpenoids [66].

Artemisinin is an effective drug against chloroquine-resistant *Plasmodium falciparum* strains and cerebral malaria. Cheng et al. carried out molecular docking simulations to probe the interactions of artemisinin and its analogues with hemin and then performed 3D-QSAR study employing comparative molecular force fields analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). Docking simulations provided probable bioactive conformations of

artemisinin analogues. The partial least squares (PLS) analysis suggests that the calculate binding energies correlate well with the activity values. The results of molecular docking and 3D-QSAR help to explain the binding model and activity of new synthesized artemisinin derivatives [67].

Kuriyama et al. performed comparative molecular field analysis (CoMFA) with thirteen seco-prezizaane terpenoids isolated from star anise species (*Illcium floridanum, Illcium parviflorum, and Illcium verum*) for their ability to inhibit the specific binding of [^{3}H]4'-ethynyl-4-n-propylbicycloorthobenzoate (EBOB), a non-competitive antagonist of gamma-aminobutyric acid (GABA) receptors, to housefly-head and rat-brain membranes. This 3D-QSAR study demonstrated that seco-prezizaane terpenoids can bind to the same site as do picrotoxane terpenoids such as picrotoxinin and picrodendrins. The CoMFA maps identified the parts of the molecules essential to high activity in housefly GABA receptors [68].

For the development of non toxic analogues of artemisinin, Avery et al. utilized comparative molecular field analysis (CoMFA) and hologram QSAR (HQSAR) with a series of 211 artemisinin analogues with known in vitro antimalarial activity. The bioactive conformation of artemisinin and its analogues were found by two ways: (i) X-ray structure of artemisinin and (ii) the hemin-docked conformation. The generated CoMFA and HQSAR models have excellent statistically properties and possessed good predictive ability for test set compounds [69].

Ghafourian et al. investigated the structural requirements of penetration enhancers using the Quantitative Structure-Activity Relationship (QSAR) technique. The possible mechanism of skin penetration of 5-fluorouracil, diclofenac sodium (DFS), hydrocortisone (HC), estradiol and

benazepril by naturally occurring terpenes, pyrrolidinone and N-acetylprolinate derivatives have been discussed [70].

Guha et al. developed QSAR models to predict the biological activity of 179 artemisinin analogues. They generated topological, geometric, and electronic descriptors for linear (multiple linear regression) and nonlinear (computational neural network) models to link the structures and biological activity. The best nonlinear model is superior to the best linear model in terms of pure predictive ability [71].

Schmidt et al. conducted a QSAR study of the seco-prezizaane-type sesquiterpenes pseudoanisatin and parviflorolide from *Illicium*. The compounds are noncompetitive antagonists at housefly (*Musca domestica*) γ-aminobutyric acid (GABA) receptors and show selectivity toward the insect receptor. This QSAR study provided insight into the structural basis of selectivity of seco-prezizaane-type sesquiterpenes and properties of the binding sites at GABA receptor-coupled chloride channels of insects and mammals [72].

Cortes-Selva carried out comparative molecular similarity indices analysis (CoMSIA), a three-dimensional quantitative structure-activity relationships with a number of dihydro-beta-agarofuran sesquiterpenes as modulators at the P-glycoprotein-like transporter. This 3D-QSAR study was employed to characterize the steric (contribution of 5.4%), electrostatic (58.9%), lipophilic (10.0%), and hydrogen-bond-donor (13.3%) and acceptor (7.5%) requirements of these sesquiterpenes with the receptor [73].

Siedle et al. investigated a set of 103 different sesquiterpene lactones for their NF-kappaB inhibiting properties and their activity values were submitted to a QSAR study [74].

Macias et al. carried out a quantitative structure-activity study with 34 guaianolides having different numbers of hydroxyl groups and ester side chains of variable length and structure to evaluate the effect of lipophilia/aqueous solubility on etiolated wheat coleoptiles elongation. Data show a strong influence of logP values [75].

Zhu et al. performed three dimensional quantitative structure-activity relationship studies employing comparative molecular fields analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA) and hologram QSAR (HQSAR) with a number of Ginkgolide analogues and their bioactivities against PAF receptor. Three rational and predictive QSAR models were constructed with high q^2 released ranging from 0.583 to 0.684. This study helps to understand the possible binding mechanism between ginkgolides and human PAF receptor and should be very useful in discovering new drugs as PAF antagonists [76].

Ishihara et al. observed that the cytotoxicity of betulinic acid derivatives can be predicted by heat of formation, hydrophobicity (log P), water solubility, ionization potential, electron affinity and dipole moment but not by molecular size. These parameters were determined by semi-empirical molecular-orbital method [77].

Two and three dimensional quantitative structure activity relationship models were constructed for a series of 20 butitaxel analogues (paclitaxel and docetaxel) to investigate the properties associated with microtubule assembly and stabilization. A CoMFA model was built using steric and electrostatic fields with $r^2 = 0.943$ and a cross-validated r^2 (i.e. q^2) = 0.376. Using the same data, HQSAR generated an $r^2 = 0.919$ and a $q^2 = 0.471$. All the analogues were docked into beta-tubulin model and their docking pose were nearly identical with paclitaxel bound to the protein.

A modest correlation (r^2 = 0.53) was found between activity and docking energy of all the butitaxel analogues in the dataset [78].

Comparative molecular field analysis (CoMFA), a three-dimensional quantitative structure-activity relationship (3D-QSAR) paradigm was used to observe effect of argentatin B derivatives on the growth of K562 cancer cell lines. This study indicated that the activity depends on bulky group at C-2, a C1-C2 double bond, and low electronic density at C-25 [79].

Sung et al. constructed QSAR models between oleanolic acid analogues and the inhibition activities of protein tyrosine phosphatase-1B (PTP1B) using 2D-QSAR and HQSAR methodologies. From the analysis results of these two models, the HQSAR was statistically better than 2D-QSAR [80].

Rasulev et al. studied quantitative structure-estrogenic activity relationship in a series of terpenoid esters with aromatic and aliphatic acid substituent isolated from Ferula plants. In this QSAR approach the data generated from three-dimensional structures of terpenoids and from quantum-chemical calculations at the B3LYP/6-31G(d, p) level of theory. A significant QSAR model with r^2=0.892 showed that the estrogenic activity depends on the parameters such as molecular shape, number of phenolic groups, surface polarity and the energy of the highest occupied molecular orbital [81].

Hemmateenejad et al. developed a quantitative structure property relationship (QSPR) for the Kovats retention indices of a large number of terpenoids [82].

Scotti et al. investigated a QSAR study of 37 different sesquiterpene lactones from the Asteraceae family with their cytotoxic activities. A single model was constructed using 3D

molecular descriptors and genetic algorithms which explained the important properties for the inhibition potency [83].

Chang et al. developed a quantitative structure-activity relationship model of eleven terpenoids with significant neuroprotective activity. This study indicated that the activity was mainly governed by lipophilicity, shape index, and electrostatic property [84].

Multidrug resistance is one of the major challenges in many diseases. Reyes et al. reported the inhibitory activity of a number of 76 dihydro-beta-agarofuran sesquiterpenes, tested on NIH-3T3 cells expressing the human P-glycoprotein (Pgp) multidrug transporter. A three-dimensional quantitative structure-activity relationship model using the comparative molecular similarity indices analysis (CoMSIA) was used to understand the structural basis for inhibitory activity and guide the design of more potent Pgp inhibitors [85].

The structural requirements of terpenes and terpenoids as penetration enhancers have been investigated by Kang et al. using the Quantitative Structure-Activity Relationship (QSAR) technique. In this work, the human skin penetration effect of 49 terpenes and terpenoids were compared by the in vitro permeability coefficients of haloperidol through excised human skin. They suggested that an ideal terpene enhancer must have at least one or combinations of the following properties: hydrophobic, in liquid form at room temperature, with an aldehyde or ester functional group but not acid group, and is neither a triterpene nor tetraterpene [86].

Setzer has been used molecular docking techniques to examine the potential binding sites of a number of triterpenoids inhibitors of topoisomerase II. The molecular docking results reveal that most of the triterpenoid ligands preferentially bind to the DNA binding site of topoisomerase II,

while a few also bind to the ATP binding site. This study gives some idea about the cytotoxicity of these natural products [87].

Breast cancer is one of the leading causes of death among women. Verma et al. used four series of taxane derivatives to correlate their inhibitory activities against breast cancer cells with their hydrophobic and steric properties to understand their chemical-biological interactions. This study suggested that the inhibitory activities of these compounds against breast cancers are mainly dependent either on their hydrophobicity or the hydrophobic/molar refractivity descriptor of their substituents [88].

Hansch et al. discussed the interaction of taxanes with the tubulin/microtubule system by the formulation of six QSARs. They suggested that hydrophobicity of the substituents and steric parameters are the important determinants of the activity. They believed that two QSARs models can help to provide guidance in design compounds that may have high biological activities [89].

Cardoso et al. studied the activity of artemisinin and some derivatives against D-6 strains of *Plasmodium falciparum*. They used molecular electrostatic potential maps to identify important features of the compounds and then use those to propose new artemisinin derivatives. The partial least squares (PLS) method was then utilized to produce a predictive model. They found that the important descriptors for construction of the model were the highest occupied molecular orbital energy, atomic charges on the atoms O1 and C3, molecular volume, and hydrophilic index [90].

Wang et al. synthesized a series of terpenoid compounds from alpha- and beta-pinenes. The antifeedant activities of these compounds were tested on the aphid, *Lipaphis erysimi* (Kalt.). The statistically best QSAR model suggested that the relative number of O atoms, HOMO-LUMO

energy gap, molecular volume, and total charge on the positively charged fragments were the important indices to predict the antifeedant activity [91].

Setzer was carried out quantum chemical calculations at the B3LYP/6-31G* level of theory on 20 celastroid triterpenoids to get a set of molecular electronic properties and to correlate these with cytotoxic activities. The cytotoxic activities of these triterpenoids may be roughly correlated with the energies of the frontier molecular orbitals (E_{HOMO} and E_{LUMO}), the HOMO-LUMO energy gap, the dipole moment, the charge on C(6), and the electrophilicity on C(6) [92].

In order to find out new taxane derivatives with fewer side-effects and improved anticancer activity, Verma et al. correlated their inhibitory activities against lung cancer cells with hydrophobic and steric descriptors for understanding their chemical-biological interactions. On the basis of this QSAR study, six compounds were suggested as potential synthetic targets [93].

Steinmetz et al. conducted 3D-QSAR studies of trichothecene mycotoxins, toxic natural products of fungi from the family *Hypocreaceae*, are potent inhibitors of protein synthesis. They explained the role of electrostatics and steric factors in the activity of the toxins and show that the conformation of the macrolide ring influences the toxicity of the macrolide toxins [94].

Schmidt et al. conducted a QSAR study on a set of 40 sesquiterpene lactones against *T. brucei rhodesiense* (which causes East African sleeping sickness), *T. cruzi*, *Leishmania donovani* and *Plasmodium falciparum* [95].

Hansch et al. conducted QSAR studies on a series of C2-modified 10-deacetyl-7-propionyl cephalomannines with respect to their binding affinities toward beta-tubulin and cytotoxic activities against both drug-sensitive and drug-resistant tumor cells. The drug resistance is mediated through either P-glycoprotein overexpression or beta-tubulin mutation mechanisms.

This study suggested that hydrophobicity and molar refractivity are the important parameters for the activity. They believed that two QSAR models may provide guidance in design and synthesis of cephalomannine derivatives that may have high biological activities [96].

Saiz-Urra et al. performed quantitative structure-activity relationship (QSAR) studies using a topological sub-structural molecular design (TOPS-MODE) approach of twenty-three clovane derivatives with their in vitro antifungal activity against the phytopathogenic fungus *Botrytis cinerea*. The most important parameters were the spectral moments weighted by bond dipole moment, hydrophobicity, and the combined dipolarity/polarizability Abraham molecular descriptor [97].

Little et al. performed docking and QSAR study of a series of artemisinin derivatives. The Heme molecule receptors were primarily selected to represent the changing binding and oxidation states of the molecule and relate these results to the observed biological activity [98].

Xu et al. conducted 2D and 3D QSAR studies on andrographolide derivatives as α-glucosidase inhibitors. They used 25 andrographolide derivatives as a training set and recommended that the combination of 2D and 3D QSAR models might be useful in predicting the alpha-glucosidase inhibiting activity of andrographolide derivatives [99].

Badawy et al. conducted QSAR studies of monoterpenes against the two-spotted spider mite, *Tetranychus urticae*. The QSAR model showed brilliant agreement between the predicted and experimentally measured toxicity parameter for the tested monoterpenes [100].

For the development of taxane analogues with improved anticancer activity and fewer side effects, Verma et al. performed QSAR modeling of taxane derivatives against colon cancer. The results of the study suggested that the steric and hydrophobic parameters of the substituents are

the two most important determinants for the activities of taxane analogues against colon cancers [101].

McGovern et al. constructed two CoMFA models of Salvinorin A analogs substituted at the C-2 position at the kappa-opioid receptor. They employed three alignment methods: a receptor-docked alignment derived from GOLD algorithm, the ligand-based alignment from FlexS algorithm, and a rigid realignment of the poses from the receptor-docked alignment. The first algorithm i.e. receptor-docked alignment produced statistically better results compared to either the FlexS alignment or the realignment. From the CoMFA contour maps, the binding mode of amine-containing Salvinorin A analogs was proposed and suggested that the beta-epimers (R-configuration) of protonated amines at the C-2 position have a higher affinity than the corresponding alpha-epimers (S-configuration) [102].

Hassan et al. proposed 3D-QSAR studies of eunicellin-based diterpenoids which showed significant anti-migratory and anti-invasive activities against prostate cancer in wound-healing and Cultrex invasion models. They created a valid 3D-QSAR model for guiding the design of potent eunicellin diterpenes cancer migration inhibitors [103].

Bharate et al. carried out QSAR study for a series of phloroglucinol-terpene adducts exhibiting anti-leishmanial activity and suggested that the lipophilic character (CLogP), isoelectric point, Haray index and Platt index play important role in anti-leishmanial activity [104].

A number of betulinic acid and betulin derivatives with anti-HIV-1 activities were used to perform 3D-QSAR studies by using CoMFA and CoMSIA. The analysis A and analysis B of this study were related to two activity indexes EC_{50} and TI (therapeutic index) respectively. The analysis A which was resulted from 41 molecules produced r_{cv}^2 values of 0.664 and

0.718, r^2 values of 0.979 and 0.955, respectively. The analysis B resulted from 41 molecules provided r_{cv}^2 values of 0.570 and 0.559, r^2 values of 0.938 and 0.933, respectively. The contour maps illustrated the regions in space where interactive fields may influence the activity. The results may be used for the design of potential betulinic acid and betulin derivatives with better anti-HIV-1 activity [105].

In order to predict the structural features responsible for α-glucosidase inhibitory activity, Moorthy et al. performed a QSAR analysis on a series of andrographolide derivatives. They used subdivided surface area, adjacency, surface volume and shape, partial charge descriptors and found a high correlation with the inhibitory activity [106].

Lan et al performed 3D-QSAR and molecular docking studies of betulinic acid derivatives that are responsible for the anti-HIV activity [107].

Wei et al. performed 3D-QSAR studies of a set of 43 natural sesquiterpene polyol esters with optimal narcotic or insecticidal activities. The 3D-QSARmodels suggested that the electronic effect governs the narcotic activities where as the combination of electrostatic and hydrophobic interactions are more influential in the insecticidal activities of the molecules [108].

Zhao et al. studied a QSAR study of a series of tanshinone compounds with cytotoxicity against murine leukemia cell lines P-388 using density functional theory. They used four indices: the maximum molecular electrostatic potential at the SAS surface, the average nucleophilic superdelocalizability, the dihedral between ring A and B and the net atomic charge of C (12). They constructed QSAR equation via multiple linear regression analysis and based on this model they designed compounds theoretically [109].

Bartalis et al. conducted QSAR studies to evaluate cucurbitacins (Cucs) liver protective activity in vitro against lipophilicity and ab initio descriptors [110].

Liang et al. performed receptor-based 3D-QSAR studies of 106 naturally occurring pentacyclic triterpenes as glycogen phosphorylase inhibitors. This study suggested that the elongated or bulky substitutions in C17 position and/or C2, C3 positions of pentacyclic triterpenes are favorable. They synthesized 56 compounds and evaluate their activity [111].

Maurya et al. performed QSAR modeling and docking studies of two triterpenoids ursolic acid and lupeol isolated from *Eucalyptus tereticornis* and *Gentiana kurroo*. This study suggested that both the triterpenoids show anti-inflammatory activity due to high binding affinity to human receptors viz., NF-kappaB p52 and may be considered as potential immunomodulatory drug-like molecules [112].

Tong et al. determined two QSAR models of monoterpenoids which have insecticidal potency on pest insects. This study suggested that the hydrophobicity and stability of monoterpenoid molecules were strongly involved in binding activities to the housefly GABA receptor [113].

Kalani et al. developed QSAR models for predicting the activities of ursolic acid analogs against human lung (A-549) and CNS (SF-295) cancer cell lines by a forward stepwise multiple linear regression method using a leave-one-out approach. This study indicated that the LUMO energy, ring count, and solvent-accessible surface area were strongly correlated with anticancer activity. Some ursolic acid analogs were semisynthesized based on QSAR results and tested in vitro against the human lung (A-549) and CNS (SF-295) cancer cell lines [114].

Abbasitabar et al. derived QSAR models for 179 analogues of artemisinin, a potent antimalarial agent. The reactive and partial equalization of orbital electronegativity descriptors represented the highest impact on the antimalarial activity [115].

Wang et al. performed CoMFA studies on 37 betulinic acid and betulin derivatives and their in vitro anti-cancer activity results against HT29 human colon cancer cells. The study provided a leave-one-out cross-validation q2 value of 0.722 and a non-cross-validation r^2 value of 0.974, which suggested that the model has good predictive ability (q2 > 0.2). The contour maps suggested that bulky and electron-donating groups at the C-28 site and a moderately bulky and electron withdrawing group near the C-3 site would improve the activity. Three betulin derivatives were designed and synthesized and their in vitro cytotoxicity was consistent with the predicted values. Thus the present topomer CoMFA model could guide the synthesis of new betulin derivatives with high anti-cancer activity [116].

Song et al. studied the interactions between low-toxicity terpenoid mosquito repellents and lactic acid at the HF and B3LYP level. This study suggests that the repellent-lactic acid complexes may play an important role [117].

Sousa et al. applied QSAR methodology to identify the most relevant molecular features of macrocyclic diterpenes with P-glycoprotein inhibitory activity. They developed a QSAR model for a set of 51 bioactive diterpenic compounds which includes lathyrane and jatrophane-type diterpenes and another model just for jatrophanes [118].

A QSAR study was performed of the antimalarial agent artemisinin and some of its derivatives using DFT based descriptors such as hardness, chemical potential, electrophilicity index, Fukui function, and local philicity. Multiple regression analysis was performed to construct QSAR

model using these descriptors against the chloroquine-resistant, mefloquine-sensitive Plasmodium falciparum W-2 clone [119].

Schomburg et al. investigated a QSAR model of natural sesquiterpene lactones as inhibitors of Myb-dependent gene expression. This QSAR approach was based on flexible alignment method [120].

Liao et al. in 2014 studied molecular interactions between terpenoid mosquito repellents and three typical human-secreted attractants, ammonia, 1-octen-3-ol, and formic acid. Relative energies, bond distances, and bond angles of the molecular interactions at HF level were used in the study in order to understand the relationship among mosquito repellents and attractants secreted by human hosts [121].

Foudah et al. conducted pharmacophore modeling and 3D-QSAR studies of sipholane triterpenoids as breast cancer migration and proliferation inhibitors [122].

Rudnitskaya et al. developed 3D-QSAR models of fifteen sesquiterpenoids, relating the hepatoprotection activity with molecular properties. Different chemical features such as shape, branching, symmetry, and presence of electronegative fragments can regulate the hepatoprotective activity of these sesquiterpenoids [123].

Trossini et al. constructed HQSAR models of 40 sesquiterpene lactones with activity against *T. brucei, T. cruzi, L. donovani and P. falciparum* (antiprotozoal activities) and also with their cytotoxicity. The differences between the most and least potent compounds were found from HQSAR contribution maps. This study also suggested as previous QSAR study that the presence of the α,β-unsatured carbonyl groups is fundamental to biological activity of sesquiterpene lactones [124].

Cheng et al. synthesized 21 novel sesquiterpenoids, trichodermin derivatives containing conjugated oxime ester and screened for in vitro antifungal activity. They performed QSAR analysis with these compounds and found that log P and hardness were two critical parameters for the biological activities [125].

Thanashankar et al. have performed the pharmacophore modeling and 3D-QSAR studies of a series of amino alkyl rupestonates (Rupestonic Acid) derivatives which inhibit H1N1, H3N2 and influenza B virus. In order to improve the biological activity of these compounds, a four point pharmacophore model with one acceptor and three hydrophobic regions was developed. On the basis of pharmacophore hypothesis, the 3D-QSAR model was constructed which provided an invaluable insight into structure activity correlation [126].

Appell et al. evaluated 35 trichothecenes using density functional theory at the 6-311+ G (d,P) level of theory. Type A and type B trichothecenes have distinct quantum-based differences including their frontier molecular orbital and natural bond orbital properties. QSAR models were constructed using one and two dimensional descriptors to describe cytotoxicity, phytotoxicity, and detection cross-reactivity. The important components of the models were topological indices and electronegativity [127].

Andrographolide, the labdane diterpene, isolated from *Andrographis paniculata*, has several pharmacological activities including immuno-stimulatory, cytotoxic, anti-inflammatory, anticancer, hypotensive, cardio-protective and anti-HIV. Mondal et al. synthesized a number of andrographolide derivatives and performed 2D QSAR study which indicates that steric effects and van der Waals interactions play major roles in the determination of antiproliferative activity

of these compounds. 3D QSAR study revealed that the benzyl substitution at N20 position may be important for higher steric interaction [128].

Artemisinin possesses anticancer activity through anti-angiogenic effects. Efferth et al. performed molecular docking of 52 artemisinin derivatives to vascular endothelial growth factor receptors (VEGFR1 VEGFR2), and VEGFA using Autodock4. They also performed QSAR study with these compounds [129].

Tiwari et al. performed QSAR study with a number of gymnemic acid analogues against PPARγ, a promising drug target for diabetes. In this study they found that chemical descriptors viz., dipole moment, electron affinity, dielectric energy, secondary amine group count and LogP correlated well with the activity and provides an insight into the therapeutics for diabetes mellitus [130].

A series of 30 compounds which are structurally related to geranyl acetone, nerolidol, farnesal, farnesol and farnesyl acetate, are potentially useful in fragrance compositions as antimicrobial agents and showed better or comparable activity to parent terpenoids. The generated pharmacophore models, obtained by 3D QSAR modeling, indicate significant steric factors which determine the antimicrobial activity of the compounds [131].

Essential oils and their constituents are known for their wide variety of biological activities such as antibacterial, antifungal, antiparasitic and antimycobacterial properties. Rivera-Chavira et al. was evaluated the descriptor of the molecular properties and the structural characteristics responsible for antimycobacterial activity of the tested compounds and developed QSAR models. These descriptors provide insight into the mechanisms of action of the active molecules and help to synthesize active compounds [132].

2.3. References

[1] S.B. Mahato, S. Sen, Advances in triterpenoid research, 1990-1994, Phytochemistry. 44 (1997) 1185-1236.

[2] I.H. Hall, K.H. Lee, C.O. Starnes, Y. Sumida, R.Y. Wu, T.G. Waddell, J.W. Cochran, K.G. Gerhart, Anti- inflammatory activity of sesquiterpene lactones and related compounds, J Pharm Sci. 68 (1979) 537-542.

[3] C.H. Chen, L.M. Yang, T.T. Lee, Y.C. Shen, D.C. Zhang, D.J. Pan, A.T. McPhail, D.R. McPhail, S.Y. Liu, D.H. Li, Antitumor agents-CLI. Bis(helenalinyl)glutarate and bis(isoalantodiol- B)glutarate, potent inhibitors of human DNA topoisomerase II, Bioorg Med Chem. 2 (1994) 137-145.

[4] K. Hayashi, H. Hayashi, N. Hiraoka, Y. Ikeshiro, Inhibitory activity of soyasaponin II on virus replication in vitro, Planta Med. 63 (1997) 102-105.

[5] M.C. Recio, R.M. Giner, L. Uriburu, S. Manez, M. Cerda, J.R. De la Fuente, J.L. Rios, In vivo activity of pseudoguaianolide sesquiterpene lactones in acute and chronic inflammation, Life Sci. 66 (2000) 2509-2518.

[6] C.L. Cantrell, S.G. Franzblau, N.H. Fischer, Antimycobacterial plant terpenoids, Planta Med. 67 (2001) 685-694.

[7] N. Osaki, T. Koyano, T. Kowithayakorn, M. Hayashi, K. Komiyama, M. Ishibashi, Sesquiterpenoids and plasmin-inhibitory flavonoids from Blumea balsamifera, J Nat Prod. 68 (2005) 447-449.

[8] D.B. Stierle, A.A. Stierle, T. Girtsman, K. McIntyre, J. Nichols, Caspase-1 and -3 inhibiting drimane sesquiterpenoids from the extremophilic fungus Penicillium solitum, J Nat Prod. 75 (2012) 262-266.

[9] Q.Y. Ma, Y.C. Chen, S.Z. Huang, Z.K. Guo, H.F. Dai, Y. Hua, Y.X. Zhao, Two new guaiane sesquiterpenoids from Daphne holosericea (Diels) Hamaya, Molecules. 19 (2014) 14266-14272.

[10] Q.Q. Tao, K. Ma, L. Bao, K. Wang, J.J. Han, J.X. Zhang, C.Y. Huang, H.W. Liu, New sesquiterpenoids from the edible mushroom Pleurotus cystidiosus and their inhibitory activity against α-glucosidase and PTP1B, Fitoterapia. 111 (2016) 29-35.

[11] K. Dimas, C. Demetzos, M. Marsellos, R. Sotiriadou, M. Malamas, D. Kokkinopoulos, Cytotoxic activity of labdane typed diterpenes against human leukemic cell lines in vitro, Planta Med. 64 (1998) 208-211.

[12] R. Batista, E. Chiari, A.B. de Oliveira, Trypanosomicidal kaurane diterpenes from Wedelia paludosa, Planta Med. 65 (1999) 283-284.

[13] M. Singh, M. Pal, R.P. Sharma, Biological activity of the labdane diterpenes, Planta Med. 65 (1999) 2-8.

[14] M. DellaGreca, A. Fiorentino, M. Isidori, P. Monaco, A. Zarrelli, Antialgal ent-labdane diterpenes from Ruppia maritime, Phytochemistry. 55 (2000) 909-913.

[15] R. Tanaka, H. Ohtsu, M. Iwamoto, T. Minami, H. Tokuda, H. Nishino, S. Matsunaga, A. Yoshitake, Cancer chemopreventive agents, labdane diterpenoids from the stem bark of Thuja standishii (Gord.) Carr, Cancer Lett. 161 (2000) 165-170.

[16] A. Navarro, B. de las Heras, A.M. Villar, Andalusol, a diterpenoid with anti-inflammatory activity from Siderits foetens Clemen, Z Naturforsch. 52c (1997) 844-849.

[17] M.A. Fernandez, M.P. Tornos, M.D. Garcia, B. de las Heras, A.M. Villar, M.T. Saenz, Anti-inflammatory activity of abietic acid, a diterpene isolated from Pimenta racemosa var. grissea, J Pharm Pharmacol. 53 (2001) 867-872.

[18] X. Chen, J. Ding, Y.M. Ye, J.S. Zhang, Bioactive abietane and seco-abietane diterpenoids from Salvia prionitis, J Nat Prod. 65 (2002) 1016-1020.

[19] M. Na, W.K. Oh, Y.H. Kim, X.F. Cai, S. Kim, B.Y. Kim, J.S. Ahn, Inhibition of protein tyrosine phosphatase 1B by diterpenoids isolated from Acanthopanax koreanum, Bioorg Med Chem Lett. 16 (2006) 3061-3064.

[20] A. Leverrier, M.T. Martin, C. Servy, J. Ouazzani, P. Retailleau, K. Awang, M.R. Mukhtar, F. Guéritte, M. Litaudon, Rearranged diterpenoids from the biotransformation of ent-trachyloban-18-oic acid by Rhizopus arrhizus, J Nat Prod. 73 (2010) 1121-1125.

[21] L. Yang, L. Qiao, C. Ji, D. Xie, N.B. Gong, Y. Lu, J. Zhang, J. Dai, S. Guo, Antidepressant abietane diterpenoids from chinese eaglewood, J Nat Prod. 76 (2013) 216-222.

[22] J.X. Zhao, C.P. Liu, W.Y. Qi, M.L. Han, Y.S. Han, M.A. Wainberg, J.M. Yue, Eurifoloids A-R, structurally diverse diterpenoids from Euphorbia neriifolia, J Nat Prod. 77 (2014) 2224-2233.

[23] G. Luo, Q.Ye, B. Du, F. Wang, G. Zhang, Y. Luo, Iridoid glucosides and diterpenoids from Caryopteris glutinosa, J Nat Prod. 79 (2016) 886-893.

[24] H. Nishino, K. Yoshioka, A. Iwashima, H. Takizawa, S. Konishi, H. Okamoto, H. Okabe, S. Shibata, H. Fujiki, T. Sugimura, Glycyrrhetic acid inhibits tumor promoting activity of teleocidin and 12-O-tetradecanoylphorbol-13-acetate in two-stage mouse skin carcinogenesis, Jpn J Cancer Res. 77 (1986) 33-38.

[25] H. Tokuda, H. Ohigashi, K. Koshimizu, Y. Ito, Inhibitory effects of ursolic and oleanolic acid on skin tumor promotion by 12-O-tetradecanoylphorbol-13-acetate, Cancer Lett. 33 (1986) 279-285.

[26] M.C. Recio, R.M. Giner, S. Manez, J.L. Rios, Structural requirements for the anti-inflammatory activity of natural triterpenoids, Planta Med. 61 (1995) 182-185.

[27] F. Soler, C. Poujade, M. Evers, J.C. Carry, Y. Henin, A. Bousseau, T. Huet, R. Pauwels, E. De Clercq, J.F. Mayaux, J.B. Le Pecq, N. Dereu, Betulinic acid derivatives: a new class of specific inhibitors of human immunodeficiency virus type 1 entry, J Med Chem. 39 (1996) 1069-1083.

[28] H. Ito, S. Onoue, Y. Miyake, T. Yoshida, Iridal-type triterpenoids with ichthyotoxic activity from Belamcanda chinensis, J Nat Prod. 62 (1999) 89-93.

[29] A. Rajic, G. Kweifio-Okai, T. Macrides, R.M. Sandeman, D.S. Chandler, G.M. Polya, Inhibition of serine proteases by anti-inflammatory triterpenoids, Planta Med. 66 (2000) 206-210.

[30] C.L. Cantrell, S.G. Franzblau, N.H. Fischer, Antimycobacterial plant terpenoids, Planta Med. 67 (2001) 685-694.

[31] T. Fujioka, Y. Kashiwada, R.E. Kilkuskie, L.M. Cosentino, L.M. Ballas, J.B. Jiang, W.P. Janzen, I.S. Chen, K.H. Lee, Anti-AIDS agents, 11. Betulinic acid and platanic acid as anti-HIV principles from Syzigium claviflorum, and the anti-HIV activity of structurally related triterpenoids, J Nat Prod. 57 (1994) 243-247.

[32] Z.Z. Ma, Y. Hano, T. Nomura, Y.J. Chen, Three new triterpenoids from Peganum nigellastrum, J Nat Prod. 63 (2000) 390-392.

[33] S. Wada, A. Iida, R. Tanaka, Screening of triterpenoids isolated from Phyllanthus flexuosus for DNA topoisomerase inhibitory activity, J Nat Prod. 64 (2001) 1545-1547.

[34] K. Yoshikawa, M. Inoue, Y. Matsumoto, C. Sakakibara, H. Miyataka, H. Matsumoto, S. Arihara, Lanostane triterpenoids and triterpene glycosides from the fruit body of

Fomitopsis pinicola and their inhibitory activity against COX-1 and COX-2, J Nat Prod. 68 (2005) 69-73.

[35] P.T. Thuong, C.H. Lee, T.T. Dao, P.H. Nguyen, W.G. Kim, S. J. Lee, W.K. Oh, Triterpenoids from the leaves of Diospyros kaki (persimmon) and their inhibitory effects on protein tyrosine phosphatase 1B, J Nat Prod. 71 (2008) 1775-1778.

[36] M. Litaudon, C. Jolly, C.L. Callonec, D. D. Cuong, P. Retailleau, O. Nosjean, V.H. Nguyen, B. Pfeiffer, J.A. Boutin, F. Guéritte, Cytotoxic pentacyclic triterpenoids from Combretum sundaicum and Lantana camara as inhibitors of Bcl-xL/BakBH3 domain peptide interaction, J Nat Prod. 72 (2009) 1314-1320.

[37] C. Chen, S. Qiang, L. Lou, W. Zhao, Cucurbitane type triterpenoids from the stems of Cucumis melo, J Nat Prod. 72 (2009) 824-829.

[38] P. Wang, S. Ownby, Z. Zhang, W. Yuan, S. Li, Cytotoxicity and inhibition of DNA topoisomerase I of polyhydroxylated triterpenoids and triterpenoid glycosides, Bioorg Med Chem Lett. 20 (2010) 2790-2796.

[39] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[40] J.D.S. Mpetga, Y. Shen, P. Tane, S.F. Li, H.P. He, H.K. Wabo, M. Tene, Y. Leng, X.J. Hao, Cycloartane and friedelane triterpenoids from the leaves of Caloncoba glauca and their evaluation for inhibition of 11β- hydroxysteroid dehydrogenases, J Nat Prod. 75 (2012) 599-604.

[41] L. Xiong, M. Zhu, C. Zhu, S. Lin, Y. Yang, J. Shi, Structure and bioassay of triterpenoids and steroids isolated from Sinocalamus affinis, J Nat Prod. 75 (2012) 1160-1166.

[42] C.L. Zhang, Y. Wang, Y.F. Liu, G. Ni, D. Liang, H. Luo, X.Y. Song, W.Q. Zhang, R.Y. Chen, N.H. Chen, D.Q. Yu, Iridal-type triterpenoids with neuroprotective activities from Iris tectorum, J Nat Prod. 77 (2014) 411-415.

[43] G.W. Wang, C. Lv, X. Fang, X.H. Tian, J. Ye, H.L. Li, L. Shan, Y.H. Shen, W.D. Zhang, Eight pairs of epimeric triterpenoids involving a characteristic spiro-E/F ring from Abies faxoniana, J Nat Prod. 78 (2015) 50-60.

[44] L.P. Hammett, Some relations between reaction rates and equilibrium constants, Chem Rev. 17 (1935) 125-136.

[45] L.P. Hammett, Physical Organic Chemistry, 2nd ed., McGraw-Hill, New York, 1970.

[46] R.W. Taft Jr., Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters[1], J Am Chem Soc. 74 (1952) 3120-3128.

[47] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients, Nature. 194 (1962) 178-180.

[48] C. Hansch, T. Fujita, ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure, J Am Chem Soc. 86 (1964) 1616-1626.

[49] C. Hansch, Quantitative approach to biochemical structure-activity relationships, Acc Chem Res. 2 (1969) 232-239.

[50] S.M. Free Jr., J.W. Wilson, A mathematical contribution to structure-activity studies, J Med Chem. 7(1964) 395-399.

[51] Z. Simon, Specific interactions. Intermolecular forces, steric requirements, and molecular size, Angew Chem Int Ed Eng. 13 (1974) 719-727.

[52] L.H. Hall, L.B. Kier, Structure-activity studies using valence molecular connectivity, J Pharm Sci. 66 (1977) 642-644.

[53] L.B. Kier, L.H. Hall, Molecular structure description: the electrotopological state, Academic Press, San Diego, CA, 1999.

[54] W. Tong, D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sleehan, Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor, J Chem Inf Comput Sci.38 (1998) 669-677.

[55] S.J. Cho, W. Zheng, A. Tropsha, Focus-2D: a new approach to the design of targeted combinatorial chemical libraries, Pac Symp Biocomput. (1998) 305-316.

[56] H. Gao, J. Bajorath, Comparison of binary and 2D QSAR analyses using inhibitors of human carbonic anhydrase II as a test case, J Mol Diversity. 4 (1998) 115-130.

[57] H. Gao, C. Williams, P. Labute, J. Bajorath, Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands, J Chem Inf Comput Sci. 39 (1999) 164-168.

[58] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J Am Chem Soc. 110 (1988) 5959-5967.

[59] H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in drug design, ligand-protein interactions and molecular similarity (Vol. 2) and recent advances (Vol. 3), Kluwer Academic Publishers, the Netherlands, 1997.

[60] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J Med Chem. 37 (1994) 4130-4146.

[61] D.D. Robinson, P.J. Winn, P.D. Lyne, W.G Richards, Self-organizing molecular field analysis: a tool for structure-activity studies, J Med Chem. 42 (1999) 573-583.

[62] H.L. Jiang, K.X. Chen, H.W. Wang, Y. Tang, J.Z. Chen, R.Y. Ji, 3D-QSAR study on ether and ester analogs of artemisinin with comparative molecular field analysis, Zhongguo Yao Li Xue Bao. 15 (1994) 481-487.

[63] J.Z. Chen, L.H. Hu, H.L. Jiang, J.D. Gu, W, L. Zhu, Z.L. Chen, K.X. Chen, R.Y. Ji, A 3D-QSAR study on ginkgolides and their analogues with comparative molecular field analysis, Bioorg Med Chem Lett. 8 (1998) 1291-1296.

[64] J.R. Woolfrey, M.A. Avery, A.M. Doweyko, 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 Des. 12 (1998) 165-181.

[65] M. Jalali-Heravi, M.H. Fatemi, Artificial neural network modeling of Kováts retention indices for noncyclic and monocyclic terpenes, J Chromatogr A. 915 (2001) 177-183.

[66] J.A. Grodnitzky, J.R. Coats, QSAR evaluation of monoterpenoids' insecticidal activity, J Agric Food Chem. 50 (2002) 4576-4580.

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

[68] T. Kuriyama, T.J. Schmidt, E. Okuyama, Y. Ozoe, Structure-activity relationships of seco-prezizaane terpenoids in gamma-aminobutyric acid receptors of houseflies and rats, Bioorg Med Chem. 10 (2002) 1873-1881.

[69] M.A. Avery, M. Alvim-Gaston, C.R. Rodrigues, E.J. Barreiro, F.E. Cohen, Y.A. Sabnis, J.R. Woolfrey, Structure-activity relationships of the antimalarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies, J Med Chem. 45 (2002) 292-303.

[70] T. Ghafourian, P. Zandasrar, H. Hamishekar, A. Nokhodchi, The effect of penetration enhancers on drug delivery through skin: a QSAR study, J Control Release. 99 (2004) 113-125.

[71] R. Guha, P.C. Jurs, Development of QSAR models to predict and interpret the biological activity of artemisinin analogues, J Chem Inf Comput Sci. 44 (2004) 1440-1449.

[72] T.J. Schmidt, M. Gurrath, Y. Ozoe, Structure-activity relationships of seco-prezizaaneandpicrotoxane/picrodendrane terpenoids by Quasar receptor-surface modeling, Bioorg Med Chem. 12 (2004) 4159-4167.

[73] F. Cortés-Selva, M. Campillo, C.P. Reyes, I.A. Jiménez, S. Castanys, I.L. Bazzocchi, L. Pardo, F. Gamarro, A.G. Ravelo, SAR studies of dihydro-beta-agarofuran sesquiterpenes as inhibitors of the multidrug-resistance phenotype in a Leishmania tropica line overexpressing a P-glycoprotein-like transporter, J Med Chem. 47 (2004) 576-587.

[74] B. Siedle, A.J. García-Piñeres, R. Murillo, J. Schulte-Mönting, V. Castro, P. Rüngeler, C.A. Klaas, F.B. Da Costa, W. Kisiel, I. Merfort, Quantitative structure-activity relationship of sesquiterpene lactones as inhibitors of the transcription factor NF-kappaB, J Med Chem. 47 (2004) 6042-6054.

[75] F.A. Macías, R.F. Velasco, D. Castellano, J.C. Galindo, Application of Hansch's model to guaianolide ester derivatives: a quantitative structure-activity relationship study, J Agric Food Chem. 53 (2005) 3530-3539.

[76] W. Zhu, G. Chen, L. Hu, X. Luo, C. Gui, C. Luo, C.M. Puah, K. Chen, J.H. Jiang, QSAR analyses on ginkgolides and their analogues using CoMFA, CoMSIA, and HQSAR, Bioorg Med Chem. 13 (2005) 313-322.

[77] M. Ishihara, H. Sakagami, W.K. Liu, Quantitative structure-cytotoxicity relationship analysis of betulinic acid and its derivatives by semi-empirical molecular-orbital method, Anticancer Res. 25 (2005) 3951-3955.

[78] S.L. Cunningham, A.R. Cunningham, B.W. Day, CoMFA, HQSAR and molecular docking studies of butitaxel analogues with beta-tubulin, J Mol Model. 11 (2005) 48-54.

[79] H. Parra-Delgado, C.M. Compadre, T. Ramírez-Apan, M.J. Muñoz-Fambuena, R.L. Compadre, P. Ostrosky-Wegman, M. Martínez-Vázquez, Synthesis and comparative molecular field analysis (CoMFA) of argentatin B derivatives as growth inhibitors of human cancer cell lines, Bioorg Med Chem. 14 (2006) 1889-1901.

[80] Y.H. Chung, S.C. Jang, S.J. Kim, N.D. Sung, 2D-QSAR and HQSAR on the inhibition activity of protein tyrosine phosphatase 1B with oleanolic acid analogues, J Appl Biol Chem. 50 (2007) 52-57.

[81] B.F. Rasulev, A.I. Saidkhodzhaev, S.S. Nazrullaev, K.S. Akhmedkhodzhaeva, Z.A. Khushbaktova, J. Leszczynski, Molecular modelling and QSAR analysis of the estrogenic activity of terpenoids isolated from Ferula plants, SAR QSAR Environ Res. 18 (2007) 663-673.

[82] B. Hemmateenejad, K. Javadnia, M. Elyasi, Quantitative structure-retention relationship for the Kovats retention indices of a large set of terpenes: a combined data splitting-feature selection strategy, Anal Chim Acta. 592 (2007) 72-81.

[83] M.T. Scotti, M.B. Fernandes, M.J. Ferreira, V.P. Emerenciano, Quantitative structure-activity relationship of sesquiterpene lactones with cytotoxic activity, Bioorg Med Chem. 15 (2007) 2927-2934.

[84] H.J. Chang, H.J. Kim, H.S. Chun, Quantitative structure-activity relationship (QSAR) for neuroprotective activity of terpenoids, Life Sci. 80 (2007) 835-841.

[85] C.P. Reyes, F. Muñoz-Martínez, I.R. Torrecillas, C.R. Mendoza, F. Gamarro, I.L. Bazzocchi, M.J. Núñez, L. Pardo, S. Castanys, M. Campillo, I.A.E Jiménez, Biological evaluation, structure-activity relationships, and three-dimensional quantitative structure-activity relationship studies of dihydro-beta-agarofuran sesquiterpenes as modulators of P-glycoprotein-dependent multidrug resistance, J Med Chem. 50 (2007) 4808-4817.

[86] L. Kang, C.W. Yap, P.F. Lim, Y.Z. Chen, P.C. Ho, Y.W. Chan, G.P. Wong, S.Y. Chan, Formulation development of transdermal dosage forms: quantitative structure-activity relationship model for predicting activities of terpenes that enhance drug penetration through human skin, J Control Release. 120 (2007) 211-219.

[87] W.N. Setzer, Non-intercalative triterpenoid inhibitors of topoisomerase II: A molecular docking study, TOBCJ. 1 (2008) 13-17.

[88] R.P. Verma, C. Hansch, Taxane analogues against breast cancer: a quantitative structure-activity relationship study, ChemMedChem. 3 (2008) 642-652.

[89] C. Hansch, R.P. Verma, Understanding tubulin/microtubule-taxane interactions: a quantitative structure-activity relationship study, Mol Pharm. 5 (2008) 151-161.

[90] F.J. Cardoso, A.F. de Figueiredo, M. da Silva Lobato, R.M. de Miranda, R.C. de Almeida, J.C. Pinheiro, A study on antimalarial artemisinin derivatives using MEP maps and multivariate QSAR, J Mol Model. 14 (2008) 39-48.

[91] Z. Wang, J. Song, Z. Han, Z. Jiang, W. Zheng, J. Chen, Z. Song, S. Shang, Quantitative structure-activity relationship of terpenoid aphid antifeedants, J Agric Food Chem. 56 (2008) 11361-11366.

[92] W.N. Setzer, A theoretical investigation of cytotoxic activity of celastroid triterpenoids, J Mol Model. 15 (2009) 197-201.

[93] R.P. Verma, C. Hansch, Taxane analogues against lung cancer: a quantitative structure-activity relationship study, Chem Biol Drug Des. 73 (2009) 627-636.

[94] W.E. Steinmetz, C.B. Rodarte, A. Lin, 3D QSAR study of the toxicity of trichothecene mycotoxins, Eur J Med Chem. 44 (2009) 4485-4489.

[95] T.J. Schmidt, A.M. Nour, S.A. Khalid, M. Kaiser, R. Brun, Quantitative structure-antiprotozoal activity relationships of sesquiterpene lactones, Molecules. 14 (2009) 2062-2076.

[96] C. Hansch, R.P. Verma, Overcoming tumor drug resistance with C2-modified 10-deacetyl-7-propionyl cephalomannines: a QSAR study, Mol Pharm. 6 (2009) 849-860.

[97] L. Saiz-Urra, J.C. Racero, A.J. Macías-Sáchez, R. Hernández-Galán, J.R. Hanson, M. Perez-Gonzalez, I.G. Collado, Synthesis and quantitative structure-antifungal activity relationships of clovane derivatives against Botrytis cinerea, J Agric Food Chem. 57 (2009) 2420-2428.

[98] R.J. Little, A.A. Pestano, Z. Parra, Modeling of peroxide activation in artemisinin derivatives by serial docking, J Mol Model. 15 (2009) 847-858.

[99] J. Xu, S. Huang, H. Luo, G. Li, J. Bao, S. Cai, Y. Wang, QSAR Studies on andrographolide derivatives as α-glucosidase inhibitors, Int J Mol Sci. 11 (2010) 880-895.

[100] M.E. Badawy, S.A. El-Arami, S.A. Abdelgaleil, Acaricidal and quantitative structure activity relationship of monoterpenes against the two-spotted spider mite, Tetranychus urticae, Exp Appl Acarol. 52 (2010) 261-274.

[101] R.P. Verma, C. Hansch, QSAR modeling of taxane analogues against colon cancer, Eur J Med Chem. 45 (2010) 1470-1477.

[102] D.L. McGovern, P.D. Mosier, B.L. Roth, R.B. Westkaemper, CoMFA analyses of C-2 position salvinorin A analogs at the kappa-opioid receptor provides insights into epimer selectivity, J Mol Graph Model. 28 (2010) 612-625.

[103] H.M. Hassan, A.Y. Elnagar, M.A. Khanfar, A.A. Sallam, R. Mohammed, L.A. Shaala, D.T. Youssef, M.S. Hifnawy, K.A. El Sayed, Design of semisynthetic analogues and 3D-QSAR study of eunicellin-based diterpenoids as prostate cancer migration and invasion inhibitors, Eur J Med Chem. 46 (2011) 1122-1130.

[104] S.B. Bharate, I.P. Singh, Quantitative structure-activity relationship study of phloroglucinol-terpene adducts as anti-leishmanial agents, Bioorg Med Chem Lett. 21 (2011) 4310-4315.

[105] P. Lan, W.N. Chen, P.H. Sun, W.M. Chen, 3D-QSAR studies on betulinic acid and betulin derivatives as anti-HIV-1 agents using CoMFA and CoMSIA, Med Chem Res. 20 (2011) 1247-1259.

[106] N.S. Moorthy, M.J. Ramos, P.A. Fernandes, Prediction of the relationship between the structural features of andrographolide derivatives and α-glucosidase inhibitory activity: a

quantitative structure-activity relationship (QSAR) study, J Enzyme Inhib Med Chem. 26 (2011) 78-87.

[107] P. Lan, W.N. Chen, Z.J. Huang, P.H. Sun, W.M. Chen, Understanding the structure-activity relationship of betulinic acid derivatives as anti-HIV-1 agents by using 3D-QSAR and docking, J Mol Model. 17 (2011) 1643-1659.

[108] S.P. Wei, Z.Q. Ji, H.X. Zhang, J.W. Zhang, Y.H. Wang, W.J. Wu, Isolation, biological evaluation and 3D-QSAR studies of insecticidal/narcotic sesquiterpene polyol esters, J Mol Model. 17 (2011) 681-693.

[109] M.L. Zhao, J.J. Yin, M.L. Li, Y. Xue, Y. Guo, QSAR study for cytotoxicity of diterpenoid tanshinones, Interdiscip Sci. 3 (2011) 121-127.

[110] J. Bartalis, F.T. Halaweish, In vitro and QSAR studies of cucurbitacins on HepG2 and HSC-T6 liver cell lines, Bioorg Med Chem. 19 (2011) 2757-2766.

[111] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[112] A. Maurya, F. Khan, D.U. Bawankule, D.K. Yadav, S.K. Srivastava, QSAR, docking and in vivo studies for immunomodulatory activity of isolated triterpenoids from Eucalyptus tereticornis and Gentiana kurroo, Eur J Pharm Sci. 47 (2012) 152-161.

[113] F. Tong, J.R. Coats, Quantitative structure-activity relationships of monoterpenoid binding activities to the housefly GABA receptor, Pest Manag Sci. 68 (2012) 1122-1129.

[114] K. Kalani, D.K. Yadav, F. Khan, S.K. Srivastava, N. Suri, Pharmacophore, QSAR, and ADME based semisynthesis and in vitro evaluation of ursolic acid analogs for anticancer activity, J Mol Model. 18 (2012) 3389-3413.

[115] F. Abbasitabar, V. Zare-Shahabadi, Development predictive QSAR models for artemisinin analogues by various feature selection methods: a comparative study, SAR QSAR Environ Res. 23 (2012) 1-15.

[116] W. Ding, M. Sun, S. Luo, T. Xu, Y. Cao, X. Yan, Y. Wang, A 3D QSAR study of betulinic acid derivatives as anti-tumor agents using topomer CoMFA: model building studies and experimental verification, Molecules. 18 (2013) 10228-10241.

[117] J. Song, Z. Wang, A. Findlater, Z. Han, Z. Jiang, J. Chen, W. Zheng, S. Hyde, Terpenoid mosquito repellents: A combined DFT and QSAR study, Bioorg Med Chem Lett. 23 (2013) 1245-1248.

[118] I.J. Sousa, M.J. Ferreira, J. Molnár, M.X. Fernandes, QSAR studies of macrocyclic diterpenes with P-glycoprotein inhibitory activity, Eur J Pharm Sci. 48 (2013) 542-553.

[119] S. Rajkhowa, I. Hussain, K.K. Hazarika, P. Sarmah, R.C. Deka, Quantitative structure-activity relationships of the antimalarial agent artemisinin and some of its derivatives - a DFT approach, Comb Chem High Throughput Screen. 16 (2013) 590-602.

[120] C. Schomburg, W. Schuehly, F.B. Da Costa, K.H. Klempnauer, T.J. Schmidt, Natural sesquiterpene lactones as inhibitors of Myb-dependent gene expression: structure-activity relationships, Eur J Med Chem. 63 (2013) 313-320.

[121] S. Liao, J. Song, Z. Wang, J. Chen, G. Fan, Z. Song, S. Shang, S. Chen, P. Wang, Molecular interactions between terpenoid mosquito repellents and human-secreted attractants, Bioorg Med Chem Lett. 24 (2014) 773-779.

[122] A.I. Foudah, A.A. Sallam, M.R. Akl, K.A. El Sayed, Optimization, pharmacophore modeling and 3D-QSAR studies of sipholanes as breast cancer migration and proliferation inhibitors, Eur J Med Chem. 73 (2014) 310-324.

[123] J. Vinholesa, A. Rudnitskayab, P. Gonçalvesc, F. Martelc, M.A. Coimbraa, S.M. Rocha, Hepatoprotection of sesquiterpenoids: a quantitative structure–activity relationship (QSAR) approach, Food Chem. 146 (2014) 78-84.

[124] H.G. Trossini, V.G. Maltarollo, T.J. Schmidt, Hologram QSAR studies of antiprotozoal activities of sesquiterpene lactones, Molecules. 19 (2014) 10546-10562.

[125] J.L. Cheng, M. Zheng, T.T. Yao, X.L. Li, J.H. Zhao, M. Xia, G.N. Zhu, Synthesis, antifungal activity, and QSAR study of novel trichodermin derivatives, J Asian Nat Prod. 17 (2015) 47-55.

[126] K. Muthusamy, P. Kirubakaran, G. Krishnasamy, R.R. Thanashankar, Computational insights into the inhibition of influenza viruses by rupestonic acid derivatives: pharmacophore modeling, 3D-QSAR, CoMFA and COMSIA studies, Comb Chem High Throughput Screen. 18 (2015) 63-74.

[127] M. Appell, W.B. Bosma, Assessment of the electronic structure and properties of trichothecene toxins using density functional theory, J Hazard Mater. 288 (2015) 113-123.

[128] A. Hazra, C. Mondal, D. Chakraborty, A.K. Halder, Y.P. Bharitkar, S.K. Mondal, S. Banerjee, T. Jha, N.B. Mondal, Towards the development of anticancer drugs from andrographolide: semisynthesis, bioevaluation, QSAR analysis and pharmacokinetic studies, Curr Top Med Chem. 15 (2015) 1013-1026.

[129] M.E. Saeed, O. Kadioglu, E.J. Seo, H.J. Greten, R. Brenk, T. Efferth, Quantitative structure-activity relationship and molecular docking of artemisinin derivatives to vascular endothelial growth factor receptor 1, Anticancer Res. 35 (2015) 1929-1934.

[130] P. Tiwari, P. Sharma, F. Khan, N.S. Sangwan, B.N. Mishra, R.S. Sangwan, Structure Activity Relationship Studies of Gymnemic Acid Analogues for Antidiabetic Activity Targeting PPARγ, Curr Comput Aided Drug Des. 11 (2015) 57-71.

[131] R. Bonikowski, P. Świtakowska, M. Sienkiewicz, M. Zakłos-Szyda, Selected compounds structurally related to acyclic sesquiterpenoids and their antibacterial and cytotoxic activity, Molecules. 20 (2015) 11272-11296.

[132] S. Andrade-Ochoa, G.V. Nevárez-Moorillón, L.E. Sánchez-Torres, M. Villanueva-García, B.E. Sánchez-Ramírez, L.M. Rodríguez-Valdez, B.E. Rivera-Chavira, Quantitative structure-activity relationship of molecules constituent of different essential oils with antimycobacterial activity against Mycobacterium tuberculosis and Mycobacterium bovis, BMC Complement Altern Med. 15 (2015) 332.

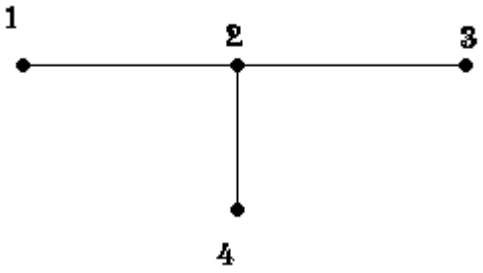
CHAPTER 3

Methodology

This chapter describes various classical as well as quantum indices and various physicochemical parameters in terms of which we would be presenting our analysis.

