STUDY OF GENETIC VARIATIONS IN PODOPHYLLUM HEXANDRUM AND COMPUTATIONAL SCREENING OF PODOPHYLLOTOXIN ANALOGUES

by

Md. Afroz Alam

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

DOCTOR OF PHILOSOPHY

IN

BIOINFORMATICS

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

MAY 2009

Dr. Pradeep Kumar Naik
Senior Lecturer
Biotechnology & Bioinformatics

Jaypee University of
Information Technology
Waknaghat-173 215
Solan, Himachal Pradesh
Phone No.: 91-1792-239227
Fax No.: 91-1792-245362

CERTIFICATE

This is to certify that the thesis entitled "Study of genetic variations in *Podophyllum hexandrum* and computational screening of podophyllotoxin analogues" submitted by Mr. Md. Afroz Alam to the Jaypee University of Information Technology, Waknaghat in fulfillment of the requirement for the award of the degree of Doctor of Philosophy in Bioinformatics (Science) is a record of bona fide research work carried out by him under my guidance and supervision and no part of this work has been submitted for any other degree or diploma.

(Dr. P. K. Naik)

DECLARATION

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

Md. Afroz Alam

ACKNOWLEDGEMENT

The great desire to acquire higher qualifications and pursue research drove me to promise my parents, teachers, brothers and my sister a useful research work. My promise was to live up to their expectations and never to let them down. And I kept my word.

I express my heartfelt gratitude to all those who have contributed directly or indirectly towards obtaining my doctorate degree and at the same time I cherish the years spent in the department of Bioinformatics and Biotechnology. I am highly indebted to my esteemed supervisor, Dr. Pradeep Kumar Naik, who has guided me through thick and thin. This project would not have been possible without his guidance and active support. His positive attitude towards research and zest for high quality research work has prompted me for its timely completion. I deem it a privilege to be a scholar doing research under Dr. P. K. Naik who has endeared himself to his students and scholars.

I am indebted to Dr. Y. Medury (Vice Chancellor, JUIT), Brig. (Retd) Balbir Singh (Registrar, JUIT) and Prof. R. S. Chauhan (Head of the department) for having provided all kinds of facilities to carry out research. My sincere thanks go to Dr. Gyan Prakash Mishra (Scientist DRDO, Leh), Harvindar Singh and the faculty, Dept. of Biotechnology & Bioinformatics, for their moral support.

It gives me immense pleasure to record my profound regards and sincere gratitude to Dr. T. Mohapatra, Principal Scientist (IARI, New Delhi) for his support, inspiration and advice during the course of my AFLP analysis.

I also thank the librarians Mr. Shree Ram and Mr. Ashok Sharma for their support in providing me with journals and articles during the period. I would also like to thank technical staff as well as staff members of JUIT for their support.

I can never ever forget my seniors Dr. Vinod Goyal, Dr. Sanjay Kalia, Mr. Swarup and junior colleagues Mani, Sree Krishna, Jatin and friends, who have always been a constant

source of inspiration and helped me in a numerous ways. I would also like to thank members of the lab and colleagues from other labs for their constant support. It was a pleasure to work with them. I also thank Somlata, Ismail, Ravikant, Baleshwar, Mamta for their assistantance. I specially thank my special friends, Mujahid, Dr. Mamnoon, Dr. Khalid Rahman, Iqbal Razi, Ashok Thakur, Rakesh, Tanzeer, Vinay Singh, Ambika, Shalendra, Govind for being with me throughout the research work.

As is usual I adore my parents as Gods. My loving and caring brothers and sisters have been quite supportive during the research work. I can not but appreciate their kind gesture. I would fail in my duty if I don't make a mention of Sangeeta Thakur who stood by me in the hour of need providing support and guidance.

Sincere Thanks to my school and college teachers specially Dr. Sayed Hasan (Padam Shree awardee) who has always been inspirational for putting in hard work.

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

This work has been supported by funds from Department of Science and Technology, Government of India, in the form of young scientist scheme to Dr. Naik and is duly acknowledged.

(Md.Afroz Alam)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Podophyllum hexandrum Royle (Berberidaceae), more popularly known as Indian Mayapple is a herbaceous, rhizomatous species of great medicinal importance. It has long been used by the Himalayan natives and the American Indians. The rhizomes of several *Podophyllum* species have been found to be the source of podophyllotoxin, 4'-demethyl podophyllotoxin and podophyllotoxin 4-O-glucoside. These compounds have been used for the treatment of lung and testicular cancers as well as certain types of leukemia. Podophyllotoxin is also the precursor of semi synthetic chemotherapeutic drugs such as etoposide, teniposide and etophos. The Indian species *P. hexandrum* contains three times more podophyllotoxin than its American counterpart *P. peltatum*. The rhizomes are being indiscriminately harvested in large quantities from the wild to meet the ever increasing demand for the crude drug. As a result of this and a lack of organized cultivation, *P. hexandrum* has been reported as a threatened species in the Himalayan region. Some of the populations in certain pockets have virtually disappeared owing to anthropogenic activities and overexploitation. As the species is already endangered and exploitation of its underground parts continues to exceed the rate of natural regeneration, it needs immediate attention for conservation. Studies of its population biology and genetic diversity are important for successful development of conservation strategies. Podophyllotoxin content varies greatly between populations of *P. hexandrum* from the Northwestern Himalayas. Therefore, it would be inefficient, environmentally destructive and economically unsound to randomly harvest Mayapple. For the rapid development of *P. hexandrum* as an economically viable alternative crop, it would be essential to evaluate the natural populations of *P. hexandrum* for podophyllotoxin content and agronomic traits in the Himalayan region. Appropriate methods of conservation management should be adopted including *in situ* conservation and germplasm collection from the remaining populations with great genetic variations.

Twenty-eight populations of *P. hexandrum* were collected from 28 sites covering 11 forest divisions in the interior range of the Northwestern Himalayan region, Himachal Pradesh, India for study of genetic variations using molecular markers such as RAPD, ISSR and AFLP.

Nineteen random decamer primers (Operon Tech USA) A, B, C and D series were used for RAPD analysis. A total of 28 plant samples were fingerprinted using 19 RAPD makers. 131 (an average of 6.89 bands per primer) RAPD loci were scored, out of which 121 (92.37%) were polymorphic and only 10 (7.63%) were monomorphic loci. The number of amplification fragments produced per primer as well as their size range from 250 bp to 3,100 bp which is analytically appropriate and in conformity with those recorded with certain other plants examined analogously. The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variations in the *Podophyllum* population. The resolving power of the 19 RAPD primers ranged from 2.0 for primer OPC15 to a maximum of 16.571 for primer OPA18. All the 28 *Podophyllum hexandrum* populations were distributed into 12 main clusters as per their forest divisions with similarity value ranging from 0.61 to 0.96. Over half of the total variations in the studied populations: 53% and 47% could be accounted for by differences among the forest divisions and between populations within a forest division, respectively. The total gene diversity (Ht) among populations was 0.338 and within populations (Hs) was 0.104. Shannon's information index was 0.501 and estimated gene flow was found to be 0.110 among the 28 *P. haxandrum* populations. The RAPD study indicates that *P. hexandrum* populations in the northwestern Himalayan region are genetically highly diverse.

