

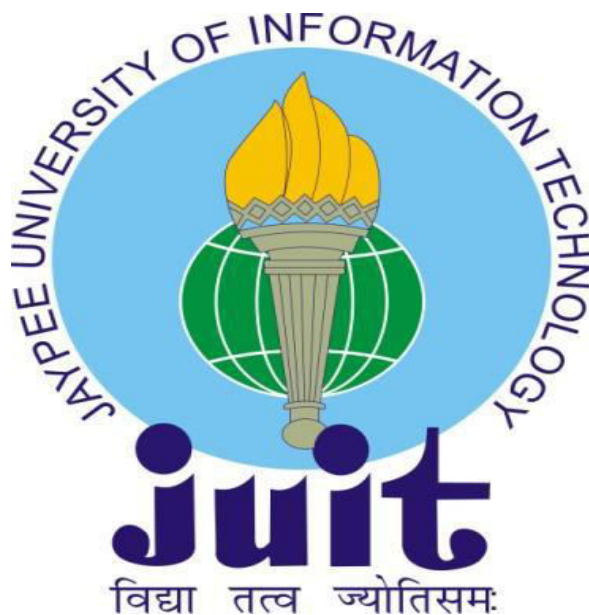
**DEVELOPMENT OF NOVEL QSPR STRATEGY AND
MODELS FOR THE SOLUBILITY PREDICTION AND
IN-SILICO BIOAVAILABILITY STUDY OF
PACLITAXEL PRODRUGS**

Thesis submitted in fulfilment for the requirements of the degree of

DOCTOR OF PHILOSOPHY

By

NUPUR



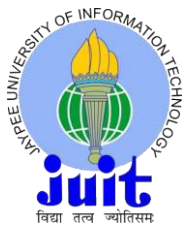
Department of Biotechnology and Bioinformatics

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, DISTRICT SOLAN, H.P., INDIA

August, 2020

Copyright
@
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT
AUGUST, 2020
ALL RIGHTS RESERVED



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

(Established by H.P. State Legislative Vide Act no. 14 of 2002)
Waknaghat, P.O. Dumehar Bani, Kandaghat, Distt. Solan – 173215 (H.P.) INDIA

Website: www.juit.ac.in

Phone No. (91) 01792-257999(30 Lines).

Fax: (91) 01792 245362

DECLARATION

I certify that:

- The work contained in this thesis is original and has been done by me under the guidance of my supervisor.
- The work has not been submitted to any other organization for any degree or diploma.
- Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

Nupur

Date: 19/08/2020

Enrollment No. 136501

Department of Biotechnology & Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan, India



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

(Established by H.P. State Legislative Vide Act no. 14 of 2002)

Waknaghat, P.O. Dumehar Bani, Kandaghat, Distt. Solan – 173215 (H.P.) INDIA

Website: www.juit.ac.in

Phone No. (91) 01792-257999(30 Lines).

Fax: (91) 01792 245362

CERTIFICATE

This is to certify that the thesis entitled, “**Development of novel QSPR strategy and models for the solubility prediction and *in-silico* strategy bioavailability study of Paclitaxel prodrugs**” which is being submitted by **Nupur (Enrollment No. 136501)** in fulfillment for the award of degree of **Doctor of Philosophy in Bioinformatics** at **Jaypee University of Information Technology, Waknaghat, India** is the record of candidate’s own work carried out by her under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Tiratha Raj Singh

Associate Professor

Dept. of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan, H.P. India-173234

Date:

ACKNOWLEDGEMENT

The people I need to thank will not opportune to a single acknowledgement section. It is my right and honor to express my gratitude to the following without whose support and conduct I would not been able to complete my PhD voyage.

Firstly, I would like to express my profound gratitude to my supervisor, **Dr. Tiratha Raj Singh**. His supervision, guidance and faith in me are baffling. He is the patron who gives optimism and advice vital for me to proceed through my doctoral program and complete my dissertation. My sincere thanks go to **Dr. Singh** for his tireless stimulus. He has been substantial and ancillary adviser to me during my PhD journey.

I would also like to thank **Late Dr. Chittaranjan Rout** for his interest in the study, persistent guidance, motivation, criticism that help me evolve in this competitive world during my PhD journey .

I am acutely indebted to my parents, **OP Munjal** and **Krishna Munjal** and my sisters **Alka Sukhija, Monika Munjal Mehta and Shelley Karan Palta** for their love, patience, support and belief in me that provided me inspiration and they were the motivation for successful completion of my work. And to my little chaps **Divij Bharat, Prerika** and **Veer Karan** their innocence make me smile every time. I behold everything to them and i have no words for their love and affection.

Mr. Sameer Narwal has been a great pillar of support for me in all these years. His confidence, understanding and reinforcement have provided me the courage to go through many challenges on the way. Together with my family he is the most relieved one. He is recipient of all the laurels as i would not be me without him.

I sincerely express my gratitude to JUIT administration, **Prof (Dr.) Vinod Kumar** (Vice Chancellor), **Prof (Dr.) S.D. Gupta** (Director&Dean, Academics & Research), **Maj. Gen (Retd.) Rakesh Bassi** (Registrar), **Prof (Dr.) Y. Medury** (Former COO, JES), **Prof (Dr.) S.K. Kak** (Former vice chancellor), **Brig (Retd.) K.K. Marwah** (Former Registrar), **Brig (Retd.) Balbir Singh** (Former Registrar), **Prof (Dr.) T.S. Lamba** (Former Dean, Academic & Research), **Prof (Dr.) R.S. Chauhan** (Former Head, BT & BI), **Prof (Dr.) Sudhir Kumar** (Head, BT & BI) for giving me the opportunity to pursue the doctorate

degree, teaching assistantship and advanced lab infrastructure to accomplish the scientific endeavor of my life.

I am also highly beholden to **Prof (Dr.) Jayaram, Dr. Shashank Shekhar and Dr. Rahul Kaushik** from **Indian Institute of Technology, Delhi** for providing me supercomputing facility, valuable suggestions and tremendous cooperation. I am highly thankful to **Dr. Goutam Mukherjee** Indian Institute of Technology, Delhi for helping me in learning Gaussian software and carrying out the structure optimization for my work.

I am also pleased to thank **Prof (Dr.) SP Ghreera, Dr. PK Gupta, Dr. Udaybhanu M, Dr. Jayshree Ramanna, and Dr. Raj Kumar** for their consistent support and valuable suggestions at various stages of this research work. I express my immense appreciation to **Dr. Ragothaman Yenamalli** and **Dr. Narendra Kumar** for providing congenial research environment and to all the faculty members of Department of Biotechnology and Bioinformatics for encouragement and comfortable practice.

I am also thankful to all the members of technical and non-technical staff of the department, especially **Mrs. Somlata Sharma, Mr. Baleshwar, Mrs. Sonika, Mrs. Mamta and Mr. Ismail** for their assistance and valuable contributions.

My heartfelt thanks to all my friends **Dr. Shifa, Dr. Garima, Kartiki, Dr. Smriti, Dr. Ritika, Poonam, Dr. Bandana, Dr. Kavita, Dr. Kusum, Dr. Imran, Dr. Ankita, Dr. Ashwini, Dr. Charu, Dr. Seneha, Dr. Manika, Shikha, Pulkit, Swati, Rohit, Mahima, Kanishka, Nadia, Arvind**, for their ever needed cooperation. I express my gratitude to all research scholars of the Biotechnology & Bioinformatics Department for keeping me blessed with best wishes.

Last, I thank the one above all of us, omnipresent God, for listening to my prayers and making capable to face each and every phase of my life.

All may not be mentioned, but no one is forgotten.

Nupur

ABSTRACT

Cancer is one of the most deadly diseases prevailing now-a-days. One of the common characteristic that all cancer cells share is the abnormal growth of cells. Chemotherapy is a universal way for treating cancer. However, technical advances in computational methods, simple yet very sensitive assays and the accompaniment of combinatorial chemistry artistry has now made the backbone of medicinal chemistry research field.

Taxol having the encouraging anti-cancer activity but there is delay in the development of this drug because of its low aqueous solubility and hypersensitivity posed by its formulation in cremophor EL. To avoid this hypersensitivity and obtain better clinical use of paclitaxel, developing a new co-solvent and improving the formulation for paclitaxel delivery systems has become important. In addition to developing better formulation for paclitaxel, there were ways to achieve more efficient modified paclitaxel that is prodrug formation.

Different approaches to build up paclitaxel prodrugs were introduced making it more soluble or bioactive before its clinical use. But the approach of formation of these prodrugs was not rational as it is very tedious process and many undesirable compounds were also synthesized. Theoretical models such as quantitative structure property-relationship (QSPR) can provide a set of predictors for any molecular property using the structure of the molecule. Delving into the cost- and time-effective computational strategies for the solubility prediction of prodrugs was performed. Also to increase the bioavailability, study with human metabolic enzymes must be done. With respect to these research gaps, the proposed thesis work has been conducted and the objectives are defined in the three different chapters (Chapter 2, 3 & 4). In the first objective, a novel QSPR strategy (QSPR-sPL) has been developed for the solubility prediction of small datasets as from pharmaceutical perspective small datasets are important for the solubility prediction. In the second objective a dataset of paclitaxel prodrug was and the substituent groups were taken for the solubility prediction implementing the pipeline developed in the first objective. Also to increase the bioavailability, study with human metabolic enzymes hence in the third objective the metabolic study of phosphate paclitaxel prodrugs was performed *in-silico* to inspect their oral bioavailability and found the highly soluble isotaxel to be more bioavailable that could be given orally also. Fulfillment of these objectives contributed to the solubility property prediction of low soluble drugs and solubility prediction of paclitaxel prodrugs

TABLE OF CONTENTS

CONTENT	PAGE NO.
INNER FIRST PAGE	i
COPYRIGHT INFORMATION	ii
TABLE OF CONTENTS	iii-v
DECLARATION	vi
SUPERVISOR'S CERTIFICATE	vii
ACKNOWLEDGEMENT	viii-ix
ABSTRACT	x
LIST OF ABBRIVIATIONS	xi-xiii
LIST OF FIGURES	xiv-xvii
LIST OF TABLES	xviii-xix
CHAPTER 1: INTRODUCTION AND REVIEW OF LITERATURE	1-45
1. Introduction	
1.1. Current Cancer Treatments	
1.2. Promising anticancer drug target: Microtubule Structure	
1.2.1. Microtubule structural organization and its dynamics	
1.2.2. Microtubule targeting drugs	
1.2.3. Drugs inhibiting polymerization of tubulin	
1.2.4. Drugs promoting polymerization of tubulin	
1.3. Taxol and its history	
1.3.1. Physicochemical properties of Taxol	
1.3.2. Source of Taxol	
1.3.3. Synthesis and semi-synthesis of Taxol	
1.4. Pharmacology of Taxol	
1.4.1. Mechanism of Action Taxol	
1.4.2. Structure Activity features of Taxol	
1.4.3. Resistance possessed by Taxol	
1.5. Preclinical and clinical information of Taxol	
1.5.1. Anti-tumor activity of Taxol	
1.5.2. Clinical trials of Taxol anti-cancer activity	
1.5.3. Animal Toxicity of Taxol	
1.5.4. Combination Chemotherapy of Taxol	
1.5.5. Human toxicity	
1.5.6. Metabolism of Taxol	
1.6. Pharmacokinetics of Taxol	
1.7. Pharmaceutics of Taxol	
1.7.1. Formulation of Taxol	

- 1.7.2. Stability of Taxol
- 1.7.3. Toxicity caused by Cremophor EL in the Taxol formulation
- 1.7.4. Cremophor EL and multi-drug resistance
- 1.8. Prodrugs of Paclitaxel
 - 1.8.1. Aqueous solubility: a major limitation of Paclitaxel
 - 1.8.2. Small molecule Paclitaxel prodrugs
- 1.9. Data mining techniques and QSPR models
- 1.10. Problem Identification
- References

CHAPTER 2: DEVELOPMENT OF NOVEL QSPR STRATEGY FOR SOLUBILITY PREDICTION

46-63

- 2.1. Introduction
- 2.2. Material and methods
 - 2.2.1. Solubility data collection
 - 2.2.2. Geometry optimization
 - 2.2.3. Structural descriptors by *DRAGON7*
 - 2.2.4. Structural descriptor reduction
- 2.3. Results and discussion
- 2.4. Conclusion
- References

CHAPTER 3: QSPR MODELS FOR THE SOLUBILITY PREDICTION OF PACLITAXEL PRODRUGS AND *IN-SILICO* MODELS FOR THE METABOLIC STUDIES OF HYDROPHILLIC SUBSTITUENT GROUPS

64-94

- 3.1. Introduction
- 3.2. Material and methods
 - 3.2.1. Paclitaxel prodrugs and hydrophilic substituent groups' datasets
 - 3.2.2. Solubility data
 - 3.2.3. Geometry optimization of the structures
 - 3.2.4. Dragon7 and SiRMS descriptor extraction and reduction
 - 3.2.5. Non-linear regression determination for descriptor selection
 - 3.2.6. Formation of QSPR models and their validation
 - 3.2.7. ADMET prediction of substituent groups
 - 3.2.8. CYP1A2 binding site
 - 3.2.9. Molecular docking
 - 3.2.10. Molecular dynamics simulation study of best docked complexes
 - 3.2.11. Principal Component Analysis
 - 3.2.12. Computing MM-PBSA binding-free energy

3.3. Results and discussion
3.4. Conclusion
References

CHAPTER 4: MOLECULAR DOCKING AND SIMULATION STUDIES OF PACLITAXEL PRODRUGS WITH CYTOCHROME 3A4 TO CORRELATE SOLUBILITY AND BIOAVAILABILITY THROUGH PHYSICOCHEMICAL CHARACTERIZATION 95-118

4.1. Introduction
4.2. Material and methods
 4.2.1. Protein & Paclitaxel phosphate prodrugs data collection and geometry optimization
 4.2.2. Interpretation of CYP3A4 binding site
 4.2.3. ADMET Prediction of Paclitaxel phosphate prodrugs
 4.2.4. Molecular Docking study
 4.2.5. Conformation Stability Analysis
 4.2.6. Principal Component Analysis
 4.2.7. Computing MM-PBSA Binding free energy
4.3. Preclinical and clinical information of Taxol
 4.3.1. Pharmacokinetics analysis
 4.3.2. Molecular docking studies
 4.3.3. Molecular Dynamics Simulation
 4.3.4. MM-PBSA binding free energy analysis
4.4. Conclusion
References

CHAPTER 5: CONCLUSION AND FUTURE DIRECTION 119-126

5.1. Conclusion
5.2. Future prospects

APPENDIX A 127-136

APPENDIX B 137-151

PUBLICATIONS 152-153

LIST OF ABBREVIATIONS

1D	1-dimensional
2D	2-dimensional
2-ME	2- Methoxyestradiol
3D	3-dimensional
AIC	Akiake Information Criteria
AM1	Austin method 1
AQUAFAC	AQUeous Functional group Activity Coefficient
ASR	Age Standardized Rate
AUC	Areas under concentration versus time curve
C _m	Maximum plasma concentrations
Cremophor EL	Polyethoxylated castor oil
CYP1A2	Cytochrome 1A2
CYP2C	Cytochrome 2C
CYP3A	Cytochrome 3A
CYPs	Cytochrome P450
DFT	density-functional theory
EPR	Retention effects
F	Fisher static
GDP	Guanosine Diphosphate
GIT	Gastrointestinal cancer
Globocan	Global Cancer Observatory
GSE	General solubility equation
GTP	Guanosine Triphosphate
H-bonds	Hydrogen bonds
ICSAS	Informatics and Computational Safety Analysis Staff
IFNc	Interferon c
IP	Intra-peritoneal
LINCS	Variance Of Wavelet Coefficients

log (ely)	Logarithm of solubility
$\log_{10}(S)$	Logarithm of solubility
$\log X_p$	Logarithm of positive descriptor value
MAP	Map kinase
MDR	Multi-drug resistance
MDS	Molecular Dynamics Simulation
MM-PBSA	Molecular Mechanic/Poisson-Boltzmann Surface Area
MT	Microtubule
NCI	National Cancer Institute
NDA	New Drug Application (NDA)
ns	Nano second
NSCLC	Non small cell lung cancer
PCA	Principal Component Analysis
PCs	Principal Components
PDB	Protein Data Bank
PHCHV	Positively High Correlated with High Variance
PHCLV	Positively High Correlated with Low Variance
PM6	Parametrization method 6
PMCHV	Positively Moderately Correlated with High Variance
PMCLV	Positively Moderately Correlated with Low Variance
PME	Particle Mesh Ewald
PO	Per os
ps	Pico second
PVHCHV	Positively Very High Correlated with High Variance
PVHCLV	Positively Very high Correlated with Low Variance
Q^2	Coefficient of correlation
QSAR	Quantitative Structure Activity Relationship
QSPR	Quantitative Structure Property Relationship
R^2	Coefficient of determination

REACH	Registration, Evaluation, And Authorization of Chemicals
R _g	Radius of gyration
RMSD	Root Mean Square Deviation
RMSE	Root mean square error
RMSF	Root Mean Square Fluctuation
SAR	Structure Activity Relationship
SASA	Solvent Accessible Surface Area
SC	Sub-cutaneous
SCLC	Small Cell Lung Cancer
SiRMS	Simplex Representation of Molecular Structure
SMILE	Simplified Molecular-Input Line Entry System
t _{1/2}	Half life
UV	Ultra-violet
V _{d_c}	Mean central
V _{d_{ss}}	steady state volumes
VIF	Variance Inflation Factor
VMD	Visual Molecular Dynamics
VSA	Van der Waals surface area
X _n	Value of negative descriptor
ZCHV	Zero Correlated with High Variance
ZCLV	Zero Correlated with Low Variance
β _n	Coefficient for negative value descriptor
β _p	Coefficient of positively valued descriptors
β _p	Coefficient for positive value descriptor

LIST OF FIGURES

Figure No.	Title
Figure 1.1	Distribution of cancer cases continental wise. (a) Visual depiction of cancer cases by Globocan 2019. (b) Data compilation from panel A. This data clearly indicates Asia as the leader with 48 percent of cancer cases. This data also indicates that cancer is not the disease of only rich but the prevalence of cancer is at global level
Figure 1.2	Estimated no. of cancer cases in both the sexes projected in the next twenty years
Figure 1.3	Estimated age-standardized incidence rates (world) in 2018, India, both sexes, all ages
Figure 1.4	Proportions of various mechanisms/drugs currently in use for cancer cure
Figure 1.5	Molecular structure of tubulin hetero-dimer comprising α - and β - subunit. Ribbon representation of X-ray crystallographic data (3.7 Å) for tubulin depicting α - and β - monomers. Alpha helix, beta strands, and loops are forming the tubulin structure. GTP is bind to α -subunit while the GDP and taxol are shown to be bind to β -subunit. The arrow is illustrating the advancement of the protofilaments in the course of microtubule formation
Figure 1.6	Lateral associations of 13 protofilaments forming the hollow rod-like structure of the microtubule which is a long run of head-to-tail association of tubulin dimer. Incurring head-to-tail association of $\alpha\beta$ subunits, protofilament structure has α subunit exposed at one end and the other end is open to β unit forming minus and plus ends respectively
Figure 1.7a	Structure of taxol showing configuration of different functional groups
Figure 1.7b	Structure of taxol core ring. It consists of ring A (cyclohexene), ring B (cyclooctane), and ring C (cyclohexane). 14 positions are ranging in order from B-C-A
Figure 1.8	Structure activity relationship of taxol
Figure 1.9	Bioactivation of paclitaxel prodrugs (<i>in-vivo</i>) by chemical or enzymatic modifications
Figure 1.10	Traditional and virtual screening of lead compounds
Figure 1.11	Schematic representation of QSAR/QSPR predictive modelling

Figure 2.1	Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility
Figure 2.2	Stage-wise workflow of descriptor reduction
Figure 2.3	Experimental solubility (l_y) vs predicted solubility (e_{ly}) of paclitaxel prodrugs. a. Performance of 2D Autocorrelation descriptors with AM1 geometry dataset; b. Performance of GETAWAY descriptors with AM1 geometry dataset; c. Performance of 2D Autocorrelation descriptors with PM6 geometry dataset; and d. Performance of WHIM descriptors with PM6 geometry dataset
Figure 2.4	Experimental solubility (l_y) vs predicted solubility (e_{ly}) of paclitaxel prodrugs. a. AM1 optimized geometry dataset with ten 2D Autocorrelation and GETAWAY descriptors; b. PM6 optimized geometry dataset with four 2D Autocorrelation and WHIM descriptors; c. AM1 optimized geometry dataset with 11 descriptors from a and b; d. PM6 optimized geometry dataset with 11 descriptors from a and b
Figure 3.1	Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility
Figure 3.2	Geometry optimization, descriptor extraction, selection, development and validation of QSPR model
Figure 3.3	Model optimization
Figure 3.4	Experimental solubility (l_y) vs predicted solubility ($logS$) of paclitaxel prodrugs. Performance of descriptors selected with AIC&VIF indicators for paclitaxel prodrugs with [a] AM1 geometry dataset; [b] with PM6 geometry dataset; [c] Performance of descriptor selected with AIC&VIF indicators for substructures with AM1 geometry dataset; and [d] for substructures with PM6 geometry dataset
Figure 3.5	Experimental solubility (l_y) vs predicted solubility ($logS$) of paclitaxel prodrugs. Performance of descriptors selected with MATLAB 'stepwise-fit' for paclitaxel prodrugs with [a] AM1 geometry dataset; [b] with PM6 geometry dataset; [c] Performance of descriptor selected with MATLAB 'stepwise-fit' for substructures with AM1 geometry dataset; and [d] for substructures with PM6 geometry dataset
Figure 3.6	Ligand Interaction diagram [a] Bisulphite 2' acryoltaxol with CYP1A2; [b] Fluorescent taxol with CYP1A2; [c] Paclitaxel with carbamate linkage to β -glucuronic acid with CYP1A2 [d] Sulphate of Paclitaxelwith CYP1A2

Figure 3.7	Molecular dynamics simulation. (a) RMSD of the C α backbone of substituent groups over the 50 ns MDS at 300 K, (b) RMSF of residues of substituent groups during MDS, In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-Paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2- Sulphate of Paclitaxel (magenta)
Figure 3.8	Molecular dynamics simulation. (a) Rg vs time of substituent groups during MDS. (b) Number of hydrogen bonds interaction between protein and ligand during simulation time scale for ligand complexes. In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-Paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2-Sulphate of Paclitaxel (magenta)
Figure 3.9	Principal component analysis. (a) Projection of the motion of the protein in phase space along the PC1 and PC2 (b) Plot of eigenvalues vs. eigenvector index. First 50 eigenvectors were considered. In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-Paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2- Sulphate of paclitaxel (magenta)
Figure 3.10	Solvent accessible surface area. (a) The SASA plot for CYP1A2 and substituent group complexes in water with respect to time. (b) The Residue SASA plot for CYP1A2 and substituent group complexes. In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acroyltaxol (red), CYP1A2-Flourescent taxol (blue), CYP1A2-Paclitaxel with carbamate linkage to β -glucuronic acid (dark cyan), CYP1A2- Sulphate of paclitaxel (magenta)
Figure 3.11	Gibbs free energy Landscape. The free energy landscape calculated for PC1 and PC2. In the figure, (a) CYP1A2, (b) CYP1A2-Bisulphite of 2' acroyltaxol, (c) CYP1A2-Flourescent taxol, (d) CYP1A2-Paclitaxel with carbamate linkage to β -glucuronic acid, (e) CYP1A2-Sulphate of Paclitaxel
Figure 4.1	Pharmacokinetic prediction, molecular docking and molecular dynamics simulation study of CYP3A4 with the paclitaxel phosphate prodrugs
Figure 4.2	Ligand interaction diagrams. [a] Paclitaxel with CYP3A4; [b] 2' Phosphonoxy methyl carbonate paclitaxel with CYP3A4; [c] 2' Phosphonoxy methyl ether derivative of paclitaxel with CYP3A4 [d] Isotaxel with CYP3A4

Figure 4.3	Molecular dynamics simulation. (a) RMSD of the C α backbone of paclitaxel prodrugs over the 1000 ns MDS at 300 K, (b) RMSF of residues of paclitaxel prodrugs during MDS. In all panels the color code is: CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), and CYP3A4-Isotaxel (magenta)
Figure 4.4	Molecular dynamics simulation. (a) Number of hydrogen bonds interaction between protein and ligand during simulation time scale for paclitaxel and other ligand complexes, (b) Rg vs time of paclitaxel prodrugs during MDS. In all panels the color code is: CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta)
Figure 4.5	Solvent accessible surface area. (a) The SASA plot for paclitaxel and all other complexes in water with respect to time, (b) The Residue SASA plot for paclitaxel and other ligand complexes. In all panels, the color code is CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta)
Figure 4.6	Principal component analysis. (a) The plot of eigenvalues vs. eigenvector index. First 50 eigenvectors were considered (B) Projection of the motion of the protein in phase space along the PC1 and PC2, In all panels the color code is: CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta)
Figure 4.7	Gibbs free energy Landscape. The free energy landscape calculated for PC1 and PC2. In the figure, (a) CYP3A4 (black) and the ligands (b) CYP3A4-Paclitaxel (red), (c) CYP3A4-2'-Phosphonoxy methyl carbonate (yellow), (d) CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), (e) CYP3A4-Isotaxel (magenta)
Figure 5.1	Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility
Figure 5.2	QSPR Pipeline

LIST OF TABLES

Table No.	Title
Table 1.1	Currently used drugs for cancer treatment
Table 1.2	Tubulin polymerization promoters in clinical development
Table 1.3	Different tumors and their response rates against taxol
Table 2.1	Regression (R^2) and 4-fold cross validation correlation (Q^2) coefficients of the QSPR model developed using dragon molecular descriptor groups
Table 2.2	Regression (R^2) and 4-fold cross validation correlation (Q^2) coefficients of the QSPR model determined through combination of descriptors
Table 2.3	Regression (R^2) and validation correlation (Q^2) coefficients of the QSPR model for <i>Huuskonen</i> dataset
Table 2.4	Comparison among the results obtained from literature methods and the model formed in the study
Table 3.1	Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models for paclitaxel prodrugs and substructures with PM6 and AM1 optimized geometry dataset
Table 3.2	Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models formed from quasi-mixture descriptors for the substructures with PM6 and AM1 optimized geometry dataset
Table 3.3	Docking results of the selected substituent groups with PM6 optimized geometry against CYP1A2 showing the binding affinity and the interacting residues
Table 3.4	Table represents the Van der Waals, electrostatic, polar solvation, SASA and binding energy in $\text{kJ}\cdot\text{mol}^{-1}$ for substituent groups
Table 4.1	Docking study result of the selected paclitaxel prodrugs with PM6 optimized geometry against CYP3A4 showing the binding affinity and the interacting residues
Table 4.2	Table represents the Van der Waals, electrostatic, polar solvation, SASA and binding energy in $\text{kJ}\cdot\text{mol}^{-1}$ for control compound and prodrugs
Table 5.1	Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of

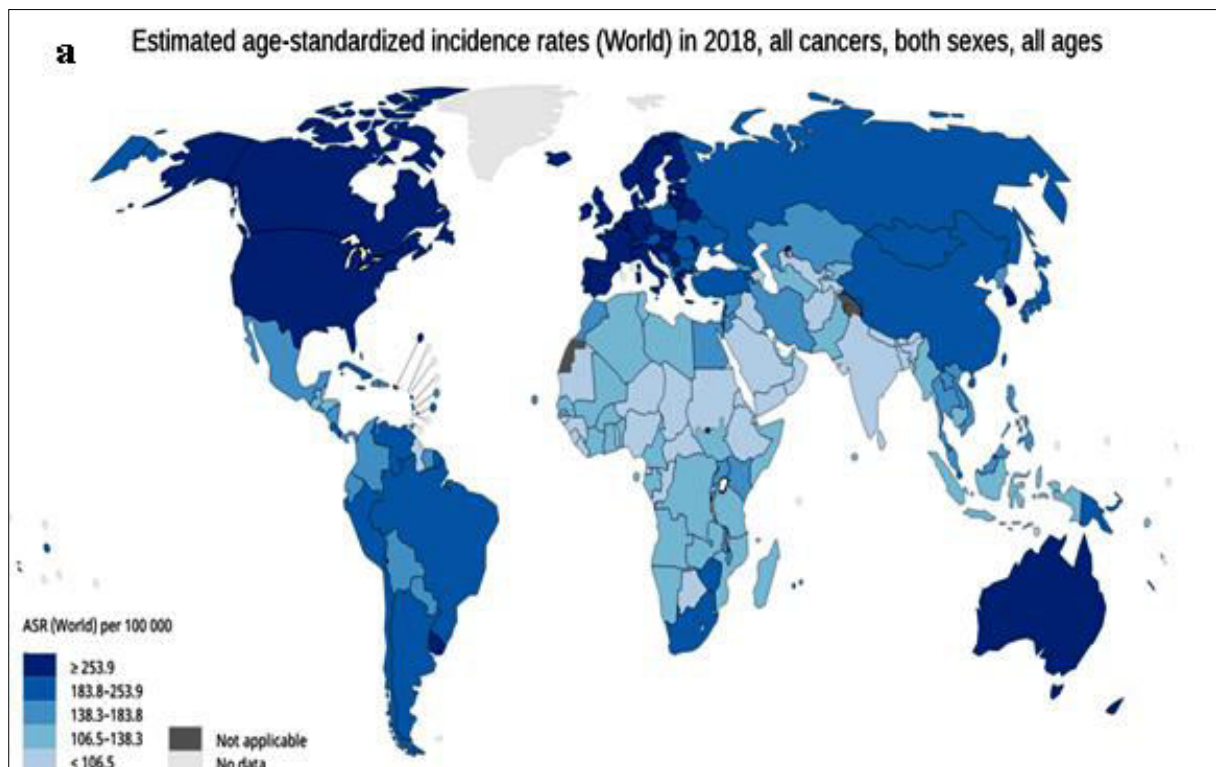
	the QSPR models for paclitaxel prodrugs and substructures with PM6 and AM1 optimized geometry dataset
--	---

CHAPTER 1

INTRODUCTION AND REVIEW OF LITERATURE

1 INTRODUCTION

Cancer is one of the most deadly diseases prevailing now-a-days. Cancer comprises a class of hundred different diseases. One of the common characteristic that all cancer cells share is the abnormal growth. Cancer cells shows uncontrolled growth which have the potential to intrude other body parts, destroy adjacent tissues, and can spread to a different body part. This disease may affect the people of all the ages but risk for all the ages for various cancers increases throughout the world (Figure 1.1). From the very beginning the disease had taken the centre stage of healthcare providers but still we are in search of cure for this devastating disease. Although cure for the disease is still not in our hands but many advances has been achieved for restraining adversity and progression of the disease.



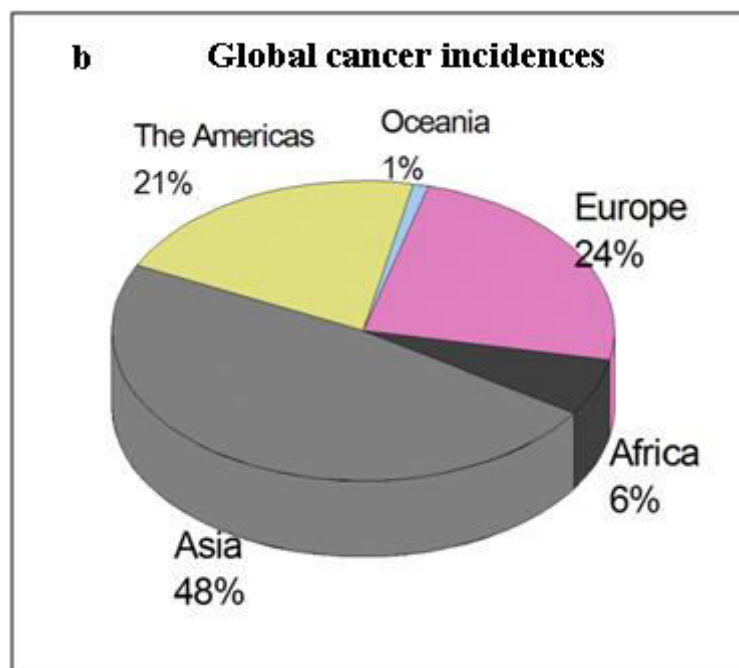


Figure 1.1: Distribution of cancer cases continental wise. (a) Visual depiction of cancer cases by Globocan 2019. (b) Data compilation from panel A. This data clearly indicates Asia as the leader with 48 percent of cancer cases. This data also indicates that cancer is not the disease of only rich but the prevalence of cancer is at global level; Data source: Globocan 2018; <http://gco.iarc.fr>.

As reported in September 2018 by Global Cancer Observatory (Globocan), International agency for research on cancer of World Health Organization (WHO), globally there is an increase from 14.1 million to 18.1 million new cancer cases and 8.2 million to 9.6 million cancer deaths from year 2012 to 2018 [1]. According to the Globocan there has been prevalence of 35 types of cancer in 185 countries. Globocan 2018 report provided information about the diagnosis of the most common cancers worldwide; including lung cancer (2.09 million, 11.58% of the total), breast cancer (2.08 million, 11.55% of the total), colorectal cancer (1.84 million, 10.23% of the total). The most common cause of cancer deaths were lung cancer (1.76 million, 18.4% of total cancer deaths), followed by stomach cancer (0.782 million, 8.2% of total cancer deaths), and liver cancer (0.781 million, 8.2% of total cancer deaths) [2]. Also, cardiovascular diseases contribute to the world mortality rate the most followed by cancer [3]. In relation to India, cancer is the second most leading cause of mortality with about 0.78 million deaths per year after cardiovascular diseases [4]. According to the cancer data compiled by Globocan in 2012, the number of total cancer patients, male patients, and female cancer patients were 9,79,786, 4,62,408 and 5,17,378 respectively [5]. Similarly, in 2018 the total number of cancer patients were 11,57,294 with 5,70,045 and 5,87,249 males and females respectively [6]. These figures clearly indicated that the cancer

cases are being increased gradually (Figure 1.2). All cancer types including lung, breast, rectum, skin, stomach, liver, oesophagus, prostate, bladder etc. were reported in indian population. Male population in india was found to be mainly suffering from lip & oral cavity, lung, stomach, colorectal, prostate and oesophagus cancers whereas in women most frequent cancers were breast, cervix uteri, ovary, lip & oral cavity, and colorectal (Figure 1.3) [7]. The age group of incidences of breast cancer among indian women was established less than a decade as that in the USA.

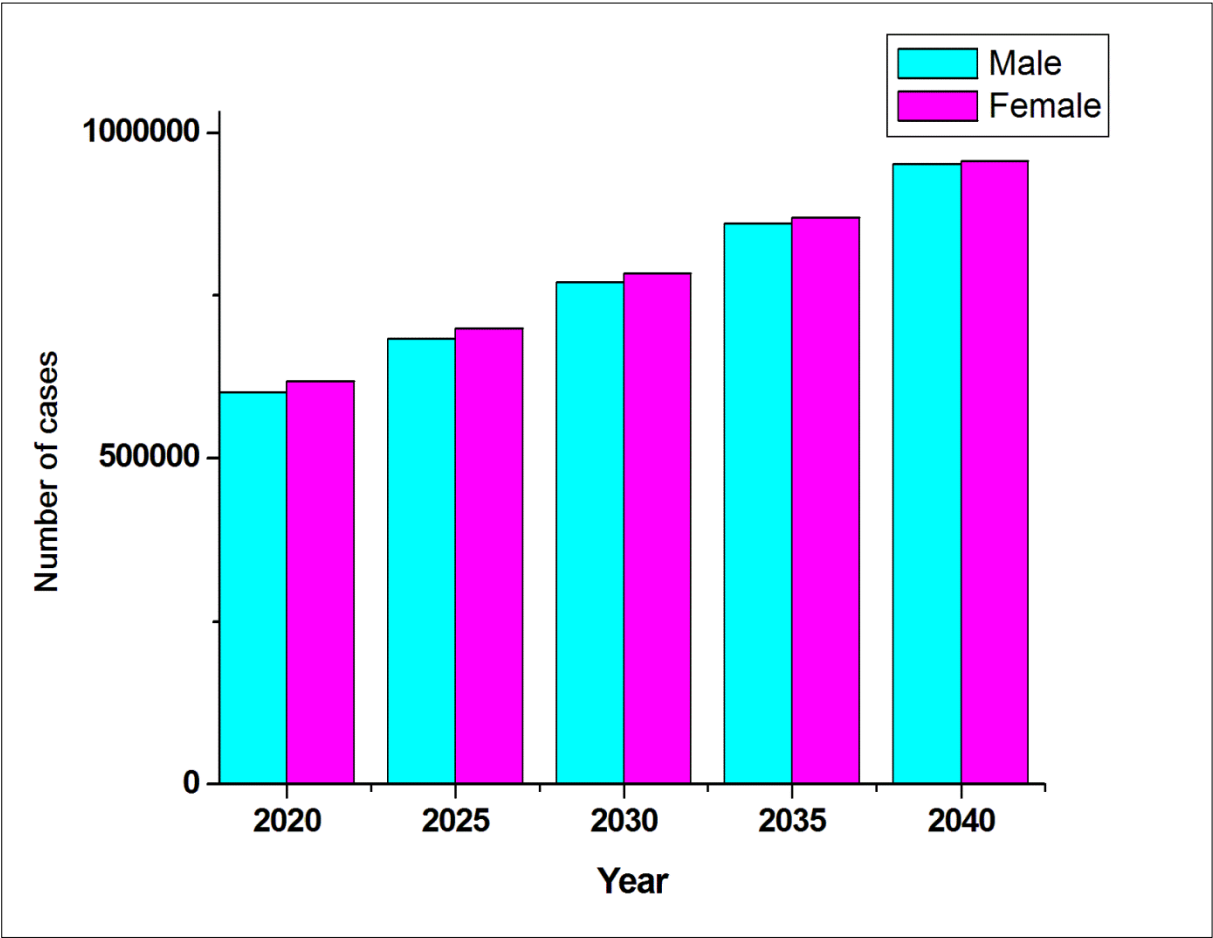


Figure 1.2: Estimated no. of cancer cases in both the sexes projected in the next twenty years. (Data source: Globocan 2018; <http://gco.iarc.fr>).

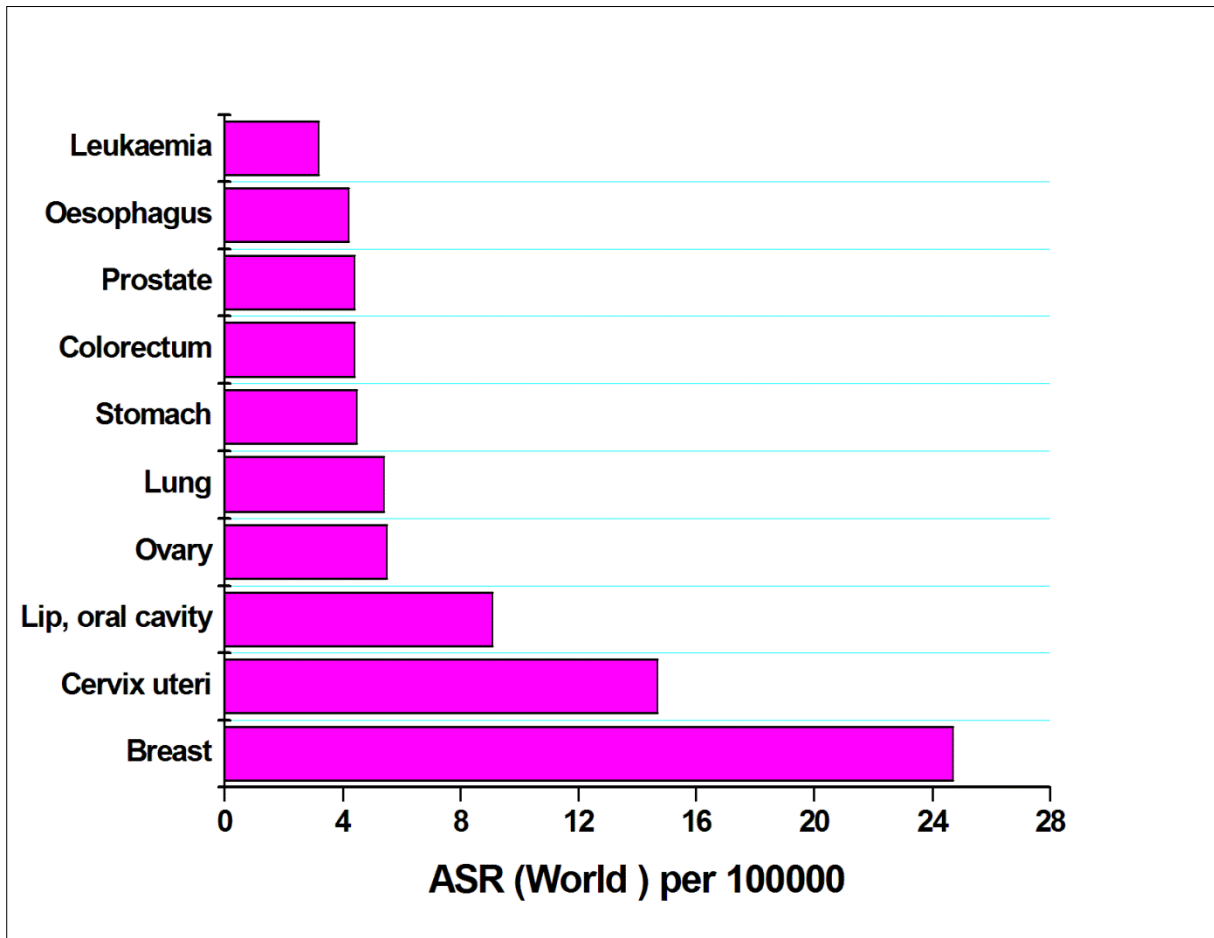


Figure 1.3: Estimated age-standardized incidence rates (World) in 2018, India, both sexes, all ages. (Data source: Globocan 2018; <http://gco.iarc.fr>).

1.1 Current cancer treatments

From decades various cancer treatments are being employed to curb cancer. Due to ample work in the cancer research field we had gained phenomenal knowledge about the cancer cells growth, division, and advancements. This knowledge provided scientists insight to improve or apply new ways to stop or reverse the growth of the cancerous cells. Till now the most common types of cancer treatment are: Surgery, Chemotherapy, Radiation therapy and some other targeted therapies. Among all the treatment regimens, chemotherapy is most commonly used for treatment, prevention of reoccurrence and palliative care for a variety of cancer types. The relative contribution of various cancer target curative regimens is shown in the Figure 1.4.

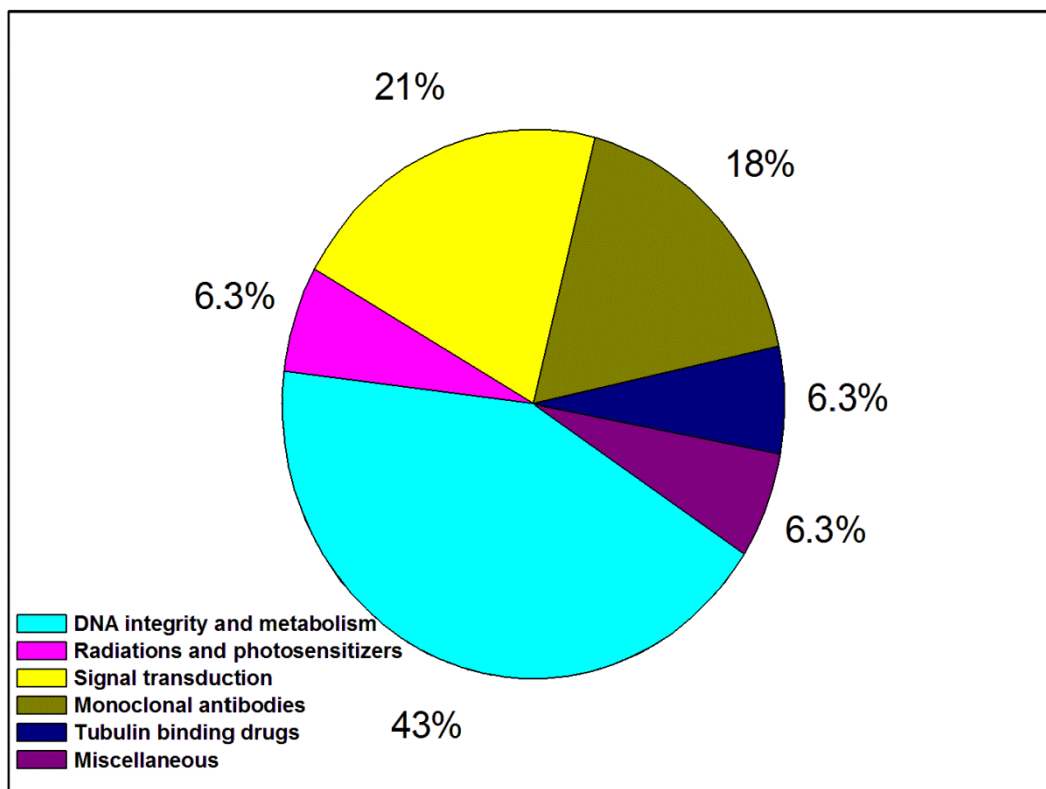


Figure 1.4: Proportions of various mechanisms/drugs currently in use for cancer cure.

Chemotherapy is the term used for the treatment of cancer with any anticancer drug. These drugs play their role by any of the mechanisms like interrupting with cell division, DNA metabolism and signal transduction. As these drugs are administered systemically and having generalized bioavailability, so they can act on the cancer cells from distant sites other than the site of origin for cancer. Chemotherapy is a universal way for treating cancer; control the disease from spreading, and providing the palliative care for various cancers. Therefore, notably more than half of the patients diagnosed with cancer got chemotherapy regimens. Traditionally, these agents were determined from the study of naturally occurring compounds or of chemically synthesized compounds after assessing their cytotoxicities so as to select new promising lead compounds. However, technical advances in computational methods, simple yet very sensitive assays along with the advancement in combinatorial chemistry artistry has now made the backbone of this research field. Likewise, advances in prodrug design techniques, and advances in the drug delivery systems have opened ingenious view in this research field. Essentially, the compounds that are modestly selected for cytotoxicities against cancer cells not only target cancer cells exclusively but are also having equally adverse effects on the healthy cells. Despite having high cytotoxicities, many of the drugs still are most effective across the cancerous cells of ovary and breast cancers as well as for some of the haematological malignancies, Table 1.1 [8-11].

Table 1.1 Currently used drugs for cancer treatment

Mode of Inhibition	Drugs and their targets
<p>DNA integrity and metabolism</p>	<p>Folic acid metabolism: Dihydrofolate reductase inhibitor (Aminopterate, Methotrexate, Pemetrexed); Thymidylate synthase inhibitor (Raltitrexed, Pemetrexed).</p> <p>Purine metabolism: Adenosine deaminase inhibitor (Pentostatin); Halogenated/ribonucleotide reductase inhibitors (Cladribine, Clofarabine, Fludarabine); Thiopurine (Thioguanine, Mercaptopurine).</p> <p>Pyrimidine metabolism: Thymidylate synthase inhibitor (Fluorouracil, Capecitabine, Tegafur, Carmofur, Floxuridine), DNA polymerase inhibitor (Cytarabine); Ribonucleotide reductase inhibitor (Gemcitabine); Hypomethylating inhibitor (Azacitidine, Decitabine).</p> <p>Deoxuribonucleotides metabolism: Ribonucleotide reductase inhibitor (Hydroxyurea).</p> <p>Topoisomerase I inhibitor: Camptotheca (Camptothecin, Topotecan, Irinotecan, Rubitecan, Belotecan).</p> <p>Topoisomerase II inhibitor: Podophyllum (Etoposide, Teniposide).</p> <p>Topoisomerase II + intercalation: Anthracyclines (Aclarubicin, Daunorubicin, Doxorubicin, Epirubicin, Idrubicin, Amrubicin, Pirarubicin, Valrubicin, Zorubicin); Anthracenediones (Mitoxantrone, Pixantrone).</p> <p>Cross linkers: Nitrogen mustards: Mechlorethamine; Cyclophosphamide (Ifosfamide, Trofosfamide); Chlorambucil (Melphalan, Prednimustine); Bendamustine; Uramustine; Estramustine.</p> <p>Nitrosoureas: Carmustine; Lomustine (Semustine); Fotemustine; Nimustine; Ranimustine; Streptozocin.</p> <p>Alkyl sulfonates: Busulfan (Mannosulfan, Treosulfan).</p> <p>Aziridines: Carboquone; ThioTEPA; Triaziquone; Triethylenemelamine, Platinum (Carboplatin, Cisplatin, Nedaplatin, Oxaliplatin, Triplatin, Tetranitrate, Satraplatin); Hydrazines (Procarbazine); Triazines (Dacarbazine, Temozolomide); Altretamine; Mitobronitol; Streptomyces (Actinomycin, Bleomycin, Mitomycin, Plicamycin).</p>
<p>Radiation and photosensitizers</p>	<p>Radiationtherapy: High-energy radiation from x-rays, gamma rays, neutrons, protons and other sources.</p> <p>Photosensitizers: Aminolevulinic acid/Methylaminolevulinate; Etoposide; Porphyrin derivatives (Porfimer sodium, Talaporfin, Temoporfin, Verteporfin).</p>
	<p>Receptor tyrosine kinase inhibitor: ErBb: HER1/EGFR (Erlotinib,</p>

<p style="text-align: center;">Signal Transduction</p>	<p>Gefitinib, Lapatinib, Vandetanib, Neratinib), HER2/neu (Lapatinib, Neratinib)</p> <p><i>RTK class III:</i> C-kit (Axitinib, Sunitinib, Sorafenib); FLT3 (Lestaurtinib); PDGFR (Axitinib, Sunitinib, Sorafenib); VEGFR (Vandetanib, Semaxanib, Cediranib, Axitinib, Sunitinib, Sorafenib).</p> <p><i>Non-receptor tyrosine kinase inhibitor:</i> bcr-abl (Imatinib, Nilotinib, Dasatinib); Src (Bosutinib); Janus kinase 2 (Lestaurtinib).</p> <p><i>Enzyme inhibitors:</i> Farnesyl transferase FI (Tipifarnib); CDK inhibitors (Alvocidib, Seliciclib); Proteome inhibitor PrI (Bortezomib); PDE II inhibitor PhI (Anagrelide); Imp dehydrogenase inhibitor IMPDI (Tiazofurine); Lipooxygenase inhibitor L1 (Masoprocol).</p> <p><i>Receptor antagonists/hormones:</i> Endothelial receptor antagonist ERA (Atrasentan); Retinoid X receptor (Bexarotene); Sex steroid (Testolactone).</p>
<p style="text-align: center;">Monoclonal Antibodies</p>	<p><i>Receptor tyrosine kinase:</i> ErBb: HER1/EGFR (Cetuximab, Panitumumab); HER2/neu (Transtuzumab).</p> <p><i>Solid tumors:</i> EpCAM (Catumaxomab, Edrecolomab); VEGF-A (Bevacizumab).</p> <p><i>Lymphoid:</i> CD20 (Rituximab, Tositumomab, Ibritumomab).</p> <p><i>Myeloid:</i> CD52 (Alemtuzumab).</p> <p><i>Lymphoid + Myeloid:</i> CD33 (Gemtuzumab).</p> <p><i>Others:</i> Afutuzumab, Alemtuzumab, Bevacizumab/Ranibizumab, Bivatuzumab mertansine, Cantuzumab mertansine, Citatuzumab bogatoz, Dacetuzumab, Etaracizumab, Farletuzumab, Gemtuzumab, Ozogamicin, Inotuzumab ozogamicin, Labetuzumab, Lintuzumab, Milatuzumab, Nimotuzumab, Oportuzumab monatox, Pertuzumab, Sibrotuzumab, Sontuzumab, Tacatuzumab tetraxentan, Tigatuzumab.</p>
<p style="text-align: center;">Tubulin binding drugs</p>	<p><i>Inhibit microtubule assembly:</i> Vinca alkaloids (Vinblastine, Vincristine, Vinflunine, Vindesine, Vinorelbine).</p> <p><i>Promote microtubule assembly:</i> Taxanes (Pacxlitaxel. Docetaxel, Larotaxel, Ortataxel, Tesetaxel); Epothilones (Ixabepilone).</p>
<p style="text-align: center;">Miscellaneous</p>	<p>Amsacrine; Trabectedine; Retinoids (Aletretinoin, Tretinoin); Arsenic trioxide; Asparagine depleter (Asparaginase, Pegapargase); Celecoxib; Demecolcine; Elesclomol; Elsamitrucin; Etoglucid; Lonidamine; Lucanthone; Mitoguazone; Mitotane; Oblimersen; Temsirolimus; Vorinostat.</p>

1.2 PROMISING ANTICANCER DRUG TARGET: MICROTUBULE STRUCTURE

1.2.1 Microtubule structural organization and its dynamics

Microtubules are prevalent tubular structures that form the cytoskeleton of the cell. Microtubule structures are responsible for the formation, and preservation of mitotic spindle that play major role in the separation of duplicated chromosomes in the course of cell division [12-14]. MTs are hollow rod shaped structures having a diameter of 25 nm, made of tubulin subunit which is a dimer of α and β subunits that shows enduring disassembly, and rearrangement in the cell. Electron crystallography studies had provided the three dimensional structure of the microtubule (Figure 1.5).