3.1. Topological indexes

A topological index also known as connectivity index is a numerical parameter that characterized molecular structure using graph theoretic formalism. The molecular graph is generally represented as G =(V,E) where V is a set of vertexes which represents the set of atoms in a molecule and E is a unordered pairs of elements of the set V which symbolizes covalent bonds between adjacent atoms. Most of the topological indices are derived from adjacency matrix A(G) and the distance matrix D(G) of the graph G. The hydrogen suppressed graph of isobutene is given below (Figure 3.1):



G

Figure 3.1. Hydrogen suppressed graph of isobutene

$$\text{Adjacency matrix A(G)} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} (1) & (2) & (3) & (4) \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$\text{Distance matrix D(G)} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} (1) & (2) & (3) & (4) \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{bmatrix} \end{array}$$

3.1.1. Wiener index

This was the first index based on graph theory to model the boiling point of hydrocarbons. The Wiener index (W) is calculated as half-sum of all the elements d_{ij} of the distance matrix [1]:

$$W = \frac{1}{2}\sum_{ij} d_{ij} = \sum_h h.g_h \quad \ldots\ldots. (3.1)$$

Where g_h is the number of unordered pairs of vertices whose distance is h.

3.1.2. Harary index

Harary index (H) is derived from the reciprocal of the distance matrix. It measures the molecular compactness as it is increases with increasing molecular size and branching [2].

$$H = \frac{1}{2}\sum_{ij} d_{ij}^{-1} \quad \ldots\ldots. (3.2)$$

Here d_{ij} is equal to distance between vertices v_i and v_j in G.

3.1.3. Randić connectivity index

The degree of the ith vertex (δ_i) is calculated as the sum of all entries in the ith row of the adjacency matrix with n vertices:

$$\delta_i = \sum_{j=1}^{n} a_{ij} \quad \ldots\ldots. (3.3)$$

Zero order connectivity index ($^0\chi$) is defined as [3]

$$^0\chi = \sum_i (\delta_i)^{-1/2} \quad \ldots\ldots. (3.4)$$

Randić connectivity index ($^1\chi$) is defined as [4]

$$^1\chi = \sum_{all\ edges} (\delta_i \delta_j)^{-1/2} \quad \ldots\ldots. (3.5)$$

A generalized connectivity index ($^h\chi$) can be defined as [3]:

$$^h\chi = \sum (\delta_{v_o} \delta_{v_1} \ldots \delta_{v_h})^{-1/2} \quad \ldots\ldots.. (3.6)$$

Where the summation is taken over all possible path of lengths 0,1, …,h.

3.1.4. Information-theoretic topological indices

Basak et al, developed information theoretic indices which take into account all atoms including hydrogens in the constitutional formula [5]. There are three types of informational indices such as IC (mean information content), CIC (complementary information content) and SIC (structural information content). An appropriate set A of n-elements is derived from a molecular graph G depending on various classes of atoms in their topological neighborhood. The set A is partitioned into equivalence classes Ai of order n_i (i=1, 2,… h; $\Sigma_i n_i$=n). A probability distribution is then assigned to the set of equivalence classes:

$A1, A2, \ldots, A_h$

$p_1, p_2, \ldots, p_h$

Where $p_i = n_i/n$, n_i and n are the cardinalities of Ai and A respectively.

On the basis of Shannon information theory, the mean information content is defined as [6]:

$$IC = -\sum_{i=1}^{h} p_i \, log_2 p_i \quad \ldots\ldots. (3.7)$$

The binary logarithm is taken to measure the information content in bites. The total information content is then n times IC.

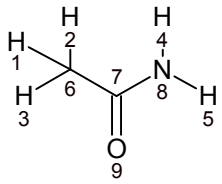
The division of atoms into different order of neighborhoods depends upon the coordination sphere taken into account [7, 8]. This leads to the indices of different order r.

$$IC_r = -\sum_{i} p_i \, log_2 p_i \quad \ldots\ldots. (3.8)$$

$$SIC_r = \frac{IC_r}{log_2 n} \quad \ldots\ldots.. (3.9)$$

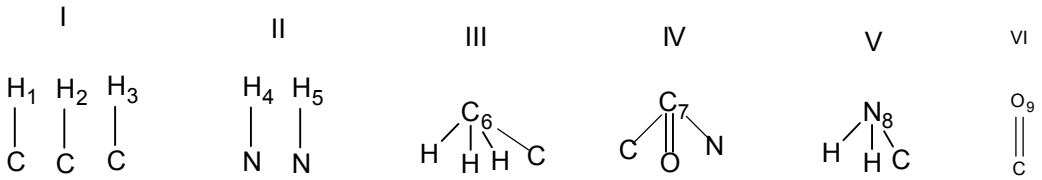
$$CIC_r = log_2 n - IC_r \quad \ldots\ldots.. (3.10)$$

For the equation (3.8), (3.9) and (3.10) the summation spans the range from i=1 to i=r, where r = 0, 1, 2… ρ, ρ is the radius of the molecular graph G and n is the total number of vertices of the graph i.e. total number of atoms in the molecule. Figure 3.2 gives a sample calculation of IC_1, SIC_1 and CIC_1.



Acetamide

First order neighbors:



	I	II	III	IV	V	VI
Subset :	$(H_1\text{-}H_3)$	$(H_4\text{-}H_5)$	C_6	C_7	N_8	O_9
Probability :	3/9	2/9	1/9	1/9	1/9	1/9

$$IC_1 = 4 * \frac{1}{9} * log_2 9 + \frac{2}{9} * log_2 \frac{9}{2} + \frac{3}{9} * log_2 \frac{9}{3} = 2.419\ bits$$

$$SIC_1 = \frac{IC_1}{log_2 9} = 0.763\ bits$$

$$CIC_1 = log_2 9 - IC_1 = 0.751\ bits$$

Figure 3.2. Labeled graph of acetamide and sample calculation of IC_1, SIC_1 and CIC_1

3.2. Molar refraction

H. A. Lorenz and L. V. Lorentz independently deduced a relation between density (d) and refractive index (n) of a substance [9].

$$R_s = \frac{n^2-1}{n^2+2} * \frac{1}{d} \quad \text{.....(3.11)}$$

The Rs is called the specific refraction of the substance and it is independent of temperature. Multiplying both side of equation (3.11) by mol. wt. M of the substance, we get molar refraction (R_M)

$$R_M = M * R_s = \frac{n^2-1}{n^2+2} * \frac{M}{d} \quad \text{..... (3.12)}$$

Here M/d is the molar volume and the molar refraction is a unit of volume since refractive index (n) is a dimensionless quantity. The molar refraction depends on the number and nature of atoms present and binding between atoms. However it is almost independent of pressure, temperature and the state of aggregation of the substance. Hence the molar refraction is partly additive and partly and partly constitutive. Table 3.1 shows molar refraction (R_M) of some atoms and structures.

Table 3.1. Molar refraction (R_M) at 589 nm, (cm^3mol^{-1})

H	1.100	O (carbonyl)	2.211	N (Tertiary amines)	2.840
C	2.418	O (ether)	1.644	Double bond (C=C)	1.733
Cl	5.967	O (hydroxyl)	1.522	Triple bond (C≡C)	2.398
Br	8.865	N (Primary amines)	2.322		
I	13.900	N (Secondary amines)	2.499		

3.3. Molar volume

The molar volume (M_V) may be defined as the volume of a gm mole of a substance at a given temperature and pressure [9]. By definition

$$M_V = \frac{M}{d} \ldots\ldots. (3.13)$$

Molar volume (M_V) can be calculated from additive increments. The additive atomic increments were achieved using a database of density (d) and calculated mol. wt. M.

3.4. Solvent accessible surface area

Molecules are often represented as a set of overlapping spheres of the constituent atoms. Lee and Richards defined solvent accessible surface and solvent exclude surface as [10]:

"The solvent accessible surface (SAS) is traced out by the center of the probe representing a solvent molecule. The solvent excluded surface (SES) is the topological boundary of the union of all possible probes which do not overlap with the molecule".

Accurate molecular surface areas are calculated using formulae given by Connolly [11] and these surface computations are based on the use of the reduced surface introduced by Sanner [12]. The algorithms which have been implemented in MSMS program are (i) computation of the reduced surface of a molecule (ii) analytical representation of the solvent excluded surface which may be self intersecting (iii) removing of all self-intersecting parts (iv) the last algorithm produces a triangulation of the SES [13].

3.5. Partition coefficient

The partition coefficient is defined as the ratio of concentrations of unionized solute between two immiscible solvents at equilibrium. Normally one of the solvent is water and the second one is octanol. Partition coefficient (P) is normally calculated in the form of its logarithm to base 10, because it ranges from 10^{-4} to 10^{8}.

$$logP = log \frac{(concentration\ of\ solute\ in\ octanol)}{(concentration\ of\ solute\ in\ water)} \quad \ldots\ldots\ldots (3.14)$$

Partition coefficient measures the hydrophilicity or hydrophobicity of a compound. The hydrophobic drugs are localized in the hydrophobic environment such as lipid bilayers of cells while hydrophilic drugs are distributed in the hydrophilic environment, such as blood serum [14]. LogP can be determined experimentally or predicted from structural data. The standard experimental procedure for logP estimation is the shake flask method, ranging from -2 to 4 logP values. High performance liquid chromatography (HPLC) may be used for more hydrophobic compounds ranging from 0 to 6 logP values. However logP values can be calculated computationally by five major methods: substituent methods, fragments methods, methods based on atomic contribution and/or surface areas, methods based on molecular properties, and, finally, methods based on solvatochromic parameters [15].

3.6. Quantum chemical descriptors

Quantum chemical techniques are generally used to get accurate molecular properties such as energy of the highest occupied molecular orbital (E_{HOMO}), energy of the lowest unoccupied molecular orbital (E_{LUMO}), dipole moment (μ), electronegativity (χ) etc. The three main

approaches to calculating molecular properties are semi-empirical methods, the density functional method and ab-initio method.

The starting point of any quantum chemical discussion is the time-independent Schrödinger equation which is often written in a compact form as,

$$H\psi=E\psi \quad \ldots\ldots. (3.15)$$

Where H is the Hamiltonian operator corresponding to the total energy of the system i.e. sums of kinetic and potential energies, ψ is the wave function and E is the energy of the molecule. A number of solutions exist for equation (3.15) and the solution of lowest energy represents the ground state. The equation (3.15) is an eigenvalue equation.

The important idea of Hartree–Fock theory is to consider the wavefunction as a series of molecular orbitals with differing electronic occupations and one of these sets of molecular orbitals will be the lowest energy ground state. i.e.

$$\psi = \varphi 1 \varphi 2 \varphi 3 \ \ldots \varphi n \quad \ldots\ldots. (3.16)$$

$$H(\varphi 1 \varphi 2 \varphi 3 \ldots \varphi n) = E(\varphi 1 \varphi 2 \varphi 3 \ldots \varphi n) \quad \ldots\ldots. (3.17)$$

where φi is the ith molecular orbital. Hartree–Fock theory uses the so-called basis functions which are one-electron mathematical functions representing the atomic orbitals. Both semi-empirical and density functional methods make use of basis functions which is called basis sets. The incomplete treatment of exchange–correlation effects is the major drawback in the Hartree–Fock formalization when evaluating the energy of the wavefunction.

The time consuming steps in Hartree–Fock formalization is the manipulation of the mathematical representations of the molecular orbitals. In contrast, semi-empirical AM1 method (Austin model

1) [16] deals only with the valence electrons, thus reducing the computation time. Again the use of parameterized functions for some of the terms in the Hamiltonian reduces the computational time in AM1 method. These parameterized functions are derived using experimental data. In 1989, Stewart gives PM3 method (parametric method 3) [17-19] in which one-center electron repulsion integrals are taken as parameters to be optimized rather than being found from atomic spectra data.

In density functional theory (DFT), the molecular electronic energy is calculated from the molecular electron probability density. DFT provides a more complete electronic structure description than that from Hartree–Fock theory and is also more complete than semi-empirical methods. The commonly used functional in DFT calculation is B3LYP [20, 21].

An ab initio calculation uses Hamiltonian with complete representation of all nonrelativistic interactions between the nuclei and electrons in a molecule [22]. Hence ab initio calculations are limited by the types of atoms and size of molecules [23].

HOMO and LUMO energies are important quantum chemical descriptors that play a major role in governing many chemical reactions and determining electronic band gaps in solids. The energies of HOMO and LUMO are directly related to the ionization potential and electron affinity respectively. HOMO energy characterizes the susceptibility of the molecule toward attack by electrophiles while LUMO energy characterizes the susceptibility of the molecule toward attack by nucleophiles [24]. The gap energy i.e. energy difference between HOMO and LUMO is an important parameter used in QSAR study as it describes the stability of molecules [25].

Electronegativity is defined as the negative of the chemical potential (δ) i.e. negative of the partial derivative of energy (E) with respect to the number of electrons (N) at constant external potential (V) of an atomic or molecular system [26].

$$\chi = -\delta = -(\frac{\partial E}{\partial N})_V \quad \ldots\ldots\ldots (3.18)$$

By combining the work of Iczkowski and Margrave [27] with equation (3.18) and assuming a quadratic relationship between E and N, we get

$$\chi_{koopmans} = \frac{E_{HOMO}+E_{LUMO}}{2} \quad \ldots\ldots\ldots (3.19)$$

The polarity of the molecule is often described by the term dipole moment and it is an important parameter used in QSAR study. The classical expression for the electric dipole moment (μ) of a set of discrete charges Q_i is given by [28]

$$\mu = \sum_i Q_i r_i \quad \ldots\ldots (3.20)$$

where r_i is the position vector from the origin of the *i*th charge.

The quantum mechanical quantity that corresponds to the dipole moment (μ) of the system in the absence of an applied electric field is given by [28]

$$\mu = \int \psi^{(0)*} \hat{\mu} \psi^{(0)} d\tau \quad \ldots\ldots\ldots (3.21)$$

where $\psi^{(0)}$ is the unperturbed wave function and $\hat{\mu}$ is the electric dipole moment operator.

3.7. Molar entropy

Statistical mechanics provides molar entropy of an ideal gas as the sum of translational, rotational, vibrational and electronic contributions. The contribution of translational term

depends only on the molar mass of the gas. The rotational term depends on the principal moments of inertia and symmetry number. The vibrational contribution depends on the molecular vibrational frequencies and the electronic contribution depends on the ground state electronic degeneracy and in a few cases on the energies of any low-lying electronic states [29].

3.8. Molecular docking

Molecular docking is an important tool for drug discovery in which the interaction between a ligand and a protein at the atomic level are noticed to characterize the behavior of ligand in the protein binding site [30]. Molecular docking involves two main steps: prediction of the ligand pose in the active site of the protein and assessment of the binding affinity. These two steps are related to the sampling methods and scoring schemes respectively.

Examples of some popular protein-ligand docking systems include AutoDock [31], GOLD [32], DOCK [33], GLIDE [34], ICM [35] and FlexX [36]. In our case, the molecular docking simulation was carried out using the Autodock 4.2.

New released AutoDock has three search methods: simulated annealing, Monte Carlo simulated annealing (genetic algorithm) and the Lamarckian genetic algorithm and the Lamarckian genetic algorithm is the most efficient, and reliable. AutoDock predicts the binding free energies from an empirical binding free energy force field which has been calibrated using a large set of diverse protein–ligand complexes.

The free energy of binding is estimated in two steps. At the beginning ligand and protein have unbound conformation. The first step measures the intramolecular energetics of the transition from these unbound states to the bound conformation for each of ligand and protein separately. The second step calculates the intermolecular energetics of combining the ligand and protein

together into the bound complex. The force field contains six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} - \Delta S_{conf}) \text{ ….. (3.22)}$$

where L and P refer to the ligand and protein respectively in a protein-ligand complex. In the unbound state ligand and protein are sufficiently distant from one another i.e. $V_{unbound}^{P-L} = 0$. As we did not allow motion in a protein, the difference of intramolecular energy between bound and unbound state of the protein is zero.

For two atoms i, j, the pair-wise atomic terms include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

$$V = W_{vdw} \sum_{ij} (\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}) + W_{hbond} \sum_{ij} E(t) (\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}) + W_{elec} \sum_{ij} \frac{q_i q_j}{\in(r_{ij}) r_{ij}} + W_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{-(\frac{r_{ij}^2}{2\sigma^2})} \text{ …… (3.23)}$$

W are weighted factors for calibrate the empirical free energy based on a set of experimentally characterized complexes. Parameters A and B were collected from the Amber force field [37] and parameters C and D are assigned to give a maximal well depth of 5 kcal/mol at 1.9 Å for hydrogen bonds with oxygen and nitrogen, and a depth of 1 kcal/mol at 2.5 Å for hydrogen bonds with sulpher. The first term is a 12-6 dispersion/repulsion term and the second term is a directional hydrogen bond term on a 10/12 potential where E(t) is a directional weight based on the angle, *t*, between the probe and the target atom. The third term is a screened Coulombic electrostatic potential. The calculation of the final term i.e. desolvation potential was most challenging and is based on the volume (V) of the atoms surrounding a given atom being

weighted by a solvation parameter (S) and a distance based exponential term . The distance weighting factor σ is set to 3.5 Å.

The loss of torsional entropy upon binding (ΔSconf) is directly proportional to the number of rotatable bonds in the molecule (Ntors):

$$\Delta S_{conf} = W_{conf} N_{tors} \quad \ldots\ldots\ldots \quad (3.24)$$

Rotatable bonds in the molecule contain all torsional degrees of freedom, including rotation of polar hydrogen atoms on hydroxyl groups and the like [31, 38].

3.9. Regression analysis

The statistical method to find out mathematical models that depicts relationships between two or many variables and the use of these relationships thus modeled for the purpose of prediction and other statistical inferences [39]. Historically, the word "regression" was first used by sir Francis Galton, who studied the parent and son height relationship. He published the results of his studies in a paper "Regression toward mediocrity in hereditary stature". Today the word regression is used in many areas of scientific investigationswithout any reference to biostatistic [40, 41].

Linear relationship between two variables is represented by a straight line. The line of average relationship is another name for a regression line. When a regression equation is to be specified, n paired observations are plotted, setting the vertical scale for dependent variable Y and horizontal scale for independent variable X. This diagram is called scatter diagram. The scatter diagram may be considered as a basis of deciding the type of regression equation, suited for the relationship between the two variables, Y and X. In natural science, social science and

economics, the number of parameters necessarily is not confined to only two variables. A large number of domains exist where involvement of more than two variables would be dictated by the problem. In those areas of study, we often need to give actual relationship between three or more variables. For such domains, multivariate regression and correlation are important tools. If we want to establish the relationship between dependent and independent variables, a mathematical equation can be formulated for tackling the situation. The equation pertaining to such a relationship may be of various types. But here we will only deal with a linear relationship which represents a plane according to the number of variables involved.

Let a mathematical model with dependent variable Y and k independent variables $X_1, X_2, \ldots\ldots X_k$, be,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + - - - - + \beta_k X_k + e \quad \ldots\ldots\ldots (3.25)$$

This type of regression equation is also known as multiple regression equation or prediction equation, where Y is predictant and $X_1, X_2, \ldots X_k$ are predictors. e is the error in formulating the linear model. This error is distributed normally with mean 0 and variance σ^2 , i.e. e $\approx N(0, \sigma^2)$. We have to estimate the parameter $\beta_0, \beta_1, \beta_3, \ldots \beta_k$ on the basis of n sample observations in which each observation is (k+1)-touple, n composite sample observations can be presented in the following format.

Composite observation no.	Variables		
	Y	X_1	X_2……..X_3……..X_k
1	Y1	x11	x21….xj1…….xk1
2	Y2	x12	x22….xj2…….xk3
.			
.			
I	Yi	x1i	x2i….xji…….xki
.			
.			
n	Yn	x1n	x2n….xjn…….xkn
Total	$\sum_i yi$	$\sum_i yi$	$\sum_i x1i$ $\sum_i xji$ xki

Estimation of β's by least squre method: For the i-th-touple the regression model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + e_i \quad \ldots\ldots\ldots (3.26)$$

Thus

$$e_i^2 = (yi - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \cdots - \beta_k x_{ki})^2$$

Taking the sum overall n-touples, we obtain

$$\sum_i e_i^2 = \sum_i (yi - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \cdots - \beta_k x_{ki})^2$$

For minimization, we have considered the square of errors summed over all observation. Let $\sum_i e_i^2 = Q$. To minimize Q, the overall squared error, we partially differentiate Q with respect to $\beta_0, \beta_1, \beta_3, \ldots \beta_k$ respectively and equate them to zero. Let the estimated values of $\beta_0, \beta_1, \beta_3, \ldots \beta_k$ be b_0, b, b_3. In this way we get (k+1) normal equations in (k+1) unknowns. Solving

these equations, we get the expressions for b_0, b, b_3 in terms of observed values. By substituting these estimates, we obtained the estimated equation.

Normal equations are

$$\sum_i y_i = \sum_1 b_0 + b_i \sum_i x_{1i} + b_2 \sum_i x_{2i} + \ldots + b_k \sum_i x_{ki}$$

$$\sum_i x_{1i} y_i = b_0 \sum_i x_{1i} + b_i \sum_i x_{1i}^2 + b_2 \sum_i x_{1i} x_{2i} + \ldots + b_k \sum_i x_{1i} x_{ki}$$

$$\sum_i x_{2i} y_i = b_0 \sum_i x_{2i} + b_i \sum_i x_{1i} x_{2i} + b_2 \sum_i x_{2i}^2 + \ldots + b_k \sum_i x_{2i} x_{ki}$$

……

$$\sum_i x_{ki} y_i = b_0 \sum_i x_{ki} + b_i \sum_i x_{1i} x_{ki} + b_2 \sum_i x_{2i} x_{ki} + \ldots + b_k \sum_i x_{ki}^2 \quad \ldots\ldots\ldots (3.27)$$

where i=1,2,3,…n.

If we replace the observed data in the form of vectors and matrices, the set of normal equations given by (3.26) can be written as follows.

$$Y_{nX1} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix} \quad ; \qquad \beta_{(k+1)X1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ . \\ . \\ . \\ \beta_n \end{bmatrix} ; \qquad B_{(k+)X1} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ . \\ . \\ . \\ b_n \end{bmatrix}$$

$$\text{And } X_{nX(k+1)} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{k1} \\ 1 & x_{12} & x_{21} & x_{k1} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & x_{1n} & x_{2n} & x_{kn} \end{bmatrix}$$

And $e' = (e_1, e_2 \ldots e_n)$

The regressions for i= 1,2,…n in matrix notation are,

$Y = X\beta + e$ …… (3.28)

And its estimated equation is

$$Y = XB$$

Whereas the set of normal equations (3.28) in matrix notation is

$$X'Y = X'XB \quad \ldots\ldots (3.29)$$

Or $B = (X'X)^{-1}X'Y$ …… (3.30)

Provided $(X'X)$ is a nonsingular matrix.

$X'Y$ is the left hand side of (3.29) will be as

$$X'X = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2j} \ldots \ldots \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{1i} \ldots \sum x_{1i}x_{1i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \ldots\ldots x_{2i}x_{ki} \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ \sum x_{1i} & \sum x_{1i}\, x_{ki} & \sum x_{2i}\, x_{ki} \ldots\ldots \sum x_{ki}^2 \end{bmatrix}$$

The above matrix is of the order (k+1)X(k+1).

From the equation set we can obtain (3.31),

$\sum_i b_0 = \sum_i y_i - b_1 \sum_i x_{1i} - b_2 \sum_i x_{2i} - \ldots - b_k \sum_i x_{ki}$

Since $\sum_i x_{ij} = n\bar{x}_j$ for j=1,2,…,k

And $\sum_i y_{ij} = n\bar{y}_j$,

$nb_0 = n(\bar{y} - b_1\overline{x_1} - b_2\overline{x_2} - \ ... \ - \ b_k\overline{x_k})$

Or $\quad b_0 = \bar{y} - b_1\overline{x_1} - b_2\overline{x_2} - \ ... \ - \ b_k\overline{x_k}$ …… (3.31)

Substituting the value of $\sum_i b_0$ in the second equation of the set of normal equations, we get

$\sum_i x_{1i}y_i = n\overline{x_1}\,(\bar{y} - b_1\overline{x_1} - b_2\overline{x_2} - \ ... \ - \ b_k\overline{x_k}) + b_1\sum_i x_{1i}^2 + \ b_2\ \sum_i x_{1i}x_{2i} + \ b_k\ \sum_i x_{1i}x_{ki}$

Or,

$\sum_i x_{1i}y_i - n\overline{x_1}\,y \ = -\,nb_1\overline{x^2{}_1} - nb_2\overline{x_1x_2} - \ ... \ - \ nb_k\overline{x_1x_k} + b_1\sum_i x_{1i}^2 + \ b_2\ \sum_i x_{1i}x_{2i} +$

$b_k\ \sum_i x_{1i}x_{ki}$

$\sum_i(x_{1i} - \bar{x}_i)(y_i - \bar{y}) = \ b_1\sum_i(x_{1i} - \bar{x}_1)^2 + \ b_2\sum_i(x_{1i} - \bar{x}_1)(\ x_{2i} - \bar{x}_2) + \ldots + b_k\sum_i(x_{1i} -$

$\bar{x}_1)(\ x_{ki} - \bar{x}_k)$

Suppose $x_{ji} - x_j = u_{ji}, y_i - \bar{y} = v_i$

For i=1,2,…,n and j=1,2,…,k.

The above equation is,

$\sum_i u_{1i}v_i = \ b_1\sum_i u_{1i}^2 + \ b_2\sum u_{1i}u_{2i} + \cdots + b_k\sum u_{1i}u_{ki}$

Similarly the other equations of the set are

$$\sum_i u_{2i}v_i = b_2\sum_i u_{2i}^2 + b_1\sum u_{1i}u_{2i} + \cdots + b_k\sum u_{ki}u_{ki}$$

……

$$\sum_i u_{ki}v_i = b_k\sum_i u_{ki}^2 + b_2\sum u_{1i}u_{ki} + \cdots + b_1\sum u_{1i}u_{ki} \quad \ldots\ldots\ldots (3.32)$$

In the matrix notation the set of normal equations can be represented by

$$\begin{bmatrix} \sum_i u_{1i}v_i \\ \sum_i u_{2i}v_i \\ . \\ . \\ . \\ . \\ \sum_i u_{ki}v_i \end{bmatrix} = \begin{bmatrix} \sum_i u_{1i}^2 & \sum u_{1i}u_{2i} \ldots\ldots\ldots\ldots \sum u_{1i}u_{ki} \\ \sum u_{1i}u_{2i} & \sum_i u_{2i}^2 \ldots\ldots\ldots \sum u_{2i}u_{ki} \\ & . \\ & . \\ & . \\ & . \\ \sum u_{1i}u_{ki} & \sum u_{2i}u_{ki} \qquad\qquad \sum_i u_{ki}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ . \\ b_k \end{bmatrix} \quad \ldots\ldots\ldots (3.33)$$

$Y_{k\times 1}$ $\qquad\qquad$ $A_{k\times k}$ $\qquad\qquad$ $B_{k\times 1}$

Thus the set of above equations (3.33) may be written as follows

Y=AB

In the above equations,

$$\sum_i u_{1i}^2 = \sum_i x_{ji}^2 - \left(\sum_i x_{ji}\right)^2/n$$

$$\sum_i u_{ji}u_{j'j} = \sum_i x_{ji}x_{j'j} - \left(\sum_i x_{ji}\right)\left(\sum_i x_{j'j}\right)/n$$

For j≠ j'

$$\sum_i u_{ji}v_i = \sum_i x_{ji}y_i - \left(\sum_i x_{ji}\right)\left(\sum_i y_i\right)/n$$

The matrix A is known as the coefficient matrix. From Y=AB, we can get the solution $B = A^{-1}Y$

Where A^{-1} *is* the inverse of the non-singular matrix A. Let $[c_{jj}]$ be the inverse matrix of A and j,j' =1,2,…,k.

$$A^{-1}=\begin{bmatrix} c_{11} & c_{12} & \cdot\cdot & c_{1k} \\ c_{21} & c_{22} & \cdot\cdot & c_{2k} \\ . & . & \cdot\cdot & . \\ . & . & \cdot\cdot & . \\ . & . & \cdot\cdot & . \\ c_{k1} & . & \cdot\cdot & c_{kk} \end{bmatrix}$$

Matrix A^{-1} is a symmetric matrix.

Thus we can write in the expanded form as

$$\begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_k \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdot & \cdot & c_{1k} \\ c_{21} & c_{22} & \cdot & \cdot & c_{2k} \\ . & . & \cdot & \cdot & . \\ . & . & \cdot & \cdot & . \\ . & . & \cdot & \cdot & . \\ . & . & \cdot & \cdot & . \\ c_{k1} & . & \cdot & . & c_{kk} \end{bmatrix} \begin{bmatrix} \sum_i u_{1i} v_i \\ \sum_i u_{2i} v_i \\ . \\ . \\ . \\ . \\ \sum_i u_{ki} v_i \end{bmatrix}$$

From the above relation partial regression coefficients are

$$b_1 = c_{11} \sum_i u_{1i} v_i + c_{12} \sum_i u_{2i} v_i + \ldots + c_{1k} \sum_i u_{ki} v_i$$

$$.b_2 = c_1 \sum_i u_{1i} v_i + c_{22} \sum_i u_{2i} v_i + \ldots + c_{1k} \sum_i u_{ki} v_i$$

……

$$b_j = c_{j1} \sum_i u_{1i} v_i + c_{j2} \sum_i u_{2i} v_i + \ldots + c_{jk} \sum_i u_{ki} v_i$$

……

$$b_k = c_{k1} \sum_i u_{1i} v_i + c_{12} \sum_i u_{2i} v_i + \cdots + c_{kk} \sum_i u_{ki} v_i$$

Thus by obtained values of $b_1, b_2, \ldots$ the final regression equation will be

$$(Y - \bar{y}) = b_1(X_1 - \bar{x}_1) + b_2(X_2 - \bar{x}_2) + \cdots + b_k(X_k - \bar{x}_k)$$

$$Y = R + b_1X_1 + b_2X_2 + \cdots + b_kX_k \quad \ldots\ldots\ldots (3.34)$$

$$\text{And} \quad R = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - \cdots - b_k\bar{x}_k$$

Where the estimated value of Y will be obtained by substituting the given values $X_1, X_2, \ldots\ldots\ldots, X_k$ in the prediction Equation (3.34).

3.10. The statistical features used for predicting the best model

3.10.1. Co-efficient of correlation

The extent or degree of relationship between the two variables is measured interms of another parameter called co-efficient of correlation. It is a measure of the closeness between the two variables. It lies between -1 to +1. The correlation is perfect and positive if r=1 and it is perfect and negative if r=-1. If r=0 then there is no correlation between the two variables and said to be independent.

3.10.2. Cross validated coefficient

The cross validate coefficient is the leave one out scheme, a model is build with n-1 compounds and the n^{th} compound is predicted. Each compound is left out of the model derivation and predicted in turn. An indication of the performance of the model is obtained from the cross validated coefficient.

$$r_{CV}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_p)^2}{\sum_{i=1}^{n}(y_i - y_a)^2} \quad \ldots\ldots\ldots (3.35)$$

Where y_i is the actual experimental activity, y_p is the predicted activity of compound i and $y_a$is the average of actual experimental activity.

3.10.3. F test

Fischer statistics is the ratio between explained and unexplained variance for a given number of degree of freedom and is defined as:

$$F = \frac{\{n-(m+1)\}*r^2}{(1-r^2)*m} \quad \ldots\ldots\ldots (3.36)$$

Where n is the number of data set, m is number of descriptors used in the QSAR equation and r is the co-efficient of correlation. Higher values of the F test indicate the significance of the QSAR model.

3.10.4. Quality factor

Quality factor is defined as

$$Q = \frac{r}{s} \quad \ldots\ldots\ldots (3.37)$$

Where r is the co-efficient of correlation and s is the standard deviation. High value of Q indicate high predictive power of the QSAR model.

3.11. References

[1] H. Wiener, Structural determination of paraffin boiling points, J Am Chem Soc. 69 (1947) 17-20.

[2] F. Harary, Graph theory, Addison-Wesley, Reading, MA, 1969.

[3] L.B. Kier, L.H. Hall, Molecular connectivity in structure activity analysis, Research studies press Letchworth, Hertfordshire, U. K., 1986.

[4] M. Randić, Characterization of molecular branching, J Am Chem Soc. 97 (1975) 6609-6615.

[5] S.C. Basak, Use of molecular complexity indices in predictive pharmacology and toxicology: a QSAR approach, Med Sci Res. 15 (1987) 605-609.

[6] C.E. Shannon, A mathematical theory of communication, Bell Syst Tech J. 27 (1948) 379-423.

[7] S.C. Basak, S. Bertelsen, G.D. Grunwald, Application of graph theoretical parameters in quantifying molecular similarity and structure-activity studies, J Chem Inf Comput Sci. 34 (1994) 270-276.

[8] S.C. Basak, G.D. Grunwald, Predicting mutagenicity of chemicals using topological and quantum chemical parameters: a similarity based study, Chemosphere. 31 (1995) 2529-2546.

[9] R. Todeschini, V. Consonni, Handbook of molecular descriptors, Wiley, New York, 2008.

[10] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J Mol Biol. 55 (1971) 379-400.

[11] M.L. Connolly, Analytical molecular surface calculation, J Appl Cryst.16 (1983) 548-558.

[12] M.F. Sanner, Ph.D. dissertation thesis: Modeling and applications of molecular surfaces, Université de Haute-Alsace, France, 1992.

[13] M.F. Sanner, A.J. Olson, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers. 38 (1996) 305-320.

[14] J. Kujawski, H. Popielarska, A. Myka, B. Drabińska, M.K. Bernard, The log P Parameter as a molecular descriptor in the computer-aided drug design – an overview, CMST. 18 (2012) 81-88.

[15] D. Eros, I. Kovesdi, L. Orfi, K. Takacs-Novak, G. Acsády, G. Kéri, Reliability of logP predictions based on calculated molecular descriptors: a critical review, Curr Med Chem. 9 (2002) 1819-1829.

[16] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J Am Chem Soc. 107 (1985) 3902-3909.

[17] J.J.P. Stewart, Optimization of parameters for semiempirical methods.1. method, J Comput Chem. 10 (1989) 209-220.

[18] J.J.P. Stewart, Comments on a comparison of AM1 with the recently developed PM3 method- reply, J Comput Chem. 11 (1990) 543-544.

[19] J.J.P. Stewart, Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi, J Comput Chem. 12 (1991) 320-341.

[20] J.B. Foresman, A. Frisch, Exploring chemistry with electronic structure methods, Gaussian Inc., Pittsburgh, 1996.

[21] W. Koch, M.C. Holthausen, A chemist's guide to density functional theory, Wiley-VCH, Weinheim, 2000.

[22] C. Gruber, V. Buss, Quantum-mechanically calculated properties for the development of quantitative structure-activity relationships (QSAR'S). pKa-values of phenols and aromatic and aliphatic carboxylic acids, Chemosphere. 19 (1989) 1595-1609.

[23] T. Sotomatsu, Y. Murata, T.J. Fujita, Correlation analysis of substituent effects on the acidity of benzoic acids by the AM1 method, J Comput Chem. 10 (1989) 94-98.

[24] M. Karelson, V.S. Lobanov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, Chem Rev. 96 (1996) 1027-1043.

[25] D.F. Lewis, C. Ioannides, D.V. Parke, Interaction of a series of nitriles with the alcohol-inducible isoform of P450: computer analysis of structure-activity relationships, Xenobiotica. 24 (1994) 401-408.

[26] R.G. Parr, R.A. Donnelly, M. Levy, W.E. Palke, Electronegativity: the density functional viewpoint, J Chem Phys. 68 (1978) 3801-3807.

[27] R.P. Iczkowski, J.L. Margrave, Electronegativity, J Am Chem Soc. 83 (1961) 3547-3551.

[28] I.N Levine, Quantum chemistry, Pearson Education, Singapore, 2000.

[29] I.N Levine, Physical chemistry, Mc-Graw Hill, New York, 2011.

[30] B.J. McConkey, V. Sobolev, M. Edelman, The performance of current methods in ligand-protein docking, Curr Sci. 83 (2002) 845-855.

[31] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J Comput Chem. 19 (1998) 1639-1662.

[32] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R,D. Taylor, Improved protein-ligand docking using GOLD, Proteins. 52 (2003) 609-623.

[33] T.J. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, J Comput Aided Mol Des. 15 (2001) 411-428.

[34] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1 Method and assessment of docking accuracy, J Med Chem. 47 (2004) 1739-1749.

[35] R. Abagyan, M. Totrov, D. Kuznetsov, ICM-A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J Comput Chem. 15 (1994) 488-506.

[36] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J Mol Biol. 261 (1996) 470-489.

[37] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J Am Chem Soc. 106 (1984) 765-784.

[38] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, J Comput Chem. 28 (2007) 1145-1152.

[39] T.W. Anderson, An introduction to multivariate analysis, John Wiley, New York, 1958.

[40] A.M. Kshirsagar, Multivariate analysis, Marcel Dekker, New York, 1972.

[41] C.R. Rao, Advanced statistical inference and its applications, Wiley Eastern, New Delhi, 1973.

CHAPTER 4

QSAR study and Molecular docking of 23-hydroxybetulinic acid derivatives as RMGPa and HeLa cells inhibitors

4.1. Introduction

The phosphorolytic cleavage of α-1, 4-linked glucosyl units in glycogen into α-D-glucose-1-phosphate is mediated by glycogen phohphorylase (GP) with the help of a debranching enzyme and plays an important role for controlling hepatic glucose production [1]. GP has three tissue specific isoforms which are brain, liver, and muscle according to their expression patterns. The muscle isoform supplies energy for muscle contraction while the brain isoform provides glucose during the periods of severe hypoglycemia. In glycogenolysis, liver enzyme plays a rate limiting role and hence it is exploited as an attractive target for the treatment of type 2 diabetes [2-5]. GP exists in two interconvertible forms: the phosphorylated high activity glycogen phosphorylase a (GPa) and the dephosphorylated low activity glycogen phohphorylase b (GPb). Allosteric effectors can promote equilibrium between a less active GPb and an active GPa. The active conformation is stabilized by phosphorylation of Ser 14 and binding of AMP [6, 7]. GP contains at least six potential regulatory binding sites: glucose analogues at the catalytic site, azasugar inhibitors, lactones at the allosteric site (AMP), caffeine at the purine inhibitor site, indole-2-carboxamide at the indole binding site and cyclodextrins at the glycogen storage site [8,9]. The X-ray analysis indicates that pentacyclic triterpenes bind at the allosteric site [10].

A series of 23-hydroxybetulinic acid derivatives were reported as potent GPa inhibitors [11]. Some derivatives of these triterpenes display anti-tumor activities against a variety of tumor cell

lines and the mechanism of action may be associated to its effects on the proliferation, migration, cell cycle and apoptosis of tumor cells [12, 13].

Apoptosis or programmed cell death plays an important role in regulating development and homeostasis. In many diseases including cancer, apoptosis is suppressed. Telomerase, a ribonucleoprotein, maintains chromosome lengths by adding telomeres to the chromosome ends repeatedly. Telomerase activation has been found in about 90% of tumor tissues, but with very low, almost undetectable activity in somatic cells. Apoptosis is regulated by a number of cellular genes including B cell leukemia/lymphoma 2 (bcl-2). The stable expression of bcl-2 in human cancer cells raises telomerase activity and resistance to apoptosis. The 23-hydroxybetulinic acid induces apoptosis through simultaneous inhibition of bcl-2 expression and telomerase activity [14- 17].

Furthermore, 23-hydroxybetulinic acid improves anti-tumor activity of doxorubicin in vitro and in vivo. The synergism is associated with increasing doxorubicin concentration in tumor tissue brought about by 23-hydroxybetulinic acid. Hence 23-hydroxybetulinic acid has the prospect to be developed as a novel chemosensitizer [18].

In this study, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa [11]. The other set of equations predicts the antiproliferative activities against HeLa cells. We have also performed docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa.

4.2. Materials and methods

The development of QSAR models are described in the following section as: dataset preparation, parameters and statistical methods.

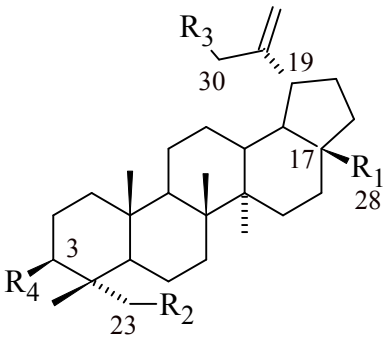
4.2.1. Dataset and parameters

The dataset which is selected from the literature [11, 12] contains 47 compounds of 23-hydroxybetulinic acid derivatives. The biological property of this data set is reported as IC_{50} values. The IC_{50} values were converted into log IC_{50} and taken as the response variable for QSAR modeling. Structural details of the 47 compounds and their biological activity are listed in Table 4.1.

The initial structures of 47 compounds used in this study were constructed by ChemSketch [19]. We attempted several descriptors (data not shown) and it has been found that quantum chemical descriptors (EH, EL, μ), molar refractivity (MR), molar volume (MV) and topological indices such as Structural Information Content (SIC), Complementary Information Content (CIC), Wiener index (W), Harary index (H), Randiac's connectivity index of first order (χ^0) [20-26] can better represent the biological activity of compounds.

The quantum chemical properties (EH, EL, μ) of the studied molecules have been determined by DFT/B3LYP calculation and the basis set 6-31G* was used. All quantum chemical calculations were performed by Gamess [27]. Molar refractivity (MR) and molar volume (MV) were determined using ChemSketch software. The graph theoretical descriptors such as SIC, CIC, W, H, χ^0 were computed using program written by us in Fortran-77.

Table 4.1. Structures and activity of 47 compounds of 23-hydroxybetulinic acid derivatives



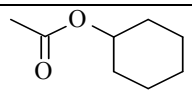
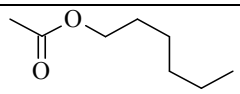
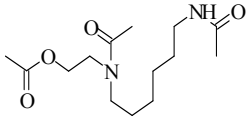
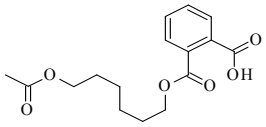
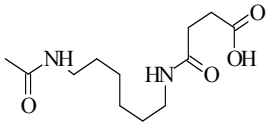
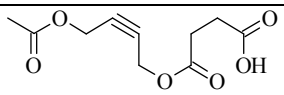
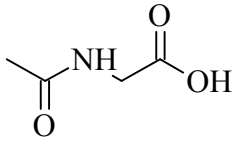
Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
1	-COOH	-OH	-H	-OH	103	46.22
2		-OH	-H	-OH	50.4	-
3		-OH	-H	-OH	69.1	-
4		-OAc	-H	-OAc	96.6	-
5		-OAc	-H	-OAc	69.2	-
6		-OAc	-H	-OAc	194	-
7		-OAc	-H	-OAc	94.5	-
8		-OH	-H	-OH	67.7	-

Table 4.1 (continued)

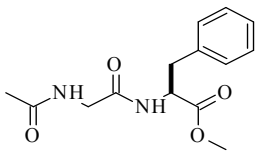
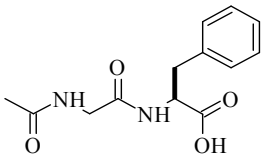
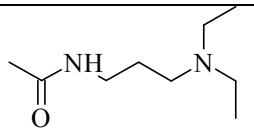
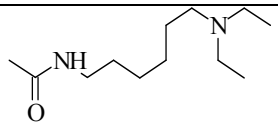
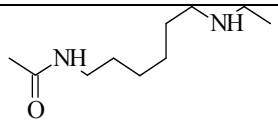
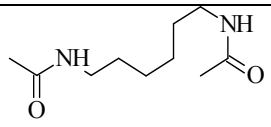
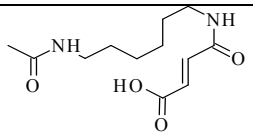
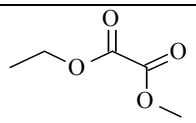
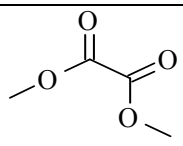
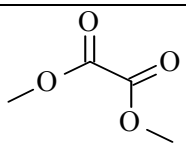
Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
9		-OH	-H	-OH	40.6	-
10		-OH	-H	-OH	129	-
11		-OAc	-H	-OAc	72.8	-
12		-OAc	-H	-OAc	55.6	-
13		-OAc	-H	-OAc	29.5	-
14		-OAc	-H	-OAc	34.1	-
15		-OAc	-H	-OAc	54.8	-
16			-H		30.1	-

Table 4.1 (continued)

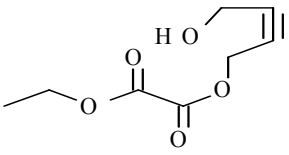
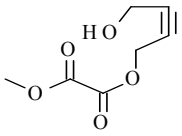
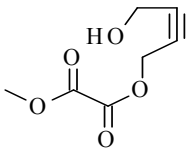
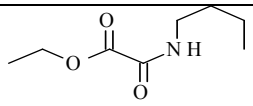
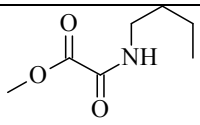
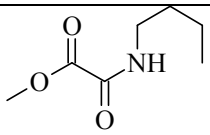
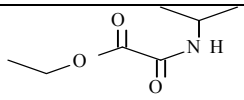
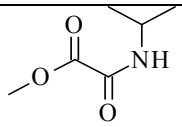
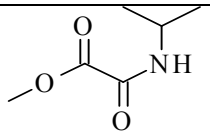
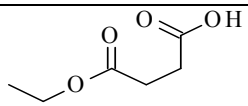
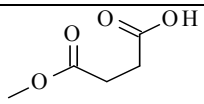
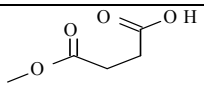
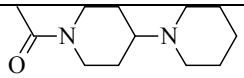
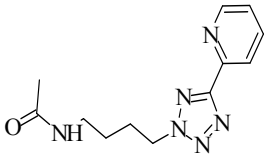
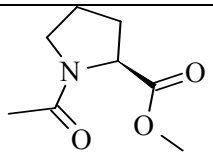
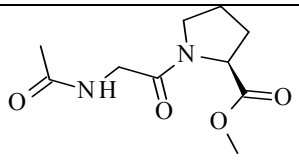
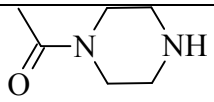
Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
17			-H		19.5	-
18			-H		34.7	-
19			-H		35.9	-
20			-H		97.8	-
21		-OAc	-H	-OH	-	10.80
22		-OAc	-H	-OH	-	17.87
23		-OAc	-OH	-OH	-	13.08
24		-OH	-H	-OH	-	4.84
25		-OAc	-H	-OH	-	18.84

Table 4.1 (continued)

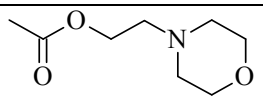
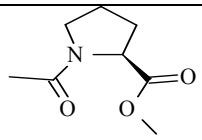
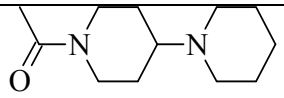
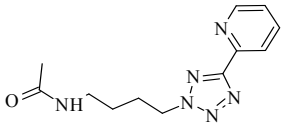
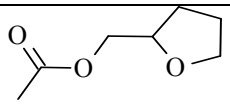
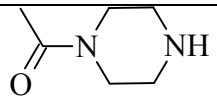
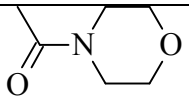
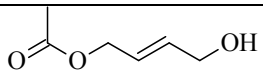
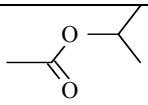
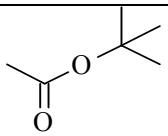
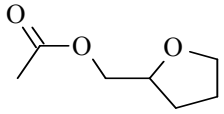
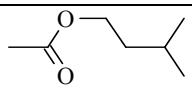
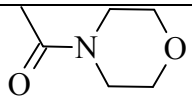
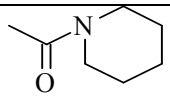
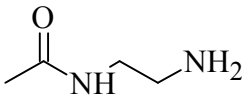
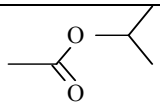
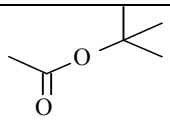
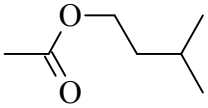
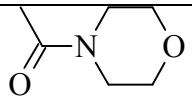
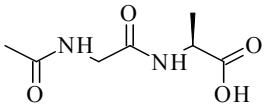
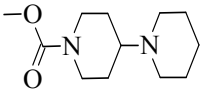
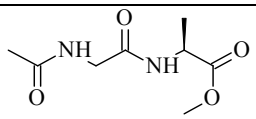
Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
26		-OAc	-H	-OH	-	9.18
27		-OAc	-H	-OH	-	9.78
28		-OH	-H	-OH	-	8.42
29		-OH	-H	-OH	-	46.26
30		-OH	-OH	-OH	-	79.35
31		-OH	-H	-OH	-	7.12
32		-OH	-H	-OH	-	34.29
33		-OH	-H	-OH	-	44.44
34		-OAc	-OH	-OH	-	19.46
35		-OAc	-OH	-OH	-	23.28
36		-OAc	-OH	-OH	-	17.79

Table 4.1 (continued)

Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
37		-OAc	-OH	-OH	-	20.84
38		-OAc	-OH	-OH	-	18.31
39		-OAc	-OH	-OH	-	11.02
40		-OAc	-H	-OH	-	7.39
41		-OH	-OH	-OH	-	22.27
42		-OH	-OH	-OH	-	30.65
43		-OH	-OH	-OH	-	20.85
44		-OAc	-H	-OH	-	7.47
45		-OH	-H	-OH	-	22.10
46	-COOH		-H	-OH	-	12.52
47		-OH	-H	-OH	-	8.27

4.2.2. Statistical methods

Multiple linear regression (MLR) analysis was used to build up QSAR models. Different combinations of parameters were used to develop these models. Statistical qualities of MLR equations were judged by parameters like correlation coefficient (R), square of the correlation coefficient (R^2), cross validated coefficient (R^2_{cv}), standard deviation of the regression (S), Fischer statistics (F) and quality factor (Q). MLR program written by ourselves in Fortran-77 is used.

4.2.3. Molecular docking

The coordinates of RMGPa in complex with CHI (1LWO.pdb) [6] were obtained from the RCSB protein data bank (www.rcsb.org). The 23-hydroxy betulinic acid derivatives were docked into the active pocket of the enzyme by using docking program Autodock 4.0 [28-30]. Initially the structures of the ligands have been optimized with AM1 method and the hydrogen atoms were added to the enzyme. The Lamarckian genetic algorithm (LGA) was applied to look out for the best conformers. A grid map with 80x80x80 points and 0.375 Å spacing was used in Autogrid program to evaluate the binding energies between the inhibitors and RMGPa. The grid centre was set at the active site position 28.672, 0.621 and 53.033 and the default settings were used. For each compound ten docking poses were saved and ranked by binging energy. The lowest energy docking pose of 23-hydroxybetulinic acid was nearly identical to the Asiatic acid and Maslinic acid orientation in crystal structure with GPb [10]. So the pose with highest negative binding energy was selected for analyzing the type of interactions. The binding site was analyzed with molegro molecular viewer software [31].

4.3. Results and discussion

The data set of 47 compounds was divided into two groups. The first group of molecules contains 20 compounds having inhibitory activity of RMGPa and the second group of molecules contains 28 compounds with antiproliferative activities against HeLa cells.

The 20 compounds of the first group were subdivided into two parts: 15 molecules in training set (1,2,3,4,5,7,8,10,11,12,15,16,18,19,20) and 5 molecules in test set (6,9,13,14 and 17). A cross correlation matrix (Table 4.2) between descriptors and the $logIC_{50}$ values of training set demonstrates that SIC_1 and EL have positive correlation with activity, CIC_1, μ, lnMR and lnMV shows negative correlation and EH does not show any significant correlation. SIC_1, CIC_1, molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in Table 4.3.

Table 4.2. Correlation matrix of 15 training RMGPa inhibitors

	SIC_1	CIC_1	EH	EL	μ	ln MR	ln MV	$logIC_{50}$
SIC_1	1	-0.8949	-0.0991	-0.0641	-0.177	-0.3421	-0.4055	0.3677
CIC_1	-0.8949	1	0.1147	-0.2029	0.368	0.7222	0.7682	-0.3963
EH	-0.0991	0.1147	1	0.2071	0.1069	0.0393	0.0507	0.0105
EL	-0.0641	-0.2029	0.2071	1	-0.1946	-0.5701	-0.5533	0.6553
μ	-0.177	0.368	0.1069	-0.1946	1	0.4259	0.4537	-0.4433
ln MR	-0.3421	0.7222	0.0393	-0.5701	0.4259	1	0.9932	-0.2648
ln MV	-0.4055	0.7682	0.0507	-0.5533	0.4537	0.9932	1	-0.3051
$logIC_{50}$	0.3677	-0.3963	0.0105	0.6553	-0.4433	-0.2648	-0.3051	1

Table 4.3. SIC_1, CIC_1, quantum chemical descriptors, molar refractivity and molar volume of 15 training RMGPa inhibitors

Comp no.	SIC_1	CIC_1	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	0.4253	3.6537	-0.2310	0.0089	6.1915	134.70	427.30
2	0.3449	4.3331	-0.2290	-0.0230	3.1764	160.84	501.90
3	0.3411	4.3779	-0.2352	-0.0063	2.2178	162.70	535.20
4	0.3572	4.5137	-0.2289	-0.0299	6.5763	217.62	698.00
5	0.3760	4.3538	-0.2338	-0.0701	5.0497	219.55	680.80
7	0.3823	4.2048	-0.2330	-0.0518	1.8782	192.86	606.90
8	0.4090	3.8272	-0.2255	-0.0302	3.6113	147.59	466.30
10	0.4243	3.8966	-0.2271	-0.0451	1.8798	189.55	582.80
11	0.3602	4.3878	-0.2277	-0.0274	4.1665	192.49	618.80
12	0.3412	4.5893	-0.2160	-0.0229	7.8094	206.39	666.80
15	0.3898	4.2294	-0.2263	-0.0850	3.2123	207.84	651.80
16	0.3809	4.1733	-0.2332	-0.1259	5.0089	183.01	586.90
18	0.3658	4.5019	-0.2298	-0.0881	9.1787	230.42	735.10
19	0.3812	4.3319	-0.2340	-0.0781	8.6982	216.46	689.00
20	0.3800	4.2516	-0.2350	-0.0361	6.2042	196.29	608.70

Among the generated QSAR equations; three equations were finally selected for predicting the inhibitory activity of RMGPa. Three best models are given below:

Model 4.1

Log IC_{50}= 0.8954+(2.8856)SIC_1+(3.5920)EL

Where, N=15, R=0.773, R^2=0.598, R^2_{cv}=0.323, S=0.165, F=8.925, Q=4.685

Model 4.2

Log IC_{50}= 3.4496+(4.1769)lnMR+(-3.7611)lnMV+(-3.7737)EH+(4.5525)EL+(-0.0256)μ

Where, N=15, R=0.864, R^2=0.746, R^2_{cv}=0.256, S=0.172, F=5.287, Q=5.023

Model 4.3

Log IC_{50}= -8.4488+(12.1049)SIC_1+(1.1185)CIC_1+(-6.0974)EH+(5.3929)EL+(-0.0388)μ

Where, N=15, R=0.946, R^2=0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227

By using model number 4.1, 4.2 and 4.3 the theoretical log IC_{50} values of 15 training compounds are given in Table 4.4 together with experimental log IC_{50}.

Table 4.4. List of experimental and predicted $logIC_{50}$ of 15 training RMGPa inhibitors

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (By model 4.1)	Predicted $logIC_{50}$ (By model 4.2)	Predicted $logIC_{50}$ (by Model 4.3)
1	2.0128	2.1546	1.9002	2.0023
2	1.7024	1.8080	1.9600	1.7217
3	1.8395	1.8570	1.8905	1.8909
4	1.9850	1.8187	1.8639	1.9029
5	1.8401	1.7285	1.8688	1.8239
7	1.9754	1.8125	1.9211	1.9504
8	1.8306	1.9671	1.8205	1.8548
10	2.1106	1.9577	2.0098	2.1142
11	1.8621	1.8363	1.8723	1.8980
12	1.7451	1.7977	1.7660	1.7050
15	1.7388	1.7149	1.7548	1.7970
16	1.4786	1.5423	1.4111	1.3783
18	1.5403	1.6345	1.5798	1.5844
19	1.5551	1.7148	1.6357	1.6789
20	1.9903	1.8622	1.9520	1.9039

The model 4.3 with the R=0.946, R^2 =0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227 turns out to be the best fit model. The indices of the 5 test compounds are presented in Table 4.5.