For ISSR, eleven primers (after being screened out of 30 primers) were selected and were commercially synthesized by Sigma Inc. Eleven ISSR primers used in the study generated a total of 68 ISSR loci (an average of 6.18 bands per primer) out of which 57 were polymorphic (83.82%). It differed substantially within the discrete groups of plants with an average of 15.05% and was found to be in between 5.88% (Kinnaur forest division) and a maximum of 27.94% (Bharmaur forest division). The genetic diversity was high (percentage of polymorphic loci = 83.82%; Shannon's information index, I = 0.441) at the population level. The mean coefficient of gene differentiation (Gst) was 0.630, indicating that 29.44% of the genetic diversity resided within the population. An overall value of mean estimated number of gene flow (Nm = 0.147) indicated that there was limited gene flow among the sampled populations. The high Gst value (0.630) and the low Nm value (0.147) both indicated rapid genetic differentiation among the 28 populations, especially among the

regions. About 62% of the genetic variation in the samples can be attributed to variation among populations, indicating that among populations genetic diversity is higher than within population genetic diversity (38%) in the studied populations. All the 28 populations were grouped under 11 main clusters as per their forest division with the similarity index values ranged from 0.57 to 0.96. This indicated that gene flow in the populations of *P. hexandrum* occurred mainly within the same forest division rather than among divisions.

In order to facilitate reasoned scientific decisions on its management and conservation and prepare for selective breeding programme, genetic analysis of 28 populations was performed using amplified fragment length polymorphism (AFLP) markers. The thirteen pairs of AFLP primers (*Eco*RI/*Mse*I) generated a total of 551 loci out of which 466 (84.4%) were polymorphic. The mean coefficient of gene differentiation (Gst) was 0.51, indicating that 26% of the genetic diversity resided within the population. Analysis of molecular variance (AMOVA) indicated that 64% of the genetic diversity among the studied populations was attributed to geographical location while 36% was attributed to differences in their habitats. An overall value of mean estimated number of gene flow (Nm = 0.24) indicated that there was limited gene flow among the sampled populations. The high levels of population differentiation detected suggest that provenance source is an important factor in the conservation and exploitation of *P. hexandrum* genetic resources.

Populations from 11 forest divisions were clustered into region-specific groups using RAPD, ISSR and AFLP based DNA profiling with the exception of Kullu forest division. The high genetic variation in *P. hexandrum* may be attributed partly to the cross- pollinated nature or clonal propagation of *P. hexandrum.* The low level of genetic diversity within the population and low gene flow among populations detected in this study point towards the possibility of a single isolated population possessing unique genotypes not found in other populations. It is, therefore, imperative for conservation planners in designing conservation strategies for wild populations of *P. hexandrum*, to ensure that many possible separate populations are targeted for conservation rather than conserving a few selected populations.

For conservation aspects it is very important to assess the potential of *P. hexandrum* populations for podophyllotoxin production. Total synthesis of podophyllotoxin is an expensive process and availability of the compound from the natural resources is an important issue for pharmaceutical companies that manufacture anticancer drugs. It was found that the podophyllotoxin content in the root of the plants obtained from the Lahaul forest division (at altitude 4300 m) was high (8.857 to 9.533% on dry weight basis) compared to the root samples collected from other forest divisions. The lowest values obtained were samples from Parvati forest division (at altitude 1300 m) 3.020 to 4.753% on dry weight basis. For populations in the same forest division as well as other forest divisions, the podophyllotoxin content increased with the increase in altitude. The existing variations in podophyllotoxin content among the populations were proved to be coupled with altitude and environmental variables but not with genetic diversity. Thus the study demands the optimization of soil nutrient and environmental factors in order to increase the rate of production of podophyllotoxin from collected populations.

In order to facilitate reasoned scientific decisions on its domestication, conservation and sustainable utilization, the effects of soil nutrients and environmental factors on podophyllotoxin content in the rhizome of *P. hexandrum* were investigated in the Northwestern Himalayan region, Himachal Pradesh, India. The variation in podophyllotoxin content is highly dependent on climatic factors. It was seen to be related positively with humidity; $r = 0.825$ (at after noon) and $r = 0.844$ (at fore noon) and it reached statistical significance level $P < 0.001$. The correlation coefficient between podophyllotoxin content was -0.595 (significant at $P < 0.01$) with maximum rainfall and 0.717 (significant at $P < 0.001$) with minimum rainfall. The linear correlation coefficient (r) was -0.720 for maximum temperature (significant at $P < 0.001$) and -0.635 ($P < 0.001$) for minimum temperature and are negatively correlated with podophyllotoxin content. The podophyllotoxin content reached higher than 6.62% of root dry weight when soil pH value was about 4.82, soil organic carbon was higher than 3.23% and nitrogen content was higher than 2.7% of soil dry weight. However, soil available phosphorous content higher than 0.419% and potassium content higher than 1.56% resulted in low podophyllotoxin content. The strong and linear relationship detected between podophyllotoxin as well as soil nutrients, environmental

factors and altitude suggested that further optimization of these factors are very important in the conservation and exploitation of *P. hexandrum.* In this regard the prediction model like artificial neural network (ANN) and multiple linear regression (MLR) developed in this study to map the effect of these factors on podophyllotoxin yield will be of great help. The ANN prediction model revealed better prediction of yield ($r^2 = 0.9905$) than MLR prediction model ($r^2 = 0.9302$). Low root mean square error (RMSE) for ANN model (0.0399) is less than MLR model (0.2939) with respect to the experimental measurement which establishes the ANN method as an efficient tool for optimization of soil nutrients and climatic factors for podophyllotoxin yield.

Podophylltoxin and its structural derivatives are well known for their anti-cancer activity since long. However, the clinical application of podophyllotoxin and its analogues in the treatment of cancer has been limited by severe toxic side effects during administration of the drugs. Further, improvement in their clinical efficacy by overcoming the drug resistance, myelosuppresion and poor bioavailability problems associated with podophyllotoxin, still continues to be challenging. Consequently, there has been considerable increase in the number of podophyllotoxin analogues. In this regard it is necessary to investigate and prepare new, more potent and less toxic analogues with better therapeutic indices. In a number of experimental studies during the past decade, a large number of analogues have been designed, synthesized and evaluated. These include modification of the A ring, B ring, C ring, D ring and E ring in the scaffold structure of podophyllotoxin. These studies provided important insights into development of the structure-activity relationship models between analogues. This has resulted in a paradigm shift to identify such problems early in the drug discovery process. Virtual screening, using a computational approach to assess the interaction of an *in silico* library of small molecules and the structure of a target macromolecule has arisen as an alternative method for the rapid identification of new drug leads. Virtual screening models such as QSAR, Docking, MM-GB/SA and SGB-LIE developed in this work can be extended to facilitate the search for the potential drugs with low toxicity and better biological activity against cancer based on podophyllotoxin skeleton.

The virtual library of podophyllotoxin analogues contains 154 compounds divided into 5 sub libraries: tetralinelactones consists of 52 compounds; non-lactonic tetralines consist of 45 compounds; pyrazolignans and isoxazolignans consist of 23 compounds; lactonic and non-lactonic naphthalene consist of 6 compounds and aza-podophyllotoxin analogues consist of 28 compounds. All these compounds belong to different ring modifications. The virtual library of podophyllotoxins was built from the scaffolds by different ring modifications and substitutions of various functional groups.