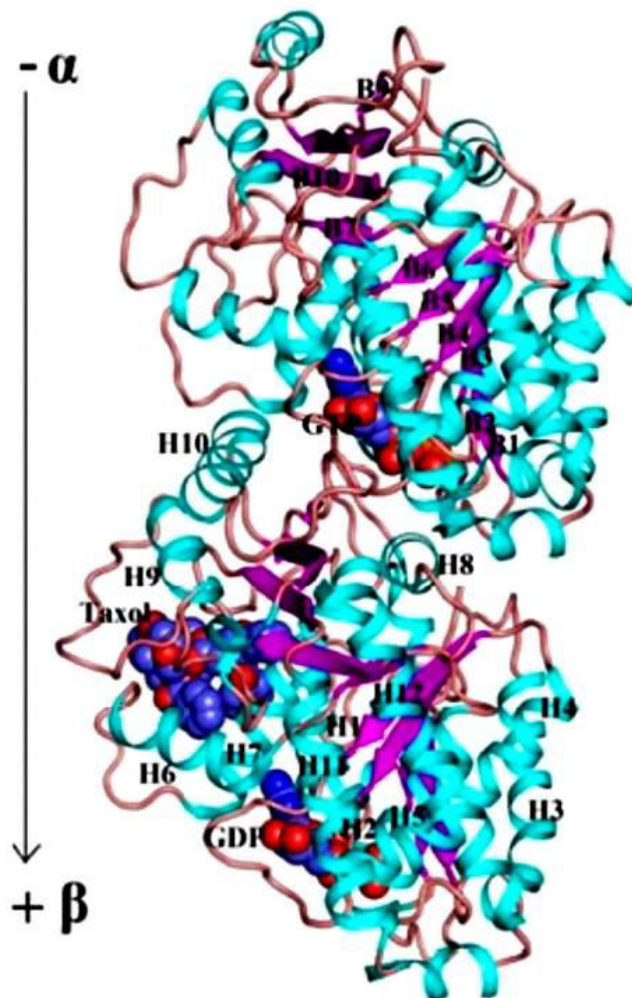


Figure 1.5: Molecular structure of tubulin hetero-dimer comprising α - and β - subunit. Ribbon representation of X-ray crystallographic data, PDB id 1TUB (3.7 Å) for tubulin depicting α - and β - monomers. Alpha helix, beta strands and loops are forming the tubulin structure. GTP is bind to α -subunit while the GDP and taxol are shown to be bind to β -subunit. The arrow is illustrating the advancement of the protofilaments in the course of microtubule formation.

Each monomer subunit is asymmetric with nearly width, height, depth of 46 x 40 x 65 Å respectively [15]. Every tubulin is segregated into three functional domains: amino terminal,

intermediate terminal and carboxy terminal. Amino terminal domain is having the nucleotide binding region and the carboxy terminal domain controls the binding of the drugs such as vincristine, paclitaxel [16-18]. Each MT consists of 13 protofilaments associated laterally, forming hollow rod like structure and consequently forming head-to-tail body of the tubulin dimers (Figure 1.6). Incurring head-to-tail association of $\alpha\beta$ subunits, protofilament structure has α subunit exposed at one end and the other end is open to β unit forming minus and plus ends respectively (Figure 1.6).

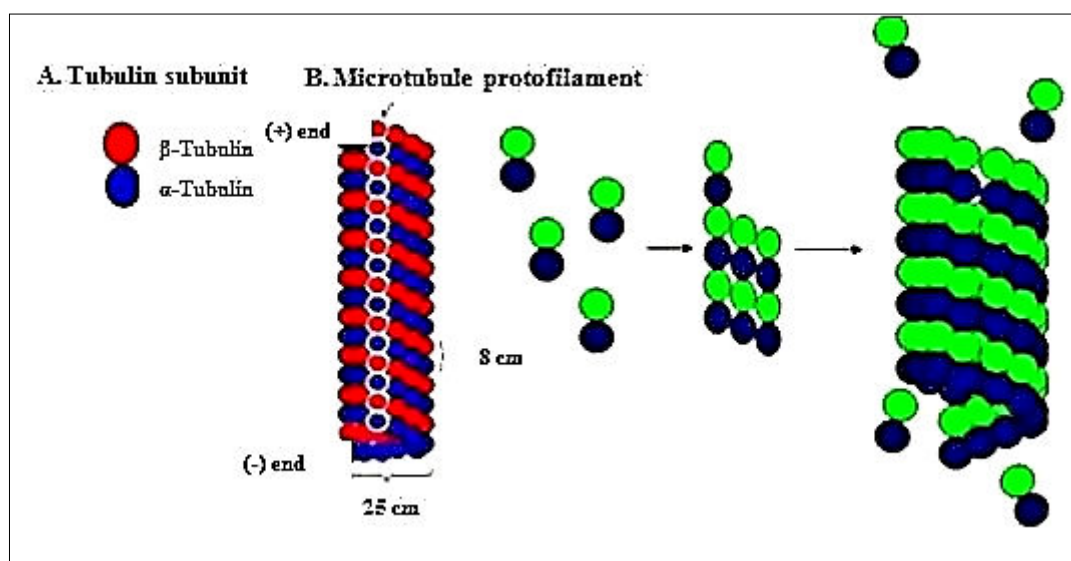


Figure 1.6: Lateral associations of 13 protofilaments forming the hollow rod- like structure of the microtubule which is a long run of head-to-tail association of tubulin dimer. Incurring head-to-tail association of $\alpha\beta$ subunits, protofilament structure has α subunit exposed at one end and the other end is open to β unit forming minus and plus ends respectively.

1.2.2 Microtubule targeting drugs

Microtubule targeting drugs are broadly divided into two groups. One group consists of drugs inhibiting tubulin polymerization and stimulating MT disassembly. Some of the important drugs of this category are vinca alkaloids, combretastatin and colchicine. Taxanes, discodermolides, and laulimalides represent the second class of microtubule targeting drugs that causes over assembly of MT by promoting tubulin polymerization. In spite of current progress in the study of binding interactions of laulimalides, paclitaxel, and colchicine with MTs, the specific nature of site-directed binding interactions of anti-MT drugs has been an area of profound research. In 1998, Downing and co-workers reported the three dimensional electron crystallographic structure of heterodimer $\alpha\beta$ tubulin at 3.7 Å resolutions. Several binding site illustrating techniques had established three main binding sites on tubulin: (a) “colchicine” binding site [19] (b) “vinca alkaloid” binding site [20] and (c) the “taxane”

binding site on the MT [21]. The cell-cycle got arrested at the prometaphase of mitosis by these compounds. These drugs showed significant effects on MT dynamics also. They alter the growth rate and duration or shorten the rate and duration. Also, they alter the transition frequency between the growth and shortening phases and also the pause state duration when no growth or shortening is evidenced. Some of the reported drugs are explained in the next section.

1.2.3 Drugs inhibiting polymerization of tubulin

Drugs for vinca alkaloid-binding site: Vinca-alkaloids drugs are derived from naturally occurring periwinkle plant, *Catharanthus roseus*. These were the first class effective anti-cancer drugs those bind to the microtubules. A few of the vinca-alkaloids, vinorelbine, vincristine, vinflunine and vindesine are also used for the treatment of Hodgkin's and non-Hodgkin's lymphomas. They bind to β -tubulin, obstruct MT polymerization and stop the mitosis at prometaphase, [22-24] finally causes cell death [24]. Cryptophycins a recently identified group of compounds that might bind at "vinca" binding sites and acts against tumor cells at a low concentrations than other anticancer drugs [25].

Drugs for colchicine-binding site: A toxic, naturally occurring, secondary metabolite *colchicine* is extracted from *Colchicum autumnale* plant. It binds to $\alpha\beta$ -tubulin at a different site than "vinca-alkaloid binding site". Some of the colchicine-binding site drugs such as indanocine, 2-aryolindoles and methoxyestradiol (2-ME) binds tubulin, break mitotic spindle and arrest the cell cycle in mitosis [26].

1.2.4 Drugs promoting polymerization of tubulin

Drugs for Taxane-binding site: Paclitaxel (Taxol), a forbearer member of this family, a natural product originated from the bark of *Taxus brevifolia* (pacific yew tree) [27]. It acts by binding to MT and stabilizing the depolymerisation of MTs thus interrupting their dynamic stability resulting in cell death [28,29]. Compounds of this family; paclitaxel and docetaxel (taxotere) are very well studied. Both the compounds bind to the β -tubulin within the lumen of MT thus stabilizing the inter protofilament 3 and blocks disassembly of $\alpha\beta$ tubulin [30-34]. Taxanes stabilizes the existing MT, additionally it also nucleate newer assembly of other polymers. Consequently, the vigorous rearrangement of bipolar spindle arrays is not attained and the cell cycle is arrested in the pro-metaphase of mitosis leading to cell death [35]. The members of taxane family shows anticancer activity for a range of solid tumor growth such as ovarian cancer, breast cancer, non small cell lung cancer (NSCLC), gastro oesophageal cancer and also for head and neck cancer [30,36,37]. In the related manner, the disorder caused by

taxanes and also by many other anti-MT drugs disturbs the cell's haemostasis which in turn is reported to affect the genes encoding inflammatory mediators like interleukins, tumor necrosis factor alpha and nitric-oxide synthase, and also cyclooxygenase 2 enzymes [38-41]. Epothilones forms a newer class of drugs in the taxane family. Being a macrolide inhibitor this class represent novel anti-MT drugs. Epothilones also bind to MT in the same manner and in the same binding site as taxanes [42]. The progress gained by taxanes, epothilones, and other anti-MT drugs in curbing cancer cells had instilled the exploration of new anti-MT drugs [43,44]. Such as, the useful anticancer drugs discodermolide and dicoumarol create an enhanced effect by working synergistically with the paclitaxel [45,46]. Also, naturally occurring compounds laulimalide and peloruside A had open ring structure just like epothilone and they also stabilize MT, and arrest the cell division [47,48]. The present developmental status of anti-cancer drugs of this class is shown in Table 1.2.

Table 1.2 Tubulin polymerization promoters in clinical development

Drug	Highest Phase	Indication	Drug type
AI 850	Phase-I	Solid tumors	Paclitaxel, in a novel polyethoxylated castor oil-free hydrophobic microparticle delivery system
ANG1005	Phase-I	Brain cancer (Metastatic disease), Glioma cancer	Taxane derivative
ARC 100 (TPI-287)	Phase-II	Pancreatic cancer, Prostate cancer (Hormone refractory)	3 rd generation cancer to overcome drug resistance
BMS 188797	Phase-I	Cancer	Taxane
BMS 275183	Phase-II	Non-small cell lung cancer, Prostate cancer, Solid tumors	Taxane
BMS 310705	Phase-I	Cancer	Epothilone analog
BMS 753493	Phase-I/II	Solid tumors	Folate receptor targeted epothilone
Cabazitaxel	Phase-III	Breast cancer, Prostate cancer	Taxane

		(Hormone refractory)	
Docetaxel	Launched	Cancer	Taxane
Docetaxel emulsion	Phase-I	Breast cancer	Taxane
DTS 301	Phase-II	Glioblastoma, Oesophageal cancer, Pancreatic cancer	Paclitaxel delivered in copolymer gel ReGel
EndoTAG I	Phase-II	Cancer	Positively charged lipid complex to transport
Ixabepilone	Launched	Cancer	Epothilone
KOS 1584	Phase-II	Non-small cell lung cancer, Solid tumors	Epothilone
Larotaxel	Phase-III	Cancer	Taxane
Liposome encapsulated docetaxel	Phase-I	Solid tumors	Taxane
Milataxel	Phase-II	Cancer	Taxane analog
NK 105	Phase-I	Solid tumors	Paclitaxel-incorporating polymeric micellar nanoparticle (85 nm in size)
OAS PAC 100	Phase-III	Ovarian cancer (Combination therapy: in combination with carboplatin)	Paclitaxel micellar
Orataxel	Phase-III	Non-hodgkin's lymphoma, Renal cancer, Solid tumors	Taxane
Paclitaxel – Bristol Myers Squibb	Launched	Cancer	Paclitaxel
Paclitaxel- Yew tree pharmaceutica	Launched	Breast cancer, Ovarian cancer	Taxane

I			
Paclitaxel-angiotech	Launched	Inflammation and cancer	Taxane
Paclitaxel-Aphios	Preclinical	Cancer	Taxane
Paclitaxel-Hanmi	Phase-I	Cancer	Taxane
Paclitaxel-SuperGen	Preclinical	Solid tumors	Taxane
Paclitaxel nanoparticle-Dabur Pharma	Launched	Cancer	Taxane
Paclitaxel nanoparticles-BioAlliance pharma	Precilinal	Solid tumors	Taxane
Albumin bind Paclitaxel	Launched	Cancer	Taxane
Patupilone	Phase-III	Cancer	Epothilone B
PEGylated Docetaxel	Phase-I	Solid tumors	Taxane
Sagopilone	Phase-II	Cancer	Epothilone
Simotaxel	Phase-I	Solid tumors	Taxane
Tesetaxel	Phase-II	Cancer	Taxane
TL-310	Phase-I	Cancer	Taxane Analog

1.3 TAXOL AND ITS HISTORY

Taxol a natural product and a precursor of taxane, is showing anti-cancer activity [49]. Even though, history shows the scientific traces of taxol thirty eight years ago, it had been used in the traditional medicine in india as a clarified butter preparation having European yew (*Taxus baccata*) [50]. In 1962, taxol was selected as one of the naturally occurring agent for the cytotoxicity screening program led by National Cancer Institute of the United States (NCI) [51]. For the first time it was isolated from the bark of western yew (*Taxus brevifolia*) in 1967. Wani *et al*, in 1971 reported specific structure and cytotoxic nature of taxol; in late 1970s it gained the interest of researchers [52]. It poses a unique mechanism of action which was founded in late 1970s and found to be different amid other plant alkaloids and anti-cancer

drugs [53]. Preclinical studies approved taxol's anti-cancer activity against various solid tumors, accomplished in early 1980's [54]. Despite having the encouraging anti-cancer activity ascertained by clinical trials there was delay in the development of this drug because of its low aqueous solubility and hypersensitivity [55]. Thereafter, these problems were overcome by the formulation of taxol in cremophor EL and also directing anti-hypersensitivity intermediate before each taxol dose.

In the mid 1980's Phase-II clinical trials of taxol were started as taxol a single agent in the patients those responded well in Phase-I. All the patients in these trials were pre-treated with anti-hypersensitivity agents. Some tumors profoundly responded to taxol that they were showing the objective response rates even higher than 50 percent as in breast cancer, 30 percent or higher in ovarian cancer and more than 20 percent in NSCLC. The response rates achieved were comparable or better than the standard chemotherapy [56]. Since then in 1992 taxol got the approval for the treatment of advanced ovarian cancer in US. Clinical results from Phase-II trials proposed taxol as potential anti-cancer drug. Establishing taxol efficacy for treating ovarian cancer, NCI along with Bristol-Myers Squibb pharmaceuticals approved taxol for New Drug Application (NDA). Taxol became available in market under the name paclitaxel in a 5 ml vials having 30 mg taxol in cremophor EL as a single dose [57]. Ongoing trials were devised for exploring combination of taxol with other anticancer drugs, overcoming toxicities, pharmacokinetics, and pharmacodynamics asserting its formulation and administration route.

1.3.1 Physicochemical properties of Taxol

Taxol compound, with molecular formula ($C_{46}H_{49}O_{14}N$) having the chemical structure as shown in Figure 1.7a. Chemically taxol was discovered as a complex diterpenoid having a taxane nucleus and a phenylisoserine side chain at C-13 of its molecule [58]. Taxol was also found to share structural similarities with other alkaloids such as vincristine and vinblastine but the nucleus of taxol was found to be different as shown in Figure 1.7b [59]. Structure of taxane nucleus was bring into being from four rings A, B, C, D. Ring A is cyclohexene, ring B is cyclooctane, ring C is cyclohexane and ring D is oxetane ring, Figure 1.7b. Physicochemical structure of taxol was divided in two segments; a polar segment and a non-polar segment. These two segments were found to be quite important in the interaction of taxol with its receptor, Figure 1.7a.

The commercial availability of taxol was formed as a white, amorphous, odourless, hygroscopic powder with a molecular weight of 859.3 g/mol. It was discovered to be highly

insoluble in water and showed a maximum UV absorption at 227 nm. The melting point was observed to be 205-208°C and octanol/buffer partitioning was found to be more than 99 [60].

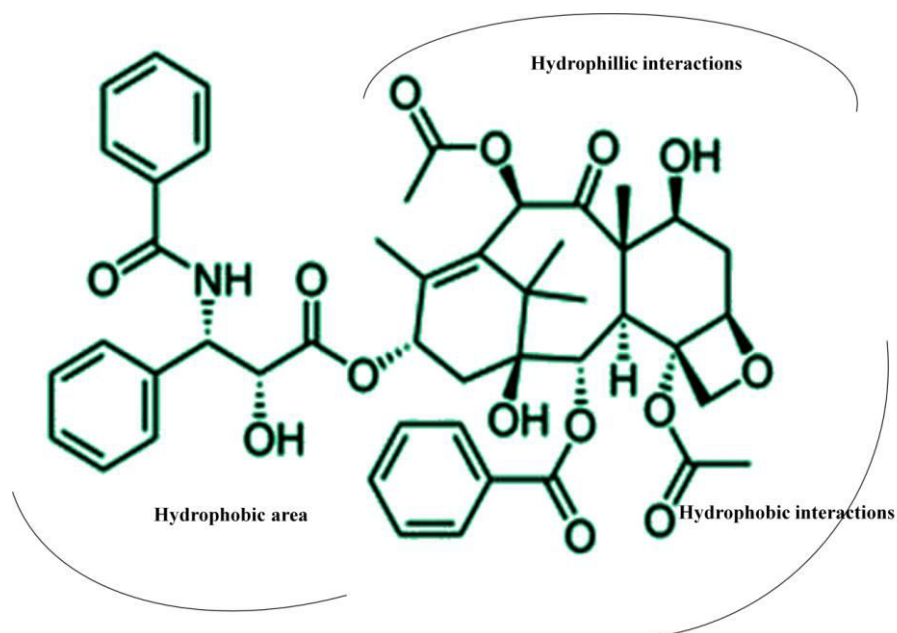


Figure 1.7a: Structure of taxol showing configuration of different functional groups.

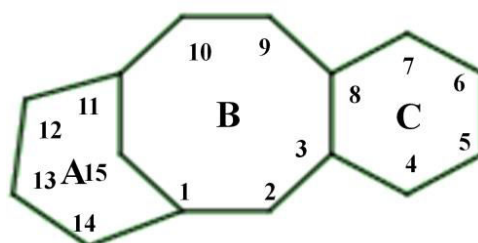


Figure 1.7b: Structure of taxol core ring. It consists of ring A (cyclohexene), ring B (cyclooctane), ring C (cyclohexane). 14 positions are ranging in order from B-C-A.

1.3.2 Source of taxol

In 1967, taxol was first isolated from the bark of *Taxus brevifolia*, commonly known as Pacific yew or western yew. Pacific yew is a shrub of 20-40 feet height, 12-15 inches diameter and having life span of 250-300 years. About 0.00002- 0.057% of taxol amount is present in different part of *Taxus brevifolia* but the maximum concentration of taxol is found in the dried bark (0.01-0.057%). The production of taxol from the tree is so low that minimum three trees of 100 years of lifespan are needed for the treatment of one cancer patient. Research groups across the world are working on production of taxol from biomass collection of wild species, cell culture, semi synthesis, and total synthesis [61].

1.3.3 Synthesis and semi-synthesis of taxol

Solution to taxol supply problem can be overcome by semi-synthesis. Semi-synthesis of taxol represents some major problems. 10-deacetyl baccatin III can be extracted from the needles of different yew plants and the attachment of lengthy side chain of taxol provides a solution to the limited supply of taxol [62]. This approach brings supply surveillance and quality assurance of paclitaxel [63,64].

Advancement in the supply of paclitaxel has been achieved by the complete synthesis of compound. The semi-synthesis and complete synthesis of Paclitaxel has been achieved since 1994, but the approach is very complicated and is burdensome to scale-up the process. Therefore, replacing those biochemical assays for the formation of semi-synthetic or synthetic taxol did not suggest a beneficial commercial solution.

1.4 PHARMACOLOGY OF TAXOL

1.4.1 Mechanism of action of taxol

Taxol displays similar intracellular effects as other anti-cancer drugs but the mechanism of action of taxol is quite different from other drugs. It targets microtubule polymers of dimeric protein tubulin. Taxol is well known to increase the over-polymerization of microtubules which results in non-functional, assembled, and aberrant microtubules that are responsible for cell apoptosis [65]. Taxol binds to microtubule at a low concentration as $0.05\mu\text{mol/L}$ and make less concentration of tubulin for the rearrangement of microtubule even in the absence of factors such as GTP or MAPs that are customary for the dynamics of microtubules [66]. Taxol binds to the microtubules at a different site than colchicine, vinblastine and podophyllotoxin, yet the definite binding site of taxol on microtubule is not identified but it had been seen that taxol choose to bind to microtubules than tubulin dimers [67].

Microtubules which are treated with taxol found to be stable even if subjected to calcium or low temperature, those conventionally advocates disassembly [68]. Clinical concentrations of taxol ($0.1-10\mu\text{mol/L}$), can be attained by its prolonged infusions, apparently produce two morphological affects. First, in the vicinity of taxol there occurs the formation of ample arrays of disordered microtubules laterally aligned. This bundle formation occurs in every phase of cell cycle [69], interruption of mitotic spindle causes the capture of cell division in the G2/M phase. Secondly, taxol provide exceptional stability to microtubule structures that can afflict the effects of fundamental cellular regularity elements in this arrangement which finally modify microtubule dynamics and divest cell of its capability to control and arrange the

cytoskeleton. Due to effects caused by taxol, cell seems to disrupt the mitotic spindle during cell division and also the uprightness of interphase microtubules that is crucial for many cellular functions.

1.4.2 Structure activity features of taxol

Taxol has a unique structure, which explains its unusual clinical and pharmacological features and also its mechanism of action [70]. Taxol was examined for both as an anti-cancer drug as well a tool to understand the various cellular functions of microtubules. Structural requirements for obtaining taxol's effects are based on many studies those were designed to learn structure activity features of taxol [71]. The results obtained were suggestive in the following ways (Figure 1.7).

- i) C₂ and C₃ rings configuration are essential for the microtubule assembly.
- ii) Oxetane ring is also responsible for microtubule assembly.
- iii) Activation of taxol is dependent on esterification of C₁₃ hydroxyl group.
- iv) Major loss in the activity is noticed by the replacement of 3' phenyl by methyl.
- v) C₂ and C₇ hydroxyl group's esterification seems to increase the solubility of taxol.
- vi) Side chain configuration of taxol is least affected by C₃ phenyl group and C₂ hydroxyl group but are necessary for taxol microtubule interaction.
- vii) Efficacy of taxol is increased by the amide substitution at C₃ by hydrophobic substituent.
- viii) In vitro, cytotoxicity is not affected by C₁₀ acetyl group.

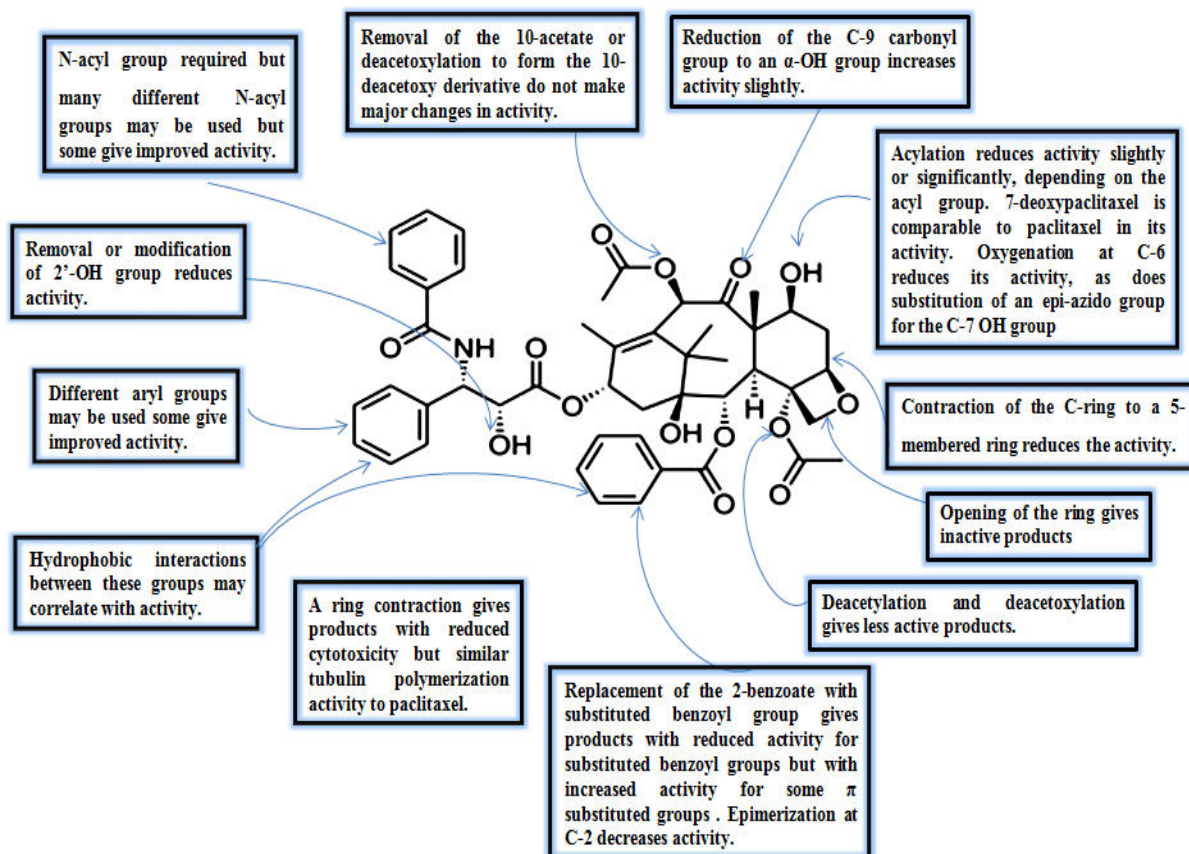


Figure 1.8: Structure activity relationship of taxol.

1.4.3 Resistance possessed by taxol

Two mechanisms of acquiring the resistance to taxol were observed *in vitro*. First, cell may get mutated when grown in the absence of taxol, lacking normal microtubules in their mitotic spindles hence cells become completely or partly reliant on the drug for their normal growth. Secondly, taxol may be involved in the multidrug resistance phenotype which advised varying degree of cross-resistance to taxol and also to other natural products [53].

1.5 PRECLINICAL AND CLINICAL INFORMATION OF TAXOL

1.5.1 Antitumor activity of taxol

Cytotoxic activity of taxol had been reported since 1971 by NCI after a large panel screening. Taxol showed dramatic anti-cancer activity against nearly resistant murine B16 melanoma. It showed good anti-cancer activity against a lot of other human tumor xenografts including MX-1 mammary tumor when given intravenously, showed moderate activity for human CX-1 colon, LX-1 lung xenografts, and L1210 murine leukaemia as screened by the NCI panel. In 1986 Riondel and Jacrot reported that taxol brought significant delay in growth of human brain, tongue, ovarian, endometrial, and lung xenografts [72]. Prolonged and frequent drug

administration schedules were found to be superior. Combination of taxol with cisplatin produced critical delay in the growth of tumor [73].

1.5.2 Clinical trials of taxol anti-cancer activity

Taxol (Paclitaxel) was approved by FDA for the treatment of ovarian and breast cancer patients [74]. According to NCI, taxol was in clinical studies in different areas; in phase-I, phase-II, and phase-III. Results from completed and ongoing phase I, II, III trials are compiled in Table 1.3.

Taxol was proved to be a salvage therapy for patients having advanced or metastatic breast, ovarian, head and neck, and NSCLC cancers. Taxol treated tumor patients exhibited influential response rates because the patients were pre-treated heavily and also patients were prior obstinate towards standard chemotherapy. Taxol activity against cervical, prostate, colon, renal cancers was not found to be promising at that time. Also, for bladder, stomach, oesophageal cancers as well for paediatric malignancies use of taxol was investigated.

Table 1.3: Different tumors and their response rates against taxol

Target tumor	Response rate (in percent)	References
Advanced ovarian cancer	36	[75], [76]
Metastatic breast cancer (with one prior chemotherapy)	59	[77], [78]
Metastatic breast cancer (without any treatment)	62	[79]
Untreated Non Small Cell Lung Cancer (NSCLC)	22 (21-24)	[80], [81]
Small Cell Lung Cancer (SCLC)	34	[82]
Malignant melanoma	16 (12-33)	[83]
Head and Neck	40	[84]

1.5.3 Animal toxicity of taxol

Toxicity of taxol was determined using animal models like rats, mice, and pigs. Dose-volume restrain in rodents was raised due to the low aqueous solubility of taxol, also the toxicity is imparted by cremophor EL (polyoxyethylated castor oil) and formulation of taxol in ethanol. Intra-peritoneal (ip) course of administration of drug was used for the toxicological studies

during the screening program [85]. Tissues which were having an expeditious cell turnover like lymphatic, reproductive, gastrointestinal tissues seemed to have noticeable toxic effects than hepatic, nervous, and renal tissues which showed no deposition of end-organ damage in post-mortem examinations. Taxol containing vehicle showed innate toxicity [86]. In toxicological studies, cremophor EL shows vasodilations, lethargy, hypotension, and laboured breathing. Although in the repeated dose regimen vehicle was better tolerated and did not showed any cumulative toxicity [87]. Some animals were shown to be more sensitive to the hypotensive effects of the cremophor and these results suggested some alternative for the insolubility of the taxol.

1.5.4 Combination chemotherapy of taxol

Due to the promising results obtained by single-agent taxol in various cancers, the next probable step in the development of this drug was to make combination of taxol with other active agents in these tumors. Two mainly studied combinations were cisplatin and taxol and doxorubicin and taxol. Other combinations like taxol/carboplatin, taxol/etoposide, taxol/ifosfamide, taxol/cyclophosphamide, taxol/hexamethylmelamine, and taxol/R-verapamil were also studied. These studies clearly defined the role of taxol in ovarian cancer; also these studies defined the role of this agent in breast cancer as the combination of taxol with other drugs was optimized. Dose response relationships within the range of current clinical dosages were important to resolve. Important issues in the development of taxol were its role as primary therapeutics in solid tumors, its optimal infusion length, clears understanding of its toxicities, and its low solubility.

1.5.5 Human toxicity

Taxol's major toxicities were depicted in the phase I trials of the drug. In these trials patients were pre-treated with anti-hypersensitivity regime comprising an antihistamine, H₁ blocker and a H₂ blocker. Neutropenia was shown to be dose limiting side effect in most of the cases. A proportional increase in the severe neutropenia was observed with the increase in the dosage. Also, other dose related toxicities like alopecia, myalgias, nausea, mucositis were observed in all patients. Hypersensitivity reactions and cardio toxicity were also seen in the patients but these were not dose related toxicities. Although dose-dependent toxicities were found but dose-response relationship was not possible to define for taxol [88].

1.5.6 Metabolism of taxol

Predisposition of the entire taxol metabolism was found to be hepatic metabolism and biliary metabolism. High concentrations of taxol and hydroxylated metabolites were found in bile of

both rat and human [89, 90, 91]. Glucourinated and sulfated metabolites were not identified in the human or rat metabolite. Baccatin III was recognised as a minor metabolite in the rat bile. With the exception of Baccatin III all other biliary metabolites identified were hydroxylated derivatives.

In vitro metabolism studies of taxol had shown the role of cytochrome P450s (CYPs) in the metabolism of taxol in the humans. CYPs from CYP2C family were shown to be responsible for the formation of major human hydroxylated derivatives and CYPs from CYP3A family were subjected to the formation of metabolite hydroxylated at position C₁₃ [92].

1.6 PHARMACOKINETICS OF TAXOL

Pharmacokinetic conduct of paclitaxel was studied during the phase I clinical trials. Most of these studies included 6 and 24 hours infusion schedules because of the high occurrences of major intense hypersensitivity reactions due to the shorter infusions as suggested by the early clinical trials. Studies showed that maximum plasma taxol concentrations (C_m) and areas under concentration versus time curve (AUC) were well correlated with taxol dose in 6 and 24 hours administration schedule [93,94]. During the early studies of taxol, biexponential elimination model was used for the study of taxol escape from the plasma. Mean alpha and beta half lives of taxol were found to 0.34 hour and 4.9 hour respectively. Also, Beijnen reported the equity of taxol from plasma by a three compartment model. The half lives were described as 10 minutes ($t_{1/2} \alpha$), 2 hours ($t_{1/2} \beta$) and 15 hours ($t_{1/2} \gamma$). Mean central (Vd_c) and steady state volumes (Vd_{ss}) of the distribution were found to large, having values of mean Vd_c and mean Vd_{ss} 13.81 l/m² 87 l/m² respectively [95]. These large values of Vd_{ss} that was even larger than the total volume of the body indicated taxol bound to proteins or some other tissue elements. Plasma protein binding was reported to be 95% to 97% for a wide range of drug concentrations [96].

Although, taxol showed high binding affinity to the plasma proteins but it was also readily eliminated from the plasma showing reversible binding of lower affinity. Early phase I trials and pharmacological studies found the systemic clearance values to be 496 ml/min/m². Also, the total urinary excretion of the taxol was 5.5%. This data gave the insight that the bulk amount of administered taxol dose disappearance was largely done by the metabolism, biliary excretion, and/or done by extensive tissue binding and was less contributed by systemic clearance and urinary excretion of the taxol was minimum [97].

Bioavailability studies done by Fujita H *et al*, 1994 had suggested that the dose given orally gave AUC (oral)/AUC (intravenous) ratio in tumor, liver, bile, plasma was 0.83, 6.71, 18.91 and 0.45% respectively. Eiseman *et al.*, 1994 defined the pharmacokinetics of paclitaxel after ip, po, and sc administration of 22.5 mg/kg of paclitaxel to mice to be 10%, 0 and 0 respectively.

Cremophor augments to the nonlinear pharmacokinetics of paclitaxel as shown by Sparreboom *et al.*, 1996. In the results given by them, for the given concentration of paclitaxel increase in the cremophor amount in the formulation decreases the taxol clearance. Also, they exemplified that although the plasma concentration for a fixed dose of taxol formulated in the cremophor were higher than the dimethylacetamide and tween 80 formulations but the concentration of paclitaxel in the tissues was same for all of the three tested preparations.

1.7 PHARMACEUTICS OF TAXOL

1.7.1 Formulation of taxol

Taxol is a non-polar compound having low-solubility in water. Paclitaxel, which is a formulated form of taxol is a clear, colourless, concentrated, and viscous solution consisting of 6 mg/ml of taxol in 50% cremophor EL (polyoxyethylated castor oil) and 50% dehydrated alcohol. Paclitaxel is infused intravenously after dilution to a concentration of 0.3 to 1.2 mg/ml with suitable parenteral fluid over 24 hours every 3 week. Cremophor EL was examined to cause many alarming hypersensitive reactions in the humans as well as animals. Many efforts were made to increase the paclitaxel solubility by chemical modifications while holding its anticancer activity but none of the modification was able to replace the taxol: cremophor formulation [98].

1.7.2 Stability of taxol

Paclitaxel has been observed to convert spontaneously into 7-epitaxol in the saline solution at a temperature of 37⁰ C after 48 hours. Tissue culture medium of J774.2 murine macrophage cells gave similar results for the drug, giving 50% of parent drug conversion to 7-epitaxol after 72-hours of the drug treatment [99,100]. Paclitaxel hydrolysis to several products including 7-epitaxol, baccatin-III, 10-deacetyl baccatin III pursued by the formation of 10-deacetyl taxol [89]. Diluted cremophor EL formulation of paclitaxel has been shown to form a particulate matter overtime suggesting the need of in-line filtration of the solution during dispensation. The stability of paclitaxel as suggested by bristol-myers company marked the

0.03-1.2 mg/ml concentration of paclitaxel in either NS, D5W, D5W in ringer's solution passed by 0.22 μm filter had not lost the potency in 27 hours kept in storage at normal room temperature and normal room lightning conditions.

1.7.3 Toxicity caused by cremophor EL in the taxol formulation

Several hypersensitive reactions were observed in the patients treated with cremophor-taxol formulation. These acute hypersensitive reactions like, shortening of breath, flushing, chest pain, urticaria started immediately after the administration of the formulation. Other observed side-effects of cremophor EL were increase in the blood platelet count and serum lipid values, and decrease in the percentage of alpha-lipoproteins in the blood. Increase in the tissue-lipids in spleen and lymph nodes was also observed due to the cremophor EL as observed by histopathological examination of the organs [101].

1.7.4 Cremophor EL and multidrug resistance

Mammalian cancer cells were shown to display MDR because of the production of more concentrations of p-glycoprotein, a membrane protein that behave as a broad-spectrum pump and efflux a number of anti-tumor drugs, like vinca alkaloids, daunorubicin, and doxorubicin from the cell. Paclitaxel has also been shown to induce the MDR effect of the cell. Dose-limiting toxicities imposed by many agents restricted their use but they were effective in reversing the MDR. As predicted by the *in-vitro* experiments that clinically readily available concentration of the cremophor EL reverses the MDR effect in the cells [102]. A variety of surface active agents some of them were used as pharmaceutical agents were compared to cremophor EL and found that Triton X-100 and Tween-80 showed the similar activity as that shown by cremophor EL in reversing the MDR effect but cremophor showed the lowest inherent cytotoxicity. Thus blocking the drug efflux and the enhanced fluidity of membrane did not involve cell lysis. Enhanced membrane fluidity may damage the ability of p-glycoprotein to efflux the drug from the cell. Though cremophor is found responsible for reversing MDR *in-vitro* and *in-vivo* but there was also need of high concentration of cremophor for resistance modulation. Although cremophor itself was not found to be major vesicant for the tumor cells but the formulation of taxol:cremophor may possess the vesicant activity [103].

1.8 PACLITAXEL PRODRUGS

Chemotherapy is an effective way to treat many types of cancer. Although paclitaxel is directed against various cancers but it is confronted with the limitation of extremely low

aqueous solubility. The formulation of paclitaxel gave a major impediment because of its low aqueous solubility. For that purpose it is always administered with cremophor EL and ethanol that may cause serious hypersensitivity in the patients. Hence, to circumvent the hypersensitivity and for obtaining better clinical use of the drug improvement in the formulation of paclitaxel has become important. Along with the development of new formulations for paclitaxel there occurred the development of paclitaxel prodrugs which were efficient modified form of paclitaxel.

1.8.1 Aqueous Solubility: A major limitation of paclitaxel

Aqueous solubility of a drug is the most important property that affects not only drug's formulation but also the bioavailability of the drug. Prodrugs are the chemically modified form of the pharmacologically active drug that must be subjected to *in-vivo* transformation to release the active drug [104]. Prodrug formation of paclitaxel is the approach to improve the physicochemical, biopharmaceutical, and pharmacokinetic properties of the drug thereby increasing its advancement and its usefulness as an anti-cancer drug [105]. Paclitaxel prodrugs formation for increasing its solubility also increases its bioavailability and may also represent a life-cycle management opportunity with the improved properties of this established drug.

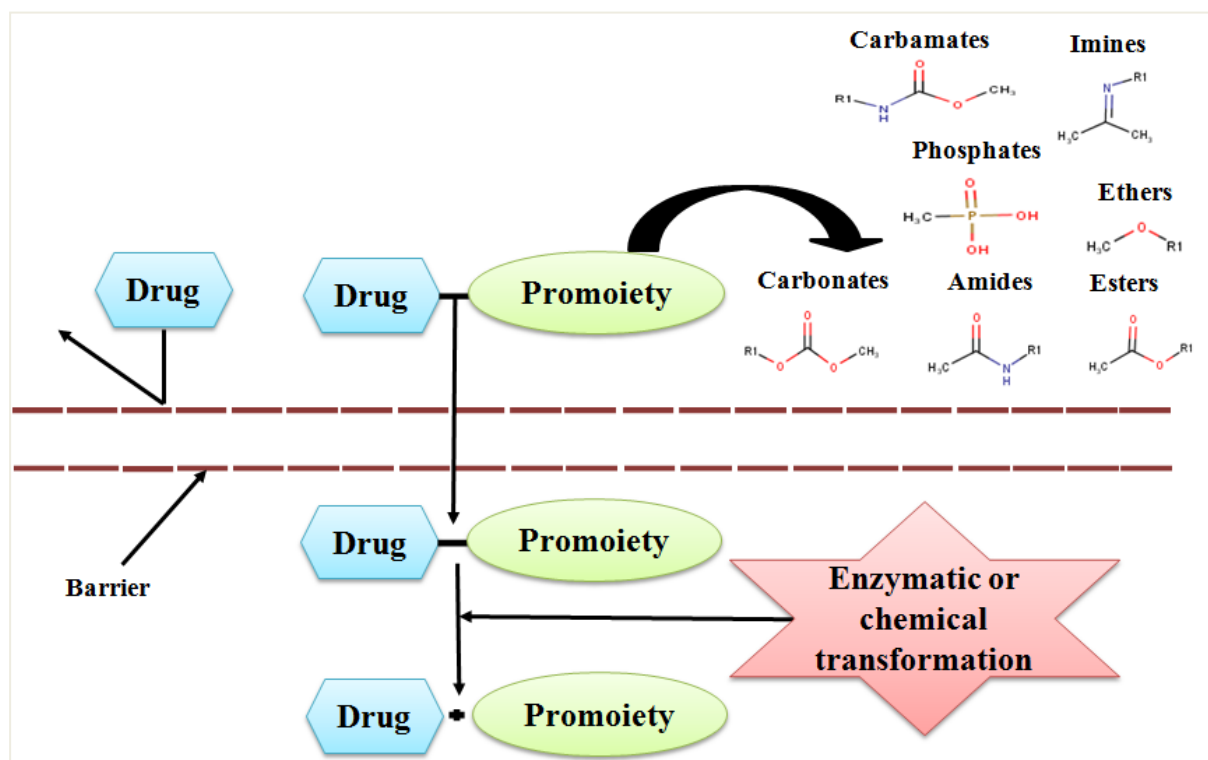


Figure 1.9: Bioactivation of paclitaxel prodrugs (*in-vivo*) by chemical or enzymatic modifications (adapted from Chung *et al*, 2015).

Besides extremely low aqueous solubility paclitaxel also come up with multiple drug resistance like other anti-cancer drugs. Paclitaxel prodrug formation strategy overcomes the imperfections of the drug to a large extent. The prodrug formation included the small molecule and macro molecule, small molecule prodrug formation overcomes the low aqueous solubility of paclitaxel and macro-molecule prodrug included the conjugation of paclitaxel with polymers or with proteins. Macromolecule prodrug formation of paclitaxel could generally enhance the permeability and retention effects (EPR) as for other prodrug strategies [106]. In view of this, many prodrugs of paclitaxel were formed for more effective use of this highly active anti-tumor drug. As anti-cancer mechanism of paclitaxel was unfolded it was crucial to make it more hydrophilic so as to be more bioavailable. Many of the groups worked for making it more hydrophilic, two different approaches were followed to make it more bioavailable. Firstly, by adding a soluble group to the molecule and forming more soluble drug. Another way was making drug hydrophobic which was acceptable for long half-life hydrophobic formulation.

For the formation of paclitaxel prodrugs one has to understand structure-activity relationship of the molecule. Early work showed that the central part of the molecule is rigid to accept any structural modification but the tail part is flexible for making any structural change. The central part of paclitaxel holding 1-OH and 2-benzoyloxy are the vital components for its anti-cancer activity. Also, the stereochemistry of C2' and C3' in the side chain affects the activity of drug. Some studies suggested C2'-OH as binding site of the drug to tubulin. Therefore it was proposed to be most important functional group in the side chain which is a vital component of the drug. Hence early development of the prodrugs mainly focus on C2' and C3' positions. Though C-7 and C-10 does not interact with the tubulin directly but it was suggested that these were involved in the binding of the drug with P-glycoprotein which was subjected to MDR phenotype. The complex structure of paclitaxel allows modification of the molecule at different positions but for the formation of prodrugs C2'-OH was commonly accepted as it is the binding site of paclitaxel to tubulin. Also C7'-OH was also accepted by some groups for the formation of prodrugs [107], see chapter 3 for complete details.

1.8.2 Small Molecule Paclitaxel prodrugs

Small molecule paclitaxel prodrugs were formed to increase the solubility of the drug. In the small molecule prodrug formation phosphate esters of paclitaxel were formed which is a common strategy to increase water solubility of the molecule, also which provides a site to alkaline phosphatase enzymes for the release of the parent drug. Some phosphonooxymethyl

ethers of paclitaxel were formed by the addition of phosphate group to the C2'-OH and C7'-OH of paclitaxel molecule. The prodrugs formed showed great improvement in water solubility. Ueda and co-workers, introduced the phosphate group to the 7'-OH of 2'-ethoxycarbonyl paclitaxel which raised the water solubility of molecule to 2.5 mg/ml, and the formed complexes showed the *in-vivo* cytotoxicity comparable to paclitaxel [108]. Likewise, Damen *et al*, synthesized two paclitaxel esters of malic acids at 2'-OH and 7'-OH respectively. The resulting prodrugs were 50 times more soluble than paclitaxel and showed equal anti-tumor activity as that of paclitaxel [109]. Besides, Wohl and co-workers synthesized 2'-OH silicated Paclitaxel prodrugs and obtained improved the water solubility [110].

In the initial stage of paclitaxel's development approach, studies were mainly targeted on its low aqueous solubility. In order to achieve more competent paclitaxel many research group formed various prodrugs of paclitaxel. With these preparations several prodrugs with improved solubility as well as with equivalent anti-cancer properties were formed, those were directed for pre-clinical and clinical studies and were subsequently marketed and used for various cancers.

1.9 DATA MINING TECHNIQUES, QSPR MODELS AND VARIOUS COMPUTATIONAL TOOLS

Computational models are imperative for eloquence of biology from descriptive to a predictive science. Now-a-days computational approaches are gaining acceptance, practice, and recognition in the drug exploration and discovery. Computational techniques bring about the use of computing power for streamlining the drug development process; it takes advantage of chemical and biological knowledge of ligands, targets and assists in the identification, and optimization of novel drugs. Newly designed *in-silico* filters eliminate the compounds with poor abominable properties and aids in selecting the most likely compounds. Rapid development in the computer-aided lead identification and development occurred in the last decade because of the continuous progress, development, and fineness in the computational software and hardware power. [111]. These computer-aided techniques are being utilized for the identification of active drug candidates, most likely candidates for the further evaluation, and optimizing biologically active compounds into drugs by enhancing their physicochemical, pharmacokinetic, and pharmaceutical properties. Implicitly the approach is being utilized for discovering new drug candidates designed with a reduction in the chemical space and thus allowing the focus on the assuring candidates. Particularly, the *in-silico* techniques allow us to

necessarily curtail the time, and resources of the synthetic chemists and of biological testing. Along with the identification of new promising leads it is essential in the drug discovery process that the drug should not undergo attrition.

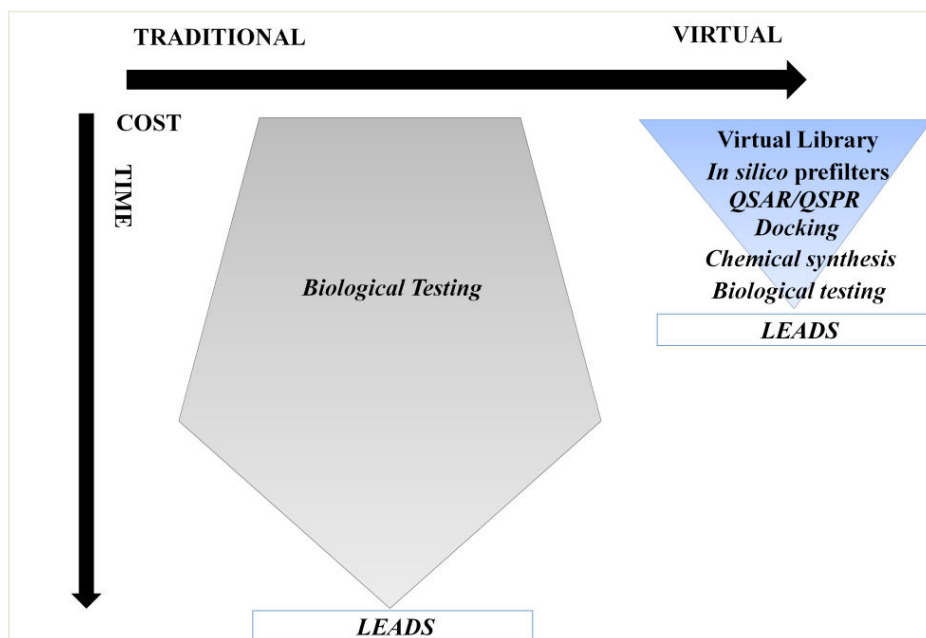


Figure 1.10: Traditional and virtual screening of lead compounds.

From decades toxicity of the drug is the major cause for drug attrition and is one of the main causes for the drug failure. Preclinical animal studies are time consuming, expensive and often provide defined predictability for human effects. Hence, there is a need of time for the development of *in-silico* predictive assessment of toxicity and other pharmacokinetic properties so as to minimize the animal testing. It is expected that the *in-silico* predictive model will aid in escaping the resource wastage, will reduce the regulatory review responsibility, and will predict the detrimental effects which were not predicted by the animal model such as nausea, cognitive impairment, dizziness etc [112]. European policy for the evaluation of chemicals (REACH: Registration, Evaluation, And Authorization of Chemicals) strongly supported the idea of an alternate *in-silico* mechanism for predicting the chemical toxicity, and other pharmacokinetics of the chemicals so as to lessen the animal trials, and sustain the resources and time.

Quantitative structure activity relationship (QSAR) and Quantitative structure property relationship (QSPR) are the frequently used computational predictive methods for the toxicology and the pharmacokinetic properties [113]. Both the methods share the same principle but the frame of reference in QSAR is biological activity and in QSPR it is bio-

physico-chemical properties. Molecular descriptors like topological, conformational, spatial, electronic etc. are the representative of dependent variables in the QSAR/QSPR models. The idea of structure-activity relationship came into existence in eighteenth century but later in the late 60s Corwin Hansch had demonstrated the versatility of this predictive approach and led to its advancing usage [114]. The field holds the interest in the regulatory arena and was being continuously evaluated. The regulatory body of Informatics and Computational Safety Analysis Staff (ICSAS) within the CDER at FDA is actively involved in the assessment of these predictive models used for toxicology and pharmacokinetic properties [115]. They developed various databases of different clinical endpoints, developed various databases, and had evaluated the data mining and QSAR techniques. QSAR validation is mostly done using the internal dataset or external validation data set, where the validation principles include accuracy and the rank ordering [116]. However, there still remain concerns regarding these predictive QSAR/QSPR models which should be further addressed through the present studies, and can provide promising results. It is also believed that the predictions based on direct human data would decrease the reliance on animal studies that are lengthy and time consuming.