Table 4.5. Quantum chemical descriptors, SIC_1 and CIC_1 of 5 test RMGPa inhibitors

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	SIC_1	CIC_1
6	-0.2266	-0.0315	0.5028	0.3742	4.3521
9	-0.2317	-0.0532	3.9045	0.4216	3.9376
13	-0.2299	-0.0355	4.5568	0.3516	4.4704
14	-0.2253	-0.0291	3.7998	0.3604	4.4021
17	-0.2329	-0.1254	4.1689	0.4135	4.0854

Using the model number 4.3, we calculated the theoretical log IC_{50} values of the test set (R=0.891) are given in Table 4.6.

Table 4.6. List of experimental and predicted logIC_{50} of 5 test RMGPa inhibitors

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (by Model 4.3)
6	2.2878	2.1409
9	1.6085	2.0331
13	1.4698	1.8409
14	1.5328	1.9068
17	1.2900	1.7081

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 4.3) are presented in Figure 4.1 and 4.2 respectively.

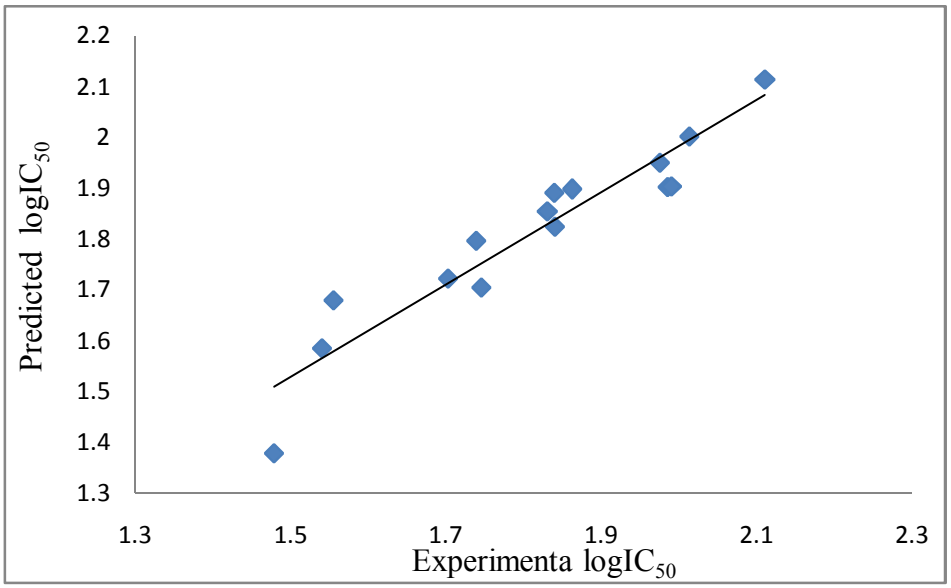


Figure 4.1. A plot between the predicted and the experimental activities for the training set of RMGPa inhibitors

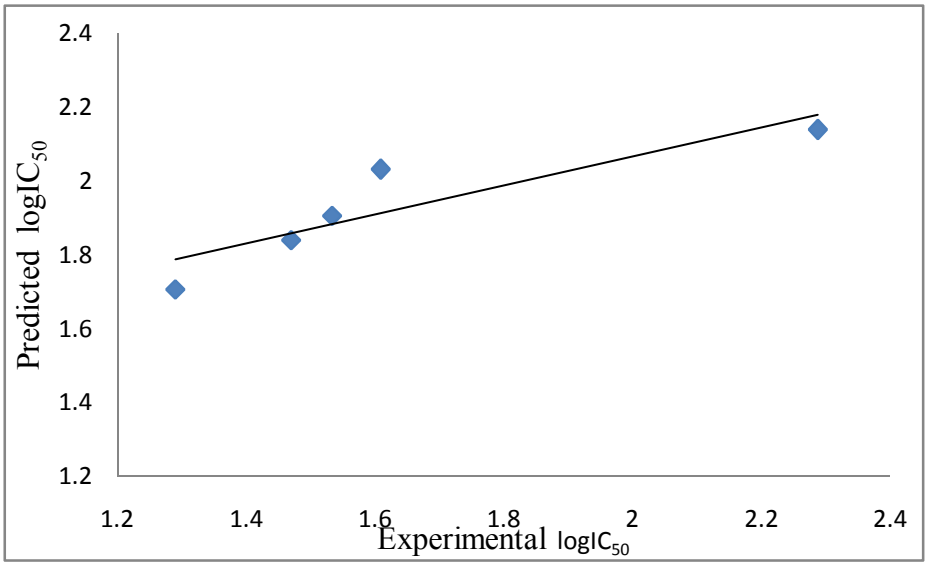


Figure 4.2. A plot between the predicted and the experimental activities for the test set of RMGPa inhibitors

The second group containing 28 compounds were divided into two parts: 20 molecules in training set (1,21,22,25,26,27,28,32,33,34,35,36,37,38,39,41,42,43,45,46) and test set of 8 molecules (23,24,29,30,31,40,44,47). The correlation matrix of electronic properties and topological indices with cytotoxic activity are presented in Table 4.7 and 4.8 respectively.

Table 4.7. Correlation matrix of 20 training antiproliferative compounds against HeLa cells with quantum chemical, molar refractivity and molar volume parameters

	EH	EL	μ	MR	MV	$logIC_{50}$
EH	1	-0.0285	0.0279	0.2584	0.2895	-0.4013
EL	-0.0285	1	0.4369	-0.099	0.0282	-0.3483
μ	0.0279	0.4369	1	0.1258	0.2688	-0.3627
MR	0.2584	-0.099	0.1258	1	0.9244	-0.7583
MV	0.2895	0.0282	0.2688	0.9244	1	-0.8578
$logIC_{50}$	-0.4013	-0.3483	-0.3627	-0.7583	-0.8578	1

Table 4.8. Correlation matrix of 20 training antiproliferative compounds against HeLa cells with different topological indices

	SIC_1	CIC_1	ln W	ln H	χ^0	$logIC_{50}$
SIC_1	1	-0.9307	-0.1859	-0.2095	-0.0872	0.457
CIC_1	-0.9307	1	0.5237	0.5436	0.4364	-0.6751
ln W	-0.1859	0.5237	1	0.9917	0.9856	-0.7546
ln H	-0.2095	0.5436	0.9917	1	0.977	-0.7818
χ^0	-0.0872	0.4364	0.9856	0.977	1	-0.6966
$logIC_{50}$	0.457	-0.6751	-0.7546	-0.7818	-0.6966	1

Molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in Table 4.9 and their topological indices are presented in Table 4.10.

Table 4.9. Quantum chemical descriptors, molar refractivity and molar volume of 20 training antiproliferative compounds against HeLa cells

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	-0.2310	0.0089	6.1915	134.70	427.30
21	-0.1999	-0.0134	5.1567	191.96	610.10
22	-0.2045	-0.0428	1.6927	202.81	560.70
25	-0.2170	-0.0013	3.8836	166.02	534.90
26	-0.2074	-0.0010	6.0763	175.31	572.80
27	-0.1566	-0.0016	7.3747	173.63	558.80
28	-0.2060	0.0125	7.3125	182.45	571.10
32	-0.2179	-0.0252	3.4896	154.78	490.90
33	-0.1968	-0.0230	2.9099	150.40	476.80
34	-0.2225	0.0001	2.4967	159.80	522.60
35	-0.1948	-0.0245	3.2973	164.44	538.80
36	-0.1963	-0.0375	4.3380	168.64	545.00
37	-0.2111	-0.0179	5.4231	169.07	555.60
38	-0.1693	-0.0068	4.6924	165.82	527.40
39	-0.1731	-0.0013	1.2905	168.82	536.40
41	-0.1986	-0.0012	3.8606	150.30	483.60
42	-0.1978	-0.0255	3.3445	154.94	499.80
43	-0.2032	-0.0367	5.9112	159.56	516.60
45	-0.1892	-0.0141	4.5361	165.07	522.10

Table 4.9 (continued)

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
46	-0.2008	0.0058	7.5438	5.2403	6.3820

Table 4.10. Topological indices of 20 training antiproliferative compounds against HeLa cells

Comp no.	SIC_1	CIC_1	ln W	H	χ^0
1	0.4253	3.6537	7.9649	162.1636	24.8970
21	0.3517	4.4462	8.9289	251.9431	34.2775
22	0.4288	3.9028	9.0789	263.7028	37.1059
25	0.3872	4.0716	8.5865	214.7929	30.2943
26	0.3720	4.2252	8.7828	228.4780	32.4156
27	0.3826	4.1372	8.7686	235.9327	32.7419
28	0.3484	4.4273	8.7749	233.0269	31.9930
32	0.3794	4.0679	8.3622	195.8768	28.0099
33	0.3866	4.0112	8.4016	191.0617	28.4325
34	0.3739	4.1325	8.5254	5.3306	30.1730
35	0.3714	4.1765	8.5916	213.8546	31.0957
36	0.3743	4.1747	8.7221	225.2267	31.7085
37	0.3621	4.2655	8.6630	217.3454	31.5872
38	0.3919	4.0402	8.6484	221.8434	31.0014
39	0.3749	4.1707	8.6484	221.8434	31.0014
41	0.3756	4.0734	8.2890	187.6982	27.8885
42	0.3732	4.1177	8.3612	194.8093	28.8112

Table 4.10 (continued)

Comp no.	SIC_1	CIC_1	ln W	H	χ^0
43	0.3630	4.2135	8.4670	198.6210	29.3028
45	0.4187	3.8536	8.6769	215.3581	31.7504
46	0.3665	4.3284	8.9253	245.4236	34.2775

Among the generated QSAR models; three models were finally selected. Model summary of three best models are given below.

Model 4.4

$\text{Log } IC_{50}$= 19.6957+(-4.4880)SIC_1+(-0.8479)CIC_1+(0.0427)lnW+(-3.0400)lnH+(0.0885)χ0

Where, N=28, R=0.859, R^2=0.738, R^2_{cv}=0.288, S=0.197, F=7.887, Q=4.360

Model 4.5

$\text{Log } IC_{50}$= 13.3115+(-0.1637)lnMR+(-1.8648)lnMV+(-2.2260)EH+(-4.4968)EL+(-0.0005)μ

Where, N=28, R=0.933, R^2=0.870, R^2_{cv}=0.594, S=0.205, F=18.738, Q=4.551

Model 4.6

$\text{Log } IC_{50}$= 13.4441+(-11.8969)SIC_1+(-2.0004)CIC_1+(-2.7582)EH+(-4.5601)EL+(-0.0010)μ

Where, N=28, R=0.916, R^2=0.839, R^2_{cv}=0.698, S=0.201, F=14.591, Q=4.557

Model 4.4 and model 4.5 are constructed by using topological indices and electronic parameters respectively. But the quality of the equation is much increased when we use both the topological and electronic indices simultaneously i.e. Model 4.6. By using model number 4.4, 4.5 and 4.6 the

theoretical log IC_{50} values of 20 training compounds are given in Table 4.11 together with experimental log IC_{50}. The indices of the 8 test compounds are given in Table 4.12.

In these models, N is the number of data points; R is the correlation coefficient between experimental values and calculated values from the equation. R^2 is the square of the correlation coefficient and it measures the goodness of fit of the regression equation. Cross validated coefficient (R^2_{cv}) gives an idea of the performance of the model. S is the standard deviation of the regression. Fischer statistics (F) is a ratio of variances between calculated and observed activity. The larger value of F test signifies the QSAR model. Q is the quality factor. Q value measures predictive power of the QSAR models.

Table 4.11. List of experimental and predicted logIC_{50} of 20 training antiproliferative compounds against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (By model 4.4)	Predicted logIC_{50} (by Model 4.5)	Predicted logIC_{50} (By model 4.6)
1	1.6648	1.7631	1.6839	1.6658
21	1.0334	0.9534	0.9934	0.9731
22	1.2521	1.1862	1.2860	1.2931
25	1.2751	1.2294	1.2467	1.2933
26	0.9628	1.1759	1.0864	1.1369
27	0.9903	1.1337	1.0231	1.0481
28	0.9253	1.0126	0.9836	0.9467
32	1.5352	1.3361	1.5281	1.5054
33	1.6478	1.4666	1.5305	1.4655

Table 4.11 (continued)

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (By model 4.4)	Predicted $logIC_{50}$ (by Model 4.5)	Predicted $logIC_{50}$ (By model 4.6)
34	1.2891	1.3430	1.3031	1.3399
35	1.3670	1.2959	1.2901	1.3166
36	1.2502	1.1867	1.3258	1.3481
37	1.3189	1.2595	1.2338	1.2620
38	1.2627	1.2020	1.1914	1.1929
39	1.0422	1.1676	1.1423	1.1229
41	1.3477	1.4644	1.4096	1.3766
42	1.4864	1.4093	1.4510	1.4256
43	1.3191	1.3629	1.4456	1.4187
45	1.3444	1.3977	1.2881	1.3357
46	1.0976	1.0663	0.9696	0.9452

Table 4.12. Quantum chemical descriptors, SIC_1 and CIC_1 of 8 test antiproliferative compounds against HeLa cells

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	SIC_1	CIC_1
23	-0.2102	-0.0159	6.4319	0.3970	4.0484
24	-0.1916	-0.0062	5.8572	0.4044	4.0072
29	-0.2086	-0.0356	5.4176	0.4292	3.8631
30	-0.2017	-0.0261	4.0357	0.3762	4.1168
31	-0.2222	0.0191	6.1811	0.3853	4.0382
40	-0.1571	0.0018	5.2365	0.4034	3.9284

Table 4.12 (continued)

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	SIC_1	CIC_1
44	-0.1915	-0.0211	3.9464	0.3785	4.1204
47	-0.2078	-0.0112	1.5058	0.4189	3.8771

We calculated the theoretical log IC_{50} of the test set (R=0.751) by model number 4.6 which appeared in Table 4.13.

Table 4.13. List of experimental and predicted logIC_{50} of 8 test antiproliferative inhibitors against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (by Model 4.6)
23	1.1166	1.2684
24	0.6848	1.1678
29	1.6652	1.3425
30	1.8995	1.4045
31	0.8525	1.3018
40	0.8686	1.2064
44	0.8733	1.3191
47	0.9175	1.3274

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 4.6) are presented in Figure 4.3 and 4.4 respectively.

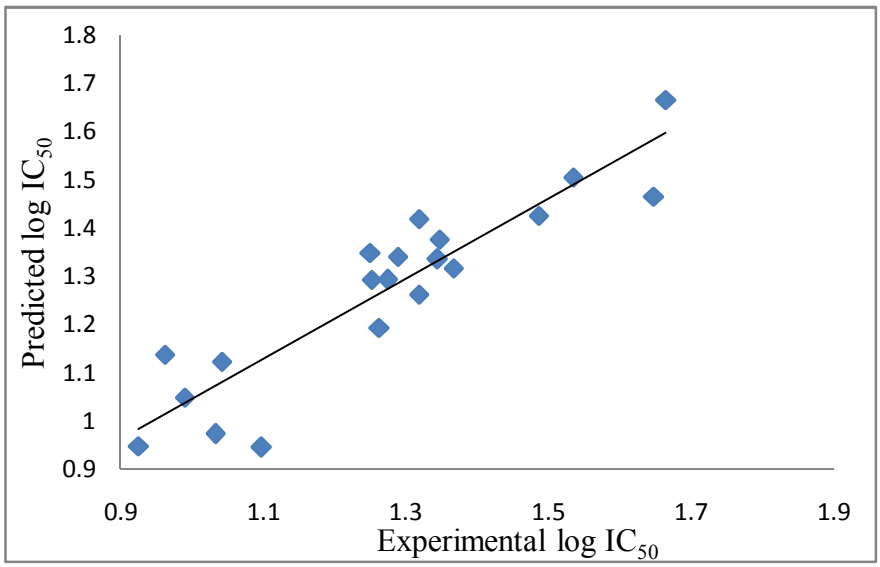


Figure 4.3. A plot between the predicted and the experimental activities for the training antiproliferative inhibitors against HeLa cells

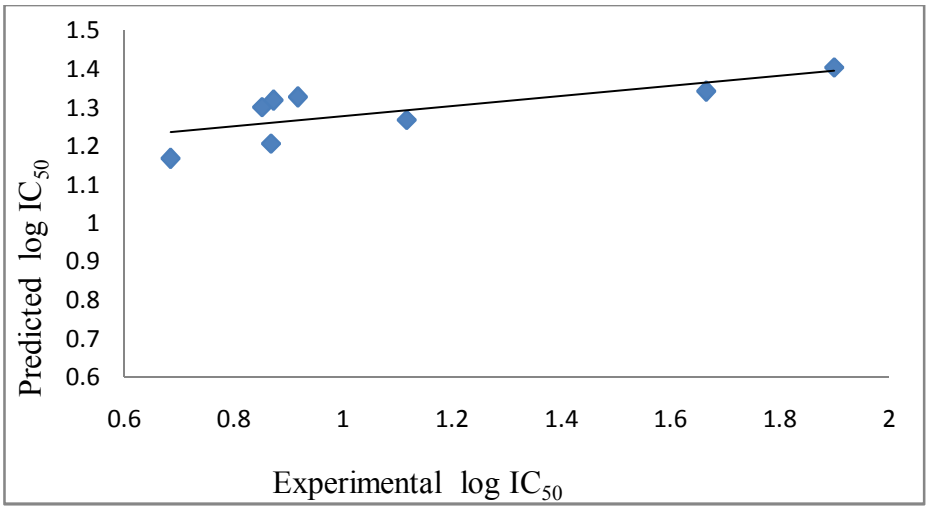


Figure 4.4. A plot between the predicted and the experimental activities for the test antiproliferative inhibitors against HeLa cells

We also calculated the antiproliferative activity (from compound number 2 to 20) and inhibitory activity of RMGPa (from compound number 21 to 47) by model number 4.6 and 4.3 respectively which were not determined experimentally and are shown in Table 4.14.

Table 4.14. Predicted Antiproliferative activity/ RMGPa inhibitory activity of studied compounds

Comp. No.	HeLa $IC_{50}(\mu M)$	RMGPa $IC_{50}(\mu M)$	Comp. No.	HeLa $IC_{50}(\mu M)$	RMGPa $IC_{50}(\mu M)$
2	25.48	-	20	16.79	
3	20.12	-	21	-	53.46
4	8.44	-	22	-	114.14
5	16.63	-	23	-	67.81
6	11.34	-	24	-	68.55
7	22.99	-	25	-	90.72
8	47.69	-	26	-	63.61
9	26.87	-	27	-	29.51
10	27.02	-	28	-	57.51
11	13.50	-	29	-	86.53
12	7.88	-	30	-	43.87
13	12.88	-	31	-	89.11
14	12.61	-	32	-	56.37
15	22.63	-	33	-	47.91
16	59.79	-	34	-	91.14
17	36.50	-	35	-	44.26
18	12.97	-	36	-	37.83
19	17.23	-	37	-	48.47

Table 4.14 (continued)

Comp. No.	HeLa $IC_{50}(\mu M)$	RMGPa $IC_{50}(\mu M)$	Comp. No.	HeLa $IC_{50}(\mu M)$	RMGPa $IC_{50}(\mu M)$
38	-	42.42	43	-	29.49
39	-	56.57	44	-	43.88
40	-	39.14	45	-	67.81
41	-	51.11	46	-	61.92
42	-	41.03			

The binding energies of 47 docked compounds are given in Table 4.15 and they are ranges between -4.58 and -13.2kcal/mol. The docking study shows both polar (GLN7, ARG10, LYS11, SEP14, ARG69, GLN71, GLN72, TYP74, TYR75, GLU76, ARG81, TYR155, LYS191, ARG193, GLU195, THR240, ARG242, ARG306, ARG309, AGR310, LYS312, SER313, SER314) and non polar (ILE13, VAL15, LEU18, ILE63, VAL64, TRP67, ILE68, PHE196, ASP227) amino acids make important interactions to the inhibitors.

Table 4.15. The binding energies of 47 docked compounds

Comp. No.	Binding Energy kcal/mol	Comp. No.	Binding Energy kcal/mol
1	-10.22	7	-7.21
2	-9.1	8	-7.30
3	-8.39	9	-6.74
4	-6.79	10	-7.49
5	-5.19	11	-5.67
6	-6.49	12	-7.30

Table 4.15 (continued)

Comp. No.	Binding Energy kcal/mol	Comp. No.	Binding Energy kcal/mol
13	-6.45	31	-7.58
14	-7.06	32	-7.51
15	-6.99	33	-7.56
16	-13.2	34	-8.27
17	-4.58	35	-7.77
18	-6.08	36	-9.05
19	-6.10	37	-8.15
20	-7.18	38	-7.59
21	-7.24	39	-7.51
22	-7.37	40	-6.65
23	-6.07	41	-7.49
24	-6.87	42	-7.31
25	-8.60	43	-9.56
26	-7.57	44	-8.51
27	-6.62	45	-7.55
28	-7.50	46	-8.97
29	-8.32	47	-7.41
30	-8.36		

In our study, we used different topological and quantum chemical indices to obtain phenogram based on Unweighted Pair Group Method with Arithmetic mean (UPGMA) of 47 compounds. The phenogram of the compounds is presented in Figure 4.5. Compounds having similar

molecular properties are in the same clade though in some cases their binding energies and activity are different.

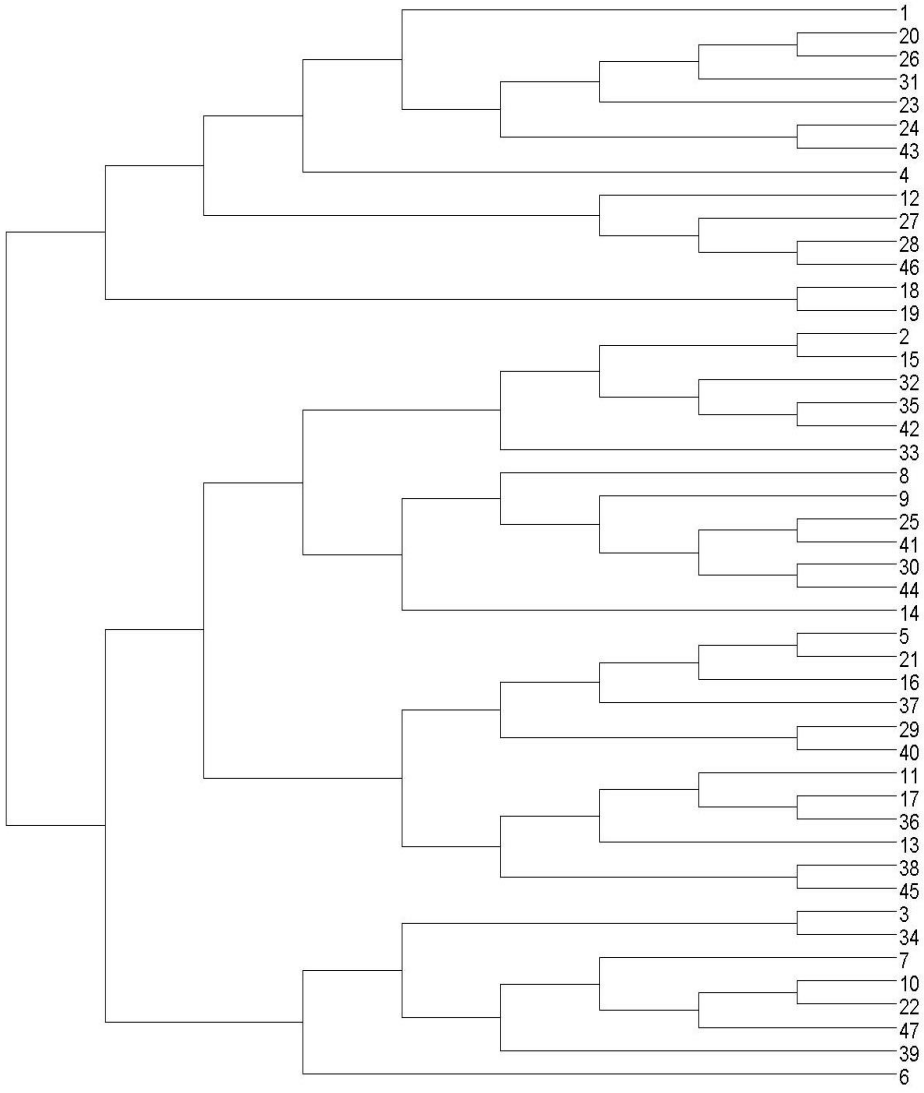


Figure 4.5. Phenogram using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) of 47 compounds of 23-hydroxy betulinic acid derivatives

Compound 1, 20, 26, 31, 23, 24, 43, 4 are in same clade. Ligand 1 (23-hydroxy betulinic acid) was used as a model drug (Figure 4.6a) as a reference for other compounds and have three hydrogen bonds. The carboxylate group at C-28 forms two hydrogen bonds at 1.743Å and 1.821Å with ARG310 and ARG 242 respectively. The –OH group at C-3 also forms another hydrogen bond with ILE68 (1.988 Å). In this clade, compound 20 and 26 remain as a pair and they have almost same binding energy. The binding energy of 20 and 26 are -7.38 and -7.56kcal/mol respectively. Compound 24 and 43 have different binding energies though they remain as a pair. Their predicted IC_{50} values are 68.50μM and 29.49μM respectively. In case of ligand 43 (Figure 4.6b), –OH group present at C-30 and the ester group at C-28 form two hydrogen bonds with SER313 (2.181 Å) and ARG310 (1.933 Å) respectively. Ligand 24 also forms two hydrogen bonds with ARG81 and GLN71 but at the same time the N-substituted amide group at C-28 may increase the chance of steric bumps with ARG309, which weakens the interaction between ligand and enzyme causing decrease in bioactivity (Figure 4.6c).

Compound 12, 27, 28, 46 are in same clade in which 46 has highest negative binding energy. In case of ligand 46 (Figure 4.6d), the carboxylate group at C-28 forms hydrogen bond with ARG193 (1.937Å). Compound 18 and 19 remain as a pair in the next clade and have comparable binding energy. Compound 2, 15, 32, 35, 42, 33 are present in same clade. Though compound 2 and 15 are in a pair, their binding energies are different. The IC_{50} values of ligand 2 and 15 are 50.4μM and 54.8μM respectively. Compound 2 lies well inside the protein (Figure 4.6e) and the –OH group at C-23 and C=O group at C-28 make two hydrogen bonds with GLN72 (2.675Å) and ARG310 (2.015Å) respectively. Although the amide group at C-28 of ligand 15 makes important interactions with the enzyme but the major part of the ligand is outside the protein (Figure 4.6f). Compound 35 and 42 are in a pair in same clade and they have almost same

binding energy. In another clade, compound 8, 9, 25, 41, 30, 44, 14 are present and their binding energies are almost as same as the other.

Compound 5, 21, 16, 37, 29, 40 are in same clade. Compound 16 has highest negative binding energy than the other compounds and forms one hydrogen bond with ARG310 (Figure 4.6g). In the next clade compound 17 and 36 are in a pair. The binding energy of 36 is more negative than 17. In ligand 36, the –OH group at C-3 and O atom at C-23 form hydrogen bonds with ARG310 at 2.002Å and 2.223Å respectively (Figure 4.6h). The acetyl group at C-23 also forms hydrogen bond with ARG242 (1.87Å). Ligand 17 forms hydrogen bonds with ARG69, GLN71 and TYR155. In spite of the steric bump formation between the long chains at C-3 with ARG310, this compound possesses good inhibitory activity due to hydrogen bonds (Figure 4.6i). Compound 3, 34 and 10, 22 are in pairs and they have comparable binding energy.

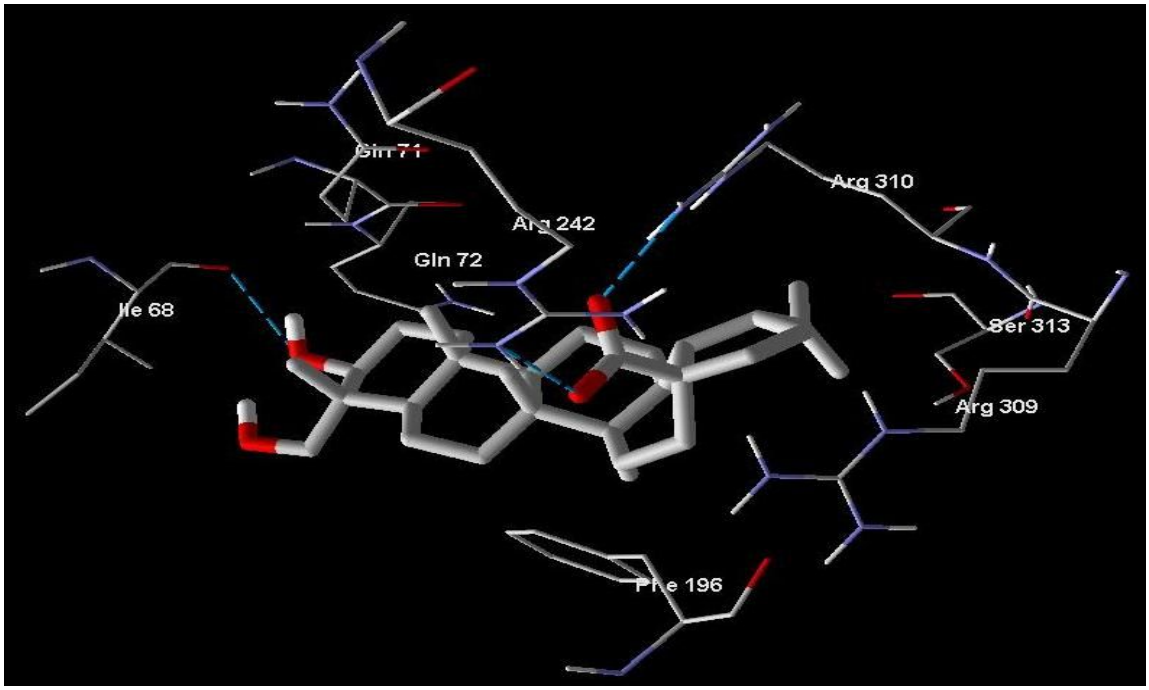


Figure 4.6a. Docked conformation of ligand 1 along with the important amino acid residues of RMGPa.

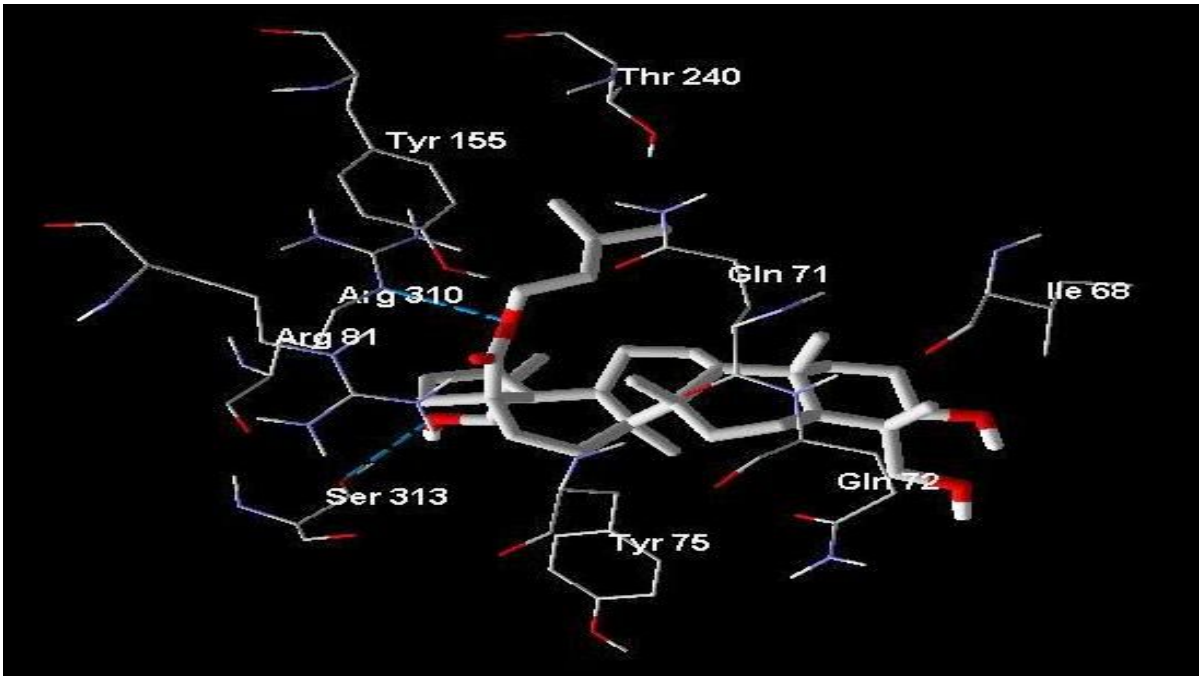


Figure 4.6b. Docked conformation of ligand 43 along with the important amino acid residues of RMGPa

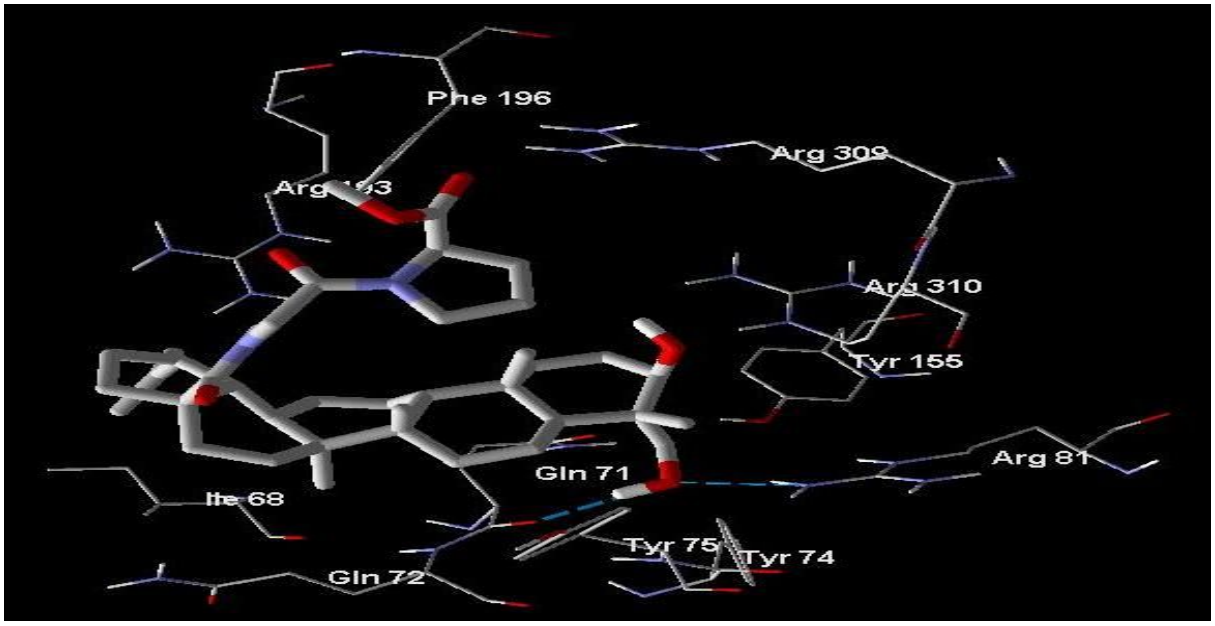


Figure 4.6c. Docked conformation of ligand 24 along with the important amino acid residues of RMGPa

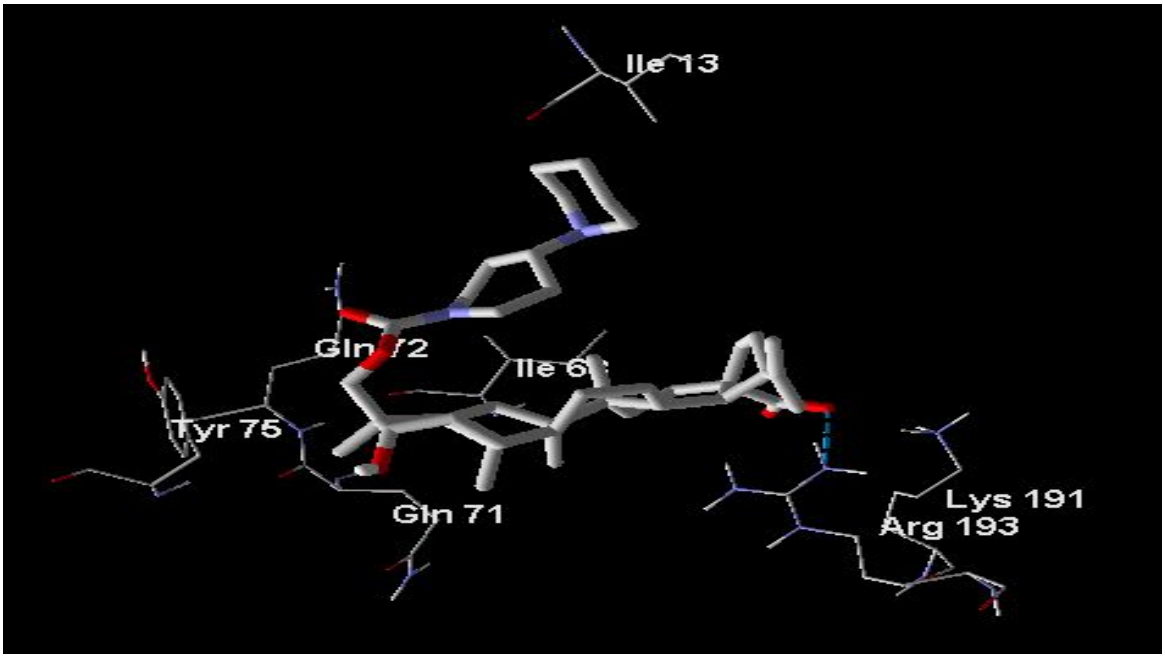


Figure 4.6d. Docked conformation of ligand 46 along with the important amino acid residues of RMGPa.

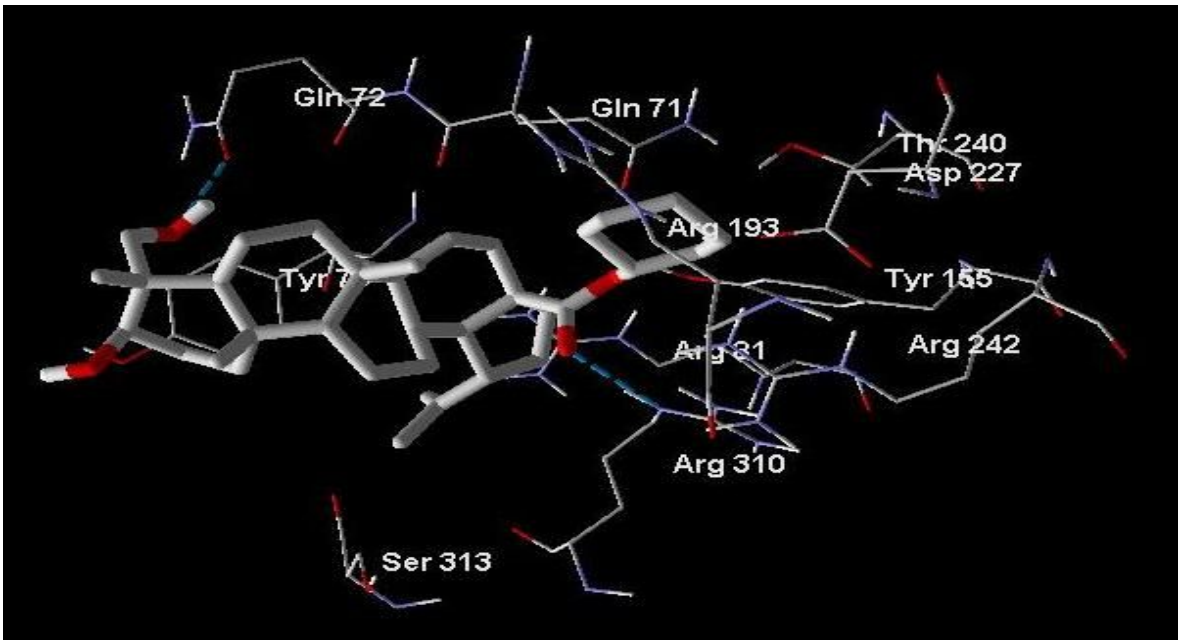


Figure 4.6e. Docked conformation of ligand 2 along with the important amino acid residues of RMGPa.

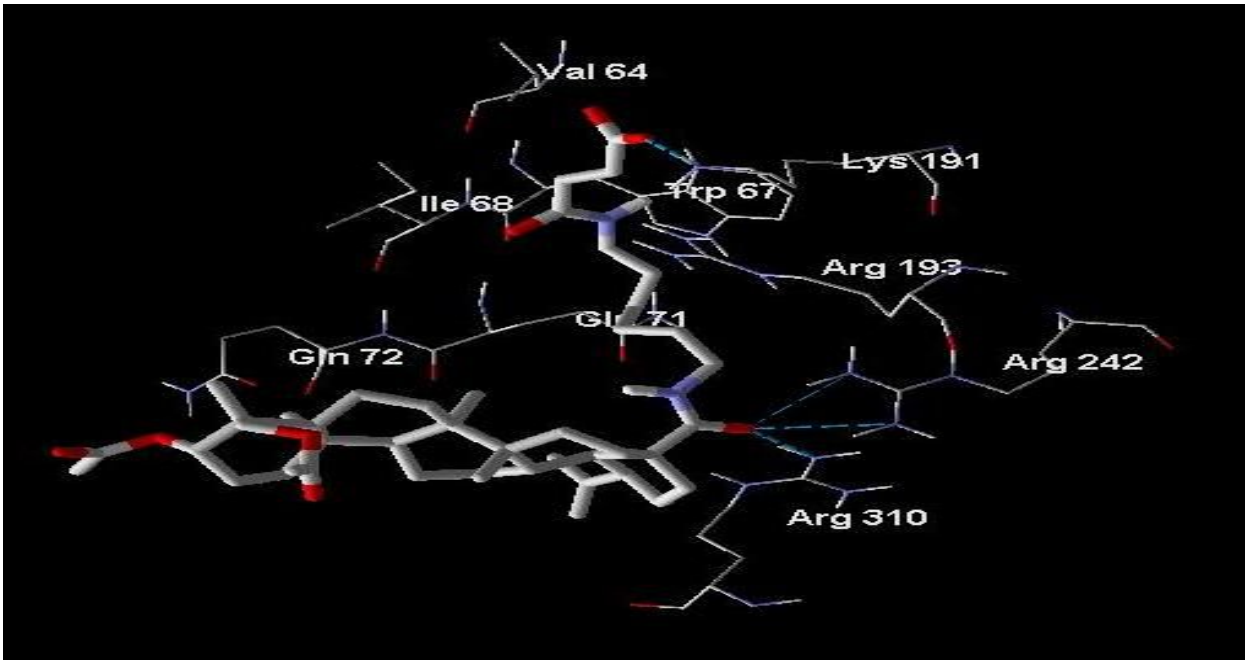


Figure 4.6f. Docked conformation of ligand 15 along with the important amino acid residues of RMGPa.

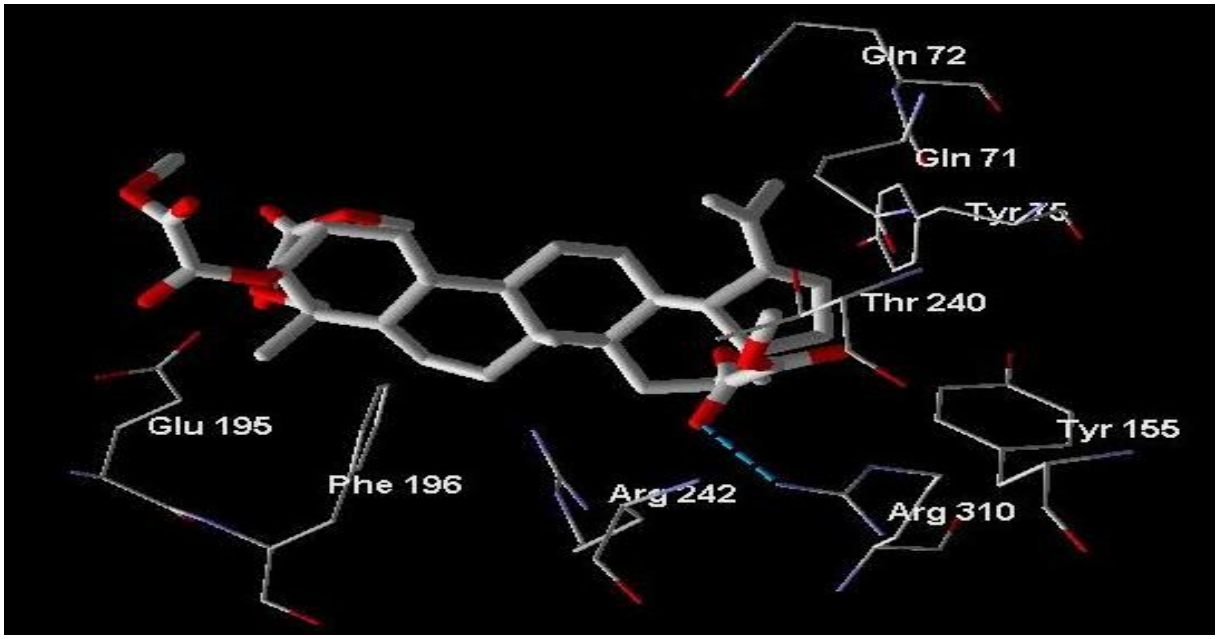


Figure 4.6g. Docked conformation of ligand 16 along with the important amino acid residues of RMGPa.

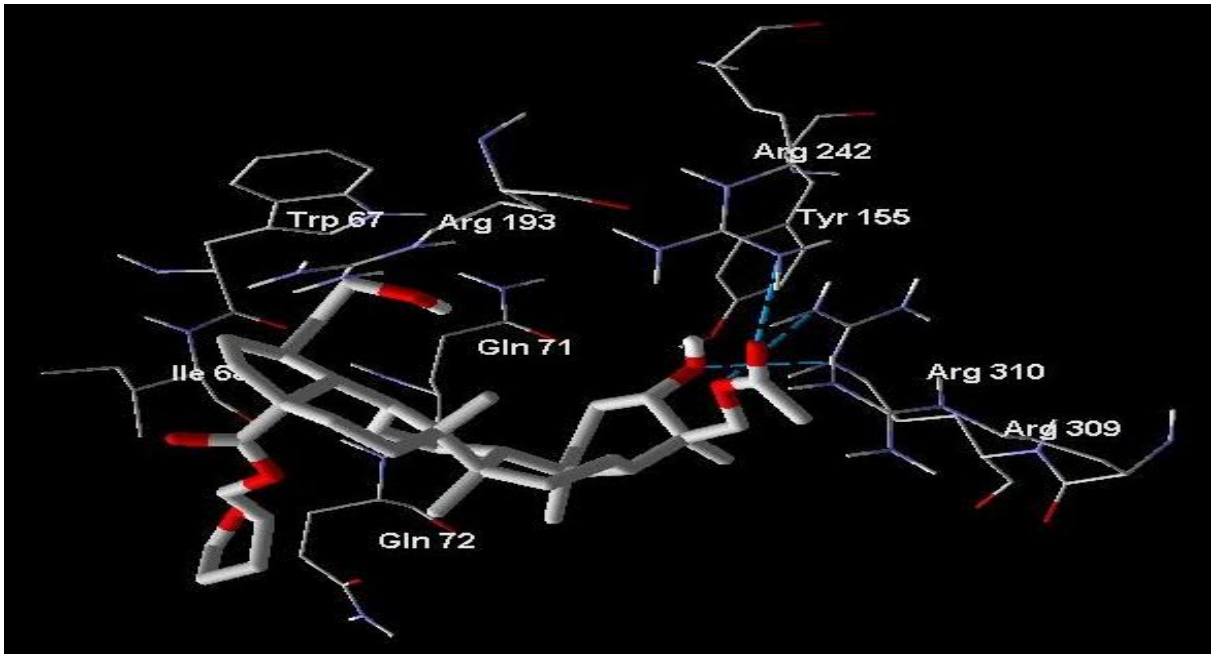


Figure 4.6h. Docked conformation of ligand 36 along with the important amino acid residues of RMGPa

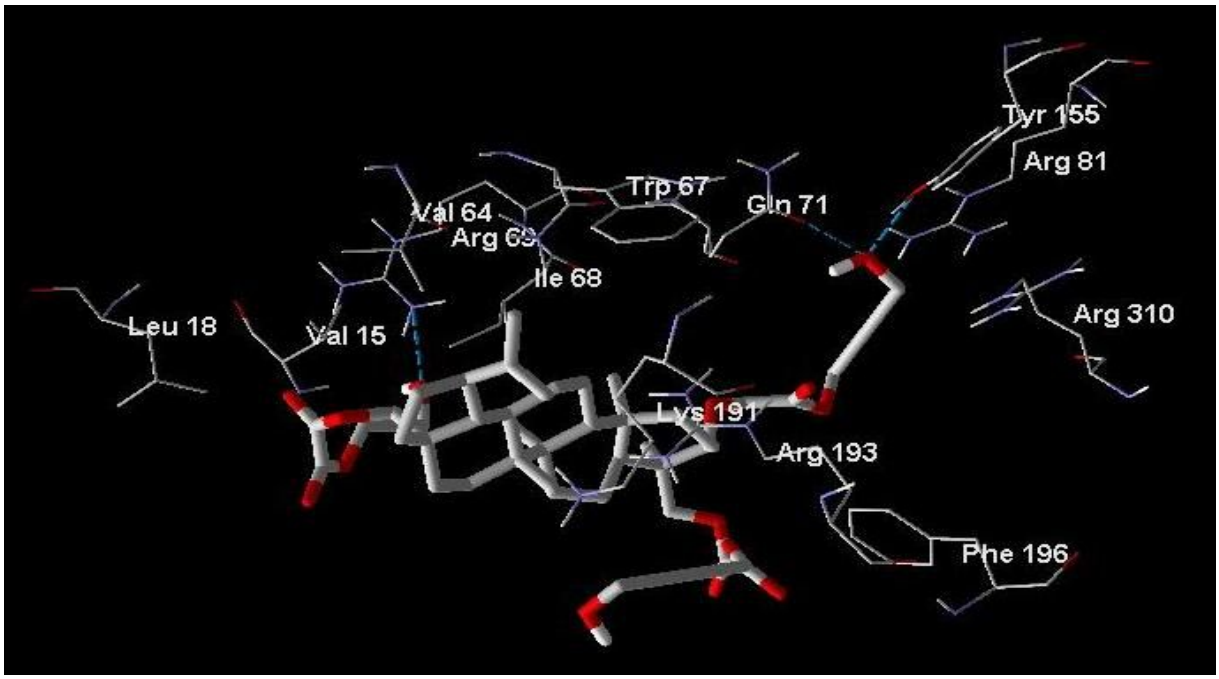


Figure 4.6i. Docked conformation of ligand 17 along with the important amino acid residues of RMGPa.

4.4. Conclusion

In conclusion, this QSAR study has shown that topological indices (e.g. SIC, CIC) and quantum chemical descriptors (e.g. EH, EL, μ) are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. Model 4.3 and model 4.6 are the best equation for predicting the inhibitory activity of RMGPa and the antiproliferative activities against HeLa cells respectively and these QSAR models may be used in prediction of activity of designed compounds. The docking study shows that the important interacting amino acids present in the active site are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the ligands can form hydrogen bonds with ARG193 and/or ARG310. The –OH group at C-3 and C-23 can increase the hydrogen bond interaction between ligands and enzyme. However, acetylation or esterification of the –OH group at the C-3 and C-23 not only decreases the number of hydrogen bond, but may also increase the unfavorable steric clashes. Thus binding energy may decrease. Large substituent at C-17 may increase the chance of steric bumps, thus lowering the inhibitory activity of the ligand.

4.5 References

[1] M. Bollen, S. Keppens, W. Stalmans, Specific features of glycogen metabolism in the liver, Biochem J. 336 (1998) 19-31.

[2] H.G. Hers, The control of glycogen metabolism in the liver, Annu Rev Biochem. 45 (1976) 167-190.

[3] W. Stalmans, The interaction of liver phosphorylase *a* with glucose and AMP, Eur J Biochem. 49 (1974) 415-427.

[4] W. Pimenta, N. Nurjhan, P.A. Jansson, M. Stumvoll, J. Gerich, M. Korytkowski, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia. 37 (1994) 697-702.

[5] M.K. Hellerstein, R.A. Neese, P. Linfoot, M. Christiannsen, S. Turner, A. Letscher, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest. 100 (1997) 1305-1319.

[6] N.G. Oikonomakos, E.D. Chrysina, M N. Kosmopoulou, D.D. Leonidas, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 A resolution, Biochim Biophys Acta. 1647 (2003) 325-332.

[7] V.L. Rath, M. Ammirati, D.E. Danley, J.L. Ekstrom, E.M. Gibbs, T.R. Hynes, A.M. Mathiowetz, R.K. Mcpherson, T.V. Olson, J.L. Treadway, D.J. Hoover, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol. 7 (2000) 677-682.

[8] R. Kurukulasuriya, J.T. Link, D.J. Madar, Z. Pei, S.J. Richards, J.J. Rohde, A.J. Souers, B.G. Szczepankiewicz, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem. 10 (2003) 123-153.

[9] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[10] X. Wen, H. Sun, J. Liu, K. Cheng, P. Zhang, L. Zhang, J. Hao, L. Zhang, P. Ni, S.E. Zographos, D.D. Leonidas, K.M. Alexacou, T. Gimisis, J.M. Hayes, N.G. Oikonomakos, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem. 51 (2008) 3540-3554.

[11] P. Zhu, Y. Bi, J. Xu, Z. Li, J. Liu, L. Zhang, W. Ye, X. Wu, Terpenoids. III: Synthesis and biological evaluation of 23-hydroxybetulinic acid derivatives as novel inhibitors of glycogen phosphorylase, Bioorg Med Chem Lett. 19 (2009) 6966-6969.

[12] P. Lan, J. Wang, D.M. Zhang, C. Shu, H.H. Cao, P.H. Sun, X.M. Wu, W.C. Ye, W.M. Chen, Synthesis and antiproliferative evaluation of 23-hydroxybetulinic acid derivatives, Eur J Med Chem. 46 (2011) 2490-2502.

[13] Y. Bi, J. Xu, F. Sun, X. Wu, W. Ye, Y. Sun, W. Huang, Synthesis and biological activity of 23-hydroxybetulinic acid C-28 ester derivatives as antitumor agent candidates, Molecules. 17 (2012) 8832-8841.

[14] W. Ding, M. Sun, S. Luo, T. Xu, Y. Cao, X. Yan, Y. Wang, A 3D QSAR study of betulinic acid derivatives as anti-tumor agents using topomer CoMFA: model building studies and experimental verification, Molecules. 18 (2013) 10228-10241.

[15] S. Fulda, K.M. Debatin, Betulinic acid induces apoptosis through a direct effect on mitochondria in neuroectodermal tumors, Med Pediatric Oncol. 35 (2000) 616-618.

[16] Z.N. Ji, W.C. Ye, G.G. Liu, W.L. Wendy Hsiao, 23-Hydroxybetulinic acid-mediated apoptosis is accompanied by decreases in Bcl-2 expression and telomerase activity in HL-60 Cells, Life Sci. 72 (2002) 1-9.

[17] M. Mandal, R. Kumar, Bcl-2 Modulates telomerase activity, J Biol Chem. 272 (1997) 14183-14187.

[18] Y Zheng, F. Zhou, X. Wu, X. Wen, Y. Li, B. Yan, J. Zhang, G. Hao, W. Ye, G. Wang, 23-Hydroxybetulinic acid from Pulsatilla chinensis (Bunge) Regel synergizes the antitumor activities of doxorubicin in vitro and in vivo, J Ethnopharmacol. 128 (2010) 615-622.