The molecular modeling techniques: docking-molecular mechanics based on generalized Born/surface area (MM-GBSA) solvation model was used for virtual screening of podophyllotoxin derivatives for energetically favorable interaction with the tubulin which is the therapeutic targets for podophyllotoxin and the energy score has been used to build models for prediction of the inhibitory activity (pIC_{50}). All the 154 podophyllotoxin analogues have been well fitted to the defined binding pocket and were found to be the good binders with tubulin. Quantitative structure activity relationships were developed between the cytotoxic activity (pIC_{50}) of these compounds and molecular descriptors like docking score and binding free energy. For both the cases the r^2 was in the range of 0.642-0.728 indicating good data fit and r^2_{cv} was in the range of 0.642-0.728 indicating that the predictive capabilities of the models were acceptable. In addition, a linear correlation was observed between the predicted and experimental pIC_{50} for the validation data set with correlation coefficient r^2 of 0.806 and 0.887, suggesting that the docked structure orientation and the interaction energies are reasonable. Low levels of root mean square error for the majority of inhibitors, established the docking and prime/MM-GBSA based prediction model as an efficient tool for generating more potent and specific inhibitors of tubulin by testing rationally designed lead compounds based on podophyllotoxin derivatization.

Podophyllotoxin and its analogues have important therapeutic value in the treatment of cancer due to their ability to induce apoptosis in cancer cells in a proliferation-independent manner. These ligands bind to colchicine binding site of tubulin near the α- and β-tubulin interface and interfere with tubulin polymerization. The structure-based linear interaction energy method implementing a surface generalized Born (SGB-LIE) continuum

model for solvation was used to build a binding affinity model for estimating the free energy of binding for a diverse set of podophyllotoxin analogues. The analogues were docked into the colchicine binding site of tubulin and the docked complexes were used for subsequent SGB-LIE calculations. The molecular dynamics (MD) technique has been used for LIE conformation space sampling in the present work. A training set of 76 podophyllotoxin analogues was used to build a binding affinity model for estimating the free energy of binding for 36 inhibitors (test set) with diverse structural modifications. The root mean square error (RMSE) between the experimental and predicted binding energy values was 0.56 kcal/mol which is an indicator of the robustness of the fit. The quality of the fit can also be judged by the value of the squared correlation coefficient (r^2) which was 0.8714 for the training set. SGB-LIE model developed in this study is statistically (q^2 = 0.8647, r^2 = 0.8714, F = 166.77) best fitted and consequently used for prediction of cytotoxic activities (pIC_{50}) of training and test sets of molecules. The squared correlation coefficient between experimental and SGB-LIE estimates for the free energy of the test set compounds is also significant (r^2 = 0.733). On the basis of the analysis of the binding energy, we propose that the three-dimensional conformation of the A, B, C and D rings is important for interaction with tubulin. On the basis of this insight, 12 analogues of varying ring modifications were taken, tested with LIE methodology and then validated with their experimental potencies of tubulin polymerization inhibition. Low levels of RMSE for the majority of inhibitors establish the structure-based LIE method as an efficient tool for generating more potent and specific inhibitors of tubulin by testing rationally designed lead compounds based on podophyllotoxin derivatives. The magnitude of free energy changes upon binding of inhibitors to tubulin directly correlates with the experimental potency of these inhibitors. Thus fast and accurate estimation of binding free energies provides a means to screen the compound libraries for lead optimization and rational design. Hence, this could bring about the development of new and more effective drugs.

Podophyllotoxin and some derivatives with cytotoxic activity against P388 cell line and tubulin polymerization inhibition were studied using classical quantitative structure activity relationship (QSAR). Though the podophyllotoxin derivatives have received a great deal of attention by medicinal chemists, the literature survey resulted in little evidence on

them being investigated by means of molecular modeling techniques. A series of 119 podophylltoxin derivatives with known experimental biological activities have been selected for studies in order to trace structure-activity relationship. The fundamental hypothesis of the QSAR methodology is that the biological activity is a function of the molecular structure. This method is used to find empirical relationships in a set of compounds (the instructional set) that are known to have interesting properties. In this instance, this is the biological activity (cytotoxicity). A set of 372 molecular descriptors which includes E-state indices, electronic, topology, thermodynamics, structural, information content and lead likeness have been used to derive a QSAR. We used a more systematic way of selecting the meaningful descriptors which includes in order of missing value test → zero test → simple correlation test → multicollinearity test → genetic algorithm among the large quantity of potentially possible initial ones. We have built a robust QSAR model with high values of q^2 (0.888) for training sets and predictive r^2 (0.918) for test set including descriptors like shadow area 4 and 5, 5^{th} order chain molecular connectivity valence, maximum free radical superdelocalizability, difference between minimum and maximum E-state values, dipole moment, minimum nucleophilic superdelocalizability, length to breadth ratio, angle strain energy of molecule, mass weighted length to breadth ratio and relative positive charge surface area. The model was also tested successfully for external validation criteria. The high predictive ability of the model allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates.

Therefore, the computational models such as docking-MMGBSA, LIE-SGB and QSAR developed in this study should facilitate the rational design of novel derivatives, guide the design of focused libraries based on the podophyllotoxin skeleton and facilitate the search for related structures with similar biological activity from large databases.

CHAPTER 1

Introduction

Indian Podophyllum (*Podophyllum hexandrum*, Royle Berberidaceae) is a rhizomatous perennial native to India. The name *Podophyllum* is derived from ancient Greek words *podos* and *phyllon*, meaning "foot leaves". The genus *Podophyllum* is also called Mayapple because its fruit ripen in spring. *Podophyllum* has been known to the Hindu physicians since ancient times and has been variously described in Indian vernacular names: *Sanskrit*-Laghu Patra; *Hindi*-Ban-Kakan, Papra, Papri, Bhavan-bakara; *Bengali*-papra, *Kashmiri*-Banwangan; *Punjabi*-Ban Kakri; *Marathi*-Padwall. The possibility of Himalayan *Podophyllum hexandrum* of being "Aindri"- a divine drug of the ancient Indians had found mention in Susruta samhita. In a broad sense the genus *Podophyllum* is represented by two species: *Podophyllum hexandrum*, commonly distributed in the Himalayan regions and popularly called Himalayan mayapple while the other species is *Podophyllum peltatum,* commonly distributed in the Atlantic North America and popularly known as American mayapple. *Podophyllum hexandrum,* an endangered medicinal herb grows wild in the interior Himalayan ranges of India (Gupta and Sethi, 1983). The population size of *P. hexandrum* is very low (40-700 plants per location) and is declining every year. Some of the populations in certain pockets have virtually disappeared owing to anthropogenic activities and overexploitation (Bhadula et al., 1996). Thus, there is a need to conserve genetic diversity of this prized medicinal plant which may become extinct if its reckless exploitation continues. It is recognized for its anti-cancer properties. The rhizomes and roots of *P. hexandrum* contain anti-tumor lignans such as podophyllotoxin, 4'-demethyl podophyllotoxin and podophyllotoxin 4-o-glucoside (Tyler et al., 1988; Broomhead and Dewick, 1990). Among these lignans, podophyllotoxin is the most important for its use in the semisynthesis of anti-cancer drugs etoposide and teniposide (Issell et al., 1984). Podophyllotoxin content in Himalayan mayapple is high (4.3%) compared with other species of *Podophyllum,* notably *P. peltatum* (0.25%) (Jackson and Dewick, 1984).