In general data mining and QSPR models are dynamic techniques that are able to correlate the molecular properties with the physicochemical properties of the molecules [117]. QSAR/QSPR techniques can be defined as the application of data analysis and statistical methods for the formation of models that could accurately predict the activity or biological property of the compounds based on their structures. It finds some empirical relationships (QSAR/QSPR) of the form $P_i = k'(M_1, M_2, M_3, \dots, M_n)$ where P_i is the property of interest of the molecules, $M_1, M_2, M_3, \dots, M_n$ are the calculated structural properties (descriptors) or experimentally determined properties of the molecules and k_i is some mathematically defined empirical function. QSAR/QSPR modelling aims to establish a trend among the descriptor values which parallels the trend in the biological property [118]. The QSAR/QSPR approach implies a simple principle of “similarity” that for a long time had established a foundation for the medicinal chemistry: compounds having similar structure shows similarity in their activities/properties also. QSAR/QSPR model development and validation includes a complete workflow that depicts the predictive QSAR/QSPR modelling, proclaiming a validated model, and finally some computational hits those could be further experimentally verified as shown in Figure 1.10.

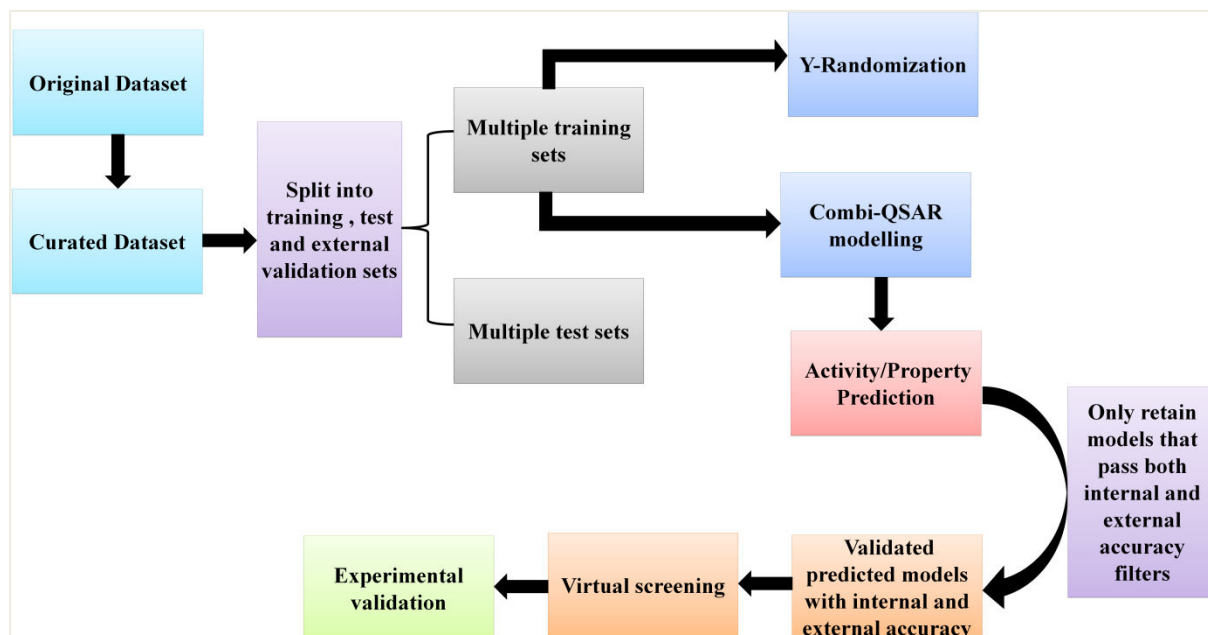


Figure 1.11: Schematic representation of QSAR/QSPR predictive modelling.

The formation of QSAR/QSPR model includes the step wise process, which starts by the careful curation of chemical structures and associated biological activities/properties to form a dataset for the subsequent computations. A part of compounds (mostly, 10-20%) is selected from the dataset as an external validation dataset. Subsequently, there is employment of multiple QSAR techniques based upon the combinatorial exploration of all possible pairs of descriptor sets and employing various supervised data analysis techniques “combi-QSAR/QSPR” and finally select models which represented the higher accuracy of prediction in the training and test sets [119].

1.9.1 Tools used in the study for the formation of QSPR pipeline (QSPR-sPL), QSPR models formation, and validation.

Gaussview: Gaussview is a graphical user interface, designed for the preparation of molecules to give as input to the Gaussian for molecular optimization and also to visualize the output [120]. It provides the benefit of advanced visualization facility, it makes easy to set up Gaussian calculations, also it let to examine the results implementing a variety of graphical techniques. Structure of prodrugs of paclitaxel are large and complex as the structure involves a eight-membered ring, four six-membered rings, one four-membered ring, and also the side carbon chain with branches (Figure 1.7a). This complex structure needs an advanced visualization and drawing tool for the drawing of prodrugs structures. paclitaxel prodrugs

taken from literature are drawn and their three dimensional coordinates are saved in Gaussview. The 3D structure is further taken for the structure optimization in Gaussian09.

Gaussian 09: Gaussian is a computer program that utilizes laws of fundamental mechanics for the prediction of molecular structure, energies, and spectroscopic data [121]. It is used for the electronic structure modelling. Gaussian 09 provides the capability that will enable to model molecular systems with increasing size, more accuracy, and with a broader range of real biological situations. It provides comprehensive investigation of molecules by predicting the transition state structures, and verifying local minima of the structures. For the optimization of molecular structures of paclitaxel prodrugs which is a molecule with higher molecular weight and for verifying the global minimum structure vibration frequencies are taken into account for which Gaussian 09 provides the accurate structures.

Dragon7: The study implemented Dragon 7.0 software for the calculation of molecular descriptors [122]. Molecular descriptor set must be functionally complete set that expresses all the features of various molecules so that it can be used to construct predictive models for different activities and properties. Dragon 7.0 calculates 5250 molecular descriptors that are organized in different logical blocks covering most of the various theoretical approaches. Descriptor list of Dragon 7.0 includes simplest atom types, functional groups, geometrical and topological descriptors, three dimensional descriptors, and also various properties estimation such as logP and drug-like and lead-like alerts. Broad range of different approaches and theories are used for descriptors calculation, and their implementation correctness and precision is ensured. Most of the common organic and inorganic compounds either charged or uncharged are correctly processed for molecular descriptors in Dragon. Paclitaxel prodrugs taken for the development of predictive QSPR model for the solubility prediction in the study are complex structures to study. Combining molecular descriptors from multiple resources might lose the integrity of different descriptor systems, introduce redundancy, and sometimes forfeit causation. Hence for the descriptor calculation Dragon 7 was used as it will give a complete and clean non redundant set which can be used for the formation of QSPR model for paclitaxel prodrugs.

WEKA: WEKA is popular open source software that has tools for data pre-processing, implementation of various algorithm and visualization tools also so that machine learning

techniques can be developed and applied to the data mining [123]. Regression is an important class in predictive modelling. Linear regression is an approach for modelling relationship between a scalar dependent variable and more explanatory non-dependent variable. Also, dissolving a compound in aqueous solution is a complex process, involving multiple factors accounting for solute-solute, solvent-solvent, and solute-solvent interactions. Therefore, it is reasonable to introduce nonlinearity to the QSPR models to tackle the complex problem with correspondingly improved complexity of the models. Hence for the development of QSPR model for solubility prediction of paclitaxel prodrugs for which a descriptor set of 5250 descriptors is calculated using Dragon 7 software and AIC and VIF multicollinearity indicators are used to reduce the collinearity in the data. Then WEKA which is comprehensive software for the application of various machine learning algorithms and comparison of different outputs was used for the formation of QSPR model for the paclitaxel prodrugs.

MATLAB: MATLAB also known as matrix laboratory is a computing environment and programming language. Although MATLAB allows various computations like plotting of functions, matrix manipulations, implementation of algorithm, but primarily MATLAB is used for numerical computing. Regression is a statistical modelling technique used to describe a continuous response variable as a function of one or more predictor variables [124]. It can help to understand and predict the behaviour of complex systems or analyze experimental biological data. For multiple and multivariate regression model MATLAB provides regression methods that accommodate multiple response variables and enables multivariate regression to generate predictions. In the development of QSPR strategy for solubility prediction of paclitaxel prodrugs, a pipeline is developed and to compare the results obtained from the developed pipeline, regression model is also developed from MATLAB stepwiselm tool. It has been established that R^2 and Q^2 values obtained from QSPR-sPL are comparable to the R^2 and Q^2 values obtained with the MATLAB stepwiselm tool.

y-scrambling: y-scrambling is used to verify the absence of chance correlation between the independent variables and the aqueous solubility in the training set and further validate the model. When the compounds in the external test set fall into the applicability domain of the training set, the model will provide a decent prediction to the test set despite the occurrence of over fitting [125]. To check the model avoid over fitting, the solubility model was subjected to a different validation test. Y-scrambling constructs a new model after shuffling the

modelled property, usually termed y vector, and to compare the performances of the Y -randomized model with the original model. A large deviation in the performances of the two models is a good indication of a statistically solid model. Y -scrambling completely demolished the correlation between the independent variables and the dependent variables, demonstrating that the relationships extracted in the QSPR modelling are true relationships. Therefore, the Y -scrambling test clearly verified the absence of chance correlation in the QSPR model.

admetSAR: admet structure activity relationship (admetSAR) server is a comprehensive knowledge and tool for predicting Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) properties of drug compounds and environmental molecules [126]. ADMET properties of molecules play an important role in the drug discovery process and also in the assessment of environmental risk. admetSAR 2.0 describes several properties that should be taken for the screening of drug molecules for the metabolic study of molecules. In the study of bioavailability of paclitaxel prodrugs, molecules those are representing more favourable properties are further taken for docking and molecular dynamics simulation studies.

AutoDock: AutoDock is a molecular docking tool, designed for the prediction of small molecule and substrate interaction [127]. It has been widely used and examples of its successful application have been seen in extensive literature available on PubMed and other scientific resources. It is very fast, effective and accurate tool and provides high quality predictions of ligand conformations. It also shows good correlations between experimental and predicted inhibition constants. It uses free-energy scoring function, the AMBER force field, and a larger dataset of diverse protein-ligand complexes with known inhibition constants. In the AutoDock diverse and larger set of HIV-1 protease complexes are used for the validation of best protein-ligand model which confirms the standard error of 2.5 kcal/mol. It is enough for the discrimination of lead molecules with milli, micro, nano, and molar inhibition constants. For the docking studies we have used AutoDock where the molecules were screened on the basis of binding energies and the conformation poses.

AutoDock Vina: AutoDock Vina shows improved average accuracy of the binding mode predictions as compared to AutoDock [128]. It is found to be more improved than many other docking programs. AutoDock vina is used for finding the consensus in the results obtained

from AutoDock. For screening the molecules for further molecular dynamics simulation studies the molecules which are showing consensus among AutoDock and AutoDock Vina are taken into account.

Gromacs: Gromacs is molecular dynamics tool used for carrying out the molecular dynamics simulation of proteins, lipids, and nucleic acids [129]. It is very fast optimizing tool and also supports various kind of force fields. All the molecules attain convergence at the end of simulation period. It is used in the study for carrying out the molecular dynamics simulation study of paclitaxel prodrugs with cytochromes, to study the bioavailability of the prodrugs.

1.10 PROBLEM IDENTIFICATION

Paclitaxel prodrug formation for the increase in solubility was employed and prodrugs with increased solubility were also obtained. Although many of the prodrugs of paclitaxel were formed in the past but the process of formation was not rational. The formation process includes many of the experimental set-ups along with the formation of many side compounds that are not cost- and time- effective thus hampering the growth of this nano-molar active drug. On the other side there are techniques QSAR/QSPR methods for the prediction of biological activities or the properties for the biological or chemical compounds [130]. A rational solution for the formation of paclitaxel prodrugs with improved solubility can be found. Also no such QSPR model has been suggested for the solubility prediction of paclitaxel prodrugs, which has a very bulky molecule and has complex structure. Looking into research gap between the two commendable techniques; the predictive QSAR/QSPR techniques and the prodrug formation of the Paclitaxel the current study tried to fill this gap of by employing the computational and data analysis methods for the formation QSPR model for the solubility prediction of Paclitaxel prodrugs.

As discussed previously the formation of prodrugs of Paclitaxel includes the addition of a hydrophobic group to the chemical structure of the Paclitaxel and made the whole compound inactive, this inactive compound becomes active in the human body and release the parent drug in its native form so that it can show its desired effect. Hence, the promoiety that is added to the Paclitaxel for the formation of prodrug must be pharmacologically affable, exhibit clinical effect, and impeded the anti-cancer property of the Paclitaxel. The current study tried to put efforts for the formation of a QSPR strategy and models for the solubility prediction of Paclitaxel and also the formation of QSPR model for the substructures that were

added for the formation of prodrugs. Metabolic study of the prodrugs is also carried computationally so that we can contribute a comprehensive view of predictive information for the solubility and bioavailability of paclitaxel prodrugs.

References

- [1] World Health Organization. Global Health Observatory. Geneva: World Health Organization; 2018. who.int/gho/database/en/. Accessed June 21, 2018.
- [2] F. Bray, et al., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68(6), pp. 394-424, 2018.
- [3] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Pineros, and F. Bray, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144(8), pp. 1941-1953, 2019.
- [4] N. Jain, G. Gupta, S. Bichpuria, J. Rai, M. A. Mishra, and S. Ojha, "REVIEW OF RECENT CANCER STATISTICS: AN INDIAN PERSPECTIVE," *International Journal of Scientific Research*, vol. 8(8), 2019.
- [5] R. Dikshit, P. C. Gupta, C. Ramasundarahettige, et al., "Cancer mortality in India: a nationally representative survey," *Lancet*, vol. 379(9828), pp. 1807-1816, 2012.
- [6] M. P. H. Siegel, L. Rebecca, D. Kimberly Miller, and J. Ahmedin, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69(1), pp. 7-34, 2019.
- [7] P. K. Dhillon, P. Mathur, A. Nandakumar, C. Fitzmaurice, G. A. Kumar, R. Mehrotra, and J. S. Thakur, "The burden of cancers and their variations across the states of India: the Global Burden of Disease Study 1990–2016," *The Lancet Oncology*, vol. 19(10), pp. 1289-1306, 2018.
- [8] K. C. Zorn, O. N. Gofrit, G. D. Steinberg, and A. L. Shalhav, "Evolution of robotic surgery in the treatment of localized prostate cancer," *Current treatment options in oncology*, vol. 8(3), pp. 197-210, 2007.
- [9] L. R. Medeiros, D. D. Rosa, M. C. Bozzetti, J. M. Fachel, S. Furness, R. Garry, and A. T. Stein, "Laparoscopy versus laparotomy for benign ovarian tumor," *Cochrane Database of Systematic Reviews*, vol. 2, 2009.
- [10] S. P. Luh, and H. P. Liu, "Video-assisted thoracic surgery—the past, present status and the future," *Journal of Zhejiang University Science B*, vol. 7(2), pp. 118-128, 2006.

- [11] D. E. Gerber, and T. A. Chan, "Recent advances in radiation therapy," *American Family Physician*, vol. 78(11), pp. 1254-1262, 2008.
- [12] A. Desai, and T. J. Mitchison "Microtubule polymerization dynamics," *Annual review of cell and developmental biology*, vol. 13(1), pp. 83-117, 1997.
- [13] J. Howard, and A. A. Hyman, "Dynamics and mechanics of the microtubule plus end," *Nature*, vol. 422(6933), pp. 753, 2003.
- [14] M. A. Jordan, and L. Wilson, "Microtubules as a target for anticancer drugs," *Nature Reviews Cancer*, vol. 4(4), pp. 253, 2004.
- [15] E. Nogales, S. G. Wolf, and K. H. Downing, "Correction: Structure of the $\alpha\beta$ tubulin dimer by electron crystallography", *Nature*, vol. 393(6681), pp.191, 1998.
- [16] S. Chakraborty, S. Gupta, T. Sarkar, A. Poddar, J. Pena, R. Solana, and B. Bhattacharyya, "The B-ring substituent at C-7 of colchicine and the α -C-terminus of tubulin communicate through the "tail-body" interaction," *Proteins: Structure, Function, and Bioinformatics*, vol. 57(3), pp. 602-609, 2004.
- [17] S. S. Rai, and J. Wolff, "The C terminus of β -tubulin regulates vinblastine-induced tubulin polymerization," *Proceedings of the National Academy of Sciences*, vol. 95(8), pp. 4253-4257, 1998.
- [18] B. Bhattacharyya, D. Panda, S. Gupta, and M. Banerjee, "Anti-mitotic activity of colchicine and the structural basis for its interaction with tubulin," *Medicinal research reviews*, vol. 28(1), pp. 155-183, 2008.
- [19] R. F. Williams, C. L. Mumford, G. A. Williams, et al., "A photoaffinity derivative of colchicine: 6'-(4'-azido-2'-nitrophenylamino) hexanoyldeacetylcolchicine. Photolabeling and location of the colchicine-binding site on the alpha-subunit of tubulin," *Journal of Biological Chemistry*, vol. 260, pp. 13794-13802, 1985.
- [20] A. R. Safa, E. Hamel and R. L. Felsted, "Photoaffinity labeling of tubulin subunits with a photoactive analogue of vinblastine," *Biochemistry*, vol. 26, pp. 97-102, 1987.
- [21] S. B. Horwitz, "Taxol (paclitaxel): mechanisms of action," *Annals of Oncology*, vol. 5, pp. S3-S6, 1994.
- [22] A. Jordan, J. A. Hadfield, N. J. Lawrence, et al., "Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle," *Medicinal Research Reviews*, vol. 18, pp. 259-296, 1998.
- [23] J. J. Correia, and S. Lobert, "Physiochemical aspects of tubulin-interacting antimitotic drugs," *Current Pharmaceutical Design*, vol. 7, pp. 1213-1228, 2001.

- [24] V. K. Ngan, K. Bellman, D. Panda, et al., "Novel actions of the antitumor drugs vinflunine and vinorelbine on microtubules," *Cancer Research*, vol. 60, pp. 5045-5051, 2000.
- [25] C. Shih, & B. A. Teicher, "Cryptophycins: a novel class of potent antimetabolic antitumor depsipeptides," *Current Pharmaceutical Design*, vol. 7, pp. 1259-1276, 2001.
- [26] L. M. Leoni, E. Hamel, D. Genini, et al., "Indanocine, a microtubule-binding indanone and a selective inducer of apoptosis in multidrug-resistant cancer cells," *Journal of National Cancer Institute*, vol. 92, pp. 217-224, 2000.
- [27] M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon, and A. T. McPhail, "Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*," *Journal of the American Chemical Society*, vol. 93(9), pp. 2325-2327, 1971.
- [28] V. Walsh, J. Goodman, "From Taxol to Taxol®: The changing identities and ownership of an anti-cancer drug," *Medical anthropology*, vol. 21(3-4), pp. 307-336, 2002.
- [29] P. B. Schiff, J. Fant, and S. B. Horwitz, "Promotion of microtubule assembly in vitro by taxol," *Nature*, vol. 277, pp., 665-667, 1979.
- [30] P. B. Schiff, and S. B. Horwitz, "Taxol assembles tubulin in the absence of exogenous guanosine 5'-triphosphate or microtubule-associated proteins," *Biochemistry*, vol. 20, pp. 3247-3252, 1981.
- [31] P. B. Cassidy, P. J. Moos, R. C. Kelly, et al., "Cyclooxygenase-2 induction by paclitaxel, docetaxel, and taxane analogues in human monocytes and murine macrophages: structure- activity relationships and their implications," *Clinical Cancer Research*, vol. 8, pp. 846-855, 2002.
- [32] E. Nogales, "A structural view of microtubule dynamics," *Cellular and Molecular Life Sciences*, vol. 56, pp. 133-142, 1999.
- [33] J. P. Snyder, J. H. Nettles, B. Cornett, et al., "The binding conformation of Taxol in B-tubulin: A model based on electron crystallographic density," *Proceedings of the National Academy of Sciences of United States of America*, vol. 98, pp. 5312-5316, 2001.
- [34] S. S. Bacus, A. V. Gudkov, M. Lowe, et al., "Taxol-induced apoptosis depends on MAP kinase pathways (ERK and p38) and is independent of p53," *Oncogene*, vol. 20, pp. 147-155, 2001.
- [35] L. Wilson, and M. A. Jordan, "Microtubule dynamics: taking aim at a moving target," *Chemistry and Biology*, vol. 2, pp. 569-573, 1995.
- [36] J. Parness, D. G. Kingston, R. G. Powell, C. Harracksingh, et al., "Structure-activity study of cytotoxicity and microtubule assembly in vitro by taxol and related taxanes," *Biochemical and Biophysical Research Communications*, vol. 105, pp. 1082-1089, 1982.
- [37] E. K. Rowinsky, "The development and clinical utility of the taxane class of antimicrotubule chemotherapy agents," *Annual Review of Medicine*, vol. 48, pp. 353-374, 1997.

- [38] C. Bogdan, and A. Ding, "Taxol, a microtubule-stabilizing antineoplastic agent, induces expression of tumor necrosis factor alpha and interleukin-1 in macrophages," *Journal of Leukocyte Biology*, vol. 52, pp. 119-21, 1992.
- [39] P. J. Moos, and F. A. Fitzpatrick, "Taxane-mediated gene induction is independent of microtubule stabilization: induction of transcription regulators and enzymes that modulate inflammation and apoptosis," *Proceedings of the National Academy of Sciences of United States of America*, vol. 95, pp. 3896-3901, 1998.
- [40] C. M. White, B. K. Martin, L. F. Lee, and et al., "Effects of paclitaxel on cytokine synthesis by unprimed human monocytes, T lymphocytes, and breast cancer cells," *Cancer Immunology, Immunotherapy*, vol. 46, pp. 104-112, 1998.
- [41] E. Napoleone, F. Zurlo, M. C. Latella, et al., "Paclitaxel downregulates tissue factor in cancer and host tumor-associated cells," *European Journal of Cancer*, vol. 45, pp. 4707, 2009.
- [42] R. Altaha, T. Fojo, E. Reed, et al., "Epothilones: a novel class of non-taxane microtubule-stabilizing agents," *Current Pharmaceutical Design*, vol. 8, pp. 1707-1712, 2002.
- [43] K. H. Altmann, "Microtubule-stabilizing agents: a growing class of important anticancer drugs," *Current Opinion in Chemical Biology*, vol. 5, pp. 424-431, 2001.
- [44] L. He, G. A. Orr, and S. B. Horwitz, "Novel molecules that interact with microtubules and have functional activity similar to Taxol," *Drug Discov Today*, vol. 6, pp. 1153-1164, 2001.
- [45] R. J. Kowalski, P. Giannakakou, S. P. Gunasekera, et al., "The microtubule-stabilizing agent discodermolide competitively inhibits the binding of paclitaxel (Taxol) to tubulin polymers, enhances tubulin nucleation reactions more potently than paclitaxel, and inhibits the growth of paclitaxel-resistant cells," *Molecular Pharmacology*, vol. 52, pp. 613-622, 1997.
- [46] H. Madari, D. Panda, L. Wilson, et al., "Dicoumarol: a unique microtubule stabilizing natural product that is synergistic with Taxol," *Cancer Research*, vol. 63, pp. 1214-1220, 2003.
- [47] D. E. Pryor, A. O'Brate, G. Bilcer, et al., "The microtubule stabilizing agent laulimalide does not bind in the taxoid site, kills cells resistant to paclitaxel and epothilones, and may not require its epoxide moiety for activity," *Biochemistry*, vol. 41, pp. 9109-9115, 2002.
- [48] K. A. Hood, L. M. West, B. Rouwe, & et al., "Peloruside A, a novel antimetabolic agent with paclitaxel-like microtubule-stabilizing activity," *Cancer Research*, vol. 62, pp. 3356-3360, 2002.
- [49] R. E. Perdue, "Search for plant sources of anticancer drugs," *Morris Arboretum Bulletin*, 1969.
- [50] A. Hoernel, "Plants used against cancer:-a Survey," *Jonathan L. Hartwell, Quarterman Publications, Inc. Lawrence, MA*, vol. 8, pp. 438, 1982.
- [51] R. E. Gregory, and A. F. DeLisa, "Paclitaxel: a new antineoplastic agent for refractory ovarian cancer," *Clinical Pharmacy*, vol. 12(6), pp. 401-415, 1993.

- [52] M. E. Wall, and M. C. Wani, "Camptothecin and taxol: from discovery to clinic," *Journal of ethnopharmacology*, vol. 51(1-3), pp. 239-254, 1996.
- [53] E. K. Rowinsky, N. Onetto, R. M. Canetta, and S. G. Arbuck, "Taxol: the first of the taxanes, an important new class of antitumor agents," *In Seminars in oncology*, vol. 19(6), pp. 646-662, 1992.
- [54] P. H. Wiernik, E. L. Schwartz, J. J. Strauman, J. P. Dutcher, R. B. Lipton, and E. Paietta, "Phase I clinical and pharmacokinetic study of taxol," *Cancer research*, vol. 47(9), pp. 2486-2493, 1987.
- [55] J. M. Koeller, H. A. Burris, R. T. Dorr, R. S. Finley, B. R. Goldspiel, J. G. Kuhn, and E. K. Rowinsky, "Pharmaceutical issues of paclitaxel," *Annals of Pharmacotherapy*, vol. 28(5), pp. S5-S36, 1994.
- [56] E. K. Rowinsky, P. J. Burke, J. E. Karp, R. W. Tucker, D. S. Ettinger, and R. C. Donehower, "Phase I and pharmacodynamic study of taxol in refractory acute leukemias," *Cancer research*, vol. 49(16), pp. 4640-4647, 1989.
- [57] G. M. Cragg, "Paclitaxel (Taxol®): a success story with valuable lessons for natural product drug discovery and development," *Medicinal research reviews*, vol. 18(5), pp. 315-331, 1998.
- [58] M. J. O'Neil (ed), "The Merck Index-An encyclopedia of Chemicals, Drugs, and Biologicals," *Whitehouse Station, NJ: Merck and Co., Inc.*, pp. 1204, 2006.
- [59] C. R. Wilson, J. M. Sauer, and S. B. Hooser, "Taxines: a review of the mechanism and toxicity of yew (*Taxus* spp.) alkaloids," *Toxicon*, vol. 39(2-3), pp. 175-185, 2001.
- [60] D. Mastropaolo, A. Camerman, Y. Luo, G. D. Brayer, and N. Camerman, "Crystal and molecular structure of paclitaxel (taxol)," *Proceedings of the National Academy of Sciences*, pp. 92(15), pp. 6920-6924, 1995.
- [61] A. E. Koepp, M. Hezari, J. Zajicek, B. S. Vogel, R. E. LaFever, N. G. Lewis, and R. Croteau, "Cyclization of Geranylgeranyl Diphosphate to Taxa-4 (), 11 ()-diene Is the Committed Step of Taxol Biosynthesis in Pacific Yew," *Journal of Biological Chemistry*, vol. 270(15), pp. 8686-8690, 1995.
- [62] R. A. Holton, "U.S. Patent No. 5, 175, 315," *Washington, DC: U.S. Patent and Trademark Office*, 1992.
- [63] K. V. Rao, J. B. Hanuman, C. Alvarez, M. Stoy, J. Juchum, R. M. Davies, and R. Baxley, "A New Large-Scale Process for Taxol and Related Taxanes from *Taxus brevifolia*," *Pharmaceutical research*, vol. 12(7), pp. 1003-1010, 1995.
- [64] R. A. Holton, C. Somoza, H. B. Kim, F. Liang, R. J. Biediger, P. D. Boatman, et al., "The total synthesis of paclitaxel starting with camphor," *ACS Symposium Series*, vol. 583, pp. 288-301, 1995.

- [65] N. Kumar, "Taxol-induced polymerization of purified tubulin Mechanism of action," *Journal of Biological Chemistry*, vol. 256(20), pp. 10435-10441, 1981.
- [66] J. J. Manfredi, and S. B. Horwitz, "Taxol: an antimetabolic agent with a new mechanism of action," *Pharmacology & therapeutics*, vol. 25(1), pp. 83-125, 1984.
- [67] J. Parness, and S. B. Horwitz, "Taxol binds to polymerized tubulin in vitro," *The Journal of cell biology*, vol. 91(2), pp. 479-487, 1981.
- [68] W. C. Thompson, L. Wilson, and D. L. Purich, "Taxol induces microtubule assembly at low temperature," *Cell motility*, vol. 1(4), pp. 445-454, 1981.
- [69] J. R. Roberts, E. K. Rowinsky, R. C. Donehower, J. O. E. L. Robertson, and D. C. Allison, "Demonstration of the cell cycle positions of taxol-induced "asters" and "bundles" by sequential measurements of tubulin immunofluorescence, DNA content, and autoradiographic labeling of taxol-sensitive and-resistant cells," *Journal of histochemistry & cytochemistry*, vol. 37(11), pp.1659-1665, 1989.
- [70] D. B. McKAY, "Structure-activity study on the actions of taxol and related taxanes on primary cultures of adrenal medullary cells," *Journal of Pharmacology and Experimental Therapeutics*, vol. 248(3), pp. 1302-1307.1989.
- [71] F. Gueritte-Voegelein, D. Guenard, F. Lavelle, M. T. Le Goff, L. Mangatal, and P.Potier, "Relationships between the structure of taxol analogs and their antimetabolic activity," *Journal of medicinal chemistry*, vol. 34(3), pp. 992-998, 1991.
- [72] J. Riandel, M. Jacrot, F. Picot, H. Beriel, C. Mouriquand, and P. Potier, "Therapeutic response to taxol of six human tumors xenografted into nude mice," *Cancer chemotherapy and pharmacology*, vol. 17(2), pp. 137-142, 1986.
- [73] V. Chaudhry, E. K. Rowinsky, S. E. Sartorius, R. C. Donehower, and D. R. Cornblath, "Peripheral neuropathy from taxol and cisplatin combination chemotherapy: clinical and electrophysiological studies," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 35(3), pp. 304-311, 1994.
- [74] D. P. Stull, T. A. Scales, R. Daughenbaugh, N. A. Jans, and D. T. Bailey, "Taxol®(Paclitaxel)," *Applied Biochemistry and Biotechnology*, vol. 54(1-3), pp. 133-140, 1995.
- [75] G. Sarosy, E. Kohn, D. A. Stone, M. Rothenberg, J. D. O. Jacob, Adamo, and E. Reed, "Phase I study of taxol and granulocyte colony-stimulating factor in patients with refractory ovarian cancer," *Journal of clinical oncology*, vol. 10(7), pp. 1165-1170, 1992.

- [76] A. I. Einzig, P. H. Wiernik, J. Sasloff, C. D. Runowicz, and G. L. Goldberg, "Phase II study and long-term follow-up of patients treated with taxol for advanced ovarian adenocarcinoma," *Journal of Clinical Oncology*, vol. 10(11), pp. 1748-1753, 1992.
- [77] G. N. Hortobagyi, F. A. Holmes, R. L. Theriault, and A. U. Buzdar, "Use of Taxol (paclitaxel) in breast cancer," *Oncology*, vol. 51(Suppl. 1), pp. 29-32, 1994.
- [78] A. D. Seidman, C. A. Hudis, G. Raptis, J. Baselga, D. Fennelly, and L. Norton, "Paclitaxel for breast cancer: the memorial Sloan-Kettering cancer center experience," *Breast Cancer*, vol. 11(3), 1997.
- [79] A. Seidman, B. Reichman, J. Crown, C. Begg, R. Heelan, T. Hakes, and C. Hudis, "Activity of taxol with recombinant granulocyte colony stimulating factor (GCSF) as first chemotherapy of patients with metastatic breast cancer," in *Proc ASCO*, vol. 11, pp. 59, 1992.
- [80] W. K. Murphy, F. V. Fossella, R. J. Winn, D. M. Shin, H. E. Hynes, H. M. Gross, and I. H. Krakoff, "Phase II study of taxol in patients with untreated advanced non-small-cell lung cancer," *JNCI: Journal of the National Cancer Institute*, vol. 85(5), pp. 384-388, 1993.
- [81] C. M. Tsai, K. T. Chang, R. P. Perng, T. Mitsudomi, M. H. Chen, C. Kadoyama, and A. F. Gazdar, "Correlation of intrinsic chemoresistance of non-small-cell lung cancer cell lines with HER-2/neu gene expression but not with ras gene mutations," *JNCI: Journal of the National Cancer Institute*, vol. 85(11), pp. 897-901, 1993.
- [82] D. S. Ettinger, D. M. Finkelstein, R. Sarma, and D. H. Johnson, "Phase II study of Taxol in patients (pts) with extensive-stage small cell lung cancer (SCLC): an Eastern Cooperative Oncology Group Study," in *Proceedings of American Society of Clinical Oncology*, vol. 12, pp. 329, 1993.
- [83] N. T. Sklarin, P. H. Wiernik, W. R. Grove, L. Benson, A. Mittelman, J. A. Maroun, and J. Jolivet, "A phase II trial of CI-921 in advanced malignancies. *Investigational new drugs*," vol. 10(4), pp. 309-312, 1992.
- [84] A. A. Forastiere, "Paclitaxel (Taxol) for the treatment of head and neck cancer," In *Seminars in oncology*, vol. 21 (5 Suppl 8), pp. 49-52, 1994.
- [85] W. Lorenz, H. J. Reimann, A. Schmal, P. Dormann, B. Schwarz, E. Neugebauer, and A. Doenicke, "Histamine release in dogs by Cremophor EL® and its derivatives: Oxethylated oleic acid is the most effective constituent," *Agents and actions*, vol. 7(1), pp. 63-67, 1977.

- [86] G. Cavaletti, G. Tredici, M. Braga, and S. Tazzari, "Experimental peripheral neuropathy induced in adult rats by repeated intraperitoneal administration of taxol," *Experimental neurology*, vol. 133(1), pp. 64-72, 1995.
- [87] R. B. Weiss, R. C. Donehower, P. H. Wiernik, T. Ohnuma, R. J. Gralla, D. L. Trump, and B. Leyland-Jones, "Hypersensitivity reactions from taxol", *Journal of clinical oncology*, vol. 8(7), pp.1263-1268, 1990.
- [88] P. H. Wiernik, E. L. Schwartz, J. J. Strauman, J. P. Dutcher, R. B. Lipton, and E. Paietta, "Phase I clinical and pharmacokinetic study of taxol," *Cancer research*, vol. 47(9), pp. 2486-2493, 1987.
- [89] B. Monsarrat, E. Mariel, S. Cros, M. Gares, D. Guenard, F. Gueritte-Voegelein, and M. Wright, "Taxol metabolism. Isolation and identification of three major metabolites of taxol in rat bile," *Drug metabolism and disposition*, vol. 18(6), pp. 895-901, 1990.
- [90] B. Monsarrat, P. Alvinerie, M. Wright, J. Dubois, F. Gueritte-Voegelein, D. Guenard, and E. K. Rowinsky, "Hepatic metabolism and biliary excretion of Taxol in rats and humans," *Journal of the National Cancer Institute. Monographs*, vol. 15, pp. 39-46, 1993.
- [91] B. Monsarrat, M. Wright, J. Dubois, E. Gueritte-Voegelien, D. Guenard, E. K. Rowinsky, and R. C. Donehower, "Taxol metabolism in rat and human bile," *Annals of Oncology*, vol. 3(Suppl 1), pp. 123, 1992.
- [92] A. Rahman, K. R. Korzekwa, J. Grogan, F. J. Gonzalez, and J. W. Harris, "Selective biotransformation of taxol to 6 α -hydroxytaxol by human cytochrome P450 2C8," *Cancer research*, vol. 54(21), pp. 5543-5546, 1994.
- [93] S. M. Longnecker, R. C. Donehower, A. E. Cates, T. L. Chen, R. B. Brundrett, L. B. Grochow, and M. Colvin, "High-performance liquid chromatographic assay for taxol in human plasma and urine and pharmacokinetics in a phase I trial," *Cancer treatment reports*, vol. 71(1), pp. 53-59, 1987.
- [94] T. Tamura, Y. Sasaki, T. Shinkai, K. Eguchi, Y. Ohe, M. Nishio, and H. Nakashima, "Phase I and pharmacokinetic study of taxol by a 24-hour intravenous infusion," *In Proc Am Soc Clin Oncol*, vol. 12, pp. 143, 1993.
- [95] T. A. Willey, E. J. Bekos, R. C. Gaver, G. F. Duncan, L. K. Tay, J. H. Beijnen, and R. H. Farmen, "High-performance liquid chromatographic procedure for the quantitative determination of paclitaxel (Taxol®) in human plasma," *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 621(2), pp. 231-238, 1993.

- [96] P. H. Wiernik, E. L. Schwartz, A. I. Einzig, J. J. Strauman, R. B. Lipton, and J. P. Dutcher, "Phase I trial of taxol given as a 24-hour infusion every 21 days: responses observed in metastatic melanoma," *Journal of Clinical Oncology*, vol. 5(8), pp. 1232-1239, 1987.
- [97] P. H. Wiernik, E. L. Schwartz, J. J. Strauman, J. P. Dutcher, R. B. Lipton, and E. Paietta, "Phase I clinical and pharmacokinetic study of taxol," *Cancer research*, vol. 47(9), pp. 2486-2493, 1987.
- [98] H. Gelderblom, J. Verweij, K. Nooter, and A. Sparreboom, "Cremophor EL: the drawbacks and advantages of vehicle selection for drug formulation," *European journal of cancer*, vol. 37(13), pp. 1590-1598, 2001.
- [99] I. Ringel, and S. B. Horwitz, "Taxol is converted to 7-epitaxol, a biologically active isomer, in cell culture medium," *Journal of Pharmacology and Experimental therapeutics*, vol. 242(2), pp. 692-698, 1987.
- [100] T. Brown, K. Havlin, G. Weiss, J. Cagnola, J. Koeller, J. Kuhn, and D. Von Hoff, "A phase I trial of taxol given by a 6-hour intravenous infusion," *Journal of clinical oncology*, vol. 9(7), pp. 1261-1267, 1991.
- [101] D. L. Howrie, R. J. Ptachcinski, B. P. Griffith, R. J. Hardesty, J. T. Rosenthal, G. J. Burckart, and R. Venkataramanan, "Anaphylactoid reactions associated with parenteral cyclosporine use: possible role of Cremophor EL," *Drug intelligence & clinical pharmacy*, 19(6), 425-427, 1985.
- [102] D. M. Woodcock, M. E. Linsenmeyer, G. Chojnowski, A. B. Kriegler, V. Nink, L. K. Webster, and W. H. Sawyer, "Reversal of multidrug resistance by surfactants," *British journal of cancer*, vol. 66(1), pp. 62, 1992.
- [103] J. Rizzo, C. Riley, D. von Hoff, J. Kuhn, J. Phillips, and T. Brown, "Analysis of anticancer drugs in biological fluids: determination of taxol with application to clinical pharmacokinetics," *Journal of pharmaceutical and biomedical analysis*, vol. 8(2), pp. 159-164, 1990.
- [104] A. Albert, "Chemical aspects of selective toxicity," *Nature*, vol. 182(4633), pp. 421-423, 1958.
- [105] Z. Meng, Q. Lv, J. Lu, H. Yao, X. Lv, F. Jiang, and G. Zhang, "Prodrug strategies for paclitaxel", *International journal of molecular sciences*, vol. 17(5), pp. 796, 2016.
- [106] V. P. Podduturi, I. B. Magana, D. P. O'Neal, and P. A. Derosa, "Simulation of transport and extravasation of nanoparticles in tumors which exhibit enhanced permeability

and retention effect,” *Computer methods and programs in biomedicine*, vol. 112(1), pp. 58-68, 2013.

[107] G. Samaranayake, N. F. Magri, C. Jitrangri, and D. G. Kingston, “Modified taxols. 5. Reaction of taxol with electrophilic reagents and preparation of a rearranged taxol derivative with tubulin assembly activity,” *The Journal of Organic Chemistry*, vol. 56(17), pp. 5114-5119, 1991.

[108] Y. Ueda, J. D. Matiske, A. B. Mikkilineni, V. Farinal, J. O. Knipe, W. C. Roset, A. M. Casazzat, D. M. Vyas, “Novel, water-soluble phosphate derivatives of 2'-ethoxy carbonylpaclitaxel as potential prodrugs of paclitaxel: Synthesis and antitumor evaluation,” *Bioorganic and Medicinal Chemistry Letters*, vol. 5, pp. 247-252, 1995.

[109] E. W. P. Damen, P. H. G. Wiegerinck, L. Braamer, D. Sperling, D. de Vos, H. W. Scheeren, “Paclitaxel esters of malic acid as prodrugs with improved water solubility,” *Bioorganic Medicinal Chemistry*, vol. 8, pp. 427-432, 2000.

[110] A. R. Wohl, A. R. Michel, S. Kalscheuer, C. W. Macosko, J. Panyam, T. R. Hoye, “Silicate esters of paclitaxel and docetaxel: Synthesis, hydrophobicity, hydrolytic stability, cytotoxicity, and prodrug potential,” *Journal of Medicinal Chemistry*, vol. 57, pp. 2368-2379, 2014.

[111] N. Adams, and U. S. Schubert, “From data to knowledge: chemical data management, data mining, and modeling in polymer science,” *Journal of combinatorial chemistry*, vol. 6(1), pp. 12-23, 2004.

[112] P. Gramatica, “Principles of QSAR models validation: internal and external,” *QSAR & combinatorial science*, vol. 26(5), pp. 694-701, 2007.

[113] T. Puzyn, J. Leszczynski, and M. T. Cronin, “Recent advances in QSAR studies: methods and applications,” *Springer Science & Business Media*, vol. 8, 2010.

[114] C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, F. Geiger, and M. Streich, “The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients,” *Journal of the American Chemical Society*, vol. 85(18), pp. 2817-2824, 1963.

[115] O. OECD, “Principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models,” in *37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology, Organisation for Economic Co-Operation and Development*, Paris, France, 2004.

- [116] T. Puzyn, A. Mostrag-Szlichtyng, A. Gajewicz, M. Skrzynski, and A. P. Worth, "Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models," *Structural Chemistry*, vol. 22(4), pp. 795-804, 2011.
- [117] A. P. Toropova, A. A. Toropov, E. Benfenati, D. Leszczynska, and J. Leszczynski, "QSAR model as a random event: a case of rat toxicity," *Bioorganic & medicinal chemistry*, vol. 23(6), pp. 1223-1230, 2015.
- [118] P. Gramatica, "On the development and validation of QSAR models," In *Computational toxicology*, pp. 499-526, Humana Press, Totowa, NJ, 2013.
- [119] A. Tropsha, and A. Golbraikh, "Predictive QSAR modeling workflow, model applicability domains, and virtual screening," *Current pharmaceutical design*, vol. 13(34), pp. 3494-3504, 2007.
- [120] A. Frish, A. Nielsen, and A. Holder, "Gauss view user manual. Gaussian Inc.," Pittsburgh, PA, 2001.
- [121] R. A. Kwiecien, M. Rostkowski, A. Dybała-Defratyka, and P. Paneth, "Validation of semiempirical methods for modeling of corrinoid systems," *Journal of Inorganic Biochemistry*, vol. 98, pp. 1078-1086, 2004.
- [122] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "Dragon software: An easy approach to molecular descriptor calculations," *Match*, vol. 56, pp. 237-248, 2006.
- [123] R. R. Bouckaert, E. Frank, and M. A. Hall, et al., "WEKA: Experiences with a Java Open-Source Project," *Journal of Machine Learning Research*, vol. 11, pp. 2533-2541, 2010.
- [124] MATLAB Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.
- [125] S. Wold, L. Eriksson, "In Chemometric methods in molecular design van de Waterbmd", H., ed.; VCH: Weinheim, pp. 309, 1995.
- [126] <http://Immd.ecust.edu/admetsar2>.
- [127] D. S. Goodsell, G. M. Morris, and A. J. Olson, "Automated docking of flexible ligands: applications of AutoDock," *Journal of Molecular Recognition*, vol. 9(1), pp. 1-5, 1996.
- [128] O. Trott, and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of computational chemistry*, vol. 31(2), pp. 455-461, 2010.
- [129] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, Flexible, and Free," *Journal of Computational Chemistry*, vol. 26(16), pp. 1701-1718, 2005.

[130] H. Hong, M. Chen, H. W. Ng, and W. Tong, W. “QSAR models at the US fda/nctr,” in *In Silico Methods for Predicting Drug Toxicity*, pp. 431-459, Humana Press, New York, NY, 2016.

CHAPTER 2

DEVELOPMENT OF NOVEL QSPR STRATEGY FOR SOLUBILITY PREDICTION

SUMMARY

Paclitaxel, a nanomolar active anti-cancer drug with a major limitation of poor solubility had been studied from years and several prodrugs with better solubility than paclitaxel had been synthesized. Prodrug design is a widely known molecular modification strategy that aims to optimize the physicochemical and pharmacological properties of drugs to improve their solubility and pharmacokinetic features and decrease their toxicity. Theoretical models such as Quantitative Structure Property-Relationship (QSPR) models and data mining methods are efficient techniques that correlate molecular characteristics with physicochemical properties of molecules. However, the availability of physicochemical parameters for all molecules is difficult, molecular descriptors determined from these structures were explored for the development of QSPR models. Therefore, a QSPR model can provide a set of predictors for any molecular property using the structure of the molecule. With the advent of computational approaches the rational solution for the prodrug formation of paclitaxel has been attempted. Robust QSPR models for the solubility prediction are formed using robust statistical measures AIC and VIF multicollinearity indicators for the selection of independent and significant descriptors. Geometry optimization of these paclitaxel prodrugs was performed at the parametrization method 6 (PM6) and Austin model 1 (AM1) levels of geometry optimization by the Gaussian software. Four descriptor groups, 2D Autocorrelation, CATS_3D, WHIM, and GETAWAY administered initial QSPR models with an average accuracy for both the optimized geometry datasets. Multi-tiered descriptor selection process was performed for obtaining the most favourable models which contain five and four descriptors for PM6 and AM1 optimized geometry datasets respectively. Thus the formed QSPR pipeline (QSPR-sPL) was implemented for the formation of QSPR models and obtained R^2 & Q^2 values were 0.86 & 0.83 and 0.87 & 0.86 for PM6 and AM1 geometries respectively. QSPR-sPL was also implemented on *Huuskonen* small dataset and the final formed QSPR model had only two descriptors with the R^2 and Q^2 values 0.87 and 0.85, respectively which was found to be comparable to the results of paclitaxel prodrugs dataset. This approach can be applicable to different datasets for the solubility prediction which further can guide the synthesis of molecule with better solubility.

2.1 Introduction

Aqueous solubility is one of the important physicochemical properties that in turn affect the bioavailability of a drug by the way of dissolution and absorption. Efficient drug formulation carries a risk due to the poor solubility of the drug. Lead molecules with superior solubility generally show good pharmacokinetic properties; thereby increasing the chance for providing better outcomes in clinical trials. General solubility equation (GSE) and Aqueous Functional group Activity Coefficient (AQUAFAC) [1,2] are two of the most popular aqueous solubility prediction models presently available in the literature but they are of limited use as they utilize experimental data for the solubility prediction. Quantitative Structure Property Relationship (QSPR) models and data mining methods are efficient techniques that are able to correlate molecular characteristics with the physicochemical properties of molecules [3]. There are QSPR models those are efficient in the prediction of aqueous solubility of a diverse set of organic compounds [4-14]. Also, for the solubility prediction of drug-like organic molecules the practiced QSPR model gave a correlation coefficient of 0.99 [15]. Though the described QSPR models for solubility prediction were of excellent accuracy, but the descriptors/parameters used to develop these models were dependent upon the compound series [16]. Positive results were given by the models those were mostly developed on large number of molecules [6,15] and the prediction accuracy of the QSPR model got scaled down significantly when a model is developed by utilizing a small size dataset [17]. But taking into account the pharmaceutical perspective, a QSPR model for the solubility prediction needs to be developed from the small dataset to assist the synthesis of better soluble molecules in future [18,19]. In this objective thesis emphasis on the development of a robust QSPR strategy (QSPR-sPL) for the solubility prediction using the solubility data of twenty two paclitaxel prodrugs.

Paclitaxel, a western yew isolated in 1971 from *Taxus brevifolia* [20]. (Figure 2.1) is a complex diterpene which provides antitumor activity against ovarian, breast, lung and prostate cancers [21]. The anticancer activity of this drug emanates from its capability to stabilize the microtubule by promoting its assembly that in turn arrest cell cycle at the G₂/M phase, that also develops organizational and functional alteration of nuclear envelope and bring cell death [22]. Despite this, use of this drug is limited because of poor aqueous solubility, and hence lesser bioavailability. It is commonly administered intravenously in a vehicle containing cremophor EL (polyethoxylated castor oil) that leads to many significant

side-effects like severe anaphylactoid hypersensitive reactions, hyperlipidemia, aggregation of erythrocytes, etc. [23].

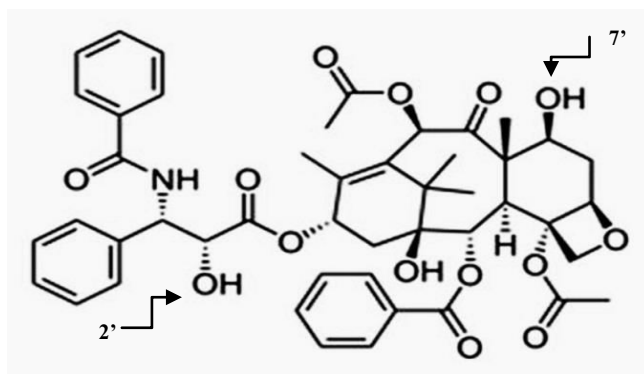


Figure 2.1: Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility.

Recognizing poor solubility as a major obstacle for this highly active drug, many prodrugs for increasing the solubility of paclitaxel were synthesized by the addition of hydrophilic groups such as carboxylic acids, phosphates, sugar derivatives, etc. [24].

Structure Activity-Relationship (SAR) study of paclitaxel showed that 2' and/or 7' hydroxyl groups in the structure as the potential sites for the addition of hydrophilic groups (Figure 2.1). Though increase in water solubility of paclitaxel prodrugs was obtained [25-29], but the advent of formation of the prodrugs was not judicious as in many cases undesirable compounds were formed and also it is a very time consuming process. A QSPR model can guide the synthesis of paclitaxel prodrugs with better solubility. Though, there is difficulty in finding the physicochemical parameters for all molecules hence, molecular descriptors determined from the structures were explored for the development of QSPR models. The study delved into the development of QSPR-sPL that is implied for the formation of QSPR models for the solubility prediction from a small dataset of 22 paclitaxel prodrugs (Appendix A, Table A.1). Dragon molecular descriptors were extracted for each prodrug molecule and the multicollinearity indicators like Akiake Information Criteria (AIC) and Variance Inflation Factor (VIF) [30] were implied for the extraction of independent descriptors. The developed QSPR models for the prediction of aqueous solubility were validated using 4-fold cross-validation.

2.2 Material and methods

2.2.1 Solubility data collection

Structures of twenty two diverse paclitaxel prodrugs were collected from the literature having either better or comparable solubility with the paclitaxel [25-29]. Initial coordinates of paclitaxel were taken from PubChem (www.ncbi.nlm.nih.gov) and the structures of different prodrugs were drawn in GaussView [31].