[19] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[20] E. Eroglu, Some QSAR Studies for a Group of Sulfonamide Schiff Base as Carbonic Anhydrase CA II Inhibitors, Int J Mol Sci. 9 (2008) 181-197.

[21] L.B. Kier, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci. 69 (1980) 807-810.

[22] S.C. Basak, D.K. Harriss, V.R. Magnuson, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci. 73 (1984) 429-437.

[23] H. Wiener, Structural determination of paraffin boiling points, J Am Chem Soc. 69 (1947) 17-20.

[24] D. Playsic, S. Nikolic, N. Trinajstic, Z. Mihalic, On the Harary index for the characterization of chemical graphs, J Math Chem. 12 (1993) 235-250.

[25] M. Randiac, On the characterization of molecular branching, J Am Chem Soc. 97 (1975) 6609-6615.

[26] L.B. Kier, L.H. Hall, Molecular connectivity in structure-activity analysis, Research studies press: Letchworth, Hertfordshire, U K, 1986.

[27] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, GAMESS Version= 24 Mar 2007 (R1) from Iowa State University, J Comput Chem. 14 (1993) 1347-1363.

[28] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J Comput Chem. 30 (2009) 2785-2791.

[29] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, J Comput Chem. 28 (2007) 1145-1152.

[30] R. Huey, D.S. Goodsell, G.M. Morriss, A.J. Olson, Grid-based hydrogen bond potentials with improved directionality, Lett Drug Des Discov. 1 (2004) 178-183.

[31] R. Thomsen, M.H. Christensen, MolDock: a new technique for high-accuracy molecular docking, J Med Chem. 49 (2006) 3315-3321.

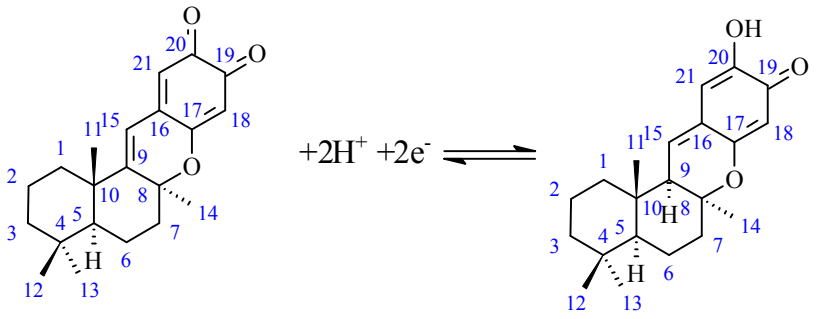
CHAPTER 5

Computational study on redox reaction of puupehenone in aqueous solution by density functional theory

5.1. Introduction

Lipoxygenases are non-heme iron containing oxidative enzymes, occurring in a number of plants and animals [1, 2]. These enzymes catalyze oxygenation of naturally occurring poly-unsaturated fatty acids (PUFAs) such as arachidonic acid and linoleic acid [3]. More importantly, lipoxygenases are involved in the regulation of inflammatory responses that can promote human disease. For example, human 5-lipoxygenase (5-HLO), human 12- lipoxygenase (12-HLO) and human 15- lipoxygenase (15-HLO) are involved in several diseases like asthma, arthritis, allergy, psoriasis, atherosclerosis and tumorigenesis [4-9].

The mechanism of lipoxygenase inhibition by inhibitors are classified into two groups, redox and nonredox inhibition. The redox active compound reduces lipoxygenase from ferric oxidation state to its inactive ferrous form where as allosteric inhibition can occure in nonredox mechanism [10, 11]. Puupehenone, a biologically active terpenoid, is a redox inhibitor of lipoxygenases most likely due to its relationship with the o-quinone (puupehedienone) shown in Figure 5.1 [12].



Puupehedienone (p) puupehenone (PH_2)

Figure 5.1. The two protons, two electron redox process of puupehenone

In this study, we have calculated the redox potential of puupehenone (PH_2) and its oxidized form i.e. puupehedienone (P) by using DFT method. In aqueous medium, puupehedienone and puupehenone are capable of forming hydrogen bonds with water. Hence the influence of hydrogen-bond on the redox reaction was also investigated.

5.2. Computational details

All quantum chemical calculations were performed using Firefly [13]. To get the redox potential, it is required to calculate the standard free energy change (ΔG^0) for reaction (5.1).

$$P\ (Sol) + 2H^+\ (Sol) + 2e \rightarrow PH_2\ (Sol) \quad \ldots\ldots\ldots\ (5.1)$$

The relation between ΔG^0 and absolute reduction potential is given by

$$E^0 = -\Delta G^0/nF \quad \ldots\ldots\ldots\ (5.2)$$

Where n is the number moles of electrons transferred in the reaction, which is equal to 2 for reaction (5.1), and F is the Faraday (96485coulomb/mole).

To get ΔG^0 from computation, the following thermodynamic cycle (Figure 5.2) is used. This thermodynamic cycle involved all the species in the reaction (5.1) from gas to solution phase. Using this thermodynamic cycle, ΔG^0 (total) can be written as

$$\Delta G^0 \text{ (total)} = \Delta G^0 \text{ (g)} + \Delta G^0 \text{ (solv, } PH_2\text{)} - \Delta G^0 \text{ (solv, P)} - 2\Delta G^0 \text{ (solv, } H^+\text{)} \quad \ldots\ldots\ldots \quad (5.3)$$

$$\Delta G^0 \text{ (total)}$$

$$P\,(aq) + 2H^+\,(aq) + 2e \rightarrow PH_2\,(aq)$$

$$\Delta G^0 \text{ (solv,P)} \uparrow \qquad \uparrow \Delta G^0 \text{ (solv,} 2H^+\text{)} \qquad \uparrow \Delta G^0 \text{ (solv,} PH_2\text{)}$$

$$P\,(g) + 2H^+\,(g) + 2e \rightarrow PH_2\,(g)$$

$$\Delta G^0 \text{ (g)}$$

Figure 5.2. Thermodynamic cycle for obtaining ΔG^0 (total)

Where ΔG^0 (g) is the Gibbs free energy of reaction (5.1) in gas phase. ΔG^0 (solv, PH_2), ΔG^0 (solv, P) and ΔG^0 (solv, H^+) are the solvation Gibbs free energy of PH_2, P and H^+ respectively. The standard Gibbs free energy of each molecule in the standard state at gas phase is obtained by equation (5.4) [14].

$$\Delta G^0_{(g)} = E_{0k} + ZPE + \Delta\Delta G_{0\rightarrow 298} \quad \ldots\ldots\ldots \quad (5.4)$$

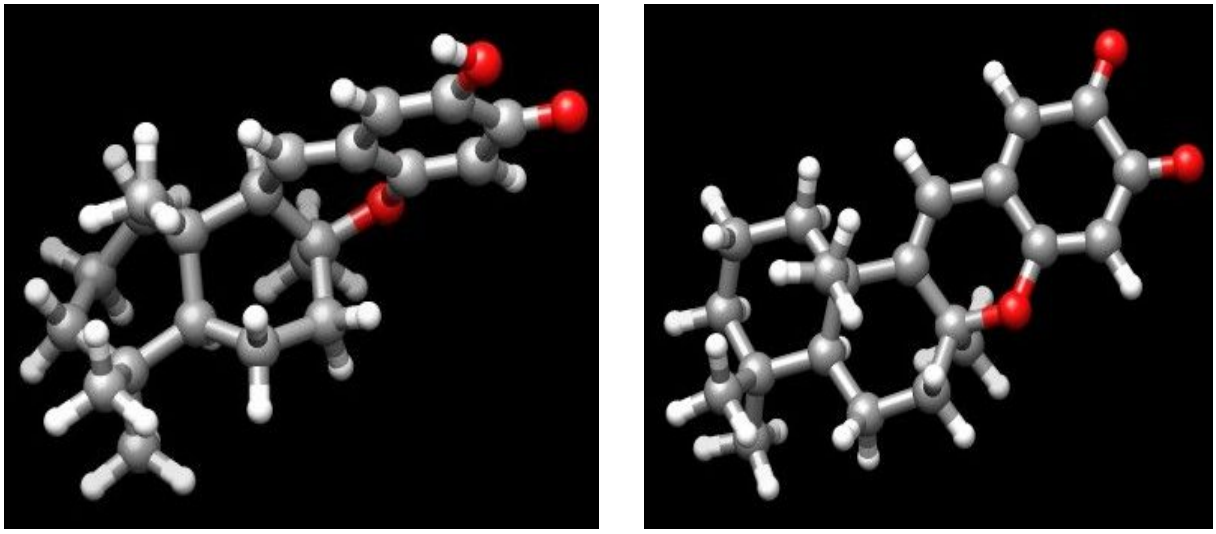
Where E_{0k} and ZPE are the energy at 0K and zero point energy respectively. $\Delta\Delta G_{0\rightarrow 298}$ is the Gibbs free energy change from 0 to 298K at 1 atm. An extra term RT ln(24.46) should be added in equation (5.4) to convert $\Delta G^0_{(g)}$ state from 1 atm to 1 M.

$$\Delta G^0_{(g)}\,(1M) = \Delta G^0_{(g)}\,(1\text{ atm}) + RT\ln(24.46) = \Delta G^{0\rightarrow *} \quad \ldots\ldots\ldots \quad (5.5)$$

The $\Delta G^0_{(solv)}$ can be calculated as the subtraction of the standard free energy of aqueous phase, $\Delta G^0_{(aq)}$ and gas phase, $\Delta G^0_{(g)}$.

$$\Delta G^0_{(solv)} = \Delta G^0_{(aq)} - \Delta G^0_{(g)} \ldots\ldots\ldots (5.6)$$

To calculate $\Delta G^0_{(g)}$, we optimized the molecular structure of P and PH_2 at the DFT- B3LYP/6-311G(d,p) level of theory separately (Figure 5.3). Frequency calculations were performed at the same level of theory and basis set to verify that structure lies in the global minima and obtains the free energy of P and PH_2.The standard free energy of electron is obtained by using its energy (3.720kJ.mol^{-1}) and entropy (0.022734 kJ mol^{-1}K^{-1}) at 298 K[15]. The reported value of Gibbs free energy of H^+ (g) to be -26.3 kJ.mol^{-1}[16].

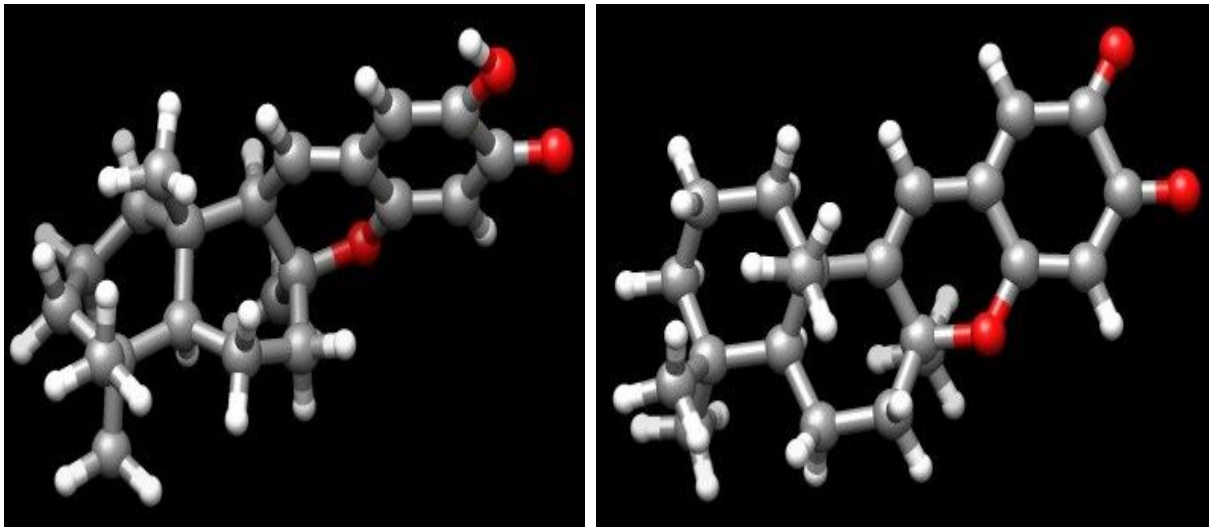


(a) (b)

Figure 5.3. Optimized geometries of PH_2 (a) and P (b) at B3LYP/6-311G(d,p) level in gas phase

In order to compute $\Delta G^0_{(aq)}$, the molecular structure of P and PH_2 were re-optimized by PCM model using water as a solvent at the same level of theory and basis set (Figure 5.4). Vibrational frequency calculations were performed to the optimized structures to get free energy of P and PH_2. We have used the literature value of -1104.6 kJ.mol^{-1} for ΔG^0 (solv, H^+) [17]. It should be mentioned that this value is the change in the standard Gibbs free energy of reaction (5.1) in

solution in the standard state of gas phase (1 atm)[18]. To obtain the standard free energy of relation (5.3) in solution (1 mol/L) from gas phase (1 atm), it is necessary to add $\Delta n\Delta G^{0\rightarrow *}$ to ΔG^{0} (total). In reaction (5.1) Δn is equal to -2 and $\Delta G^{0\rightarrow *}$ is 7.9 kJ.mol^{-1}.



(c) (d)

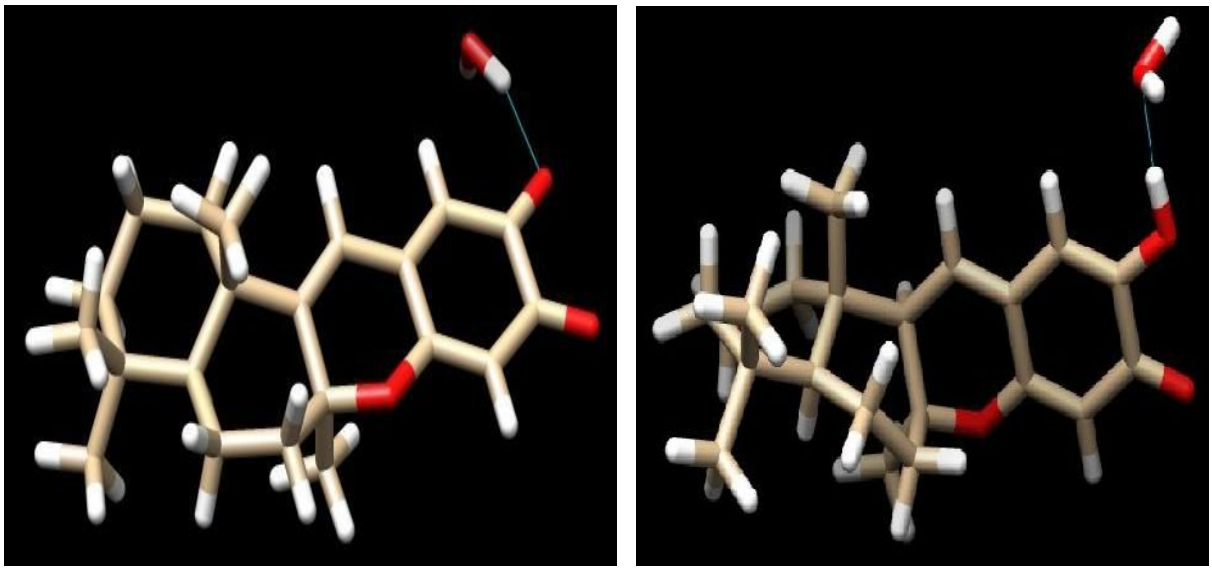
Figure 5.4. Optimized geometries of PH_2 (c) and P (d) at B3LYP/6-311G(d,p) level in water

To investigate the effect of H-bonding interactions on the redox potential of P/PH_2 system, we have microsolvated both puupehedienone (P) and puupehenone (PH_2) on the carbonyl or hydroxyl group at C-20 with one to three water molecule(s). The hydrated complexes were optimized at the B3LYP/6-31G(d) level of theory in gas phase and performed the frequency calculations for the optimized low energy structures to determine the true local minima. The interaction energy (ΔE), which is defined as ΔE= EM…n(w) - E_M - nEw, is calculated with each hydrated complex . Also, to predict the extra stability of the hydrated PH_2 complex than P complex, the difference in the interaction energy (ΔE_{diff}) and the difference in the solvation free energy (ΔG_{diff}) are calculated. ΔE_{diff} and ΔG_{diff} are given by:

$$\Delta E_{diff} = \Delta E(PH_2 \text{ complex}) - \Delta E(P \text{ complex}) \quad \ldots\ldots\ldots (5.7)$$

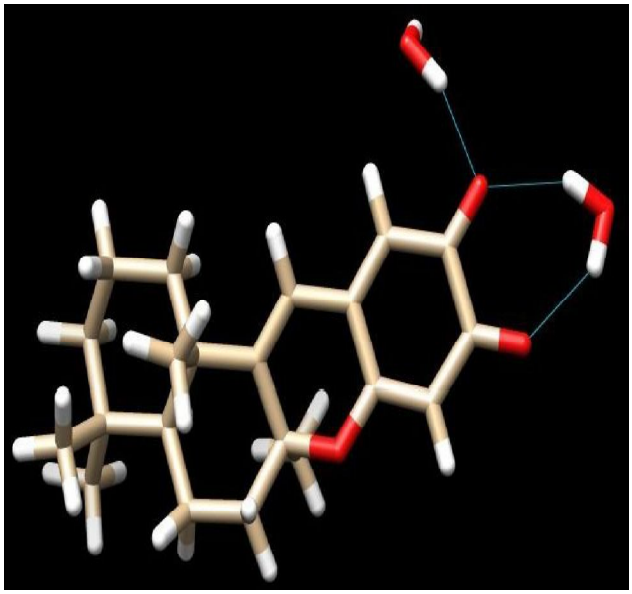
$$\Delta G_{diff} = \Delta G(solv, PH_2 \text{ complex}) - \Delta G(solv, P \text{ complex}) \quad \ldots\ldots\ldots (5.8)$$

Optimized geometries of complexes along with the interaction energy (ΔE) are given in Figure 5.5 to 5.7. The molecular plots were produced using the UCSF Chimera 1.9.

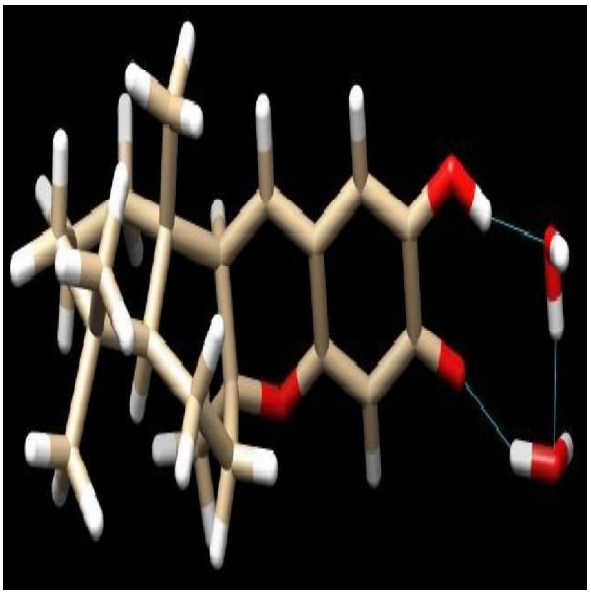


P-$(H_2O)_1$:ΔE=-39.04 PH_2-$(H_2O)_1$:ΔE=-43.00

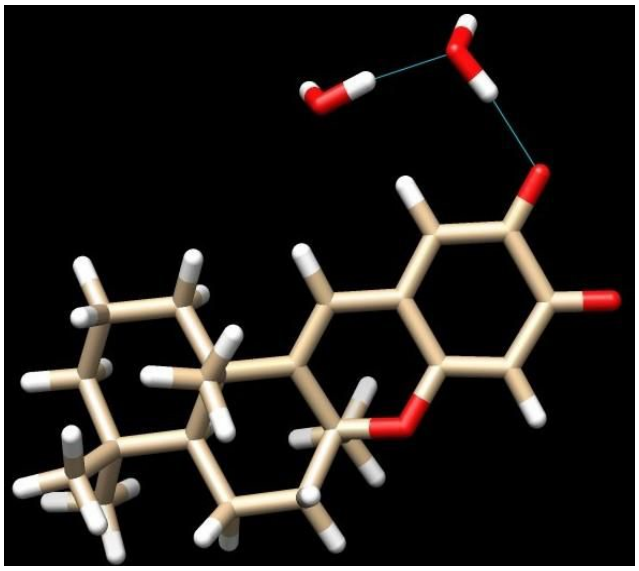
Figure 5.5. Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with one water molecule along with ΔE ($kJmol^{-1}$)



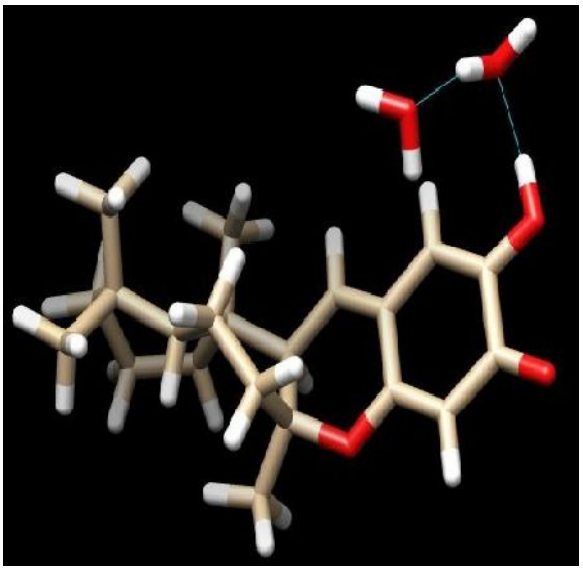
P-$(H_2O)_{2a}$:ΔE=-74.22



PH2-$(H_2O)_{2a}$:ΔE=-140.64



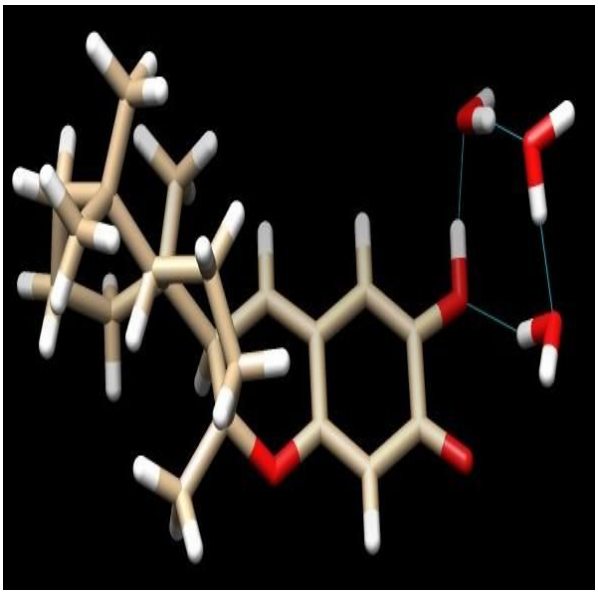
P-$(H_2O)_{2b}$:ΔE=-92.81



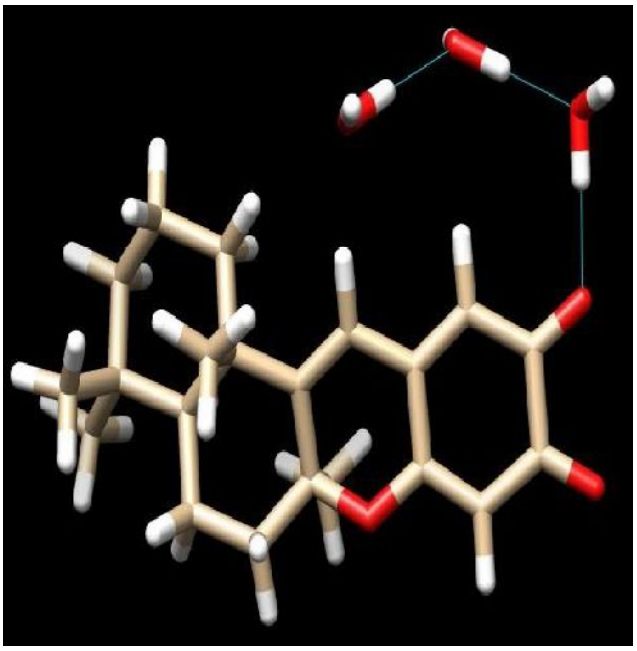
PH_2-$(H_2O)_{2b}$:ΔE=-94.89

Figure 5.6. Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with two water molecules at different configurations along with ΔE ($kJmol^{-1}$)

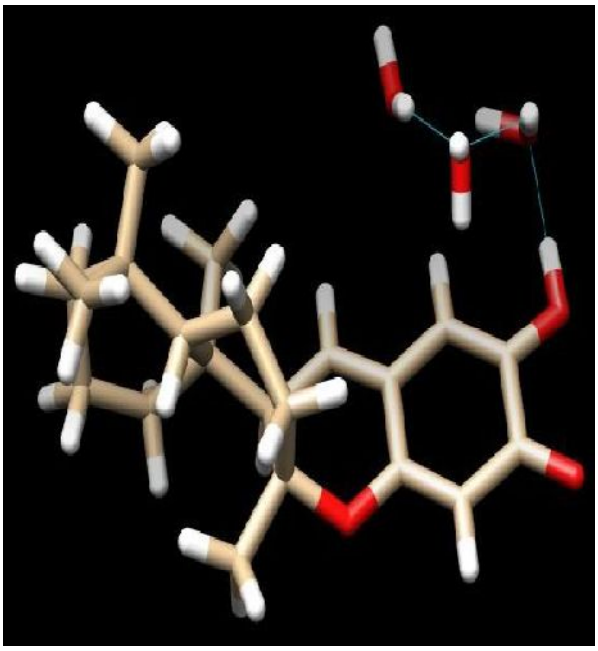
P-$(H_2O)_{3a}$:ΔE=-157.57



PH_2-$(H_2O)_{3a}$:ΔE=-190.72



P-$(H_2O)_{3b}$:ΔE=-152.19



PH_2-$(H_2O)_{3b}$:ΔE=-146.12

Figure 5.7. Optimized geometries of puupehedienone (P) and puupehenone (PH_2) with three water molecules at different configurations along with ΔE ($kJmol^{-1}$)

5.3. Results and discussion

In redox reaction, the thermodynamic cycle linking the process in the gas phase with that in solvent can be used to evaluate the reaction free energy. The free energy of the studied molecules and the redox potential are tabulated in Table 5.1. The standard free energy change of reaction (5.1) in solution, ΔG^0 (total) is equal to -785.481907 kJ.mol-1 (at 6-311G** level). Using the value of ΔG^0 (total) and relation (5.2), the absolute reduction potential of P/PH_2 has been calculated in the order 4.07V. The absolute redox potential of SHE is 4.44V, so the E^0 value of P/PH2 system is in the order -0.370V.

Table 5.1. The Gibbs free energy of P and PH_2 in gas phase and solution, together with solvation free energies of species calculated at 6-31G* and 6-311G** basis set.

Basis set used	6-31G*		6-311G**	
Compounds	P	PH_2	P	PH_2
B3LYP Free Energy (g)/ (a.u)	-1040.711281	-1041.855358	-1040.994739	-1042.147231
B3LYP Free Energy (aq)/ (a.u)	-1040.729337	-1041.877011	-1041.015038	-1042.171998
ΔG^0 (solv)/kJmol^{-1}	-47.406023	-56.849946	-53.295019	-65.025752
ΔG^0 (g)/kJmol^{-1}	-2945.057593		-2967.151174	
ΔG^0 (total)/kJmol^{-1}	-761.101516		-785.481907	
Absolute redox potential	3.944		4.070	
E^0/V	-0.496		-0.370	

The reduction potential of the ferric ion in soybean lipoxygenase-1 versus normal hydrogen electrode has been estimated to be in excess of 0.5 V [19]. Hence the inhibitors in this class must

be weak reducing agents to reduce the ferric ion to the inactive ferrous state and puupehenone may serve this purpose well.

The next objective of our study is to investigate the effect of hydrogen bonds on redox potentials. Figure 5.5 depicts the optimized geometry of puupehedienone (P) and puupehenone (PH_2) with one water molecule at B3LYP/6-31G* level of theory. Both puupehedienone (P) and puupehenone (PH_2) form one hydrogen bond with interaction energies are -39.04 $kJmol^{-1}$ and -43 $kJmol^{-1}$ respectively. Hence the difference in the interaction energy (ΔE_{diff}) is -3.96 $kJmol^{-1}$. Figure 5.6 depicts the optimized geometry of puupehedienone (P) and puupehenone (PH_2) with two water molecules at different configurations and it was found that the difference in the interaction energy is greater in P-$(H_2O)_{2a}$ /PH_2-$(H_2O)_{2a}$ than P-$(H_2O)_{2b}$ /PH_2-$(H_2O)_{2b}$. Figure 5.7 depicts the optimized geometry of puupehedienone (P) and puupehenone (PH_2) with three water molecules at different configurations. The difference in the interaction energy of P-$(H_2O)_{3a}$ /PH_2-$(H_2O)_{3a}$ and P-$(H_2O)_{3b}$ /PH_2-$(H_2O)_{3b}$ complexes are -33.15 $kJmol^{-1}$ and 6.07 $kJmol^{-1}$ respectively. Difference in the interaction energies (ΔE_{diff}), difference in solvation free energy (ΔG_{diff}), ΔG^0 (total) and absolute E^0 are displayed in Table 5.2. The trends (cf. Table 5.2) suggest that ΔG^0 (total) or absolute E^0 of P/PH_2 couple is highly dependent on ΔE_{diff} and ΔG_{diff}.

Table 5.2. Difference in the interaction energies (ΔE_{diff}), difference in solvation free energy (ΔG_{diff}), ΔG^0 (total) and absolute E^0 of puupehedienone (P) and puupehenone (PH_2) complexes with different water molecules

Couple	ΔE_{diff} kJmol^{-1}	ΔG_{diff} kJmol^{-1}	ΔG^0 (total) kJmol^{-1}	Absolute E^0/V
P/PH_2-$(H_2O)_1$	-3.96	-4.94	-756.598784	3.921
P/PH_2-$(H_2O)_{2a}$	-66.42	-54.59	-806.247371	4.178
P/PH_2-$(H_2O)_{2b}$	-2.08	-5.47	-757.126509	3.924
P/PH_2-$(H_2O)_{3a}$	-33.15	-31.24	-782.895789	4.057
P/PH_2-$(H_2O)_{3b}$	6.07	-5.73	-445.923501	3.865
P/PH_2-(PCM)	-4.25	-9.44	-761.101516	3.944

5.4. Conclusion

The standard reduction potential of Fe^{+3}/ Fe^{+2} couple is 0.77V and it is expected that the reduction potentials of lipoxygenases are lower than this value. However the exact reduction potential of the ferric ion in 5-HLO, 12-HLO and 15-HLO are not known. Hence this study helps to predict the E^0 value of different lipoxygenases. Since puupehedienone and puupehenone are capable of forming hydrogen bond with water, the absolute value of E^0 of P/PH_2 couple is highly dependent on ΔE_{diff} and ΔG_{diff}.

5.5 References

[1] E.I. Solomon, J. Zhou, F. Neese, E.G. Pavel, New insights from spectroscopy into the structure/function relationships of lipoxygenases, Chem Biol. 4 (1997) 795-808.

[2] H.W. Gardner, Biological roles and biochemistry of the lipoxygenase pathway, HortScience. 30 (1995) 197-205.

[3] R. Wisastra, F.J. Dekker, Inflammation, cancer and oxidative lipoxygenase activity are intimately linked, Cancers. 6 (2014) 1500-1521.

[4] L.A. Dailey, P. Imming, 12-Lipoxygenase: classification, possible therapeutic benefits from inhibition, and inhibitors, Curr Med Chem. 6 (1999) 389-398.

[5] V.E. Steele, C.A. Holmes, E.T. Hawk, L. Kopelovich, R.A. Lubet, J.A. Crowell, C.C. Sigman, G.J. Kelloff, Lipoxygenase inhibitors as potential cancer chemopreventives, Cancer Epidemiol Biomarkers Prev. 8 (1999) 467-483.

[6] B. Samuelsson, S.E. Dahlen, J.A. Lindgren, C.A. Rouzer, C.N. Serhan, Leukotrienes and lipoxins: structures, biosynthesis, and biological effects, Science. 237 (1987) 1171-1176.

[7] X.Z. Ding, W.G. Tong, T.E. Adrian, 12-Lipoxygenase metabolite 12(S)-HETE stimulates human pancreatic cancer cell proliferation via protein tyrosine phosphorylation and ERK activation, Int J Cancer. 94 (2001) 630-636.

[8] J.A. Cornicelli, B.K. Trivedi, 15-Lipoxygenase and its inhibition: a novel therapeutic target for vascular disease, Curr Pharm Des. 5 (1999) 11-20.

[9] U.P. Kelavkar, J.B. Nixon, C. Cohen, D. Dillehay, T.E. Eling, K.F. Badr, Overexpression of 15-lipoxygenase-1 in PC-3 human prostate cancer cells increases tumorigenesis, Carcinogenesis. 22 (2001) 1765-1773.

[10] C. Kemal, P. Louis-Flamberg, R. Krupinski-Olsen, A.L. Shorter, Reductive inactivation soybean lipoxygenase-1 by catechols: a possible mechanism for regulation of lipoxygenase activity, Biochemistry. 26 (1987) 7064-7072.

[11] R. Mogul, E. Johansen, T.R. Holman, Oleyl sulfate reveals allosteric inhibition of soybean lipoxygenase-1 and human 15-lipoxygenase, Biochemistry. 39 (2000) 4801-4807.

[12] T. Amagata, S. Whitman, T.A. Johnson, C.C. Stessman, C.P. Loo, E.L. Jonclardy, P. Crews, T.R. Holman, Exploring sponge-derived terpenoids for their potency and selectivity against 12-human, 15-human, and 15-soybean lipoxygenases, J Nat Prod. 66 (2003) 230-235

[13] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[14] Y.H. Jang, W.A. Goddard III, K.T. Noyes, L.C. Sowers, S. Hwang, D.S. Chung, p*K*a Values of guanine in water: density functional theory calculations combined with poisson-boltzmann continuum-solvation model, J Phys Chem B. 107 (2003) 344-357.

[15] J.E. Bartmess, Thermodynamics of the electron and the proton, J Phys Chem. 98 (1994) 6420-6424.

[16] M.D. Liptak, K.G. Gross, P.G. Seybold, S. Feldgus, G.C. Shields, Absolute pK(a) determinations for substituted phenols, J Am Chem Soc. 124 (2002) 6421-6427.

[17] P. Winget, E.J. Weber, C.J. Cramer, D.G. Truhlar, Computational electrochemistry: aqueous one-electron oxidation potentials for substituted anilines, Phys Chem Chem Phys. 2 (2000) 1231-1239.

[18] N.S. Babu, Computational studies for oxidation reduction reactions of cinnoline - 4(1H)-one, in aqueous phase by density functional theory, BJAST. 4 (2014) 465-476.

[19] M.J. Nelson, D.G. Batt, J.S. Thompson, S.W. Wright, Reduction of the active-site iron by potent inhibitors of lipoxygenases, J Biol Chem. 266 (1991) 8225-8229.

CHAPTER 6

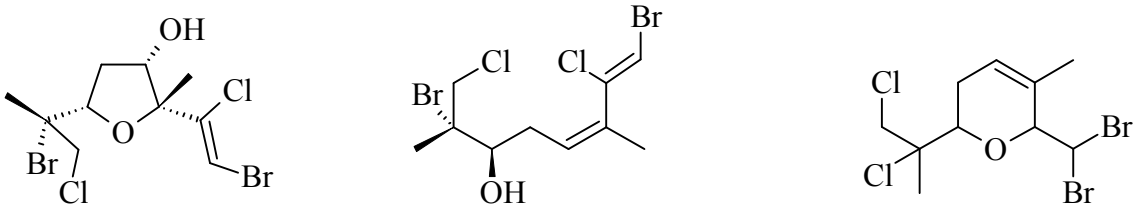
A theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum

6.1. Introduction

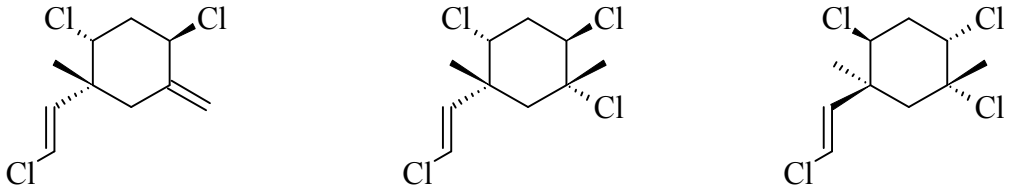
Marin algae considered to have wide applications, such as antibacterial, antiviral, insecticidal and antitumor activities [1-3]. Nine halogenated monoterpenoids furoplocamioid C (1a), pre furoplocamioid (1b), pirene (1c), and the cyclohexanes (1d-1i), including mertensene (1g) and violacene (1h) were isolated from the red alga plocamium cartilagineum have exhibited notable cytotoxic activity [4]. The cytotoxic effects of these compounds have been tested on the tumor cell lines CT26 (murin colon adenocarcinoma), SW480 (human colon adenocarcinoma), HeLa (human cervical adenocarcinoma) and SkMel28 (human malignant melanoma) with several multidrug resistance mechanism against the mammalian non tumor cell line CHO (Chinese hamster ovary cells) [5].In this work, we have tried to give an explanation about the cytotoxic activity of the studied molecules using electronic properties such as the highest occupied molecular orbital (HOMO) energies, lowest unoccupied molecular orbital (LUMO) energies, LUMO-HOMO energy gap, dipole moment and stereo chemical structure.

6.2. Computational studies

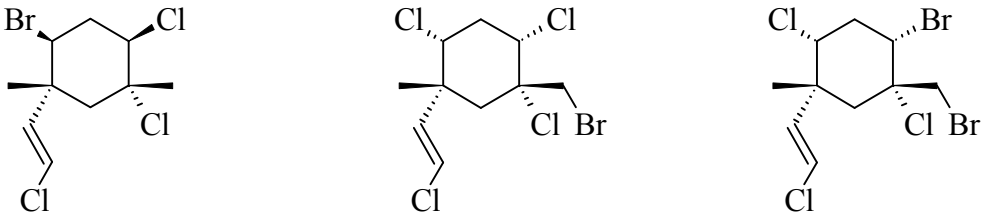
The structures of the molecules (Figure 6.1) under investigation were constructed using ACD/ChemSketch, version 12.01 [6]. All quantum chemical calculations were performed with the Firefly [7]. The ground-state geometries and electronic properties of the studied molecules have been determined at the B3LYP/6-31G*.



1a 1b 1c



1d 1e 1f



1g 1h 1i

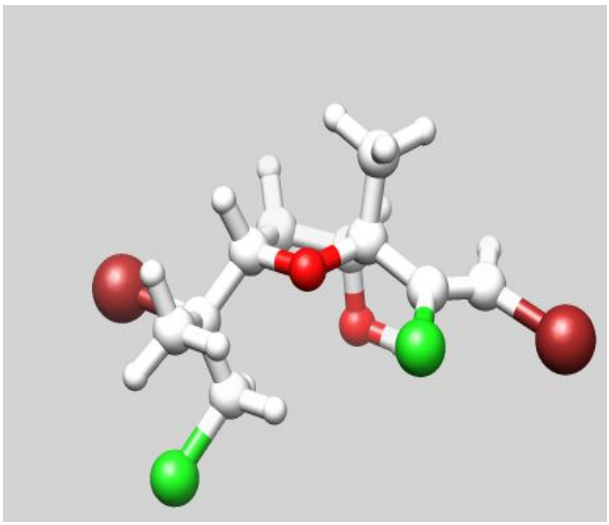
Figure 6.1. Chemical structures of furoplocamioid C (1a), prefuroplocamioid (1b), pirene (1c), cyclohexanes (1d-1i), including mertensene (1g), and violacene (1h).

6.3. Results and discussion

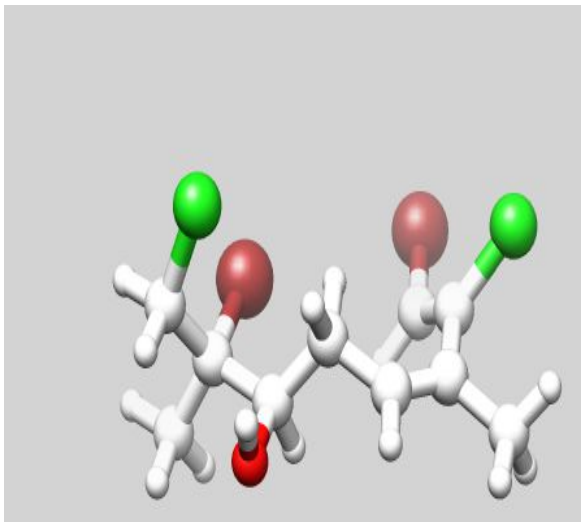
The minimal inhibitory concentration (MIC) of the studied compounds which produce a cytotoxic effect on the different cell lines along with their molecular electronic properties are summarized in Table 6.1 and their optimized geometry structures are illustrated in Figure 6.2.

Table 6.1. Minimal inhibitory concentration (MIC) and selected molecular electronic properties of the studied compounds

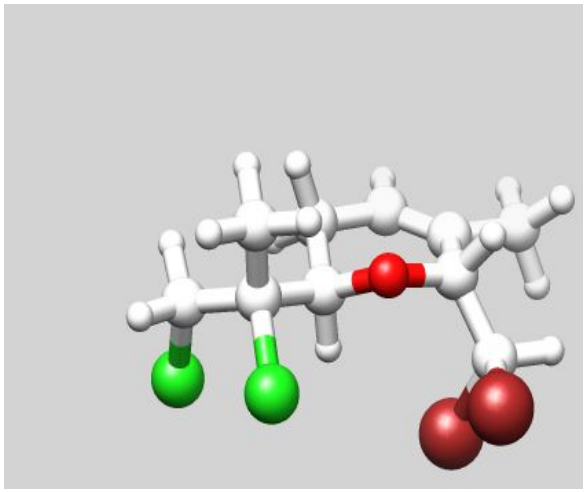
Comp. No.	$MIC_{(CHO)}$ µM	$MIC_{(CT26)}$ µM	$MIC_{(SW480)}$ µM	$E_{(HOMO)}$ eV	$E_{(LUMO)}$ eV	ΔE_{gap} eV	Dipole (D)	Total energies (hartree)
1a	126	63	126	-7.0668	-0.9415	6.1253	4.89	-6603.7885
1b	132	66	66	-6.479	-0.8626	5.6164	3.99	-6528.5601
1c	262	262	131	-6.9307	-0.9225	6.0082	4.94	-6528.583
1d	3.3	6.52	3.3	-6.9389	-0.283	6.6559	2.19	-1769.4516
1e	23	181	5.7	-6.7321	-0.6585	6.0736	2.98	-2230.2748
1f	362	362	362	-7.0341	-0.2503	6.7838	0.8	-2230.2751
1g	39	78	78	-7.1158	-0.7293	6.3865	1.54	-4341.8244
1h	141	141	141	-6.9961	-0.6694	6.3267	4.95	-4801.4049
1i	63	125	125	-6.9743	-0.8735	6.1008	4.83	-6912.9557



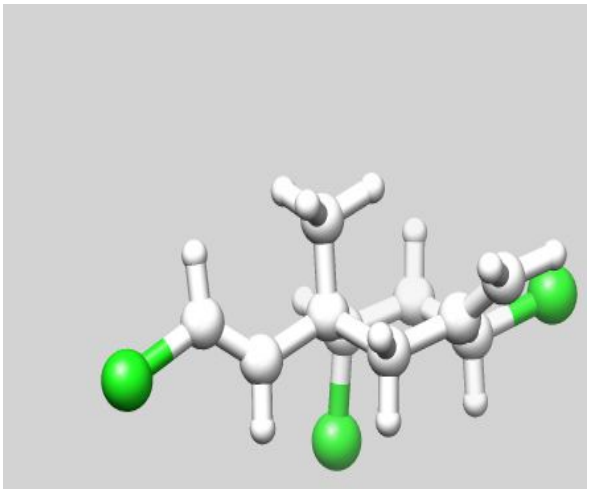
1a



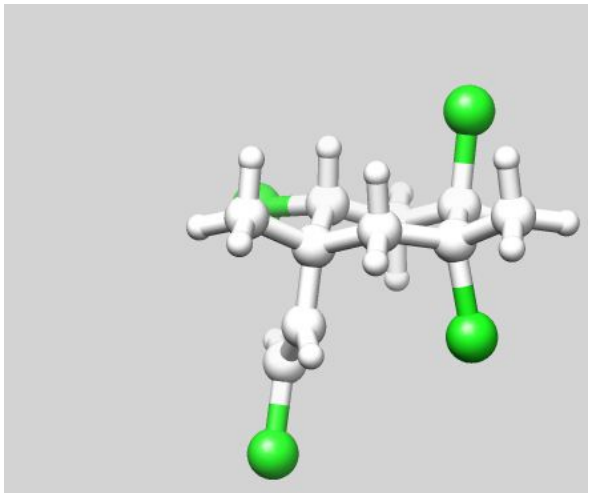
1b



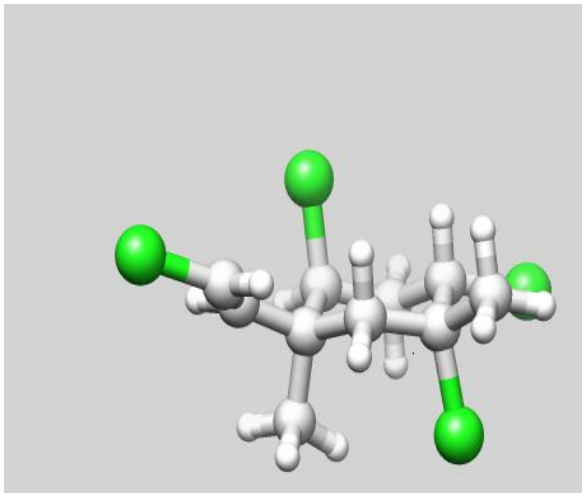
1c



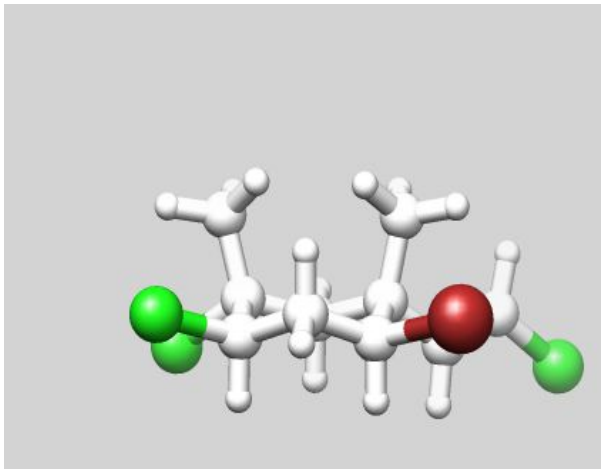
1d



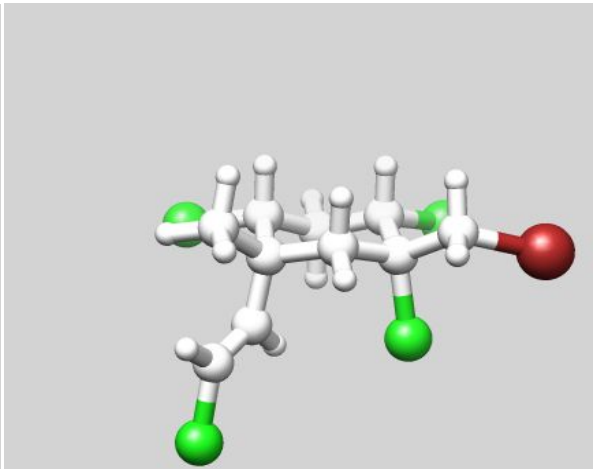
1e



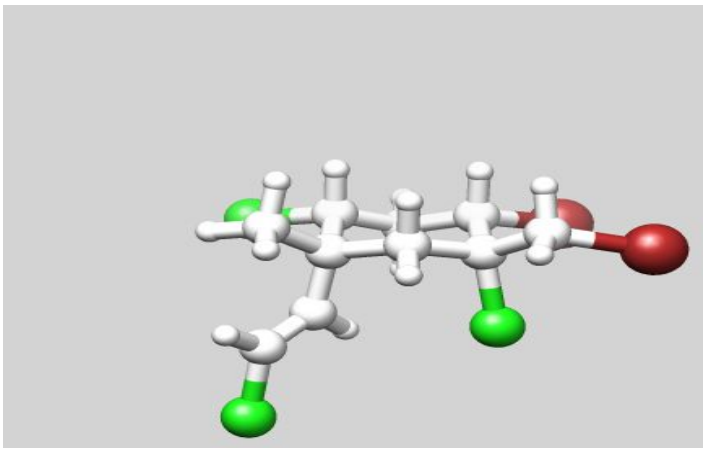
1f



1g



1h



1i

Figure 6.2. Optimized structures of studied molecules obtained by B3LYP/6-31G* level.

Compounds 1a, 1b, 1c and 1e had selective activity against cancer cells versus CHO cells. Compounds 1a and 1c exhibited selective cytotoxicity to CT26 and SW480 cell lines respectively, with MIC values of 63 µM and 131 µM. Compound 1b produced a selective cytotoxic effect on CT26 and SW480 cells with MIC value of 66 µM. Interestingly, compound 1e was the most active and exhibited cytotoxicity against SW480 cell lines with MIC value of 5.70 µM [5].

The energies of HOMO and LUMO of the inhibitor molecule are important. A high energy HOMO means weakly held electrons while a low energy LUMO indicates a more stable orbital for electrons. A molecule with a high HOMO may act as a donor while with a low LUMO may act as an acceptor. The ΔE_{gap} provides a measure for the stability of the formed complex on the metal surface. Thus the complex stability increases with decreasing the value of ΔEgap. Compounds 1a, 1b, 1c, and 1e have lower value of ΔE_{gap} (ranges from 5.6 to 6.1) compared to the compounds 1d, 1f, 1g and 1h. A cross correlation matrix (Table 6.2) between electronic descriptors and the MIC (SW480) values demonstrates ΔE_{gap} is positively correlated with

$MIC_{(SW480)}$. But $E_{(HOMO)}$, $E_{(LUMO)}$, and dipole moment are weakly correlated with $MIC_{(SW480)}$. Thus the stereo chemical features of these compounds also play an important role towards activity.

Table 6.2. Correlation matrix of $MIC_{(SW480)}$ and the electronic descriptors for the studied compounds

	$MIC_{(SW480)}$	$E_{(HOMO)}$	$E_{(LUMO)}$	ΔE_{gap}	dipole(D)
$MIC_{(SW480)}$	1.000	-0.389	0.279	0.423	-0.235
$E_{(HOMO)}$	-0.389	1.000	-0.176	-0.689	0.168
$E_{(LUMO)}$	0.279	-0.176	1.000	0.835	-0.779
ΔE_{gap}	0.423	-0.689	0.835	1.000	-0.668
dipole(D)	-0.235	0.168	-0.779	-0.668	1.000

Compound 1f is a diastereomer of 1e though the cytotoxic activity of 1f is lower than 1e. This is due to high gap energy (6.78 eV) and low dipole moment (0.80 D) of the molecule. In compound 1f, the resultant bond moment of the –Cl and -CH=CHCl groups at one side of the molecule is in the opposite direction to the resultant moment of the two –Cl groups on the other side. Hence, the bond moment is nearly cancelled out. Again the LUMO energy of the compound 1f is high compared to the other. These result suggest that 1f should be a lesser charge acceptor and hence less potent than other studied compounds. This is also an agreement with the experimental results.

6.4. Conclusion

Compounds 1a, 1b, 1c and 1e had selective activity against cancer cells versus CHO cells. Their selective cytotoxic activity depends mainly on the gap energy and the stereo chemical features of these compounds.

6.5. References

[1] H. Harada, Y. Kamei, Dose-dependent selective cytotoxicity of extracts from marine green alga, Cladophoropsis vaucheriaeformis, against mouse leukemia L1210 cells, Biol Pharm Bull. 21 (1998) 386-389.

[2] K.E. Apt, P.W. Behrens, Commercial developments in microalgal biotechnology, J Phycol. 35 (1999) 215-26.

[3] K. Harada, M. Suomalainen, H. Uchida, H. Masui, K. Ohmura, J. Kiviranta, M.L. Niku-Paavola, T. Ikemoto, Insecticidal compounds against mosquito larvae from Oscillatoria agardhii strain 27, Environ Toxicol. 15 (2000) 114-119.

[4] V.H. Argandoña, J. Rovirosa, A. San-Martín, A. Riquelme, A.R. Díaz-Marrero, M. Cueto, J. Darias, O. Santana, A. Guadaño, A. González-Coloma, Antifeedant effects of marine halogenated monoterpenes, J Agric Food Chem. 50 (2002) 7029-7033.

[5] C.D. Inés, V.H. Argandoña, J. Rovirosa, A. San-Martín, A.R. Díaz-Marrero, M. Cueto, A.G. Coloma, Cytotoxic activity of halogenated monoterpenes from Plocamium cartilagineum, Z Naturforsch. 59c (2004) 339-344.

[6] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[7] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

CHAPTER 7

Molecular docking and DFT based QSAR study on oleanolic acid derivatives as Protein–tyrosine phosphatase 1B inhibitors

7.1. Introduction

Protein–tyrosine phosphatase 1B (PTP1B) is an attractive target for the treatment of type 2 diabetes and is found in a wide variety of human tissues [1, 2]. The removal of the phosphoryl group from phosphotyrosine residue(s) in protein substrates by Protein–tyrosine phosphatases (PTPs) and the reverse action by protein tyrosine kinases is a common mechanism for the control of biological pathways [3-5].

PTP1B is the prototypical intracellular PTPs serves as a key negative regulator of insulin signaling pathway and is overexpressed in human breast cancer [6, 7]. Knock-out studies suggest that the lack of PTP1B would result in increased insulin sensitivity and suppression of weight gain in mice [8].

Oleanane type triterpenes possess exciting pharmacological properties, including the anti-inflammatory, hypolipidemic, antioxidant, antidiabetic, microbicid and antiatherosclerotic actions [9-11]. They interfere in the neuro degenerative disorders and in the development of different types of cancer [12]. Inhibition of PTP1B by oleanolic acid improves insulin sensitivity and stimulates glucose uptake [13]. Molecular docking studies indicate that triterpenes bind in the aryl phosphate binding site not in the catalytic site [14, 15].

In this article, we have performed QSAR study followed by molecular docking with a series of oleanolic acid derivatives to explore the important properties of potent and selective PTP1B inhibitors.

7.2. Materials and methods

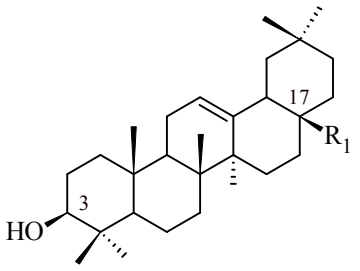
7.2.1. Molecular docking of the oleanolic acid derivatives to PTP1B enzyme

A total of 35 oleanolic acid derivatives published from the literature were used for the molecular docking and QSAR studies [16]. The initial structures of 35 compounds used in this study were generated by ChemSketch [17]. The structure coordinates of PTP1B in complex with OAI (1C83.pdb) were obtained from the RCSB protein data bank (www.rcsb.org). The oleanolic acid derivatives were docked into the active pocket of the enzyme by using docking program Autodock 4.0 [18]. Initially the structure of the ligands has been optimized with Austin Model 1 (AM1) parameterization and the hydrogen atoms were added to the enzyme. The Lamarckian genetic algorithm (LGA) was applied to search for the best conformers. A grid map with 60x50x40 points and 0.375 Å spacing was used in Autogrid program to evaluate the binding energies between the inhibitors and PTP1B. The grid centre was set at the active site position 47.411, 9.703 and 4.79 and the default settings were used. For each compound ten docking poses saved and ranked by binging energy. The lowest free energy conformation was chosen for analyzing the type of interactions. Visualization of the protein-ligand complex was performed using Molegro molecular viewer software [19]. The lowest energy geometry of the inhibitors obtained from docking was used for the QSAR study.