1.1 Ecology, Taxonomy and Biology of the Himalayan Mayapple

Podophyllum hexandrum is believed to have originated in the Himalayan regions. Himalayan mayapple has been found growing in the wild in the Himalayan regions of the Indian subcontinent and different parts of North and South America. In India the presence of this species has been reported from Zanskar & Suru valleys of Ladakh, Kashmir, Lahaual &

Spiti valleys, Kangra, Chamba, Kinnaur of Himachal Pradesh, Sikkim and Arunachal Pradesh. The populations of *P. hexandrm* are found growing on open slopes and alpine pastures in moist humid condition. These also grow in the forest rich in humus and decayed organic matter as well as in cultivated fields between 2,600 and 4,500m altitude. The species is generally associated with the species of *Rhodenddron, Salix, Juniperus* and *Viburnum.* The soft leaves of *Podophyllum hexandrum* indicate that the plant could be sensitive to heat and direct sunlight mostly in the dry season. The deep-growing root system needs sufficiently deep soil to reach the ground water table. The sensitivity to frost and the late sprouting in spring are the characteristics which *Podophyllum hexandrum* shares with other plants that prefer sites with an early, high and long lasting snow cover. The genus *Podophyllum* shows a great range of variations. As a result, there are different views regarding the inclusion of *Podophyllum* and its relatives in a distinct family of 'Podophyllaceae' or as a subfamily of Berberidaceae. The genus *Podophyllum* is classified according to Benthum & Hooker's system of classification as follows:

Kingdom: Plantae
Division: Magnoliophyta
Class: Magnoliopsida
Order: Ranunculales
Family: Berberidaceae
Genus: *Podophyllum*
Species*:* *hexandrum* L.

The plant is an erect, glabrous, succulent herb, 35-60 cm. tall with creeping, perennial rhizome, bearing numerous roots, leaves 2 or 3 orbicular periform, palmate, peltate with lobed segments, flower solitary, white or pink, cup-shaped, fruit an oblong or elliptic berry, 2.5-5.0 cm. diameter, orange or red, containing many seeds embedded in the pulp. Mayapple is described as self-incompatible but some researchers believe that colonies in the wild may come from a single seedling. Thus one genotype grows in clonal patches (Laverty and Plowright, 1988). In contrast, Policansky (1983) reported that mayapple colonies comprised more than one genotype and intra-population crosses resulted in lower seed set than inter-population crosses. This evidence suggests that mayapple is at least partially self-incompatible. Plants remain juvenile for 4 to 5 years and during this time there is upward growth of the underground bud. The rhizome grows horizontally from that single

bud and ends with a terminal bud after reaching maturity. Each terminal bud produces a shoot next season. The rhizome continues to develop annually, producing elongated internodes between nodes. Each node is a complex structure composed of a compressed stem, a main bud that develops into the next season's growth and minor buds that can develop and continue rhizome growth if the terminal bud is lost. Roots develop at the base of the node and can also arise from the internodal tissue near the terminal bud of the rhizome. A single shoot arises from each node of the rhizome and is either asexual, producing a single leaf or sexual, producing a forked petiole with two leaves and a solitary flower (Geber et al., 1997). Both asexual and sexual growths emerge in early spring before trees produce leaves and then senesce by midsummer. Landa et al., (1992) showed that photoassimilates produced during the growing season and stored in the roots, rhizomes and nodes at the onset of leaf senescence are translocated to newly developing growing points in the following spring. Plants senescing later produce longer and heavier rhizomes (de Kroon et al., 1991). Senescence of plant is affected by several factors such as vigor of the rhizome system, the genotype and the environment to which the plant is exposed. Vegetative propagation of mayapple is by rhizome cuttings or by micropropagation using the terminal bud as the source of explant-inducing adventitious buds with 70% to 90% success rate in soil acclimatization (Moraes-Cerdeira et al., 1998). As for growing mayapple, one may expect that shade will provide the best condition for growth since numerous reports on colonies describe the plant as an understory species. The populations of *Podophyllum hexandrum* shows a range of variation in plant height, number of leaves (1 to 4 leaved), leaf insertion, size and shape of the fruit and the colour of seeds. The rhizomes are externally yellowish brown to earthy brown in colour with characteristic odour, somewhat bitter and acrid in taste. On the basis of leaf number, four morphotypes viz. single leaved, double leaved, triple leaved and four leaved have been reported by Bhadula et al., (1996) from Garhwal Himalaya.

Figure 1.1. The plant *Podophyllum hexandrum* and its different parts.

1.2 Importance of podophyllotoxin and its derivatives

The Mayapple has been known for its medicinal uses since long by the Himalayan natives and the American Indians as a canthartic cholagog (Anon, 1970). It has been found listed along with its resin, "podophyllin" in most of the European, South American and Asian pharmacopoeias. Among the plethora of physiological activities and potential medicinal and agricultural applications, the antineoplastic and antiviral properties of podophyllotoxin congeners and their derivatives are arguably the most eminent from a pharmacological perspective (Ayres et al., 1990). The whole plant has also got great importance in traditional systems of medicine including Ayurveda, Unani and Tibetan for curing several diseases. The recent studies conducted by the Institute of Nuclear Medicine & Allied Sciences (DRDO) reveal that the podophyllotoxin has got the property of radioprotection (Kumar et al., 2005). Podophyllotoxin is the most important as an antitumor agent due to its biological activity blocking mitosis (Loike et al., 1978) and its use as the starting compound of the semi-synthetic chemotherapeutic drugs such as etoposide, teniposide and etopophos (Stahelin and Wartburg, 1991) (Figure 1.2).

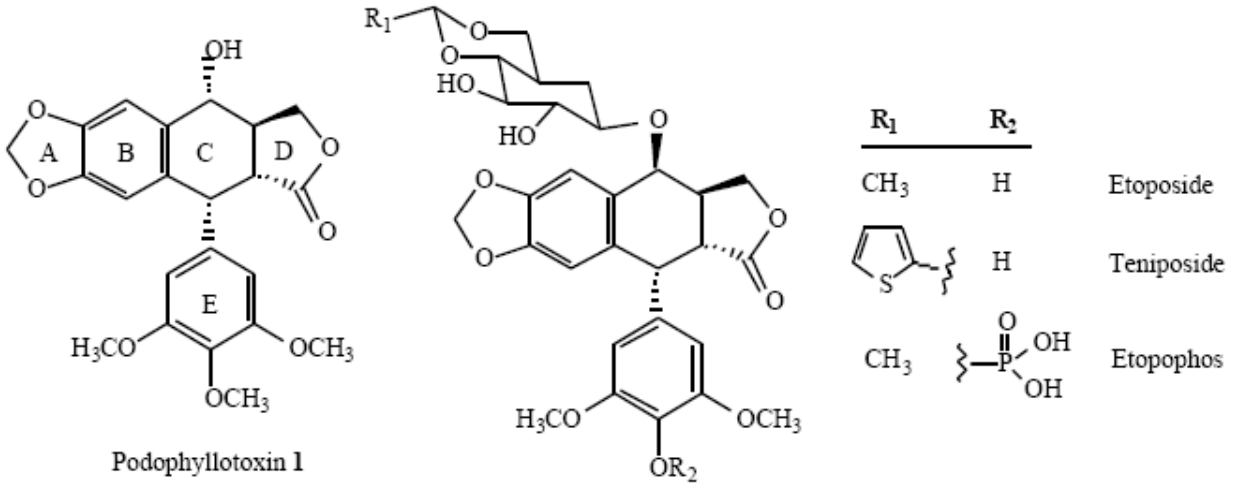


Figure 1.2. Structures of podohyllotoxin and its derivatives.