2.2.2 Geometry optimization

Despite the experimental advents, computational approaches those of quantum mechanical calculations are adopted for determining the microscopic properties of the molecules. A molecule is characterized by the combination of electronic wave functions which represents each atom forming the molecule. Several parameters such as radial and angular parts of the molecule which were further defined by bond angles, bond lengths, and dihedral angles of rotation about single bond of the molecule forms the electronic wave function of a polyatomic molecule. Energy of the molecule is defined by the Schrodinger wave equation ($H\psi = E\psi$). Different energies are found for the molecules of different configurations. Molecular energies and properties are majorly calculated by means of four methods: molecular mechanics, *ab-initio*, semi-empirical and density-functional theory (DFT) methods. Semi-empirical methods make use of simpler approximate Hamiltonian operator, and uses empirical parameters whose values are adjusted to fit the experimental values in contrary to the *ab-initio* calculations that are based on the correct Hamiltonian without making use of the experimental data. The principle behind the DFT is based on the electron probability density, ρ and it is used for the calculation of molecular electronic energy. The DFT method uses wave function that includes fewer variables for the calculation of energy and other properties. Molecular mechanics methods recognize the molecule as a set of atoms and articulated the molecular energy as a total of bond bending, and stretching energies.

Basis function

Electronic orbital or electronic wave function of atoms/molecules is represented by the mathematical function known as *basis set*. These functions are used in Hartee-Fock method and density function theory methods for the conversion of partial differential equation generated from a molecule into algebraic equations for the appropriate effective implementation on the computer. There are several types of atomic orbitals such as Slatter type, Gaussian-type, numerical and different classes of basis sets such as minimal, split valence, correlation-consistent, pople basis, plane-wave, Karlsruhe etc. Minimum energy configuration of a molecule from different configurations is defined as energy minimization. Mainly, the algorithms used for the geometry optimization are steepest descent and conjugate gradient algorithm. Steepest descent is used in the initial steps of the geometry optimization

and afterwards conjugate gradient method is employed for obtaining the global energy minima of the structures. All positive vibrational frequencies illustrate the global energy minima for the structures. Semi-empirical methods are based on three approximations; the elimination of the core electrons from the calculation, the use of minimum number of basis sets, and the reduction of number of two-electron integrals. Semi-empirical methods are efficient methods for obtaining energy minimized structures for larger systems.

In the present study, due to the high molecular weight of the paclitaxel structures of all the drawn molecules were optimized at the levels of parametrization method 6 (PM6) and austin model 1 (AM1) [32] using Gaussian 09 quantum chemistry software [33]. A global energy minimum of each structure was verified by analyzing vibrational frequencies. The energy minimized structures were further used for the descriptor extraction.

2.2.3 Structural descriptor by DRAGON 7

Molecular descriptors are being used for the prediction of biological or physiochemical properties of the molecules implying QSAR/QSPR and for the automatic evaluation of molecular libraries. Descriptors are obtained by treating the molecules as real objects, and translated into a molecular representation that enables mathematical regimen. The information content of the molecular descriptors obviously depends on the character of its representation and also on the algorithm used for the calculation of descriptors. All the molecular descriptors should contain chemical information, should follow some invariance properties and should be determined from well-established procedure so that molecular descriptors could be extracted for any kind of datasets.

It is trivial to assume that a single descriptor or a small set of molecular descriptor will be able to describe a whole molecule. DRAGON 7 software [34] is adept to provide variety of descriptors by making use of different molecular representations. DRAGON 7 allows the calculation of 5250 descriptors that includes simple molecular descriptors determined by simply counting different atom types or structural fragments of the molecules 1-dimensional (1D), additionally descriptors derived by algorithm that is applied on the topological representation of the molecules are mostly called 2-dimensional (2D) descriptors. Also, there are descriptors derived from the spatial arrangement of molecule known as to be geometrical or 3-dimensional (3D) descriptors. All the 5250 descriptors derived from DRAGON 7 belong to thirty different molecular descriptor groups (Appendix A, Table A.2). All the 1D, 2D, and 3D descriptors were calculated for each paclitaxel prodrug using the geometry optimized structures.

2.2.4 Structural descriptor reduction

The descriptors were obtained for the dataset in thirty different descriptor groups which were not reduced using the standard protocol (mainly the reductionist approach) but were reduced implementing multicollinearity indicators, following the formation of QSPR models. The reductionist approach was not able to provide the efficient QSPR models. We had observed that the QSPR models for both the geometry optimized datasets were inconsistent as compared to the results of some individual descriptor groups. For the descriptor reduction, the standard protocol was not followed in exact sense. The approach described here identified only significant and unbiased descriptors through feature reduction and selection in stepwise manner that helps to optimize the model.

In statistics, the dependence between two random variables commonly refers to the degree of linear relationship between the variables is measured by the correlation and the measure of a set of number from its average termed as variance, were taken into consideration for the descriptors in an individual descriptor group and on that basis each descriptor group was segregated into eight independent subgroups: “positively very high correlated with high variance (pvhchv), positively very high correlated with low variance (pvhclv), positively high correlated with high variance (phchv), positively high correlated with low variance (phclv), positively moderately correlated with high variance (pmchv), positively moderately correlated with low variance (pmclv), zero correlated with high variance (zchv) and zero correlated with low variance (zclv)”. The compilation of these eight subgroups for each individual group was executed and accomplished in R-3.2.2 (cran.r-project.org). Descriptors showing no variance with solubility property were expelled from the pool of descriptors for the dataset [35,36]. Thereafter, on each sub-group of each descriptor group akaike information criteria (AIC) [37] was applied that eliminated the most dependent response indicators from each sub-group. AIC follows penalized-likelihood criteria. It is commonly used for choosing the best predictor subsets in regression analysis and is frequently used for comparing models, which cannot be obtained using ordinary statistical tests. The AIC for a model is customarily written in the mathematical form as, $[-2\log L + kp]$, where L is the likelihood function, p is the number of parameters in the model, and k is 2 for AIC. It was calculated for each subgroup; also ‘p’ value is calculated for each sub-group. All the subgroups in each descriptor groups were then optimized for the value of $p < 0.05$ and thus the predictor variables were determined. The process was stopped when the p-value of response variable becomes less than the value of intercept. Various other statistical characteristics for the models like coefficient of determination R^2 , fisher static F , and p value for each descriptor’s coefficient were also calculated.

The variance inflation factor (VIF) multi-collinearity indicator was also applied for the descriptor reduction and to predict only independent variables. VIF is based on square of the multiple correlation coefficients resulting from regression of a predictor variable against all other predictor variables. VIF for a variable would be large if that variable has strong linear relationship with at least one other variable and the correlation coefficient would be close to 1. In each sub-group VIF value less than 5 was taken and the descriptors which were showing value of more than 5 were discarded. The whole process of descriptor reduction is shown in Figure 2.2.

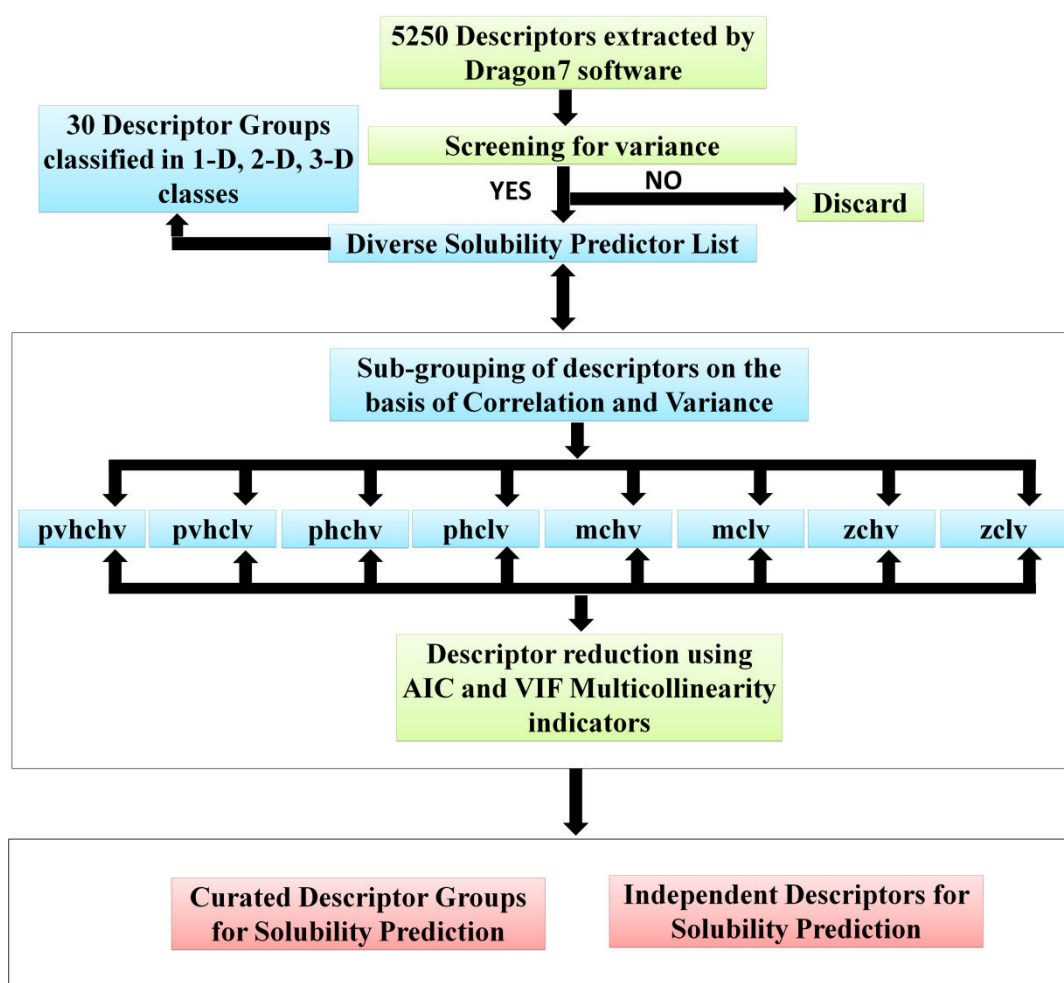


Figure 2.2: Stage-wise workflow of descriptor reduction

2.2.5 Non linear multiple regression study

The descriptors obtained from the DRAGON 7 that gave a pool of 5250 descriptors categorized in thirty different descriptor groups. Descriptor values for some of the descriptors were positive and others are negative. It indicated that the extracted structural descriptors those were having positive values vary linearly with $\log(e_{ly})$ (logarithm of solubility)

whereas descriptors with negative values vary exponentially with log (ely). Therefore, the following regression equation (Eqn. 1) was proposed:

$$\log(ely) = \log\beta_0 + \beta_p \log X_p + \beta_n X_n \quad \text{.. Eqn. 1}$$

where $\log\beta_0$ is a constant, β_p is the coefficient of positively valued descriptors, $\log X_p$ is the logarithm of descriptor value, β_n is the coefficient of negatively valued descriptor, and X_n is the negative descriptor value.

2.2.6 Formation of QSPR model

Descriptor reduction and selection advent was followed to identify independent descriptors for QSPR model development. WEKA-3.6.11 software [38] was implemented using classifier linear regression to perform the step-wise regression analysis for the formation of QSPR models from the selected descriptors for both PM6 and AM1 geometry optimized datasets. However, 2-fold, 3-fold, 4-fold and 5-fold cross-validations were observed for the QSPR models but 4-fold cross-validation gave R^2 and Q^2 values much closer for all analyses. Therefore, only 4-fold results were reported exclusively for all correlation (Q^2) analyses. To identify any chance correlation in the formed QSPR models Y-Randomization test was executed in QSPR models. DTC_yRandomization program (http://teqip.jdvu.ac.in/QSAR_Tools/) was practiced with the generation of ten random models and R^2_{yrand} & Q^2_{yrand} values were obtained for the QSPR models. This yRandomization approach is based on the obtained values of Q^2_{yrand} and R^2_{yrand} and their relationships as those given by Eriksson and Wold [39], defined in the following section:

- If, $Q^2_{\text{yrand}} < 0.2$ and $R^2_{\text{yrand}} < 0.2 \rightarrow$ no chance correlation;
- any Q^2_{yrand} and $0.2 < R^2_{\text{yrand}} < 0.3 \rightarrow$ negligible chance correlation;
- any Q^2_{yrand} and $0.3 < R^2_{\text{yrand}} < 0.4 \rightarrow$ tolerable chance correlation;
- any Q^2_{yrand} and $R^2_{\text{yrand}} > 0.4 \rightarrow$ recognized chance correlation.

For passing this test, the cR_p^2 parameter should be more than 0.5.

$$cR_p^2 = R^* (R^2 - (\text{Average } R_r)^2)^{1/2}$$

where "Average R_r "= average 'R' from random models.

2.3 Result and discussion

In the current study, the QSPR solubility prediction pipeline QSPR-sPL from a small dataset of paclitaxel prodrugs was developed by implementing AIC and VIF multicollinearity indicators for the identification of independent and significant descriptors. Subsequently, the

QSPR-sPL was implemented for both AM1 and PM6 geometry optimized datasets for the identification of independent descriptors. The QSPR models were formed and reviewed for any chance correlation. Initially, the models were developed on each individual thirty descriptor groups. The descriptor groups those were depicting successful models initially are described in Table 2.1. The significant descriptors from these descriptor groups were further combined and selected for the formation of optimal QSPR models (Table 2.2).

Individual thirty groups of descriptors were assessed for R^2 and Q^2 values. From the thirty descriptor groups only four groups namely 2D Autocorrelation, CATS_3D, WHIM, and GETAWAY were showing correlation with solubility (Table 2.1, Fig. 2.3(a-d)). 2D Autocorrelation descriptor group provided favourable results for both the optimized geometries though descriptors from GETAWAY and WHIM groups showed good correlations with solubility for only AM1 and PM6 optimized geometry structures respectively.

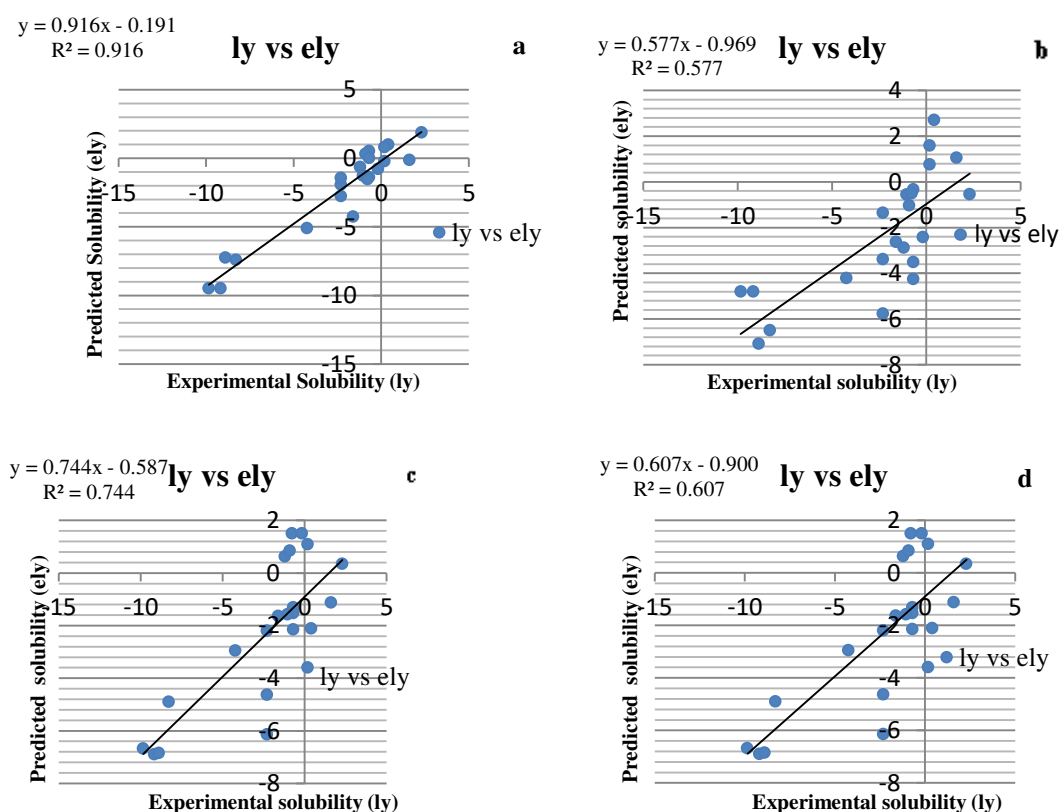


Figure 2.3: Experimental solubility (ly) vs predicted solubility (ely) of paclitaxel prodrugs. a. Performance of 2D Autocorrelation descriptors with AM1 geometry dataset; b. Performance of GETAWAY descriptors with AM1 geometry dataset; c. Performance of 2D Autocorrelation descriptors with PM6 geometry dataset; and d. Performance of WHIM descriptors with PM6 geometry dataset.

Table 2.1: Regression (R^2) and 4-fold cross validation correlation (Q^2) coefficients of the QSPR model developed using Dragon molecular descriptor groups

S. No.	Descriptor Group	AM1 optimized geometry dataset				PM6 optimized geometry dataset			
		No. of descriptors [#]	R ² / [§] R ² _{yrand}	Q ² / [§] Q ² _{yrand}	RMSE	No. of descriptors [#]	R ² / [§] R ² _{yrand}	Q ² / [§] Q ² _{yrand}	RMSE
1	2D Autocorrelation	10 (7)	0.91/0.31	0.62/-0.82	3.39	6 (2)	0.74/0.06	0.72/-0.23	2.07
2	CATS_3D	5 (1)	0.37/0.04	0.30/-0.17	4.78	5 (3)	0.62/0.13	0.30/-0.36	3.38
3	GETAWAY	8 (3)	0.57/0.12	0.65/-0.35	2.61	13 (3)	0.37/0.13	0.15/-0.28	3.73
4	WHIM	5 (1)	0.23/0.03	0.15/-0.17	3.75	5 (2)	0.60/0.05	0.65/-0.25	2.68

[#]QSPR equation indicating descriptors are provided in Appendix A (Eqn A1-A8). Parenthesis indicates the no. of significant descriptors included in QSPR model.

[§]R²_{yrand} & Q²_{yrand} values are calculated to determine statistical significance of the model (Section 2.2.6).

Measured R² & Q² values for the respective descriptor groups (Table 2.1) was average; hence QSPR model for paclitaxel prodrugs was developed by applying the QSPR-sPL on the whole pool of 5250 descriptors belonging to three descriptor classes (1D, 2D and 3D). Individual descriptor class was segregated into eight subgroups (pvhchv, pvhclv, phchv, phclv, pmchv, pmclv, zchv and zclv) on the basis of correlation and variance among the descriptors. Thereafter, the selected descriptors from same subgroup those belong to three different classes were merged to determine eight subgroups of descriptors. Then for selecting independent and significant descriptors, AIC and VIF multicollinearity indicators were applied on each subgroups. The developed QSPR models from the significant descriptors gave R² and Q² values 0.90 & 0.93 and 0.66 & 0.40 for PM6 and AM1 optimized geometries respectively (Appendix A, Table A.3). Thereby, the obtained lower values of Q² (0.66 & 0.40) signified that the formed QSPR model was erratic if compared with the obtained results from some individual descriptor groups (Table 2.1). Thereof, the independent and significant descriptors from four descriptor groups; 12 for AM1 and 10 for PM6 geometry optimized structures respectively were merged. Subsequently, from these descriptors significant descriptors were selected by removing the dependent descriptors implementing the multicollinearity indicators. The formed QSPR models attained reasonably better values of R² & Q², but the accuracy achieved were proportionate with that obtained by significant 2D Autocorrelation descriptors (Table 2.1 & Appendix A, Table A.4).

QSPR models lacking the reasonable Q² values can be explained as a result of low correlation coefficients of CATS_3D, WHIM (in AM1 optimized geometry) and GETAWAY (in PM6 optimized geometry) descriptor groups (Table 2.1). Accordingly, only descriptors those were providing reasonable correlation were used for the formation of two different QSPR models.

2D Autocorrelation descriptors were providing good correlation for both the geometry optimized datasets. The selected 2D Autocorrelation descriptors were combined with significant GETAWAY and WHIM descriptors for AM1 and PM6 geometry optimized structures respectively and the QSPR model was determined. The formed resultant models gave reasonable correlations for both geometry optimized datasets (Table 2.2, Fig. 2.4(a-b)). Further for obtaining more consistent and optimal QSPR models, 11 significant descriptors sorted from above step (combination descriptors from AM1 and PM6 geometries) were benefitted, and the regression (R^2) and correlation (Q^2) coefficients for these descriptors were calculated for both geometries. Rational QSPR models with equitable regression and correlation coefficients were determined (Model I & II, Table 2.2 and Figure 2.4(c-d)). Moreover yRandomization results indicated the statistical significance of these models.

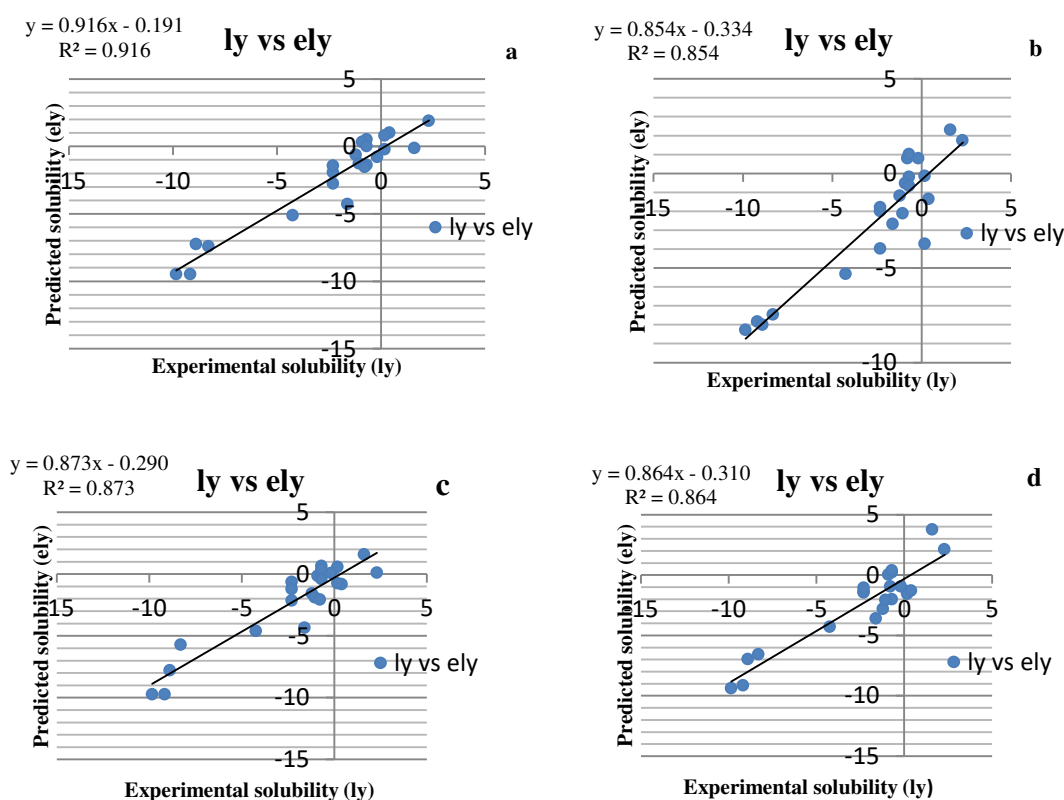


Figure 2.4: Experimental solubility (ly) vs predicted solubility (ely) of paclitaxel prodrugs. a. AM1 optimized geometry dataset with ten 2D Autocorrelation and GETAWAY descriptors; b. PM6 optimized geometry dataset with four 2D Autocorrelation and WHIM descriptors; c. AM1 optimized geometry dataset with 11 descriptors from a and b; d. PM6 optimized geometry dataset with 11 descriptors from a and b.

Table 2.2: Regression (R^2) and 4-fold cross validation correlation (Q^2) coefficients of the QSPR model determined through combination of descriptors

S. No.	AM1 optimised geometry dataset	PM6 optimised geometry dataset
--------	--------------------------------	--------------------------------

	No. of descriptors [%]	^{\$} R ²	^{\$} Q ²	RMSE	No. of descriptors [%]	^{\$} R ²	^{\$} Q ²	RMSE
1	^a 10 (7)	0.92	0.82	2.45	^b 4 (4)	0.85	0.80	2.11
2.	^c 11 (4)	0.87/0.19	0.86/-0.38	1.79	^c 11 (5)	0.86/0.29	0.83/-0.56	1.91

[%]QSPR equations indicating the descriptors are provided in Appendix A (Eqn. A11-A12) and Model I & II. Parenthesis indicates the no. of significant descriptors in QSPR equation. These selected descriptors were used to develop enhanced QSPR model. ^a Seven 2D Autocorrelation and three GETAWAY descriptors from Table 2.1 were used; ^b Two 2D Autocorrelation and two WHIM descriptors from Table 2.1 were used; ^c Seven and four selected descriptors explained in previous two steps (^a & ^b), respectively are used to make QSPR model.

^{\$}R²_{yrand} & Q²_{yrand} values are calculated to determine statistical significance of the model (Section 2.2.6).

QSPR Models:

$$ly = -9.5239 - 172.1078 * MATS4i - 38.9269 * MATS8e + 25.3947 * MATS7s \\ + 1155.1885 * JGI8$$

.....Model I

$$ly = 111.2416 + 45.625 * MATS7e - 239.2606 * MATS4i - 48.2859 * GATS3p \\ - 50.5689 * GATS5i - 5.94 * Kv$$

.....Model II

To validate the efficacy of proposed pipeline QSPR-sPL, *Huuskonen* dataset [7] comprising twenty one molecules was taken and the whole pipeline was followed for the solubility prediction and QSPR model was developed. Descriptors obtained from individual descriptor groups were analyzed and for the dataset five groups provided significant R² and Q² values (Table 2.3).

Table 2.3: Regression (R²) and validation correlation (Q²) coefficients of the QSPR model for *Huuskonen* dataset

S. No.	Descriptor Group	No. of descriptors ^{\$}	[*] R ²	[*] Q ²	RMSE
1	P_VSA	3 (2)	0.87 (0.08)	0.85 (-0.26)	0.95
Results of individual descriptor groups					
1	P_VSA-like descriptors	55 (2)	0.87	0.83	0.952
2	3D autocorrelations	80 (3)	0.91	0.81	0.816
3	RDF descriptors	210 (6)	0.86	0.75	1.006
4	Molecular properties	20 (6)	0.91	0.80	0.819
5	Drug-like indices	28 (1)	0.77	0.75	1.262

*Values in italic indicate R²_{yrand} (0.08) & Q²_{yrand} (-0.26) values which are calculated using ten random models. The cR_p² value is 0.73 (Section 2.2.6). ^{\$}Parenthesis indicates the no. of significant descriptors included in QSPR model.

These descriptors were merged following the selection of independent and significant descriptors. Two suggestive descriptors from P_VSA descriptor group provided the QSPR model with R² and Q² values 0.87 and 0.85, respectively with model measures comparable to

earlier models (Table 2.2). yRandomization results also decipher no chance correlation for the formed QSPR model (Table 2.3).

In the literature various QSPR models for the solubility prediction, those based on different methodologies with diverse datasets were reported [8-13]. Results obtained in the current study were compared to previously reported QSPR solubility models (Table 2.4). Klopman and Hou's solubility prediction model was trained on 1168 organic compounds with 118 descriptors [9]. Also, the solubility prediction model developed by McElroy and Jurs had descriptors from topological, geometrical, electronic, and hybrid group extracted from 298 compounds [10]. The proposed QSPR of Tetko *et al.* was formed by 879 diverse organic compounds and had 34 descriptors which were sorted from 94 descriptors and were related to electrotopological characteristics of the molecules [11]. Though Cheng and Merz QSPR model did not need 3D structures of the molecules, and required only 2D descriptors [12]. Similarly, Delaney QSPR approach did not necessitate 3D representation of the molecules and the model had only 4 descriptors describing the solubility property [13]. Hou *et al.* used 76 descriptors to build a QSPR model from 878 organic compounds with correlation coefficient 0.95 which is highest among all the proposed QSPR models [14]. Model developed by Ghasemi *et al.* used 110 drug organic compounds and it is based on four descriptors for the solubility prediction [15].

Table 2.4: Comparison among the results obtained from literature methods and the model formed in the study

Model	TdN*	dN*	Dataset (N)	R ²	Q ²	Data series*	Descriptor group*	Descriptor extracting software
Klopman and Hou	NA	118	1168	0.95	NA	Organic chemicals including drug like	Structural descriptors	MPAR
McElroy and Jurs	229	11	298	0.96	0.92	Organic compounds	Topological, Geometric, Electronic, Hybrid descriptors	ADAPT toolkit
Tetko <i>et al.</i>	94	34	879	0.93	0.89	Diverse organic compounds	Molecular connectivity, Shape, Flexibility, Indicator indices, E-state indices	NA

Cheng and Merz	139	8	775	0.89	0.83	Small organic compounds	MW and Electrotopological indices	Cerius package
Delaney	9	4	1144	NA	NA	Low MW organic compounds	Physiochemical properties	SMILES
Hou <i>et al.</i>	NA	76	878	0.96	0.95	Organic compounds	2D Molecular topology	Drug-LOGS
Ghasemi <i>et al.</i>	56	4	110	0.99	NA	Drug organic compounds	MW, HB, HLB, PSA	MMPP
Current study	5250	5	22	0.86	0.83	Paclitaxel prodrugs (PM6 geometry)	2D-Autocorrelation, WHIM	Dragon 7
Current study	5250	4	22	0.87	0.86	Paclitaxel prodrugs (AM1 geometry)	2D Autocorrelation	Dragon 7

*TdN: Total no. of descriptors; dN: Descriptors used for model formation; *MW: Molecular weight; NA: Not Available

The QSPR models proposed in the current study are comparably efficient as those present in the literature for the solubility prediction (Table 2.4). As compared to other methods, our optimal QSPR models required only 4 and 5 descriptors in Model I and Model II respectively for solubility prediction. The efficacy of the QSPR pipeline for solubility prediction was also evaluated through different measures. It was implemented on *Huuskonen* dataset, consists of twenty one molecules that provided similar result as obtained for paclitaxel prodrug dataset (Table 2.2 & 2.3). Twelve more paclitaxel prodrugs were also added to earlier datasets of 22 molecules (Appendix A, Table A.5). This dataset of thirty four prodrugs was divided into twenty two and twelve as training and test datasets, respectively. QSPR model was built on the training dataset and the model was evaluated on test dataset, the Q^2 values for independent test set were 0.80 and 0.60 for PM6 and AM1, respectively (Appendix A, Table A.6). These outcomes proved that proposed pipeline; QSPR-sPL is effective in the solubility prediction.

2.4 Conclusion

Aqueous solubility is a fundamental property for therapeutic efficacy of a drug. Studies done in the past revealed that the QSPR models are data series specific. Therefore, the current study proposed a QSPR pipeline, QSPR-sPL for the solubility prediction which is applicable to smaller dataset having potential to provide significant accuracy. Paclitaxel prodrugs were

subjected to the QSPR-Pac and the QSPR models with significant accuracies were developed, those were further validated for any chance correlation by the yRandomization test. The developed models were comparable to solubility QSPR models those were reported earlier. The proposed protocol was also implemented on *Huuskonen* small dataset and the final QSPR model was obtained with two descriptors. For this dataset the R^2 and Q^2 values were 0.87 and 0.85, respectively. The proposed QSPR pipeline, QSPR-Pac for the formation of QSPR model for solubility prediction can be applied to different datasets and that may provide promising result. Hence, it can be used to direct the prediction of solubility thus assist the synthesis of future molecule with better solubility.

References

- [1] Y. Ran, N. Jain, and S. H. Yalkowsky, "Prediction of aqueous solubility of organic compounds by the general solubility equation (GSE)," *Journal of Chemical Information and Computer Science*, vol. 41, pp. 1208-1217, 2001.
- [2] P. Myrdal, G. H. Ward, R. M. Dannenfelser, D. Mishra, and S. H. Yalkowsky, "AQUAFAC 1: Aqueous functional group activity coefficients; application to hydrocarbons," *Chemosphere*, vol. 24, pp. 1047-1061, 1992.
- [3] N. Jain, and S. H. Yalkowsky, "Estimation of the aqueous solubility I: Application to organic nonelectrolytes," *Journal of Pharmaceutical Sciences*, vol. 90, pp. 234-252, 2001.
- [4] T. Ghafourian, A. H. Bozorgi, "Estimation of drug solubility in water, PEG 400 and their binary mixtures using the molecular structures of solutes," *European Journal of Pharmaceutical Sciences*, vol. 40, pp. 430-440, 2010.
- [5] C. I. Cappelli, S. Manganelli, A. Lombardo, A. Gissi, and E. Benfenati, "Validation of quantitative structure-activity relationship models to predict water-solubility of organic compounds," *Science of the total Environment*, vol. 463, pp. 781-789, 2013.
- [6] J. Huuskonen, "Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology," *Journal of Chemical Information and Computer Science*, vol. 40, pp. 773-777, 2000.
- [7] J. Huuskonen, "Estimation of water solubility from atom-type electrotopological state indices," *Environmental Toxicology and Chemistry*, vol. 20, pp. 491-497, 2001.
- [8] J. Huuskonen, J. Rantanen, and D. Livingstone, "Prediction of aqueous solubility for a diverse set of organic compounds based on atom-type electrotopological state indices," *European Journal of Medicinal Chemistry*, vol. 35, pp. 1081-1088, 2000.

- [9] G. Klopman, and H. Zhu, "Estimation of the aqueous solubility of organic molecules by the group contribution approach," *Journal of Chemical Information and Computer Science*, vol. 41, pp. 439-445, 2001.
- [10] N. R. McElroy, and P. C. Jurs, "Prediction of aqueous solubility of heteroatom-containing organic compounds from molecular structure," *Journal of Chemical Information and Computer Scienc.*, vol. 41, pp. 1237-1247, 2001.
- [11] I. V. Tetko, V. Y. Tanchuk, T. N. Kasheva, and A. E. Villa, "Estimation of aqueous solubility of chemical compounds using E-state indices," *Journal of Chemical Information and Computer Science*, vol. 41, pp. 1488-1493, 2001.
- [12] A. Cheng, and K. M. Merz, "Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships," *Journal of Medicinal Chemistry*, vol. 46, pp. 3572-3580, 2003.
- [13] J. S. Delaney, "ESOL: estimating aqueous solubility directly from molecular structure," *Journal of Chemical Information and Computer Science.*, vol. 44, pp. 1000-1005, 2004.
- [14] T. Hou, K. Xia, W. Zhang, and X. Xu, "ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach," *Journal of Chemical Information and Computer Science*, vol. 44, pp. 266-275, 2004.
- [15] J. Ghasemi, and S. Saaidpour, "QSPR prediction of aqueous solubility of drug-like organic compounds," *Chemical and Pharmaceutical Bulletin*, vol. 55, pp. 669-674, 2007.
- [16] L. Du-Cuny, J. Huwyler, M. Wiese, and M. Kansy, "Computational aqueous solubility prediction for drug-like compounds in congeneric series," *European Journal of Medicinal Chemistry*, vol. 43, pp. 501-512, 2008.
- [17] M. Gonzalez, C. Teran, M. Teijeira, and M. Gonzalez-Moa, "GETAWAY descriptors to predicting A_{2A} adenosine receptors agonists" *European Journal of Medicinal Chemistry*, vol. 40, pp. 1080-1086, 2005.
- [18] S. Vahdani, and Z. Bayat, "A quantitative structure activity study of Anti-cancer Drugs," *Der Chemica Sinica*, vol. 2, pp. 235-243, 2011.
- [19] E. Baher, and N. Darzi, "QSAR studying of oxidation behavior of Benzoxazines as an important pharmaceutical property," *Iran Journal of Pharmaceutical Research*, vol. 6, pp. 146, 2017.
- [20] M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon, and A. T. McPhail, "Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*," *Journal of American Chemical Society*, vol. 93, pp. 2325-2327, 1971.

- [21] B. J. Roth, R. Dreicer, and L. H. Einhorn, et al., "Significant activity of paclitaxel in advanced transitional-cell carcinoma of the urothelium: a phase II trial of the Eastern Cooperative Oncology Group," *Journal of Clinical Oncology*, vol. 12, pp. 2264-2270, 1994.
- [22] F. Lavelle, "Taxoids: a new class of antimitotic agents," *Current Opinion in Investigational Drugs*, vol. 2, pp. 627-635, 1994.
- [23] E. K. Rowinsky, L. A. Cazenave, and R. C. Donehower, "Taxol: a novel investigational antimicrotubule agent," *Journal of National Cancer Institute*, vol. 82, pp. 1247-1259, 1990.
- [24] C. E. Muller, "Prodrug approaches for enhancing the bioavailability of drugs with low solubility" *Chemistry & Biodiversity*, vol. 6, pp. 2071-2083, 2009.
- [25] K. Nicolaou, C. Riemer, M. Kerr, D. Rideout, and W. Wrasidlo, "Design, synthesis and biological activity of protaxols," *Nature*, vol. 364, pp. 464-466, 1993.
- [26] K. C. Nicolaou, R. K. Guy, E. N. Pitsinos, and W. Wrasidlo, "A Water-Soluble Prodrug of Taxol with Self-Assembling Properties" *Angewandte Chemie International Edition (English)*, vol. 33, pp. 1583-1587, 1994.
- [27] R. B. Greenwald, A. Pendri, D. Bolikal, and C. W. Gilbert, "Highly water soluble taxol derivatives: 2'-polyethyleneglycol esters as potential prodrugs", *Bioorganic Medicinal Chemistry Letters*, vol. 4, pp. 2465-2470, 1994.
- [28] R. B. Greenwald, C. W. Gilbert, A. Pendri, C. D. Conover, J. Xia, and A. Martinez, "Drug delivery systems: water soluble taxol 2'-poly (ethylene glycol) ester prodrugs design and in vivo effectiveness," *Journal of Medicinal Chemistry*, vol. 39, pp. 424-431, 1996.
- [29] A. A. Moosavi-Movahedi, S. Hakimelahi, and J. Chamani, et al., "Design, synthesis, and anticancer activity of phosphonic acid diphosphate derivative of adenine-containing butenolide and its water-soluble derivatives of paclitaxel with high antitumor activity", *Bioorganic Medicinal Chemistry*, vol. 11, pp. 4303-4313, 2003.
- [30] M. Cama, I. Cristi Nicu, C. Conoscenti, G. Queneherve, and M. Maerker, "The role of multicollinearity in landslide susceptibility assessment by means of Binary Logistic Regression: comparison between VIF and AIC stepwise selection," *EGU General Assembly Conference Abstracts*, pp. 16839, 2016.
- [31] A. Frish, A. Nielsen, and A. Holder, "Gauss view user manual. Gaussian Inc," Pittsburgh, PA, 2001.
- [32] R. A. Kwiecien, M. Rostkowski, A. Dybała-Defratyka, and P. Paneth, "Validation of semiempirical methods for modeling of corrinoid systems," *Journal of Inorganic Biochemistry*, vol. 98, pp. 1078-1086, 2004.

- [33] M. Frish, G. Trucks, and H.Schiegel, "Gaussian 09 [CP]. Revision A. 02," *Wallingford CT: Gaussian*, 2009.
- [34] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "Dragon software: An easy approach to molecular descriptor calculations", *Match*, vol. 56, pp. 237-248, 2006.
- [35] L. R. Taylor, "Aggregation, variance and the mean," *Nature*, vol. 189(4766), pp. 732-735, 1961
- [36] H. Scheffe, "The analysis of variance," vol. 7, John Wiley & Sons, 1999.
- [37] W. Pan, "Akaike's information criterion in generalized estimating equations. *Biometrics*," vol. 57(1), pp. 120-125, 2001.
- [38] R. R. Bouckaert, E. Frank, and M. A. Hall, et al., "WEKAâExperiences with a Java Open-Source Project," *Journal of Machine Learning Research*, vol. 11, pp. 2533-2541, 2010.
- [39] S. Wold, L. Eriksson, "In Chemometric methods in molecular design van de Waterbmd", H., ed.; VCH: Weinheim, pp. 309, 1995.

CHAPTER 3

QSPR MODELS FOR THE SOLUBILITY PREDICTION OF PACLITAXEL PRODRUGS AND *IN-SILICO* MODELS FOR THE METABOLIC STUDIES OF HYDROPHILLIC SUBSTITUENT GROUPS

SUMMARY

Paclitaxel drug is directed for the treatment of various cancers such as ovarian, breast and also lung cancer. Regardless of being a nanomolar active drug, there is limited use of paclitaxel as it has poor aqueous solubility. As a result of this limitation many prodrugs of paclitaxel with increased solubility were formed in the past but synthesis process was not rational. The QSPR pipeline developed for the solubility prediction is used for the solubility prediction of paclitaxel prodrugs. In this study, eight QSPRs with statistical significance for four different datasets were developed for the solubility prediction. All the molecules were geometry optimized at PM6 and AM1 levels, followed by descriptor extraction from Dragon 7 that is giving a pool of 5250 descriptors. Additionally, quasi-mixture descriptors were extracted for the same dataset. For the formation of parsimonious QSPRs only independent descriptors those were selected using the pipeline developed in the previous study and QSPR models were obtained with 12 and 10 descriptors and R^2 & Q^2 values of 0.78 & 0.60 and 0.80 & 0.69 for AM1 and PM6 optimized geometry datasets respectively. Also, for substituent group dataset QSPRs having 8 and 9 descriptors with R^2 & Q^2 0.82 & 0.76 and 0.93 & 0.83 values were determined for AM1 and PM6 optimized geometry dataset respectively. Quasi-mixture descriptors for substituent group datasets as they were providing the more statistical inference as substituent groups are easy to interpret as compared to the whole molecular structure of paclitaxel prodrug. It provided the QSPR models with R^2 & Q^2 values of 0.70 & 0.58 and 0.69 & 0.52 for AM1 and PM6 optimized geometries respectively. Eventually to study the metabolism of the substituent group dataset the *in-silico* docking and MD simulation studies were performed and a computational model was set for the QSPR, docking, MDS based studies. It is expected that the proposed QSPR models developed in the study would be useful for medicinal and chemical synthesist, as one of the most major area where the computational chemist can contribute to the synthesis of new therapeutic drug formation is study of SAR.

3.1 Introduction

A naturally occurring alkaloid paclitaxel (taxol), belongs to taxane group. These dynamic diterpenes are cytotoxic in nature and are isolated from *Taxus brevifolia* [1]. Precisely, paclitaxel is a clinically successful anticancer agent [2] and most effective in the treatment of different cancers like ovarian, breast, head and neck tumors, non-small cell lung cancer, and Kaposi's sarcoma [3-6]. The major bottleneck in its efficacy as an antitumor agent is its poor aqueous solubility; affected by this limitation there have been continuous efforts from the past for the formation of surfactants, emulsions, liposomes, cyclodextrins, and polymers so the solubility of paclitaxel can be improved [7-10].

Aqueous solubility that tends to be one of the most imperative physicochemical properties is one of the biological factors that affect the bioavailability of a biological molecule hence imparting a risk in the potent drug formulation [11]. Taxol that is the commercial formulation of paclitaxel constitutes two components; a surfactant, cremophor EL (polyethoxylated castor oil) and water-free ethanol. Both these excipients are used for enhancing the solubility of paclitaxel and the prepared formulation is given intravenously to the patients [12]. Along with that cremophor EL imparts many clinical side effects such as hyperlipidemia, anaphylactoid hypersensitive reactions, aggregation of erythrocytes, etc [13].

Prodrugs are modified form of drugs that undergo biological conversion in the body before it shows pharmacological effects [14]. Paclitaxel prodrug formation for increased solubility has also been explored to obtain its effective application [15]. The chemical structure of paclitaxel has a bulky, composed and fused ring together with number of hydrophobic substituent groups making it highly lipophilic in nature. Hence, a lot of hydrophilic substituent groups like carboxylic acids, phosphates, sugar derivatives etc. were added to paclitaxel structure for the formation of prodrugs with increased aqueous solubility [16-20]. Structure Activity-Relationship (SAR) studies explains 2' and/or 7' hydroxyl groups are the successful distinctive sites for making an addition to paclitaxel structure of hydrophilic groups for the formation of prodrugs (Figure 3.1). Nevertheless formation of paclitaxel prodrugs increases the solubility of drug to many folds but also the formation process associates with the synthesis of undesirable compounds those are summing to the cost of new prodrug evolution. Also, in the synthesis process of prodrugs, pharmacokinetics and pharmacodynamics properties of the drug molecule could be altered. Though the cytotoxic activities of paclitaxel prodrugs formed were comparable to drug molecule. Hence there is an

immediate need for some alternative computational methods for the solubility prediction of paclitaxel prodrugs so that the development of prodrugs could be cost and time effective.

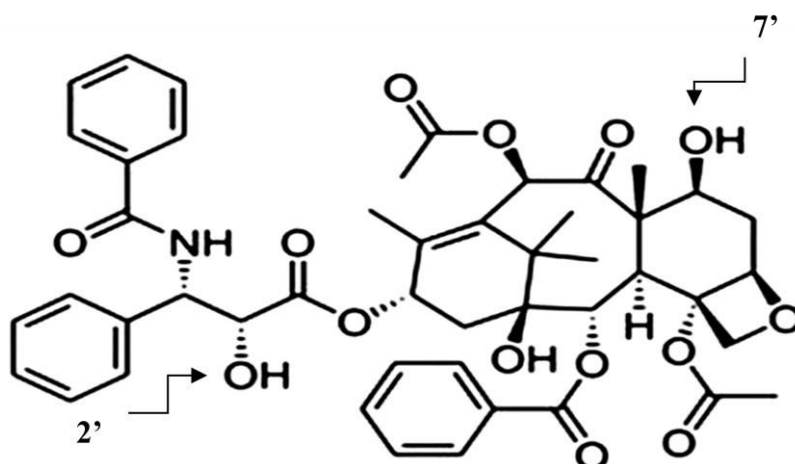


Figure 3.1: Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility.

Quantitative Structure Property-Relationship (QSPR) and data mining studies are efficient means to analyze molecular properties. Descriptors describe the structure and detailed description of molecular components that in turn model the molecular property being studied [21]. Hence molecular descriptors obtained from paclitaxel prodrug structures are used and QSPR models for the solubility prediction have been proposed. The hydrophilic substituent groups in the structure of molecule that are significantly responsible for solubility are difficult to portray. Whilst the descriptors derived from the substituent groups' molecular structures and subsequently the QSPR models formed will be easier to define. Therefore QSPR models from the substituent groups that were added to the paclitaxel for the formation of more soluble prodrugs were also formed and included in this study. In addition to Dragon descriptors, the quasi-mixture simplex representation of molecular structure (SiRMS) descriptors was also calculated for the substituent group dataset. SiRMS are based on the theory of 11 fundamental tetratomic fragments with connectivity and composition [22]. Independent and significant descriptors from the datapool of 5250 descriptors and the SiRMS descriptors were selected by implementing robust statistical measures like Akaike information criteria (AIC) and variance inflation factor (VIF), multicollinearity indicators [23-24]. Descriptor selection was also carried out using 'stepwise fit' in MATLAB (<https://in.mathworks.com/help/matlab/>) to review the descriptor selection and false positives could be removed. Cross-validation of all the models was accomplished. For the validations

of the developed models and to identify the chance correlations if any, Y-randomization test was performed.

Also, the metabolic studies of hydrophilic substituent groups would be substantial in the synthesis of new prodrugs, thus a molecular docking and molecular dynamics simulation (MDS) based model for the substituent groups' dataset, characterizing the most promising hydrophilic groups added for the formation of paclitaxel prodrugs. Cytochromes P450 (CYPs) liver metabolic enzymes would likely to characterize the kinetics of the substituent groups. CYPs are the major members of heme-thiolate protein family subsisting sixty different human isoforms which metabolizes ninety percent of the xenobiotics. The six major human isoforms of CYPs were subjected to docking with the substituent group dataset and the most favourable results were obtained from the Cytochrome 1A2 (CYP1A2), and the four substituent groups those were then subjected to MDS.

In such a way, QSPR modelling, molecular docking, and MDS based computational model supports a better view for the hydrophilic substituent groups those could be added to the paclitaxel drug molecule for the development of prodrugs with increased solubility and those could further be metabolised by the human body.

3.2. Materials and Methods

Development of QSPR models correlating aqueous solubility of paclitaxel prodrugs and hydrophilic substituent groups with Dragon descriptors and with quasi-mixture SiRMS descriptors is described below (Figure 3.2).

3.2.1 Paclitaxel prodrugs and hydrophilic substituent groups' datasets

A dataset of eighty diverse, more or comparably soluble paclitaxel prodrugs was formed. Also, a dataset comprising of sixty five unique, hydrophilic substituent groups those were added to the paclitaxel drug molecule in order to form the prodrugs in the past studies was collected from the literature (Appendix B, Table B.1) [16-20]. The solubility and activity values of the formed prodrugs were also collected from the literature. As there is diversity in the structures and in efficacy of the molecules; 1-dimensional (1D), 2-dimensional (2D) and 3-dimensional (3D) descriptors describing the whole structures, correlating with aqueous solubility were extracted for both the datasets.

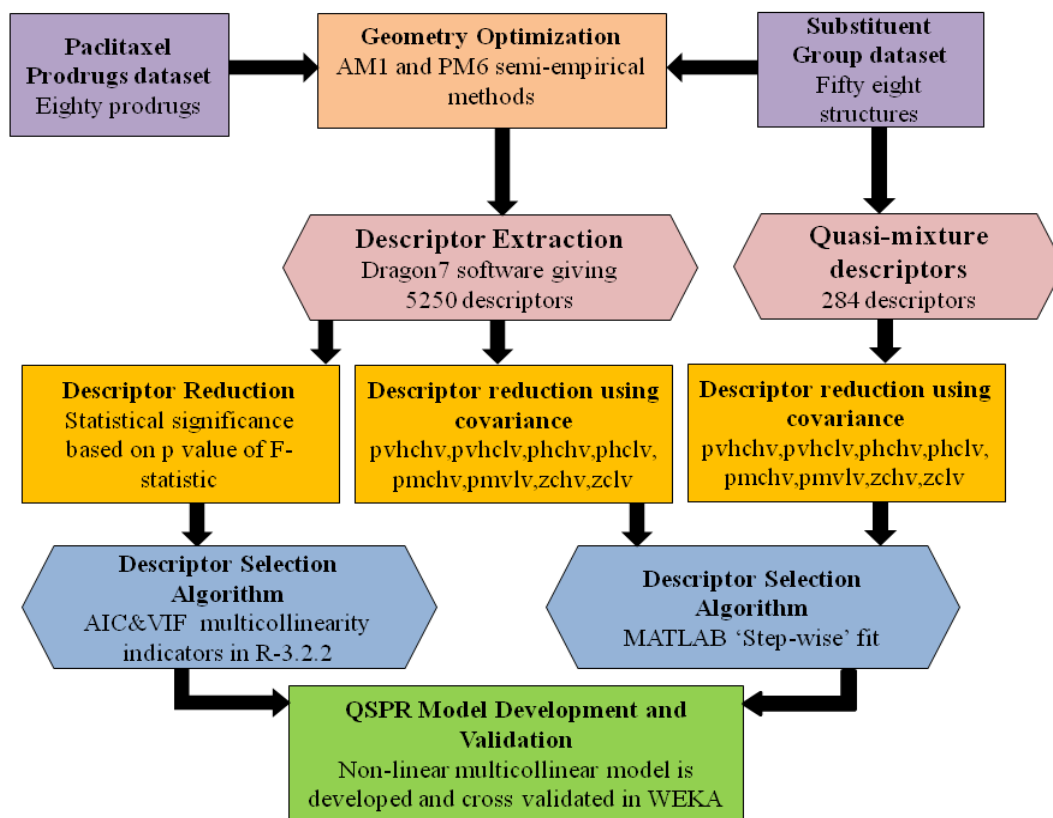


Figure 3.2: Geometry optimization, descriptor extraction, selection, development and validation of QSPR model.

3.2.2 Solubility data

Solubility values of the previously formed prodrugs were taken from the literature after that the solubility values were converted in $\log_{10}(S)$ for studying the correlation of solubility with the molecular descriptors. This step is required for organizing the solubility of the data in a linear manner.

3.2.3 Geometry optimization of the structures

Paclitaxel prodrugs and hydrophilic substituent groups' initial structures were drawn in Gauss View [23]. For optimizing the geometries of the structures semi-empirical methods were employed and the molecular structures were optimized levelled at parameterization method 6 (PM6) and Austin Model 1 (AM1) in Gaussian 09 quantum chemistry software [24]. As described in the chapter 2 in accordance with the QSPR-sPL.