7.2.2. Descriptors and data set for QSAR

The biological property of this data set is reported as IC_{50} (μM) values. This value was changed to the minus logarithmic scale [pIC_{50}] and used for subsequent QSAR analysis as the response variable. Structural details of the 35 compounds and their biological activity are listed in Table 7.1. We attempted several descriptors (data not shown) and it is found that binding energy (EB), HOMO energy (EH), LUMO energy (EL), dipole moment (μ), molar refractivity (MR), molar volume (MV), solvent accessible surface area (SASA) and the octanol/water partition coefficient (logP) can better represent the biological activity of the selected compounds.

Table 7.1. Structural feature of oleanolic acid and its derivatives having PTP1B inhibitory activity



Comp no	R_1	IC_{50} (μM)
1	COOH	3.37
2	$(CH_2)_2$-COOH	2.10
3	$(CH_2)_4$-COOH	1.33
4	$(CH_2)_8$-COOH	0.78
5	$CONH_2$	4.76
6	$(CH_2)_{10}$-COOH	0.72

Table 7.1 (continued)

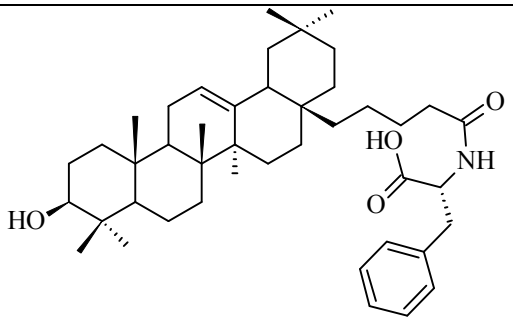
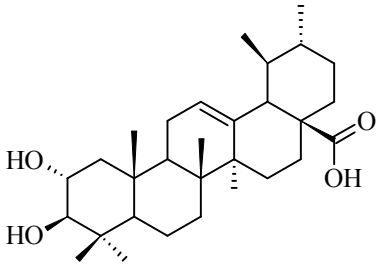
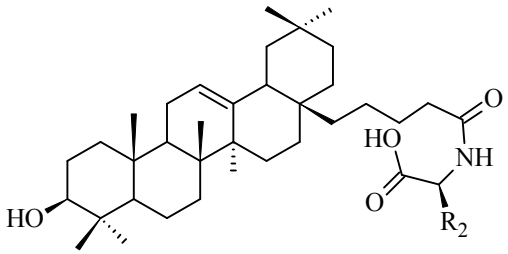
Comp no	R_1	IC_{50} (μM)
7*	COOMe	4.44
8*	$(CH_2)_{12}$-COOH	0.59
9		0.74
10*		5.49

Table 7.1 (continued)



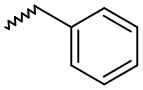
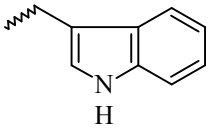
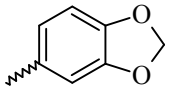
Comp. no.	R_2	IC_{50} (μM)
11		0.57
12		0.59
13	$(CH_2)_2$-SMe	0.55
14	2-Cl-Ph	0.56
15	3-Cl-Ph	0.51
16	4-Cl-Ph	0.61
17	4-F-Ph	0.57
18	2-Me-Ph	0.55
19	4-NO_2-Ph	0.45
20	2-OMe-Ph	0.53
21	3-OMe-Ph	0.52
22	4-OMe-Ph	0.60
23		0.44

Table 7.1 (continued)

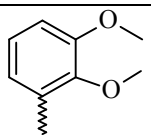
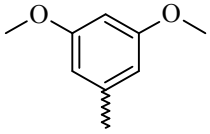
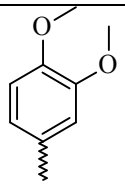
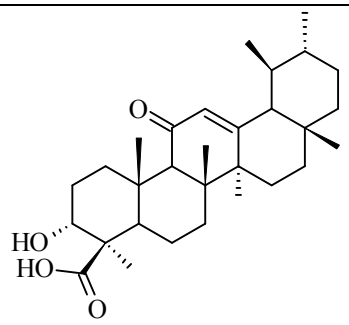
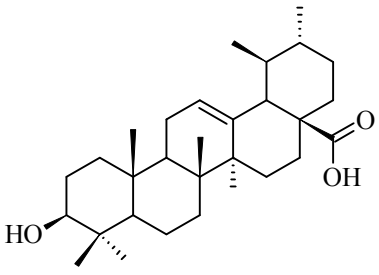
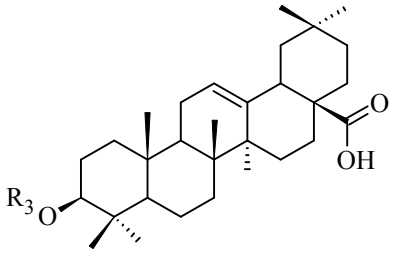
Comp. no.	R_2	IC_{50} (μM)
24		0.66
25		0.63
26*		0.82
27*		8.04
28*		3.08

Table 7.1 (continued)



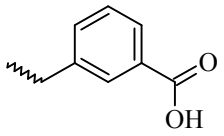
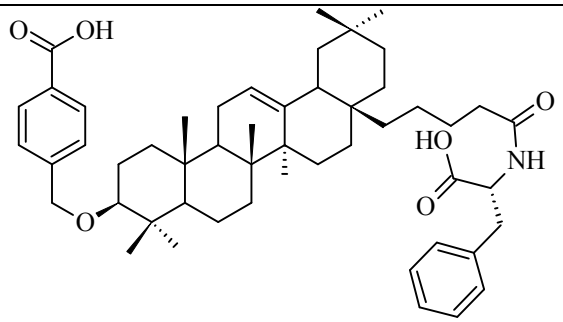
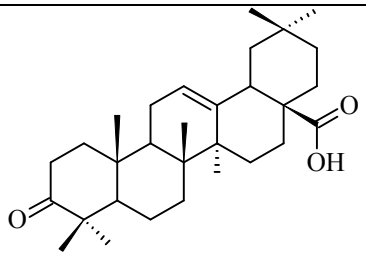
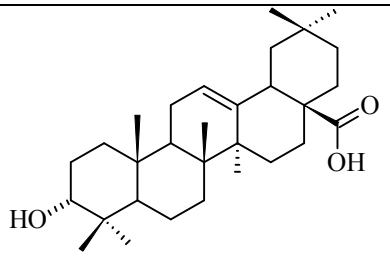
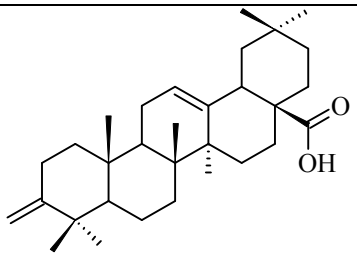
Comp. no.	R_3	IC_{50} (μM)
29		0.62
30	COCOOH	2.86
31	$COCH_2C(Me)_2COOH$	2.33
32*		0.15
33		5.32

Table 7.1 (continued)

34		5.05
35		2.85

Ph=phenyl, Me=methyl, Et=ethyl

*indicates test set compounds

The quantum chemical properties (EH, EL, μ) of the studied molecules have been determined by DFT/B3LYP calculation and the basis set 6-31G* was used. All quantum chemical calculations were performed with the Firefly [20]. Molar refractivity (MR), molar volume (MV) and partition coefficient (logP) were determined using ChemSketch software [17]. The binding energies (EB) of different ligands obtained from the docking study and solvent accessible surface area (SASA) of different inhibitors were calculated by Autodock Tools 1.5.6 [21].

7.2.3. Statistical methods

Multiple linear regression (MLR) analysis was used to build up QSAR models. Different combinations of parameters were tried to develop these models. Statistical qualities of MLR equations were judged by parameters like correlation coefficient (R), square of the correlation coefficient (R^2), cross validated coefficient (R^2_{cv}), standard deviation of the regression (S),

Fischer statistics (F) and quality factor (Q). MLR program written by ourselves in Fortran-77 is used.

7.3. Results and discussion

The binding energies of 35 ligands are ranges between -6.04 and -12.43 kcal/mol. The docking study shows both polar (TYR20, GLN21, ARG24, SER28, TYR46, ASP48, ASP181, ARG254, GLN262, THR263) and non polar (ALA27, VAL49, PHE182, ALA217, ILE219, MET258, GLY259) amino acids make important interactions to the inhibitors. Most of the ligands can form hydrogen bonds with ARG24 and/or ARG254.

Oleanolic acid (ligand 1) was used as a model drug (Figure 7.1a). The –COOH group at C-17 forms two hydrogen bonds with ARG24 (1.885Å) and ARG254 (1.901Å). Substitution of –COOH group by $–CONH_2$ and –COOMe results ligands 5 and 7 have lower biological activities. This is due to the fact that ligand 1 has higher –EB compared to ligands 5 and 7. Again the $–CONH_2$ group of ligand 5 (Figure 7.1b) and –COOMe group of ligand 7 (Figure 7.1c) do not make any hydrogen bond interaction with the enzyme.

The biological activity increases with increasing the carbon chain length at C-17 in ligands 2, 3, 4, 6 and 8. Except ligand 3, binding energy decreases with increasing chain size but their lipophilic efficiency increases. Again compound 8 has lower value of ΔEgap compared to the compounds 2, 3, 4 and 6 which suggest that complex formed between enzyme and ligand 8 (Figure 7.1d) is more stable than other. Compound 9 is an isomer of 11 though the biological activity of 9 is lower than 11. This is due to the ligand 9 has lower -EB than ligand 11(Figure 7.1e).

For the compounds in the high bioactive range, such as compounds 11 to 26 (IC_{50}<1 μM), there exists hydrogen bond(s) between amide backbone (especially with ARG24 and/or ARG254) and – $(CH_2)_4CONHCH(R_2)COOH$ group. Ligands 29, 30 and 31 are obtained from compound 1 by the substitution at the C-3 position and have greater biological activity. The biological activity of compound 29 (Figure 7.1f) is greater than 30 and 31 due to higher lipophilic efficiency.

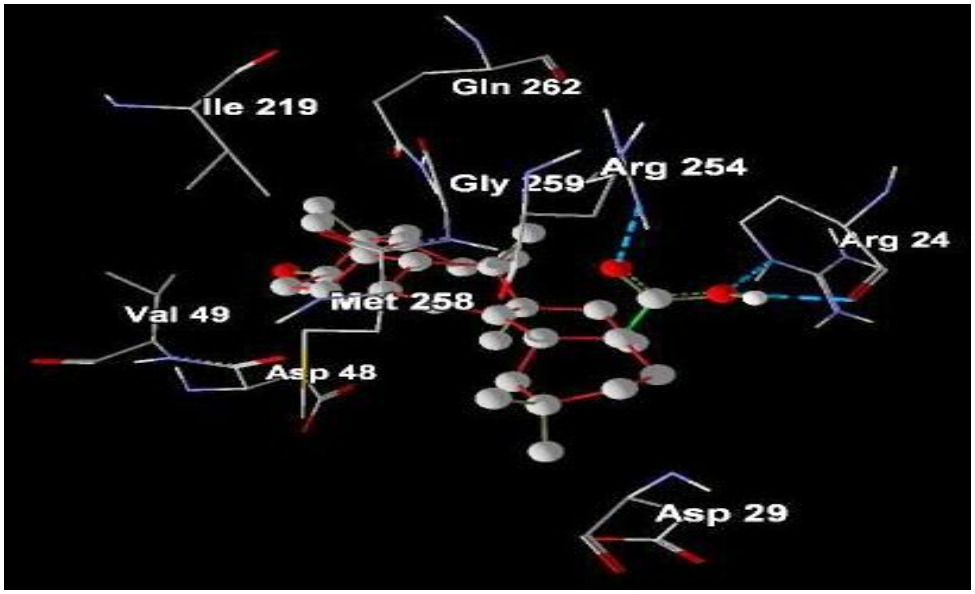


Figure 7.1a. Docked conformation of ligand 1 along with the important amino acid residues of PTP1B

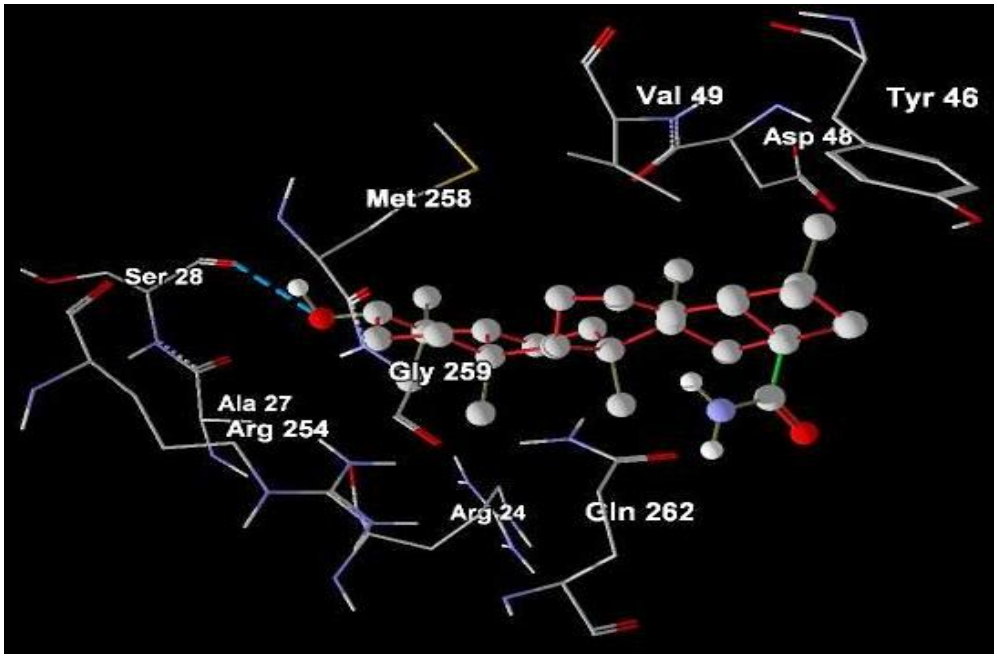


Figure 7.1b. Docked conformation of ligand 5 along with the important amino acid residues of PTP1B

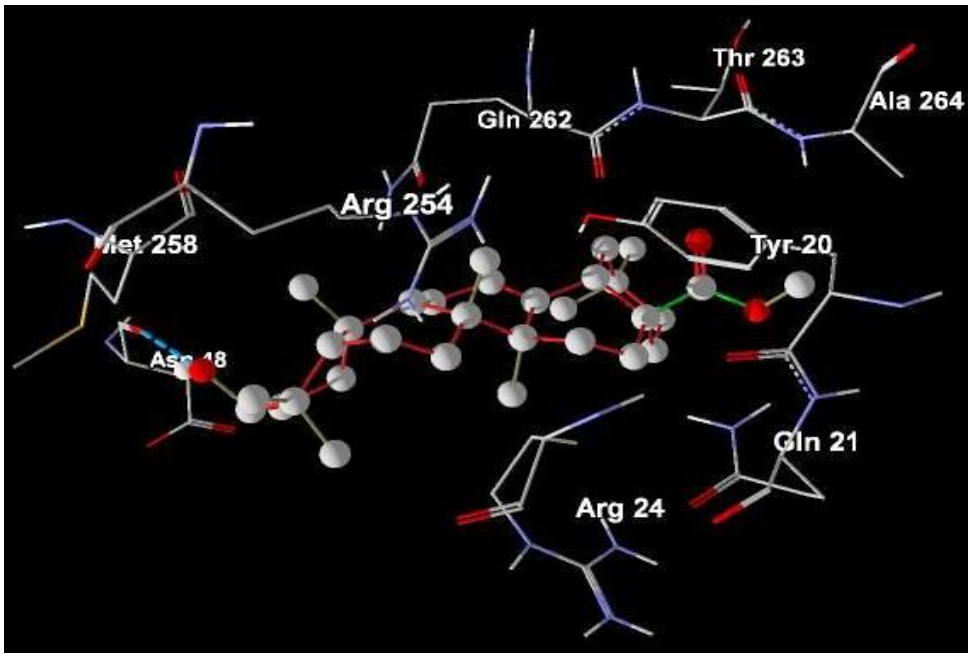


Figure 7.1c. Docked conformation of ligand 7 along with the important amino acid residues of PTP1B

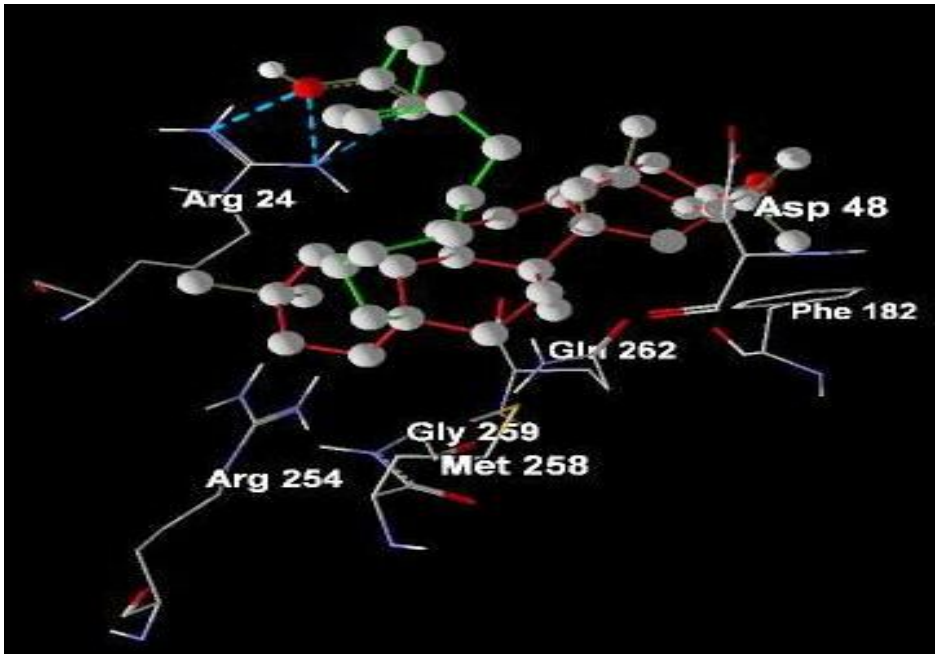


Figure 7.1d. Docked conformation of ligand 8 along with the important amino acid residues of PTP1B

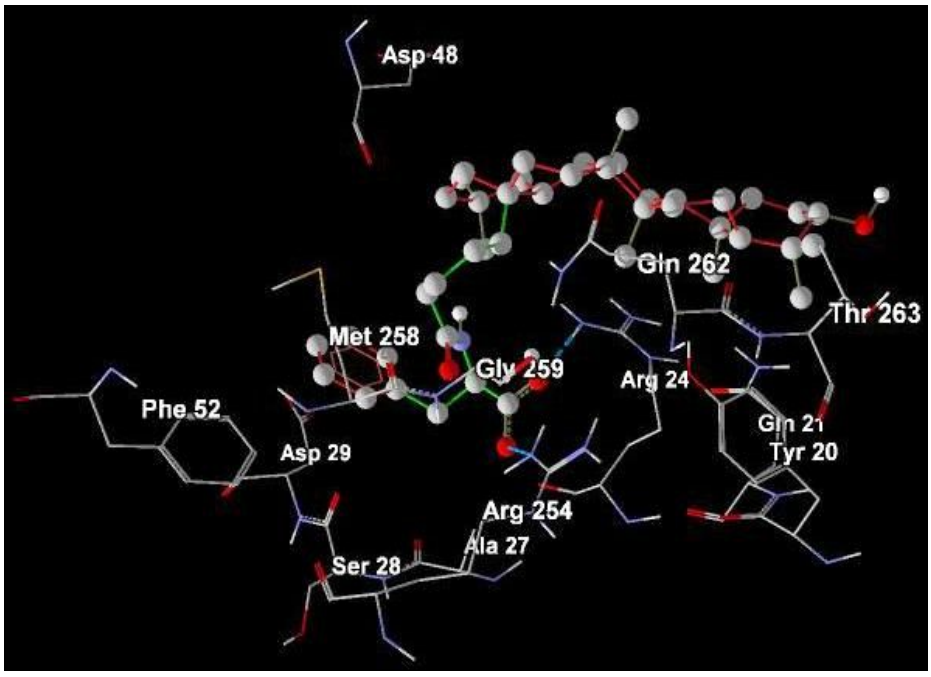


Figure 7.1e. Docked conformation of ligand 11 along with the important amino acid residues of PTP1B

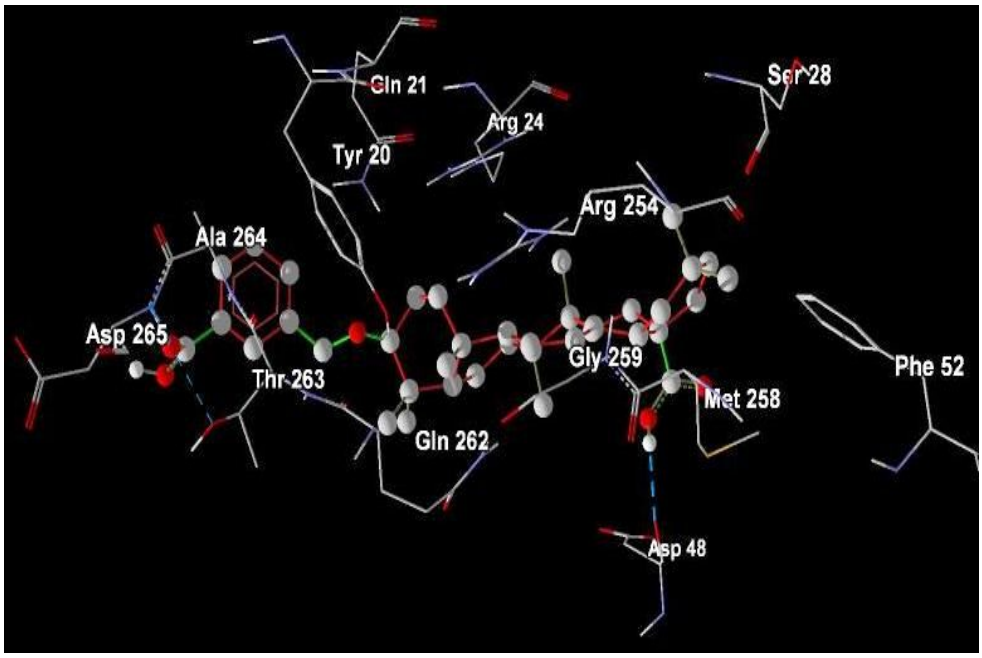


Figure 7.1f. Docked conformation of ligand 29 along with the important amino acid residues of PTP1B

The data set of 35 compounds was divided into two groups. The training sets constitute 28 compounds (1,2,3,4,5,6,9,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,29,30,31,33,34,35) and the remaining 7 compounds (7,8,10,26,27,28,32) are part of the test sets. The list of the descriptors of training and test compounds are presented in Table 7.2.

Table 7.2. Binding energy (EB), Solvent accessible surface area (SASA), Molar refractivity (MR), Molar volume (MV), Partition coefficient (logP), HOMO energy (EH), LUMO energy (EL) and Dipole moment (μ) of 41 PTP1B inhibitors

Sl	EB kcal/mol	SASA	MR (cm^3)	MV (cm^3)	logP	EH (hartree)	EL (hartree)	μ (debye)
1	-9.95	693.68	133.57	414.90	9.06	-0.2092	-0.0371	4.8603
2	-8.30	727.81	142.83	447.00	10.05	-0.2238	-0.0292	3.5940
3	-9.04	797.18	152.09	479.20	11.08	-0.2186	-0.0137	4.0082
4	-7.37	848.88	170.62	543.40	13.20	-0.2208	-0.0079	4.7519
5	-8.62	694.83	135.66	421.1	8.11	-0.2267	0.0109	4.4407
6	-7.24	929.64	179.88	575.50	14.27	-0.2190	0.0004	6.3005
7	-8.60	712.99	138.41	439.70	9.52	-0.2163	-0.0128	2.0653
8	-6.72	834.31	189.14	607.60	15.33	-0.2054	-0.0134	4.4429
9	-7.13	952.31	194.45	590.10	12.14	-0.2402	-0.0173	6.0856
10	-9.11	696.21	135.03	413.50	7.82	-0.2205	-0.0328	3.1187
11	-9.44	919.25	194.45	590.10	12.14	-0.2379	-0.0210	8.6168
12	-8.29	967.79	205.93	602.50	12.06	-0.2786	0.1185	6.1512
13	-6.86	940.36	187.01	577.90	11.17	-0.2175	-0.0486	3.7634
14	-9.32	996.47	194.64	584.80	12.55	-0.3437	0.1061	5.6645
15	-8.46	993.24	194.64	584.80	12.55	-0.3437	0.1061	5.6645
16	-9.11	999.19	194.64	584.80	12.55	-0.3437	0.1061	5.6645
17	-7.05	989.79	189.93	578.40	12.01	-0.3217	0.1186	2.1511
18	-8.97	945.86	194.44	589.70	12.42	-0.3083	0.1188	9.2002
19	-9.64	945.02	195.85	584.70	11.69	-0.3139	0.0314	6.0850
20	-6.04	934.37	196.18	595.50	11.87	-0.3209	0.1188	5.4226

Table 7.2 (continued)

Sl	EB kcal/mol	SASA	MR (cm^3)	MV (cm^3)	logP	EH (hartree)	EL (hartree)	μ (debye)
21	-8.19	958.45	196.18	595.50	11.87	-0.3090	0.1137	4.1579
22	-8.79	965.52	196.18	595.50	11.87	-0.3113	0.1276	1.8820
23	-8.49	985.10	195.87	582.20	11.82	-0.3095	0.1498	3.1091
24	-8.34	964.21	202.54	617.00	11.69	-0.3055	0.1230	7.3918
25	-12.43	905.20	202.54	617.00	11.67	-0.3146	0.1115	1.8242
26	-6.78	974.30	202.54	617.00	11.69	-0.3026	0.1113	6.3092
27	-9.42	691.71	133.69	412.50	7.10	-0.2211	-0.0540	4.4538
28	-10.12	673.80	133.52	415.70	9.01	-0.2250	-0.0124	3.9526
29	-8.69	890.55	169.40	507.50	11.41	-0.3155	0.0891	1.9924
30	-10.59	752.05	144.81	446.50	9.17	-0.3375	0.0759	2.5937
31	-10.59	828.68	163.34	511.20	10.27	-0.3769	0.1443	4.0873
32	-6.04	1073.27	230.27	682.00	14.49	-0.3161	0.0811	3.1218
33	-9.09	683.58	132.17	413.80	8.48	-0.3441	0.1472	5.5761
34	-8.65	684.99	133.57	414.9	9.06	-0.3354	0.1487	4.6350
35	-9.36	693.12	136.23	428.2	11.20	-0.3281	0.1532	5.2573

Among the generated QSAR models; two models were finally selected. Model summary of two best models are given below:

Model 7.1

pIC_{50}=-17.510236+(-0.0088)BE+(2.6299)lnSASA+(1.1996)EH+(0.1447)EL+(-0.0053)μ

N=28, R=0.96, R^2=0.92, R^2_{cv}=0.87, F=50.60, S=0.35, Q=2.74

Model 7.2

pIC_{50}=-9.718794+(0.9222)lnSASA +(2.3374)lnMR+(-1.7038)lnMV+(0.8755)logP

N=28, R=0.95, R^2=0.90, R^2_{cv}=0.78, F=51.75, S=0.31, Q=3.06

In these models, N is the number of data points; R is the correlation coefficient between experimental values and calculated values from the equation. R^2 is the square of the correlation coefficient and it measures the goodness of fit of the regression equation. Cross validated coefficient (R^2_{cv}) gives an idea of the performance of the model. S is the standard deviation of the regression. Fischer statistics (F) is a ratio between variances calculated and observed activity. The larger value of F test signifies the QSAR model. Q is the quality factor. Q value measures predictive power of the QSAR models.

By using model number 7.1 and 7.2 the theoretical pIC_{50} values of 28 training compounds are given in Table 7.3 together with experimental pIC_{50}. Using the model number 7.1 and 7.2, we calculated the theoretical pIC_{50} of the test set which appeared in Table 7.4.

Table 7.3. List of experimental and predicted pIC_{50} of 28 training compounds

Comp no.	Experimental pIC_{50}	Predicted pIC_{50} (By model 7.1)	Predicted pIC_{50} (By model 7.2)
1	-0.5276	-0.4838	-0.3975
2	-0.3222	-0.3896	-0.3235
3	-0.1239	-0.1373	-0.2111
4	0.1079	0.0106	-0.0988
5	-0.6776	-0.5120	-0.3849
6	0.1427	0.2506	0.0109

Table 7.3 (continued)

Comp no.	Experimental pIC_{50}	Predicted pIC_{50} (By model 7.1)	Predicted pIC_{50} (By model 7.2)
9	0.1308	0.2876	0.1724
11	0.2441	0.2179	0.1399
12	0.2291	0.2941	0.2858
13	0.2596	0.2793	0.1053
14	0.2518	0.3019	0.2319
15	0.2924	0.2859	0.2289
16	0.2147	0.3071	0.2344
17	0.2441	0.2907	0.1872
18	0.2596	0.2042	0.1671
19	0.3468	0.2010	0.1976
20	0.2757	0.1312	0.1600
21	0.284	0.2312	0.1834
22	0.2218	0.2532	0.1902
23	0.3565	0.3053	0.2437
24	0.1805	0.2525	0.2030
25	0.2007	0.1117	0.1448
29	0.2076	0.0345	0.0452
30	-0.4564	-0.4196	-0.2592
31	-0.3674	-0.2117	-0.1190
33	-0.7259	-0.6919	-0.4312
34	-0.7033	-0.6798	-0.4091
35	-0.4548	-0.6337	-0.4060

Table 7.4. List of experimental and predicted pIC_{50} of 7 test compounds

Comp no.	Experimental pIC_{50}	Predicted pIC_{50} (By model 7.1)	Predicted pIC_{50} (By model 7.2)
7	-0.6474	-0.4297	-0.5332
8	0.2291	-0.0327	0.2077
10	-0.7396	-0.5013	-0.6805
26	0.0862	0.2673	0.2471
27	-0.9053	-0.5266	-0.7911
28	-0.4886	-0.5856	-0.6220
32	0.15	0.5117	0.6540

Statistical significance of these two models (model 7.1 & 7.2) were further supported by a plot of predicted pIC_{50} vs. experimental pIC_{50} (Figure 7.2 & Figure 7.3) of training set inhibitors and give an idea about how fit model was trained and how well it predict the activity of the test set compounds (Figure 7.4 & Figure 7.5).

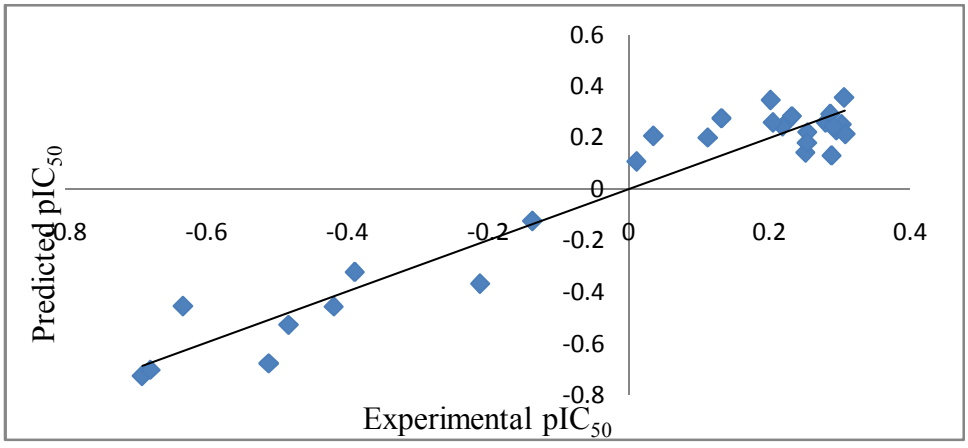


Figure 7.2. A plot between the predicted and the experimental pIC_{50} for the training set by model 7.1

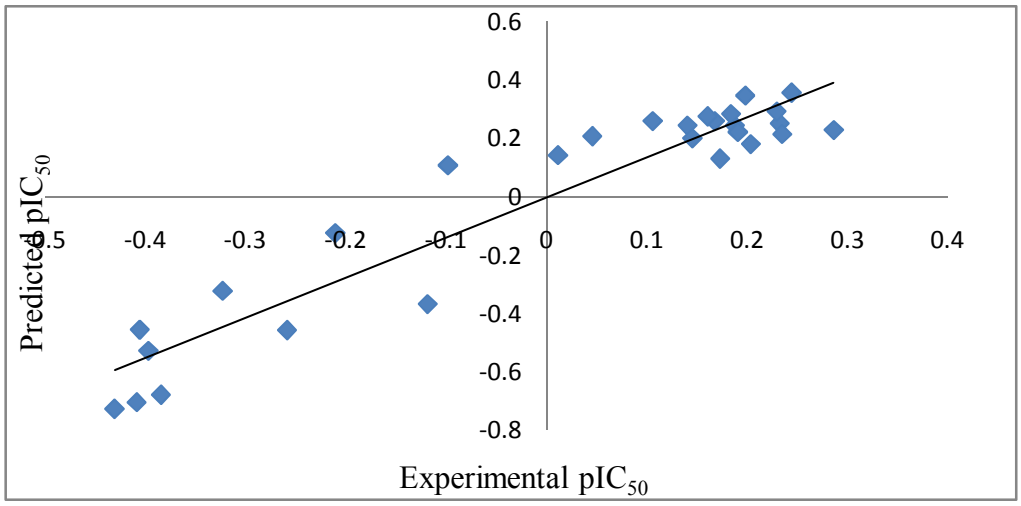


Figure 7.3. A plot between the predicted and the experimental pIC_{50} for the training set by model 7.2

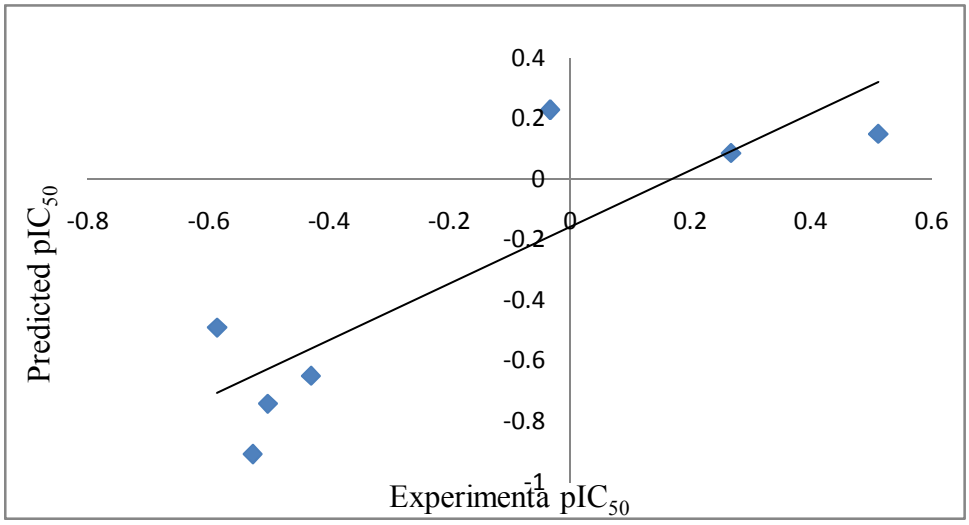


Figure 7.4. A plot between the predicted and the experimental pIC_{50} for the test set by model 7.1

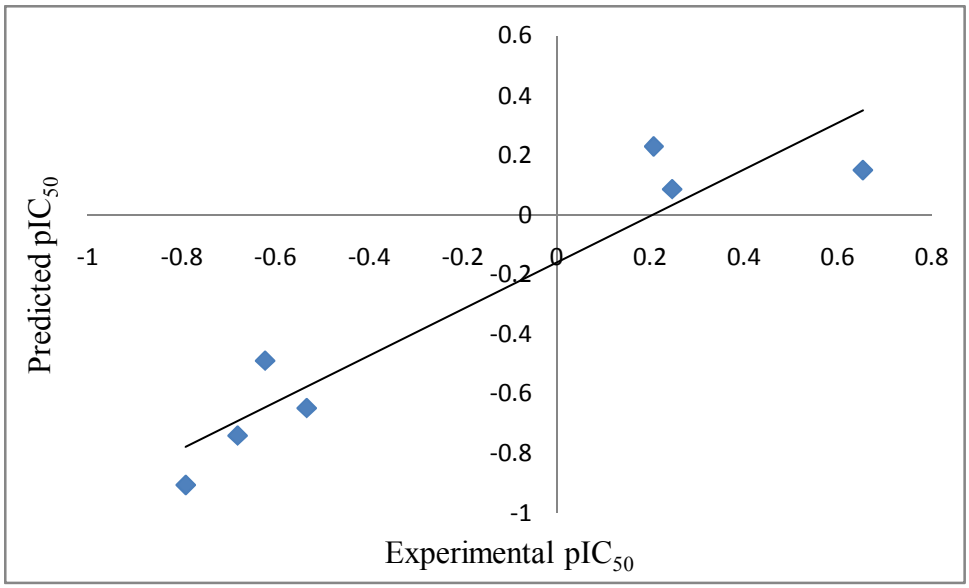


Figure 7.5. A plot between the predicted and the experimental pIC_{50} for the test set by model 7.2

Model 7.1 revealed that solvent accessible surface area (SASA), HOMO energy (EH) and LUMO energy (EL) were contributed positively to the model where binding energy (EB) and dipole moment (μ) were contributed negatively to the model. Solvent accessible surface area (SASA), molar refractivity (MR), and partition coefficient (logP) were contributed positively where molar volume (MV) was contributed negatively to the model 7.2.

7.4. Conclusion

In conclusion, this QSAR study has shown that binding energy (EB), HOMO energy (EH), LUMO energy (EL), dipole moment (μ), molar refractivity (MR), molar volume (MV), solvent accessible surface area (SASA) and partition coefficient (logP) are the important parameters for determining the activity of oleanolic acid derivatives. Model 7.1 and model 7.2 are the best equation for predicting the inhibitory activity of Protein–tyrosine phosphatase 1B and these QSAR models may be used in prediction of activity of designed compound. The docking study shows that the important interacting amino acids present in the active site are TYR20, GLN21,

ARG24, ALA27, SER28, TYR46, ASP48, VAL49, ASP181, PHE182, ALA217, ILE219, ARG254, MET258, GLY259, GLN262, THR263. Most of the ligands can form hydrogen bonds with ARG24 and/or ARG254. Binding energies and partion coefficient (logP) play an important role for predicting the activity of the inhibitors.

7.5 References

[1] D. Barford, A.J. Flint, N.K. Tonks, Crystal structure of human protein tyrosine phosphatase 1B, Science. 263 (1994) 1397-1404.

[2] Z.Y. Zhang, Structure, mechanism, and specificity of protein-tyrosine phosphatases, Curr Top Cell Regul. 35 (1997) 21-68.

[3] D. Barford, A.K. Das, M.P. Egloff, The structure and mechanism of protein phosphatases: insights into catalysis and regulation, Annu Rev Biophys Biomol Struct. 27 (1998) 133-164.

[4] Z.Y. Zhang, Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis, Crit Rev Biochem Mol Biol. 33 (1998) 1-52.

[5] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, Protein tyrosine phosphatases in the human genome, Cell. 117 (2004) 699-711.

[6] K.A. Kenner, E. Anyanwu, J.M. Olefsky, J. Kusari, Protein-tyrosine phosphatase 1B is a negative regulator of insulin- and insulin-like growth factor-I-stimulated signaling, J Biol Chem. 271 (1996) 19810-19816.

[7] J.R. Wiener, B.J. Kerns, E.L. Harvey, M.R. Conaway, J.D. Iglehart, A. Berchuck, R.C. Bast Jr, Overexpression of the protein tyrosine phosphatase PTP1B in human breast cancer: association with p185c-erbB-2 protein expression, J Natl Cancer Inst. 86 (1994) 372-378.

[8] M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A.L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.C. Chan, C. Ramachandran, M.J. Gresser, M.L. Tremblay, B.P. Kennedy, Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene, Science. 283 (1999) 1544-1548.

[9] J. Liu, Pharmacology of oleanolic acid and ursolic acid, J Ethnopharmacol. 49 (1995) 57-68.

[10] J. Liu, Oleanolic acid and ursolic acid: research perspective, J Ethnopharmacol. 100 (2005) 92-94.

[11] P. Dzubak, M. Hajduch, D. Vydra, A. Hustova, M. Kyasnica, D. Biedermann, L. Markova, M. Urban, J. Sarek, Pharmacological activities of natural triterpenoids and their therapeutic implications, Nat Prod Rep. 23 (2006) 394-411.

[12] R. Martín, J. Carvalho-Tavares, M. Hernández, M. Arnés, V. Ruiz-Gutiérrez, M.L. Nieto, Beneficial actions of oleanolic acid in an experimental model of multiple sclerosis: a potential therapeutic role, Biochem Pharmacol. 79 (2010) 198-208.

[13] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic

activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[14] Y.A. Puius, Y.U. Zhao, M. Sullivan, D.S. Lawrence, S.C. Almo, Z.Y. Zhang, Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design, Proc Natl Acad Sci USA. 94 (1997) 13420-13425.

[15] J.M. Castellano, A. Guinda, T. Delgado, M. Rada, J.A. Cayuela, Biochemical basis of the antidiabetic activity of oleanolic acid and related pentacyclic triterpenes, Diabetes. 62 (2013) 1791-1799.

[16] Y.N. Zhang, W. Zhang, D. Hong, L. Shi, Q. Shen, J.Y. Li, J. Li, L.H. Hu, Oleanolic acid and its derivatives: new inhibitor of protein tyrosine phosphatase 1B with cellular activities, Bioorg Med Chem. 16 (2008) 8697-8705.

[17] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[18] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J Comput Chem. 19 (1998) 1639-1662.

[19] Molegro molecular viewer – version 2.5.0. http://www.molegro.com/index.php.

[20] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[21] M.F. Sanner, Python: a programming language for software integration and development, J Mol Graphics Mod. 17 (1999) 57-61.

CHAPTER 8

Quantum chemical study of Halomon by the DFT methods

8.1. Introduction

Halomon [6(R)-bromo-3(S)-(bromomethyl)-7-methyl-2,3,7-trichloro-1-octene], is a poly halogenated acyclic monoterpine, was originally isolated from the red alga portieria hornemannii [1]. Halomon exhibited highly differential cytotoxicity in vitro against brain tumor, renal and colon tumor cell lines in the national Cancer institute's screen and the compound was selected for preclinical drug development [2-6]. Recent research suggested that halomon is inhibitor of the enzyme DNA methyl transferase-1 (DNMT-1). DNMT-1 is responsible for methylation by catalyzing the transfer of a methyl group from S-adenosylmethionine to the 5′ position on cytosine phosphodiester-linked guanine dinucleotide (CpG) sites. In many cancers, hypermethylation at CpG sites silence the promoters of tumor suppressor genes. Thus, the inhibition of DNMT-1 could potentially reverse tumor growth [7, 8]. The molecular structure of halomon is given in Figure 8.1.

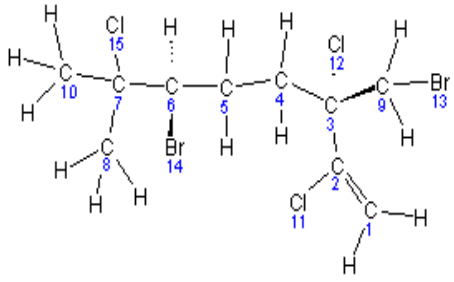


Figure 8.1. Molecular structure and atom labeling of halomon (hydrogen atoms are not numerated)

The primary aim of this study is to identify all possible conformers of halomon and their relative stability is determined by using DFT calculation. The structure of most stable conformer is compared with the experimental values. Finally harmonic vibrational frequency calculations were performed at the B3LYP / 6-31+G(d,p) level.

8.2. Computational methods

The initial structure of halomon was obtained from Cambridge Crystallographic Data Centre (www.ccdc.cam.ac.uk) and other conformations are constructed by molecular dynamics calculations using the PM7 semi-empirical potential, followed by a geometry optimization using the same semi-empirical method. The molecular dynamics simulations were done using the Velocity Verlet algorithm and other default settings were used. Similar geometries were omitted (two geometries are considered similar if the difference between their energies is less than 1 kJ/mol). The conformational isomers of halomon obtained from the PM7 semi-empirical calculations were fully re-optimized by using the DFT/B3LYP method with 6-31+G(d,p)basis set to determine the relative energies and the structures of five distinct conformations. Vibrational analysis was carried out to ensure that each optimization located a true minimum energy structure (no imaginary frequencies). Semi-empirical calculations were done by MOPAC2016 [9]. The DFT calculations were performed with the Firefly [10].

8.3. Results and discussion

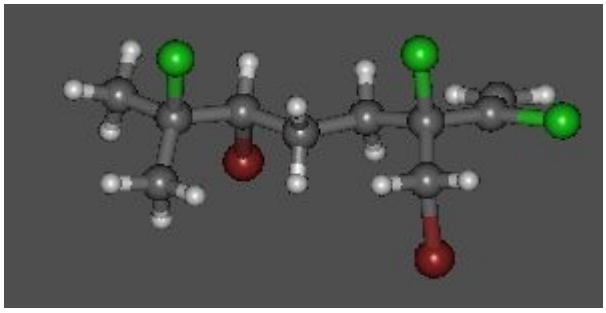
During the DFT optimization, some conformations are converted into the most stable conformers. Table 8.1 shows the total and relative energies of the five conformers of halomon calculated by the DFT B3LYP/6-31+G(d,p).

Table 8.1. DFT calculated energies of the various conformers of halomon and the relative energies with respect to the most stable conformation

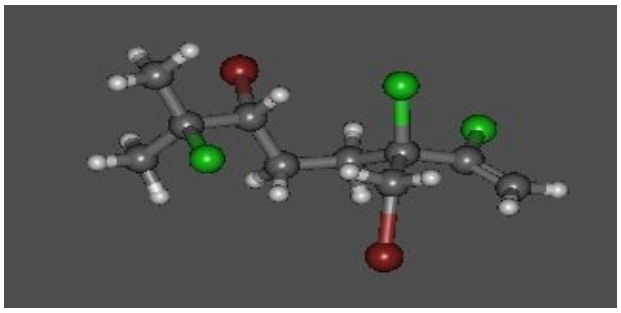
Conformer	a	b	c	d	e
Total energy[a]	-6914.2546	-6914.2509	-6914.2505	-6914.2471	-6914.2465
Relative energy[b]	0.00	9.71	10.76	19.69	21.27

[a]The unit of DFT energy is in a.u. [b]The unit of relative energy is (kJ/mol)

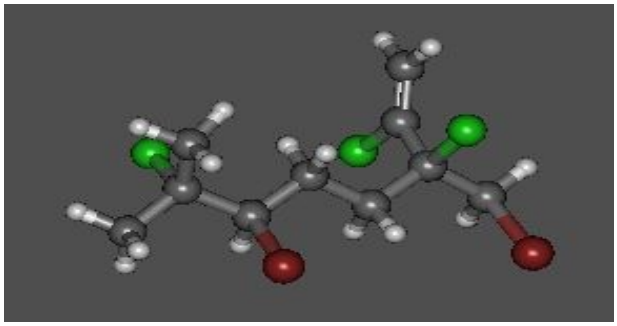
Table 8.1 suggest that conformer a is 9.71 kJ/mol more stable than conformer b, about 10.76 kJ/mol more stable than conformer c, about 19.69 kJ/mol more stable than conformer d and 21.27 kJ/mol more stable than conformer e, respectively. Figure 8.2 shows the DFT B3LYP/6-31+G(d,p) optimized stable structures of the various conformers of halomon. The molecular plots were produced using Gabedit 2.4.8 [11].



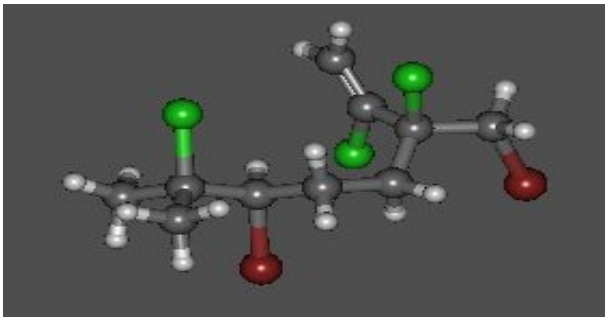
a



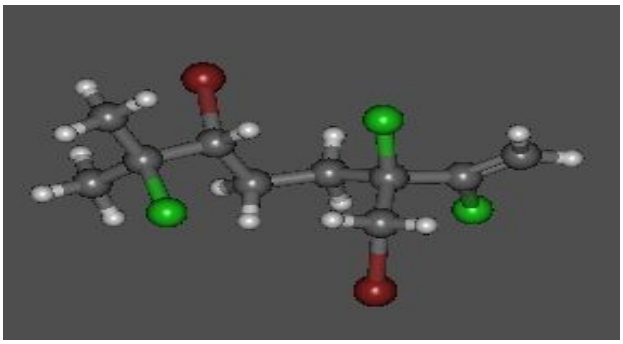
b



c



d



e

Figure 8.2. Conformers of halomon

The calculated bond lengths, bond angles and dihedral angles of most stable conformer of halomon (a) by B3LYP/6-31+G(d,p) and B3PW91/6- 31+G(d,p) are listed in Table 8.2 along with the experimental data obtained by X-ray crystal analysis. The comparison between the theoretical and experimental data on bond lengths, bond angels and dihedral angles in Table 8.2 shows a good agreement between them. In the experiments, the shortest bond in halomon was observed for C1-C2 while the longest bonds are C9-Br13 and C6-Br14. Our theoretical results confirm these experimental findings as shown in Table 8.2.

Table 8.2. Optimized structural parameters of most stable conformer of halomon with experimental data

Structural parameters	Experimental	B3LYP/6- 31+G(d,p)	B3PW91/6- 31+G(d,p)
Bond length (Å)			
C1-C2	1.388	1.333	1.332
C2-C3	1.517	1.520	1.517
C3-C4	1.388	1.540	1.534
C4-C5	1.495	1.536	1.530
C5-C6	1.570	1.527	1.522

Table 8.2 (continued)

Bond length (Å)	Experimental	B3LYP/6- 31+G(d,p)	B3PW91/6- 31+G(d,p)
C6-C7	1.624	1.548	1.544
C7-C8	1.503	1.528	1.523
C3-C9	1.582	1.531	1.527
C7-C10	1.500	1.530	1.525
C2-Cl11	1.718	1.767	1.753
C3-Cl12	1.818	1.869	1.845
C9-Br13	2.002	1.970	1.949
C6-Br14	1.985	2.005	1.981
C7-Cl15	1.854	1.872	1.894
Bond angle (^{0})			
C1-C2-C3	123.40	126.26	126.08
C2-C3-C4	117.34	112.00	111.93
C3-C4-C5	115.45	114.90	114.69
C4-C5-C6	113.23	112.60	112.46
C5-C6-C7	113.57	116.38	116.19
C6-C7-C8	106.60	112.68	112.40
C1-C2-Cl11	115.94	118.29	118.59
C2-C3-Cl12	105.66	107.23	107.42
C9-C3-Cl12	99.88	103.31	103.62
C3-C9-Br13	109.81	110.65	110.44
C7-C6-Br14	105.16	108.93	109.07
C6-C7-Cl15	100.28	104.15	104.32

Table 8.2 (continued)

Dihedral angles (^{0})	Experimental	B3LYP/6-31+G(d,p)	B3PW91/6-31+G(d,p)
C1-C2-C3-Cl12	-118.53	-113.57	-113.60
C4-C3-C9-Br13	58.53	63.46	64.15
C5-C6-C7-Cl15	100.28	66.00	65.73
Br14-C6-C7-Cl15	-178.26	-171.08	-178.95
Cl12-C3-C9-Br13	-177.39	-179.12	-171.08

The main differences are the torsion angle formed by C5-C6-C7-Cl15 atoms (100.28° in crystallography, 66.00° in B3LYP and 65.73° in B3PW91) and these differences are probably due to packing effects. The IR spectra for the most stable conformer of halomon are given in Figure 8.3.

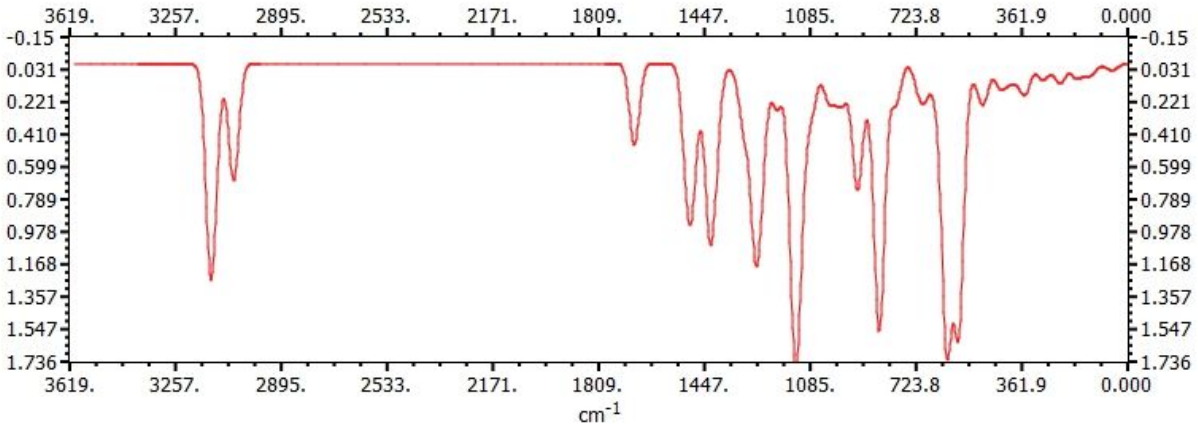


Figure 8.3. IR spectra of the most stable conformer of Halomon in gas phase

Due to the use of the harmonic approximation and incomplete inclusion of electron correlation computed harmonic vibrational frequencies are usually somewhat larger than experimental fundamentals. The B3LYP/6-31G+(d,p) spectra show C-H stretching modes in the 3000-3200 cm^{-1} range.

8.4. Conclusions

The present quantum chemical investigation based on DFT methods provide useful information about the structure of halomon. The harmonic vibrational frequencies for halomon were calculated at the B3LYP/6-31+G(d,p).

8.5. Reference

[1] R.W. Fuller, J.H. Cardellina II, Y. Kato, L.S. Brinen, J. Clardy, K.M. Snader, M.R. Boyd, A pentahalogenated monoterpine from the red alga Portieria hornemannii produces a novel cytotoxicity profile against a diverse panel of human tumor cell lines, J Med Chem. 35 (1992) 3007-3011.

[2] R.W. Fuller, J.H. Cardellina II, J. Jurek, P.J. Scheuer, B.A. Lindner, M. Mcguire, G.N. Gray, J.R. Steiner, J. Clardy, E. Menez, R.H. Shoemaker, D.J. Newman, K.M. Snader, M.R. Boyd, Isolation and structure/activity features of halomon - related antitumor monoterpines from the red alga Portieria hornemannii, J Med Chem. 37 (1994) 4407-4411.

[3] M.J. Egorin, D.L. Sentz, D.M. Rosen, M.F. Ballesteros, C.M. Kearns, P.S. Callery, J.L. Eiseman, Plasma pharmacokinetics, bioavailability, and tissue distribution in CD_2F_1 mice of halomon, an antitumor halogenated monoterpene isolated from the red alga Portieria hornemannii, Cancer Chemother Pharmacol. 39 (1996) 51-60.

[4] M.J. Egorin, M.D. Rosen, S.E. Benjamin, P.S. Callery, D.L. Sentz, J.L. Eiseman, In vitro metabolism of mouse and human liver preparations of halomon, an antitumor halogenated monoterpene, Cancer Chemother Pharmacol. 41 (1997) 9-14.