These antineoplastic pharmaceuticals block DNA toposisomerase II (Horwitz and Loike, 1977; Minocha and Long, 1984) and have been used for the treatment of small and large cell lung, refractory testicular, stomach, pancreatic cancers and myeloid leukemias (Ekstrom et al., 1998; Holm et al 1998; Ajani et al., 1999). The successful derivatization of podophyllotoxin into potent antineoplastic drugs like etoposide and teniposide has generated interest in structure optimization to produce new derivatives with superior pharmacological profiles and broader therapeutic uses. Numerous derivatives leading to change in the basic structure of etoposide have been proposed, synthesized and clinically tested. Etopophos is a new etoposide phosphate designed to overcome the limitations associated with the poor solubility of etoposide. Etopophos can be administered intravenously in higher doses and rapidly converted to etoposide by phosphatase in the blood plasma leading to an improvement in the treatment (Schacter, 1996). Further, NK 611, TOP 53 and GL 311 are among the most promising derivatives developed so far that attempt to increase the biological activities as more potent drugs than etoposide. The administration of podophyllotoxin-derived drugs causes complex physiological reactions beyond inhibition of DNA topoisomerase and tubulin polymerization. A mixture of benzylidinated podophyllotoxin glycosides is a new drug (CPH 82) for the treatment of rheumatoid arthritis and psoriasis. Arthritis patients treated with CPH 82 have shown a reduction of the inflammatory process within three months of therapy in the first and second phases of clinical trials in Europe (Lerndal and Svensson, 2000). These results suggest that CPH 82 is a safe and efficacious drug for rheumatoid arthritis with gastrointestinal inconveniences as side effects (Bjorneboe et al., 1998). The therapeutic value of podophyllotoxins as mitosis inhibitors has other medicinal applications including uses as anti-malarial and anti-fungal agents with immune modulator activities (Leander and Rosen, 1988; Pugh et al., 2001). Podophyllin from *P. hexandrum* contains the major active constituent and the amount ranges from 32 to 54 percent; a number of other related compounds and their glucosides have been isolated from the resin (Table 1.1). It also contains quercetin (8%), kaempferol, astragalin (kaempferol-3-glucoside) and an essential oil (3.7%) responsible for the odour of podophyllin, wax (8.6%) and mineral salts. Podophyllin is considered as a cholagogue, purgative, alternative, emetic and bitter tonic. Owing to its cytotoxic action it is used as

paint in the treatment of soft venereal and other warts. An ointment of podophyllin is employed to remove warts in animals (Kaplan, 1942).

Table 1.1. Lignan compounds from *Podophyllum hexandrum.*

Lignan Compound	Molecular Formula	Meting Point (^{0}C)
Podophylltoxin	C22H22O8	183-84
Podophylltoxin β D-glucose	C26H32O13	149-52
Picropodophyllin 1,b	C22H22O8	235-36
1-O-(β-D-glucopyranosyl)-picropodophyllin	C28H32O13	237-38
4'-Demethylpodophylltoxin	C21H20O8	250-52
4'-Demethylpodophylltoxin β D-glucose	C27H20O8	165-67
Dehydropodophyllotoxin	C20H18O8	272-74
Podophyllol	C20H22O8.H2O	115 decomp.
Podophyllic acid	C22H24O9	163-65 decomp.
4'-Demethyldeosxypodophylltoxin β D-glucose	C27H30O12	146-58

Due to ever increasing demand of podophyllotoxin, many synthetic chemists have devoted their efforts in developing new routes to the total synthesis of podophyllotoxin. This is, however, a low yield process due to the large number of steps involved (Bush and Jones, 1995). Currently, the preferred source of podophyllotoxin is the Indian *Podophyllum* species and it has acquired the status of a man-made endangered species due to extensive collection and lack of cultivation. To secure podophyllotoxin supply, in this work we have examined the populations of Indian mayapple (*P. hexandrum* L.) and its potential for podophyllotoxin production.

1.3 Sataus and Conservation

Podophyllum hexandrum is one of the important high value medicinal plants found in the high altitude and cold desert trans-Himalayas. The rhizome yields a very valuable drug 'Podophyllotoxin' which has a great demand in pharmaceutical industries throughout the world (Tyler et al., 1988; Broomhead and Dewick, 1990). In addition, the drug obtained

from Himalayan mayapple contains resin of superior quality in comparison to American mayapple (Jackson and Dewick, 1984). So *Podophyllum hexandrum* is being uprooted unscientifically for commercial trade. As a result of this and lack of organized cultivation, *P. hexandrum* has the status of an endangered & threatened species from the Himalayan region (Gupta and Sethi, 1983). No major steps have been taken for commercial cultivation of this very useful plant. During the ethno botanical survey, the authors have reported that the Himalayan mayappple grows with a very limited population in Lahaul & Kinnaur areas of Himachal Pradesh. The populations of this species have been decreasing at an alarming rate in nature due to unscientific exploitation by traders as well as traditional healers (Bhadula et al., 1996). The mode of regeneration of *P. hexandrum* in nature is by seed as well as by rhizome. Since it is subjected to harsh and extended winter conditions it lies dormant for most of the year. The period of growth is confined only to summer months. Flowering and fruiting are erratic; germination is not uniform and seeds of the same batch may take a few months to several years for germination and seedling establishment (Badhwar and Sharma, 1963) resulting in poor regeneration. As the species is already endangered and exploitation of its underground parts continues to exceed the rate of natural regeneration, it needs immediate attention for conservation. Although conservation oriented management is being prioritized on a massive scale in the high altitude & cold desert Himalayas, the knowledge on the design parameters such as shape, size and acreage of conservation areas remain poor. The advantages associated with conservation oriented management need to be objectively analyzed not only to enrich science for conservation but also to motivate people's participation. Sustainable management is the need of the hour that means using/managing of a species, a group of species or a population or an ecosystem at the rate within its capacity for renewal and in a manner compatible with conservation of the diversity. Studies of its population biology and genetic diversity are important for successful development of conservation strategies.

1.4 Genetic diversity of *Podophyllum hexandrum* populations

For the purpose of efficient conservation and successful breeding programmes, it will be prudent to study the populations of *P. hexandrum* at genetic and molecular levels. Study within and between population variations at the molecular level provides an efficient

tool for taxonomic and evolutionary studies and for devising strategies to protect the genetic diversity of the plant. Genetic variability also can be exploited to select useful genotypes that could be utilized as cultivars to avoid batch-to-batch variations in extraction of standard drugs. Considerable variations in morphological characters such as plant height, leaf characteristics, fruit weight, seed weight and colour, etc. and in biochemical characters such as podophylloresin and podophyllotoxin content in rhizomes have been reported in *P. hexandrum* plants from the Garhwal Himalayas (Bhadula et al., 1996; Airi et al., 1997; Purohit et al., 1999). At least four distinct morphological variants with 1, 2, 3, and 4 leaves have been reported (Purohit et al., 1999). Polypeptide profiles in seeds of these variants indicate that these may be genetically distinct from one another. Polypeptide patterns and esterase isozyme analysis have indicated the existence of high inter and intra population variations in *P. hexandrum* from the Garhwal Himalayas (Bhadula et al., 1996). However, morphological and protein markers are influenced by the stages of plant growth as well as environmental factors and hence may give erroneous results. DNA markers such as RFLPs (Botstein et al., 1980), RAPD (Williams et al., 1990), ISSR (Morgante and Olivieri, 1993) and AFLP (Vos et al., 1995) on the other hand are quite stable and highly polymorphic in nature. These markers are more advanced and have many advantages over the conventional markers. Therefore, morphological characters need the support of molecular information to clarify population structures.