3.2.4 Dragon7 and SiRMS descriptor extraction and reduction

For each molecule a large pool of 5250 descriptors were obtained by Dragon7 software [25], also 2D SiRMS descriptors for substituent group dataset were calculated by SiRMS master [26]. Dragon7 descriptors were grouped in thirty different descriptor groups and were broadly categorized in three classes *i.e.* 1-dimensional (1D), 2-dimensional (2D) and 3-dimensional

(3D) (Appendix A, Table A.2). Following the QSPR-sPL for descriptor reduction it was performed on the basis of variance and correlation. Various other statistical characteristics for the models like coefficient of determination R^2 , Fisher static F and p value for each descriptor's coefficient were also calculated. The whole process of descriptor reduction for all the datasets is shown in Figure 3.3.

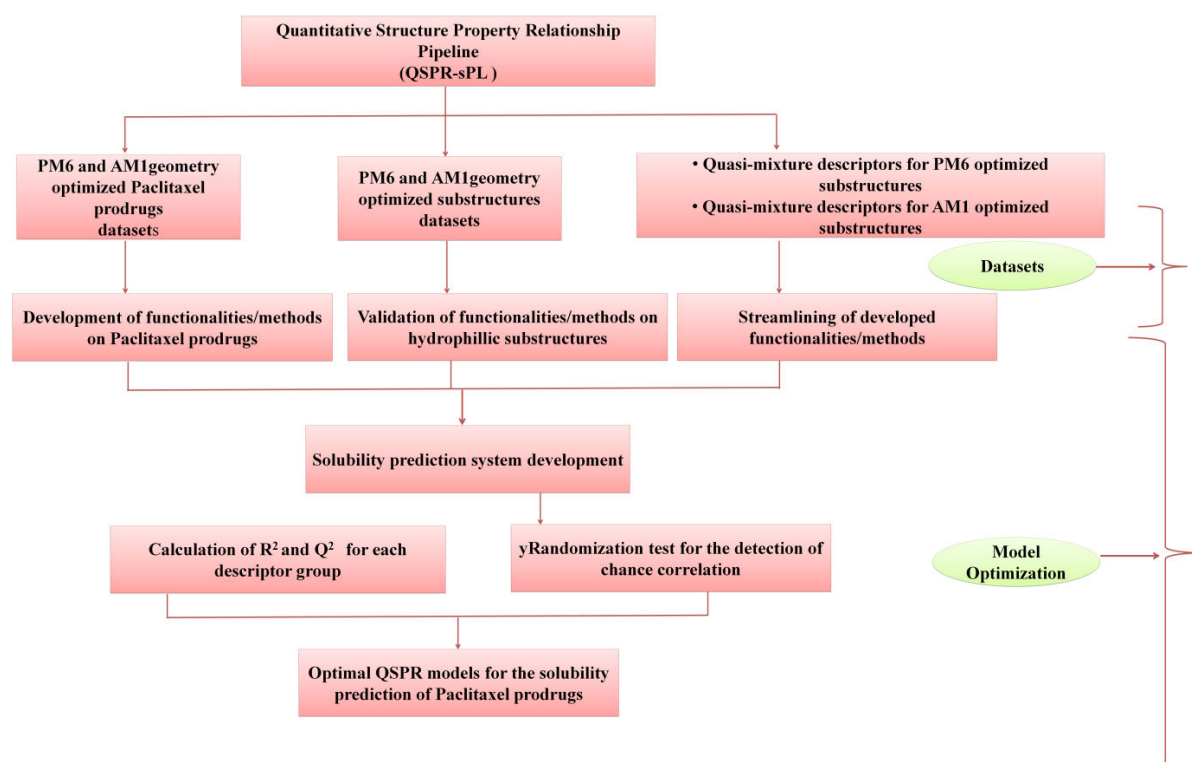


Figure 3.3: Model optimization

The 2D quasi-mixture SiRMS descriptors obtained for the substituent group dataset were also reduced implementing AIC and VIF multicollinearity indicators. MATLAB stepwiselm function was also implemented for the selection of descriptors having covariance value ($p < 0.05$) were selected from each descriptor group and the individual descriptor groups were subjected for the calculation of regression and correlation values. Significant descriptors from different descriptor groups were combined and again significant variables were selected. The optimal QSPR model was attained by combining the critical descriptors from the descriptor groups.

3.2.5 Non-linear regression determination for descriptor selection

Descriptor numerical values obtained for all the molecules differ from positive to negative as attained from the descriptor extraction. Hence, the equation which is proposed in QSPR-sPL was used for the formation of regression models.

$$\log_{10}(S) = \log\beta_0 + \beta_p \log X_p + \beta_n X_n \quad \text{Eqn 1}$$

In the equation, $\log\beta_0$ is a constant; β_p serve as the coefficient of descriptors with positive value; $\log X_p$ represents the logarithm of positive descriptor value; X_n shows the value of negative descriptor; and β_n serve the coefficient for negative descriptor. Significant descriptors obtained from each descriptor group were combined and then again chosen until there optimum R^2 and Q^2 values were obtained for each dataset. Finally, optimal models were formed and introduced for prediction of solubility for paclitaxel prodrugs and the hydrophilic substituent groups.

3.2.6 Formation of QSPR models and their validation

The QSPR models from dselected Dragon descriptors and also from the selected quasi-mixture SiRMS descriptors were developed and validated implementing QSPR-sPL pipeline.

3.2.7 ADMET Prediction of substituent groups

Computation of pharmacokinetic properties of substituent group dataset comprising sixty five molecules those were added to paclitaxel molecule for prodrugs formation was carried out using admetSAR 2.0 (<http://Immd.ecust.edu/admetSar2>), a web based server. Simplified Molecular-Input Line Entry System (SMILE) format for all the molecules was given input to admetSAR 2.0 that computed forty six properties for the dataset. The server does the prediction of various pharmacokinetics properties. From the dataset molecules those were showing favourable pharmacokinetic properties were further employed to molecular docking studies against CYP1A2 enzyme.

3.2.8 CYP1A2 binding site study

CYP1A2 protein structure was downloaded from the PDB (PDB ID: 2HI4, 1.95 Å). CYP1A2 is a major human CYPs family enzyme that metabolizes various small drugs, procarcinogens and also endogenous compounds. Planar active site of the enzyme is reported with the dimensions that are in agreement with its substrates [27]. “Thr118, Ser122, Thr124, Phe125, Thr223, Phe226, Phe256, Asn257, Phe260, Asp313, Ala317, Phe319, Thr321, Leu382 and Ile386” [28] amino acid residues highlights the active site in the structure of CYP1A2. Docking of all the substituent groups was performed in the described active site cavity.

3.2.9 Molecular docking

Pharmacokinetic aspect of the molecules was studied by ADMET properties. From the substituent group dataset thirty four substituent groups were selected and were implied to molecular docking against CYP1A2 in the above defined active site in AutoDock 4.2.6 [29].

AutoDock makes use of Lamarckian genetic algorithm for carrying out molecular docking and uses empirical scoring function for ranking the docked molecules. Semi-empirical free energy force field is used in regard of the calculation of ligand binding conformation. Protein-ligand binding conformation was used for the calculation of intermolecular energy starting from the unbound protein-ligand conformation. Protein's active site residues were employed to specify the grid and a grid box of dimensionality X=68, Y=86 and Z=74 grid points including the grid spacing of 0.397 Å was generated. All the molecules of substituent group dataset were set for docking with the population size of 150, number of maximum generations 27,000, maximum evaluations of 2,50,000, with a gene mutation rate of 0.02 and crossover rate of 0.8. All the molecules were implemented for the generation of 100 binding poses for the respective docking. Binding affinity was taken as the parameter for determining the docking and the best docked poses.

3.2.10 Molecular dynamics simulation study of best docked complexes

Four docked complexes with highest binding affinity were selected and were further subjected to molecular dynamics (MD) simulation for studying the stability of the docked complexes. Gromacs v5.1.2, [30] was used for performing the MD simulation as long as 50 ns to validate the stability of protein-ligand complexes. Prodrgr server [31] was used for obtaining the ligand topologies for all the five substituent groups using GROMOS 96 53a6 force field parameters [32]. The initial structures of all the five systems were solvated in a cubic box of SPC water model (three point model having three interaction sites that corresponds to three atoms of water molecule) and were charge neutralized by eleven chloride ions. The obtained structures were energy minimized for 50,000 steps with steepest descent algorithm, pursued by NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) equilibration of 0.1 ps, each one using the leap-frog integrator algorithm [33]. The Brendensen thermostat [34] algorithm was used to control the simulation temperature, and Parrinello-Rahman isotropic pressure coupling [35] was practiced for controlling pressure.

GROMOS 96 53a6 force field [36] was used for the MD simulation of apo-CYP1A2, CYP1A2-Bisulphite 2' acryoltaxol, CYP1A2-Fluorescent taxol, CYP1A2-Complex of paclitaxel with carbamate linkage to β-glucuronic acid, and CYP1A2-sulphate of paclitaxel for 50-ns time-scale to observe time dependent trajectories of systems in explicit solvent. The production MD was set again to time of 0 ps and the entire run was for 50 ns, with coordinates, velocities, energies stored at every 10 ps. The integration time-step was 2 fs,

treating all atoms explicitly and were coupled with LINCS constraint algorithm for restraining hydrogen and heavy atom bonds, the van der Waal cut-off was observed at 0.9 nm, and the Verlet cut-off scheme was used with Coloumb, with particle mesh ewald (PME) long range electrostatics [37]. Initial velocities were allowed on the basis of Maxwell distribution at 300 K with 1 nm minimum distance enforced in all the three directions within the periodic boundary conditions. The root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), hydrogen bonds (H-bonds), and solvent accessible surface area (SASA) calculations were done using the in-built gmx rms, gmx rmsf, gmx gyration, gmx h_bond and gmx sasa tools, respectively. The final MD trajectories were obtained using visual molecular dynamics (VMD) [38] and Chimera v1.13.1 [39] and Origin Pro 6.0 was used for plotting the respective graphs.

3.2.11 Principal Component Analysis

Projection of first two principal components (PCs) for the docked complexes kept on MD simulation for 50 ns was also analysed using the built-in principal component analysis (PCA) or essential dynamics method of Gromacs, also eigenvectors, eigenvalues, eigRMSF were evaluated. The correlated motions of the proteins that majorly define the protein functions were observed by the PC analysis and the principal components were grouped according to the protein functions [40]. All the translational and rotational movements of the complexes were removed and covariance matrices for all the complexes were formed. The following equation was used for the computation of elements of the positional covariance matrix C:

$$C_{ij} = (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \quad (i, j = 1, 2, \dots, 3N) \quad \text{Eqn 3}$$

The i^{th} C_{α} atom was represented by the cartesian coordinate q_{ij} and N being the number of atoms. The equilibrated trajectory that was superimposed on a reference structure, overall translations and rotations were removed using ‘Least-square fit’ method. All the matrices were diagonalized adopting the orthogonal co-ordinate transformation matrix Λ to predict the remaining eigenvectors and eigenvalues λ_i .

$$\Lambda = T^T C_{ij} T \quad \text{Eqn 4}$$

In the above equation, eigenvectors corresponds to the direction of relative motion of $\langle q_i \rangle$ were represented in the columns. Each eigenvector associated with the eigenvalues representing the total mean-square fluctuation of the system along the corresponding

eigenvector. The obtained last 30 ns trajectories were analysed for the calculation of eigenvector and eigenvalues using *gmx covar* and *gmx ana eig* in-built tools of Gromacs. The free energy of the systems implementing the first two principal components was performed using *gmx sham* tool [30].

3.2.12 Computing MM-PBSA Binding free energy

Molecular Mechanic/Poisson-Boltzmann Surface Area (MM-PBSA) method was used for the computation of binding free energy of the protein-ligand interactions. *g_mmpbsa* tool [41] of the Gromacs was used for the calculation of binding free energy for the last 10 ns trajectories snapshots obtained from MDS, that is an average of three energy terms; potential energy in the vacuum, polar-solvation energy and non-polar solvation energy in the case of protein-ligand binding, the binding energy of the bound system is an average.

3.4 Results and discussion

Individual descriptor groups' performance was determined and only the descriptors from each group providing significant R^2 & Q^2 values (Appendix B, Table: B.2-B.5) were coupled and again chosen so as to develop consistent QSPR models. The process of descriptor selection was followed till the valid R^2 & Q^2 values were obtained for both the datasets. The descriptors selection process for each group was also performed in MATLAB using 'stepwise fit'. Significant descriptors from descriptor groups were selected for both AM1 and PM6 geometry optimized datasets (Appendix B Tables: B.6–B.9) and the QSPR models with optimum R^2 & Q^2 values were developed.

QSPR models (Model I - VIII, Table 1, Figure 3.4 (a-d) and Figure 3.5 (a-d)) results from both the datasets provided good correlation values. The description of the models conduct and nature are provided in the Table 3.1.

QSPR Models:

- I. Model developed using molecular descriptor selected with AIC&VIF indicators for paclitaxel prodrugs with AM1 optimized geometry dataset.

$$\log_{10}(S) = 0.4564(On0V) - 0.5017(MSD) - 75.9343(PW5) - 0.1402(HTe) - 56.6473(HATS3i) + 207676(R4p) + 0.153(PVSA_{SS}) - 2.5513(NssNH) - 0.0002(Mor01u) + 6.8766(E1v) + 0.432(C-006) + 3.395(CATS2D_{05DP}) + 25.1779$$

- II. Model developed using molecular descriptor selected with AIC&VIF indicators for paclitaxel prodrugs with PM6 optimized geometry dataset.

$$\log_{10}(S) = 0.1189(ATSC1m) - 0.555(ATSC1v) + 0.5481(NssCH2) + 1.0881(NssNH) + 1.1215(N - 068) - 2.4335(CATS2D_{05DP}) + 1.142(xCATS2D_{03NL}) - 0.0136(T(N..N)) + 11.2824(B05[N - N]) + 0.1785(F08[C - N]) - 1.8115$$

- III. Model developed using molecular descriptor selected with AIC&VIF indicators for substructure with AM1 optimized geometry dataset.

$$\log_{10}(S) = 0.5168(C-006) + 0.0008(IDET) + 0.1916(Mor03e) + 0.6081(xMor32s) + 0.2319(ATSC4v) - 2.1004(GATS4e) - 0.2506(SM6G) - 0.0228(L2u) + 6.5945$$

- IV. Model developed using molecular descriptor selected with AIC&VIF indicators for substructure with PM6 optimized geometry dataset.

$$\log_{10}(S) = -2.2096(Mor32m) + 0.1218(Mor05e) + 0.5745(Mor12p) - 1.21 + 0.8687(Mor25p) + 0.4239(Mor31i) + 0.0975(Mor21s) + 0.0517(Tu) + 0.0016(Vs) + 2.7316$$

Table 3.1: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models for paclitaxel prodrugs and substructures with PM6 and AM1 optimized geometry dataset

S. No.	Descriptor Group	Descriptor selection algorithm	AM1 optimized geometry dataset				PM6 optimized geometry dataset			
			No. of descriptors	$R^2 \backslash R^2_{\text{yrand}}$	$Q^2 \backslash Q^2_{\text{yrand}}$	RMSE	No. of descriptors	$R^2 \backslash R^2_{\text{yrand}}$	$Q^2 \backslash Q^2_{\text{yrand}}$	RMSE
1	Paclitaxel prodrugs	AIC & VIF Function	21 (12)	0.78\0.13	0.60\0.31	1.55	16 (10)	0.80\0.14	0.69\0.20	1.46
2	Substituent groups	AIC & VIF Function	8 (8)	0.82\0.16	0.76\0.21	1.33	9 (9)	0.93\0.17	0.83\0.26	0.55
3	Paclitaxel prodrugs	MATLAB 'Stepwise fit'	9 (9)	0.85\0.10	0.79\0.18	1.02	16 (7)	0.82\0.08	0.79\0.14	1.40
4	Substituent groups	MATLAB 'Stepwise fit'	7 (6)	0.83\0.10	0.67\0.30	1.2	16 (13)	0.88\0.26	0.83\2.2	1.11

		fit*								
--	--	------	--	--	--	--	--	--	--	--

Paranthesis indicates the no. of significant descriptors included in QSPR model for a particular group.
 *Statistical significance of the models is determined by R^2_{rand} & Q^2_{rand} values as calculated (Section 3.2.6).

- V. Model developed using molecular descriptor selected with MATLAB ‘step-wise-fit’ for paclitaxel prodrugs with AM1 optimized geometry dataset.

$$\log_{10}(S) = 0.6907(nCb) + 4.7408(nRCHO) - 1.5841(nC \cdot O(O)2) - 0.8635(nRNHR) + 1.8856(N-068) - 1.7953(CATS3D_{09_{DL}}) + 3.6031(CATS3D_{19_{AP}}) + 0.7004(CATS3D_{11_{AN}}) - 1.4474(CATS3D_{02_{NL}}) + 1.2857$$

- VI. Model developed using molecular descriptor selected with MATLAB ‘step-wise-fit’ for paclitaxel prodrugs with PM6 optimized geometry dataset.

$$\log_{10}(S) = -9.7178(ATSc6v) + 7.2219(GATS6s) + 7.4503(NsCH3) + 3.7658(NssNH) + 0.874(CATS2D_{01_{DD}}) + 1.0976(CATS_{2D_{NL}}) + 2.3893(CATS3D_{03_{AL}}) + 13.0683$$

- VII. Model developed using molecular descriptor selected with MATLAB ‘step-wise-fit’ for substructure with AM1 optimized geometry dataset. fit’ for substructure with AM1 optimized geometry dataset.

$$\log_{10}(S) = -0.3448(nCs) + 1.868(nCq) + 9.6299(nRCONH2) - 3.7002(nN = C-N <) + 5.8506(nRNHR) - 1.603(nSO2OH) - 0.4912$$

- VIII. Model developed using molecular descriptor selected with MATLAB ‘step-wise-fit’ for substructure with PM6 optimized geometry dataset.

$$\log_{10}(S) = 2.5006(Chi0_{EA}) - 0.042(P_{VSA_{LogP_5}}) - 2.1627(P_{VSA_{e_1}}) + 0.0528(P_{VSA_{S_5}}) + 9.7303(nArOCOn) + 2.9485(nRNHR) + 0.8348(nROR) - 5.4092(nSO_3) - 3.3368(SdsCH) - 3.1137(SddC) + 3.3186(NdsCH) - 0.2048(NaaCH) - 1.1079(NsssCH) - 1.1623$$

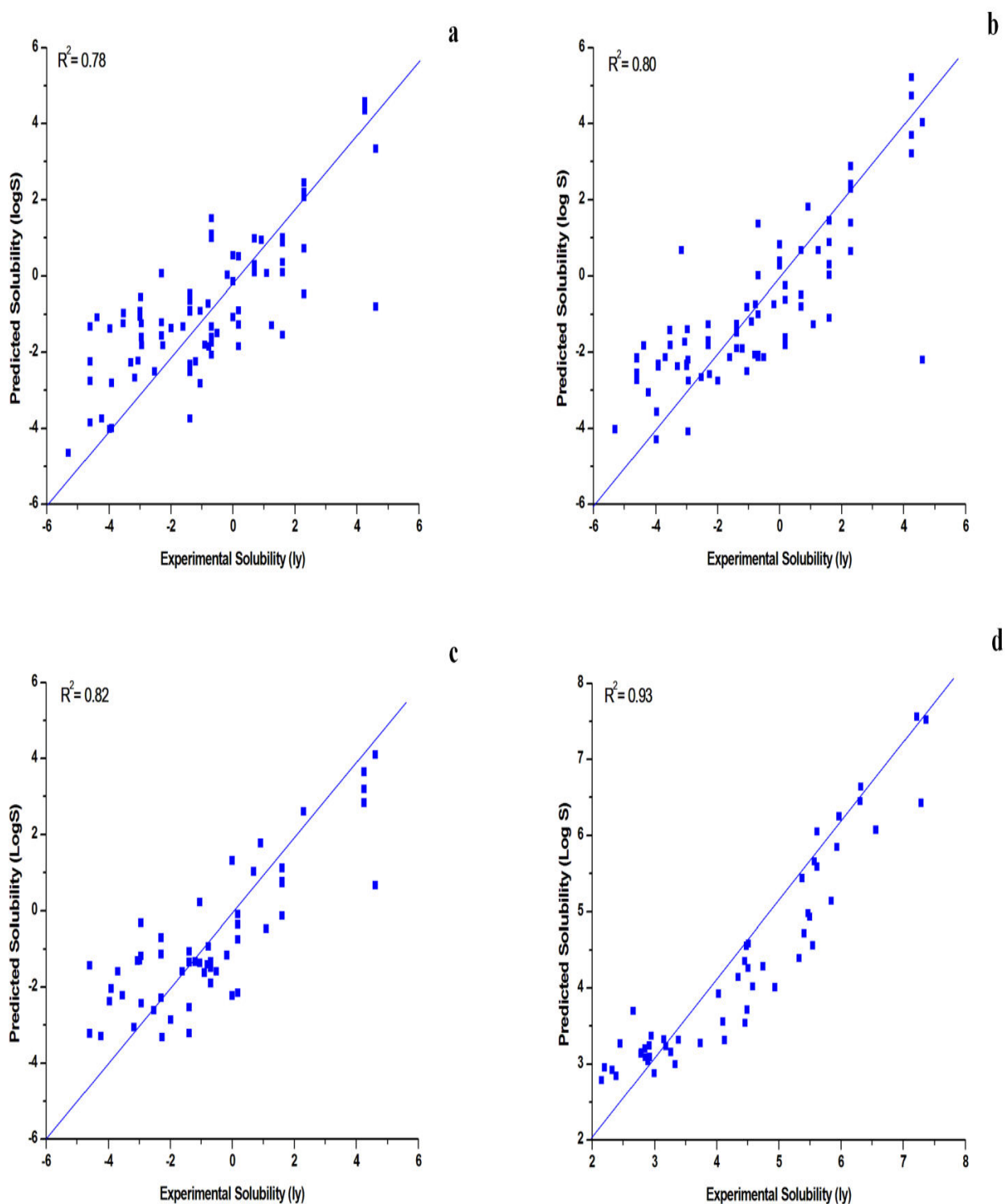


Figure 3.4: Experimental solubility (ly) vs predicted solubility (*logS*) of Paclitaxel prodrugs. Performance of descriptors selected with AIC&VIF indicators for Paclitaxel prodrugs with [a] AM1 geometry dataset; [b] with PM6 geometry dataset; [c] Performance of descriptor selected with AIC&VIF indicators for substructures with AM1 geometry dataset; and [d] for substructures with PM6 geometry dataset.

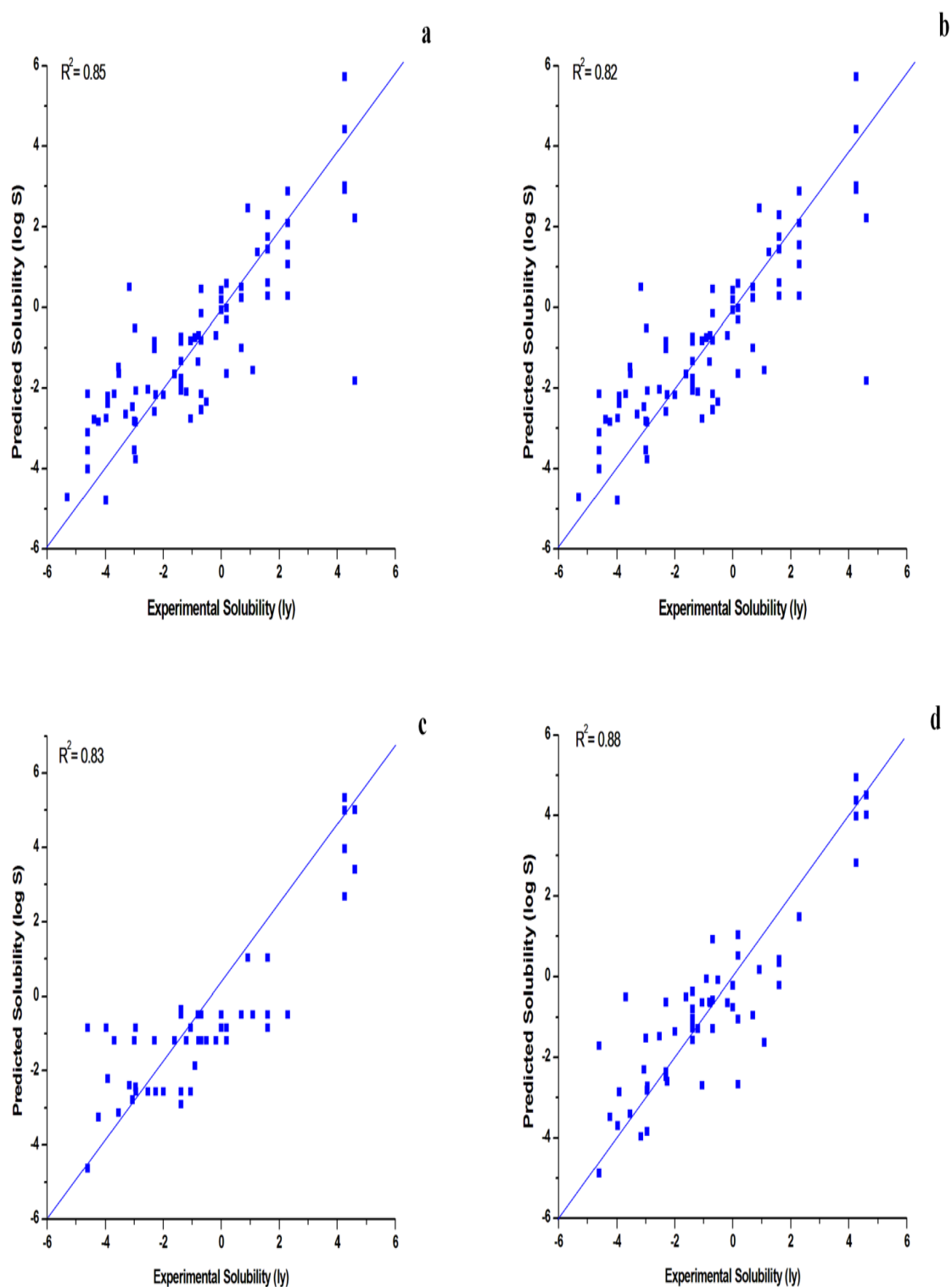


Figure 3.5: Experimental solubility (ly) vs predicted solubility ($\log S$) of paclitaxel prodrugs. Performance of descriptors selected with MATLAB ‘stepwise-fit’ for paclitaxel prodrugs with [a] AM1 geometry dataset; [b] with PM6 geometry dataset; [c] Performance of descriptor selected with MATLAB ‘stepwise-fit’ for substructures with AM1 geometry dataset; and [d] for substructures with PM6 geometry dataset.

In the current study, two different approaches were used for descriptor selection and comparable R^2 & Q^2 values were achieved. PM6 and AM1 optimized geometry datasets for the substructure molecules formed the QSPR models with more precision than the QSPR models constituted from datasets of paclitaxel prodrugs for both PM6 and AM1 optimized geometry. Therefore, additionally for the substructure dataset, the ‘quasi-mixture’ descriptors were extracted and 2D QSPR model with these descriptors for both PM6 and AM1 geometry optimized molecules were also formed. Quasi-mixture descriptors were defining the probable intermolecular interactions in pure compounds. All the common descriptors which were frequent in all the substructures were taken and making use of QSPR-sPL, QSPR models for both datasets with geometries optimized at PM6 and AM1 levels. For the formed QSPR models, comparable R^2 & Q^2 values were attained (Table 3.2).

Table 3.2: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models formed from Quasi-mixture descriptors for the substructures with PM6 and AM1 optimized geometry dataset

S. No.	AM1 optimized geometry dataset				PM6 optimized geometry dataset			
	No. of descriptors	$R^2 \backslash R^2_{\text{yrand}}$	$Q^2 \backslash Q^2_{\text{yrand}}$	RMSE	No. of descriptors	$R^2 \backslash R^2_{\text{yrand}}$	$Q^2 \backslash Q^2_{\text{yrand}}$	RMSE
1	8 (5)	0.70\0.12	0.58\0.50	1.70	7 (6)	0.69\0.09	0.52\0.32	1.71

Parenthesis indicates the no. of significant descriptors included in QSPR model.

*Statistical significance of the models is determined by R^2_{yrand} & Q^2_{yrand} values as calculated (Section 3.2.6).

Different properties of the molecule like polarity, relative aromaticity, topology, and aliphatic degree of the molecules influence the drug solubility. The set of QSPR models formed provides the descriptors that can define the solubility prediction for paclitaxel prodrugs (Appendix B Table B.2-B.5). A descriptor group like atom-type E-state indices determines the individual atom topologies as well as calculates the contribution of a distinct atomic fragment. Atom topologies can be defined by the intrinsic electronic and topological state along with the electronic field effect of a molecule on a particular atom [42]. Hence these descriptors are found to be instrumental for identifying the atoms or fragments in the molecule that are responsible for solubility. Additionally P_VSA-like descriptor group defines the sum of atomic Van der Wall surface area (VSA) contribution of each atom within a molecule [43]. The solubility of a drug can be defined as the degree of drug-solvent interaction. Thereof, the P_VSA descriptors are correlated with the solubility of the molecule. Chemical functional groups in a molecule are counted by the functional group descriptors [44]. The complex and large structure of paclitaxel benefitted the study of functional group

count descriptors for the paclitaxel descriptors, while any substituent group can be added for the formation of more soluble prodrugs. Thus the functional group count descriptors explain its association while studying the aqueous solubility of prodrugs. 3D-Morse descriptors describes the 3 dimensional structures of molecules into the numerical form based on the range of electron diffraction's scattering parameter values along with a variety of weighting schemes [45]. These descriptors played an important role in the prediction of aqueous solubility of prodrugs as there were structurally similar molecules in the dataset and these descriptors reflects a change in the values if the molecular geometry of the molecules gets differed.

Atom-centred fragments subsists a central atom encircled one or several shells of atoms that are distanced at the similar topological distance [46]. For storing the local spectral parameters different ACF fragments are designed and these descriptors thus describe the solubility property of the molecules being studied. Descriptors like WHIM (Weighted Holistic Invariant Molecular) define the various molecular features *viz.* shape, size, symmetry, and atom distribution in the three dimensional structure of a molecule. A weighting scheme is used for obtaining the specific details of the molecules in the descriptor set [47]. The descriptors are explaining the solubility property for the molecule as shape, size, and symmetry of the drug molecule affects the drug solvent interaction. Atomic properties those are distributed ahead of the molecular topology are measured numerically from 2D Autocorrelation descriptors; they imply an autocorrelation function to the different molecular regions to calculate the descriptors [48]. These descriptors measure the solubility of the molecule towards the different path lengths in the molecular structure.

Increase in the solubility of the prodrugs by the addition of substituent groups makes them a potential molecule. But the enzymatic actions in the body are also the determining factors for the potent substituent groups that could be added to the paclitaxel for the formation of prodrugs. Hence for this purpose the study includes the analyses of ADMET criterion to substituent group dataset; that consists of sixty five molecules with the help of admetSAR 2.0. (Appendix B, B.10). Further, on the basis of various calculated ADMET properties the selection of thirty four molecules was done and those were then subjected further for the molecular docking studies.

CYPs 450 liver enzymes are a major class of metabolic enzymes that metabolizes the xenobiotics. These enzymes metabolize almost ninety percent of the small molecules, drugs, and xenobiotics in the human body. Henceforth, a computational model based on molecular

docking and MD simulation was attempted for the prediction of most likely substructures for the formation of more soluble paclitaxel prodrugs. The computational metabolic property of substituent group dataset was calculated; further molecular docking of the thirty four selected molecules against CYP1A2 was performed using AutoDock 4.2.6. Taking into account the binding energies of the thirty four molecules, four top scoring molecules were chosen and were implied to MD simulation studies. Results of the molecular docking studies are presented below in Figure 3.6 and Table 3.3 In the table, the amino acid residues in bold are representing the hydrogen bond forming residues.

Table 3.3: Docking results of the selected substituent groups with PM6 optimized geometry against CYP1A2 showing the binding affinity and the interacting residues

S No.	Substituent groups	Binding energy (kJmol ⁻¹)	Interacting residues
1.	Bisulphite 2' acryoltaxol	-33.59	Arg108, Leu 123, Thr124, Trp133, Arg137, Thr385, Ile386, His388, Phe451, Gly452, Arg456, Arg457.
2.	Flourescent taxol	-38.07	Arg108, Leu123, Thr124, Arg137, Leu144, Ile200, Asp313, Ile314, Ala317, Gly318, Arg456, Gly460, Ala464, Ile459, Cys458.
3.	Paclitaxel with carbamate linkage to β -glucuronic acid	-31.42	Arg108, Leu123, Thr124, Trp133, Arg137, Leu382, Thr 385, Ile 386, Glu411, Leu450, Phe451, Gly452, Arg456, Cys458, Ile459
4.	Sulphate of paclitaxel	-41.04	Arg108, Leu123, Thr124, Trp133, Arg137, Phe226, Asp320, Thr321, Ile386, Leu450, Phe451, Gly452, Arg456, Cys458, Ile459, Leu497, Ala317

A top scoring molecule bisulphite 2' acryoltaxol-CYP1A2 with binding energy -33.59 kJ.mol⁻¹ was selected and interaction in the binding site were analyzed using LigPlot, thence amino acid residues Arg108, Leu123, His388, Arg456 were found to form hydrogen bonds, and residues Thr124, Trp133, Arg457 etc. were found to form hydrophobic contacts with CYP1A2 (Figure 3.6a and Table 3.3). In the same way, Fluorescent taxol-CYP1A2 docked complex provided the binding energy of -38.07.1 kJ.mol⁻¹. The complex was shown to have hydrogen bond formation with amino acid residues Arg108, Thr124, Arg456, Ile459, and

Arg137, Ala317, Cys458, Leu123 residues had hydrophobic interactions in the complex (Figure 3.6b and Table 3.3). Paclitaxel with carbamate linkage to β -glucuronic acid substituent group allowed the binding energy of $-31.42 \text{ kJ}\cdot\text{mol}^{-1}$, with Arg108, Trp133, Arg137 amino acids build polar contacts along with residues Leu123, Thr124, Ile386 etc. forming the hydrophobic contacts in the complex (Figure 3.6c and Table 3.3). Trp133, Arg137, Arg456 amino acid residues of sulphate of paclitaxel presented the polar interactions in the active site of CYP1A2 permitting $-41.04 \text{ kJ}\cdot\text{mol}^{-1}$ of binding energy (Figure 3.6d and Table 3.3). The results of virtual screening performed on the substituent group dataset predict the most encouraging substituent group class that could be probable in the increase of aqueous solubility of paclitaxel prodrugs. Additionally for the enrichment of the molecular docking results MD simulations were carried out and the stability of the complexes was study at detailed atomic level.

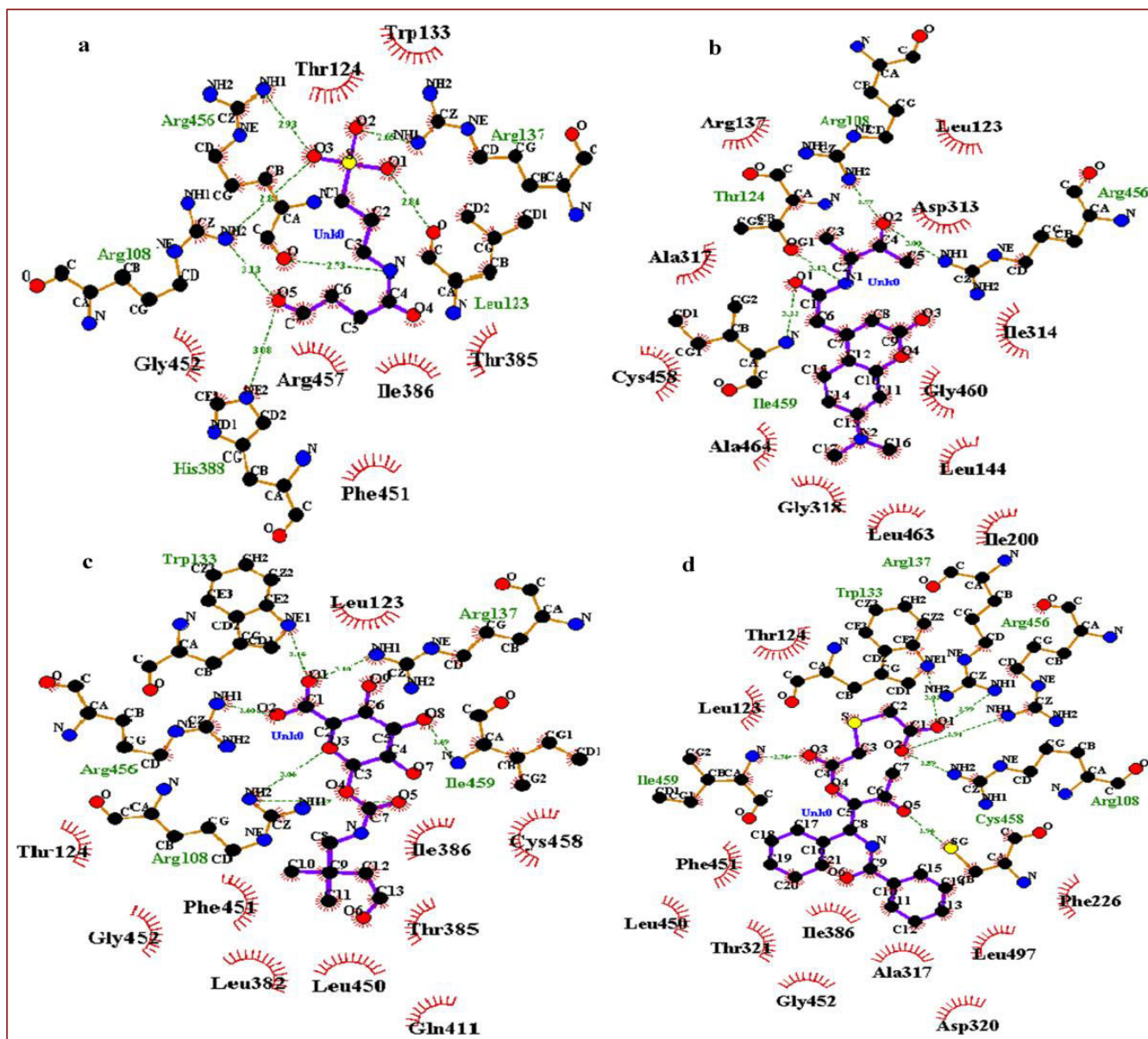


Figure 3.6: Ligand Interaction diagram [a] Bisulphite 2' acryoltaxol with CYP1A2; [b] Fluorescent taxol with CYP1A2; [c] Paclitaxel with carbamate linkage to β -glucuronic acid with CYP1A2 [d] Sulphate of paclitaxel with CYP1A2.

MD Simulation Study

The stability of docked complexes studied at the atomic level can be analyzed by the MD simulations. The dynamic study of CYP1A2 docked with four different substituent groups' complexes and one apo-protein for a time scale of 50 ns was performed on Gromacs. The equilibration state for all the five systems was acquired and equilibrated trajectories for the last 30 ns were evaluated for the computation of rmsd, rmsf, radius of gyration, SASA and binding free energy.

Root mean square deviation (RMSD)

Root mean square deviation (RMSD) describes the conformational shift occurring in the backbone of the protein for all systems over the simulation time scale. Dynamic stability of the simulated systems is defined by RMSD. RMSD values were calculated for all the five systems and the respective equilibrated trajectories were obtained after 20 ns, Figure 3.7a. Average values of “apo-CYP1A2, CYP1A2-Bisulphite of 2’ acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel” substituent groups were found to be 0.370, 0.346, 0.307, 0.320 and 0.371 nm respectively. The observed rmsd values for all the molecules infer the stability of the molecules with respect to the apo-protein. Two substituent groups bisulphite of 2’ acroyltaxol and paclitaxel with carbamate linkage to β -glucuronic acid showed the highest stability. The substituent groups’ complexes with CYP1A2 showed the stability hence got metabolised by the enzyme.

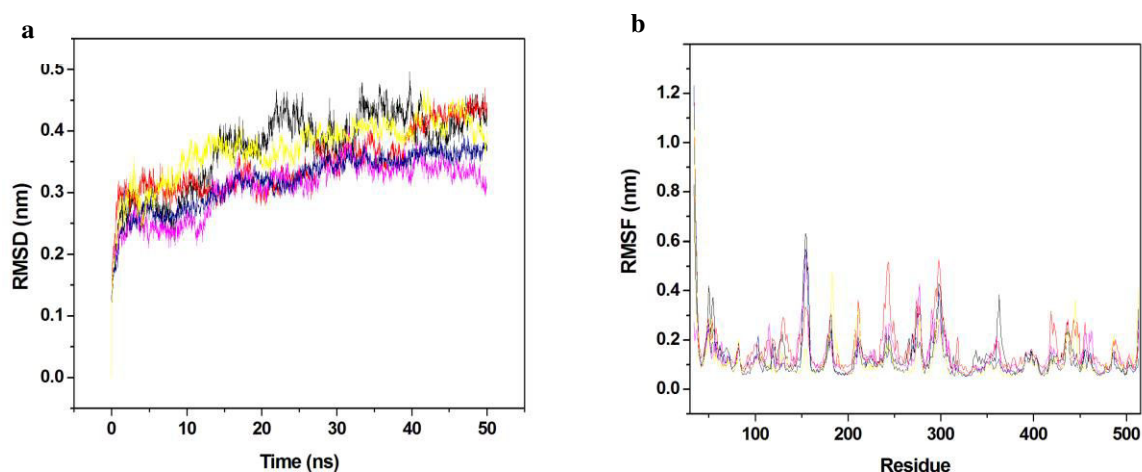


Figure 3.7: Molecular dynamics simulation. (a) RMSD of the C α backbone of substituent groups over the 50 ns MDS at 300 K, (b) RMSF of residues of substituent groups during MDS, In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2’ acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2- Sulphate of paclitaxel (magenta).

Residue Mobility Analysis

Residue mobility calculations for the systems for the last 30 ns equilibrated trajectory were noted as shown in Figure 3.7b. The average values of RMSF for “apo-CYP1A2, CYP1A2-Bisulphite of 2’ acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel” were 0.141, 0.167, 0.132, 0.117 and 0.112 nm respectively. These RMSF values showed the stability of the complexes.

According to the results obtained with these RMSF values paclitaxel with carbamate linkage to β -glucuronic acid and sulphate of paclitaxel substituent groups were found to be the most stable. The molecules forming stable complexes with CYP1A2 shows metabolism of these groups giving a good pharmacokinetics description.

Compactness Analysis

Density of a protein-ligand complex is described by Radius of gyration factor (Rg). It averages the distance between the centre of mass of the protein atoms and terminal atoms in a specific time-frame. The compact protein is considered to be more stable and there is less variation in the value of gyration while the dispersed protein structure will be less stable. In general, the compact protein shows lesser amount of variation in the gyration value while the open structure presents higher Rg value. For a given time frame, Rg values for all the complexes were plotted, Figure 3.8a. Average values of Rg for “apo-CYP1A2, CYP1A2-Bisulphite of 2’ acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel” were 2.27, 2.31, 2.30, 2.27 and 2.28 respectively. All the substituent group complexes were found to be less or comparably stable than the apo-protein, Figure 3.8a. The CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid complex presented comparable stability as that of apo-protein, hence predicting the metabolism of these groups with CYP1A2, they will be metabolised by the CYP1A2 enzyme.

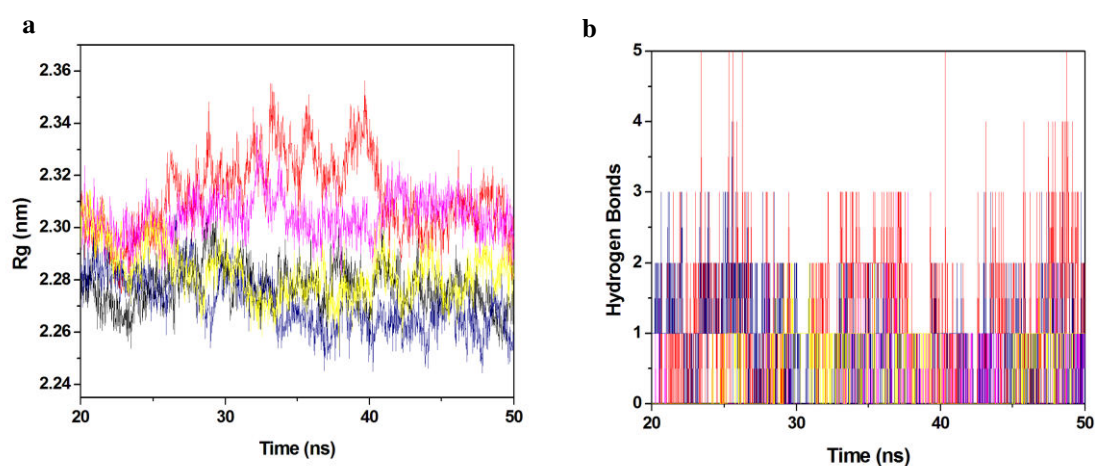


Figure 3.8: Molecular dynamics simulation. (a) Rg vs time of substituent groups during MDS. (b) Number of hydrogen bonds interaction between protein and ligand during simulation time scale for ligand complexes. In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2’ acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2- Sulphate of paclitaxel (magenta).

Hydrogen Bonding Analysis

Transient interactions in the protein-ligand complex are defined by hydrogen bonds formed between the protein-ligand providing a stable complex. In the MD simulation run calculation of the number of hydrogen bonds formed in the docked structures in a specific time-scale was observed, Figure 3.8b. Average number of hydrogen bonds formed in “CYP1A2-bisulphite of 2’ acroyltaxol, CYP1A2-flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-sulphate of paclitaxel” were 0-3, 0-1, 0-2 and 0-1 respectively. CYP1A2-bisulphite of 2’ acroyltaxol was observed to have the highest number of hydrogen bonds than other molecules. The acceptable hydrogen bonds for all the substituent groups were observed, explaining the stable interactions with the CYP1A2 enzyme.

Principal Component Analysis

Principal component analysis predicts the correlated motion of the complexes. Overall motions in the proteins are described by the principal components (PCs) that are defined by few key eigenvectors. Hence the eigenvalues of the complexes for the 50 eigenvectors were plotted Figure 3.9a. The first five eigenvectors explained 67%, 72.01%, 64.34%, 69.18%, 68.69% of the motions respectively of “apo-CYP1A2, CYP1A2-Bisulphite of 2’ acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel” for the last 30 ns equilibrated trajectory. The results explained that bisulphite of 2’ acroyltaxol, paclitaxel with carbamate linkage to β -glucuronic acid and sulphate of paclitaxel substituent groups showed few motions in comparison to the apo-protein thus forming a stable structure with CYP1A2. PCA result reveals the metabolism of these selected substituent groups by the CYP1A2 isoform of the cytochrome.

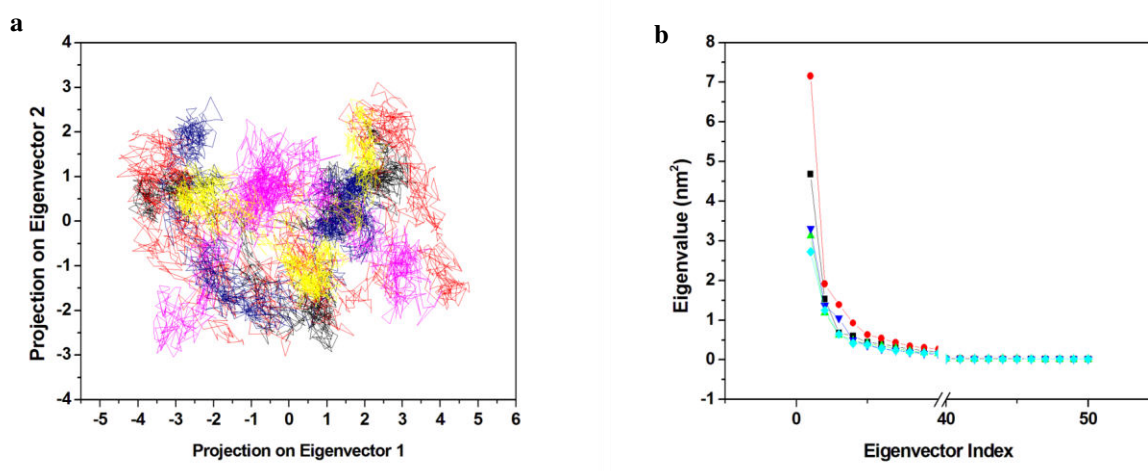


Figure 3.9: Principal component analysis. (a) Projection of the motion of the protein in phase space along the PC1 and PC2 (b) Plot of eigenvalues vs. eigenvector index. First 50 eigenvectors were considered. In all panels

the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acroyltaxol (red), CYP1A2-Flourescent taxol (navy blue), CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid (yellow), CYP1A2- Sulphate of paclitaxel (magenta).

Also, the first two eigenvectors for the complexes representing the correlated motions were plotted oppositely in a phase space in which the correlated motions are represented by the individual spectrum (Figure 3.9b). Stability of the complex is characterized by the dense nature of the complex and scattered complex showed instability. Bisulphite of 2' acroyltaxol substituent group was found to be least stable among the substituent groups.

Solvent accessible surface area (SASA)

Interactions in the protein are reflected by the conformational changes in the three dimensional structure of the protein. Relative solvent accessible surface area gives the measure of conformational changes occurring in the protein during binding of the ligand. Both polar and non-polar interactions among the protein's amino acid residues constitute the solvation free energy for the protein. SASA value for each docked complex was computed over the concluding 30 ns equilibrated trajectory, Figure 3.10a. Average SASA values obtained for "apo-CYP1A2, CYP1A2-Bisulphite of 2' acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel" were 235.69, 242.24, 237.55, 235.41 and 232.64 nm² respectively. These results showed that the substituent group paclitaxel with carbamate linkage to β -glucuronic acid and sulphate of paclitaxel substituent groups were depicting lesser or comparable SASA values as that of the protein. Thus giving the conformational stability of CYP1A2 while ligand binding. Individual residue contribution to the SASA value leading to the change in the conformation of the protein was also calculated. A graph of the residues and SASA value was plotted in Figure 3.10b. Average SASA values of the residues attained for "apo-CYP1A2, CYP1A2-Bisulphite of 2' acroyltaxol, CYP1A2-Flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-Sulphate of paclitaxel" were 0.48, 0.50, 0.49, 0.48 and 0.48 nm² respectively. Depending on the values obtained for the residue SASA it is inferred that these substituent groups would be metabolised by the enzyme CYP1A2.

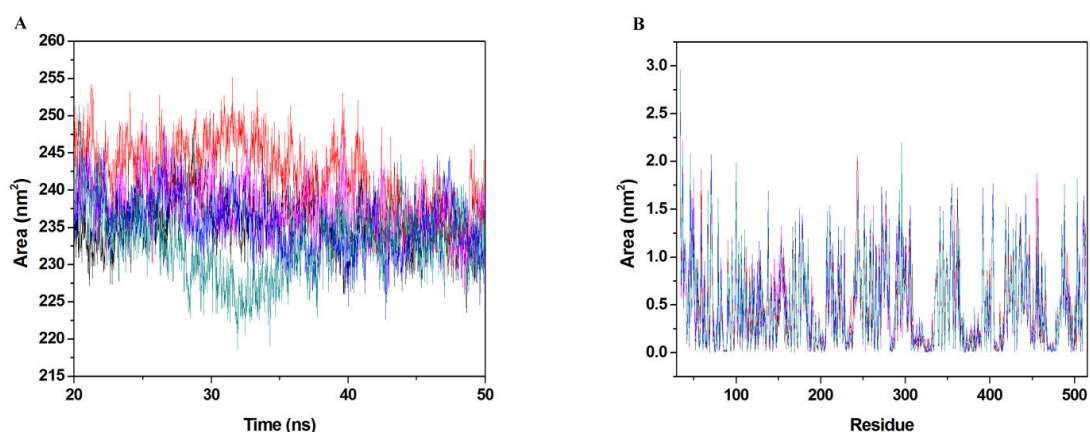


Figure 3.10: Solvent accessible surface area. (a) The SASA plot for CYP1A2 and substituent group complexes in water with respect to time. (b) The Residue SASA plot for CYP1A2 and substituent group complexes. In all panels the color code is: CYP1A2 (black) and the ligands CYP1A2-Bisulphite of 2' acrolytaxol (red), CYP1A2-Flourescent taxol (blue), CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid (dark cyan), CYP1A2-Sulphate of paclitaxel (magenta).