[5] M.E. Jung, M.H. Parker, Synthesis of several naturally occurring polyhalogenated monoterpene of the halomon class, J Org Chem. 62 (1997) 7094-7095.

[6] T. Sotokawa, T. Noda, S. Pi, M.A. Hirama, A three step synthesis of halomon, Angew Chem Int Ed Engl. 39 (2000) 3430-3432.

[7] H.D. Yoo, S.O. Ketchum, D. France, K. Bair, W.H. Gerwick, Vidalenolone , a novel phenolic metabolite from the tropical red alga Vidalia sp., J Nat Prod. 65 (2002) 51-53.

[8] E.H. Andrianasolo, D. France, S.C. Kennon, W.H. Gerwick, DNA Methyl Transferase Inhibiting Halogenated Monoterpenes from the Madagascar Red Marine Alga Portieria hornemannii, J Nat Prod. 69 (2006) 576-579.

[9] J.J.P. Stewart, MOPAC2016, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016).

[10] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[11] A.R. Allouche, Gabedit-a graphical user interface for computational chemistry softwares, J Comput Chem. 32 (2011) 174-182.

CHAPTER 9

A QSAR study of sesquiterpene lactones from Inula falconeri as potent anti-inflammatory agents

9.1. Introduction

Sesquiterpenes are a class of naturally occurring plant terpenoids with a skeleton of 15 carbons that occur as hydrocarbons or in oxygenated forms such as alcohols, ketones, aldehydes, acids, and lactones. Among them sesquiterpene lactones have been shown to exhibit a wide range of biological and pharmacological activities. Sesquiterpene lactones constitute a large and diverse group of biologically active natural products that possess anti-inflammatory and antitumor activity [1]. A number of plants from the Inula genus have rich source of sesquiterpenoids which accounted for its diverse biological activities i.e., anticancer, antibacterial, hepaprotective, cytotoxic, and anti-inflammatory activity [2]. Aim of the present study is to construct QSAR models using multiple regression method, to explore the correlations between the experimental anti-inflammatory activity and calculated molecular indexes of 16 sesquiterpene lactones from Inula falconeri.

9.2. Materials and methods

A total of 20 sesquiterpene lactones from Inula falconeri with anti-inflammatory activities via the inhibition of NO production in RAW264.7 macrophages taken from the literature [3] were used for the QSAR studies. The initial structures of 20 compounds used in this study were generated by ChemSketch [4]. The biological property of this data set is reported as IC_{50} (μM) values and this value was changed to the logarithmic scale [log IC_{50}]. Structural details of the 20 compounds

and their biological activity are listed in Table 9.1. It is found that HOMO energy (EH), LUMO energy (EL), SIC (Structural Information Content), CIC (Complementary Information Content), dipole moment (μ), entropy (S) and χ (electronegativity) can better represent the biological activity of the selected compounds.

Table 9.1. Structural feature of sesquiterpene lactones from Inula falconeri with anti-inflammatory activity

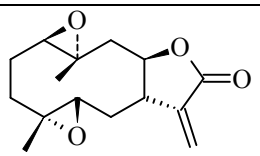
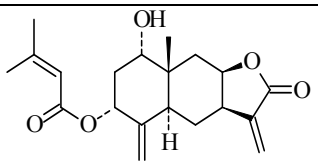
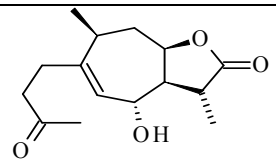
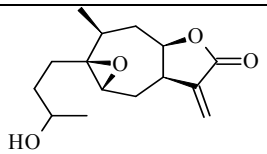
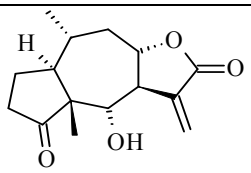
Comp. no.	Structure	IC_{50} (μM)
1		6.40
2		4.70
3		21.90
4*		18.90
5		2.18

Table 9.1 (continued)

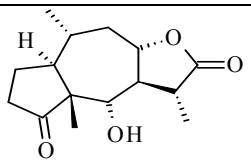
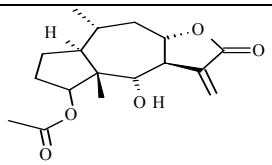
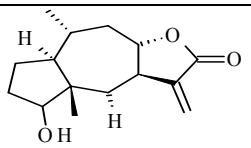
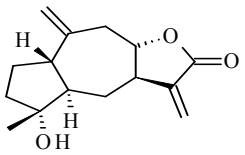
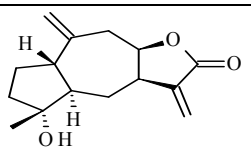
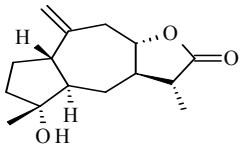
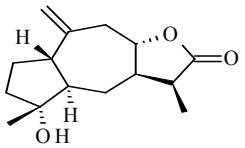
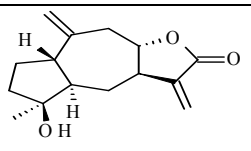
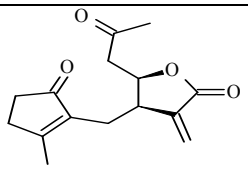
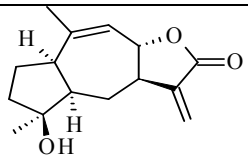
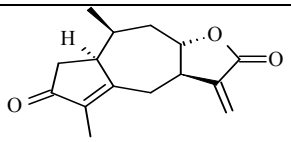
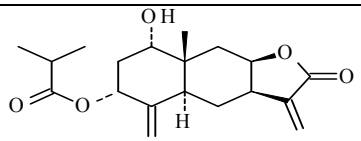
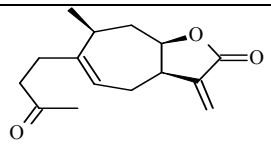
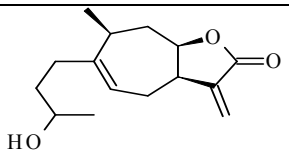
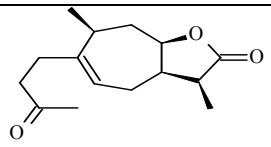
Comp. no.	Structure	IC_{50} (μM)
6		20.3
7		2.05
8		9.89
9*		9.64
10		3.94
11		41.2
12*		19.53
13		7.30

Table 9.1 (continued)

Comp. no.	Structure	IC_{50} (μM)
14		15.82
15		5.94
16		12.86
17		9.12
18*		7.14
19		5.48
20		64.9

*indicates test set compounds

The quantum chemical properties (EH, EL, μ) of the studied molecules have been determined by DFT/B3LYP calculation and the basis set 6-31G* was used [5]. All quantum chemical calculations were performed with the Firefly quantum chemistry package [6].

The average information content is defined on the basis of the Shannon information theory and is calculated as follows [7, 8]:

$$IC = -\sum_{i=1}^{h} p_i \, log_2 p_i \quad \ldots\ldots\ldots (9.1)$$

$$(p_i = n_i/n)$$

Where n_i is the number of atoms in the i^{th} class and n is a total number of atoms in the molecule. The division of atoms into different classes depends upon the coordination sphere that one has taken into account. This leads to the indices of different order k. The information content (IC) is equal to average information content multiplied by the total number of atoms. Other information content indices (SIC-structural IC, CIC-complementary IC) are defined as follows [9].

$$SIC^k = IC^k / log_2 n \quad \ldots\ldots\ldots (9.2)$$

$$CIC^k = log_2 n - IC^k \quad \ldots\ldots\ldots (9.3)$$

Entropy (S) at 298K of different compounds was calculated using semi-empirical PM6 method by Mopac [10]. Electronegativity (χ) is derived from the DFT framework and is defined as [11]:

$$\chi_{\text{koopmans}} = (EH+EL)/2 \quad \ldots\ldots\ldots (9.4)$$

Multiple linear regression (MLR) analysis was used to build up QSAR models. Different combinations of parameters were tried to develop these models. Statistical qualities of MLR equations were judged by square of the correlation coefficient (R^2), standard deviation of the regression (S), Fischer statistics (F) and quality factor (Q) [12,13]. The graph theoretical descriptors such as SIC, CIC, and MLR were computed using program written by us in Fortran-77.

9.3. Results and discussion

The data set of 20 compounds was divided into two groups. The training sets constitute 16 compounds (1,2,3,5,6,7,8,10,11,13,14,15,16,17,19,20) and the remaining 4 compounds (4,9,12,18) are part of the test sets. The list of the descriptors of training and test compounds are presented in Table 9.2.

Table 9.2. SIC_1, CIC_1, quantum chemical descriptors, entropy at 298 K and electronegativity of 20 inhibitors

Comp no.	SIC_1	CIC_1	EH (eV)	EL (eV)	μ (debye)	S (cal/M-K)	χ (eV)
1	0.4763	2.7679	-6.9988	-1.2300	5.0057	128.2796	-4.1144
2	0.4984	2.8454	-6.8734	-1.3660	3.0454	155.3138	-4.1197
3	0.4878	2.7440	-6.5416	-0.6721	4.0324	136.5074	-3.6069
4	0.4947	2.7072	-6.8600	-1.2436	4.7691	133.2406	-4.0518
5	0.5332	2.4672	-6.8328	-1.2000	5.5484	124.8293	-4.0164
6	0.4822	2.7744	-6.8165	-1.0776	5.4920	127.5972	-3.9471
7	0.4905	2.8057	-7.1375	-1.3878	3.4364	158.2125	-4.2627
8	0.4880	2.7248	-6.9171	-1.1429	5.5290	137.8948	-4.0300
9	0.5113	2.5649	-6.8872	-1.2762	4.5649	120.9682	-4.0817
10	0.5113	2.5649	-6.7702	-1.2218	4.7782	120.7365	-3.9960
11	0.4821	2.7560	-6.7838	0.1633	4.7946	123.1420	-3.3103
12	0.4821	2.7560	-6.7838	0.1633	4.7946	123.1214	-3.3103
13	0.5113	2.5649	-6.7484	-1.2789	4.2564	120.7273	-4.0137
14	0.4990	2.6100	-6.1498	-1.1592	5.0376	133.1585	-3.6545
15	0.5275	2.4795	-6.6423	-1.1320	6.3539	121.6708	-3.8872

Table 9.2 (continued)

Comp no.	SIC_1	CIC_1	EH (eV)	EL (eV)	μ (debye)	S (cal/M-K)	χ (eV)
16	0.4986	2.5923	-6.4110	-1.4123	3.7251	121.2161	-3.9117
17	0.4679	3.0480	-6.9716	-1.3442	3.9610	151.9919	-4.1579
18	0.4863	2.6958	-6.4763	-1.2599	3.2386	144.7034	-3.8681
19	0.5165	2.5729	-6.4409	-1.1973	3.9615	130.7239	-3.8191
20	0.4604	2.8716	-6.4437	-0.6095	3.1041	130.2655	-3.5266

Among the generated QSAR models; three models were finally selected. Model summary of three best models with predicted log IC_{50} are given below:

Model 9.1

$\text{Log } IC_{50}=13.208046 + 0.8186EH +0.4259LH +0.0630\mu +(-13.1295)SIC_1 + (-0.0223)CIC_1$

$N=16, R=0.96, R^2=0.92, F=23, S=0.40, Q=2.40$

Model 9.2

$\text{Log } IC_{50} = 4.2297+ (-0.3389)EL + (1.8210)\chi + (1.2849)CIC_1$

$N=16, R=0.91, R^2=0.83, F=19.53, S=0.39, Q=2.33$

Model 9.3

$\text{Log } IC_{50}= 4.2297 + -0.9710S + 0.9800\chi + (-12.3293)SIC_1$

$N=16, R=0.95, R^2=0.90, F=36, S=0.40, Q=2.38$

By using model number 9.1, 9.2 and 9.3 the predicted log IC_{50} values of 16 training inhibitors are presented in Table 9.3 together with experimental log $IC_{50.}$ The model 9.3 with the R=0.95, R^2=0.90, F=36, S=0.40, Q=2.38 turns out to be the best fit model.

Table 9.3. List of experimental and predicted logIC_{50} of 16 training compounds

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (By model 9.1)	Predicted logIC_{50} (By model 9.2)	Predicted logIC_{50} (By model 9.3)
1	0.8062	0.9550	0.7107	1.0209
2	0.6721	0.5844	0.8467	0.5576
3	1.3404	1.3551	1.4151	1.3161
5	0.3385	0.3975	0.4926	0.4419
6	1.3075	1.1222	0.9721	1.1173
7	0.3118	0.4881	0.5427	0.4969
8	0.9952	0.9393	0.7795	0.8892
10	0.5955	0.6763	0.6627	0.7643
11	1.6149	1.6356	1.6875	1.7771
13	0.8633	0.6370	0.6498	0.7470
14	1.1992	1.3877	1.3213	1.1555
15	0.7738	0.7077	0.7207	0.6637
16	1.1092	0.9890	0.9160	0.9996
17	0.9600	0.9668	1.0301	0.9172
19	0.7388	0.8364	0.9868	0.7963
20	1.8122	1.7603	1.7041	1.7780

Using the model number 9.3, we calculated the theoretical log IC_{50} of the test set (R=0.53) which appeared in Table 9.4. The correlation graph of training compounds between experimental log IC_{50} and predicted log IC_{50} (by model 9.3) are presented in Figure 9.1.

Table 9.4. List of experimental and predicted pIC_{50} of 4 test compounds

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (by Model 9.3)
4	1.2765	0.9156
9	0.9841	0.6785
12	1.2907	1.7772
18	0.8537	1.0220

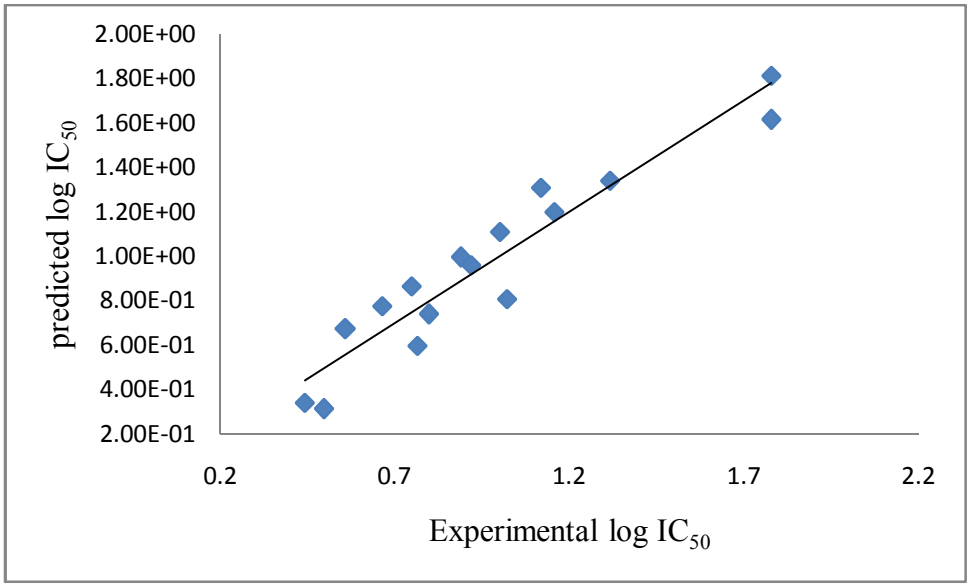


Figure 9.1. A plot between the predicted and the experimental activities for the 16 training compounds using model 9.3.

9.4. Conclusion

This QSAR study has been carried out different descriptor like first order SIC, first order CIC, HOMO energy, LUMO energy, dipole moment, entropy and electro negativity. These QSAR models may be used to find out the activity of the designed compounds.

9.5. References

[1] M.R. Orofino, O. Kreuger, S. Grootjans, M.W. Biavatti, P. Vandenabeele, K. D'Herde, Sesquiterpene lactones as drugs with multiple targets in cancer treatment: focus on parthenolide, Anticancer Drugs. 23 (2012) 883-896.

[2] Y.M. Zhao, M.L. Zhang, Q.W. Shi, H. Kiyota, Chemical constituents of plants from the genus Inula, Chem Biodivers. 3 (2006) 371-384.

[3] X. Cheng, Q. Zeng, J. Ren, J. Qin, S. Zhang, Y. Shen, J. Zhu, F. Zhang, R. Chang, Y. Zhu, W. Zhang, H. Jin, Sesquiterpene lactones from Inula falconeri, a plant endemic to the Himalayas, as potential anti-inflammatory agents, Eur J Med Chem. 46 (2011) 5408-5415.

[4] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[5] B. Bagchi, A. Chatterjee, P. Ghosh, A.K. Bothra, A theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum, JOCPR. 4 (2012) 5076-5080.

[6] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[7] E. Eroglu, Some QSAR studies for a group of sulfonamide schiff base as carbonic anhydrase CA II inhibitors, Int J mol Sci. 9 (2008) 181-197.

[8] L.B. Kier, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci. 69 (1980) 807-810.

[9] S.C. Basak, D.K. Harriss, V.R. Magnuson, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci. 73 (1984) 429-437.

[10] J.J.P. Stewart, MOPAC2012, Stewart Computational Chemistry, Colorado Springs, CO, USA.

[11] E. Eroğlu, H. Türkmen, S. Güler, S. Palaz, O. Oltulu, A DFT-based QSARs study of acetazolamide/sulfanilamide derivatives with carbonic anhydrase (CA-II) isozyme inhibitory activity, Int J Mol Sci. 8 (2007) 145-155.

[12] S. Sharma, B. Bagchi, S. Mukhopadhyay, A.K. Bothra, Theoretical study of lysophosphatidic acid acyltransferase 2 inhibitors, JOCPR. 5 (2013) 348-355.

[13] B. Das, P. Ghosh, A.K. Bothra, QSAR studies of anthrax lethal inhibitors through quantum chemical indices, JOCPR. 3 (2011) 443-449.

CHAPTER 10

Conclusion

The first chapter provides a brief description on terpenoids and the historical development of quantitative structure-activity relationship. A concise account of review of literature is included in chapter two with a view to familiarize the readers about the development of QSAR study. Determination of molecular indexes, regression analysis and statistical parameters has been the subject matter of chapter three.

Chapter four describes QSAR study and Molecular docking of 23-hydroxybetulinic acid derivatives as RMGPa and HeLa cells inhibitors. This QSAR study has shown that topological indices (e.g. SIC, CIC) and quantum chemical descriptors (e.g. EH, EL, μ) are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. Model 4.3 and model 4.6 are the best equation for predicting the inhibitory activity of RMGPa and the antiproliferative activities against HeLa cells respectively and these QSAR models may be used in prediction of activity of designed compounds. The docking study shows that the important interacting amino acids present in the active site are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the ligands can form hydrogen bonds with ARG193 and/or ARG310. The –OH group at C-3 and C-23 can increase the hydrogen bond interaction between ligands and enzyme. However, acetylation or esterification of the –OH group at the C-3 and C-23 not only decreases the number of hydrogen bond, but may also increase the unfavorable steric clashes. Thus binding energy may decrease. Large substituent at C-17 may increase the chance of steric bumps, thus lowering the inhibitory activity of the ligand.

Computational study on the redox reaction of puupehenone in aqueous solution has been given in chapter five. This study helps to predict the E^0 value of different lipoxygenases. Puupehedienone and puupehenone are capable of forming hydrogen bond with water; the absolute value of E^0 of P/PH_2 couple is highly dependent on the difference in the interaction energy and the difference in the solvation free energy.

Chapter six provides a theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum. Some compounds had selective activity against cancer cells versus CHO cells. Their selective cytotoxic activity depends mainly on the gap energy and the stereo chemical features of these compounds.

Chapter seven deals with the Molecular docking and DFT based QSAR study on oleanolic acid derivatives as Protein-tyrosine phosphatase 1B inhibitors. This QSAR study has shown that binding energy (EB), HOMO energy (EH), LUMO energy (EL), dipole moment (μ), molar refractivity (MR), molar volume (MV), solvent accessible surface area (SASA) and partition coefficient (logP) are the important parameters for determining the activity of oleanolic acid derivatives. Model 7.1 and model 7.2 are the best equation for predicting the inhibitory activity of Protein-tyrosine phosphatase 1B and these QSAR models may be used in prediction of activity of designed compound. The docking study shows that the important interacting amino acids present in the active site are TYR20, GLN21, ARG24, ALA27, SER28, TYR46, ASP48, VAL49, ASP181, PHE182, ALA217, ILE219, ARG254, MET258, GLY259, GLN262, THR263. Most of the ligands can form hydrogen bonds with ARG24 and/or ARG254. Binding energies and partion coefficient (logP) play an important role for predicting the activity of the inhibitors.

Chapter eight gives a quantum chemical study of Halomon by the DFT methods. The present study provides useful information about the structure of halomon. The harmonic vibrational frequencies for halomon were calculated at the B3LYP/6-31+G(d,p).

Chapter nine describes QSAR study of sesquiterpene lactones from Inula falconeri as potent anti-inflammatory agents. This QSAR study has been carried out different descriptor like first order SIC, first order CIC, HOMO energy, LUMO energy, dipole moment, entropy and electro negativity. These QSAR models may be used to find out the activity of the designed compounds.

BIBLIOGRAPHY

Chapter 1.

[1] G.M. Cragg, D.J. Newman, Biodiversity: A continuing source of novel drug leads, Pure Appl Chem. 77 (2005) 7-24.

[2] H.F. Ji, X.J. Li, H.Y. Zhang, Natural products and drug discovery, EMBO rep. 10 (2009) 194-200.

[3] D.J. Newman, Natural products as leads to potential drugs: an old process or the new hope for drug discovery?, J Med Chem. 51 (2008) 2589-2599.

[4] W.J. Zhang, G.H. Liu, Coevolution: A synergy in biology and ecology, Selforganizology. 2 (2015) 35-38.

[5] S.J. Cutler, H.G. Cutler, Biologically Active Natural Products: Pharmaceuticals, CRC Press, New York, 2000.

[6] B. delasHeras, B. Rodríguez, L. Boscá, A.M. Villar, Terpenoids: Sources, Structure Elucidation and Therapeutic Potential in Inflammation, Curr Top Med Chem. 3 (2003) 171-185

[7] D. Tholl, Biosynthesis and biological functions of terpenoids in plants, Adv Biochem Eng Biotechnol. 148 (2015) 63-106.

[8] Z.L. Ruzicka, The isoprene rule and the biogenesis of terpenic compounds, Experientia. 9 (1953) 357–367.

[9] J. Gershenzon, W. Kreis, Biosynthesis of monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins, Biochemistry of plant secondary metabolites,

Annual Plant Reviews, in: M. Wink (Ed), Sheffield Academic Press: Sheffield, UK, Vol. 2, 1999, pp. 222-299.

[10] J. Gershenzon, N. Dudareva, The function of terpene natural products in the natural world, Nat Chem Biol. 3 (2007) 408–414.

[11] M. Rohmer, The Discovery of a Mevalonate-Independent Pathway for Isoprenoid Biosynthesis in Bacteria, Algae and Higher Plants, Nat Prod Rep. 16 (1999) 565-574.

[12] I.L Finar, Stereochemistry and the chemistry of natural products, Pearson Education, Singapore, Vol. 2, 2004.

[13] P. Crabbé, Ord and Cd in chemistry and biochemistry, Academic Press, London, 1972.

[14] N. Harada, K. Nakanishi, Circular dichroic spectroscopy, exciton coupling in organic stereochemistry, University Science Books, Mill Valley, CA, 1983.

[15] E.L. Eliel, S.H. Wilen, Stereochemistry of organic compounds, Wiley-Interscience, New York, 1994.

[16] A.A. Saddiq, S.A. Khayyat, Chemical and antimicrobial studies of monoterpene: Citral, Pestic Biochem Physiol. 98 (2010) 89–93.

[17] C. de Bona da Silva, S.S. Guterres, V. Weisheimer, E.E.S. Schapoval, Antifungal activity of the lemongrass oil and citral against Candida spp., Braz J Infect Dis. 12 (2008) 63-66.

[18] K. Ghosh, Anticancer effect of lemongrass oil and citral on cervical cancer cell lines, Phcog Commn. 3 (2013) 41-48.

[19] W. Chaouki, D.Y. Leger, B. Liagre, J.L. Beneytout, M. Hmamouchi, Citral inhibits cell proliferation and induces apoptosis and cell cycle arrest in MCF-7 cells, Fundam Clin Pharmacol. 23 (2009) 549–556.

[20] H. Xia, W. Liang, Q. Song, X. Chen, X. Chen, J. Hong, . The in vitro study of apoptosis in NB4 cell induced by citral, Cytotechnology 65 (2013) 49-57.

[21] S. Zeng, A. Kapur, M.S. Patankar, M.P. Xiong, Formulation, characterization, and antitumor properties of trans- and cis-citral in the 4T1 breast cancer Xenograft mouse model, Pharm Res. 32 (2015) 2548-2558.

[22] P.L. Crowell, Monoterpenes in breast cancer chemoprevention, Breast Cancer Res Treat. 46 (1997) 191–197.

[23] P.L. Crowell, Prevention and therapy of cancer by dietary monoterpenes, J Nutr. 129 (1999) 775S–778S.

[24] M.N. Gould, Cancer chemoprevention and therapy by monoterpenes, Environ. Health Perspect. 105 (1997) 977–979.

[25] K.H. Wagner, I. Elmadfa, Biological relevance of terpenoids-overview focusing on mono, di and tetraterpenes, Ann Nutr Metab. 47 (2003) 95-106.

[26] H.S. Choi, H.S. Song, H. Ukeda, M. Sawamura, Radical scavenging activities of citrus essential oils and their components: detection using 1,1-diphenyl-2-picrylhydrazyl, J Agric Food Chem. 48 (2000) 4156–4161.

[27] L. Caputi, E. Aprea, Use of terpenoids as natural flavouring compounds in food industry, Recent Pat Food Nutr Agric. 3 (2011) 9-16.

[28] J.H. Joo, A.M. Jetten, Molecular mechanisms involved in farnesol-induced apoptosis, Cancer Lett. 287 (2010) 123-135.

[29] L. Kromidas, E. Perrier, J. Flanagan, R. Rivero, I. Bonnet, Release of antimicrobial actives from microcapsules by the action of axillary bacteria, Int J Cosmet Sci. 28 (2006) 103-108.

[30] A. Bahi, S.A Mansouri, E.A Memari, M.A Ameri, S.M Nurulain, S. Ojha, β-Caryophyllene, a CB_2 receptor agonist produces multiple change relevant toanxiety and depression in mice, Physiol Behav. 135 (2014) 119-124.

[31] M. Seo, T. Koshiba, Complex regulation of ABA biosynthesis in plants, trends in plant sci. 7 (2002) 41-48.

[32] T.R. Evans, S.B. Kaye, Retinoids: Present role and future potential, Br J Cancer. 80 (1999) 1-8.

[33] R. Gupta, S.K. Chakrabarty, Gibberellic acid in plant Still a mystery unresolved, Plant Signal Behav. 8 (2013) e25504.

[34] J.J. Johnson, Carnosol: A promising anti-cancer and anti-inflammatory agent, Cancer Lett. 305 (2011) 1-7.

[35] S. Birtić, P. Dussort, F.X. Pierre, A.C. Bily, M. Roller, Carnosic acid, Phytochemistry. 115 (2015) 9-19.

[36] G.E. Bartley, P.A. Scolnik, Plant carotenoids: pigments for photoprotection, visual Attraction, and human health, Plant Cell. 7 (1995) 1027-1038.

[37] J. Grassmann, Terpenoids as plant antioxidants, Vitam Horm. 72 (2005) 505-535.

[38] A.G. Atanasov, B. Waltenberger, E.M.P. Wenzig , T. Linder , C. Wawrosch, P. Uhrin, V. Temml, L. Wang, S. Schwaiger, E. Heiss, J.M. Rollinger, D. Schuster, J.M. Breuss, V. Bochkov, M.D. Mihovilovic, B. Kopp, R. Bauer, V.M. Dirsch, H. Stuppner, Discovery and resupply of pharmacologically active plant-derived natural products: a review, Biotechnol Adv. 33 (2015) 1582-1614.

[39] M. Bollen, S. Keppens, W. Stalmans, Specific features of glycogen metabolism in the liver, Biochem J. 336 (1998) 19-31.

[40] H.G. Hers, The control of glycogen metabolism in the liver, Annu Rev Biochem. 45 (1976) 167-190.

[41] W. Stalmans, M. Laloux, H.G. Hers, The interaction of liver phosphorylase a with glucose and AMP, Eur J Biochem. 49 (1974) 415-427.

[42] W. Pimenta, N. Nurjhan, P.A. Jansson, M. Stumvoll, J. Gerich, M. Korytkowski, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia. 37 (1994) 697-702.

[43] M.K. Hellerstein, R.A. Neese, P. Linfoot, M. Christiannsen, S. Turner, A. Letscher, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest. 100 (1997) 1305-1319.

[44] N.G. Oikonomakos, E.D. Chrysina, M.N. Kosmopoulou, D.D. Leonidas, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 Å resolution, Biochim Biophys Acta. 1647 (2003) 325-332.

[45] V.L. Rath, M. Ammirati, D.E. Danley, J.L. Ekstrom, E.M. Gibbs, T.R. Hynes, A.M. Mathiowetz, R.K. McPherson, T.V. Olson, J.L. Treadway, D.J. Hoover, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol. 7 (2000) 677-682.

[46] R. Kurukulasuriya, J.T. Link, D.J. Madar, Z. Pei, S.J. Richards, J.J. Rohde, A.J. Souers, B.G. Szczepankiewicz, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem. 10 (2003) 123-153.

[47] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[48] X. Wen, H. Sun, J. Liu, K. Cheng, P. Zhang, L. Zhang, J. Hao, L. Zhang, P. Ni, S.E. Zographos, D.D. Leonidas, K.M. Alexacou, T. Gimisis, J.M. Hayes, N.G. Oikonomakos, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem. 51 (2008) 3540-3554.

[49] T. Hunter, Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling, Cell. 80 (1995) 225-236.

[50] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, Protein tyrosine phosphatases in the human genome, Cell. 117 (2004) 699-711.

[51] Z.Y. Zhang, S.Y. Lee, PTP1B inhibitors as potential therapeutics in the treatment of type 2 diabetes and obesity, Expert Opin Investig Drugs. 12 (2003) 223-233.

[52] A.P. Combs, Recent advances in the discovery of competitive protein tyrosine phosphatase 1B inhibitors for the treatment of diabetes, obesity, and cancer, J Med Chem. 53 (2010) 2333-2344.

[53] S. Thareja, S. Aggarwal, T.R. Bhardwaj, M. Kumar, Protein tyrosine phosphatase 1B inhibitors: a molecular level legitimate approach for the management of diabetes mellitus, Med Res Rev. 32 (2012) 459-517.

[54] R.E. Cebula, J.L. Blanchard, M.D. Boisclair, K. Pal, N.J. Bockovich, Synthesis and phosphatase inhibitory activity of analogs of sulfircin, Bioorg Med Chem Lett. 7 (1997) 2015-2020.

[55] R.M. Chen, L.H. Hu, T.Y. An, J. Li, Q. Shen, Natural PTP1B inhibitors from Broussonetia papyrifera, Bioorg Med Chem Lett. 12 (2002) 3387-3390.

[56] C.S. Jiang, L.F. Liang, Y.W. Guo, Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades, Acta Pharmacologica Sinica. 33 (2012) 1217-1245.

[57] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[58] E.I. Solomon, J. Zhou, F. Neese, E.G. Pavel, New insights from spectroscopy into the structure/function relationships of lipoxygenases, Chem Biol. 4 (1997) 795-808.

[59] H.W. Gardner, Biological roles and biochemistry of the lipoxygenase pathway, Hortscience. 30 (1995) 197-205.

[60] R. Wisastra, F.J. Dekker, Inflammation, cancer and oxidative lipoxygenase activity are intimately linked, Cancers. 6 (2014) 1500-1521.

[61] L.A. Dailey, P. Imming, 12-Lipoxygenase: classification, possible therapeutic benefits from inhibition, and inhibitors, Curr Med Chem. 6 (1999) 389-398.

[62] V.E. Steele, C.A. Holmes, E.T. Hawk, L. Kopelovich, R.A. Lubet, J.A. Crowell, C.C. Sigman, G.J. Kelloff, Lipoxygenase inhibitors as potential cancer chemopreventives, Cancer Epidem Biomar. 8 (1999) 467-483.

[63] B. Samuelsson, S.E. Dahlen, J.A. Lindgren, C.A. Rouzer, C.N. Serhan, Leukotrienes and lipoxins: structures, biosynthesis, and biological effects, Science. 237 (1987) 1171-1176.

[64] X.Z. Ding, W.G. Tong, T.E. Adrian, 12-Lipoxygenase metabolite 12(S)-HETE stimulates human pancreatic cancer cell proliferation via protein tyrosine phosphorylation and ERK activation, Int J Cancer. 94 (2001) 630-636.

[65] J.A. Cornicelli, B.K. Trivedi, 15-Lipoxygenase and its inhibition: a novel therapeutic target for vascular disease, Curr Pharm Des. 5 (1999) 11-20.

[66] U.P. Kelavkar, J.B. Nixon, C. Cohen, D. Dillehay, T.E. Eling, K.F. Badr, Overexpression of 15-lipoxygenase-1 in PC-3 human prostate cancer cells increases tumorigenesis, Carcinogenesis. 22 (2001) 1765-1773.

[67] B.D. Roebuck, D.S. Longnecker, K.J. Baumgartner, C.D. Thron, Carcinogen-induced lesions in the rat pancreas: effects of varying levels of essential fatty acid, Cancer Res. 45 (1985) 5252-5256.

[68] J. Zhang, V.L. Go, High fat diet, lipid peroxidation, and pancreatic carcinogenesis, Adv Exp Med Biol. 399 (1996) 165-172.

[69] D.P. Rose, Dietary fatty acids and cancer, Am J Clin Nutr. 66 (1997) 998S-1003S.

[70] P.L. Zock, M.B. Katan, Linoleic acid intake and cancer risk: a review and meta-analysis, Am J Clin Nutr. 68 (1998) 142-153.

[71] W.G. Tong, X.Z. Ding, R.C. Witt, T.E. Adrian, Lipoxygenase inhibitors attenuate growth of human pancreatic cancer xenografts and induce apoptosis through the mitochondrial pathway, Mol Cancer Therapeutics. 1 (2002) 929-935.

[72] C. Kemal, P. Louis-Flamberg, R. Krupinski-Olsen, A.L. Shorter, Reductive inactivation soybean lipoxygenase-1 by catechols: A possible mechanism for regulation of lipoxygenase activity, Biochemistry. 26 (1987) 7064-7072.

[73] R. Mogul, E. Johansen, T.R. Holman, Oleyl sulfate reveals allosteric inhibition of soybean lipoxygenase-1 and human 15-lipoxygenase, Biochemistry. 39 (2000) 4801-4807.

[74] A.F.A. Cros, Ph.D. dissertation thesis: Action de l'alcohol amylique sur l'organisme, University of Strasbourg, Strasbourg, 1863.

[75] A. Crum-Brown, T.R. Fraser, On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived

from strychia, brucia, thebaia, codeia, morphia and nicotia, Trans R Soc Edinburgh. 25 (1868) 151-203.

[76] C. Richet, On the relationship between the toxicity and the physical properties of substances, Compt Rendus Seances Soc Biol. 9 (1893) 775-776.

[77] E. Overton, Osmotic properties of cells in the bearing on toxicology and pharmacology, Z Physik Chem. 22 (1897) 189-209.

[78] H. Meyer, On the theory of alcohol narcosis I. Which property of anesthetics gives them their narcotic activity?, Arch Exp Pathol Pharmakol. 42 (1899) 109-118.

[79] L.P. Hammett, Some relations between reaction rates and equilibrium constants, Chem Rev. 17 (1935) 125-136.

[80] L.P. Hammett, The effect of structure upon the reactions of organic compounds. benzene derivatives, J Am Chem Soc. 59 (1937) 96-103.

[81] R.W. Taft, Polar and steric substituent constants for aliphatic and o- Benzoate groups from rates of esterification and hydrolysis of esters[1], J Am Chem Soc. 74 (1952) 3120-3128.

[82] C. Hansch, T. Fujita, p-σ-π Analysis. A method for the correlation of biological activity and chemical structure, J Am Chem Soc. 86 (1964) 1616-1626.

[83] C. Hansch, Quantitative approach to biochemical structure-activity relationships, Acc Chem Res. 2 (1969) 232-239.

[84] S.M. Free Jr, J.W. Wilson, A Mathematical contribution to structure-activity studies, J Med Chem. 7 (1964) 395-399.

[85] T. Fujita, T. Ban, Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters, J Med Chem. 14 (1971) 148-152.

[86] L.B. Kier, Molecular orbital theory in drug research, Academic Press, New York, 1971.

[87] A.T. Balaban, F. Harary, The characteristic polynomial does not uniquely determine the topology of a molecule, J Chem Doc. 11 (1971) 258-259.

[88] M. Randić, On the recognition of identical graphs representing molecular topology, J Chem Phys. 60 (1974) 3920-3928.

[89] L.B. Kier, L.H. Hall, W.J. Murray, M. Randic, Molecular connectivity I: relationship to nonspecific local anesthesia, J Pharm Sci. 64 (1975) 1971-1974.

[90] R.H. Rohrbaugh, P.C. Jurs, Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships, Anal Chim Acta. 199 (1987) 99-109.

[91] D.T. Stanton, P.C. Jurs, Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies, Anal Chem. 62 (1990) 2323-2329.

[92] R. Todeschini, M. Lasagni, E. Marengo, New molecular descriptors for 2D- and 3D-structures, Theory J Chemom. 8 (1994) 263-273.

[93] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, J Phys Chem. 100 (1996) 10400-10407.

[94] A.M. Ferguson, T.W. Heritage, P. Jonathon, S.E. Pack, L. Phillips, J. Rogan, P.J. Snaith, EVA: a new theoretically based molecular descriptor for use in QSAR/QSPR analysis, J Comput Aided Mol Des. 11 (1997) 143-152.

[95] J. Schuur, P. Selzer, J. Gasteiger, The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity, J Chem Inf Comput Sci. 36 (1996) 334-344.

[96] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors, J Chem Inf Comput Sci. 42 (2002) 682-692.

[97] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J Am Chem Soc. 110 (1988) 5959-5967.

[98] J. Gasteiger, Handbook of chemoinformatics. From data to knowledge in 4 volumes, Wiley-VCH, Weinheim, 2003.

[99] T.I. Oprea, 3D QSAR modeling in drug design, in: P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere (Eds.), Computational medicinal chemistry for drug discovery. Marcel Dekker, New York, 2004.

[100] R. Franke, Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.

[101] K. Osmialowski, J. Halkiewicz, A. Radecki, R. Kaliszan, Quantum chemical parameters in correlation analysis of gas-liquid chromatographic retention indices of amines, J Chromatogr. 346 (1985) 53-60.

[102] K. Fukui, Theory of Orientation and Stereoselection, Springer- Verlag, New York, 1975.

[103] H. Sklenar, J. Jäger, Molecular structure-biological activity relationships on the basis of quantum-chemical calculations, Int J Quant Chem. 16 (1979) 467-484.

[104] K. Tuppurainen, S. Lötjönen, R. Laatikainen, T. Vartiainen, U. Maran, M. Strandberg, T. Tamm, About the mutagenicity of chlorine-substituted furanones and halopropenals. A QSAR study using molecular orbital indices, Mutat Res Fund Mol Mech Mut. 247 (1991) 97-102.

[105] I. Fleming, Frontier orbitals and organic chemical reactions, John Wiley & Sons, New York, 1976.

[106] D.F.V. Lewis, C. Ioannides, D.V. Parke, Interaction of a series of nitriles with the alcohol inducible isoform of P450- computer analysis of structure-activity relationships, Xenobiotica. 24 (1994) 401-408.

[107] M. Karelson, V.S. Lobanov, Quantum-Chemical Descriptors in QSAR/QSPR Studies, Chem Rev. 96 (1996) 1027-1043.

Chapter 2.

[1] S.B. Mahato, S. Sen, Advances in triterpenoid research, 1990-1994, Phytochemistry. 44 (1997) 1185-1236.

[2] I.H. Hall, K.H. Lee, C.O. Starnes, Y. Sumida, R.Y. Wu, T.G. Waddell, J.W. Cochran, K.G. Gerhart, Anti- inflammatory activity of sesquiterpene lactones and related compounds, J Pharm Sci. 68 (1979) 537-542.

[3] C.H. Chen, L.M. Yang, T.T. Lee, Y.C. Shen, D.C. Zhang, D.J. Pan, A.T. McPhail, D.R. McPhail, S.Y. Liu, D.H. Li, Antitumor agents-CLI. Bis(helenalinyl)glutarate and bis(isoalantodiol- B)glutarate, potent inhibitors of human DNA topoisomerase II, Bioorg Med Chem. 2 (1994) 137-145.

[4] K. Hayashi, H. Hayashi, N. Hiraoka, Y. Ikeshiro, Inhibitory activity of soyasaponin II on virus replication in vitro, Planta Med. 63 (1997) 102-105.

[5] M.C. Recio, R.M. Giner, L. Uriburu, S. Manez, M. Cerda, J.R. De la Fuente, J.L. Rios, In vivo activity of pseudoguaianolide sesquiterpene lactones in acute and chronic inflammation, Life Sci. 66 (2000) 2509-2518.

[6] C.L. Cantrell, S.G. Franzblau, N.H. Fischer, Antimycobacterial plant terpenoids, Planta Med. 67 (2001) 685-694.

[7] N. Osaki, T. Koyano, T. Kowithayakorn, M. Hayashi, K. Komiyama, M. Ishibashi, Sesquiterpenoids and plasmin-inhibitory flavonoids from Blumea balsamifera, J Nat Prod. 68 (2005) 447-449.

[8] D.B. Stierle, A.A. Stierle, T. Girtsman, K. McIntyre, J. Nichols, Caspase-1 and -3 inhibiting drimane sesquiterpenoids from the extremophilic fungus Penicillium solitum, J Nat Prod. 75 (2012) 262-266.

[9] Q.Y. Ma, Y.C. Chen, S.Z. Huang, Z.K. Guo, H.F. Dai, Y. Hua, Y.X. Zhao, Two new guaiane sesquiterpenoids from Daphne holosericea (Diels) Hamaya, Molecules. 19 (2014) 14266-14272.

[10] Q.Q. Tao, K. Ma, L. Bao, K. Wang, J.J. Han, J.X. Zhang, C.Y. Huang, H.W. Liu, New sesquiterpenoids from the edible mushroom Pleurotus cystidiosus and their inhibitory activity against α-glucosidase and PTP1B, Fitoterapia. 111 (2016) 29-35.

[11] K. Dimas, C. Demetzos, M. Marsellos, R. Sotiriadou, M. Malamas, D. Kokkinopoulos, Cytotoxic activity of labdane typed diterpenes against human leukemic cell lines in vitro, Planta Med. 64 (1998) 208-211.

[12] R. Batista, E. Chiari, A.B. de Oliveira, Trypanosomicidal kaurane diterpenes from Wedelia paludosa, Planta Med. 65 (1999) 283-284.

[13] M. Singh, M. Pal, R.P. Sharma, Biological activity of the labdane diterpenes, Planta Med. 65 (1999) 2-8.

[14] M. DellaGreca, A. Fiorentino, M. Isidori, P. Monaco, A. Zarrelli, Antialgal ent-labdane diterpenes from Ruppia maritime, Phytochemistry. 55 (2000) 909-913.

[15] R. Tanaka, H. Ohtsu, M. Iwamoto, T. Minami, H. Tokuda, H. Nishino, S. Matsunaga, A. Yoshitake, Cancer chemopreventive agents, labdane diterpenoids from the stem bark of Thuja standishii (Gord.) Carr, Cancer Lett. 161 (2000) 165-170.

[16] A. Navarro, B. de las Heras, A.M. Villar, Andalusol, a diterpenoid with anti-inflammatory activity from Siderits foetens Clemen, Z Naturforsch. 52c (1997) 844-849.

[17] M.A. Fernandez, M.P. Tornos, M.D. Garcia, B. de las Heras, A.M. Villar, M.T. Saenz, Anti-inflammatory activity of abietic acid, a diterpene isolated from Pimenta racemosa var. grissea, J Pharm Pharmacol. 53 (2001) 867-872.

[18] X. Chen, J. Ding, Y.M. Ye, J.S. Zhang, Bioactive abietane and seco-abietane diterpenoids from Salvia prionitis, J Nat Prod. 65 (2002) 1016-1020.

[19] M. Na, W.K. Oh, Y.H. Kim, X.F. Cai, S. Kim, B.Y. Kim, J.S. Ahn, Inhibition of protein tyrosine phosphatase 1B by diterpenoids isolated from Acanthopanax koreanum, Bioorg Med Chem Lett. 16 (2006) 3061-3064.

[20] A. Leverrier, M.T. Martin, C. Servy, J. Ouazzani, P. Retailleau, K. Awang, M.R. Mukhtar, F. Guéritte, M. Litaudon, Rearranged diterpenoids from the biotransformation of ent-trachyloban-18-oic acid by Rhizopus arrhizus, J Nat Prod. 73 (2010) 1121-1125.

[21] L. Yang, L. Qiao, C. Ji, D. Xie, N.B. Gong, Y. Lu, J. Zhang, J. Dai, S. Guo, Antidepressant abietane diterpenoids from chinese eaglewood, J Nat Prod. 76 (2013) 216-222.

[22] J.X. Zhao, C.P. Liu, W.Y. Qi, M.L. Han, Y.S. Han, M.A. Wainberg, J.M. Yue, Eurifoloids A-R, structurally diverse diterpenoids from Euphorbia neriifolia, J Nat Prod. 77 (2014) 2224-2233.

[23] G. Luo, Q.Ye, B. Du, F. Wang, G. Zhang, Y. Luo, Iridoid glucosides and diterpenoids from Caryopteris glutinosa, J Nat Prod. 79 (2016) 886-893.

[24] H. Nishino, K. Yoshioka, A. Iwashima, H. Takizawa, S. Konishi, H. Okamoto, H. Okabe, S. Shibata, H. Fujiki, T. Sugimura, Glycyrrhetic acid inhibits tumor promoting activity of teleocidin and 12-O-tetradecanoylphorbol-13-acetate in two-stage mouse skin carcinogenesis, Jpn J Cancer Res. 77 (1986) 33-38.

[25] H. Tokuda, H. Ohigashi, K. Koshimizu, Y. Ito, Inhibitory effects of ursolic and oleanolic acid on skin tumor promotion by 12-O-tetradecanoylphorbol-13-acetate, Cancer Lett. 33 (1986) 279-285.

[26] M.C. Recio, R.M. Giner, S. Manez, J.L. Rios, Structural requirements for the anti-inflammatory activity of natural triterpenoids, Planta Med. 61 (1995) 182-185.

[27] F. Soler, C. Poujade, M. Evers, J.C. Carry, Y. Henin, A. Bousseau, T. Huet, R. Pauwels, E. De Clercq, J.F. Mayaux, J.B. Le Pecq, N. Dereu, Betulinic acid derivatives: a new class of specific inhibitors of human immunodeficiency virus type 1 entry, J Med Chem. 39 (1996) 1069-1083.

[28] H. Ito, S. Onoue, Y. Miyake, T. Yoshida, Iridal-type triterpenoids with ichthyotoxic activity from Belamcanda chinensis, J Nat Prod. 62 (1999) 89-93.

[29] A. Rajic, G. Kweifio-Okai, T. Macrides, R.M. Sandeman, D.S. Chandler, G.M. Polya, Inhibition of serine proteases by anti-inflammatory triterpenoids, Planta Med. 66 (2000) 206-210.

[30] C.L. Cantrell, S.G. Franzblau, N.H. Fischer, Antimycobacterial plant terpenoids, Planta Med. 67 (2001) 685- 694.

[31] T. Fujioka, Y. Kashiwada, R.E. Kilkuskie, L.M. Cosentino, L.M. Ballas, J.B. Jiang, W.P. Janzen, I.S. Chen, K.H. Lee, Anti-AIDS agents, 11. Betulinic acid and platanic acid as anti-HIV principles from Syzigium claviflorum, and the anti-HIV activity of structurally related triterpenoids, J Nat Prod. 57 (1994) 243-247.

[32] Z.Z. Ma, Y. Hano, T. Nomura, Y.J. Chen, Three new triterpenoids from Peganum nigellastrum, J Nat Prod. 63 (2000) 390-392.

[33] S. Wada, A. Iida, R. Tanaka, Screening of triterpenoids isolated from Phyllanthus flexuosus for DNA topoisomerase inhibitory activity, J Nat Prod. 64 (2001) 1545-1547.

[34] K. Yoshikawa, M. Inoue, Y. Matsumoto, C. Sakakibara, H. Miyataka, H. Matsumoto, S. Arihara, Lanostane triterpenoids and triterpene glycosides from the fruit body of

Fomitopsis pinicola and their inhibitory activity against COX-1 and COX-2, J Nat Prod. 68 (2005) 69-73.

[35] P.T. Thuong, C.H. Lee, T.T. Dao, P.H. Nguyen, W.G. Kim, S. J. Lee, W.K. Oh, Triterpenoids from the leaves of Diospyros kaki (persimmon) and their inhibitory effects on protein tyrosine phosphatase 1B, J Nat Prod. 71 (2008) 1775-1778.

[36] M. Litaudon, C. Jolly, C.L. Callonec, D. D. Cuong, P. Retailleau, O. Nosjean, V.H. Nguyen, B. Pfeiffer, J.A. Boutin, F. Guéritte, Cytotoxic pentacyclic triterpenoids from Combretum sundaicum and Lantana camara as inhibitors of Bcl-xL/BakBH3 domain peptide interaction, J Nat Prod. 72 (2009) 1314-1320.

[37] C. Chen, S. Qiang, L. Lou, W. Zhao, Cucurbitane type triterpenoids from the stems of Cucumis melo, J Nat Prod. 72 (2009) 824-829.

[38] P. Wang, S. Ownby, Z. Zhang, W. Yuan, S. Li, Cytotoxicity and inhibition of DNA topoisomerase I of polyhydroxylated triterpenoids and triterpenoid glycosides, Bioorg Med Chem Lett. 20 (2010) 2790-2796.

[39] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[40] J.D.S. Mpetga, Y. Shen, P. Tane, S.F. Li, H.P. He, H.K. Wabo, M. Tene, Y. Leng, X.J. Hao, Cycloartane and friedelane triterpenoids from the leaves of Caloncoba glauca and their evaluation for inhibition of 11β- hydroxysteroid dehydrogenases, J Nat Prod. 75 (2012) 599-604.

[41] L. Xiong, M. Zhu, C. Zhu, S. Lin, Y. Yang, J. Shi, Structure and bioassay of triterpenoids and steroids isolated from Sinocalamus affinis, J Nat Prod. 75 (2012) 1160-1166.

[42] C.L. Zhang, Y. Wang, Y.F. Liu, G. Ni, D. Liang, H. Luo, X.Y. Song, W.Q. Zhang, R.Y. Chen, N.H. Chen, D.Q. Yu, Iridal-type triterpenoids with neuroprotective activities from Iris tectorum, J Nat Prod. 77 (2014) 411-415.

[43] G.W. Wang, C. Lv, X. Fang, X.H. Tian, J. Ye, H.L. Li, L. Shan, Y.H. Shen, W.D. Zhang, Eight pairs of epimeric triterpenoids involving a characteristic spiro-E/F ring from Abies faxoniana, J Nat Prod. 78 (2015) 50-60.

[44] L.P. Hammett, Some relations between reaction rates and equilibrium constants, Chem Rev. 17 (1935) 125-136.

[45] L.P. Hammett, Physical Organic Chemistry, 2nd ed., McGraw-Hill, New York, 1970.

[46] R.W. Taft Jr., Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters[1], J Am Chem Soc. 74 (1952) 3120-3128.

[47] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients, Nature. 194 (1962) 178-180.

[48] C. Hansch, T. Fujita, ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure, J Am Chem Soc. 86 (1964) 1616-1626.

[49] C. Hansch, Quantitative approach to biochemical structure-activity relationships, Acc Chem Res. 2 (1969) 232-239.

[50] S.M. Free Jr., J.W. Wilson, A mathematical contribution to structure-activity studies, J Med Chem. 7(1964) 395-399.

[51] Z. Simon, Specific interactions. Intermolecular forces, steric requirements, and molecular size, Angew Chem Int Ed Eng. 13 (1974) 719-727.

[52] L.H. Hall, L.B. Kier, Structure-activity studies using valence molecular connectivity, J Pharm Sci. 66 (1977) 642-644.

[53] L.B. Kier, L.H. Hall, Molecular structure description: the electrotopological state, Academic Press, San Diego, CA, 1999.

[54] W. Tong, D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sleehan, Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor, J Chem Inf Comput Sci.38 (1998) 669-677.

[55] S.J. Cho, W. Zheng, A. Tropsha, Focus-2D: a new approach to the design of targeted combinatorial chemical libraries, Pac Symp Biocomput. (1998) 305-316.

[56] H. Gao, J. Bajorath, Comparison of binary and 2D QSAR analyses using inhibitors of human carbonic anhydrase II as a test case, J Mol Diversity. 4 (1998) 115-130.

[57] H. Gao, C. Williams, P. Labute, and J. Bajorath, Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands, J Chem Inf Comput Sci. 39 (1999) 164-168.

[58] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J Am Chem Soc. 110 (1988) 5959- 5967.

[59] H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in drug design, ligand-protein interactions and molecular similarity (Vol. 2) and recent advances (Vol. 3), Kluwer Academic Publishers, the Netherlands, 1997.

[60] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J Med Chem. 37 (1994) 4130-4146.

[61] D.D. Robinson, P.J. Winn, P.D. Lyne, W.G Richards, Self-organizing molecular field analysis: a tool for structure-activity studies, J Med Chem. 42 (1999) 573-583.

[62] H.L. Jiang, K.X. Chen, H.W. Wang, Y. Tang, J.Z. Chen, R.Y. Ji, 3D-QSAR study on ether and ester analogs of artemisinin with comparative molecular field analysis, Zhongguo Yao Li Xue Bao. 15 (1994) 481-487.

[63] J.Z. Chen, L.H. Hu, H.L. Jiang, J.D. Gu, W, L. Zhu, Z.L. Chen, K.X. Chen, R.Y. Ji, A 3D-QSAR study on ginkgolides and their analogues with comparative molecular field analysis, Bioorg Med Chem Lett. 8 (1998) 1291-1296.

[64] J.R. Woolfrey, M.A. Avery, A.M. Doweyko, 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 Des. 12 (1998) 165-181.

[65] M. Jalali-Heravi, M.H. Fatemi, Artificial neural network modeling of Kováts retention indices for noncyclic and monocyclic terpenes, J Chromatogr A. 915 (2001) 177-183.

[66] J.A. Grodnitzky, J.R. Coats, QSAR evaluation of monoterpenoids' insecticidal activity, J Agric Food Chem. 50 (2002) 4576-4580.

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

[68] T. Kuriyama, T.J. Schmidt, E. Okuyama, Y. Ozoe, Structure-activity relationships of seco-prezizaane terpenoids in gamma-aminobutyric acid receptors of houseflies and rats, Bioorg Med Chem. 10 (2002) 1873-1881.

[69] M.A. Avery, M. Alvim-Gaston, C.R. Rodrigues, E.J. Barreiro, F.E. Cohen, Y.A. Sabnis, J.R. Woolfrey, Structure-activity relationships of the antimalarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies, J Med Chem. 45 (2002) 292-303.

[70] T. Ghafourian, P. Zandasrar, H. Hamishekar, A. Nokhodchi, The effect of penetration enhancers on drug delivery through skin: a QSAR study, J Control Release. 99 (2004) 113-125.

[71] R. Guha, P.C. Jurs, Development of QSAR models to predict and interpret the biological activity of artemisinin analogues, J Chem Inf Comput Sci. 44 (2004) 1440-1449.