1.5 Use of molecular markers for studying genetic diversity

In simple terms, genetic diversity is a statistical concept referring to the variations within the individual gene loci among alleles of a gene or gene combinations, between individual plants or between plant populations. The classical methods of diversity studies are based on morphological characters which are influenced by various environmental factors. However, the molecular markers which are unrestricted in number and not influenced by the environment have the ability of sampling diversity directly at the genome level. They provide increased accuracy and expanded scope of inferring genetic variability within and between populations of plant species. Traditionally, studies of population genetic structure have used proteins, isozymes, allozyme etc. (Lewontin, 1973; Gardiner et al., 1986; Puecher et al., 1996; Mauria et al., 2000). These markers provide informative genetic markers,

detecting useful levels of genetic variations within populations using straightforward laboratory procedures that are relatively rapid and inexpensive, but there are some widely recognised limitations of their use. Only a little portion of the genome was covered when protein and isozymes were analysed and most part of the genome remained unanalysed. So biochemical characterization does not reflect the entire genome. Furthermore, all genetic changes that occur at the DNA level are not detected at the protein level. The detection of genetic variation is limited to protein coding loci which may not be the representative of the entire genome, hence may lead to underestimation of genetic diversity (Schaal et al., 1991). Enzymes are tissue and species specific and change during development and differentiation. It is likely to get more variability in DNA than in proteins since much of the DNA does not code for active genes and is unexpressed. Analysis of such variations can be done in the form of DNA profiling.

The drawbacks of biochemical markers led many workers to shift to nuclear DNA markers such as RFLPs, RAPDs, ISSRs, AFLPs, etc. DNA-based molecular-marker techniques have been proved powerful in genetic diversity estimations (Lu et al., 1996). Molecular markers, unlike morphological markers are stable and have been found useful in population studies (Aitkin et al., 1994; Lakshmi et al., 1997) and phylogeny (Demeke and Adams, 1994; Adams and Demeke, 1993). Different types of marker systems have been used for biodiversity and phylogenetic analyses. These include restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR), inter simple sequence repeats (ISSR), random amplification of polymorphic DNA (RAPD) (Williams et al., 1990; Karp et al., 1997) and amplified fragment length polymorphism (AFLP) (Vos et al., 1995). RFLPs are well suited for the construction of linkage maps because of their high specificity and are also less polymorphic, more expensive and laborious compared to RAPD and ISSR. Currently AFLP is the method of choice for analysis of germplasm, genetic diversity and phylogeny, gene tagging and molecular map construction (Palacious et al., 1999; Han et al., 2000).

1.6 Various types of DNA markers

There are various types of DNA markers available presently to evaluate DNA polymorphism in sample genomes. Selection of a correct marker system depends upon the type of study to be undertaken and whether that marker system would fulfill at least a few of the mentioned characteristics such as easy availability, highly polymorphic nature, Mendelian inheritance, frequent occurrence in genome, selective neutral behavior, easy and fast assay, high reproducibility, free of epistasis and pleiotropy etc, (Welsh and Clelland, 1990). These markers are generally classified as hybridization based markers and polymerase chain reaction (PCR) based markers. In the hybridization-based markers, the DNA profiles are visualized by hybridizing the restriction enzyme digested DNA to a labeled probe which is a DNA fragment of known / unknown sequence. In case of PCR based markers, the primers of known sequence and length are used to amplify genomic and cDNA sequences which are visualised by gel electrophoresis technique. The invention of PCR which is a very versatile and extremely sensitive technique, (Saiki and Scharf et al., 1985) uses a thermostable DNA polymerase (Saiki et al., 1988) and has changed the total scenario of molecular biology and has also brought about a multitude of new possibilities in molecular marker research. Some of the PCR based marker systems used in the study have been detailed below.

1.6.1 ISSRs as molecular markers

Inter Simple Sequence Repeats (ISSR) is a type of molecular marker that can be carried out without prior knowledge of DNA sequence in the genome. Microsatellites (SSR) represent the most abundant source of polymorphism from repetitive sequences. SSR are often used as molecular markers even if this technology is time consuming and expensive. ISSR is an alternative technique to study polymorphism based on the presence of microsatellites throughout genomes (Zietkiewicz et al., 1994). ISSR markers are DNA sequences delimited by two inverted SSR sequences composed of the same units which are amplified by a single PCR primer, composed of few SSR units with or without anchored end. ISSR-PCR gives multilocous patterns which are very reproducible, abundant and polymorphic in plant genomes (Bornet et al., 2002). This approach named Inter-SSR (ISSR) employs oligonucleotides based on a simple sequence repeat anchored or not at their 5’- or

3'-end by two to four arbitrarily chosen nucleotides. This triggers site-specific annealing and initiates PCR amplification of genomic segments which are flanked by inversely oriented and closely spaced repeat sequences. The marker system called ISSRs has been developed as an anonymous RAPDs-like approach that accesses variation in the numerous microsatellite regions dispersed throughout various genomes (particularly the nuclear genome) and circumvents the challenge of characterizing individual loci that other molecular approaches require (Tautz, 1989). The resultant PCR reaction amplifies the sequence between two SSRs, yielding a multilocous marker system useful for fingerprinting, diversity analysis and genome mapping. ISSR markers are considered to be more reproducible than RAPD markers due to high annealing temperature (Bornet and Branchard, 2001; Chowdhury et al., 2002) and have been used to measure genetic diversity in potato (Bornet et al 2002), barley (Fernández et al., 2002), rice (Joshi et al., 2000), finger millet (*Eleusine coracana*) (Salimath et al., 1995), Sorghum (Yang et al., 1996) and Groundnut (Raina et al., 2001) as well as to identify cultivars in Maize (Kantety et al., 1995; Pejic et al., 1998), Wheat (Nagaoka and Ogihara, 1997), potato (Prevost and Wilkinson, 1999), oilseed rape (Charters et al., 1996) and bean (Métais et al., 2000). Huang et al., (2000) also tried to use ISSR markers to reveal genetic diversity and relationships in sweet potato and its wild relatives.

The ISSRs have several advantages for assessing genetic diversity (Zietkiewicz et al., 1994). They are advantageous because no prior genomic information is required for their use. Abundantly polymorphic and reproducible, they are a good choice for detecting genetic diversity among crop species, germplasm characterization, establishment of genomic relation and molecular phylogeny. ISSRs analyses are more specific than RAPD analyses due to the longer SSR-based primers which enable higher-stringency amplification (Wolfe et al., 1998). The high stringency reduces the problems with reproducibility, a common criticism against the low-stringency RAPD assay (Yang et al., 1996). The shortcomings of ISSR markers like RAPDs are that most bands are scored as dominant markers giving no possibility to distinguish between homozygosis and heterozygosis directly. However, ISSR studies of natural populations have recently demonstrated the hyper variable nature of these markers and their potential use in population-level studies (Ge and Sun, 1999; Culley and

Wolfe, 2001). The technique allows for dissection below the subspecies level and this gives it a good level of applicability in the study of rare or endangered plants (Blair et al., 1999).