Gibbs free energy landscape

Orientation of the binding of ligand to protein is described by Gibbs free energy landscape. During the MD simulation of substituent groups with CYP1A2 it was observed for PC1 and PC2 to view the changes in the conformational state upon ligand binding. The obtained range of energy values were; 0 to 8.66, 0 to 8.17, 0 to 9.17, 0 to 9.35 and 0 to 8.98 $\text{kJ}\cdot\text{mol}^{-1}$ for “CYP1A2, CYP1A2-bisulphite of 2' acrolytaxol, CYP1A2-flourescent taxol, CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid and CYP1A2-sulphate of paclitaxel” respectively, Figure 3.11. The substituent groups were representing the agreeable transitions of different conformations, favouring the metabolism of the groups by CYP1A2 isoform.

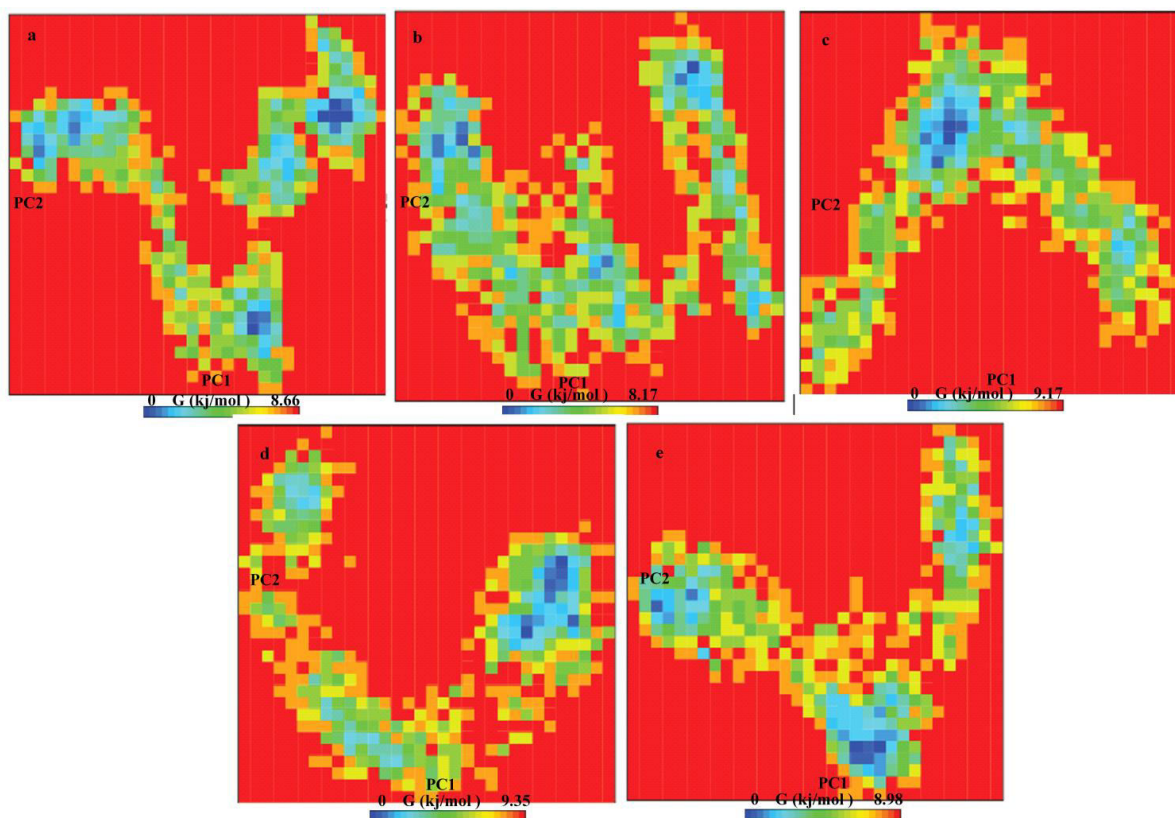


Figure 3.11: Gibbs free energy Landscape. The free energy landscape calculated for PC1 and PC2. In the figure, (a) CYP1A2, (b) CYP1A2-Bisulphite of 2' acroyltaxol, (c) CYP1A2-Flourescent taxol, (d) CYP1A2-paclitaxel with carbamate linkage to β -glucuronic acid, (e) CYP1A2-Sulphate of paclitaxel.

MMPBSA Binding Free Energy Study

Binding free energy is the determining factor for the interaction between two molecules. Ingrained MMPBSA tool in Gromacs [30], was employed for binding free energy calculations. Summation of the non-bonding interactions of complex constitutes binding free energies for the complexes which is tabulated in the Table 3.4. All the substituent groups were depicting the favourable binding free energy with CYP1A2 hence forming stable structure with the protein.

Table 3.4: Table represents the Van der Waals, electrostatic, polar solvation, SASA and binding energy in kJ.mol^{-1} for substituent groups

S. No.	Compound	Van der Waals energy	Electrostatic energy	Polar solvation energy	SASA energy	Binding energy
1.	Bisulphite of 2' acroyltaxol	-59.87 ± 66.41	-24.35 ± 29.10	62.01 ± 82.07	-6.19 ± 6.97	-28.39 ± 33.89
2.	Flourescent taxol	-196.64 ± 10.91	-22.03 ± 11.06	89.74 ± 15.28	-18.73 ± 0.85	-147.67 ± 12.57
3.	Paclitaxel with carbamate linkage to	-142.41 ± 56.29	-52.205 ± 25.42	101.93 ± 79.23	-15.52 ± 6.45	-108.21 ± 20.74

	β -glucuronic acid					
4.	Sulphate of paclitaxel	-238.83 ± 44.50	-21.46 ± 7.85	104.70 ± 24.75	-21.89 ± 4.37	-177.50 ± 32.66

QSPR models for the solubility prediction of paclitaxel prodrugs those are geometry optimized at AM1- and PM6-levels were obtained and the metabolic study of the substituent group dataset those were added to the molecule for the formation of prodrugs was also carried out to find the most efficient groups that not only increase the solubility but also can be metabolized in the body.

3.4 Conclusion

Eight new QSPR models based on the fundamentals of aqueous solubility and optimization of descriptor space from the paclitaxel prodrugs dataset and also from the substructure dataset which were optimized at both AM1- and PM6 geometry levels with good statistics were obtained. Model IV, that is obtained from substructure dataset along with molecular geometries optimized at PM6 level and the descriptor were selected with the AIC & VIF multicollinearity indicators is the most rigorous one. Model III developed from the substructure molecules optimized at AM1 level and those descriptors were selected implying AIC and VIF multicollinearity indicators can be set up after Model IV. Also, the results obtained from two different descriptor selection methods and the ‘quasi-mixture’ descriptors for the specific validation of substituent group dataset were accomplished. The study also incorporates the metabolic study of substituent group dataset in the human body. This approach will be useful for the medicinal chemists to improve the solubility profile of the future prodrugs by suggesting the structural modifications in the molecule. A systematic and robust approach of developing mixture QSPR models was also attempted and good statistical standards were obtained. It is anticipated that developed models will be used as a plausible resource for the aqueous solubility prediction for the paclitaxel prodrugs. In addition to that, the eight new proposed QSPR models will support a classic way of criterion readjustment for the future prodrugs concerning their solubility prediction. Also, the computational model based on molecular docking and MD simulation was formed for studying the binding mode of substituents with the CYP1A2 and the results affirmed the metabolism of substituent groups by the CYP1A2 enzyme. This computational model predicts the solubility of the prodrugs to be formed. It will also help the selection of potential candidates which could be predicted by the formation of a model that studies the metabolism of the potent substituent groups.

References

- [1] Lavelle, "Taxoids: a new class of antimetabolic agents," *Current Opinion in Investigational Drugs*, vol. 2, pp. 627–635, 1993.
- [2] B. J. Roth, R. Dreicer, L. H. Einhorn, D. Neuberg, D. H. Johnson, J. L. Smith, and P. J. Loehrer, "Significant activity of paclitaxel in advanced transitional-cell carcinoma of the urothelium: a phase II trial of the Eastern Cooperative Oncology Group," *Journal of Clinical Oncology*, vol. 12(11), pp. 2264–2270, 1994.
- [3] M. E. Wall, and M. C. Wani, "Camptothecin and taxol: discovery to clinic--thirteenth Bruce F. Cain Memorial Award Lecture" *Cancer Research*, vol. 55(4), pp. 753–760, 1995.
- [4] V. Walsh, and J. Goodman, "From taxol to Taxol: the changing identities and ownership of an anti-cancer drug," *Medical Anthropology a*, vol. 1(3–4), pp. 307–336, 2000.
- [5] V. Walsh, and J. Goodman, "The billion dollar molecule: Taxol in historical and theoretical perspective," *Journal of Clinical Medicine b*, vol. 66, pp. 245–267, 2000.
- [6] R. S. Tuma, "Taxol's Journey from Discovery to Use: Lessons & Updates," *Oncology Times*, vol. 25(18), pp. 52–57, 2003.
- [7] M. Ceruti, P. Crosasso, P. Brusa, S. Arpicco, F. Dosio, and L. Cattel, "Preparation, characterization, cytotoxicity and pharmacokinetics of liposomes containing water-soluble prodrugs of paclitaxel," *Journal of Controlled Release*, vol. 63(1–2), pp. 141–153, 2000.
- [8] P. Crosasso, M. Ceruti, P. Brusa, S. Arpicco, F. Dosio, and L. Cattel, "Preparation, characterization and properties of sterically stabilized paclitaxel-containing liposomes," *Journal of Controlled Release*, vol. 63(1–2), pp. 19–30, 2000.
- [9] S. C. Kim, D. W. Kim, Y. H. Shim, J. S. Bang, H. S. Oh, S. Wan Kim, and M. H. Seo, "In vivo evaluation of polymeric micellar paclitaxel formulation: toxicity and efficacy," *Journal of Controlled Release*, vol. 72(1–3), pp. 191–202, 2001.
- [10] Y. Liu, G. S. Chen, L. Li, H. Y. Zhang, D. X. Cao, and Y. J. Yuan, "Inclusion complexation and solubilization of paclitaxel by bridged bis(beta-cyclodextrin)s containing a tetraethylenepentaamino spacer," *Journal of Medicinal Chemistry*, vol. 46(22), pp. 4634–4637, 2003.
- [11] B. R. Goldspiel, "Pharmaceutical issues: preparation, administration, stability, and compatibility with other medications," *Annals of Pharmacotherapy*, vol. 28(5 Suppl), pp. S23-26, 1994.

- [12] E. K. Rowinsky, L. A. Cazenave, and R. C. Donehower, "Taxol: a novel investigational antimicrotubule agent," *Journal of National Cancer Institute*, vol. 82(15), pp. 1247–1259, 1990.
- [13] H. Gelderblom, J. Verweij, K. Nooter, and A. Sparreboom, "Cremophor EL: the drawbacks and advantages of vehicle selection for drug formulation," *European Journal of Cancer*, vol. 37(13), pp. 1590–1598, 2001.
- [14] A. Albert, "Chemical aspects of selective toxicity," *Nature*, vol. 182(4633), pp. 421–422, 1958.
- [15] C. E. Muller, "Prodrug approaches for enhancing the bioavailability of drugs with low solubility," *Chemistry & Biodiversity*, vol. 6(11), pp. 2071–2083, 2009.
- [16] K. C. Nicolaou, C. Riemer, M. A. Kerr, D. R. Rideout, and W. W. Wrasidlo, "Design, synthesis and biological activity of protaxols," *Nature*, vol. 364(6436), pp. 464, 1993.
- [17] R. B. Greenwald, A. Pendri, D. Bolikal, and C. W. Gilbert, "Highly water soluble taxol derivatives: 2'-polyethyleneglycol esters as potential prodrugs," *Bioorganic Medicinal Chemistry Letters*, vol. 4(20), pp. 2465–2470, 1994.
- [18] R. B. Greenwald, C. W. Gilbert, A. Pendri, C. D. Conover, J. Xia, and A. Martinez, "Drug Delivery Systems: Water Soluble Taxol 2'-Poly(ethylene glycol) Ester Prodrugs Design and in Vivo Effectiveness," *Journal of Medicinal Chemistry* vol. 39(2), pp. 424–431, 1996.
- [19] K. C. Nicolaou, R. K. Guy, E. N. Pitsinos, and W. Wrasidlo, "A Water-Soluble Prodrug of Taxol with Self-Assembling Properties," *Angewandte Chemie International Edition in English*, vol. 33(15–16), pp. 1583–1587, 1994.
- [20] A. A. Moosavi-Movahedi, S. Hakimelahi, J. Chamani, G. A. Khodarahmi, F. Hassanzadeh, F. T. Luo, and G. H. Hakimelahi, "Design, synthesis, and anticancer activity of phosphonic acid diphosphate derivative of adenine-containing butenolide and its water-soluble derivatives of paclitaxel with high antitumor activity," *Bioorganic Medicinal Chemistry*, vol. 11(20), pp. 4303–4313, 2003.
- [21] R. Todeschini, V. Consonni, "*Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References*. 2009. John Wiley & Sons.
- [22] V. E. Kuz'min, A. G. Artemenko, P. G. Polischuk, E. N. Muratov, A. I. Hromov, A. V. Liahovskiy, and S. Y. Makan, "Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure," *Journal of Molecular Modelling*, vol. 11(6), pp. 457–467, 2005.

- [23] M. O. Akinwande, H. G. Dikko, and A. Samson, "Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis," *Open Journal System*, vol. 5(07), pp. 754, 2015.
- [24] M. Cama, I. Cristi Nicu, C. Conoscenti, G. Queneherve, and M. Maerker, "The role of multicollinearity in landslide susceptibility assessment by means of Binary Logistic Regression: comparison between VIF and AIC stepwise selection," *EGU General Assembly Conference Abstracts*, vol. 18, 2016.
- [25] A. Frish, A. B. Nielsen, and A. J. Holder, "Gauss View User Manual, Gaussian Inc.; Pittsburg, PA", 2001.
- [26] N. Y. Golovenko, I. Y. Borisyuk, M. A. Kulinskiy, P. G. Polishchuk, E. N. Muratov, and V. E. Kuz'min, "Quantitative Structure-Property Relationship Analysis of Drugs' Pharmacokinetics Within the Framework of Biopharmaceutics Classification System Using Simplex Representation of Molecular Structure," in *Application of Computational Techniques in Pharmacy and Medicine*, pp. 461-499, Springer, Dordrecht, 2014.
- [27] S. F. Zhou, B. Wang, L. P. Yang, and J. P. Liu, "Structure, function, regulation and polymorphism and the clinical significance of human cytochrome P450 1A2," *Drug Metabolism Reviews*.
- [28] S. Sansen, J. K. Yano, R. L. Reynald, G. A. Schoch, K. J. Griffin, C. D. Stout, and E. F. Johnson, "Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2," *Journal of Biological Chemistry*, vol. 282(19), pp. 14348–14355, 2007.
- [29] D. S. Goodsell, G. M. Morris, and A. J. Olson, "Automated docking of flexible ligands: applications of AutoDock," *Journal of Molecular Recognition*, vol. 9(1), pp. 1–5, 1996.
- [30] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry* vol. 26(16), pp. 1701–1718, 2005.
- [31] A. W. Schuttelkopf, and D. M. F. van Aalten, "PRODRG: a tool for high-throughput crystallography of protein-ligand complexes," *Acta Crystallographica Section D*, vol. 60(Pt 8), pp. 1355–1363, 2004.
- [32] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6," *Journal of Computational Chemistry*, vol. 25(13), pp. 1656–1676, 2004.

- [33] W. F. Van Gunsteren, and H. J. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation*, vol. 1(3), pp. 173-185, 1998.
- [34] H. J. Berendsen, J. V. Postma, W. F. van Gunsteren, A. R. H. J. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81(8), pp. 3684-3690, 1984.
- [35] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of chemical physics*, vol. 72(4), pp. 2384-2393, 1980.
- [36] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118(45), pp. 11225-11236, 1996.
- [37] H. G. Petersen, "Accuracy and efficiency of the particle mesh Ewald method," *Journal of Chemical Physics*, vol. 103(9), pp. 3668-3679, 1995.
- [38] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14(1), pp. 33-38, 1996.
- [39] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, "UCSF Chimera--a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 23(13), pp. 1605-1612, 2004.
- [40] J. Lever, M. Krzywinski, and N. Altman, "Points of Significance: Principal component analysis," *Nature Methods*, vol. 14, pp. 641-642, 2007.
- [41] R. Kumari, R. Kumar, "Open Source Drug Discovery Consortium, & Lynn, A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations," *Journal of Chemical Information and Modelling*, vol. 54(7), pp. 1951-1962, 2014.
- [42] J. J. Huuskonen, A. E. P. Villa, I. V. Tetko, "Prediction of partition coefficient based on atom-type electrotopological state indices," *Journal of Pharmaceutical Sciences*, vol. 88(2), pp. 229-233, 1999.
- [43] P. Labute, "A widely applicable set of descriptors," *Journal of Molecular Graphics and Modelling*, vol. 18(4), pp. 464-477, 2000.
- [44] Yalkowsky SH, He Y, Jain P, He Y, Jain P. *Handbook of Aqueous Solubility Data*. 2016.
- [45] Devinyak O, Havrylyuk D, Lesyk R. 3D-MoRSE descriptors explained. *Journal of Molecular Graphics and Modelling*, vol. 54, pp. 194-203, 2014.
- [46] A. Varnek, A. Tropsha, "Chemoinformatics Approaches to Virtual Screening," *Royal Society of Chemistry*, 2008.

- [47] P. Gramatica, “WHIM Descriptors of Shape,” *QSAR and Combinatorial Science*, vol. 25(4), pp. 327–332, 2006.
- [48] A. M. Helguera, R. D. Combes, M. P. Gonzalez, M. Cordeiro, “Applications of 2D descriptors in drug design: a DRAGON tale,” *Current Topics in Medicinal Chemistry*, vol. 8(18), pp. 1628–1655, 2008.

CHAPTER 4

MOLECULAR DOCKING AND SIMULATION STUDIES OF PACLITAXEL PRODRUGS WITH CYTOCHROME 3A4 TO CORRELATE SOLUBILITY AND BIOAVAILABILITY THROUGH PHYSICOCHEMICAL CHARACTERIZATION

SUMMARY

Prodrugs are biologically inactive drug molecules that may be developed through rational drug design with an objective to improve a drug's pharmaceutical and pharmacokinetic properties. Paclitaxel a highly potent anticancer drug, is directed against many cancers like breast cancer, ovarian cancer etc. Along with its excellent antitumor activity the drug had a major limitation of low water solubility. To overcome this limitation of this nanomolar active drug many prodrugs were formed in the past. Though increase in the solubility of the drug was obtained but that does not account for its increase in bioavailability. CYP3A4 liver enzymes are responsible for the metabolism of fifty percent of the drugs and are a major metabolizing enzyme for paclitaxel. Phosphate prodrugs well are known to account the insolubility of many drugs and thus increasing their bioavailability also. In this study we calculated the ADMET properties of a dataset of twenty phosphate prodrugs of paclitaxel. On the basis of reflection of three favourable, ten prodrugs were chosen for further docking studies against CYP3A4. Finally, three prodrugs showing unfavourable binding affinities were selected for Molecular Dynamics Simulations and from this *in-silico* study identification of all the three selected prodrugs as unstable compared to paclitaxel. The instability of these prodrugs showed their lesser interaction with the CYP3A4 and hence contributing more towards its bioavailability. Thus the three suggestive prodrugs those were studied *in-silico* for oral bioavailability can be further validated for gastrointestinal cancer and this computational study gives an insight for the formation of new paclitaxel prodrugs with increased bioavailability.

4.1 Introduction

Paclitaxel a proven anticancer drug its potential is explored and found the drug to be nanomolar active for various types of cancers and specifically uses for ovarian, breast, and lung cancer [1]. Conventionally, the anti-cancer activity of paclitaxel resides in its ability to stabilize microtubules, and therefore arrest cellular proliferation. The mechanism of action of paclitaxel is different than other microtubule interacting drugs; it misleads the tumor cell into passing the G1/S checkpoint and into another cycle of DNA replication. DNA replication in the absence of cytokinesis is called endoreduplication and results in the generation of giant cells with 4N, 8N, 16N, and even 32N DNA content, this leads subsequently to delayed apoptosis [2]. Also, paclitaxel posses the ability to interact with the tubulin at low temperatures and also can interact with the beta tubulin in the absence of Guanosine-5'-triphosphate (GTP), a microtubule associated protein [3]. Apart from microtubule binding paclitaxel induce non-cytoskeletal effects also that could restrain malignancy as well. Such as, paclitaxel activates macrophages enhancing their anticancer effector functions. The pairing of paclitaxel and a “priming” signal, such as interferon c (IFNc), trigger macrophages in vivo to lyse tumor cells [4].

After establishing the potential of this nanomolar active drug towards various types of cancers and its unique mechanism of action to interact with tubulin and able to arrest cells in mitosis and later the cell undergoes apoptosis had aroused much interest for further development. Despite being nanomolar active against various types of cancers the drug posses a major limitation of low aqueous solubility [5]. The chemical nature of drug is highly hydrophobic and because of its poor solubility it has to be solubilised in unfriendly carriers such as cremophor EL that run the risk of hypersensitivity reactions; such as hyperlipidemia, aggregation of erythrocytes, peripheral neuropathy, and abnormal lipoprotein patterns [6]. To avoid this hypersensitivity and obtain better clinical use of paclitaxel, developing a new co-solvent and improving the formulation for paclitaxel delivery systems had became important. In addition, for developing better formulation for paclitaxel there were ways to achieve more efficient modified paclitaxel. Extensive research had been carried out in the past few years, for the formation of new approaches of paclitaxel delivery systems. Different carriers were involved for the formation of better delivery systems for paclitaxel like nanoparticle [7], micelles [8], liposomes [9], lipid cells [10], β -cyclodextrins [11,12], cubosomes [13], and microplate [14]. Along with these drugs formulating techniques there were ways to achieve more efficient modified paclitaxel that is prodrug formation. Prodrug design is a widely

known molecular modification strategy that aims to optimize the physicochemical, and pharmacological properties of drugs to improve their solubility, pharmacokinetic properties and decrease the toxicity of the drugs [15,16]. The prodrug formation modes can overcome the limitation of paclitaxel's low aqueous solubility. In the past, many paclitaxel prodrugs were formed, that included the formation of diverse small and large molecule prodrugs. Small molecule prodrug includes the formation of paclitaxel carbonates, phosphates, silicates, proteins, sugars and polymers were employed for the macromolecule prodrug formation [17-20]. Also, the targeting potential and retention effects (EPR) of paclitaxel could be increased with macromolecular prodrugs [21]. Different approaches to build-up paclitaxel prodrugs were introduced making it more soluble but the approach of formation of these prodrugs was not rational as many undesirable compounds were formed and also all these approaches were very tedious to follow, hence QSPR models were developed for the solubility prediction of paclitaxel prodrug [22] which may assist in the formation of new prodrugs with better solubility. The solubility of the molecule gets increased by the addition of various substituent groups; also it affects the bioavailability of the molecule.

The prodrug design is a widely known molecular modification strategy that aims to optimize the physicochemical and pharmacological properties of drugs to improve their solubility and pharmacokinetic features. Prodrug approach has been very successful for obtaining the better clinical use of paclitaxel, the development of new molecules with increased solubility hence with enhanced bioavailability of the drugs; assimilation of those was limited because of low aqueous solubility and low permeability. Paclitaxel prodrugs with increasing solubility include common hydrophilic groups such as phosphates, carbonates, silicates, etc. Phosphate esters offer a way to increase the solubility as well as the bioavailability for many sparingly soluble drugs [23]. Phosphate esters are relatively stable than the other esters those are commonly used for the formation of prodrugs, as they have low pK_a (1-2) [24,25,26] hence they can tolerate the acidic conditions prevailing in the stomach, absorbed by gut wall, and passes to the liver [27,28].

The most common liver enzymes, cytochrome P450 are major drug metabolic enzymes that are expressed mainly in liver [29]. Different variants of cytochrome P450 are extensively responsible for the kinetic delineation of any compound that depends on the phase I metabolism [30]. CYP3A4 isoform belongs to a class of heme-thiolate enzymes present in the humans, and that are responsible for the metabolism of fifty percent of the compounds [31]. CYPs carry out various reactions to accomplish the metabolism of various drugs and drug like

molecules such as aromatic oxidation, epoxidation, N-, S-, O-dealkylation, aliphatic hydroxylation, and S- N-oxidation [32]. Although, the main P450 enzymes responsible for the metabolism of paclitaxel and its prodrugs are CYP2C8 and CYP3A4 but the 6 α -hydroxypaclitaxel a major metabolite formed by CYP2C8 is not the predominant metabolite in all individuals, and CYP3A4 that is responsible for the formation of other metabolites C3'-hydroxypaclitaxel and C2-hydroxypaclitaxel may play a primary role in the metabolism of paclitaxel and its prodrugs [33].

In the present study, the interaction of paclitaxel, its phosphate prodrugs, and a highly soluble prodrug of paclitaxel; isotaxel that is 5000 times more soluble as compared to paclitaxel with CYP3A4 is studied *in-silico* and characterizes the bioavailability of the phosphate prodrugs to the intestinal cancerous cells. CYP3A4 plays a major role in drug absorption and it acts as a barrier to the prodrugs to reach the intestinal cancerous cells. For the drug to be more bioavailable to the intestinal cancerous cells they should not get metabolised in the liver by the CYP3A4 and thus reaches to the intestine where they would exhibit their anti-neoplastic effects and finally led to cell-death. ADMET properties of a set of twenty paclitaxel phosphate prodrugs were calculated using the admetSAR web server. The best way to understand the precise interaction pattern of prodrugs with the receptor is molecular docking. After the initial screening considering the different pharmacokinetic properties, molecular docking of the selected ten phosphate prodrugs those were showing more three more favourable pharmacokinetic properties; oral bioavailability, mutagenesis, and CYP3A4 inhibition was carried out. Molecular dynamics simulations (MDS) of the selected three prodrugs those were showing moderate interaction with the CYP3A4 in the docking study was performed using Gromacs 5.0. Various criterion like root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA), principal component analysis (PCA), and binding free energy were evaluated to study the interaction pattern of these prodrugs with the CYP3A4. The unstable complexes of prodrugs with CYP3A4 as revealed by the MDS study thus explained the more bioavailability of these paclitaxel phosphate prodrugs to the intestinal cancerous cells and thus can reach the site of action and subsequently lead to apoptosis. The computational calculation of ADMET properties, interaction pattern, and instability studies of paclitaxel phosphate prodrugs suggested the potential candidate prodrugs that could be delivered orally and are further suggested for effective use against gastrointestinal cancer (GIT).

4.2 Materials and Methods

The step-wise methodology followed for *in-silico* study of interaction pattern for prodrugs and CYP3A4 and characterizing them to be more bioavailable is described as follow (Figure 4.1).

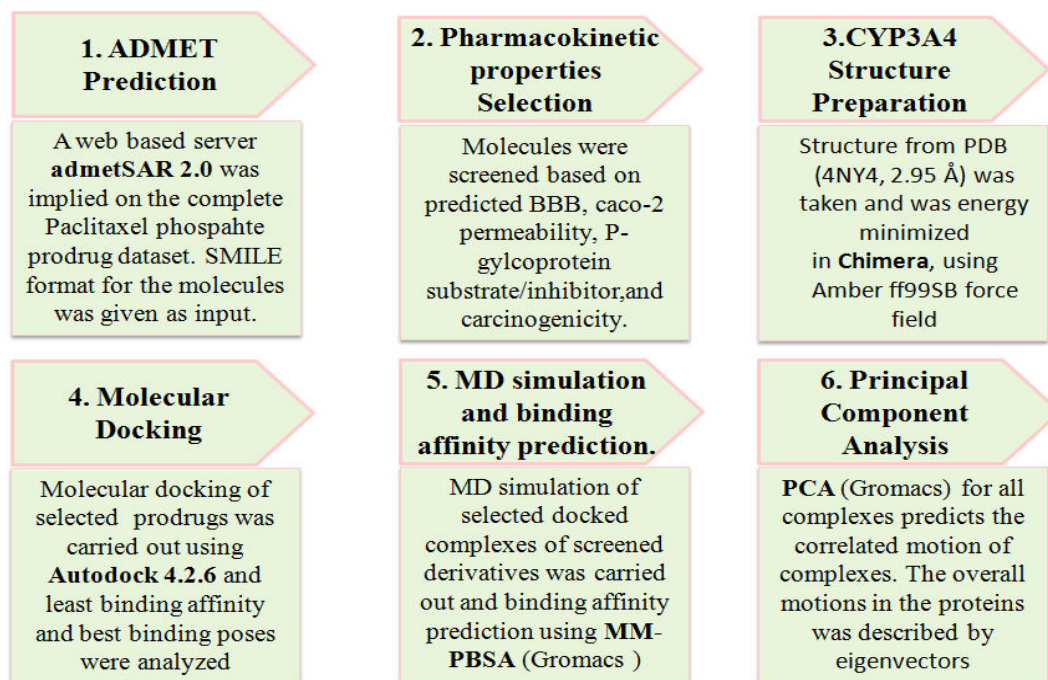


Figure 4.1: Pharmacokinetic prediction, molecular docking and molecular dynamics simulation study of CYP3A4 with the paclitaxel phosphate prodrugs.

4.2.1 Protein & paclitaxel phosphate prodrugs data collection and geometry optimization

A dataset of twenty diverse paclitaxel phosphate prodrugs having solubility better or comparable with paclitaxel was collected from literature [17-20]. The initial structures of all the prodrugs were drawn in GaussView [34] and the molecular structures were optimized at the level of parameterization method 6 (PM6) semi-empirical methods in Gaussian 09 quantum chemistry software [35].

Also, the structure of CYP3A4 (PDB ID: 2J0D, 2.75 Å) was taken from RCSB (<http://www.rcsb.org/>). Structurally, CYP3A4 is a homodimer with two identical chains. It has a small N-terminal domain that is rich in β -sheet and larger C-terminal domain that is comprised of mainly α -helices and contains the active site [36]. The secondary structure of enzyme has 268 or 55% amino acid residues engaged in forming α -helices and alone 35 or 7% residues are engaged in forming β -sheets [36]. The α -helices follow the P450 naming convention that includes the naming of helices in alphabetical designation starting from the N-terminus [37]. The structural similarity of CYP3A4 is found to be most common to the

enzyme structures of mammalian family 2 P450s [38]. But notable differences between the overall structure of CYP3A4 and other family 2 P450s are observed in the regions of helix D to H and the helix B to C. These two regions presume different conformations of the enzymes of family 2 constantly in response to the identity of bound ligands which illustrate their flexibility and structural diversity. In the CYP3A4, F and G helices that constitutes the upper bound of active site are truncated that results in extended unstructured linkers to the mediating F' and G' helices. The unique Phe cluster of six phenylalanine residues, three from the linker region interacts with phenylalanines from helix F', the B-C region, and helix I is localized in the upper region of active site and form the primary residues of peripheral binding region [39]. Also, the presence of a hydrophobic region, residues 36-50 in the N-terminal region of the globular enzyme is the principal structural deviation of CYP3A4 from other eukaryotic CYP450 superfamily enzymes. The hydrophobic region is thought to form the interaction region with microsomal membrane along with some parts of F'G' loop [39]. Chimera 1.13.1 software [40] was employed for carrying out the minimization of the complete protein structure implementing the Amber ff99SB force field.

4.2.2 Interpretation of CYP3A4 binding site

The most interesting feature of CYP3A4 is the general size and shape of active-site cavity, present adjacent to the heme group. The active site cavities of 2C8 and 3A4 are found to be 1438 Å³ and 1386 Å³ in volume respectively when cropped at the narrowest constrictions of solvent channels [41]. The comparison of active-site volume of CYP3A4 with CYP2C8 was performed as both share almost the same cavity volume to oxidize the comparably large substrates. Difference in the secondary and tertiary structures of the protein is followed by the difference in the shapes of cavities that presumably affect the substrate selectivity and enzyme catalysis [41]. The two enzymes differ in the architecture of active-sites as the structure of 3A4 is much more open in the proximity of heme group and the volume of cavity is distributed much more uniformly as compared to the undulating volume of CYP2C8 [41].

The structure of CYP3A4 presents a large substrate-binding cavity that is consistent with the sizes of substrates that are oxidized by the enzyme such as erythromycin (M_r 734), cyclosporin (M_r 1203), bromocryptine (M_r 654). The Phe cluster present at the roof of active-site contributes to the homo cooperativity of the enzyme like multiple molecules of testosterone (M_r 288) or progesterone (M_r 314) can possibly be accommodated [41]. The protein model indicated (PDB id 2J0D, 2.05 Å) the active site residues in the structure are Tyr53, Phe57, Arg106, Phe108, Met114, Arg119, Leu120, Leu210, Leu211, Arg212,

Phe213, Asp214, Phe215, Phe220, Val240, Phe241, Ile301, Phe304, Ala305, Thr309, Ile369, Ala370, Arg372, Leu373, Glu374 [39]. The prodrugs of paclitaxel ($M^f > 833$) were docked in the above described active site of the protein model CYP3A4.

4.2.3 ADMET prediction of paclitaxel phosphate prodrugs

Estimation of pharmacokinetic properties is an important step in drug discovery process. The drugs should exhibit advantageous Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) criterion. Well-known *in-vitro* and *in-vivo* based computational models are used for the estimation of ADMET parameters. admetSAR v2.0 (<http://Immd.ecust.edu/admetSar2>) [42], a web-based server was used for carrying out the prediction for the pharmacokinetic properties of paclitaxel phosphate prodrugs. Simplified Molecular-Input Line Entry System (SMILE) format for all the molecules was provided as input to admetSAR 2.0. Various pharmacokinetic properties such as caco-2 cell permeability, human intestinal absorption, blood-brain barrier, Pgp substrate/inhibitor, carcinogenicity, mutagenicity etc. were predicted by the server (Appendix C, CT1).

4.2.4 Molecular Docking study

Pharmacokinetic aspect of the molecules was studied by ADMET properties; from the paclitaxel phosphate prodrug dataset ten phosphate prodrugs were selected and were implied to molecular docking against CYP3A4 in the above defined active site in AutoDock 4.2.6 [43]. Also, AutoDock Vina [44] was used for the molecular docking of the selected prodrugs to establish the consensus in the docking study. AutoDock uses free energy of docking molecules implying 3-dimensional potential grids. AutoDock grid preparation module was used for the generation of a 3-dimensional grid box around the predefined active-site residues and a grid box of dimensionality X=50, Y=76, and Z=52 grid points including the grid spacing of 0.375 Å was generated. AutoDock employs Lamarckian genetic algorithm for carrying out molecular docking and uses empirical scoring function for the ranking of docked molecules. Protein-ligand binding conformation was used for the calculation of intermolecular energy starting from the unbound protein-ligand conformation. All the molecules of substituent group dataset were set for docking with the population size of 150, number of maximum generations 27,000, maximum evaluations of 2,50,000, with a gene mutation rate of 0.02 and crossover rate of 0.8. All the molecules were employed for the generation of 100 binding poses for the respective docking. Binding affinity was taken as the parameter for determining the docking and the best docked poses. Also, all the selected prodrugs were redocked using the fast and efficient AutoDock Vina program. AutoDock Vina

used the defined box center around the defined active-site residues. It uses the ‘Iterated Local Search global optimizer’ for optimization. The algorithm works in a number of successive steps subsists of a mutation and a local optimization, with the acceptance of each step according to the metropolis criterion. All the molecules were bestowed for the 9 runs and the binding free energy was calculated scoring function that takes into accounts both knowledge-based potentials and empirical scoring functions.

4.2.5 Conformation stability analysis

Selected docked complexes of phosphate prodrugs with CYP3A4 exhibiting less binding affinity than paclitaxel were further subjected to MDS for studying the stability of the docked complexes. Gromacs v5.1.2 [45], was used for performing the MDS as long as 1 microsecond to validate the stability of protein-ligand complexes. ProDrg server [46] was used for obtaining the protein and ligand topologies for all the five substituent groups using GROMOS96 53a6 forcefield parameters [47]. The initial structures of all the five systems were solvated in a cubic box of SPC water model (three point model having three interaction sites that corresponds to three atoms of water molecule) [48] and were charge neutralized by five chloride ions. The obtained structures were energy minimized for 50,000 steps with steepest descent algorithm, pursued by NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) equilibration of 0.1 ps, each one using the leap-frog integrator algorithm [49]. The Brendensen thermostat [50] algorithm was used to control the simulation temperature, and Parrinello-Rahman isotropic pressure coupling [51] was practiced for controlling pressure.

GROMOS96 53a6 all atom force field [52] was used for the MD simulation for 1 microsecond time-scale to observe time dependent trajectories of systems in explicit solvent. The production MD was set again to time of 0 ps and the entire run was for 1microsecond, with coordinates, velocities, energies stored at every 20 ps. The integration time-step was 2 fs, treating all atoms explicitly and were coupled with LINCS constraint algorithm for restraining hydrogen and heavy atom bonds, the van der Waal cutoff was observed at 0.9 nm, and the Verlet cutoff scheme was used with Coloumb, with particle mesh ewald (PME) long range electrostatics [53,54]. Initial velocities were allowed on the basis of Maxwell distribution at 300 K with 1 nm minimum distance enforced in all the three directions within the periodic boundary conditions. The root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), hydrogen bonds (H-bonds), and solvent accessible surface area (SASA) calculations were done using the in-built gmx rms, gmx rmsf, gmx gyration,

gmx h_bond and gmx sasa tools, respectively. The final MD trajectories were obtained using visual molecular dynamics (VMD) [55] and Chimera v1.13.1 [56] and Origin Pro 6.0 was used for plotting the respective graphs.

4.2.6 Principal Component Analysis

Projection of first two principal components (PCs) for the docked complexes kept on MD simulation for 100 ns was also analysed using the built-in principal component analysis (PCA) or essential dynamics method of Gromacs, also eigenvectors, eigenvalues, eigRMSF were evaluated. PCA is utilized to analyze the relationship between different conformers on the basis of equivalent residues of the conformers/structures. The correlated motions of the proteins that majorly define the protein functions were observed by the PC analysis and the principal components were grouped according to the protein functions [57]. All the translational and rotational movements of the complexes were removed and covariance matrices for all the complexes were formed. The following equation was used for the computation of elements of the positional covariance matrix C:

$$C_{ij} = (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \quad (i, j = 1, 2, \dots, 3N) \quad \text{Eqn 3}$$

The i^{th} C_α atom was represented by the Cartesian coordinate q_{ij} and N being the number of atoms. The equilibrated trajectory that was superimposed on a reference structure, overall translations and rotations were removed using ‘least-square fit’ method. All the matrices were diagonalized adopting the orthogonal co-ordinate transformation matrix Λ to predict the remaining eigenvectors and eigenvalues λ_i .

$$\Lambda = T^T C_{ij} T \quad \text{Eqn 4}$$

In the above equation, eigenvectors corresponds to the direction of relative motion of $\langle q_i \rangle$ were represented in the columns. Each eigenvector associated with the eigenvalues representing the total mean-square fluctuation of the system along the corresponding eigenvector. Eigenvalue corresponding to each in brief, the ensuing principal components (orthogonal eigenvectors) characterize the axes of maximal variance for structures distribution. Projection of the distribution onto the subspace defined by the best principal components concludes in representing the structural dataset in a lower dimension.

The obtained last 650 ns trajectories were analysed for the calculation of eigenvector and eigenvalues using gmx covar and gmx ana eig in-built tools of Gromacs. The free energy of

the systems implementing the first two principal components was performed using `gmx_sham` tool [58].

4.2.7 Computing MM-PBSA binding free energy

Molecular Mechanic/Poisson-Boltzman Surface Area (MM-PBSA) method was used for the computation of binding free energy of the protein-ligand interactions. `g_mmpbsa` tool [59] of the Gromacs was used for the calculation of binding free energy for the last 100 ns trajectories snapshots obtained from MDS, that is an average of three energy terms; potential energy in the vacuum, polar-solvation energy and non-polar solvation energy in the case of protein-ligand binding, the binding energy of the bound system is an average.

4.3 Results and discussion

The solubility of these prodrugs had increased but the enzymatic actions in the body are also important determinant that further determines the bioavailability of the compounds. A set of test class that is well-organized was used for providing insight that how a drug behaves in the human body. Results were analyzed systematically as explained in further sections.

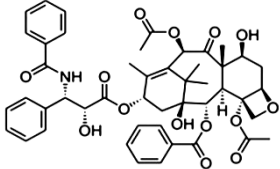
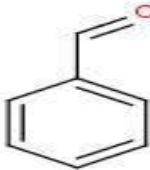
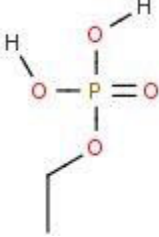
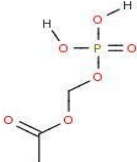
4.3.1 Pharmacokinetics analysis

The ADMET criterion is implied to paclitaxel phosphate prodrug dataset consisting of twenty molecules using `admetSAR 2.0`. Various properties were computed by the server such as BBB, HIA, Caco-2 cell permeability, Pgp substrate/inhibitor etc. On the basis favourable pharmacokinetic properties oral bioavailability, mutagenesis, and CYP3A4 inhibition ten prodrugs from the dataset were selected for further docking studies.

4.3.2 Molecular Docking studies

CYP3A4 enzyme act as major metabolic enzymes for the xenobiotics. CYP3A4 metabolize fifty percent of the drugs in the body, thus to study the bioavailability of the paclitaxel prodrugs we carried out the interaction study of CYP3A4 enzyme that is having the larger active site as compared to the other CYPs in the family as the paclitaxel molecule is a large molecule and thus so the prodrugs formed were also bulkier. After the first screening of the molecules implying ADMET properties, `Autodock 4.2.6` was used for the docking of the selected ten molecules from dataset against CYP3A4. Considering binding energies of all the molecules we choose three prodrugs: 2' phosphonoxy methyl carbonate paclitaxel, 2' phosphonoxy methyl ether derivative and isotaxel from the ten molecule dataset. Results obtained from the docking studies of molecules are presented below (Table 4.1 and Figure 4.2). Residues in bold represent the hydrogen bond forming residues (Table 4.1).

Table 4.1 Docking study result of the selected paclitaxel prodrugs with PM6 optimized geometry against CYP3A4 showing the binding affinity and the interacting residues

S No.	Paclitaxel and its Substituent groups	Substituent at 2' site * (structure2-4)	Binding energy by AutoDock (kJ.mol ⁻¹)	Binding energy by AutoDockVina (kJ.mol ⁻¹)	Interacting residues
1	Paclitaxel		-50.54	-48.28	Arg 105, Arg 106, Phe 108, Ser 119, Arg 212 , Phe 213, Phe 215, Ile 369, Ala 370 , Met 371, Arg 372, Leu 373, Pro434, Phe 435, Asn 441, Cys 442 , Gly 481, Leu 482.
2	Isotaxel		-32.71	-34.72	Phe 57, Arg 105, Arg 106, Phe 108, Ser 119, Arg 212 , Phe 213, Phe 215, Phe 304, Ala 305, Thr 309, Ile 369, Ala 370 , Met 371, Arg 372, Glu 374, Phe 435, Cys 442, Gly 481, Leu 482.
3	2' Phosphonoxy methyl ether derivative of Paclitaxel		-31.67	-32.63	Phe 57, Arg 105 , Arg 106 , Phe 108, Ser 119 , Ile 120, Arg 130, Arg 212 , Phe 215, Phe 241, Ile 301, Ile 369, Ala 370, Met 371, Leu 373, Glu 374 , Arg 375, Pro 434, Phe 435, Gly 436, Ser 437, Arg 440, Asn 441, Cys 442, Gly 481, Leu 482, Leu 483.
4	2' Phosphonoxy methyl carbonate Paclitaxel		-31.50	-43.93	Arg 105, Arg 106 , Phe 108, Ser 119, Ile 120, Arg 212, Phe 215, Ala 305, Thr 309, Ile 369, Ala 370, Arg 372, Lue 373, Glu 374 , Cys 442 , Ala 448.

The docking results obtained by two different docking algorithms AutoDock and AutoDock Vina had shown the consensus among the binding values for the chosen molecules from the dataset of phosphate prodrugs. Paclitaxel binds with differing residues of CYP3A4 and made hydrogen bonds with Arg212, Ala370, and Cys442 forming complex that gave binding energy of -50.54 kJ.mol⁻¹ and -48.28 kJ.mol⁻¹ by AutoDock and AutoDock Vina respectively. The complex formed was uphold by some non-polar interplay of Phe 108, Arg130, Gly 481, etc. residues with CYP3A4 (Figure 4.2a and Table 4.1). Phosphonoxy methyl carbonate binds with various residues of the CYP3A4 and forms hydrogen bonds with Arg106, Glu374, and Cys442 with complex giving binding energy of -31.5 Kcal.mol⁻¹ and -43.93 kJ.mol⁻¹ by AutoDock and AutoDock Vina respectively. The complex was stabilized by various hydrophobic interactions also with residues Phe215, Arg372, Leu373 etc. (Figure 4.2b and Table 4.1). Similarly phosphonoxy methyl ether derivative of Paclitaxel makes the docking

complex with CYP3A4 forming hydrogen bonds with Arg105, Arg106, Ser119, Arg212, Glu374 residues and also hydrophobic interactions giving the binding energy of $-31.67 \text{ kJ.mol}^{-1}$ and $-32.63 \text{ kJ.mol}^{-1}$ by AutoDock and AutoDock Vina respectively (Figure 4.2c and Table 4.1). The complex of isotaxel with CYP3A4 enzyme giving the binding energy of $-32.71 \text{ kJ.mol}^{-1}$ and $-34.72 \text{ kJ.mol}^{-1}$ and by AutoDock and AutoDock Vina respectively accompanying the polar interactions of residues Arg212, Phe304, Ala370 and hydrophobic interactions with the residues Phe213, Ile369, Met371 etc (Figure 4.2d and Table 4.1).

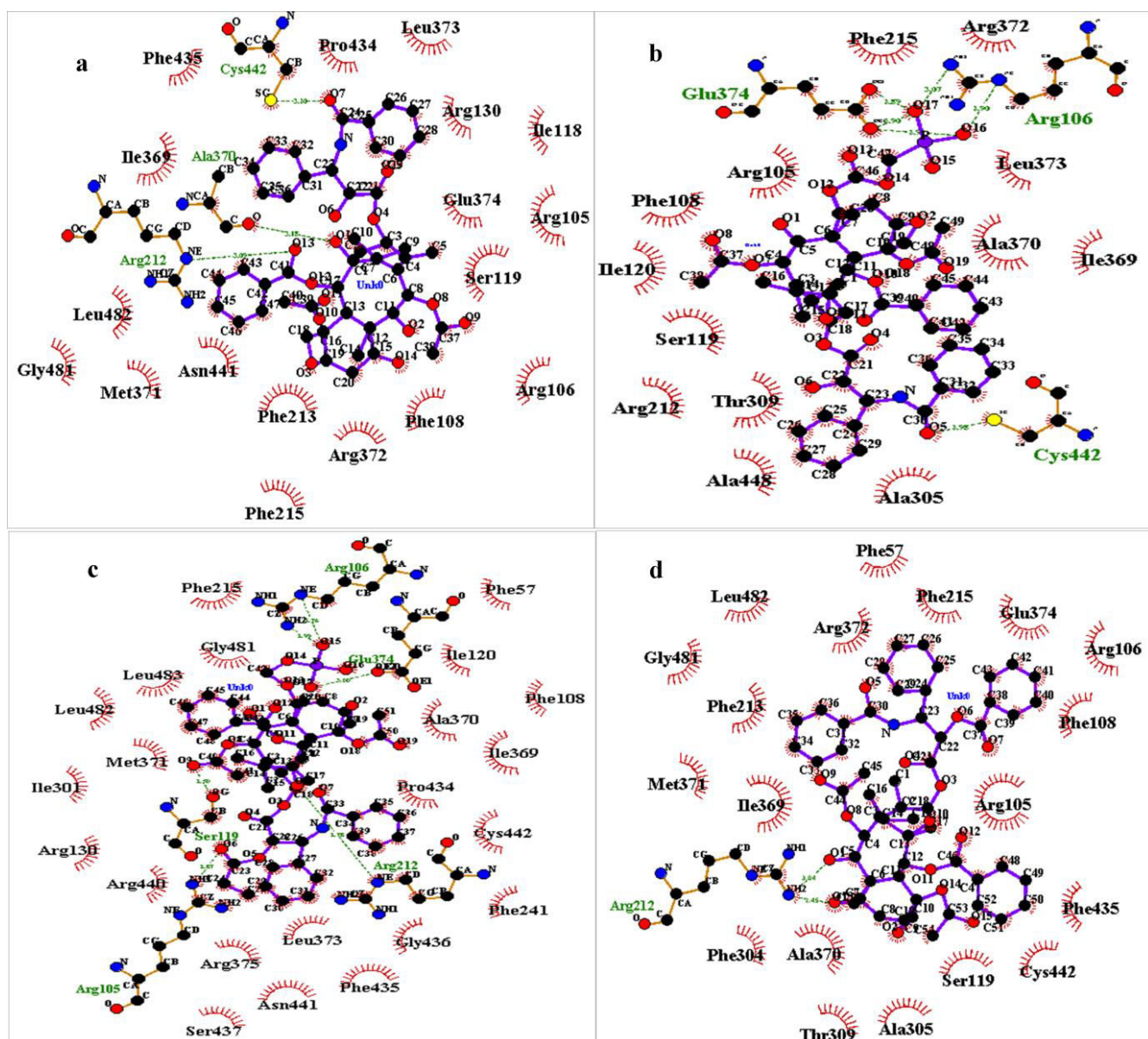


Figure 4.2: Ligand interaction diagrams. [a] Paclitaxel with CYP3A4; [b] 2' Phosphonoxy methyl carbonate paclitaxel with CYP3A4; [c] 2' Phosphonoxy methyl ether derivative of paclitaxel with CYP3A4 [d] Isotaxel with CYP3A4.

4.3.3 Molecular Dynamics Simulation

Molecular dynamics of the docked complex is a computerized method for studying the physical movements of the atoms while binding [60]. Five molecular dynamics simulation (MDS) for the four docked complexes, and one apo-protein for 1 microsecond was carried out. Equilibration state for all the systems had obtained and the trajectories were observed for RMSF, Rg and SASA H-bonds for the last 350 ns equilibrated trajectories including last 100 ns binding free energy analysis.

Root mean square deviation (RMSD)

Dynamic stability of all the systems is described by the RMSD; it calculates the changes in the protein's backbone framework during the simulation time scale. All the five systems were occupied for the calculation of RMSD values and the trajectories of all the systems were observed after 650 ns as shown in Figure 4.3a. Average values of RMSD for apo-CYP3A4, CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4-2' Phosphonoxy methyl ether derivative of paclitaxel, and CYP3A4-Isotaxel were 0.48880 nm, 0.4648 nm, 0.4417 nm, 0.4513 nm, and 0.5125 nm respectively. RMSD pattern observed for the prodrugs is more comparable to the apo-protein and paclitaxel. RMSD of the CYP3A4-isotaxel is highly deviating from the CYP3A4-Paclitaxel complex thus representing that the formed prodrug is forming less stable complex with CYP3A4 as compared to paclitaxel. CYP3A4 is a metabolizing enzyme and the formed prodrugs are representing the less stability than the paclitaxel molecule hence the result obtained from computational study proposes that the prodrugs are more bioavailable than paclitaxel as they may not get metabolized in the liver before reaching the site of action [61,62].