[72] T.J. Schmidt, M. Gurrath, Y. Ozoe, Structure-activity relationships of seco-prezizaaneandpicrotoxane/picrodendrane terpenoids by Quasar receptor-surface modeling, Bioorg Med Chem. 12 (2004) 4159-4167.

[73] F. Cortés-Selva, M. Campillo, C.P. Reyes, I.A. Jiménez, S. Castanys, I.L. Bazzocchi, L. Pardo, F. Gamarro, A.G. Ravelo, SAR studies of dihydro-beta-agarofuran sesquiterpenes as inhibitors of the multidrug-resistance phenotype in a Leishmania tropica line overexpressing a P-glycoprotein-like transporter, J Med Chem. 47 (2004) 576-587.

[74] B. Siedle, A.J. García-Piñeres, R. Murillo, J. Schulte-Mönting, V. Castro, P. Rüngeler, C.A. Klaas, F.B. Da Costa, W. Kisiel, I. Merfort, Quantitative structure-activity relationship of sesquiterpene lactones as inhibitors of the transcription factor NF-kappaB, J Med Chem. 47 (2004) 6042-6054.

[75] F.A. Macías, R.F. Velasco, D. Castellano, J.C. Galindo, Application of Hansch's model to guaianolide ester derivatives: a quantitative structure-activity relationship study, J Agric Food Chem. 53 (2005) 3530-3539.

[76] W. Zhu, G. Chen, L. Hu, X. Luo, C. Gui, C. Luo, C.M. Puah, K. Chen, J.H. Jiang, QSAR analyses on ginkgolides and their analogues using CoMFA, CoMSIA, and HQSAR, Bioorg Med Chem. 13 (2005) 313-322.

[77] M. Ishihara, H. Sakagami, W.K. Liu, Quantitative structure-cytotoxicity relationship analysis of betulinic acid and its derivatives by semi-empirical molecular-orbital method, Anticancer Res. 25 (2005) 3951-3955.

[78] S.L. Cunningham, A.R. Cunningham, B.W. Day, CoMFA, HQSAR and molecular docking studies of butitaxel analogues with beta-tubulin, J Mol Model. 11 (2005) 48-54.

[79] H. Parra-Delgado, C.M. Compadre, T. Ramírez-Apan, M.J. Muñoz-Fambuena, R.L. Compadre, P. Ostrosky-Wegman, M. Martínez-Vázquez, Synthesis and comparative molecular field analysis (CoMFA) of argentatin B derivatives as growth inhibitors of human cancer cell lines, Bioorg Med Chem. 14 (2006) 1889-1901.

[80] Y.H. Chung, S.C. Jang, S.J. Kim, N.D. Sung, 2D-QSAR and HQSAR on the inhibition activity of protein tyrosine phosphatase 1B with oleanolic acid analogues, J Appl Biol Chem. 50 (2007) 52-57.

[81] B.F. Rasulev, A.I. Saidkhodzhaev, S.S. Nazrullaev, K.S. Akhmedkhodzhaeva, Z.A. Khushbaktova, J. Leszczynski, Molecular modelling and QSAR analysis of the estrogenic activity of terpenoids isolated from Ferula plants, SAR QSAR Environ Res. 18 (2007) 663-673.

[82] B. Hemmateenejad, K. Javadnia, M. Elyasi, Quantitative structure-retention relationship for the Kovats retention indices of a large set of terpenes: a combined data splitting-feature selection strategy, Anal Chim Acta. 592 (2007) 72-81.

[83] M.T. Scotti, M.B. Fernandes, M.J. Ferreira, V.P. Emerenciano, Quantitative structure-activity relationship of sesquiterpene lactones with cytotoxic activity, Bioorg Med Chem. 15 (2007) 2927-2934.

[84] H.J. Chang, H.J. Kim, H.S. Chun, Quantitative structure-activity relationship (QSAR) for neuroprotective activity of terpenoids, Life Sci. 80 (2007) 835-841.

[85] C.P. Reyes, F. Muñoz-Martínez, I.R. Torrecillas, C.R. Mendoza, F. Gamarro, I.L. Bazzocchi, M.J. Núñez, L. Pardo, S. Castanys, M. Campillo, I.A.E Jiménez, Biological evaluation, structure-activity relationships, and three-dimensional quantitative structure-activity relationship studies of dihydro-beta-agarofuran sesquiterpenes as modulators of P-glycoprotein-dependent multidrug resistance, J Med Chem. 50 (2007) 4808-4817.

[86] L. Kang, C.W. Yap, P.F. Lim, Y.Z. Chen, P.C. Ho, Y.W. Chan, G.P. Wong, S.Y. Chan, Formulation development of transdermal dosage forms: quantitative structure-activity relationship model for predicting activities of terpenes that enhance drug penetration through human skin, J Control Release. 120 (2007) 211-219.

[87] W.N. Setzer, Non-intercalative triterpenoid inhibitors of topoisomerase II: A molecular docking study, TOBCJ. 1 (2008) 13-17.

[88] R.P. Verma, C. Hansch, Taxane analogues against breast cancer: a quantitative structure-activity relationship study, ChemMedChem. 3 (2008) 642-652.

[89] C. Hansch, R.P. Verma, Understanding tubulin/microtubule-taxane interactions: a quantitative structure-activity relationship study, Mol Pharm. 5 (2008) 151-161.

[90] F.J. Cardoso, A.F. de Figueiredo, M. da Silva Lobato, R.M. de Miranda, R.C. de Almeida, J.C. Pinheiro, A study on antimalarial artemisinin derivatives using MEP maps and multivariate QSAR, J Mol Model. 14 (2008) 39-48.

[91] Z. Wang, J. Song, Z. Han, Z. Jiang, W. Zheng, J. Chen, Z. Song, S. Shang, Quantitative structure-activity relationship of terpenoid aphid antifeedants, J Agric Food Chem. 56 (2008) 11361-11366.

[92] W.N. Setzer, A theoretical investigation of cytotoxic activity of celastroid triterpenoids, J Mol Model. 15 (2009) 197-201.

[93] R.P. Verma, C. Hansch, Taxane analogues against lung cancer: a quantitative structure-activity relationship study, Chem Biol Drug Des. 73 (2009) 627-636.

[94] W.E. Steinmetz, C.B. Rodarte, A. Lin, 3D QSAR study of the toxicity of trichothecene mycotoxins, Eur J Med Chem. 44 (2009) 4485-4489.

[95] T.J. Schmidt, A.M. Nour, S.A. Khalid, M. Kaiser, R. Brun, Quantitative structure-antiprotozoal activity relationships of sesquiterpene lactones, Molecules. 14 (2009) 2062-2076.

[96] C. Hansch, R.P. Verma, Overcoming tumor drug resistance with C2-modified 10-deacetyl-7-propionyl cephalomannines: a QSAR study, Mol Pharm. 6 (2009) 849-860.

[97] L. Saiz-Urra, J.C. Racero, A.J. Macías-Sáchez, R. Hernández-Galán, J.R. Hanson, M. Perez-Gonzalez, I.G. Collado, Synthesis and quantitative structure-antifungal activity relationships of clovane derivatives against Botrytis cinerea, J Agric Food Chem. 57 (2009) 2420-2428.

[98] R.J. Little, A.A. Pestano, Z. Parra, Modeling of peroxide activation in artemisinin derivatives by serial docking, J Mol Model. 15 (2009) 847-858.

[99] J. Xu, S. Huang, H. Luo, G. Li, J. Bao, S. Cai, Y. Wang, QSAR Studies on andrographolide derivatives as α-glucosidase inhibitors, Int J Mol Sci. 11 (2010) 880-895.

[100] M.E. Badawy, S.A. El-Arami, S.A. Abdelgaleil, Acaricidal and quantitative structure activity relationship of monoterpenes against the two-spotted spider mite, Tetranychus urticae, Exp Appl Acarol. 52 (2010) 261-274.

[101] R.P. Verma, C. Hansch, QSAR modeling of taxane analogues against colon cancer, Eur J Med Chem. 45 (2010) 1470-1477.

[102] D.L. McGovern, P.D. Mosier, B.L. Roth, R.B. Westkaemper, CoMFA analyses of C-2 position salvinorin A analogs at the kappa-opioid receptor provides insights into epimer selectivity, J Mol Graph Model. 28 (2010) 612-625.

[103] H.M. Hassan, A.Y. Elnagar, M.A. Khanfar, A.A. Sallam, R. Mohammed, L.A. Shaala, D.T. Youssef, M.S. Hifnawy, K.A. El Sayed, Design of semisynthetic analogues and 3D-QSAR study of eunicellin-based diterpenoids as prostate cancer migration and invasion inhibitors, Eur J Med Chem. 46 (2011) 1122-1130.

[104] S.B. Bharate, I.P. Singh, Quantitative structure-activity relationship study of phloroglucinol-terpene adducts as anti-leishmanial agents, Bioorg Med Chem Lett. 21 (2011) 4310-4315.

[105] P. Lan, W.N. Chen, P.H. Sun, W.M. Chen, 3D-QSAR studies on betulinic acid and betulin derivatives as anti-HIV-1 agents using CoMFA and CoMSIA, Med Chem Res. 20 (2011) 1247-1259.

[106] N.S. Moorthy, M.J. Ramos, P.A. Fernandes, Prediction of the relationship between the structural features of andrographolide derivatives and α-glucosidase inhibitory activity: a

quantitative structure-activity relationship (QSAR) study, J Enzyme Inhib Med Chem. 26 (2011) 78-87.

[107] P. Lan, W.N. Chen, Z.J. Huang, P.H. Sun, W.M. Chen, Understanding the structure-activity relationship of betulinic acid derivatives as anti-HIV-1 agents by using 3D-QSAR and docking, J Mol Model. 17 (2011) 1643-1659.

[108] S.P. Wei, Z.Q. Ji, H.X. Zhang, J.W. Zhang, Y.H. Wang, W.J. Wu, Isolation, biological evaluation and 3D-QSAR studies of insecticidal/narcotic sesquiterpene polyol esters, J Mol Model. 17 (2011) 681-693.

[109] M.L. Zhao, J.J. Yin, M.L. Li, Y. Xue, Y. Guo, QSAR study for cytotoxicity of diterpenoid tanshinones, Interdiscip Sci. 3 (2011) 121-127.

[110] J. Bartalis, F.T. Halaweish, In vitro and QSAR studies of cucurbitacins on HepG2 and HSC-T6 liver cell lines, Bioorg Med Chem. 19 (2011) 2757-2766.

[111] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[112] A. Maurya, F. Khan, D.U. Bawankule, D.K. Yadav, S.K. Srivastava, QSAR, docking and in vivo studies for immunomodulatory activity of isolated triterpenoids from Eucalyptus tereticornis and Gentiana kurroo, Eur J Pharm Sci. 47 (2012) 152-161.

[113] F. Tong, J.R. Coats, Quantitative structure-activity relationships of monoterpenoid binding activities to the housefly GABA receptor, Pest Manag Sci. 68 (2012) 1122-1129.

[114] K. Kalani, D.K. Yadav, F. Khan, S.K. Srivastava, N. Suri, Pharmacophore, QSAR, and ADME based semisynthesis and in vitro evaluation of ursolic acid analogs for anticancer activity, J Mol Model. 18 (2012) 3389-3413.

[115] F. Abbasitabar, V. Zare-Shahabadi, Development predictive QSAR models for artemisinin analogues by various feature selection methods: a comparative study, SAR QSAR Environ Res. 23 (2012) 1-15.

[116] W. Ding, M. Sun, S. Luo, T. Xu, Y. Cao, X. Yan, Y. Wang, A 3D QSAR study of betulinic acid derivatives as anti-tumor agents using topomer CoMFA: model building studies and experimental verification, Molecules. 18 (2013) 10228-10241.

[117] J. Song, Z. Wang, A. Findlater, Z. Han, Z. Jiang, J. Chen, W. Zheng, S. Hyde, Terpenoid mosquito repellents: A combined DFT and QSAR study, Bioorg Med Chem Lett. 23 (2013) 1245-1248.

[118] I.J. Sousa, M.J. Ferreira, J. Molnár, M.X. Fernandes, QSAR studies of macrocyclic diterpenes with P-glycoprotein inhibitory activity, Eur J Pharm Sci. 48 (2013) 542-553.

[119] S. Rajkhowa, I. Hussain, K.K. Hazarika, P. Sarmah, R.C. Deka, Quantitative structure-activity relationships of the antimalarial agent artemisinin and some of its derivatives - a DFT approach, Comb Chem High Throughput Screen. 16 (2013) 590-602.

[120] C. Schomburg, W. Schuehly, F.B. Da Costa, K.H. Klempnauer, T.J. Schmidt, Natural sesquiterpene lactones as inhibitors of Myb-dependent gene expression: structure-activity relationships, Eur J Med Chem. 63 (2013) 313-320.

[121] S. Liao, J. Song, Z. Wang, J. Chen, G. Fan, Z. Song, S. Shang, S. Chen, P. Wang, Molecular interactions between terpenoid mosquito repellents and human-secreted attractants, Bioorg Med Chem Lett. 24 (2014) 773-779.

[122] A.I. Foudah, A.A. Sallam, M.R. Akl, K.A. El Sayed, Optimization, pharmacophore modeling and 3D-QSAR studies of sipholanes as breast cancer migration and proliferation inhibitors, Eur J Med Chem. 73 (2014) 310-324.

[123] J. Vinholesa, A. Rudnitskayab, P. Gonçalvesc, F. Martelc, M.A. Coimbraa, S.M. Rocha, Hepatoprotection of sesquiterpenoids: a quantitative structure–activity relationship (QSAR) approach, Food Chem. 146 (2014) 78-84.

[124] H.G. Trossini, V.G. Maltarollo, T.J. Schmidt, Hologram QSAR studies of antiprotozoal activities of sesquiterpene lactones, Molecules. 19 (2014) 10546-10562.

[125] J.L. Cheng, M. Zheng, T.T. Yao, X.L. Li, J.H. Zhao, M. Xia, G.N. Zhu, Synthesis, antifungal activity, and QSAR study of novel trichodermin derivatives, J Asian Nat Prod. 17 (2015) 47-55.

[126] K. Muthusamy, P. Kirubakaran, G. Krishnasamy, R.R. Thanashankar, Computational insights into the inhibition of influenza viruses by rupestonic acid derivatives: pharmacophore modeling, 3D-QSAR, CoMFA and COMSIA studies, Comb Chem High Throughput Screen. 18 (2015) 63-74.

[127] M. Appell M, W.B. Bosma, Assessment of the electronic structure and properties of trichothecene toxins using density functional theory, J Hazard Mater. 288 (2015) 113-123.

[128] A. Hazra, C. Mondal, D. Chakraborty, A.K. Halder, Y.P. Bharitkar, S.K. Mondal, S. Banerjee, T. Jha, N.B. Mondal, Towards the development of anticancer drugs from andrographolide: semisynthesis, bioevaluation, QSARanalysis and pharmacokinetic studies, Curr Top Med Chem. 15 (2015) 1013-1026.

[129] M.E. Saeed, O. Kadioglu, E.J. Seo, H.J. Greten, R. Brenk, T. Efferth, Quantitative structure-activity relationship and molecular docking of artemisinin derivatives to vascular endothelial growth factor receptor 1, Anticancer Res. 35 (2015) 1929-1934.

[130] P. Tiwari, P. Sharma, F. Khan, N.S. Sangwan, B.N. Mishra, R.S. Sangwan, Structure Activity Relationship Studies of Gymnemic Acid Analogues for Antidiabetic Activity Targeting PPARγ, Curr Comput Aided Drug Des. 11 (2015) 57-71.

[131] R. Bonikowski, P. Świtakowska, M. Sienkiewicz, M. Zakłos-Szyda, Selected compounds structurally related to acyclic sesquiterpenoids and their antibacterial and cytotoxic activity, Molecules. 20 (2015) 11272-11296.

[132] S. Andrade-Ochoa, G.V. Nevárez-Moorillón, L.E. Sánchez-Torres, M. Villanueva-García, B.E. Sánchez-Ramírez, L.M. Rodríguez-Valdez, B.E. Rivera-Chavira, Quantitative structure-activity relationship of molecules constituent of different essential oils with antimycobacterial activity against Mycobacterium tuberculosis and Mycobacterium bovis, BMC Complement Altern Med. 15 (2015) 332.

Chapter 3.

[1] H. Wiener, Structural determination of paraffin boiling points, J Am Chem Soc. 69 (1947) 17-20.

[2] F. Harary, Graph theory. Addison-Wesley, Reading MA, 1969.

[3] L.B. Kier, L.H. Hall, Molecular connectivity in structure activity analysis, Research studies press Letchworth, Hertfordshire, U. K., 1986.

[4] M. Randić, Characterization of molecular branching, J Am Chem Soc. 97 (1975) 6609-6615.

[5] S.C. Basak, Use of molecular complexity indices in predictive pharmacology and toxicology: a QSAR approach, Med Sci Res. 15 (1987) 605-609.

[6] C.E. Shannon, A mathematical theory of communication, Bell Syst Tech J. 27 (1948) 379-423.

[7] S.C. Basak, S. Bertelsen, G.D. Grunwald, Application of graph theoretical parameters in quantifying molecular similarity and structure-activity studies, J Chem Inf Comput Sci. 34 (1994) 270-276.

[8] S.C. Basak, G.D. Grunwald, Predicting mutagenicity of chemicals using topological and quantum chemical parameters: a similarity based study, Chemosphere. 31 (1995) 2529-2546.

[9] R. Todeschini, V. Consonni, Handbook of molecular descriptors, Wiley, New York, 2008.

[10] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J Mol Biol. 55 (1971) 379-400.

[11] M.L. Connolly, Analytical molecular surface calculation, J Appl Cryst.16 (1983) 548-558.

[12] M.F. Sanner, Ph.D. dissertation thesis: Modeling and applications of molecular surfaces, Université de Haute-Alsace, France, 1992.

[13] M.F. Sanner, A.J. Olson, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers. 38 (1996) 305-320.

[14] J. Kujawski, H. Popielarska, A. Myka, B. Drabińska, M.K. Bernard, The log P parameter as a molecular descriptor in the computer-aided drug design – an overview, CMST. 18 (2012) 81-88.

[15] D. Eros, I. Kovesdi, L. Orfi, K. Takacs-Novak, G. Acsády, G. Kéri, Reliability of logP predictions based on calculated molecular descriptors: a critical review, Curr Med Chem. 9 (2002) 1819-1829.

[16] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J Am Chem Soc. 107 (1985) 3902-3909.

[17] J.J.P. Stewart, Optimization of parameters for semiempirical methods.1. method, J Comput Chem. 10 (1989) 209-220.

[18] J.J.P. Stewart, Comments on a comparison of AM1 with the recently developed PM3 method- reply, J Comput Chem. 11 (1990) 543-544.

[19] J.J.P. Stewart, Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi, J Comput Chem. 12 (1991) 320-341.

[20] J.B. Foresman, A. Frisch, Exploring chemistry with electronic structure methods, Gaussian Inc., Pittsburgh, 1996.

[21] W. Koch, M.C. Holthausen, A chemist's guide to density functional theory, Wiley-VCH, Weinheim, 2000.

[22] C. Gruber, V. Buss, Quantum-mechanically calculated properties for the development of quantitative structure-activity relationships (QSAR'S). pKa-values of phenols and aromatic and aliphatic carboxylic acids, Chemosphere. 19 (1989) 1595-1609.

[23] T. Sotomatsu, Y. Murata, T.J. Fujita, Correlation analysis of substituent effects on the acidity of benzoic acids by the AM1 method, J Comput Chem. 10 (1989) 94-98.

[24] M. Karelson, V.S. Lobanov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, Chem Rev. 96 (1996) 1027-1043.

[25] D.F. Lewis, C. Ioannides, D.V. Parke, Interaction of a series of nitriles with the alcohol-inducible isoform of P450: computer analysis of structure-activity relationships, Xenobiotica. 24 (1994) 401-408.

[26] R.G. Parr, R.A. Donnelly, M. Levy, W.E. Palke, Electronegativity: the density functional viewpoint, J Chem Phys. 68 (1978) 3801-3807.

[27] R.P. Iczkowski, J.L. Margrave, Electronegativity, J Am Chem Soc. 83 (1961) 3547-3551.

[28] I.N Levine, Quantum chemistry, Pearson Education, Singapore, 2000.

[29] I.N Levine, Physical chemistry, Mc-Graw Hill, New York, 2011.

[30] B.J. McConkey, V. Sobolev, M. Edelman, The performance of current methods in ligand-protein docking, Curr Sci. 83 (2002) 845-855.

[31] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J Comput Chem. 19 (1998) 1639-1662.

[32] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R,D. Taylor, Improved protein-ligand docking using GOLD, Proteins. 52 (2003) 609-623.

[33] T.J. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, J Comput Aided Mol Des. 15 (2001) 411-428.

[34] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide:

a new approach for rapid, accurate docking and scoring. 1 Method and assessment of docking accuracy, J Med Chem. 47 (2004) 1739-1749.

[35] R. Abagyan, M. Totrov, D. Kuznetsov, ICM-A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J Comput Chem. 15 (1994) 488-506.

[36] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J Mol Biol. 261 (1996) 470-489.

[37] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J Am Chem Soc. 106 (1984) 765-784.

[38] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, J Comput Chem. 28 (2007) 1145-1152.

[39] T.W. Anderson, An introduction to multivariate analysis, John Wiley, New York, 1958.

[40] A.M. Kshirsagar, Multivariate analysis, Marcel Dekker, New York, 1972.

[41] C.R. Rao, Advanced statistical inference and its applications, Wiley Eastern, New Delhi, 1973.

Chapter 4.

[1] M. Bollen, S. Keppens, W. Stalmans, Specific features of glycogen metabolism in the liver, Biochem J. 336 (1998) 19-31.

[2] H.G. Hers, The control of glycogen metabolism in the liver, Annu Rev Biochem. 45 (1976) 167-190.

[3] W. Stalmans, The interaction of liver phosphorylase *a* with glucose and AMP, Eur J Biochem. 49 (1974) 415-427.

[4] W. Pimenta, N. Nurjhan, P.A. Jansson, M. Stumvoll, J. Gerich, M. Korytkowski, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia. 37 (1994) 697-702.

[5] M.K. Hellerstein, R.A. Neese, P. Linfoot, M. Christiannsen, S. Turner, A. Letscher, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest. 100 (1997) 1305-1319.

[6] N.G. Oikonomakos, E.D. Chrysina, M N. Kosmopoulou, D.D. Leonidas, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 A resolution, Biochim Biophys Acta. 1647 (2003) 325-332.

[7] V.L. Rath, M. Ammirati, D.E. Danley, J.L. Ekstrom, E.M. Gibbs, T.R. Hynes, A.M. Mathiowetz, R.K. Mcpherson, T.V. Olson, J.L. Treadway, D.J. Hoover, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol. 7 (2000) 677-682.

[8] R. Kurukulasuriya, J.T. Link, D.J. Madar, Z. Pei, S.J. Richards, J.J. Rohde, A.J. Souers, B.G. Szczepankiewicz, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem. 10 (2003) 123-153.

[9] Z. Liang, L. Zhang, L. Li, J. Liu, H. Li, L. Zhang, L. Chen, K. Cheng, M. Zheng, X. Wen, P. Zhang, J. Hao, Y. Gong, X. Zhang, X. Zhu, J. Chen, H. Liu, H. Jiang, C. Luo, H. Sun,

Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem. 46 (2011) 2011-2021.

[10] X. Wen, H. Sun, J. Liu, K. Cheng, P. Zhang, L. Zhang, J. Hao, L. Zhang, P. Ni, S.E. Zographos, D.D. Leonidas, K.M. Alexacou, T. Gimisis, J.M. Hayes, N.G. Oikonomakos, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem. 51 (2008) 3540-3554.

[11] P. Zhu, Y. Bi, J. Xu, Z. Li, J. Liu, L. Zhang, W. Ye, X. Wu, Terpenoids. III: Synthesis and biological evaluation of 23-hydroxybetulinic acid derivatives as novel inhibitors of glycogen phosphorylase, Bioorg Med Chem Lett. 19 (2009) 6966–6969.

[12] P. Lan, J. Wang, D.M. Zhang, C. Shu, H.H. Cao, P.H. Sun, X.M. Wu, W.C. Ye, W.M. Chen, Synthesis and antiproliferative evaluation of 23-hydroxybetulinic acid derivatives, Eur J Med Chem. 46 (2011) 2490-2502.

[13] Y. Bi, J. Xu, F. Sun, X. Wu, W. Ye, Y. Sun, W. Huang, Synthesis and biological activity of 23-hydroxybetulinic acid C-28 ester derivatives as antitumor agent candidates, Molecules. 17 (2012) 8832-8841.

[14] W. Ding, M. Sun, S. Luo, T. Xu, Y. Cao, X. Yan, Y. Wang, A 3D QSAR study of betulinic acid derivatives as anti-tumor agents using topomer CoMFA: model building studies and experimental verification, Molecules. 18 (2013) 10228-10241.

[15] S. Fulda, K.M. Debatin, Betulinic acid induces apoptosis through a direct effect on mitochondria in neuroectodermal tumors, Med Pediatric Oncol. 35 (2000) 616–618.

[16] Z.N. Ji, W.C. Ye, G.G. Liu, W.L. Wendy Hsiao, 23-Hydroxybetulinic acid-mediated apoptosis is accompanied by decreases in Bcl-2 expression and telomerase activity in HL-60 Cells, Life Sci. 72 (2002) 1-9.

[17] M. Mandal, R. Kumar, Bcl-2 Modulates telomerase activity, J Biol Chem. 272 (1997) 14183-14187.

[18] Y Zheng, F. Zhou, X. Wu, X. Wen, Y. Li, B. Yan, J. Zhang, G. Hao, W. Ye, G. Wang, 23-Hydroxybetulinic acid from Pulsatilla chinensis (Bunge) Regel synergizes the antitumor activities of doxorubicin in vitro and in vivo, J Ethnopharmacol. 128 (2010) 615–622.

[19] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[20] E. Eroglu, Some QSAR Studies for a Group of Sulfonamide Schiff Base as Carbonic Anhydrase CA II Inhibitors, Int J Mol Sci. 9 (2008) 181-197.

[21] L.B. Kier, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci. 69 (1980) 807-810.

[22] S.C. Basak, D.K. Harriss, V.R. Magnuson, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci. 73 (1984) 429-437.

[23] H. Wiener, Structural determination of paraffin boiling points, J Am Chem Soc. 69 (1947) 17-20.

[24] D. Playsic, S. Nikolic, N. Trinajstic, Z. Mihalic, On the Harary index for the characterization of chemical graphs, J Math Chem. 12 (1993) 235-250.

[25] M. Randiac, On the characterization of molecular branching, J Am Chem Soc. 97 (1975) 6609-6615.

[26] L.B. Kier, L.H. Hall, Molecular connectivity in structure-activity analysis, Research studies press: Letchworth, Hertfordshire, U K, 1986.

[27] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, GAMESS Version= 24 Mar 2007 (R1) from Iowa State University, J Comput Chem. 14 (1993) 1347-1363.

[28] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J Comput Chem. 30 (2009) 2785-2791.

[29] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, J Comput Chem. 28 (2007) 1145-1152.

[30] R. Huey, D.S. Goodsell, G.M. Morriss, A.J. Olson, Grid-based hydrogen bond potentials with improved directionality, Lett Drug Des Discov. 1 (2004) 178-183.

[31] R. Thomsen, M.H. Christensen, MolDock: a new technique for high-accuracy molecular docking, J Med Chem. 49 (2006) 3315-3321.

Chapter 5.

[1] E.I. Solomon, J. Zhou, F. Neese, E.G. Pavel, New insights from spectroscopy into the structure/function relationships of lipoxygenases, Chem Biol. 4 (1997) 795-808.

[2] H.W. Gardner, Biological roles and biochemistry of the lipoxygenase pathway, HortScience. 30 (1995) 197-205.

[3] R. Wisastra, F.J. Dekker, Inflammation, cancer and oxidative lipoxygenase activity are intimately linked, Cancers. 6 (2014) 1500-1521.

[4] L.A. Dailey, P. Imming, 12-Lipoxygenase: classification, possible therapeutic benefits from inhibition, and inhibitors, Curr Med Chem. 6 (1999) 389-398.

[5] V.E. Steele, C.A. Holmes, E.T. Hawk, L. Kopelovich, R.A. Lubet, J.A. Crowell, C.C. Sigman, G.J. Kelloff, Lipoxygenase inhibitors as potential cancer chemopreventives, Cancer Epidemiol Biomarkers Prev. 8 (1999) 467-483.

[6] B. Samuelsson, S.E. Dahlen, J.A. Lindgren, C.A. Rouzer, C.N. Serhan, Leukotrienes and lipoxins: structures, biosynthesis, and biological effects, Science. 237 (1987) 1171-1176.

[7] X.Z. Ding, W.G. Tong, T.E. Adrian, 12-Lipoxygenase metabolite 12(S)-HETE stimulates human pancreatic cancer cell proliferation via protein tyrosine phosphorylation and ERK activation, Int J Cancer. 94 (2001) 630-636.

[8] J.A. Cornicelli, B.K. Trivedi, 15-Lipoxygenase and its inhibition: a novel therapeutic target for vascular disease, Curr Pharm Des. 5 (1999) 11-20.

[9] U.P. Kelavkar, J.B. Nixon, C. Cohen, D. Dillehay, T.E. Eling, K.F. Badr, Overexpression of 15-lipoxygenase-1 in PC-3 human prostate cancer cells increases tumorigenesis, Carcinogenesis. 22 (2001) 1765-1773.

[10] C. Kemal, P. Louis-Flamberg, R. Krupinski-Olsen, A.L. Shorter, Reductive inactivation soybean lipoxygenase-1 by catechols: a possible mechanism for regulation of lipoxygenase activity, Biochemistry. 26 (1987) 7064-7072.

[11] R. Mogul, E. Johansen, T.R. Holman, Oleyl sulfate reveals allosteric inhibition of soybean lipoxygenase-1 and human 15-lipoxygenase, Biochemistry. 39 (2000) 4801-4807.

[12] T. Amagata, S. Whitman, T.A. Johnson, C.C. Stessman, C.P. Loo, E.L. Jonclardy, P. Crews, T.R. Holman, Exploring sponge-derived terpenoids for their potency and selectivity against 12-human, 15-human, and 15-soybean lipoxygenases, J Nat Prod. 66 (2003) 230-235

[13] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[14] Y.H. Jang, W.A. Goddard III, K.T. Noyes, L.C. Sowers, S. Hwang, D.S. Chung, p*K*a Values of guanine in water: density functional theory calculations combined with poisson-boltzmann continuum-solvation model, J Phys Chem B. 107 (2003) 344-357.

[15] J.E. Bartmess, Thermodynamics of the electron and the proton, J Phys Chem. 98 (1994) 6420-6424.

[16] M.D. Liptak, K.G. Gross, P.G. Seybold, S. Feldgus, G.C. Shields, Absolute pK(a) determinations for substituted phenols, J Am Chem Soc. 124 (2002) 6421-6427.

[17] P. Winget, E.J. Weber, C.J. Cramer, D.G. Truhlar, Computational electrochemistry: aqueous one-electron oxidation potentials for substituted anilines, Phys Chem Chem Phys. 2 (2000) 1231-1239.

[18] N.S. Babu, Computational studies for oxidation reduction reactions of cinnoline - 4(1H)-one, in aqueous phase by density functional theory, BJAST. 4(2014) 465-476.

[19] M.J. Nelson, D.G. Batt, J.S. Thompson, S.W. Wright, Reduction of the active-site iron by potent inhibitors of lipoxygenases, J Biol Chem. 266 (1991) 8225-8229.

Chapter 6.

[1] H. Harada, Y. Kamei, Dose-dependent selective cytotoxicity of extracts from marine green alga, Cladophoropsis vaucheriaeformis, against mouse leukemia L1210 cells, Biol Pharm Bull. 21 (1998) 386-389.

[2] K.E. Apt, P.W. Behrens, Commercial developments in microalgal biotechnology, J Phycol. 35 (1999) 215-26.

[3] K. Harada, M. Suomalainen, H. Uchida, H. Masui, K. Ohmura, J. Kiviranta, M.L. Niku-Paavola, T. Ikemoto, Insecticidal compounds against mosquito larvae from Oscillatoria agardhii strain 27, Environ Toxicol. 15 (2000) 114-119.

[4] V.H. Argandoña, J. Rovirosa, A. San-Martín, A. Riquelme, A.R. Díaz-Marrero, M. Cueto, J. Darias, O. Santana, A. Guadaño, A. González-Coloma, Antifeedant effects of marine halogenated monoterpenes, J Agric Food Chem. 50 (2002) 7029-7033.

[5] C.D. Inés, V.H. Argandoña, J. Rovirosa, A. San-Martín, A.R. Díaz-Marrero, M. Cueto, A.G. Coloma, Cytotoxic activity of halogenated monoterpenes from Plocamium cartilagineum, Z Naturforsch. 59c (2004) 339-344.

[6] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[7] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

Chapter 7.

[1] D. Barford, A.J. Flint, N.K. Tonks, Crystal structure of human protein tyrosine phosphatase 1B, Science. 263 (1994) 1397-1404.

[2] Z.Y. Zhang, Structure, mechanism, and specificity of protein-tyrosine phosphatases, Curr Top Cell Regul. 35 (1997) 21-68.

[3] D. Barford, A.K. Das, M.P. Egloff, The structure and mechanism of protein phosphatases: insights into catalysis and regulation, Annu Rev Biophys Biomol Struct. 27 (1998) 133-164.

[4] Z.Y. Zhang, Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis, Crit Rev Biochem Mol Biol. 33 (1998) 1-52.

[5] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, Protein tyrosine phosphatases in the human genome, Cell. 117 (2004) 699-711.

[6] K.A. Kenner, E. Anyanwu, J.M. Olefsky, J. Kusari, Protein-tyrosine phosphatase 1B is a negative regulator of insulin- and insulin-like growth factor-I-stimulated signaling, J Biol Chem. 271 (1996) 19810-19816.

[7] J.R. Wiener, B.J. Kerns, E.L. Harvey, M.R. Conaway, J.D. Iglehart, A. Berchuck, R.C. Bast Jr, Overexpression of the protein tyrosine phosphatase PTP1B in human breast cancer: association with p185c-erbB-2 protein expression, J Natl Cancer Inst. 86 (1994) 372-378.

[8] M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A.L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.C. Chan, C. Ramachandran, M.J. Gresser, M.L. Tremblay, B.P. Kennedy, Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene, Science. 283 (1999) 1544-1548.

[9] J. Liu, Pharmacology of oleanolic acid and ursolic acid, J Ethnopharmacol. 49 (1995) 57-68.

[10] J. Liu, Oleanolic acid and ursolic acid: research perspective, J Ethnopharmacol. 100 (2005) 92-94.

[11] P. Dzubak, M. Hajduch, D. Vydra, A. Hustova, M. Kyasnica, D. Biedermann, L. Markova, M. Urban, J. Sarek, Pharmacological activities of natural triterpenoids and their therapeutic implications, Nat Prod Rep. 23 (2006) 394-411.

[12] R. Martín, J. Carvalho-Tavares, M. Hernández, M. Arnés, V. Ruiz-Gutiérrez, M.L. Nieto, Beneficial actions of oleanolic acid in an experimental model of multiple sclerosis: a potential therapeutic role, Biochem Pharmacol. 79 (2010) 198-208.

[13] J.J. Ramírez-Espinosa, M.Y. Rios, S.L. Martínez, F.L. Vallejo, J.L. Medina-Franco, P. Paoli, G. Camici, G. Navarrete-Vázquez, R. Ortiz-Andrade, S. Estrada-Soto, Antidiabetic activity of some pentacyclic acid triterpenoids, role of PTP-1B: in vitro, in silico, and in vivo approaches, Eur J Med Chem. 46 (2011) 2243-2251.

[14] Y.A. Puius, Y.U. Zhao, M. Sullivan, D.S. Lawrence, S.C. Almo, Z.Y. Zhang, Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design, Proc Natl Acad Sci USA. 94 (1997) 13420-13425.

[15] J.M. Castellano, A. Guinda, T. Delgado, M. Rada, J.A. Cayuela, Biochemical basis of the antidiabetic activity of oleanolic acid and related pentacyclic triterpenes, Diabetes. 62 (2013) 1791-1799.

[16] Y.N. Zhang, W. Zhang, D. Hong, L. Shi, Q. Shen, J.Y. Li, J. Li, L.H. Hu, Oleanolic acid and its derivatives: new inhibitor of protein tyrosine phosphatase 1B with cellular activities, Bioorg Med Chem. 16 (2008) 8697-8705.

[17] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[18] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J Comput Chem. 19 (1998) 1639-1662.

[19] Molegro molecular viewer – version 2.5.0. http://www.molegro.com/index.php.

[20] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[21] M.F. Sanner, Python: a programming language for software integration and development, J Mol Graphics Mod. 17 (1999) 57-61.

Chapter 8.

[1] R.W. Fuller, J.H. Cardellina II, Y. Kato, L.S. Brinen, J. Clardy, K.M. Snader, M.R. Boyd, A pentahalogenated monoterpine from the red alga Portieria hornemannii produces a novel cytotoxicity profile against a diverse panel of human tumor cell lines, J Med Chem. 35 (1992) 3007-3011.

[2] R.W. Fuller, J.H. Cardellina II, J. Jurek, P.J. Scheuer, B.A. Lindner, M. Mcguire, G.N. Gray, J.R. Steiner, J. Clardy, E. Menez, R.H. Shoemaker, D.J. Newman, K.M. Snader, M.R. Boyd, Isolation and structure/activity features of halomon - related antitumor monoterpines from the red alga Portieria hornemannii, J Med Chem. 37 (1994) 4407-4411.

[3] M.J. Egorin, D.L. Sentz, D.M. Rosen, M.F. Ballesteros, C.M. Kearns, P.S. Callery, J.L. Eiseman, Plasma pharmacokinetics, bioavailability, and tissue distribution in CD_2F_1 mice of halomon, an antitumor halogenated monoterpene isolated from the red alga Portieria hornemannii, Cancer Chemother Pharmacol. 39 (1996) 51-60.

[4] M.J. Egorin, M.D. Rosen, S.E. Benjamin, P.S. Callery, D.L. Sentz, J.L. Eiseman, In vitro metabolism of mouse and human liver preparations of halomon, an antitumor halogenated monoterpene, Cancer Chemother Pharmacol. 41 (1997) 9-14.

[5] M.E. Jung, M.H. Parker, Synthesis of several naturally occurring polyhalogenated monoterpene of the halomon class, J Org Chem. 62 (1997) 7094-7095.

[6] T. Sotokawa, T. Noda, S. Pi, M.A. Hirama, A three step synthesis of halomon, Angew Chem Int Ed Engl. 39 (2000) 3430-3432.

[7] H.D. Yoo, S.O. Ketchum, D. France, K. Bair, W.H. Gerwick, Vidalenolone , a novel phenolic metabolite from the tropical red alga Vidalia sp., J Nat Prod. 65 (2002) 51-53.

[8] E.H. Andrianasolo, D. France, S.C. Kennon, W.H. Gerwick, DNA Methyl Transferase Inhibiting Halogenated Monoterpenes from the Madagascar Red Marine Alga Portieria hornemannii, J Nat Prod. 69 (2006) 576-579.

[9] J.J.P. Stewart, MOPAC2016, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016).

[10] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[11] A.R. Allouche, Gabedit-a graphical user interface for computational chemistry softwares, J Comput Chem. 32 (2011) 174-182.

Chapter 9.

[1] M.R. Orofino, O. Kreuger, S. Grootjans, M.W. Biavatti, P. Vandenabeele, K. D'Herde, Sesquiterpene lactones as drugs with multiple targets in cancer treatment: focus on parthenolide, Anticancer Drugs. 23 (2012) 883-896.

[2] Y.M. Zhao, M.L. Zhang, Q.W. Shi, H. Kiyota, Chemical constituents of plants from the genus Inula, Chem Biodivers. 3 (2006) 371-384.

[3] X. Cheng, Q. Zeng, J. Ren, J. Qin, S. Zhang, Y. Shen, J. Zhu, F. Zhang, R. Chang, Y. Zhu, W. Zhang, H. Jin, Sesquiterpene lactones from Inula falconeri, a plant endemic to the Himalayas, as potential anti-inflammatory agents, Eur J Med Chem. 46 (2011) 5408-5415.

[4] ACD/ChemSketch version 12.01, Advanced Chemistry Development, Inc., Toronto, Ontario, 2009.

[5] B. Bagchi, A. Chatterjee, P. Ghosh, A.K. Bothra, A theoretical investigation of cytotoxic activity of halogenated monoterpenoids from plocamium cartilagineum, JOCPR. 4 (2012) 5076-5080.

[6] A.A. Granovsky, Firefly version 8, www http://classic.chem.msu.su/gran/firefly/index.html.

[7] E. Eroglu, Some QSAR studies for a group of sulfonamide schiff base as carbonic anhydrase CA II inhibitors, Int J mol Sci. 9 (2008) 181-197.

[8] L.B. Kier, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci. 69 (1980) 807-810.

[9] S.C. Basak, D.K. Harriss, V.R. Magnuson, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci. 73 (1984) 429-437.

[10] J.J.P. Stewart, MOPAC2012, Stewart Computational Chemistry, Colorado Springs, CO, USA.

[11] E. Eroğlu, H. Türkmen, S. Güler, S. Palaz, O. Oltulu, A DFT-based QSARs study of acetazolamide/sulfanilamide derivatives with carbonic anhydrase (CA-II) isozyme inhibitory activity, Int J Mol Sci. 8 (2007) 145-155.

[12] S. Sharma, B. Bagchi, S. Mukhopadhyay, A.K. Bothra, Theoretical study of lysophosphatidic acid acyltransferase 2 inhibitors, JOCPR. 5 (2013) 348-355.

[13] B. Das, P. Ghosh, A.K. Bothra, QSAR studies of anthrax lethal inhibitors through quantum chemical indices, JOCPR. 3 (2011) 443-449.

INDEX

Acetylcoenzyme A 5

Abscisic acid 9

Artemisinin 12

Arglabin 13

Apoptosis 112

Biostatistic 97

Co-evolution 2

Citral 8-9

Complementary information content 86-88

Correlation 105

Cross validate coefficient 105-106

Dimethylallyl diphosphate 6

Dipole moment 94

Electronegativity 94

Farnesyl diphosphate 6

Free and Wilson model 18

F test 106

Geranyl diphosphate 6

Geranyl farnesyl diphosphate 7

Geranyl geranyl diphosphate 6

Gibberellic acid 10,11

Hammett equation 17

Hansch equation 17-18

HOMO 21

Harary index 85

HPLC 91

Isoprene rule 3-4

Isopentenyl diphosphate 5

Ingenol mebutate 13

Limonene 9

Lycopene 12

LUMO 21

Linear relation 97

Mevalonate 5

Mean information content 86-88

Molar refraction 89

Molar Volume 90

Molar entropy 94-95

Molecular docking 95-97

Neral 8

Nanoxel 13

Optimization 152-156

Perillyl alcohol 9,10

Paclitaxel 13

Partition coefficient 91

Phenogram 136-137

Quality factor 106

Retinol 10,11

Randic connectivity index 86

Regression analysis 97-105

Secondary metabolites 3

Structural information content 86-88

Solvent Accessible Surface Area 90

γ-Terpinene 9

Telomerase 112

Unsaturated fatty acid 15,16

Volume 90

Wiener Index 85

X ray 197

Zero order connectivity index 86

Commun. Comput. Chem. Vol.3, No.3, pp.75-102
doi: 10.4208/cicc.2015.v3.n3.2 December 2015

REGULAR ARTICLE

QSAR Study and Molecular Docking of 23-hydroxybetulinic Acid Derivatives as RMGPa and HeLa Cells Inhibitors

Bhaskar Bagchi[1], Shyamal Sharma[1], Abhik Chatterjee[1], Pranab Ghosh[2], and Asim Kumar Bothra[1*]

[1] *Cheminformatics Bioinformatics Lab,Department of Chemistry, Raiganj College (University College) P.O.-Raiganj, Dist. - Uttar Dinajpur, PIN-733134, INDIA*

[2] *Department of Chemistry University of North Bengal*

Received 22 Aug 2015; Accepted (in revised version) 20 Nov 2015

Abstract: A number of 23-hydroxybetulinic acid derivatives were found to be potent glycogen phosphorylase a (GPa) inhibitors. Some derivatives of these triterpenes exhibit anti-tumor activities against a variety of tumor cell lines. In this study, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa. The other set of equations predicts the antiproliferative activities against HeLa cells. This QSAR study has shown that topological indices and quantum chemical descriptors are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. We have also performed Docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa and it has been found that the important interacting amino acids present in the active site cavity are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the compounds can form hydrogen bonds with ARG193 and/or ARG310. The unfavorable steric clashes between ligand and the protein and decrease of number of hydrogen bonds lower the inhibitory activity of the ligand.

* Corresponding author. *Email address:* asimbothra@gmail.com Telephone No. +919474441570
Fax No.-+913523242580

AMS subject classifications: 62J05, 92E10, 93A30
Key words: QSAR, Docking, RMGPa, HeLa, 23-hydroxybetulinic acid

Introduction

The breakdown of glycogen into glucose is mediated by glycogen phosphorylase (GP) with the help of a debranching enzyme and plays an important role for controlling hepatic glucose production [1]. GP has three isoforms which are brain, liver, and muscle according to their expression patterns. The muscle isoform supplies energy for muscle contraction. The brain isoform provides glucose during the periods of severe hypoglycemia. In glycogenolysis, liver enzyme plays a rate limiting role and hence it is an attractive target for the treatment of type 2 diabetes [2-5]. GP exists in two interconvertible forms: the phosphorylated high activity glycogen phosphorylase a (GPa) and the dephosphorylated low activity glycogen phosphorylase b (GPb). Allosteric effectors can promote equilibrium between a less active GPb and an active GPa. The active conformation is stabilized by phosphorylation of Ser 14 and binding of AMP [6, 7]. GP contains at least six regulatory sites: glucose analogues at the catalytic site, azasugar inhibitors, lactones at the allosteric site (AMP), caffeine at the purine inhibitor site, indole-2-carboxamide at the indole binding site and cyclodextrins at the glycogen storage site [8,9]. The X-ray analysis indicates that pentacyclic triterpenes bind at the allosteric site [10].

A number of 23-hydroxybetulinic acid derivatives were reported as potent GPa inhibitors [11]. Furthermore, some derivatives of these triterpenes display anti-tumor activities against a variety of tumor cell lines and the mechanism of action may be associated to its effects on the proliferation, migration, cell cycle and apoptosis of tumor cells [12, 13].

Apoptosis or programmed cell death plays an important role in regulating development and homeostasis. In many diseases including cancer, apoptosis is suppressed. Telomerase, a ribonucleoprotein, maintains chromosome lengths by adding telomeres to the chromosome ends repeatedly. Telomerase activation is found in about 90% of tumor tissues, but with very low, almost undetectable activity in somatic cells. Apoptosis is regulated by a number of cellular genes including B cell leukemia/lymphoma 2 (bcl-2). The stable expression of bcl-2 in human cancer cells increases telomerase activity and resistance to apoptosis. The 23-hydroxybetulinic acid induces apoptosis through concurrent inhibition of bcl-2 expression and telomerase activity [14- 17].

Furthermore, 23-hydroxybetulinic acid enhances anti-tumor activity of doxorubicin in vitro and in vivo. The synergism is associated with increase in the doxorubicin concentration in tumor tissue brought about by 23-hydroxybetulinic acid. Hence 23-hydroxybetulinic acid has the prospect to be developed as a novel chemosensitizer [18].

In this communication, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa [11]. The other set of equations predicts the antiproliferative activities against HeLa cells. We have also performed Docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa.

Materials and methods

The development of QSAR models are described in the following section as: dataset preparation, parameters and statistical methods.

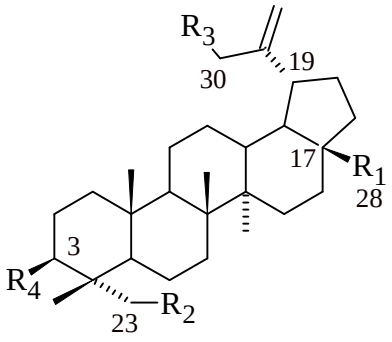
Dataset and parameters

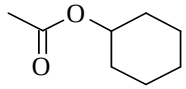
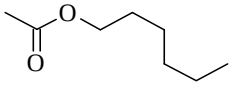
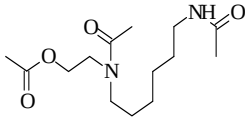
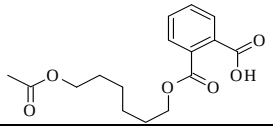
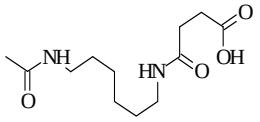
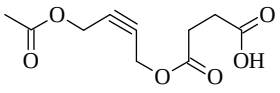
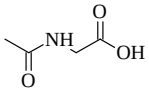
The dataset which is selected from the literature [11, 12] contains 47 compounds of 23-hydroxybetulinic acid derivatives. The biological property of this data set is reported as IC_{50} values. The IC_{50} values were converted into log IC_{50} and taken as the response variable for QSAR modeling. Structural details of the 47 compounds and their biological activity are listed in **Table 1**.

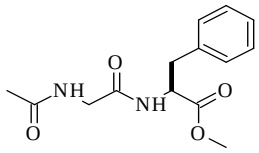
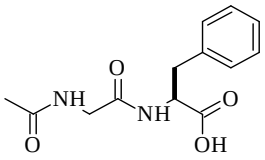
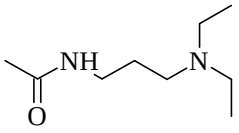
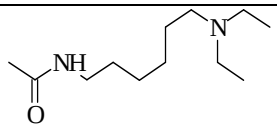
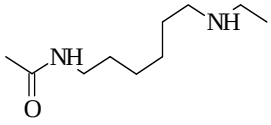
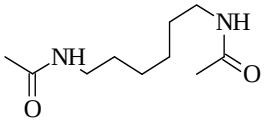
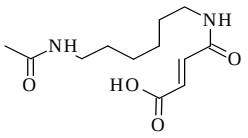
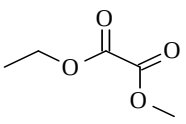
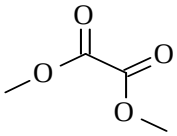
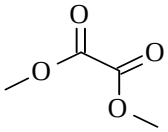
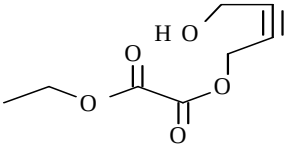
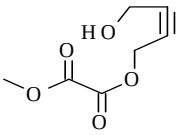
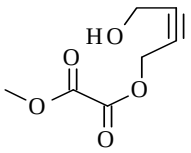
The initial structures of 47 compounds used in this study were constructed by ChemSketch [19]. We attempted several descriptors (data not shown) and it has been found that quantum chemical descriptors (EH, EL, μ), molar refractivity (MR), molar volume (MV) and topological indices such as Structural Information Content (SIC), Complementary Information Content (CIC), Wiener index (W), Harary index (H), Randiac's connectivity index of first order ($\chi 0$) [20-26] can better represent the biological activity of compounds.

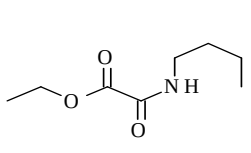
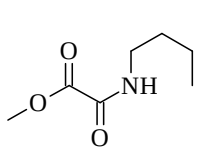
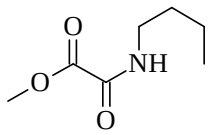
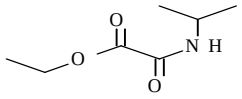
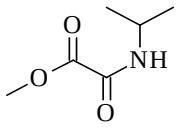
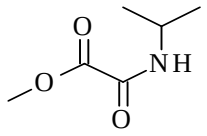
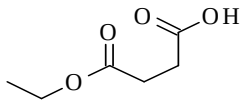
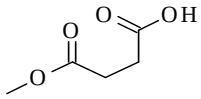
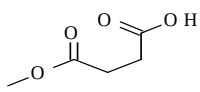
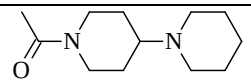
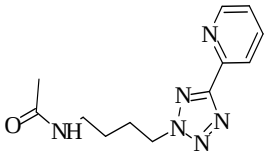
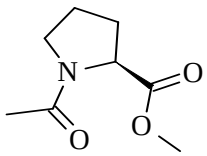
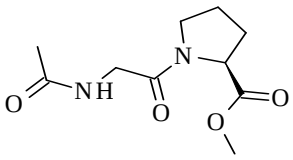
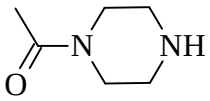
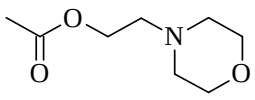
The quantum chemical properties (EH, EL, μ) of the studied molecules have been determined by DFT/B3LYP calculation and the basis set 6-31G* was used. All quantum chemical calculations were performed by Gamess [27]. Molar refractivity (MR) and molar volume (MV) were determined using ChemSketch software. The graph theoretical descriptors such as SIC, CIC, W, H, $\chi 0$ were computed using program written by us in Fortran-77.

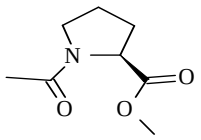
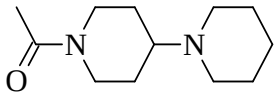
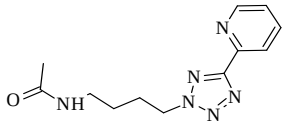
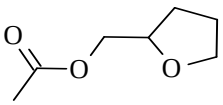
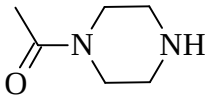
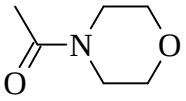
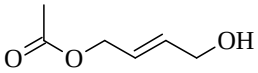
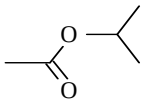
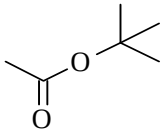
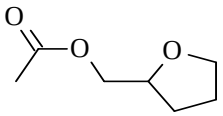
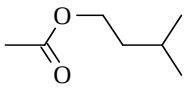
Table 1 Structures and activity of 47 compounds of 23-hydroxybetulinic acid derivatives

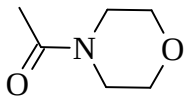
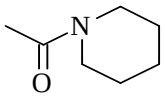
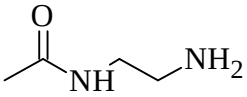
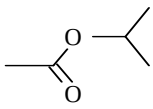
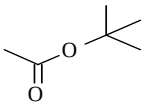
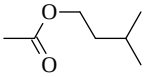
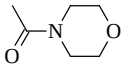
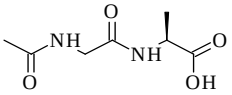
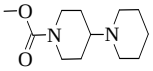
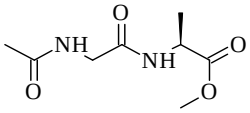


Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
1	-COOH	-OH	-H	-OH	103	46.22
2		-OH	-H	-OH	50.4	-
3		-OH	-H	-OH	69.1	-
4		-OAc	-H	-OAc	96.6	-
5		-OAc	-H	-OAc	69.2	-
6		-OAc	-H	-OAc	194	-
7		-OAc	-H	-OAc	94.5	-
8		-OH	-H	-OH	67.7	-

9		-OH	-H	-OH	40.6	-
10		-OH	-H	-OH	129	-
11		-OAc	-H	-OAc	72.8	-
12		-OAc	-H	-OAc	55.6	-
13		-OAc	-H	-OAc	29.5	-
14		-OAc	-H	-OAc	34.1	-
15		-OAc	-H	-OAc	54.8	-
16			-H		30.1	-
17			-H		19.5	-

18			-H		34.7	-
19			-H		35.9	-
20			-H		97.8	-
21		-OAc	-H	-OH	-	10.80
22		-OAc	-H	-OH	-	17.87
23		-OAc	-OH	-OH	-	13.08
24		-OH	-H	-OH	-	4.84
25		-OAc	-H	-OH	-	18.84
26		-OAc	-H	-OH	-	9.18

27		-OAc	-H	-OH	-	9.78
28		-OH	-H	-OH	-	8.42
29		-OH	-H	-OH	-	46.26
30		-OH	-OH	-OH	-	79.35
31		-OH	-H	-OH	-	7.12
32		-OH	-H	-OH	-	34.29
33		-OH	-H	-OH	-	44.44
34		-OAc	-OH	-OH	-	19.46
35		-OAc	-OH	-OH	-	23.28
36		-OAc	-OH	-OH	-	17.79
37		-OAc	-OH	-OH	-	20.84

38		-OAc	-OH	-OH	-	18.31
39		-OAc	-OH	-OH	-	11.02
40		-OAc	-H	-OH	-	7.39
41		-OH	-OH	-OH	-	22.27
42		-OH	-OH	-OH	-	30.65
43		-OH	-OH	-OH	-	20.85
44		-OAc	-H	-OH	-	7.47
45		-OH	-H	-OH	-	22.10
46	-COOH		-H	-OH	-	12.52
47		-OH	-H	-OH	-	8.27

Statistical methods

Multiple linear regression (MLR) analysis was used to build up QSAR models. Different combinations of parameters were used to develop these models. Statistical qualities of MLR equations were judged by parameters like correlation coefficient (R), square of the correlation coefficient (R^2), cross validated coefficient (R^2_{cv}), standard deviation of the regression (S), Fischer statistics (F) and quality factor (Q). MLR program written by ourselves in Fortran-77 is used.