ISSRs have been used in conjunction with RAPD data to determine the colonization history of *Olea europaea* in Macronesia along with lineages in the species complex (Hess et al., 2000); *Cicer* and cultivated chickpea (Iruela et al., 2002). The two techniques have also been utilized in examining the genetic diversity, varietal identification and phylogenetic relationships in peanut (*Archis hypogaea*) cultivars and wild species (Raina et al., 2001). The phylogenetic trees generated by different methods were largely congruent topologically. They are mostly dominant markers.

1.6.2 RAPDs as molecular markers

Random amplified polymorphic DNA is a PCR-based technique that has been applied to the study of populations (Williams et al., 1990). RAPDs are one of a family of techniques that produce arbitrary fragment length polymorphism and are collectively described as multiple arbitrary amplicon profiling (Caetano-Anolles, 1994). The RAPD technique utilizes single, arbitrary, decamer DNA oligonucleotide primers to amplify regions of the genome using PCR (Welsh and McClelland, 1990; Williams et al., 1990; Williams et al., 1993). Priming sites are thought to be randomly distributed throughout a genome and polymorphism in these regions results in different amplification products. The methodology is simple and has been widely used for the assessment of genetic diversity, genetic variation within species, determining relationships between closely related species and genotypes within a species to identify particular genotypes (cultivar identification). RAPD technique has also been used to study and investigate clonal and population structure (Kresovich et al., 1992).

The use of RAPD for determination of genetic relationships has been demonstrated in a number of crop species like maize (Welsh et al., 1991), Sorghum (Dahlberg et al., 2002) rapeseed (Forster and Knaak, 1995), pigeon pea (Ratnaparkhe et al., 1995), aromatic rice (RayChoudhury et al., 2001) and many other crops. The simplicity of the technique and the speed of data generation have attracted many researchers, particularly those interested in either genetic fingerprinting or the patterns and levels of genetic diversity (Kazan et al.,

1993 and Koller et al., 1993). In addition to the studies of genetic diversity there have been an increasing number of papers concerned with population genetics (Crochemore et al., 1996), phylogenetics (Stewart and Porter, 1995) and hybridisation/introgression (Sale et al., 1996; Comes and Abbott, 1999). The greatest attraction of this method is that it generates DNA data that are, theoretically at least, randomly scattered across the genome. Such markers are attractive for studies that involve differentiation of similar species (Zamora et al., 1996) and identification of patterns of variation (Vasconcelos et al., 1996). Rieseberg (1996) suggested that RAPDs may be useful for investigation within species or between closely related species. Many reports are available on inter and intra generic genetic diversity and molecular phylogeny using RAPDs such as, *Arachis* (Gimenes et al., 2000), *Pistacia* (Kafkas and PerlTreves, 2002) etc. The usefulness of RAPD fingerprinting was also reported for identification of Italian grape (*Vitis vinifera*) varieties (Mulcahy et al., 1995) and determining the phylogenetic relationship for 28 tropical maize varieties (Parentoni et al., 2001). All the above studies confirmed the efficiency of RAPD markers for systematic investigations. Thus, it has been suggested that RAPD fingerprinting method is simple and so powerful that one primer can distinguish between different clones while the use of multiple primers reduces fingerprint similarity and resolves discrepancies. RAPD analysis is a very good starting point for studies of relationships within and among closely related species. Only one report was so far available on RAPD characterization of *P. hexandrum* collected from Chamba and Kullu districts of Himachal Pradesh (Sharma et al., 2000).

1.6.3 AFLP as molecular markers

The Amplified Fragment Length Polymorphism (AFLP) DNA markers technique, originally known as selective restriction fragment amplification (SRFA) (Zabeau and Vos 1993), produces highly complex DNA profiles by arbitrary amplification of restriction fragments ligated to double-stranded adaptors with hemi-specific primers harboring adaptor-complementary 5' termini (Vos et al., 1995). The technique has been widely used in the construction of genetic maps containing high density of DNA marker loci. The resulting AFLP fingerprints are usually a rich source of DNA polymorphisms that can be used in mapping and general fingerprinting endeavors. Molecular marker profiles based on AFLPs can be used to detect variation at the DNA level and have proved to be extremely effective

in distinguishing closely related genotypes. Additional advantages of this technique include: reproducibility, high levels of polymorphism detection, genome-wide distribution of markers and no required prior knowledge of the genome (Prabhu and Gresshoff, 1994). AFLP has been used to assess genetic relationships of a wide range of species including somaclonal variants of *Arabidopsis thaliana* Schur. (Polanco and Ruiz, 2002) and sports of *Agave fourcroydes* Lem. (Gonzalez et al., 2003) and *Rosa* spp. (Debener et al., 2000; Vosman et al., 2004) and clones of *Rhododendron ferrugineum* L. (Escaravage et al., 1998). AFLP can be used for mapping, fingerprinting and genetic distance calculation between genotypes. The advantage of AFLP is its high multiplicity and therefore the possibility of generating high marker densities.

A study on the genetic diversity of an endangered alpine plant (*Eryngium alpinum* L.) demonstrated that AFLP markers enable a quick and reliable assessment of intra-specific genetic variability in conservation genetics. The study showed that although the endangered plant occurred in small isolated populations, these populations contained a high genetic diversity, a good indication that recovery of the species was possible. Another endangered plant, *Metrosideros bartlettii*, was studied using AFLP markers. The AFLP study of these individuals showed which site had the highest level of genetic difference between individuals and selection of sites of conservation priority (Drummond et al., 2000). AFLPs have been used in a study of clonal diversity in a dwarf bamboo (*Sasa senanensis*) population (Suyama et al., 2000). A study of clonal diversity using AFLPs was also conducted in *Rhododendron ferrugineum* populations. AFLPs used in the study of *Sticherus flabellatus*, a largely asexual bracket fern, detected significant inter-population differentiation (Keiper et al., 2000).

Zabeau et al., (1993) developed AFLP in which fingerprinting patterns are obtained by detection of genomic restriction fragments by PCR amplification. AFLP is an ingenious combination of RFLP and PCR and is extremely useful in detection of polymorphism between closely related genotypes (Saiki et al., 1988; Ehrlich et al., 1991). AFLP is based on selective amplification of restriction enzyme digested DNA fragments with specific primers. Multiple bands are generated in each amplification reaction that contains DNA markers of random origin. In this technique the DNA is digested with one or two restriction enzymes

and double stranded adapters are ligated to the fragments to generate template DNA for amplification. Thus, the sequence of the adapters and the adjacent restriction site serve as primer binding sites for subsequent amplification of the restriction fragments by the PCR. These amplified samples are analysed on denaturing polyacrylamide gel which results in the production of 50 to 100 bands per individual sample. Polymorphism detected in DNA fingerprints, obtained by restriction cleavage, can result from alterations in the DNA sequence including mutations abolishing or creating a restriction site and insertions, deletions or inversions between two restriction sites. Similar to RAPD analysis, AFLP assay also does not require prior sequence knowledge but detects a 10 fold greater number of loci than those detected by RAPD analysis. Thus, the AFLPs have the capacity to rapidly screen thousands of independent genetic loci. The AFLP markers are typically inherited in Mendelian fashion and therefore, can be used for identification, typing and mapping of genetic loci.