RMSF Fluctuation

Residue mobility calculation for the concluding 350 ns equilibrated trajectory was done after the RMSD calculation as illustrated in Figure 4.3b. The average RMSF values obtained for apo-CYP3A4, CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4-2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel were 0.1656 nm, 0.1270 nm, 0.1410 nm, 0.1352 nm and 0.2786 nm respectively. Paclitaxel showed good RMSF value and the prodrug isotaxel showed highest RMSF value amongst all including paclitaxel. Other prodrugs also showed higher RMSF values than the CYP3A4-Paclitaxel complex. It had been observed that the prodrugs were not forming well stable complex than paclitaxel and inducing the highest amount of fluctuation [61,62].

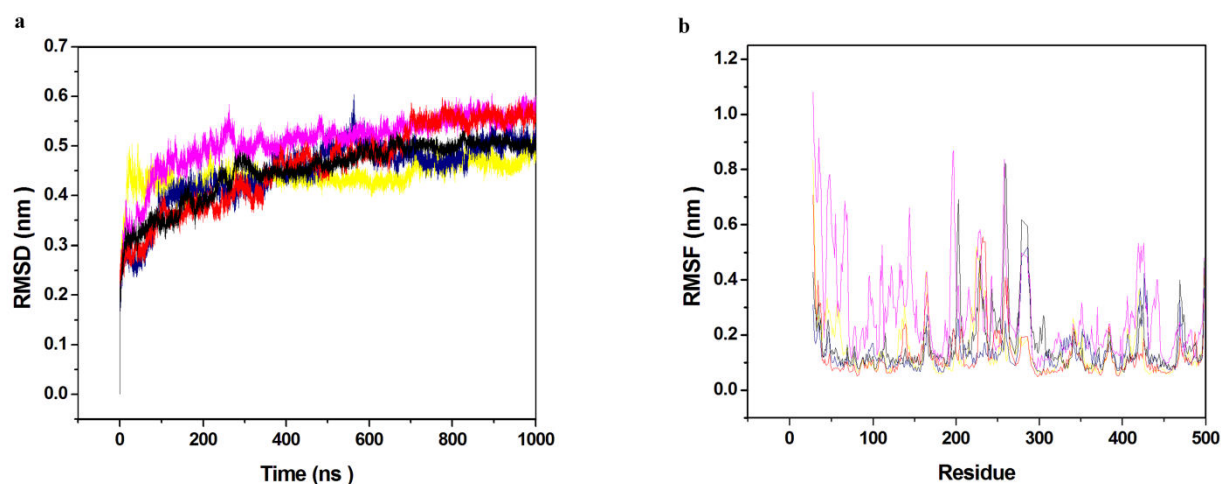


Figure 4.3: Molecular dynamics simulation. (a) RMSD of the C α backbone of paclitaxel prodrugs over the 1000 ns MDS at 300 K, (b) RMSF of residues of paclitaxel prodrugs during MDS. In all panels the color code is:

CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), and CYP3A4-Isotaxel (magenta).

Hydrogen Bonds

Protein-ligand stability is majorly defined by the hydrogen bonds as they contribute the transient interactions to the complex formation, which provide stability to the protein-ligand complex. In the current study, we had calculated the number of hydrogen bonds formed by the complexes while binding. The number of hydrogen bonds formed in the complexes in the time scale is plotted in Figure 4.4a. Moderately the number of hydrogen bonds formed for CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4- 2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel were 0-8, 0-3, 0-4 and 0-2 respectively. CYP3A4-Isotaxel shows the least number of hydrogen bonds as comparable to the CYP3A4-Paclitaxel complex and other prodrugs also showed less number of hydrogen bond formation in the complex.

Compactness Analysis

Rg factor is used to define the compactness of the protein-ligand complexes all along the simulation time-scale. It is a distance measure between the center of mass of the protein atoms with its terminal atoms in a time frame. In general, the compact protein shows a lesser amount of variation in the gyration value while the diffused structure shows higher Rg value. In the current study, we plotted the Rg values vs. time for all the complexes (Figure 4.4b). The average Rg values obtained for CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4-2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel were 22.568, 22.616, 22.776 and 23.318 respectively. All the prodrugs showed Rg

values higher than the apo-protein. It indicates that the prodrug complexes are less stable as compared to the paclitaxel. On this basis, we can infer that the modelling study suggest the prodrugs to be more bioavailable than the paclitaxel molecule which can be further validated experimentally. As CYP3A4 is a major metabolizing enzyme and the formed prodrugs were found to be less stable than the paclitaxel molecule in the computational studies hence the result suggests that the prodrugs were more bioavailable than paclitaxel as they may not get metabolized in the liver before reaching the site of action [61,62].

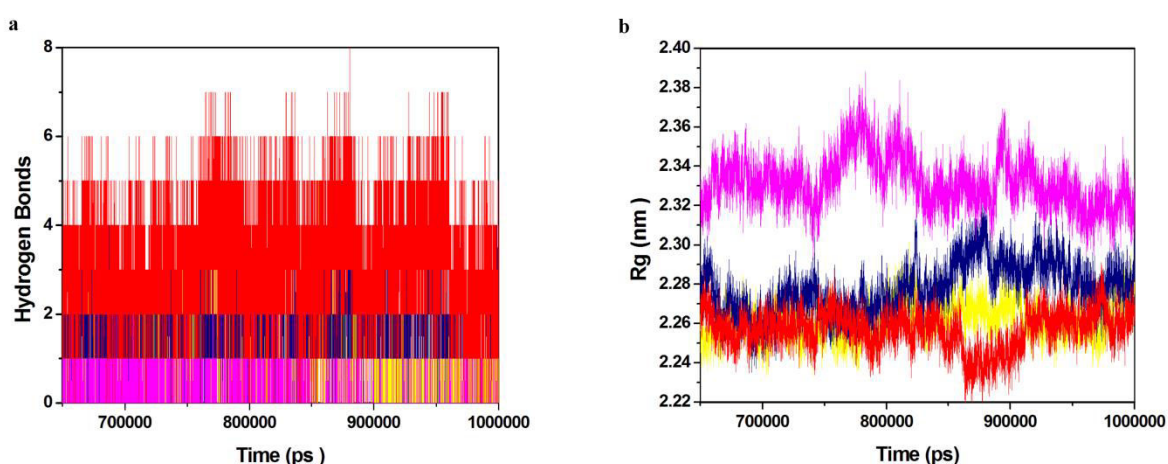


Figure 4.4: Molecular dynamics simulation. (a) Number of hydrogen bonds interaction between protein and ligand during simulation time scale for paclitaxel and other ligand complexes, (b) Rg vs time of paclitaxel prodrugs during MDS. In all panels the color code is: CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2'-Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta).

Solvent accessible surface area

Solvation free energy of a protein is the result of interaction among polar and non-polar residues of the protein. The SASA is that surface area of the protein which is monitored by the solvent molecule's probe when it seeks the Van der Waals surface of the protein. Hydrophobic residues are most responsible for the increment of the SASA value. The SASA value of each complex was calculated from the last 350 ns equilibrated trajectory and shown in Figure 4.5a. We got the average SASA values of 213.0139, 220.0228.54, 222.05273 and 230.5543 nm² for CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4-2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel respectively. The average SASA value for all the prodrugs which is higher than the paclitaxel showed that all the prodrugs are providing the conformational instability while binding of ligand. All the prodrugs showing more SASA values compared to the paclitaxel, representing

that all the prodrugs producing instability on binding to the CYP3A4 thus proposing them to be more bioavailable [61,62].

An important parameter, residue SASA value that give an understanding of conformational change per residue contribution. A graph had been plotted between residues and SASA value (Figure 4.5b). The average residue SASA values gained for apo-CYP3A4, CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4- 2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel were 0.4728, 0.4630, 0.4783, 0.4837 and 0.5012 nm² respectively. All the prodrugs showing more residual SASA as compared to the paclitaxel, thus there was reflection of instability by all the prodrugs on binding to the CYP3A4 [61,62].

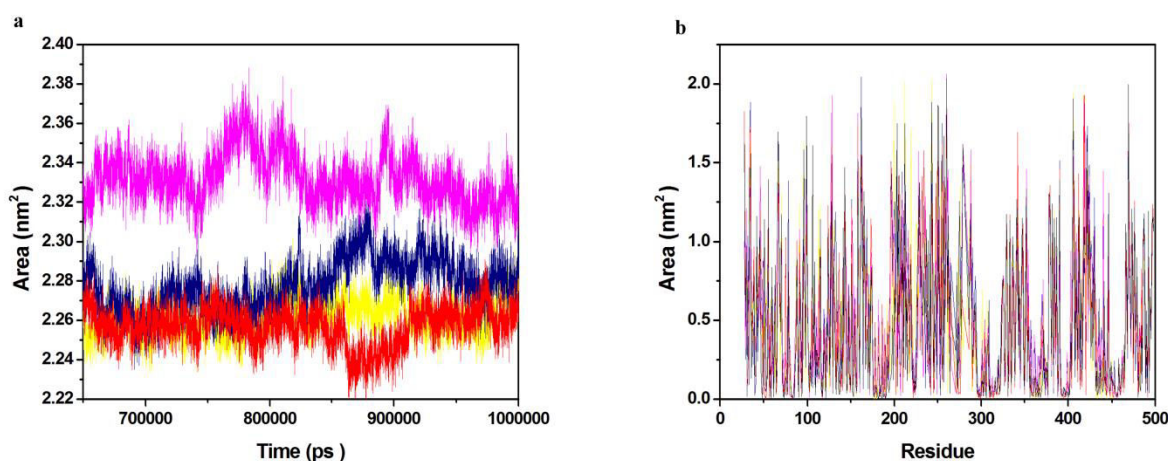


Figure 4.5 Solvent accessible surface area. (a) The SASA plot for paclitaxel and all other complexes in water with respect to time, (b) The Residue SASA plot for paclitaxel and other ligand complexes. In all panels, the color code is CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'-Phosphonoxy methyl carbonate paclitaxel (yellow), CYP3A4-2' Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta).

Principal Component Analysis

In the MDS, the principal component analysis predicts the correlated motion of the complexes, in the calculation of the Principal Components (PCs) the overall motions in the proteins is described by a few key eigenvectors. Hence in the study, we had plotted the eigenvalues with the eigenvectors for the 50 eigenvectors in Figure 4.6a. The first five eigenvectors of apo-CYP3A4, CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4- 2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel explained 63.15%, 64.91%, 61.85%, 59.47%, 64.53% of the motions respectively for the last 350 ns equilibrated trajectory. It was observed that all the prodrugs showed

comparable motions as compared to the paclitaxel. Hence, the modelled results are suggesting the instability of prodrugs towards CYP3A4.

As previously stated, we had taken 50 eigenvectors for predicting the correlated motion of the complexes and had observed that the first five eigenvectors are the key for studying overall protein dynamics. The first two eigenvectors were plotted opposite to each other in a phase space where each one of the spectrum represents the correlated motions (Figure 4.6b). The denser cluster characterizes the stability of the complex while the dispersed cluster represents the less stable cluster. It was seen from the results obtained that CYP3A4-Isotaxel showed the diffused cluster among all the other prodrugs also the other two prodrugs are less stable as compared to the CYP3A4-Paclitaxel complex. It was also in agreement with the above described results [61,62].

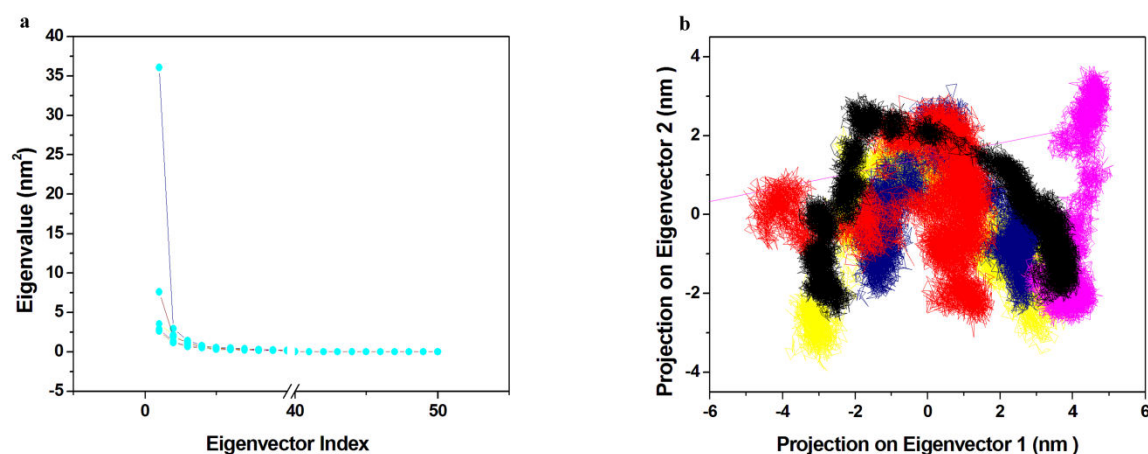


Figure 4.6 Principal component analysis. (a) The plot of eigenvalues vs. eigenvector index. First 50 eigenvectors were considered (B) Projection of the motion of the protein in phase space along the PC1 and PC2, In all panels the color code is: CYP3A4 (black) and the ligands CYP3A4-Paclitaxel (red), CYP3A4-2'Phosphonoxy methyl carbonate (yellow), CYP3A4-2'Phosphonoxy methyl ether derivative (blue), CYP3A4-Isotaxel (magenta).

Gibbs free energy landscape

The Gibbs free energy landscape can serve to give the information for the binding of a ligand to the protein. For PC1 and PC2 it was calculated to read the conformational state changes during ligand binding, Figure 4.7. The free energy values were ranged from 0 to 7.87, 0 to 8.17, 0 to 8.02, 0 to 8.43 and 0 to 10.1 $\text{kJ}\cdot\text{mol}^{-1}$ for apo-CYP3A4, CYP3A4-Paclitaxel, CYP3A4-2' Phosphonoxy methyl carbonate, CYP3A4-2' Phosphonoxy methyl ether derivative of paclitaxel and CYP3A4-Isotaxel respectively. The bluer region represents the high energy state structure. The apo-protein reflects bluer region as compared to all the prodrugs while CYP3A4-Isotaxel showed the energy range from 0 to 10.1 $\text{kJ}\cdot\text{mol}^{-1}$ that is

highest among all the prodrugs. It had been concluded from the above results that all the prodrugs are energetically unfavourable [61,62].

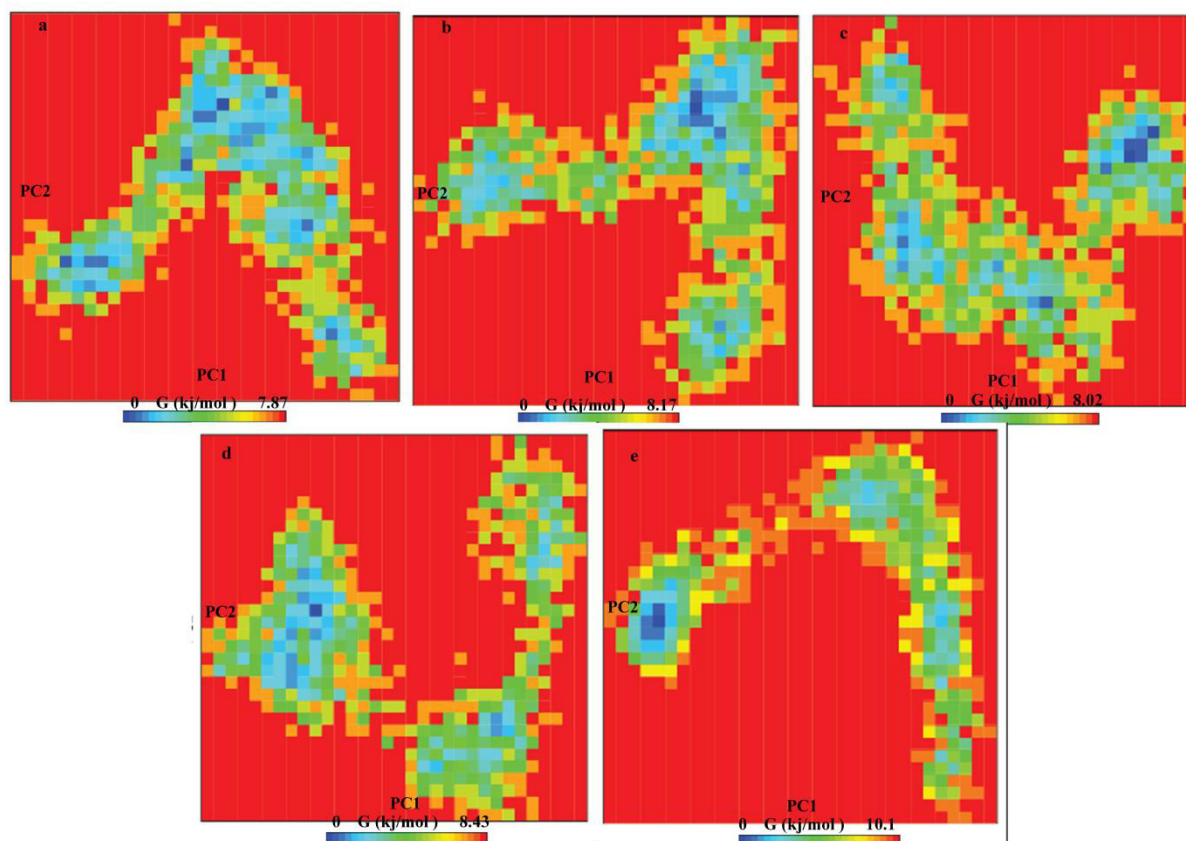


Figure 4.7: Gibbs free energy landscape. The free energy landscape calculated for PC1 and PC2. In the figure, (a) CYP3A4, (b) CYP3A4-Paclitaxel, (c) CYP3A4-2' Phosphonoxy methyl carbonate paclitaxel , (d) CYP3A4-2' Phosphonoxy methyl ether derivative, (e) CYP3A4-Isotaxel.

4.3.3 MMPBSA binding free energy analysis

The binding ability of the ligands to the protein is assessed with the help of binding free energy calculation which is the sum of non-bonded interaction energies of the complex. MMPBSA tool [59] that is directed by the Gromacs, was adopted for the calculation of binding free energy. All the binding free energies of the complexes are summarized in Table 2, Here we had concluded that all the prodrugs were showing comparable binding energy to the paclitaxel. Isotaxel showed the binding energy relatively less than paclitaxel. Hence this prodrug is most favorable with respect to the pharmacology of the compound and is most bioavailable than the other prodrugs, however other two prodrugs were also showing respectively good bioavailability as depicted from Table 2.

Table 4.2: Table represents the Van der Waals, electrostatic, polar solvation, SASA and binding energy in kJ.mol^{-1} for control compound and prodrugs

S. No	Compound	Van der Waals energy	Electrostatic Energy	Polar solvation energy	SASA energy	Binding energy
1	Paclitaxel	-426.88 ± 12.82	-35.59 ± 5.46	234.42 ± 14.49	-35.89 ± 1.18	-263.94 ± 15.94
2	Phosphonoxy methyl carbonate Paclitaxel	-439.82 ± 18.86	-60.11 ± 14.13	256.37 ± 22.76	-37.75 ± 1.67	-281.39 ± 22.82
3	2'-Phosphonoxy methyl ether derivative	-451.64 ± 16.89	-25.27 ± 7.82	160.77 ± 11.89	-39.22 ± 1.60	-355.36 ± 16.50
4	Isotaxel	-346.18 ± 15.28	-25.02 ± 6.93	151.55 ± 15.81	-37.72 ± 1.87	-257.37 ± 13.72

Molecular docking and dynamics studies of paclitaxel prodrugs with CYP3A4 for correlating bioavailability was done. Various parameters like RMSD, RMSF, Rg, SASA, PCA and binding free energy were analyzed for all the complexes using the Gromacs inbuilt tools and all the results obtained by the *in-silico* studies were in good opinion and are in agreement that the bioavailability of all the three paclitaxel prodrugs is more than the paclitaxel which can further be validated using the biochemical assays and animal models. Solubility and bioavailability could be achieved by careful selection followed by *in vitro* and *in vivo* evaluation of the prodrug.

4.4 Conclusion

Oral prodrugs should have high chemical stability, provide improved solubility, permeability, and quantitatively convert to parent drug quickly to minimize unwanted metabolism and maximize drug exposure. Paclitaxel a proven anticancer drug but with limitations for its solubility and finally bioavailability was studied through QSPR strategies and other computational investigations. Increase in solubility accounts for the bioavailability of the drug. The bioavailability of the drug is also turn to the interaction of the prodrugs with the human body enzymes. CYP3A4 which is a metabolizing enzyme accounts for the metabolism of the drugs and is responsible for the bioavailability of the prodrugs. The prodrugs must not be broken into the parent drug and the group that is added to it before reaching to the site of action. Here in our current study, we provided a docking and simulation-based model for the bioavailability prediction of the prodrugs. A series of *in-silico* studies on the prodrugs of the Paclitaxel for increasing the bioavailability of the molecules were performed. The results obtained from the docking and simulation studies favored that the isotaxel molecule which

had solubility 5000 times more than the paclitaxel is exhibiting good bioavailability, as it is showing an unstable complex with the CYP3A4. Results for other two molecules were also promising and all three molecules are suggested for the improved bioavailability. It is thus expected that the computational docking and simulation studies for the prediction of bioavailability of the prodrugs will be a great support for the synthetic chemist for the formation of new paclitaxel prodrugs.

References

- [1] M. C. Wani, H. L. Taylor, and M. E. Wall, et al, "Plant antitumour agents, VI. The isolation and structure of taxol, a novel antileukemic and antitumour agent from *Taxus Brevifolia*," *Journal of American Chemical Society*, vol. 93(9), pp. 2325-2327, 1971.
- [2] M. D. De Furia, "Paclitaxel (Taxol®): A New Natural Product with Major Anticancer Activity," *Phytomedicine*, vol 4(3), pp. 273–282, 1997.
- [3] G. Maccari, M. Mori, R. S. Javier, W. Fang, J. F. Díaz, and M. Botta, "Free Energy Profile and Kinetics Studies of Paclitaxel Internalization from the Outer to the Inner Wall of Microtubules," *Journal of Chemical Theory and Computation*, vol. 9(1), pp. 698–706, 2013.
- [4] T. M. O. Chan, L. X. Yang, "The immunological effects of taxanes," *Cancer Immunology, Immunotherapy*, vol. 49(4-5), pp. 181-185, 2000.
- [5] B. R. Goldspiel, "Pharmaceutical Issues: Preparation, Administration, Stability, and Compatibility with Other Medications," *Annals of Pharmacotherapy*, vol. 28(5 Suppl), pp. S23-S26, 1994.
- [6] H. Gelderblom, J. Verweij, K. Nooter, and A. Sparreboom, "Cremophor EL: The Drawbacks and Advantages of Vehicle Selection for Drug Formulation," *European Journal of Cancer*, vol 37(13), pp. 1590–1598, 2001.
- [7] S. K. Sahoo, and V. Labhasetwar, "Enhanced Antiproliferative Activity of Transferrin-Conjugated Paclitaxel-Loaded Nanoparticles Is Mediated via Sustained Intracellular Drug Retention," *Molecular Pharmaceutics*, vol 2(5), pp. 373–383, 2005.
- [8] Y. Shi, F. van Nostrum Cornelus, and W. E. Hennink, "Interfacially Hydrazone Cross-Linked Thermosensitive Polymeric Micelles for Acid-Triggered Release of Paclitaxel," *ACS Biomaterial Science and Engineering*, vol 1(6), pp. 393–404, 2015.
- [9] Y. Liu, L. Mei, Q. Yu, et al, "Multifunctional Tandem Peptide Modified Paclitaxel-Loaded Liposomes for the Treatment of Vasculogenic Mimicry and Cancer Stem Cells in Malignant Glioma," *ACS Applied Materials and Interfaces*, vol. 7(30), pp. 16792–16801, 2015.
- [10] N. Joshi, R. Saha, T. Shanmugam, B. Balakrishnan, P. More, and R. Banerjee, "Carboxymethyl-Chitosan-Tethered Lipid Vesicles: Hybrid Nanoblanket for Oral Delivery of Paclitaxel," *Biomacromolecules*, vol. 14(7), pp. 2272–2282, 2013.
- [11] Y. Liu, G. S. Chen, Y. Chen, D. X. Cao, Z. Q. Ge, and Y. J. Yuan, "Inclusion Complexes of Paclitaxel and Oligo(Ethylenediamino) Bridged Bis(Beta-

- Cyclodextrin)s: Solubilization and Antitumor Activity,” *Bioorganic Medicinal Chemistry*, vol. 12(22), pp. 5767–5775, 2004.
- [12] A. O. Boztas, O. Karakuzu, G. Galante, et al, “Synergistic Interaction of Paclitaxel and Curcumin with Cyclodextrin Polymer Complexation in Human Cancer Cells,” *Molecular Pharmaceutics*, vol. 10(7), pp. 2676–2683, 2013.
- [13] S. Aleandri, D. Bandera, R. Mezzenga, and E. M. Landau, “Biotinylated Cubosomes: A Versatile Tool for Active Targeting and Codelivery of Paclitaxel and a Fluorescein-Based Lipid Dye,” *Langmuir*, vol. 31(46), pp. 12770–12776, 2015.
- [14] A. V. Singh, M. Batuwangala, R. Mundra, et al, “Biomaterialized Anisotropic Gold Microplate-Macrophage Interactions Reveal Frustrated Phagocytosis-like Phenomenon: A Novel Paclitaxel Drug Delivery Vehicle,” *ACS Applied Materials and Interfaces*, vol. 6(16), pp. 14679–14689, 2014.
- [15] N. Choy, Y. Shin, P. Q. Nguyen, and D. P. Curran, et al, “Simplified Discodermolide Analogues: Synthesis and Biological Evaluation of 4-Epi-7-Dehydroxy-14,16-Didemethyl-(+)-Discodermolides as Microtubule-Stabilizing Agents,” *Journal of Medicinal Chemistry*, vol. 46(14), pp. 2846–2864, 2003.
- [16] M. D. Wittman, T. J. Altstadt, C. Fairchild, et al, “Synthesis of Metabolically Blocked Paclitaxel Analogues,” *Bioorganic and Medicinal Chemistry Letters*, vol. 11(6), pp. 809–810, 2001.
- [17] R. B. Greenwald, A. Pendri, D. Bolikal, and C. W. Gilbert, “Highly Water Soluble Taxol Derivatives: 2'-Polyethyleneglycol Esters as Potential Prodrugs,” *Bioorganic Medicinal Chemistry*, vol. 4(20), pp. 2465–2470, 1994.
- [18] R. B. Greenwald, C. W. Gilbert, A. Pendri, C. D. Conover, J. Xia, and A. Martinez, “Drug Delivery Systems: Water Soluble Taxol 2'-Poly(Ethylene Glycol) Ester Prodrugs-Design and in Vivo Effectiveness,” *Journal of Medicinal Chemistry*, vol. 39(2), pp. 424–431, 1996.
- [19] K. C. Nicolaou, C. Riemer, M. A. Kerr, D. Rideout, and W. Wrasidlo, “Design, Synthesis and Biological Activity of Protaxols,” *Nature*, vol. 364(6436), pp. 464–466, 1993.
- [20] A. A. Moosavi-Movahedi, S. Hakimelahi, J. Chamani, et al, “Design, Synthesis, and Anticancer Activity of Phosphonic Acid Diphosphate Derivative of Adenine-Containing Butenolide and Its Water-Soluble Derivatives of Paclitaxel with High Antitumor Activity,” *Bioorganic Medicinal Chemistry*, vol. 11(20), pp. 4303–4313, 2003.
- [21] V. P. Podduturi, I. B. Magana, D. Patrick O’Neal, and P. A. Derosa, “Simulation of Transport and Extravasation of Nanoparticles in Tumors Which Exhibit Enhanced Permeability and Retention Effect,” *Computer Methods and Programs in Biomedicine*, vol. 112(1), pp. 58–68, 2013.
- [22] N. S. Munjal, M. Sharma, and T. R. Singh, “Development of QSPR Strategy for the Solubility Prediction,” *Current Computer-aided drug designing*, vol. 14(4), pp. 302–309, 2018.
- [23] S. J. Hecker, and M. D. Erion, “Prodrugs of phosphates and phosphonates,” *Journal of Medicinal Chemistry*, vol. 51(8), pp. 2328–2345, 2008.

- [24] D. H. Jornada, G. F. dos Santos Fernandes, D. E. Chiba, T. R. F. De Melo, J. L. Dos Santos, and M. C. Chung, "The prodrug approach: A successful tool for improving drug solubility," *Molecules*, vol. 21(1), pp. 42, 2016.
- [25] C. Schultz, "Prodrugs of biologically active phosphate esters," *Bioorganic & medicinal chemistry*, vol. 11(6), pp. 885-898, 2003.
- [26] A. J. Wiemer, and D. F. Wiemer, "Prodrugs of Phosphonates and Phosphates: Crossing the Membrane Barrier," in *Montchamp JL. (eds) Phosphorus Chemistry I*, vol 360, Springer, Cham, 2014.
- [27] K. Ghosh, D. M. Tagore, R. Anumula, B. Lakshmaiah, P. P. B. S. Kumar, S. Singaram, and M. Ramarao, "Crystal structure of rat intestinal alkaline phosphatase—role of crown domain in mammalian alkaline phosphatases," *Journal of structural biology*, vol. 184(2), pp. 182-192. 2013.
- [28] J. P. Krise, and V. J. Stella, "Prodrugs of phosphates, phosphonates, and phosphinates," *Advanced Drug Delivery Reviews*, vol. 19(2), pp. 287-310, 1996.
- [29] Lynch T and Price A (2007). The Effect of Cytochrome P450 Metabolism on Drug Response, Interactions, and Adverse Effects. *Am Fam Physician*, 76(3), 391–396.
- [30] F. P. Guengerich, "Cytochrome p450 and chemical toxicology," *Chemical research in toxicology*, vol. 21(1), pp. 70-83 2007.
- [31] B. Monsarrat, P. Alvinerie, M. Wright, J. Dubois, F. Gueritte-Voegelein, D. Guenard, and E. K. Rowinsky, "Hepatic metabolism and biliary excretion of Taxol in rats and humans," *Journal of the National Cancer Institute. Monographs*, vol. (15), pp. 39-46, 1993.
- [32] B. Monsarrat, E. Chatelut, I. Royer, P. Alvinerie, J. Dubois, A. Dezeuse, and P. Canal, "Modification of paclitaxel metabolism in a cancer patient by induction of cytochrome P450 3A4," *Drug metabolism and disposition*, vol. 26(3), pp. 229-233, 1998.
- [33] D. S. Sonnichsen, Q. Liu, E. G. Schuetz, J. D. Schuetz, A. Pappo, and M. V. Relling, "Variability in human cytochrome P450 paclitaxel metabolism," *Journal of Pharmacology and Experimental Therapeutics*, vol. 275(2), pp. 566-575, 1995.
- [34] C. Tabti, and N. Benhalima N, "Molecular Structure, Vibrational Assignments and Non-Linear Optical Properties of 4,4' Dimethylaminocyanobiphenyl (DMACB) by DFT and ω-B3LYP/6-31G(d) HF Calculations," *Advances in Materials Physics and Chemistry*, vol. 05(07), pp. 221–228, 2015.
- [35] R. A. Kwiecien, M. Rostkowski, A. Dybała-Defratyka, and P. Paneth P, "Validation of Semiempirical Methods for Modeling of Corrinoid Systems," *Journal of Inorganic Biochemistry*, vol. 98(6), pp. 1078–1086, 2004.
- [36] P. A. Williams, J. Cosme, D. M. Vinkovic, A. Ward, H. C. Angove, P. J. Day, and H. Jhoti, "Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone," *Science*, vol. 305(5684), pp. 683-686, 2004.
- [37] M. Ekroos, and T. Sjogren, "Structural basis for ligand promiscuity in cytochrome P450 3A4," *Proceedings of the National Academy of Sciences*, vol. 103(37), pp. 13682-13687, 2006.

- [38] H. Park, S. Lee, and J. Suh, "Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4," *Journal of the American Chemical Society*, vol. 127(39), pp. 13634-13642, 2005.
- [39] J. K. Yano, M. R. Wester, G. A. Schoch, K. J. Griffin, C. D. Stout, and E. F. Johnson, "The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution," *Journal of Biological Chemistry*, vol. 279(37), pp. 38091-38094, 2004.
- [40] E. F. Pettersen, T. D. Goddard, C. C. Huang, et al, "UCSF Chimera--a Visualization System for Exploratory Research and Analysis," *Journal of Computational Chemistry*, vol. 25(13), pp. 1605-1612, 2004.
- [41] E. E. Scott, and J. R. Halpert, "Structures of cytochrome P450 3A4," *Trends in biochemical sciences*, vol. 30(1), pp. 5-7, 2005.
- [42] H. Yang, C. Lou, L. Sun, et al, "AdmetSAR 2.0: Web-Service for Prediction and Optimization of Chemical ADMET Properties," *Bioinformatics (Oxford, England)*, 2018.
- [43] D. S. Goodsell, G. M. Morris, and A. J. Olson, "Automated Docking of Flexible Ligands: Applications of AutoDock," *Journal of Molecular Recognition*, vol. 9(1), pp. 1-5, 1996.
- [44] O. Trott, and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of computational chemistry*, vol. 31(2), pp. 455-461, 2010.
- [45] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, Flexible, and Free," *Journal of Computational Chemistry*, vol. 26(16), pp. 1701-1718, 2005.
- [46] A. W. Schuttelkopf, and M. F. van Aalten Daan, "PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes," *Acta Crystallographica Section D*, vol. 60(Pt 8), pp. 1355-1363, 2004.
- [47] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, "A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6," *Journal of Computational Chemistry*, vol. 25(13), pp. 1656-1676, 2004.
- [48] M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19-25, 2015.
- [49] W. F. Van Gunsteren, and H. J. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation*, vol. 1(3), pp. 173-185, 1998.
- [50] H. J. Berendsen, J. V. Postma, W. F. van Gunsteren, A. R. H. J. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81(8), pp. 3684-3690, 1984.
- [51] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of chemical physics*, vol. 72(4), pp. 2384-2393, 1980.
- [52] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic

- liquids,” *Journal of the American Chemical Society*, vol. 118(45), pp. 11225-11236, 1996.
- [53] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, “Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides,” *The Journal of Physical Chemistry B*, vol. 105(28), pp. 6474-6487, 2001.
- [54] H. G. Petersen, “Accuracy and Efficiency of the Particle Mesh Ewald Method,” *Journal of Chemical Physics*, vol. 103(9), pp. 3668–3679, 1995.
- [55] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual Molecular Dynamics,” *Journal of Molecular Graphics*, vol. 14(1), pp. 33–38, 1996.
- [56] E. F. Pettersen, T. D. Goddard, and C. C. Huang, et al, “UCSF Chimera--a Visualization System for Exploratory Research and Analysis,” *Journal of Computational Chemistry*, vol. 25(13), pp. 1605–1612, 2004.
- [57] J. Lever, M. Krzywinski, and N. Altman, “Points of significance: Principal component analysis,” *Nature Methods*, vol. 14, pp. 641–642, 2017.
- [58] H. J. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: a message-passing parallel molecular dynamics implementation,” *Computer physics communications*, vol. 91(1-3), pp. 43-56, 1995.
- [59] R. Kumari, R. Kumar, Open Source Drug Discovery Consortium, and A. Lynn, “g_mmpbsa--A GROMACS tool for high-throughput MM-PBSA calculations,” *Journal of chemical information and modeling*, vol. 54(7), pp. 1951-1962, 2014.
- [60] S. Pall, M. J. Abraham, C. Kutzner, B. Hess, and Lindahl, “Tackling exascale software challenges in molecular dynamics simulations with GROMACS,” in *International Conference on Exascale Applications and Software*, pp. 3-27, Springer, Cham, 2014.
- [61] C. C. Ogu, and J. L. Maxa, “Drug Interactions Due to Cytochrome P450,” *Proc (Bayl Univ Med Cent)*, vol. 13(4), pp. 421–423, 2000.
- [62] H. van de Waterbeemd and E. Gifford, “ADMET in Silico Modelling: Towards Prediction Paradise?” *Nature Reviews Drug Discovery*, vol. 2(3), pp. 192–204, 2003.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Conclusion

Cancer is one of the leading causes of death now-a-days globally. Cancer cells are characterized by uncontrolled growth, invasive intrusion, destruction of adjacent tissues and spread to other body locations via blood and/or lymph. This disease may affect the people of all the ages but risk for all the ages for various cancers increases throughout the world. Cancer statistics, as reported by the Globocan 2018 report that this devastating disease affects the whole world population and all types of cancers have been reported effecting the males and female population with apparently different kinds of tumors. Since healthcare providers has gained remarkable knowledge in this field and are searching for better ways to stop the growth of tumor cells and the cure for this devastating disease. Although many advances has been achieved in treatment, prevention of reoccurrence, and obtaining the palliative care for a variety of cancers. Among all the treatment regimens, chemotherapy is widely used for the treatment of cancer. Chemotherapy is the general term for any treatment involving the use of chemical agents to stop cancer cells from growing. These chemical agents can work by any of the mechanisms such as interfering with DNA metabolism, cell division or signal transduction.

Paclitaxel, one of the most interesting anticancer drugs is being used in chemotherapy from years. This natural product drug with unique mechanism of action and dramatic anti-cancer activity is a complex diterpenoid with specific structural features vital for its activity. Paclitaxel targets a variety of tumors including ovarian, breast, lung, colon, and also melanoma and lymphoma. The major limitation with this nanomolar active drug is its extremely low solubility that prevented its entrance into the clinic widely. Many efforts in the past were made to increase the solubility of the drug while accommodating the anti-cancer activity of the molecule. However, structural modifications and pharmaceutical formulations were the main approaches for obtaining the better clinical properties of the molecule, also formation of surfactants, emulsions, liposomes, cyclodextrins, and polymers were done so the solubility of paclitaxel can be improved.

Aqueous solubility that tends to be one of the most imperative physicochemical properties is one of the biological factors that affect the bioavailability of a biological molecule hence

imparting a risk in the potent drug formulation. Taxol that is the commercial formulation of paclitaxel, constitutes two components; a surfactant, cremophor EL (polyethoxylated castor oil) and water-free ethanol. Both these excipients are used for enhancing the solubility of paclitaxel and the prepared formulation is given intravenously to the patients. Along with that cremophor EL imparts many clinical side effects such as hyperlipidemia, anaphylactoid hypersensitive reactions, aggregation of erythrocytes, etc. Hence the clinical application of paclitaxel has been widely affected by its side-effects including that is occurred due to cremophor EL. It has been reported that this type of hypersensitivity reactions is noted in all the drugs those use cremophor EL as a formulation vehicle. Therefore, an increased interest in developing new formulations of paclitaxel has elevated the need to understand the basic physicochemical properties of paclitaxel.

Prodrug design is a widely known molecular modification strategy that aims to optimize the physicochemical and pharmacological properties of drugs to improve their solubility, pharmacokinetic features, and decrease their toxicity. The chemical structure of paclitaxel has a bulky, composed and fused ring together with number of hydrophobic substituent groups making it highly lipophilic in nature. As for prodrug strategy, understanding the structure and activity relationship is the first step. Pioneer work showed that the central part of paclitaxel is rigid to a change in structure, while its side chain tail can be flexible. Structure activity relationship (SAR) studies explains 2' and/or 7' hydroxyl groups are the successful distinctive sites for making an addition to paclitaxel structure. Paclitaxel prodrug formation to increase the solubility of the drug had been sought for obtaining its effective application. Many hydrophilic substituent groups like carboxylic acids, phosphates, sugar derivatives etc. were added to paclitaxel structure for the formation of prodrugs with increased aqueous solubility. However, the approach of formation of these prodrugs was not rational as it is very tedious and in many undesirable compounds was also synthesized.

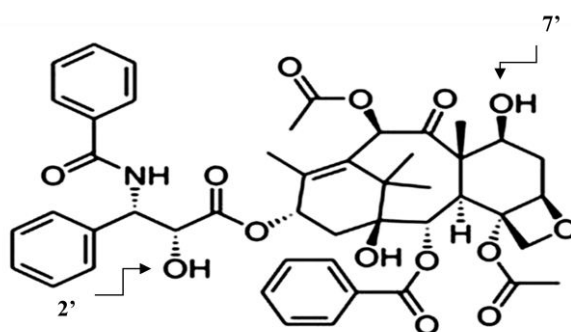


Figure 5.1: Structure of paclitaxel. 2' and 7' hydroxyl groups are the potential sites for adding hydrophilic groups to make prodrugs with better solubility.

Theoretical models such as Quantitative structure property-relationship (QSPR) models and data mining methods are efficient techniques that correlate molecular characteristics with physicochemical properties of molecules. However, the availability of physicochemical parameters for all molecules is difficult. Molecular descriptors determined from these structures were explored for the development of QSPR models. A QSPR model can provide a set of predictors for any molecular property using the structure of the molecule. Thus availability of solubility data for prodrugs of paclitaxel allows lead to develop the QSPR models for the solubility prediction of paclitaxel prodrugs and thus guide in the rational formation of more potent prodrugs of paclitaxel. In the study, a QSPR strategy has been synthesised for the solubility prediction using a small dataset implementing computational techniques that would be cost- and time- effective to screen a vast number of structural parameters and identification of structural descriptors determining the solubility of prodrugs. After establishing the fact, that poor solubility is a major bottleneck for this highly active drug and its prodrugs (derivatives), the current study emphasizes the formation of a robust QSPR model using small solubility data utilizing 22 paclitaxel prodrugs dataset. The data set used for QSPR model development in this study consists of previously reported 22 paclitaxel prodrug with different substitution on the 2'-OH and 7'-OH of paclitaxel structure. Each of these compounds had associated experimentally determined solubility. For the formation of effective QSPR strategy the geometry optimization, descriptor extraction, descriptor selection, development of QSPR model, and validation was performed in stepwise and validated manner thus establishing a strategy for the formation of a successful QSPR model for solubility prediction. Geometry optimization of prodrugs was done implementing two different methods such as PM6 and AM1 were implemented to determine the optimized molecular geometries of prodrugs using Gaussian 09 quantum chemistry software. Global energy minima of each structure were verified by analyzing vibrational frequencies. These minimized structures were eventually employed for the extraction of Dragon 7 descriptors; thus obtaining 5250 descriptors for each molecule. The descriptors were classified in three descriptor classes; 1D, 2D, 3D. There are various descriptor groups such as 2D autocorrelation, CATS_3D, GETAWAY, WHIM, 3D-Morse etc. Each descriptor group is divided into eight independent subgroups on the basis of co-relation and covariance. Then on each subgroup, AIC and VIF multi-colinearity indicators were applied for the selection of independent descriptors (p-value < 0.05). For the model formation from the selected descriptors “*Non-linear multiple regression analysis*” was performed and the models were cross validated and to detect the chance correlation yRandomization was performed.

Some descriptor values are positive and some are negative that indicated that the extracted molecular descriptors with positive values vary linearly with $\log(ely)$ (logarithm of solubility) whereas others (negative descriptors values) vary exponentially with $\log(ely)$. Therefore, the following regression equation (Eqn. 1) was proposed.

$$\log(ely) = \log\beta_0 + \beta_p \log X_p + \beta_n X_n \quad \dots \text{Eqn 1}$$

where $\log\beta_0$ is a constant, β_p is the coefficient of positively valued descriptors, $\log X_p$ is the logarithm of descriptor value, β_n is the coefficient of negatively valued descriptor, and X_n is the negative descriptor value.

Rational QSPR models with equitable regression and correlation coefficients were determined for AM1 and PM6 geometry optimized datasets with R^2 & Q^2 values 0.87 & 0.86 and 0.86 & 0.83 respectively. Moreover yRandomization results indicated the statistical significance of these models. The QSPR models related to solubility are based on different methodologies with diverse datasets. Results of the current study are comparable to earlier reported QSPR solubility models. As compared to other methods the performance of formed QSPR models is optimal with 4 and 5 descriptors. This objective concludes in the formation of an effective pipeline for the descriptor selection and formation of QSPR models with high statistical standards.

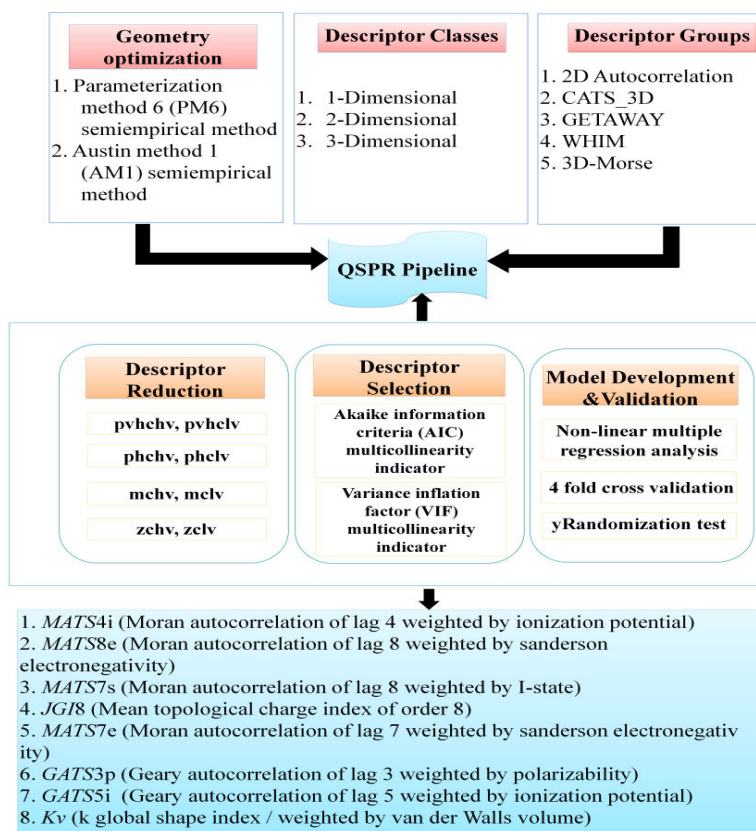


Figure 5.2: QSPR pipeline (QSPR-sPL)

Being poorly soluble, the bioavailability for paclitaxel remains very low. Hence opportunities must be explored better for the formation of more potent prodrugs could be formed. This led us to develop QSPRs which guided us *in-silico* in designing more potent prodrugs. Also our curiosity was raised to look into the precise hydrophilic groups that were incorporated in the structure for the formation of prodrugs. Understanding of molecular structure of paclitaxel revealed some important sites that would be important to improve the solubility of prodrugs and thus towards bioavailability.

Paclitaxel chemical structure includes a bulky, composed, and fused ring along with a number of hydrophobic substituent which makes its chemical nature highly lipophilic. Several prodrugs of paclitaxel were synthesized by the addition of hydrophilic substituent groups such as phosphates, carboxylic acids, sugar derivatives etc. Substituent groups that are contributing significantly towards solubility are difficult to interpret, while the descriptors derived from the substituent's molecular structures and subsequently the formed QSPR model will be easier to interpret. The metabolism of the substituent groups would be more meaningful in the modelling of new prodrugs. The kinetic description of the substituent groups would depend on the liver metabolism involving CYPs. CYPs form the extensive heme-thiolate protein family consisting of sixty different human isoforms that metabolizes ninety percent of the xenobiotics.

The best way to understand the precise interaction pattern of substituent groups with receptor is molecular docking and molecular dynamics simulation (MDS) study. Thus a docking and MDS study based model on the substituent group dataset, depicting the hydrophilic groups that are more favourable for the formation of paclitaxel prodrugs was performed. All the different isoforms of CYPs were subjected to docking with the substituent group dataset but the most favourable results were obtained from the cytochrome 1A2 (CYP1A2). In the objective, QSPR modelling, docking, and MDS based model is developed that provide a better view of the substituent groups those would be added to the paclitaxel molecule for the formation of better soluble prodrugs and also further get metabolised in the human body.

Descriptor selection was done using the QSPR pipeline (QSPR-sPL) described in the first objective. Development of consistent and optimal QSPR models was done by formation of QSPR model by combining calculated descriptors from each group those were showing reasonable R^2 and Q^2 . The descriptors were selected and combined (R^2 & Q^2) till the highest accuracy model is obtained for both the PM6 and AM1 geometry optimized datasets. yRandomization for the detection of chance correlation is done at each step using DTC_yrandomization program. Eight new QSPR models based on the fundamentals of

aqueous solubility and optimization of descriptor space from the paclitaxel prodrugs dataset and also from the substructure dataset which were optimized at both AM1- and PM6 geometry levels with good statistics were obtained.

Table 5.1: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models for paclitaxel prodrugs and substructures with PM6 and AM1 optimized geometry dataset

S. No.	Descriptor Group	Descriptor selection algorithm	AM1 optimized geometry dataset			PM6 optimized geometry dataset				
			No. of descriptors	$R^2 \setminus R^2_{\text{yrand}}$	$Q^2 \setminus Q^2_{\text{yrand}}$	RMSE	No. of descriptors	$R^2 \setminus R^2_{\text{yrand}}$	$Q^2 \setminus Q^2_{\text{yrand}}$	RMSE
1	Paclitaxel prodrugs	AIC & VIF Function	21 (12)	0.78\0.13	0.60\0.31	1.55	16 (10)	0.80\0.14	0.69\0.20	1.46
2	Substituent groups	AIC & VIF Function	8 (8)	0.82\0.16	0.76\0.21	1.33	9 (9)	0.93\0.17	0.83\0.26	0.55
3	Paclitaxel prodrugs	MATLAB 'Stepwise fit'	9 (9)	0.85\0.10	0.79\0.18	1.02	16 (7)	0.82\0.08	0.79\0.14	1.40
4	Substituent groups	MATLAB 'Stepwise fit'	7 (6)	0.83\0.10	0.67\0.30	1.2	16 (13)	0.88\0.26	0.83\0.22	1.11

The two methods mentioned above could provide consistently better results based on the multi-way comparison and are hereby suggesting the quasi-mixture descriptors also. It is based on the theory of 11 basic tetraatomic fragments with connectivity and composition for the substituent group dataset. For all the substituent molecules, the quasi-mixture simplex representation of molecular structure (SiRMS) descriptors was calculated. The results obtained from the 'quasi-mixture' descriptors for the specific validation of substituent group dataset also validated the effectiveness of the proposed QSPR models. The study also incorporates the metabolic study of substituent group dataset in the human body. The favourable results obtained by various structural parameters like RMSD, RMSF, Rg, H-bond, and Principal Component Analysis (PCA) were evaluated for the stability studies of CYP1A2 enzyme and selected substituent groups. This approach will be useful for the medicinal chemists to improve the solubility profile of the future prodrugs by suggesting the structural modifications in the molecule.

The prodrug approach has been very successful in last decade for enhancing the development of new molecules to enhance oral bioavailability of drugs whose assimilation is limited due to low aqueous solubility and permeability. Prodrugs with increasing solubility include common

hydrophilic groups such as phosphates, carbonates and silicates etc. Phosphate prodrugs show prominent increase in solubility in paclitaxel. Also, the phosphate prodrugs have low pKa (1-2).

In the last objective, interaction of CYP3A4 with paclitaxel and its phosphate prodrugs is studied *in-silico* to characterize the bioavailability of the prodrugs for its oral usage. In this study, the interaction of paclitaxel, its phosphate prodrugs, and a highly soluble prodrug of paclitaxel, isotaxel that is 5000 times more soluble as compared to paclitaxel with CYP3A4 is studied *in-silico* and characterizes the bioavailability of the phosphate prodrugs to the intestinal cancerous cells. CYP3A4 plays a major role in drug absorption and it acts as a barrier to the prodrugs to reach the intestinal cancerous cells. For the drugs to be more bioavailable to the intestinal cancerous cells they should not get metabolised in the liver by the CYP3A4 and thus reaches to the intestine where they would exhibit their anti-neoplastic effects and finally led to cell-death.