Molecular docking

The coordinates of RMGPa in complex with CHI (1LWO.pdb) [6] were obtained from the RCSB protein data bank (www.rcsb.org). The 23-hydroxy betulinic acid derivatives were docked into the active pocket of the enzyme by using docking program Autodock 4.0 [28-30]. Initially the structures of the ligands have been optimized with AM1 method and the hydrogen atoms were added to the enzyme. The Lamarckian genetic algorithm (LGA) was applied to look out for the best conformers. A grid map with 80x80x80 points and 0.375 Å spacing was used in Autogrid program to evaluate the binding energies between the inhibitors and RMGPa. The grid centre was set at the active site position 28.672, 0.621 and 53.033 and the default settings were used. For each compound ten docking poses were saved and ranked by binging energy. The lowest energy docking pose of 23-hydroxybetulinic acid was nearly identical to the Asiatic acid and Maslinic acid orientation in crystal structure with GPb [10]. So the pose with highest negative binding energy was selected for analyzing the type of interactions. The binding site was analyzed with molegro molecular viewer software [31].

Results and discussion

The data set of 47 compounds was divided into two groups. The first group of molecules contains 20 compounds having inhibitory activity of RMGPa and the second group of molecules contains 28 compounds with antiproliferative activities against HeLa cells. The 20 compounds of the first group were subdivided into two parts: 15 molecules in training set (1,2,3,4,5,7,8,10,11,12,15,16,18,19,20) and 5 molecules in test set (6,9,13,14 and 17). A cross correlation matrix (Table S1, supporting information) between descriptors and the $\log IC_{50}$ values of training set demonstrates that SIC_1 and EL have positive correlation with activity, CIC_1, μ, lnMR and lnMV shows negative correlation and EH does not show any significant correlation. SIC_1, CIC_1, molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in **Table 2**.

Table 2 SIC_1, CIC_1, quantum chemical descriptors, molar refractivity and molar volume of 15 training RMGPa inhibitors

Comp no.	SIC_1	CIC_1	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	0.4253	3.6537	-0.2310	0.0089	6.1915	134.70	427.30
2	0.3449	4.3331	-0.2290	-0.0230	3.1764	160.84	501.90
3	0.3411	4.3779	-0.2352	-0.0063	2.2178	162.70	535.20
4	0.3572	4.5137	-0.2289	-0.0299	6.5763	217.62	698.00
5	0.3760	4.3538	-0.2338	-0.0701	5.0497	219.55	680.80
7	0.3823	4.2048	-0.2330	-0.0518	1.8782	192.86	606.90
8	0.4090	3.8272	-0.2255	-0.0302	3.6113	147.59	466.30
10	0.4243	3.8966	-0.2271	-0.0451	1.8798	189.55	582.80
11	0.3602	4.3878	-0.2277	-0.0274	4.1665	192.49	618.80
12	0.3412	4.5893	-0.2160	-0.0229	7.8094	206.39	666.80
15	0.3898	4.2294	-0.2263	-0.0850	3.2123	207.84	651.80
16	0.3809	4.1733	-0.2332	-0.1259	5.0089	183.01	586.90
18	0.3658	4.5019	-0.2298	-0.0881	9.1787	230.42	735.10
19	0.3812	4.3319	-0.2340	-0.0781	8.6982	216.46	689.00
20	0.3800	4.2516	-0.2350	-0.0361	6.2042	196.29	608.70

Among the generated QSAR equations; three equations were finally selected for predicting the inhibitory activity of RMGPa. Three best models are given below:

Model 1

$\log IC_{50} = 0.8954+(2.8856)SIC_1+(3.5920)EL$

Where, N=15, R=0.773, R^2=0.598, R^2_{cv}=0.323, S=0.165, F=8.925, Q=4.685

Model 2

$\log IC_{50} = 3.4496+(4.1769)\ln MR+(-3.7611)\ln MV+(-3.7737)EH+(4.5525)EL+(-0.0256)\mu$

Where, N=15, R=0.864, R^2=0.746, R^2_{cv}=0.256, S=0.172, F=5.287, Q=5.023

Model 3

$\log IC_{50} = -8.4488+(12.1049)SIC_1+(1.1185)CIC_1+(-6.0974)EH+(5.3929)EL+(-0.0388)\mu$

Where, N=15, R=0.946, R^2=0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227

By using model number 1, 2 and 3 the theoretical log IC_{50} values of 15 training compounds are given in **Table 3** together with experimental log IC_{50}.

Table 3 List of experimental and predicted $logIC_{50}$ of 15 training RMGPa inhibitors

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (By model 1)	Predicted $logIC_{50}$ (By model 2)	Predicted $logIC_{50}$ (by Model 3)
1	2.0128	2.1546	1.9002	2.0023
2	1.7024	1.8080	1.9600	1.7217
3	1.8395	1.8570	1.8905	1.8909
4	1.9850	1.8187	1.8639	1.9029
5	1.8401	1.7285	1.8688	1.8239
7	1.9754	1.8125	1.9211	1.9504
8	1.8306	1.9671	1.8205	1.8548
10	2.1106	1.9577	2.0098	2.1142
11	1.8621	1.8363	1.8723	1.8980
12	1.7451	1.7977	1.7660	1.7050
15	1.7388	1.7149	1.7548	1.7970
16	1.4786	1.5423	1.4111	1.3783
18	1.5403	1.6345	1.5798	1.5844
19	1.5551	1.7148	1.6357	1.6789
20	1.9903	1.8622	1.9520	1.9039

The model 3 with the R=0.946, R^2 =0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227 turns out to be the best fit model. The indices of the 5 test compounds are presented in Table S2, supporting information. Using the model number 3, we calculated the theoretical log IC_{50} values of the test set (R=0.891) are given in **Table 4**.

Table 4 List of experimental and predicted $logIC_{50}$ of 5 test RMGPa inhibitors

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (by Model 3)
6	2.2878	2.1409
9	1.6085	2.0331
13	1.4698	1.8409
14	1.5328	1.9068
17	1.2900	1.7081

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 3) are presented in **Figure 1** and **Figure 2** respectively.

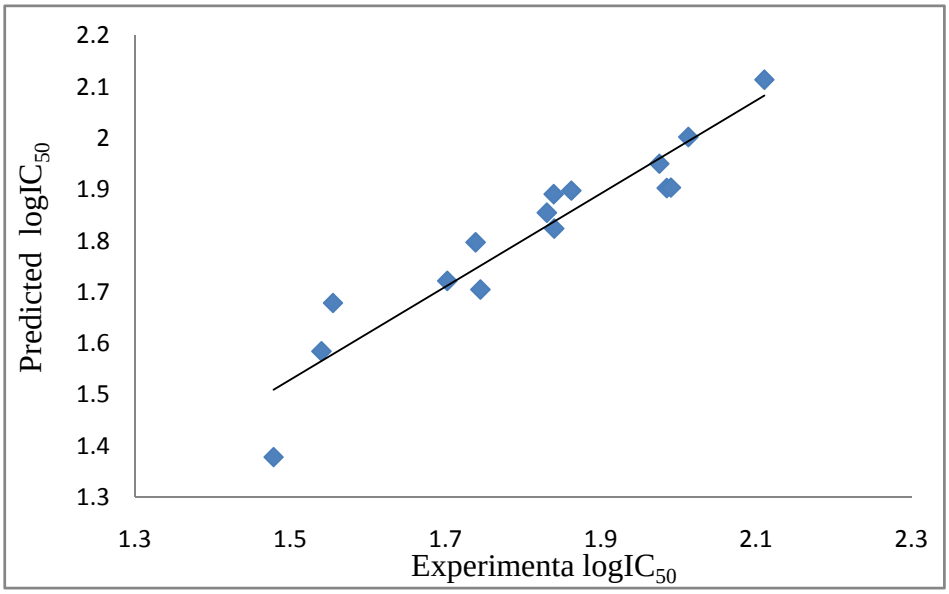


Figure 1: A plot between the predicted and the experimental activities for the training set of RMGPa inhibitors

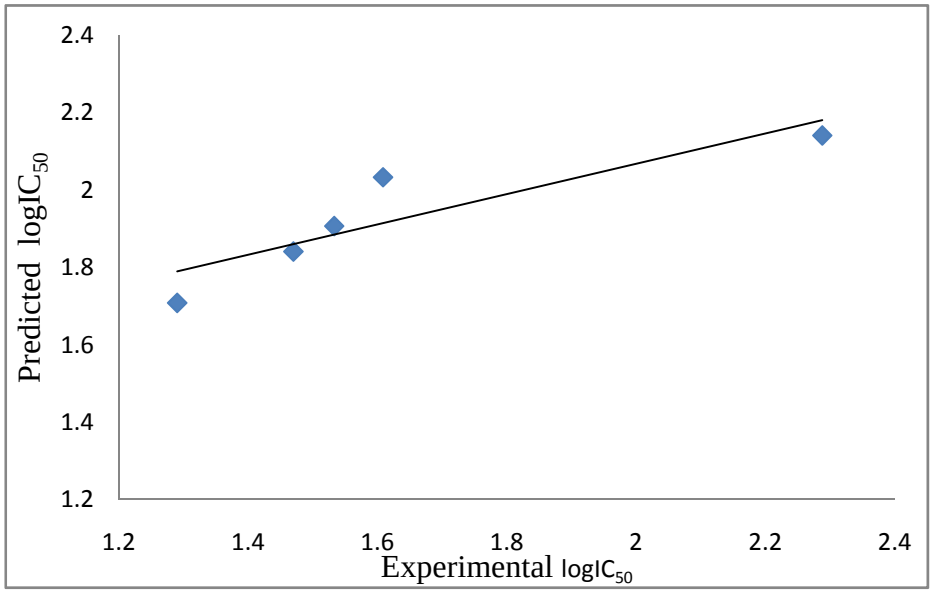


Figure 2: A plot between the predicted and the experimental activities for the test set of RMGPa inhibitors

The second group containing 28 compounds were divided into two parts: 20 molecules in training set (1,21,22,25,26,27,28,32,33,34,35,36,37,38,39,41,42,43,45,46) and test set of 8 molecules (23,24,29,30,31,40,44,47). The correlation matrix of electronic properties and topological indices with cytotoxic activity are presented in Table S3 and S4, supporting information, respectively. Molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in **Table 5** and their topological indices are presented in **Table 6**.

Table 5 Quantum chemical descriptors, molar refractivity and molar volume of 20 training antiproliferative compounds against HeLa cells

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	-0.2310	0.0089	6.1915	134.70	427.30
21	-0.1999	-0.0134	5.1567	191.96	610.10
22	-0.2045	-0.0428	1.6927	202.81	560.70
25	-0.2170	-0.0013	3.8836	166.02	534.90
26	-0.2074	-0.0010	6.0763	175.31	572.80
27	-0.1566	-0.0016	7.3747	173.63	558.80
28	-0.2060	0.0125	7.3125	182.45	571.10
32	-0.2179	-0.0252	3.4896	154.78	490.90
33	-0.1968	-0.0230	2.9099	150.40	476.80
34	-0.2225	0.0001	2.4967	159.80	522.60
35	-0.1948	-0.0245	3.2973	164.44	538.80
36	-0.1963	-0.0375	4.3380	168.64	545.00
37	-0.2111	-0.0179	5.4231	169.07	555.60
38	-0.1693	-0.0068	4.6924	165.82	527.40
39	-0.1731	-0.0013	1.2905	168.82	536.40
41	-0.1986	-0.0012	3.8606	150.30	483.60
42	-0.1978	-0.0255	3.3445	154.94	499.80
43	-0.2032	-0.0367	5.9112	159.56	516.60
45	-0.1892	-0.0141	4.5361	165.07	522.10
46	-0.2008	0.0058	7.5438	5.2403	6.3820

Table 6 Topological indices of 20 training antiproliferative compounds against HeLa cells

Comp no.	SIC_1	CIC_1	ln W	H	χ^0
1	0.4253	3.6537	7.9649	162.1636	24.8970
21	0.3517	4.4462	8.9289	251.9431	34.2775
22	0.4288	3.9028	9.0789	263.7028	37.1059
25	0.3872	4.0716	8.5865	214.7929	30.2943
26	0.3720	4.2252	8.7828	228.4780	32.4156
27	0.3826	4.1372	8.7686	235.9327	32.7419
28	0.3484	4.4273	8.7749	233.0269	31.9930
32	0.3794	4.0679	8.3622	195.8768	28.0099
33	0.3866	4.0112	8.4016	191.0617	28.4325
34	0.3739	4.1325	8.5254	5.3306	30.1730
35	0.3714	4.1765	8.5916	213.8546	31.0957
36	0.3743	4.1747	8.7221	225.2267	31.7085
37	0.3621	4.2655	8.6630	217.3454	31.5872
38	0.3919	4.0402	8.6484	221.8434	31.0014
39	0.3749	4.1707	8.6484	221.8434	31.0014
41	0.3756	4.0734	8.2890	187.6982	27.8885
42	0.3732	4.1177	8.3612	194.8093	28.8112
43	0.3630	4.2135	8.4670	198.6210	29.3028
45	0.4187	3.8536	8.6769	215.3581	31.7504
46	0.3665	4.3284	8.9253	245.4236	34.2775

Among the generated QSAR models; three models were finally selected. Model summary of three best models are given below.

Model 4

$\text{Log } IC_{50}$= 19.6957+(-4.4880)SIC_1+(-0.8479)CIC_1+(0.0427)lnW+(-3.0400)lnH+(0.0885)$\chi 0$

Where, N=28, R=0.859, R^2=0.738, R^2_{cv}=0.288, S=0.197, F=7.887, Q=4.360

Model 5

$\text{Log } IC_{50}$= 13.3115+(-0.1637)lnMR+(-1.8648)lnMV+(-2.2260)EH+(-4.4968)EL+(-0.0005)μ

Where, N=28, R=0.933, R^2=0.870, R^2_{cv}=0.594, S=0.205, F=18.738, Q=4.551

Model 6

Log IC_{50}= 13.4441+(-11.8969)SIC_1+(-2.0004)CIC_1+(-2.7582)EH+(-4.5601)EL+(-0.0010)μ
Where, N=28, R=0.916, R^2=0.839, R^2_{cv}=0.698, S=0.201, F=14.591, Q=4.557

Model 4 and model 5 are constructed by using topological indices and electronic parameters respectively. But the quality of the equation is much increased when we use both the topological and electronic indices simultaneously i.e. Model 6. By using model number 4, 5 and 6 the theoretical log IC_{50} values of 20 training compounds are given in **Table 7** together with experimental log IC_{50}. The indices of the 8 test compounds are given in Table S5, supporting information.

Table 7 List of experimental and predicted logIC_{50} of 20 training antiproliferative compounds against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (By model 4)	Predicted logIC_{50} (by Model 5)	Predicted logIC_{50} (By model 6)
1	1.6648	1.7631	1.6839	1.6658
21	1.0334	0.9534	0.9934	0.9731
22	1.2521	1.1862	1.2860	1.2931
25	1.2751	1.2294	1.2467	1.2933
26	0.9628	1.1759	1.0864	1.1369
27	0.9903	1.1337	1.0231	1.0481
28	0.9253	1.0126	0.9836	0.9467
32	1.5352	1.3361	1.5281	1.5054
33	1.6478	1.4666	1.5305	1.4655
34	1.2891	1.3430	1.3031	1.3399
35	1.3670	1.2959	1.2901	1.3166
36	1.2502	1.1867	1.3258	1.3481
37	1.3189	1.2595	1.2338	1.2620
38	1.2627	1.2020	1.1914	1.1929
39	1.0422	1.1676	1.1423	1.1229
41	1.3477	1.4644	1.4096	1.3766
42	1.4864	1.4093	1.4510	1.4256
43	1.3191	1.3629	1.4456	1.4187
45	1.3444	1.3977	1.2881	1.3357
46	1.0976	1.0663	0.9696	0.9452

In these models, N is the number of data points; R is the correlation coefficient between experimental values and calculated values from the equation. R^2 is the square of the correlation coefficient and it measures the goodness of fit of the regression equation. Cross validated coefficient (R^2_{cv}) gives an idea of the performance of the model. S is the standard deviation of the regression. Fischer statistics (F) is a ratio of variances between calculated and observed activity. The larger value of F test signifies the QSAR model. Q is the quality factor. Q value measures predictive power of the QSAR models.

We calculated the theoretical log IC_{50} of the test set (R=0.751) by model number 6 which appeared in **Table 8**.

Table 8 List of experimental and predicted logIC_{50} of 8 test antiproliferative inhibitors against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (by Model 6)
23	1.1166	1.2684
24	0.6848	1.1678
29	1.6652	1.3425
30	1.8995	1.4045
31	0.8525	1.3018
40	0.8686	1.2064
44	0.8733	1.3191
47	0.9175	1.3274

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 6) are presented in **Figure 3** and **4** respectively.

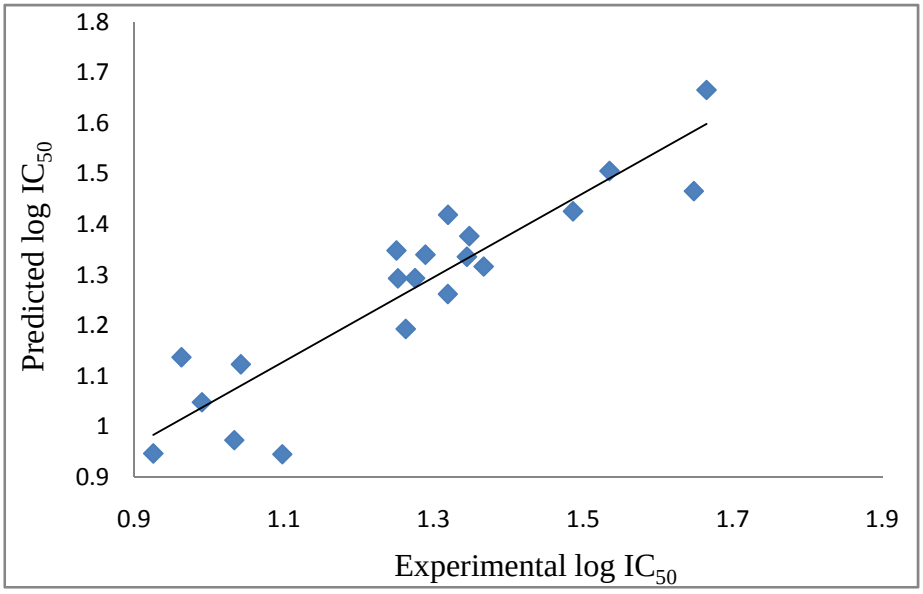


Figure 3: A plot between the predicted and the experimental activities for the training antiproliferative inhibitors against HeLa cells

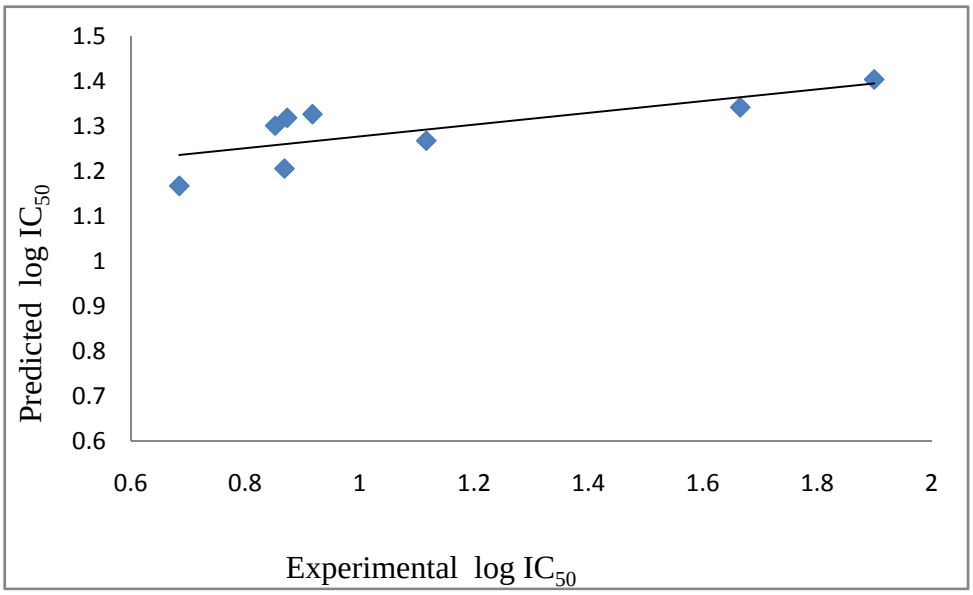


Figure 4: A plot between the predicted and the experimental activities for the test antiproliferative inhibitors against HeLa cells

We also calculated the antiproliferative activity (from compound number 2 to 20) and inhibitory activity of RMGPa (from compound number 21 to 47) by model number 6 and 3 respectively which were not determined experimentally and are shown in **Table 9**.

Table 9 Predicted Antiproliferative activity/ RMGPa inhibitory activity of studied compounds

Comp. No.	HeLa IC_{50}(μM)	RMGPa IC_{50}(μM)	Comp. No.	HeLa IC_{50}(μM)	RMGPa IC_{50}(μM)
2	25.48	-	25	-	90.72
3	20.12	-	26	-	63.61
4	8.44	-	27	-	29.51
5	16.63	-	28	-	57.57
6	11.34	-	29	-	86.53
7	22.99	-	30	-	43.87
8	47.69	-	31	-	89.11
9	26.87	-	32	-	56.37
10	27.02	-	33	-	47.91
11	13.50	-	34	-	91.14
12	7.88	-	35	-	44.26
13	12.88	-	36	-	37.83
14	12.61	-	37	-	48.47
15	22.63	-	38	-	42.42
16	59.79	-	39	-	56.57
17	36.50	-	40	-	39.14
18	12.97	-	41	-	51.11
19	17.23	-	42	-	41.03
20	16.79	-	43	-	29.49
21	-	53.46	44	-	43.88
22	-	114.14	45	-	67.81
23	-	67.81	46	-	61.92
24	-	68.55			

Docking

The binding energies of 47 docked compounds are given in Table S6, supporting information, and are ranges between -4.58 and -13.2kcal/mol. The docking study shows both polar (GLN7, ARG10, LYS11, SEP14, ARG69, GLN71, GLN72, TYP74, TYR75, GLU76, ARG81, TYR155, LYS191, ARG193, GLU195, THR240, ARG242, ARG306, ARG309, AGR310, LYS312, SER313, SER314) and non polar (ILE13, VAL15, LEU18, ILE63, VAL64, TRP67, ILE68, PHE196,

ASP227) amino acids make important interactions to the inhibitors.

In our study, we used different topological and quantum chemical indices to obtain phenogram based on Unweighted Pair Group Method with Arithmetic mean (UPGMA) of 47 compounds. The phenogram of the compounds is presented in **Figure 5**. Compounds having similar molecular properties are in the same clade though in some cases their binding energies and activity are different.

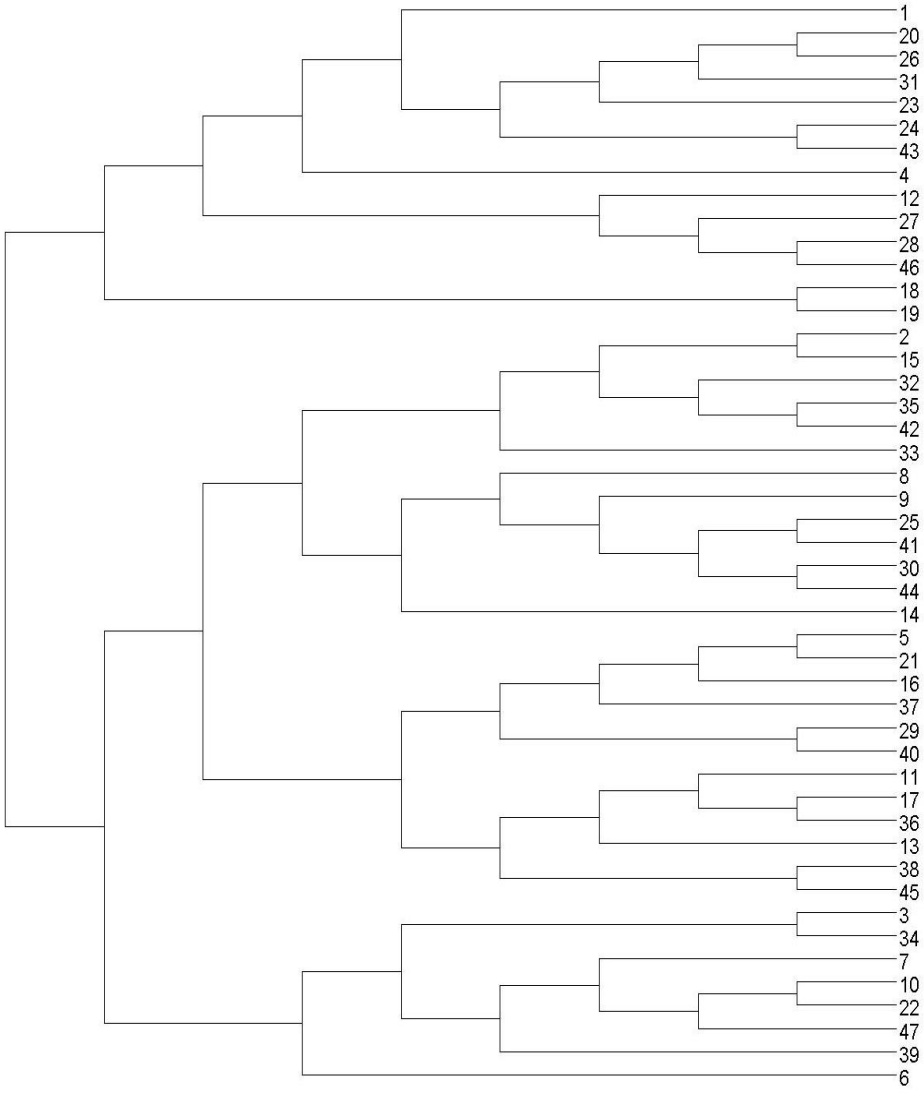


Figure 5: Phenogram using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) of 47 compounds of 23-hydroxy betulinic acid derivatives

Compound 1, 20, 26, 31, 23, 24, 43, 4 are in same clade. Ligand 1 (23-hydroxy betulinic

acid) was used as a model drug (**Figure 6a**) as a reference for other compounds and have three hydrogen bonds. The carboxylate group at C-28 forms two hydrogen bonds at 1.743Å and 1.821Å with ARG310 and ARG 242 respectively. The –OH group at C-3 also forms another hydrogen bond with ILE68 (1.988 Å). In this clade, compound 20 and 26 remain as a pair and they have almost same binding energy. The binding energy of 20 and 26 are -7.38 and -7.56kcal/mol respectively. Compound 24 and 43 have different binding energies though they remain as a pair. Their predicted IC_{50} values are 68.50μM and 29.49μM respectively. In case of ligand 43 (**Figure 6b**), –OH group present at C-30 and the ester group at C-28 form two hydrogen bonds with SER313 (2.181 Å) and ARG310 (1.933 Å) respectively. Ligand 24 also forms two hydrogen bonds with ARG81 and GLN71 but at the same time the N-substituted amide group at C-28 may increase the chance of steric bumps with ARG309, which weakens the interaction between ligand and enzyme causing decrease in bioactivity (**Figure 6c**).

Compound 12, 27, 28, 46 are in same clade in which 46 has highest negative binding energy. In case of ligand 46 (**Figure 6d**), the carboxylate group at C-28 forms hydrogen bond with ARG193 (1.937Å). Compound 18 and 19 remain as a pair in the next clade and have comparable binding energy. Compound 2, 15, 32, 35, 42, 33 are present in same clade. Though compound 2 and 15 are in a pair, their binding energies are different. The IC_{50} values of ligand 2 and 15 are 50.4μM and 54.8μM respectively. Compound 2 lies well inside the protein (**Figure 6e**) and the –OH group at C-23 and C=O group at C-28 make two hydrogen bonds with GLN72 (2.675Å) and ARG310 (2.015Å) respectively. Although the amide group at C-28 of ligand 15 makes important interactions with the enzyme but the major part of the ligand is outside the protein (**Figure 6f**). Compound 35 and 42 are in a pair in same clade and they have almost same binding energy. In another clade, compound 8, 9, 25, 41, 30, 44, 14 are present and their binding energies are almost as same as the other.

Compound 5, 21, 16, 37, 29, 40 are in same clade. Compound 16 has highest negative binding energy than the other compounds and forms one hydrogen bond with ARG310 (**Figure 6g**). In the next clade compound 17 and 36 are in a pair. The binding energy of 36 is more negative than 17. In ligand 36, the –OH group at C-3 and O atom at C-23 form hydrogen bonds with ARG310 at 2.002Å and 2.223Å respectively (**Figure 6h**). The acetyl group at C-23 also forms hydrogen bond with ARG242 (1.87Å). Ligand 17 forms hydrogen bonds with ARG69, GLN71 and TYR155. In spite of the steric bump formation between the long chains at C-3 with ARG310, this compound possesses good inhibitory activity due to hydrogen bonds (**Figure 6i**). Compound 3, 34 and 10, 22 are in pairs and they have comparable binding energy.

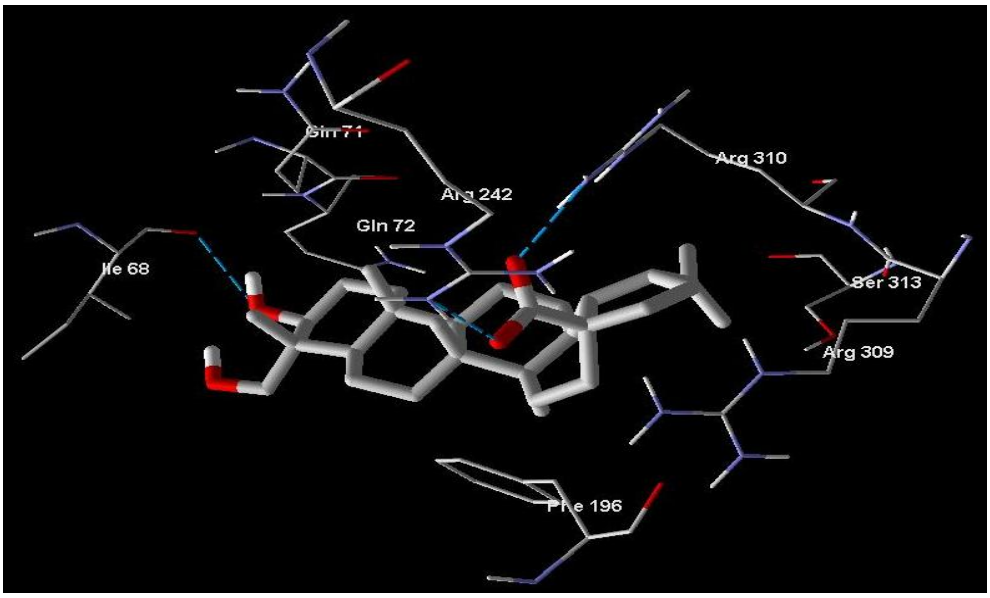


Figure 6a

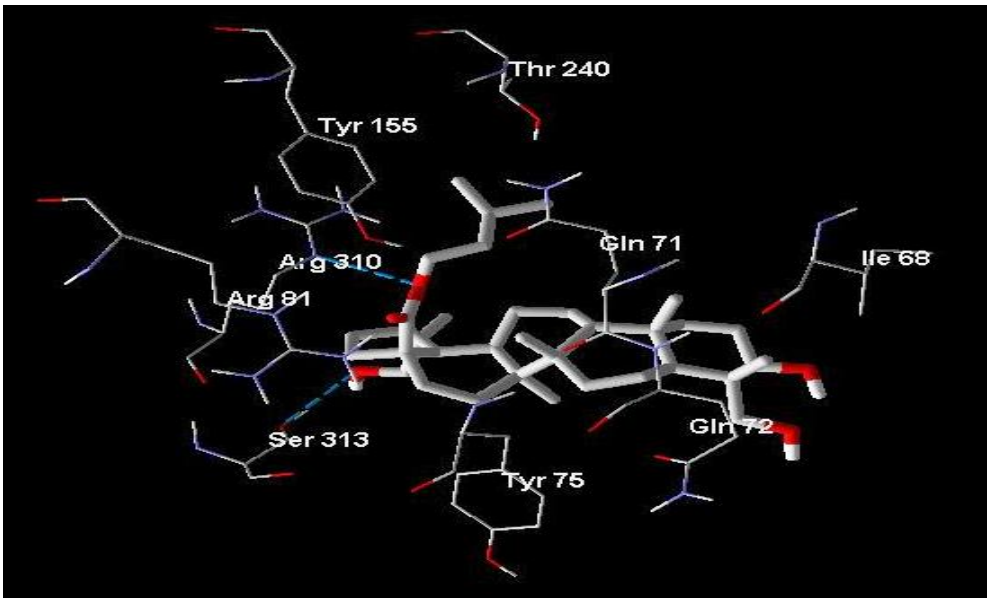


Figure 6b

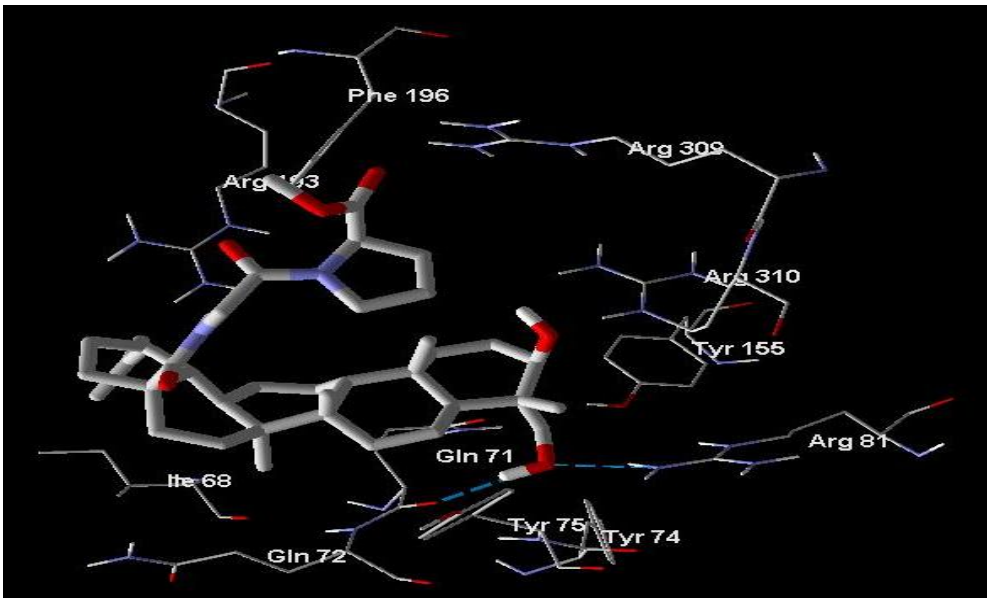


Figure 6c

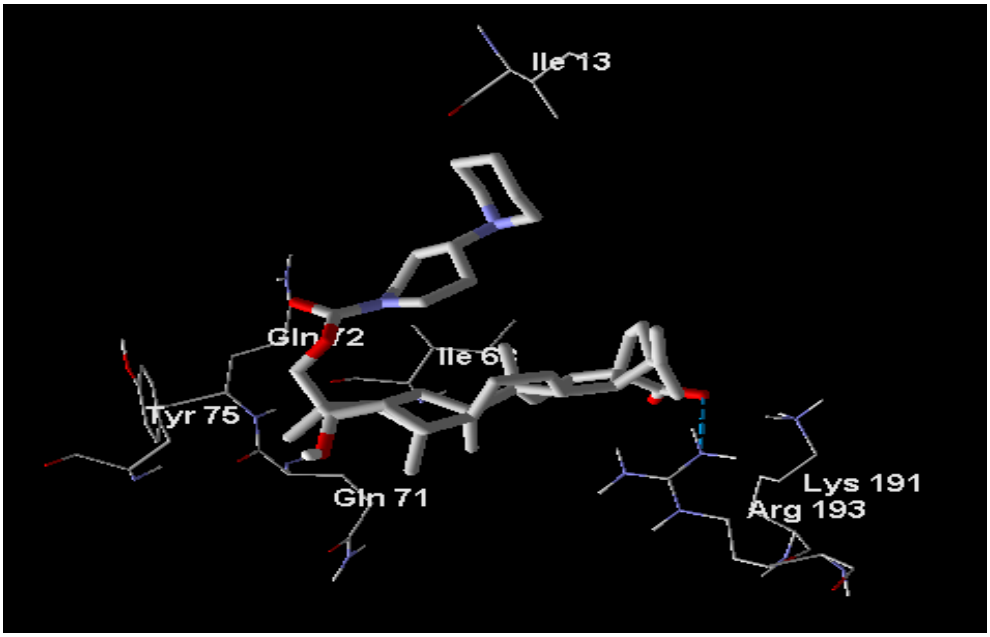


Figure 6d

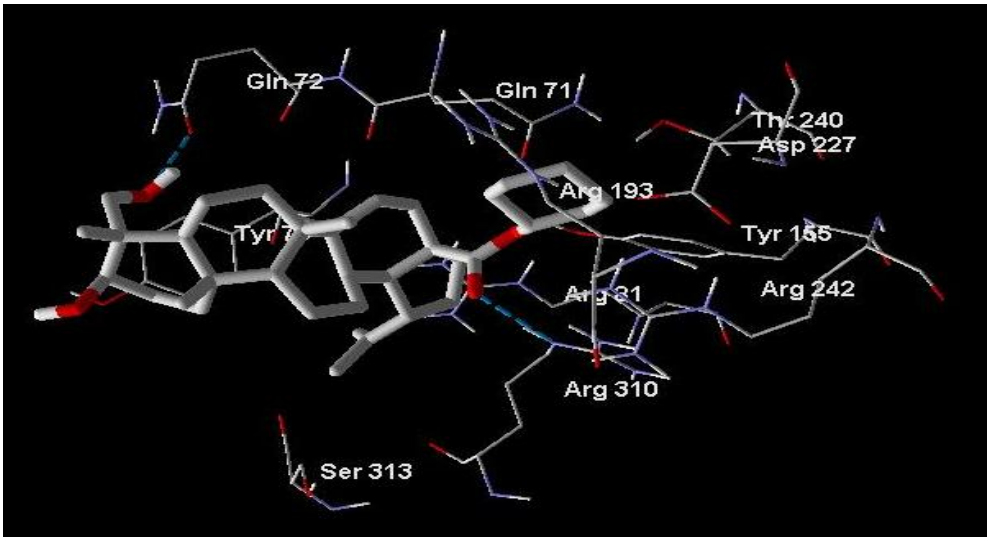


Figure 6e

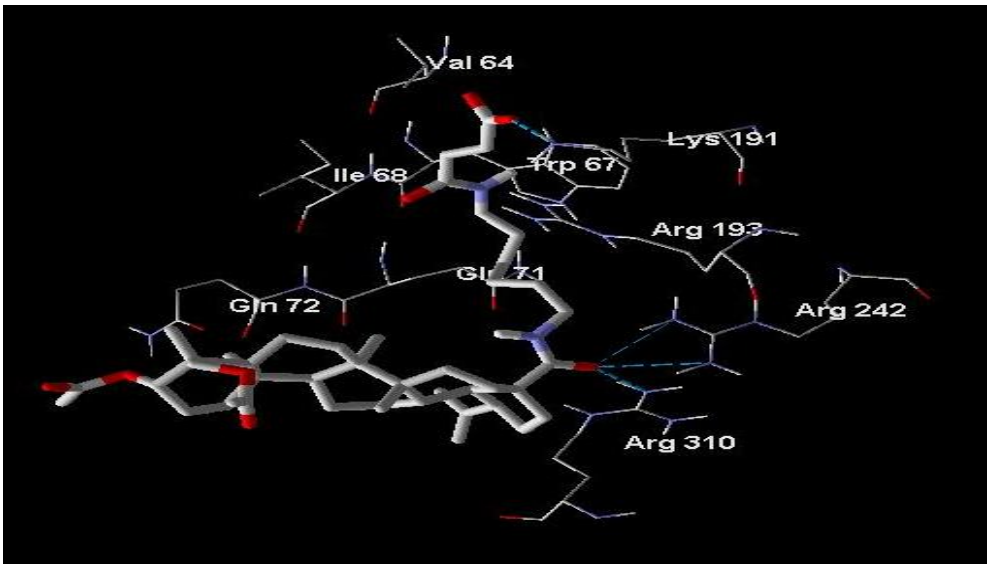


Figure 6f

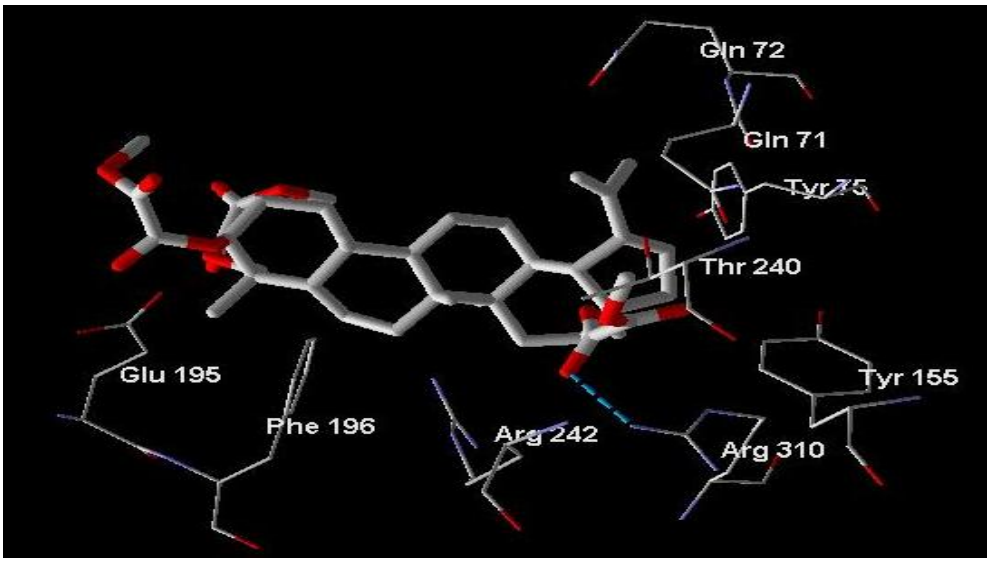


Figure 6g

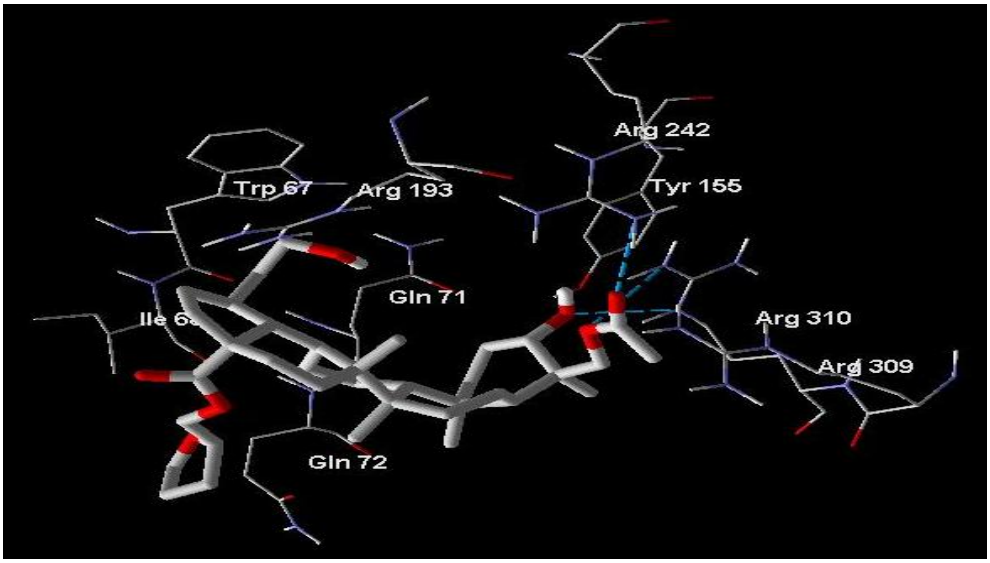


Figure 6h

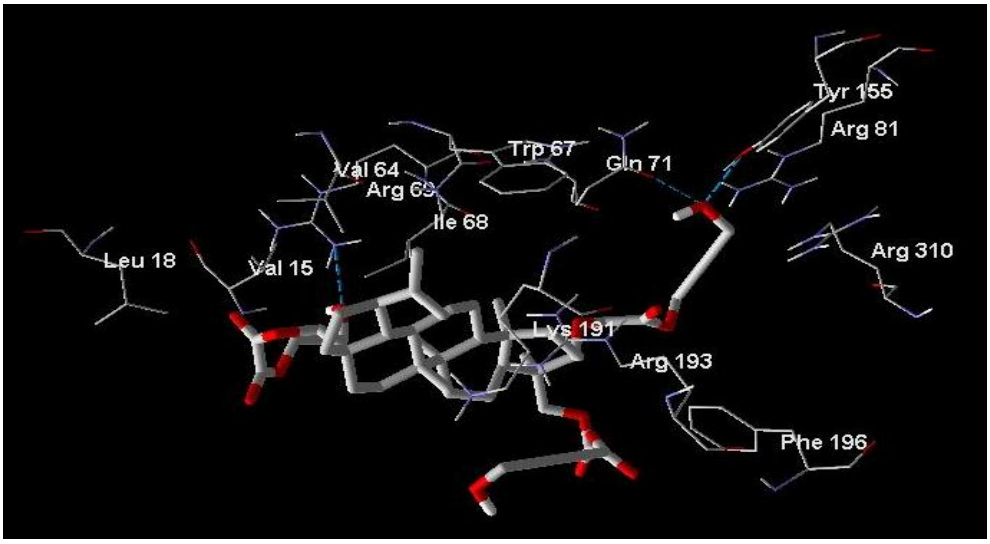


Figure 6i

Figure 6: Poses of different ligands in the active site of rabbit muscle glycogen phosphorylase a (RMGPa). (6a) Docked conformation of ligand 1 along with the important amino acid residues of RMGPa. (6b) Docked conformation of ligand 43 along with the important amino acid residues of RMGPa. (6c) Docked conformation of ligand 24 along with the important amino acid residues of RMGPa. (6d) Docked conformation of ligand 46 along with the important amino acid residues of RMGPa. (6e) Docked conformation of ligand 2 along with the important amino acid residues of RMGPa. (6f) Docked conformation of ligand 15 along with the important amino acid residues of RMGPa. (6g) Docked conformation of ligand 16 along with the important amino acid residues of RMGPa. (6h) Docked conformation of ligand 36 along with the important amino acid residues of RMGPa. (6i) Docked conformation of ligand 17 along with the important amino acid residues of RMGPa.

In conclusion, this QSAR study has shown that topological indices (e.g. SIC, CIC) and quantum chemical descriptors (e.g. EH, EL, μ) are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. Model 3 and model 6 are the best equation for predicting the inhibitory activity of RMGPa and the antiproliferative activities against HeLa cells respectively and these QSAR models may be used in prediction of activity of designed compounds. The docking study shows that the important interacting amino acids present in the active site are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the ligands can form hydrogen bonds with ARG193 and/or ARG310. The –OH group at C-3 and C-23 can increase the hydrogen bond interaction

between ligands and enzyme. However, acetylation or esterification of the –OH group at the C-3 and C-23 not only decreases the number of hydrogen bond, but may also increase the unfavorable steric clashes. Thus binding energy may decrease. Large substituent at C-17 may increase the chance of steric bumps, thus lowering the inhibitory activity of the ligand.

References

[1] M. BOLLEN, S. KEPPENS AND W. STALMANS, Specific features of glycogen metabolism in the liver, Biochem J, 336 (1998), pp. 19-31.

[2] H. G. HERS, The Control of Glycogen Metabolism in the Liver, Annu Rev Biochem, 45 (1976), pp. 167-190.

[3] W. STALMANS, The Interaction of Liver Phosphorylase a with Glucose and AMP, Eur J Biochem, 49 (1974), pp. 415-427.

[4] W. PIMENTA, N. NURJHAN, P. A. JANSSON, M. STUMVOLL, J. GERICH AND M. KORYTKOWSKI, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia, 37 (1994), pp. 697-702.

[5] M. K. HELLERSTEIN, R. A. NEESE, P. LINFOOT, M. CHRISTIANNSEN, S. TURNER AND A. LETSCHER, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest, 100 (1997), pp. 1305-1319.

[6] N. G. OIKONOMAKOS, E. D. CHRYSINA, M. N. KOSMOPOULOU AND D. D. LEONIDAS, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 A resolution, Biochim Biophys Acta, 1647 (2003), pp. 325– 332.

[7] V. L. RATH, M. AMMIRATI, D. E. DANLEY, J. L. EKSTROM, E. M. GIBBS, T. R. HYNES, A. M. MATHIOWETZ, R. K. MCPHERSON, T. V. OLSON, J. L. TREADWAY AND D. J. HOOVER, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol, 7 (2000), pp. 677-682.

[8] R KURUKULASURIYA, J. T. LINK, D. J. MADAR, Z. PEI, S. J. RICHARDS J. J. ROHDE, A. J. SOUERS AND B. G. SZCZEPANKIEWICZ, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem, 10 (2003), pp. 123-153.

[9] Z. LIANG, L. ZHANG, L. LI, J. LIU, H. LI, L. ZHANG, L. CHEN, K. CHENG, M. ZHENG, X. WEN, P. ZHANG, J. HAO, Y. GONG, X. ZHANG, X. ZHU, J. CHEN, H. LIU, H. JIANG, C. LUO AND H. SUN, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem, 46 (2011), pp. 2011-2021.

[10] X. WEN, H. SUN, J. LIU, K. CHENG, P. ZHANG, L. ZHANG, J. HAO, L. ZHANG, P. NI, S. E. ZOGRAPHOS, D. D. LEONIDAS, K. M. ALEXACOU, T. GIMISIS, J. M. HAYES AND N. G. OIKONOMAKOS, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem, 51 (2008), pp. 3540–3554.

[11] P. ZHU, Y. BI, J. XU, Z. LI, J. LIU, L. ZHANG, W. YE AND X. WU, Terpenoids. III: Synthesis and biological evaluation of 23-hydroxybetulinic acid derivatives as novel inhibitors of glycogen phosphorylase, Bioorg Med Chem Lett, 19 (2009), pp. 6966–6969.

[12] P. LAN, J. WANG, D. M. ZHANG, C. SHU, H. H. CAO, P. H. SUN, X. M. WU, W. C. YE AND W. M. CHEN, Synthesis and antiproliferative evaluation of 23-hydroxybetulinic acid derivatives, Eur J Med Chem, 46 (2011), pp. 2490-2502.

[13] Y. BI, J. XU, F. SUN, X. WU, W. YE, Y. SUN AND W. HUANG, Synthesis and biological activity of 23-hydroxybetulinic acid C-28 ester derivatives as antitumor agent candidates, Molecules, 17 (2012), pp. 8832-8841.

[14] W. DING, M. SUN, S. LUO, T. XU, Y. CAO, X. YAN AND Y. WANG, A 3D QSAR Study of Betulinic Acid Derivatives as Anti-Tumor Agents Using Topomer CoMFA: Model Building Studies and Experimental Verification, Molecules, 18 (2013), pp. 10228-10241.

[15] S. FULDA AND K. M. DEBATIN, Betulinic Acid Induces Apoptosis Through a Direct Effect on Mitochondria in Neuroectodermal Tumors, Med Pediatric Oncol 35 (2000), pp. 616–618.

[16] Z. N. JI, W. C. YE, G. G. LIU AND W. L. WENDY HSIAO, 23-Hydroxybetulinic Acid-Mediated Apoptosis is Accompanied by Decreases in Bcl-2 Expression and Telomerase Activity in HL-60 Cells, Life Sci, 72 (2002), pp. 1–9.

[17] M MANDAL AND R KUMAR, Bcl-2 Modulates Telomerase Activity, J Biol Chem, 272 (1997), pp. 14183–14187.

[18] Y ZHENG, F. ZHOU, X. WU, X. WEN, Y. LI, B. YAN, J. ZHANG, G. HAO, W. YE AND G. WANG, 23-Hydroxybetulinic acid from Pulsatilla chinensis (Bunge) Regel synergizes the antitumor activities of doxorubicin in vitro and in vivo, Journal of Ethnopharmacol, 128 (2010), pp. 615–622.

[19] ACD/ChemSketch version 12.01 Advanced Chemistry Development, Inc., Toronto, Ontario, 2009

[20] E. EROGLU, Some QSAR Studies for a Group of Sulfonamide Schiff Base as Carbonic Anhydrase CA II Inhibitors, Int J Mol Sci, 9 (2008), pp. 181-197.

[21] L. B. KIER, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci, 69 (1980), pp. 807-810.

[22] S. C. BASAK, D. K. HARRISS AND V. R. MAGNUSON, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci, 73 (1984), pp. 429-437.

[23] H. WIENER, Structural Determination of Paraffin Boiling Points, J Am Chem Soc, 69 (1947), pp. 17-20.

[24] D. PLAYSIC, S. NIKOLIC, N. TRINAJSTIC AND Z. MIHALIC, On the Harary Index for the Characterization of Chemical Graphs, J Math Chem, 12 (1993), pp. 235-250.

[25] M. RANDIAC, On the characterization of Molecular branching, J Am Chem Soc, 97 (1975), pp. 6609-6615.

[26] L. B. KIER AND L. H. HALL, Molecular connectivity in structure-activity analysis, Research studies press: Letchworth, Hertfordshire, U K (1986).

[27] M. W. SCHMIDT, K. K. BALDRIDGE, J. A. BOATZ, S. T. ELBERT, M. S. GORDON AND J. H. JENSEN, GAMESS Version= 24 Mar 2007 (R1) from Iowa State University, J Comput Chem, 14 (1993), pp. 1347-1363.

[28] G. M. MORRIS, R. HUEY, W. LINDSTROM, M. F. SANNER, R. K. BELEW, D. S. GOODSELL AND A. J. OLSON, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J Comput Chem, 30 (2009), pp. 2785–2791.

[29] R. HUEY, G. M. MORRIS, A. J. OLSON AND D. S. GOODSELL, Software News and Update A Semiempirical Free Energy Force Field with Charge-Based Desolvation, J Comput Chem, 28 (2007), pp. 1145–1152.

[30] R. HUEY, D. S. GOODSELL, G. M. MORRIS AND A. J. OLSON, Grid-based hydrogen bond potentials with improved directionality, Lett Drug Des Discov, 1 (2004), pp. 178–183.

[31] R. THOMSEN AND M. H. CHRISTENSEN, MolDock: a new technique for high-accuracy molecular docking, J Med Chem, 49 (2006), pp. 3315–3321.