ADMET properties of a set of twenty paclitaxel phosphate prodrugs were calculated using the admetSAR web server. After the initial screening considering the different pharmacokinetic properties, molecular docking of the selected ten phosphate prodrugs those were showing more favourable pharmacokinetic properties was carried out. Molecular dynamics study (MDS) of the selected three prodrugs those were showing moderate interaction with the CYP3A4 was performed at 1 microsecond. Various criterions like RMSD, RMSF, SASA, PCA, and free energy landscape were evaluated to study the interaction pattern of these prodrugs with the CYP3A4. Also, MM-PBSA binding free energy is calculated and results showed the average binding of three phosphate prodrugs with CYP3A4. The unstable complexes of prodrugs with CYP3A4 as revealed by the MDS study thus explained the more bioavailability of these paclitaxel phosphate prodrugs to the intestinal cancerous cells and thus can reach the site of action and subsequently lead to apoptosis. The computational calculation of ADMET properties, interaction, and instability studies of paclitaxel phosphate prodrugs suggested the potential candidate prodrugs that could be delivered orally and are further suggested for effective use against gastrointestinal cancer (GIT).

5.2 Future prospects:

The QSPR strategy for the solubility prediction can be further sophisticated using the more exhaustive double cross-validation techniques. Also, the QSPRs formed for the solubility prediction of paclitaxel prodrugs could be more refined by using proteochemometric

techniques. The phosphate prodrugs those were studied *in-silico* for the bioavailability analysis can be further experimentally validated. The data generated in this research work presents a compelling evidence for further preclinical and clinical evaluation. Thus the future holds promises for the better and more efficient prodrugs with not only increased solubility but also for the oral bioavailability of paclitaxel for more specific cancer types. It is anticipated that the proposed models and generated computational information will be of utmost use to the scientific community for its experimental validations and further dissemination.

APPENDIX A

A.1 Prodrug molecules of Paclitaxel incorporated for quantitative structure-property relationship (QSPR) model development.

Various sites of Paclitaxel molecule which facilitate the synthesis of different prodrugs with better solubility (Figure A.1).

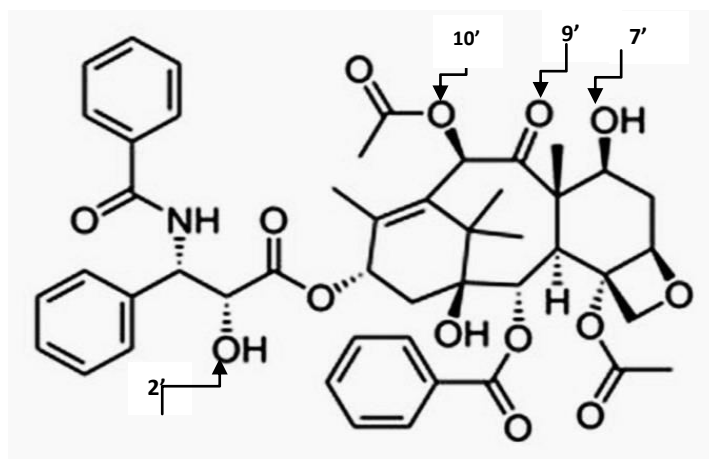
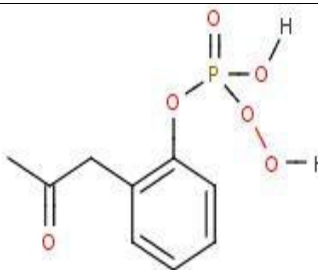
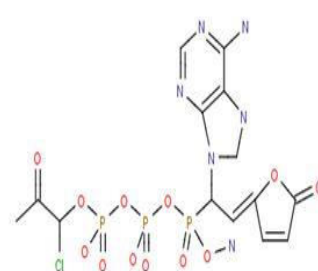
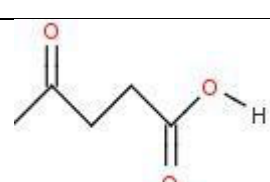
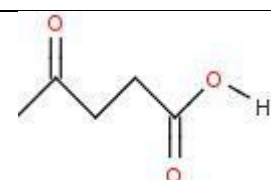
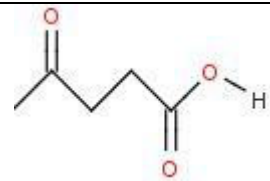
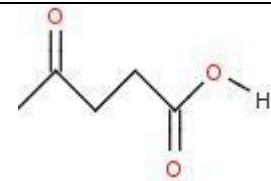
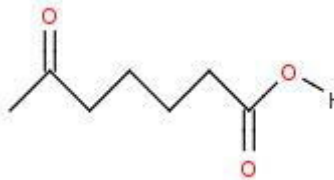
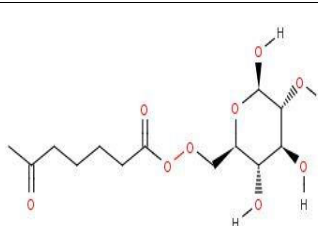


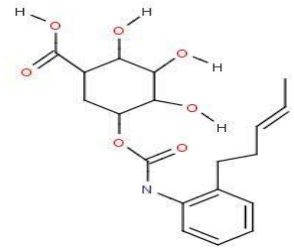
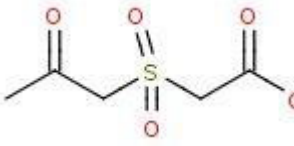
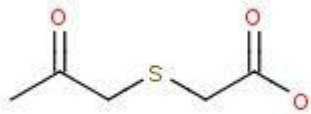
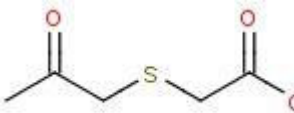
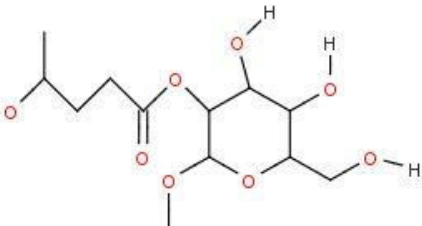
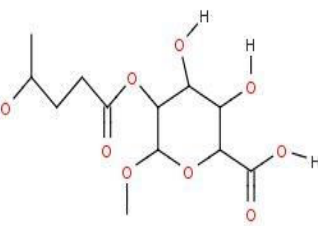
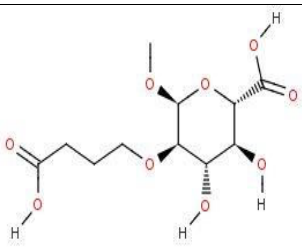
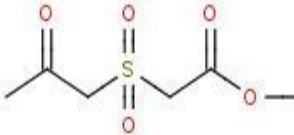
Figure A.1: Structure of Paclitaxel with 2', 7', 9' and 10' sites which assist to make prodrugs with better solubility.

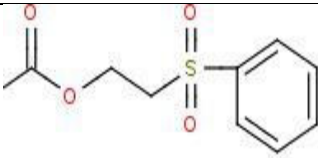
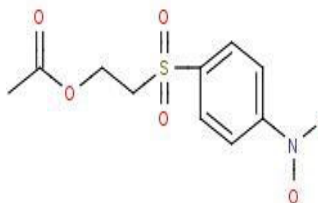
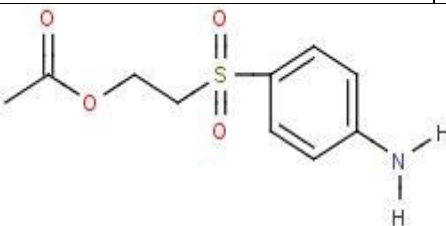
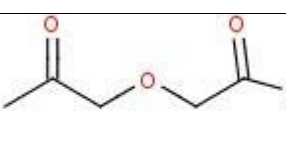
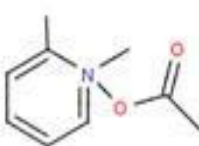
Twenty two Paclitaxel prodrugs having either better or comparable solubility to Paclitaxel were collected from literature. The details about the prodrugs and their solubility are provided in table A.1.

Table A.1: Prodrugs of Paclitaxel included for QSPR model development

S. No	Compound Name	Solubility (g/ml)	Substituent	
			2'	7'
1.	2'-Ethylcarbonate Paclitaxel	0.00025		H
2.	Ethylene glycol Paclitaxel	1.2		H

3.	Phosphate derivative of 2' ethoxy cabonyl Paclitaxel	5		H
4.	Phosphonic acid diphosphate derivative of adenine containing butenolide Paclitaxel	0.042		H
5.	2' Maly Paclitaxel	0.2		H
6.	7' Maly Paclitaxel	0.5	H	
7.	2'-7' Maly Paclitaxel	0.3		
8.	Paclitaxel-2'-adipic acid	0.40625		H
9.	Paclitaxel-2'-adipoyl glucose	0.0145		H

10.	β - Glucouronylca rbamate prodrug of Paclitaxel	0.5		H
11.	Protaxols	0.35		H
12.	Protaxols	0.46	H	
13.	Protaxols	0.84		H
14.	Ethoxy carbonyl derivative of Paclitaxel	0.000053		H
15.	Ethoxy carbonyl derivative of Paclitaxel	0.000104		H
16.	2'-O-Succinyl Paclitaxel	0.000137		H
17.	Protaxols	1.2		H

18.	Protaxols	0.1		H
19.	Protaxols	0.1		H
20.	Protaxols	0.1		H
21.	Protaxols	0.5		H
22.	Taxol-2'-methylpyridinium acetate	1.504		H

A.2 Number of Descriptors in each group provided by Dragon7 Software

The descriptors extracted from Dragon7 software form thirty groups (Table A.2). These descriptors are used for QSPR models development by using the statistical methods.

Table A.2: Descriptor groups extracted from Dragon 7

S. No.	Descriptor Group	Class	No. of Descriptors
1	Constitutional Indices	1 D	47
2	Ring descriptors	1 D	32
3	Topological Indices	2 D	75
4	Walk and path counts	2 D	46
5	Connectivity Indices	2 D	37
6	Information Indices	1 D	50
7	2D matrix-based descriptors	2 D	607
8	2D Autocorrelations	2 D	213

9	Burden eigen values	2 D	96
10	P_VSA-like descriptors	2 D	55
11	ETA indices	2 D	23
12	Edge adjacency indices	2 D	324
13	Geometrical descriptors	3 D	38
14	3D matrix- based descriptors	3 D	99
15	3D autocorrelations	3 D	80
16	RDF descriptors	3 D	210
17	3D- MoRSE descriptors	3 D	224
18	WHIM descriptors	3 D	114
19	GETAWAY descriptors	3 D	273
20	Randic molecular profiles	3 D	41
21	Functional group counts	3 D	154
22	Atom-centred fragments	3 D	115
23	Atom-type E-state indices	2 D	172
24	CATS 2D	2 D	150
25	2D Atom Pairs	2 D	1596
26	3D Atom Pairs	3 D	36
27	Charge descriptors	3 D	5
28	Molecular properties	1 D	20
29	Drug-like indices	1 D	28
30	CATS 3D	3 D	300

A.3 QSPR equations determined using individual group of descriptors provided in Table A.1.

A.3.1 AM1 optimized geometry dataset.

2D Autocorrelation descriptor group:

$$ly = -90.1437 - 4.081 * ATSC5e - 48.691 * MATS7e - 69.2967 * MATS8e + 42.1957 * MATS7s + 41.4876 * GATS5i + 3.6659 * GGI4 + 3074.5835 * JGI8 \quad \text{..Eqn. A1}$$

CATS_3D descriptor group:

$$ly = -0.1672 \pm 0.4725 * CATS3D_{15_AA} \quad \text{..Eqn. A2}$$

GETAWAY descriptor group:

$$ly = 3.7241 + 8.4763 * H3p + 0.8764 * H5s + 17.7689 * R3m \quad \text{..Eqn. A3}$$

WHIM descriptor group:

$$ly = 5.3801 - 16.3759 * kv \quad \text{..Eqn. A4}$$

A.3.2 PM6 optimized geometry dataset

2D Autocorrelation descriptor group

$$ly = 69.6424 - 215.7847 * MATS5i - 58.5471 * GATS3p \quad \text{..Eqn. A5}$$

CATS_3D descriptor group:

$$ly = -0.6135 - 3.1315 * CATS_3D_14_DD - 1.8206 * CATS3D_19_DA - 3.6055 * CATS3D_18_DN \quad \text{..Eqn.A6}$$

GETAWAY descriptor group:

$$ly = 39.0765 - 17.2143 * H3v + 13.4728 * H4v - 13.8719 * R4e \quad \text{..Eqn. A7}$$

WHIM descriptor group:

$$ly = -78.1809 + 518.9924 * Gu - 20.9545 * kv \quad \text{..Eqn. A8}$$

A.4 QSPR models developed using the complete pool 5250 descriptors

As the QSPR model's performance for individual descriptor groups was moderate, the QSPR model for Paclitaxel prodrugs was determined by applying the protocol (Figure 2) on the complete pool of 5250 descriptors which belong to three classes (1D, 2D and 3D). Further, each class of descriptors was divided into eight subgroups (pvhchv, pvhclv, phchv, phclv, pmchv, pmclv, zchv and zclv) on the basis of correlation and variance among the descriptors. Then, selected descriptors from same subgroup belong to a given class were combined. AIC and VIF multicollinearity indicators were applied on each eight subgroups of descriptors. Then only independent descriptors in each subgroup were combined which were further selected to give the final QSPR model (Table A.3). Low values of Q^2 (0.66 & 0.40) indicate that QSPR model is inconsistent as compared to results of some individual descriptor groups (Table I). The models given in the following table are statistically not significant, and appear to be over-fitted as difference between R^2 and Q^2 is quite high (Table A.3, Eqn. A9 & A10).

Table A.3: QSPR models developed using total 5250 descriptors

S. No.	Optimized Geometry	Descriptors	No. of descriptors	* R^2	* Q^2	*RMSE
1	AM1 optimised geometry dataset	Dragon 7 descriptors (5250)	15 (8)	0.93	0.40	6.0863
2	PM6 optimised geometry dataset	Dragon 7 descriptors (5250)	11 (6)	0.90	0.66	2.50

*Regression (R^2), 4-fold cross validation correlation (Q^2) coefficients and root mean square error (RMSE)

QSPR equations

$$ly = -27.0291 + 5.941 * H5e + 12.8101 * nTA - 26.0675 * MATS8e + 143.4973 * SpMin4_Bh(p) - 29.3487 * SpMax6_Bh(m) - 10.5846 * F05(O - O)x1152 - 57.4799 * WiA_G - 1.6147 * RDF150p \quad \text{..Eqn. S9}$$

$$ly = 110.9287 - 6.7539 * CATS2D_{05AA} + 46.4443 * IDDE + 3.8496 * MATS5i + 2.5137 * N\% + 325.235 * GATS5i - 97.5698 * GATS6p \quad \text{..Eqn. S10}$$

A.5 QSPR models developed by combining individual descriptor groups provided in Table A.1

The QSPR model development for each thirty groups was performed, but only four descriptor groups (2D Autocorrelation, CATS_3D, WHIM and GETAWAY) provided significant correlation with solubility for both PM6 and AM1 optimized geometry structures. We combined the significant descriptors from each group and then the QSPR models were formed for both AM1 and PM6 optimized geometry datasets in WEKA 3.6.11 using classifier linear regression module. The correlation and regression coefficients obtained for AM1 and PM6 optimized geometry datasets are provided in Table A.4 (Eqn S11& S12). The models given in the following table are statistically not significant, and appear to be over-fitted as difference between R^2 and Q^2 is quite high.

Table S4: Regression (R^2) and 4-fold cross validation correlation (Q^2) coefficients

S. No.	Optimized Geometry	Descriptor Group	No. of descriptors	R^2	Q^2	RMSE
1.	AM1 optimised geometry dataset	2D Autocorrelation + CATS_3D + GETAWAY + WHIM	12	0.90	0.60	5.42
2.	PM6 optimised geometry dataset	2D Autocorrelation + CATS_3D + GETAWAY + WHIM	10	0.91	0.46	3.32

QSPR equations

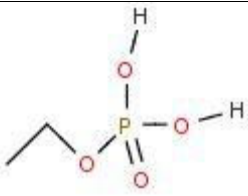
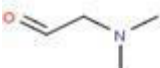
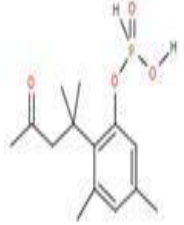
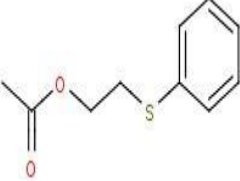
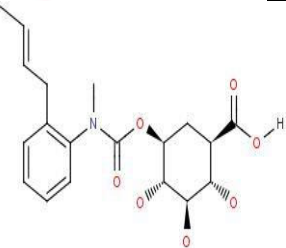
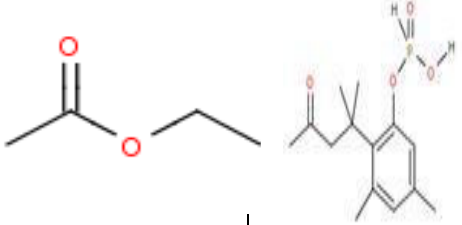
$$ly = -90.1437 - 4.081 * ATSC5e - 48.691 * MATS7e - 69.2967 * MATS8e + 42.1957 * MATS7s + 41.4876 * GATS5i + 3.6659 * GGI4 + 3074.5835 * JGI8 \quad \text{..Eqn. A11}$$

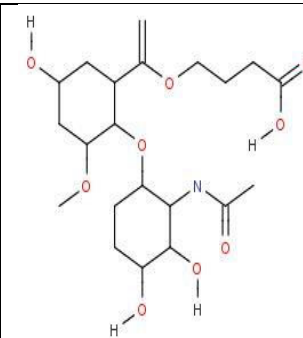
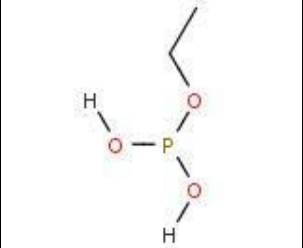
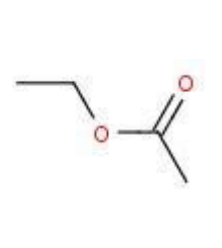
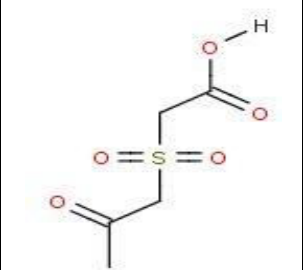
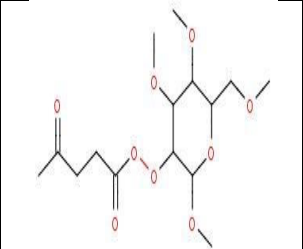
$$ly = 3.5362 - 158.6092 * MATS5i - 35.7838 * GATS3p + 269.8457 * Gu - 11.5829 * kv \quad \text{..Eqn. A12}$$

A.6 QSPR models developed for thirty four molecules

In addition to twenty two Paclitaxel prodrugs (Table A.1), twelve more similar prodrugs (Table A.5) were added to evaluate the performance of proposed protocol with more data points. Details about the substituent of prodrugs and their solubility data are provided below.

Table A.5: Additional 12 prodrugs of Paclitaxel included for evaluating the performance of QSPR model

S. No.	Compound Name	Solubility (g/ml)	Substituent	
			2'	7'
23	Phosphonoxy methyl ether derivative of Paclitaxel	0.25		H
24	2'-(N,N dimethylglycyl) taxol	2		H
25	9-dihydro taxol*	0.01875	-	-
26	10-acetyl 9-dihydro taxol*	0.0005	-	-
27	Paclitaxel Phosphates	0.5	H	
28	Protaxols	1.2		H
29	Carbamate linkage of β -glucuronic acid to Paclitaxel	0.25		H
30	Paclitaxel Phosphates	0.5		

31	Protaxols	0.01		H
32	Paclitaxel Phosphates	0.5		
33	Protaxols	1.2		H
34	Glucose conjugated Paclitaxel	0.35		H

* Hydrogen substitution at 9' position instead of acetyl group as in paclitaxel prodrug no. 25, and H substitution at 10' position instead of double bond with oxygen in prodrug no. 26.

A.6.1 QSPR models development and validation for the above mentioned thirty four molecules datasets

We had added twelve more (Table A.5) similar compounds (data points) to the model and had divided total thirty four data points into a set of twelve and twenty two as training and test sets, respectively. Descriptor selection for each thirty groups was performed following the same protocol (Figure A.2) and R^2 & Q^2 values were obtained. Descriptors providing successful models were combined and then selected to enhance QSPR models for training dataset. QSPR model developed from training dataset was evaluated on the test dataset. Ten yRandomization runs were also performed to find out significance of the descriptors (Section 1.4). The R^2 & Q^2 and R^2_{yrand} & Q^2_{yrand} values and Q^2 values for the test set were obtained (Table A.6).

Table A6: Regression (R^2), validation & cross correlation (Q^2) coefficients of the developed QSPR model

S. No.	Optimized Geometry	Descriptor Group	No. of descriptors	$R^2/\text{}^sR^2_{\text{yrand}}$	$Q^2/\text{}^sQ^2_{\text{yrand}}$	RMSE	cRp ²	Q^2 (Test set)
1	AM1	Constitutional + Burden eigen values + GETAWAY	5 (1)	0.71/0.03	0.67/-0.37	2.06	0.50	0.60
2	PM6	Geometrical +GETAWAY	3 (2)	0.89/0.12	0.80/-0.74	1.32	0.75	0.80

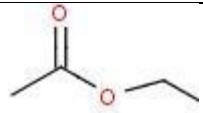
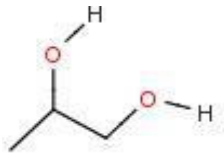
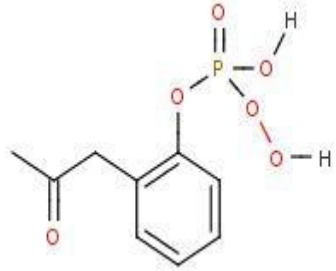
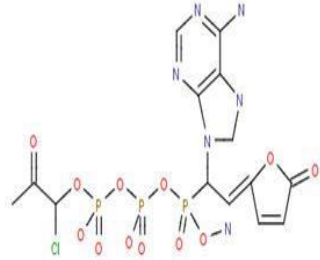
$\text{}^sR^2_{\text{yrand}}$ & $\text{}^sQ^2_{\text{yrand}}$ values are calculated to determine statistical significance of the model (Section 1.4).

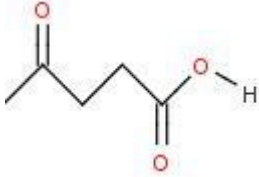
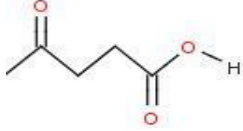
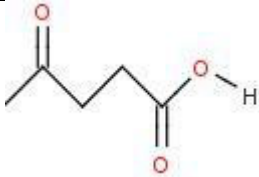
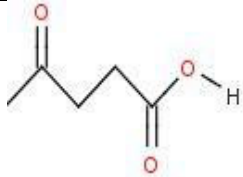
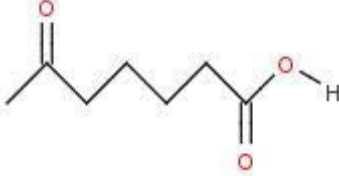
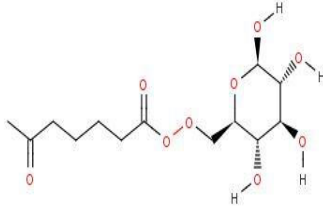
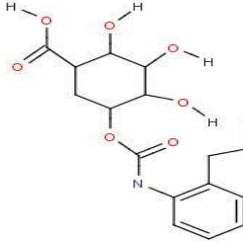
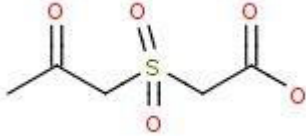
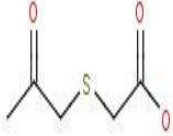
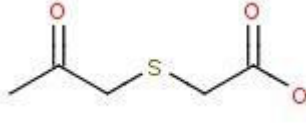
APPENDIX B

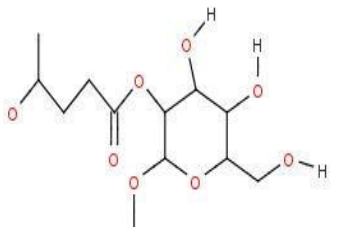
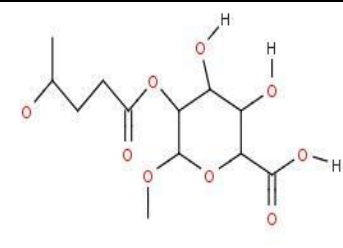
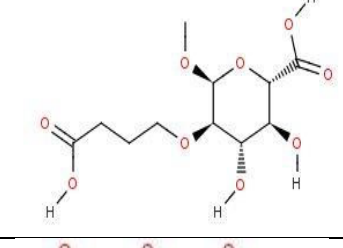
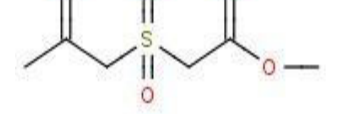
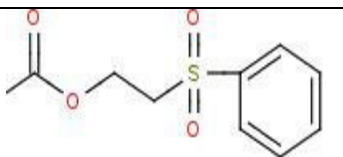
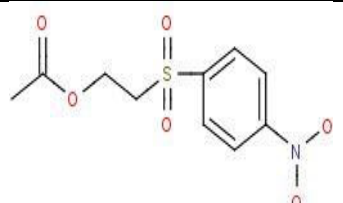
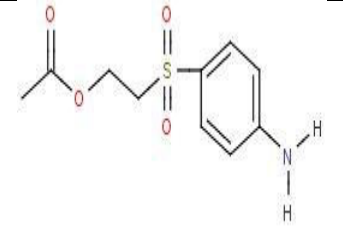
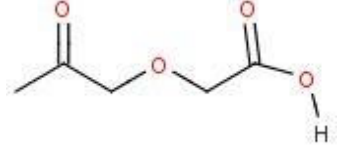
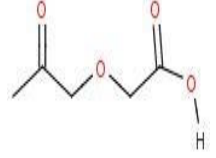
B.1 Prodrug molecules of Paclitaxel incorporated for quantitative structure-property relationship (QSPR) model development.

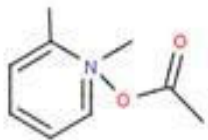
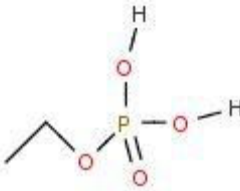
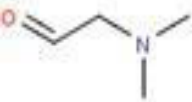
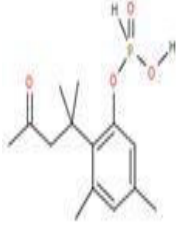
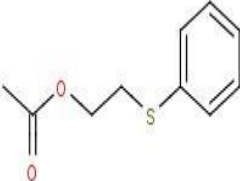
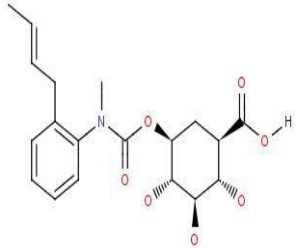
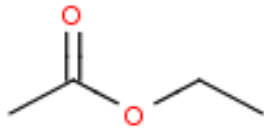
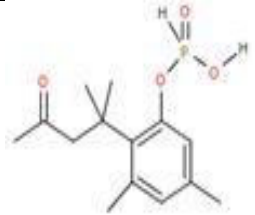
Eighty Paclitaxel prodrugs having either better or comparable solubility to Paclitaxel were collected from literature. The details about the prodrugs and their solubility are provided in Table B.1.

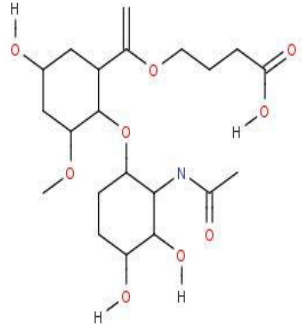
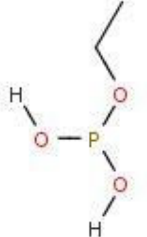
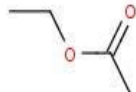
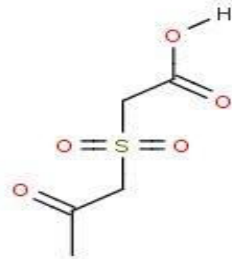
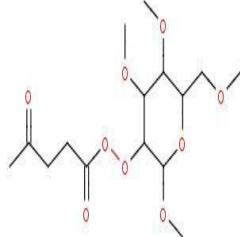
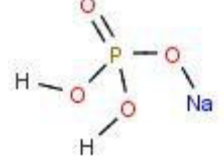
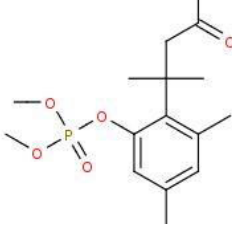
Table B.1: Prodrugs of Paclitaxel included for QSPR model development

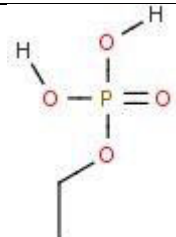
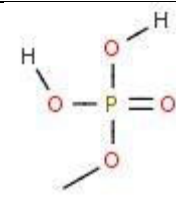
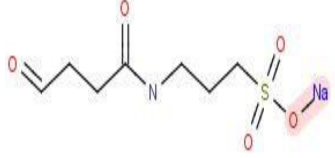
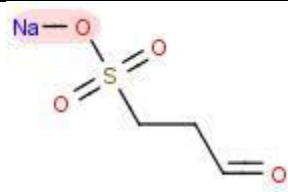
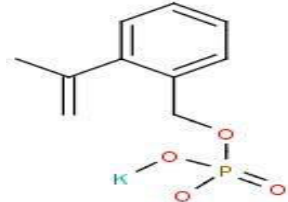
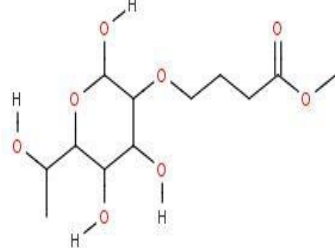
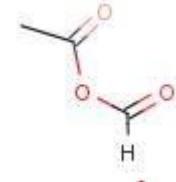
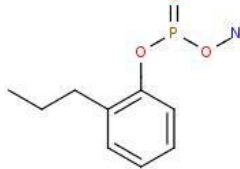
S. No.	Compound Name	Solubility (g/ml)	Substituent	
			2'	7'
1	2'-Ethylcarbonate Paclitaxel	0.00025		H
2	Ethylene glycol Paclitaxel	1.2		H
3	Phosphate derivative of 2' ethoxy cabonyl Paclitaxel	5		H
4	Phosphonic acid diphosphate derivative of adenine containing butenolide Paclitaxel	0.042		H

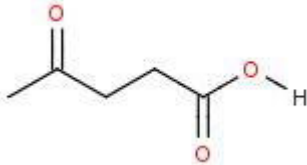
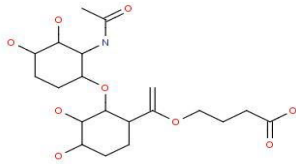
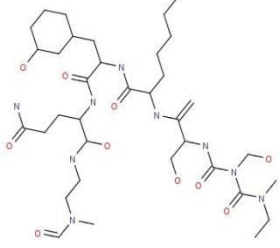
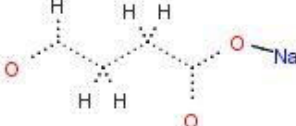
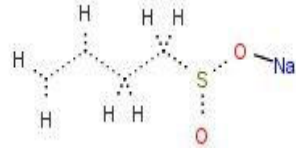
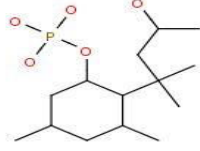
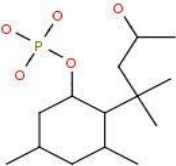
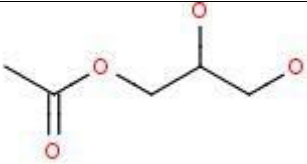
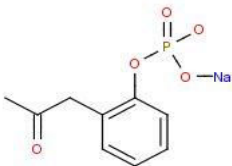
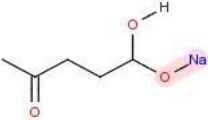
5	2' Malyl Paclitaxel	0.2		H
6	7' Malyl Paclitaxel	0.5	H	
7	2'-7' Malyl Paclitaxel	0.3		
8	Paclitaxel-2'-adipic acid	0.40625		H
9	Paclitaxel-2'-adipoyl glucose	0.0145		H
10	β -Glucouronylcarbamate prodrug of Paclitaxel	0.5		H
11	Protaxols	0.35		H
12	Protaxols	0.46	H	
13	Protaxols	0.84		H

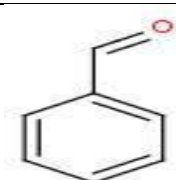
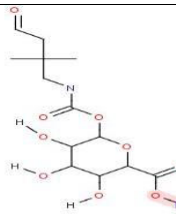
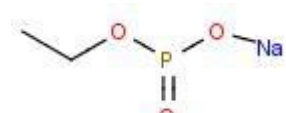
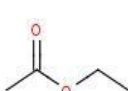
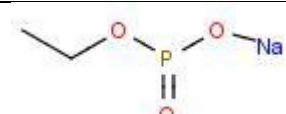
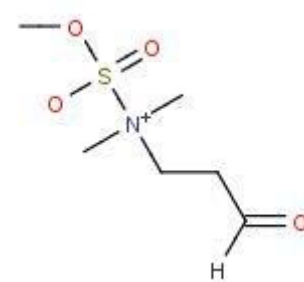
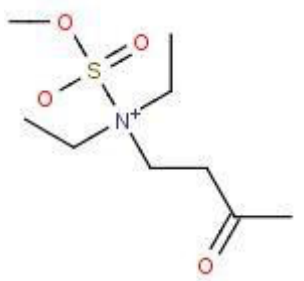
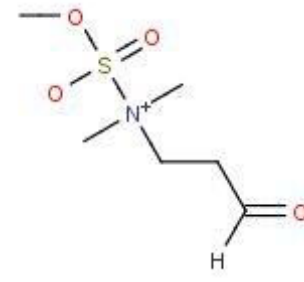
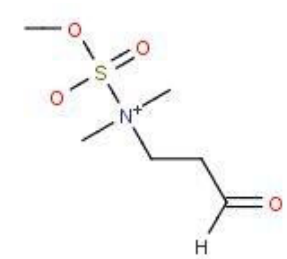
14	Ethoxy carbonyl derivative of Paclitaxel	0.000053		H
15	Ethoxy carbonyl derivative of Paclitaxel	0.000104		H
16	2'-O-Succinyl Paclitaxel	0.000137		H
17	Protaxols	1.2		H
18	Protaxols	0.1		H
19	Protaxols	0.1		H
20	Protaxols	0.1		H
21	Protaxols	0.5		

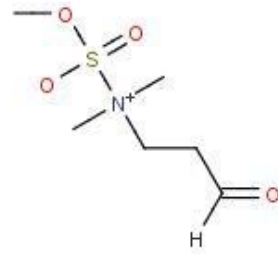
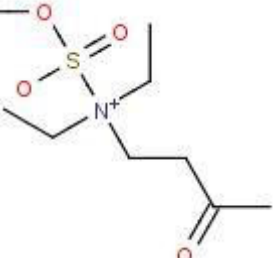
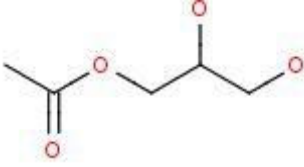
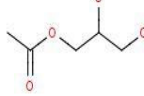
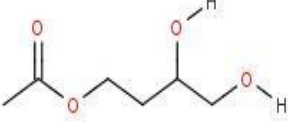
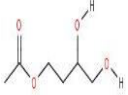
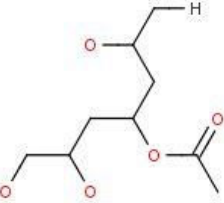
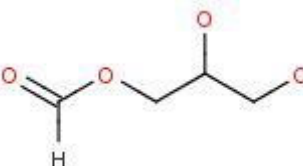
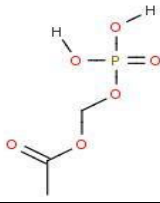
22	Taxol-2'-methylpyridinium acetate	1.504		H
23	Phosphonoxy methyl ether derivative of Paclitaxel	0.25		H
24	2'-(N,N dimethylglycyl) taxol	2		H
25	9-dihydro taxol*	0.01875	-	-
26	10-acetyl 9-dihydro taxol*	0.0005	-	-
27	Paclitaxel Phosphates	0.5	H	
28	Protaxols	1.2		H
29	Carbamate linkage of β-glucuronic acid to Paclitaxel	0.25		H
30	Paclitaxel Phosphates	0.5		

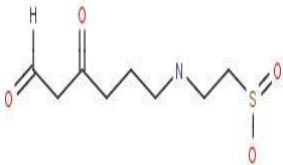
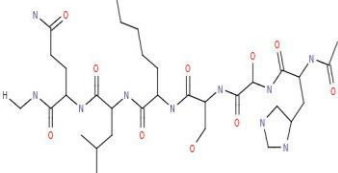
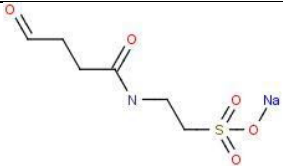
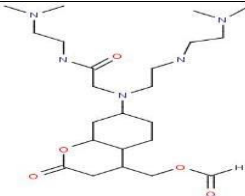
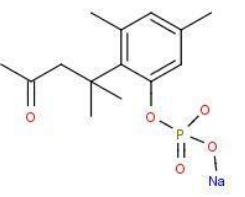
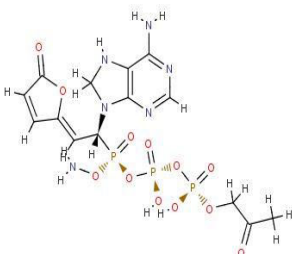
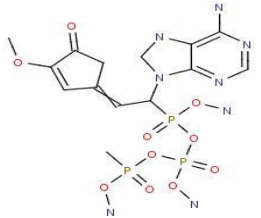

31	Protaxols	0.01	 <p>The structure shows two substituted cyclohexane rings. The top ring has a hydroxyl group, a methoxy group, and a side chain containing an ester and a hydroxyl group. The bottom ring has a hydroxyl group, a methoxy group, and a side chain containing a hydroxyl group and an amide group.</p>	H
32	Paclitaxel Phosphates	0.5	 <p>The structure shows a central phosphorus atom bonded to two hydroxyl groups and two ethoxy groups.</p>	 <p>The structure shows a phosphate group bonded to an ethoxy group and a carbonyl group.</p>
33	Protaxols	1.2	 <p>The structure shows a central sulfur atom double-bonded to two oxygen atoms and single-bonded to two carbon chains, each ending in a carboxylic acid group.</p>	H
34	Glucose conjugated Paclitaxel	0.35	 <p>The structure shows a glucose molecule with a paclitaxel moiety attached to one of its hydroxyl groups.</p>	H
35	Paclitaxel Phosphates	10	 <p>The structure shows a central phosphorus atom bonded to two hydroxyl groups, one sodium atom, and one oxygen atom.</p>	H
36	Paclitaxel Phosphates	5	 <p>The structure shows a benzene ring with a phosphate group and a side chain containing a carbonyl group.</p>	H

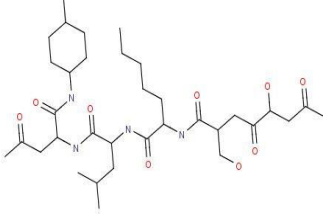
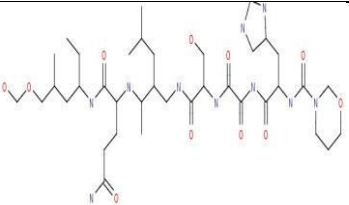
37	Phosphonoxy methyl ether derivative of Paclitaxel	0.25		H
38	Phosphonoxy methyl ether derivative of Paclitaxel	0.25		H
39	Bisulphite of 2' acrylotaxol	0.029		H
40	Bisulphite of 2' acrylotaxol	0.0525		H
41	Phosphates of Paclitaxel	3.5		H
42	Glucuronic acid conjugated Paclitaxel	0.08		H
43	Carbonates of Paclitaxel	0.051		H
44	Phosphate of Paclitaxel	1		H

45	Succinyl Paclitaxel	0.025		H
46	Heteroaromatic taxane	0.01		H
47	Carboxyl chain Paclitaxel	70		H
48	Carbonate of Paclitaxel	0.50		H
49	Sulfonated acryol taxol	0.0525		H
50	Methyl Phosphate of Paclitaxel	10		H
51	Methyl Phosphate of Paclitaxel	10	H	
52	Phospahte of Paclitaxel	5		H
53	Phospahte of Paclitaxel	5		H
54	Sodium salt of malyl Paclitaxel	0.60		H

55	Isotaxel	0.45		H
56	Paclitaxel with carbamate linkage to β -glucuronic acid	0.25		H
57	Phosphate of Paclitaxel	0.5	H	
58	Phosphate of Paclitaxel	3		
59	Sulphates of Paclitaxel	5		H
60	Sulphates of Paclitaxel	10		H
61	Sulphates of Paclitaxel	10		

62	Sulphates of Paclitaxel	2	H	
63	Sulphates of Paclitaxel	2	H	
64	Carbonates of Paclitaxel	0.01		H
65	Carbonates of Paclitaxel	0.05	H	
66	Carbonates of Paclitaxel	0.05		H
67	Carbonates of Paclitaxel	0.02	H	
68	Carbonates of Paclitaxel	0.02		H
69	Carbonates of Paclitaxel	0.0125		H
70	Phosphonoxy methyl carbonate	0.25		H

71	2'[[4-(3-sulfopropyl)-amino]-1,4 dioxobutyl]oxy taxol	0.295		H
72	Carboxyl chain Paclitaxel	70		H
73	Bisulphite of 2'acroyltaxol	0.047		H
74	Phototaxol	100		H
75	Phosphate of taxol	2.5		H
76	Phosphonic acid diphosphate derivative of adenine containing butenolide Paclitaxel	0.037		H
77	Paclitaxel 2'-triphosphono γ-(z) ethylidene-2,3 dimethoxybutenolide	0.019		H
78	Phototaxol	100		H

79	Carboxyl chain Paclitaxel	70		H
80	Carboxyl chain Paclitaxel	70		H

*Hydrogen substitution at 9' position instead of acetyl group as in paclitaxel prodrug no. 25, and H substitution at 10' position instead of double bond with oxygen in prodrug no. 26

B.2 Performance of the individual descriptor groups was assessed and only descriptors providing significant R^2 & Q^2 values (Table B.2-B.5) were combined and further selected for the development of consistent QSPR models.

Table B.2: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptors selected with AIC & VIF indicators for Paclitaxel prodrugs with AM1 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	Constitutional descriptors	1D	47 (3)	0.45	0.39
2	Ring descriptors	2D	72 (5)	0.52	0.41
3	Topological indices	2D	96 (2)	0.45	0.42
4	P_VSA-like descriptors	2D	55 (2)	0.60	0.59
5	Edge adjacency indices	2D	324 (1)	0.46	0.44
6	3D-MoRSE descriptors	3D	224 (2)	0.46	0.41
7	WHIM descriptors	3D	114 (2)	0.48	0.44
8	GETAWAY descriptors	3D	273 (5)	0.59	0.48
9	Functional group counts	3D	154 (2)	0.52	0.49

10	Atom-centered fragments	3D	115 (2)	0.47	0.39
11	Atom-type E-state indices	2D	172 (3)	0.62	0.60
12	CATS 2D	1D	150 (1)	0.45	0.43

Parenthesis indicates the no. of significant descriptors included in QSPR model from the respective group

Table B.3: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with AIC&VIF indicators for Paclitaxel prodrugs with PM6 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	Constitutional descriptors	1D	47 (2)	0.50	0.42
2	2D autocorrelations	2D	213 (8)	0.65	0.60
3	P_VSA-like descriptors	2D	55 (2)	0.65	0.63
4	GETAWAY descriptors	3D	273 (2)	0.53	0.46
5	Functional group counts	3D	154 (3)	0.65	0.63
6	Atom-centered fragments	3D	115 (2)	0.60	0.57
7	Atom-type E-state indices	2D	172 (2)	0.67	0.65
8	CATS 2D	1D	150 (3)	0.63	0.60

Parenthesis indicates the no. of significant descriptors included in QSPR model from the respective group.

Table B.4: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with AIC&VIF indicators for substructures with AM1 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	Information indices	1D	50 (3)	0.60	0.57
2	2D autocorrelations	2D	213 (3)	0.58	0.56
3	3D matrix-based	3D	99 (2)	0.57	0.56

	descriptors				
4	3D-MoRSE descriptors	3D	224 (3)	0.68	0.59
5	WHIM descriptors	3D	114 (1)	0.50	0.48
6	Functional group counts	3D	154 (2)	0.65	0.62
7	Atom-centered fragments	3D	115 (2)	0.65	0.63
8	Atom-type E-state indices	2D	172 (2)	0.61	0.60
9	2D Atom Pairs	2D	1596 (3)	0.65	0.60
10	CATS 3D	3D	300 (3)	0.60	0.53

Parenthesis indicates the no. of significant descriptors included in QSPR model from the respective group.

Table B.5: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with AIC&VIF indicators for substructures with PM6 optimized geometry dataset.

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	3D-MoRSE descriptors	3D	224 (7)	0.84	0.82
2	WHIM descriptors	3D	114 (2)	0.86	0.86
3	Functional group counts	3D	154 (3)	0.67	0.65
4	Atom-centered fragments	3D	115 (2)	0.65	0.62

Parenthesis indicates the no. of significant descriptors included in QSPR model from the respective group.

B.3 QSPR models developed from the descriptor selection done with Matlab ‘stepwise fit’ for both the datasets with the molecular geometries optimized at PM6 and AM1 methods.

The QSPR models formed using the selected descriptors with *Matlab ‘stepwise fit’* for the Paclitaxel prodrugs and the substructures added at the 2’ site of the prodrugs with both the PM6 and AM1 optimized geometry datasets were formed and validated in WEKA 3.6.11 using classifier linear regression module. The initial models were developed following the pipeline as described in Fig. 3 for all thirty descriptor groups individually for both datasets.

Performance of the individual descriptor groups is assessed and only those descriptors that are providing significant R^2 & Q^2 values (Table B.6-B.9) are combined and further selected for the development of consistent QSPR models.

Table B.6: Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with Matlab ‘stepwise fit’ for Paclitaxel prodrugs with AM1 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	2D autocorrelations	2D	213 (2)	0.95	0.93
2	Functional group counts	3D	154 (8)	0.83	0.69
3	Atom-centred fragments	3D	115 (11)	0.85	0.64

Parenthesis indicates the no. of significant descriptors included in QSPR model.

Table B.7 Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with Matlab ‘stepwise fit’ for Paclitaxel prodrugs with PM6 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	2D autocorrelations	2D	213 (5)	0.73	0.68
2	Atom-type E-state indices	2D	172 (3)	0.73	0.65
3	CATS 2D	1D	150 (6)	0.72	0.66
4	2D Atom Pairs	2D	1596 (5)	0.75	0.73
5	CATS 3D	3D	300 (9)	0.81	0.75

Parenthesis indicates the no. of significant descriptors included in QSPR model.

Table B.8 Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with Matlab ‘stepwise fit’ for substructures with AM1 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	Functional group counts	3D	154 (7)	0.87	0.83
2	CATS 2D	1D	150 (8)	0.86	0.81
3	2D Atom Pairs	2D	1596 (11)	0.90	0.79

Parenthesis indicates the no. of significant descriptors included in QSPR model.

Table B.9 Regression (R^2) and 10-fold cross validation correlation (Q^2) coefficients of the QSPR models developed using molecular descriptor selected with Matlab 'stepwise fit' for substructures with PM6 optimized geometry dataset

S No.	Group Name	1D\2D\3D	No. of Descriptors	R^2	Q^2
1	P_VSA-like descriptors	2D	55 (15)	0.62	0.64
2	Edge adjacency indices	2D	324 (13)	0.60	0.42
3	RDF descriptors	3D	211 (16)	0.53	0.51
4	Functional group counts	3D	154 (14)	0.83	0.71
5	Atom-centred fragments	3D	115 (16)	0.77	0.67
6	Atom-type E-state indices	2D	172 (16)	0.66	0.64

Parenthesis indicates the no. of significant descriptors included in QSPR model.

List of publications

1. **N. S. Munjal**, S. Dutta, M. Sharma, and C. Rout, “QSAR and QSPR model development and comparison for drugs having low solubility”, *IJETSR*, vol. 4 (12), 2017 (**CiteSeer^x : 2.1, GoogleScholar/DIIF/Q.Sensei**).
2. **N. S. Munjal**, M. Sharma, and T. R. Singh, “Development of QSPR strategy for the solubility prediction” *Current Computer Aided Drug Designing*, vol. 14(4), pp 302-309, 2018. (**IF: 1.200, SCI/Scopus Indexed**).
3. **N. S. Munjal**, R. Shukla, and T. R. Singh, “Chemometric approach to estimate kinetic properties of Paclitaxel prodrugs and their substructures for solubility prediction through molecular modelling and simulation studies” *Journal of Chemometrics* <http://dx.doi.org/10.1002/cem.3181>, 2019 (**IF: 1.847, SCI/Scopus Indexed**).
4. **N. S. Munjal**, R. Shukla, and T. R. Singh, “Molecular Docking and simulation studies of Paclitaxel prodrugs with Cytochrome 3A4 to correlate solubility and bioavailability through physicochemical characterization” (Under review in *Journal of biomolecular structure and dynamics* **IF: 3.2, SCI/Scopus Indexed**).

Conference paper

1. **N. S. Munjal**, N. Kumar, M. Sharma, and C. Rout, “QSAR Model development for the prediction of solubility,” *IEEE explore*; <http://dx.doi.org/10.1109/BSB.2016.7552139>, 2016.

Conference Abstracts

1. **N. S. Munjal**, N. Kumar, M. Sharma, and C. Rout, “QSAR Model development for the prediction of solubility,” *International Conference on Bioinformatics and Systems Biology (BSB)*, held at IIT Allahabad, March 4-6th, 2016. [Best oral presentation award in proteomics session].
2. **N. S. Munjal**, N. Kumar, M. Sharma, and C. Rout, “Computer aided model development for solubility prediction of Paclitaxel,” *R&D expo* held at Jaypee University of Information Technology, Wagnaghat, May 16th, 2016.
3. **N. S. Munjal**, M. Sharma, and C. Rout “Correlating substituent structure-solubility relationship to design prodrug with improved solubility profile,” *National Conference on Breaking Barriers through Bioinformatics and Computational Biology (BBB)*, held at IIT Delhi, July 31st-1st August, 2017.
4. **N. S. Munjal**, S. Dutta, M. Sharma, and C. Rout, “QSAR and QSPR model development and comparison for drugs having low solubility,” 6th *International Conference on New Frontiers of Engineering, Science, Management and Humanities (ICNFESMH)*, held at Institution of Electronics and Telecommunication, Chandigarh, December 9th, 2017.

5. **N. S. Munjal**, M. Sharma, and C. Rout, “Substituent structure-solubility relationship for prodrug designing with improved solubility profile,” *International Conference on Advances in Biosciences and Biotechnology (ICABB)*, held at Jaypee Institute of Information Technology, Noida, February 1st-3rd, 2018. [Third prize in poster presentation].
6. **N. S. Munjal**, B. Sharma, A. Khanduri, and C. Rout, “Development of QSAR model for ICL inhibitors,” *International Conference on Advances in Biosciences and Biotechnology (ICABB)*, held at Jaypee Institute of Information Technology, Noida, February 1st-3rd, 2018.
7. **N. S. Munjal**, and T. R. Singh, “Docking studies of Paclitaxel and Paclitaxel prodrugs with CYP enzymes to correlate solubility and bioavailability,” *Indian Conference on Bioinformatics (Inbix)*, held at Hansraj Mahila Mahavidyalay, Jalandhar, February 22nd-23rd, 2019. [Best poster stub talk award].
8. S. Dutta, **N. S. Munjal**, **T. R. Singh**, “SolPro: Database for the solubility prediction of poorly soluble drugs,” *International Conference on Recent Trends in Biotechnology and Bioinformatics (ICBAB)*, held at Jaypee University of Information Technology, Waknaghat, August 1st-3rd, 2019.