1.7 Optimization of factors and ex situ conservation of *P. hexandrum*

An alternative approach to protecting the genetic diversity of *P. hexandrum* is to standardize the synthetic protocol for artificial synthesis of podophyllotoxin in the laboratories. However, the total synthesis of podophyllotoxin is complicated due to the presence of four chiral centers, a rigid *trans*-lactone and an axial 7-aryl substituent (Gordaliza et al., 2004). Hence, *P. hexandrum* and *P. peltatum* are presently the commercial source of podophyllotoxin for the pharmaceutical industry. The demand for the compound continues to increase and thus encourages ex situ conservation of *P. hexandrum* in the Himalayan region. In an attempt at ex situ conservation Sharma et al (2000), collected root samples of *P. hexandrum* from Jalori Pass and Khajjiar (from high altitude) and raised at Palampur (low altitude). This led to reduction in podophyllotoxin content in the plant samples. Conservation strategies adopted at Y.S. Parmar University of Agricultural Science, Nauni, Himachal Pradesh, India, further resulted in reduction in podophyllotoxin content. The increased rate of reduction in podophyllotoxin content might be attributed to the difference in soil nutrients and climatic factors in different habitats. This is obvious as the plant is very adaptable to a wide range of environmental conditions. It can survive under varying conditions and adapt well from the extreme low winter temperature of the Northern

climates to the high summer temperatures and at the altitudes ranging from 1300 to 4300 m. Therefore, planting of *P. hexandrum* plants at low altitude may not be useful because of the low amount of podophyllotoxin. A number of investigations have demonstrated that the quality and quantity of several secondary metabolites have a close relationship with plant habitats (Endress, 1994). Hence, it demands management of soil nutrients and optimization of climatic factors for the successful conservation of *P. hexandrum*. We are interested in whether the soil nutrition and climatic factors contribute to the high yield of podophyllotoxin in the natural habitats. No work has been done previously in these aspects nor has any prediction model been developed for selection of conservation area for successful domestication of *P. hexandrum.*

Statistical methods such as artificial neural network (ANN) and multiple linear regressions (MLR) are very useful in these respects. MLR has been used to explain the spatial variations in soil nutrients and its impact on crop yield at field scale (Sudduth et al., 1996). However, MLR requires a normal distribution of the input variables which is not always the case (Atkinson et al., 1997). The non-linear predictors such as ANNs were widely used to solve various problems in agriculture. For example, Sudduth et al., (1998) successfully predicted corn yield with back propagation neural network models based on soil texture, topography, pH and some soil nutrient elements. This prediction model was superior to those of the nonparametric statistical benchmark methods.

1.8 Podophyllotoxin and analogues design

The aryltetralin lactone podophyllotoxin occupies a unique position among lignan natural products since its glucopyranoside derivative was recognized as a potent antitumor factor (Jardine, 1980). This discovery entails a particularly fascinating account involving multitude investigations conducted over a period of more than a century (Stahelin and von Wartburg, 1991). The studies culminated in the structure elucidation of podophyllotoxin **1**, the assessment of its biological activity and the discovery of its mode of action. Initial expectations regarding the clinical utility of podophyllotoxin were tempered largely due to its unacceptable gastrointestinal toxicity. This led chemists in the pharmaceutical research department of Sandoz, to investigate the possibility that the *Podophyllum* lignans might

occur naturally as glycosides (Stahelin and von Wartburg, 1991). Using special procedures to inhibit enzymatic degradation, these researchers indeed obtained the podophyllotoxin-β-D-glucopyranoside **5** as the main component and its 4'-demethyl derivative **6** from the Indian *Podophyllum* species. Both of these glucosides and the glucosides **8** and **10** of α- and β-peltatin **7** and **9** (Figure 3) were also isolated from the American *P. peltatum*. Being less hydrophobic, the glucosides displayed lower toxicity than the aglucones but their cytostatic activity was reduced to the same degree. The research efforts were then focused on a program to chemically modify both the glucosides and aglucones of a wide range of podophyllotoxin derivatives. Nearly 600 derivatives were prepared and tested over a period of about 20 years (Stahelin and von Wartburg, 1991). Somehow serendipitously, a radical change in the mechanism of action and a quantum step in therapeutic utility were effected by acetalization of the 4- and 6-hydroxy groups of the glucopyranose moiety using aldehydes eventually leading to the discovery of the clinically important anticancer drugs etoposide **11**, etopophos **12** (Schacter, 1996) and teniposide **13** (Figure 1.3). However, the clinical application of podophyllotoxin and its analogues in the treatment of cancer has been limited by severe toxic side effects during administration of the drugs (Keller-Juslen et al., 1971; Weiss et al., 1975).

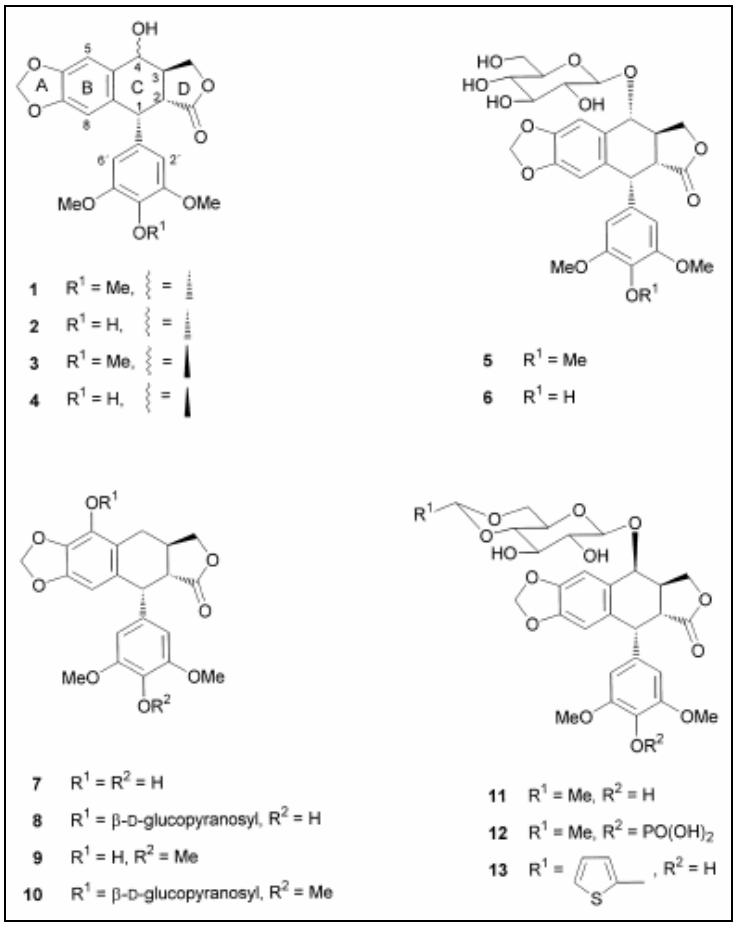


Figure 1.3. Structures of podophyllotoxin and its congeners and derivatives.

The main problem of most antineoplastic agents which substantially reduces their therapeutic usefulness lies in their scant selectivity because these substances affect cancer and normal cells alike and lead to the appearance of adverse side effects. Like most antineoplastic agents, podophyllotoxin, etoposide and teniposide do not show specificity against tumour cells; rather they affect all cells, especially those in the active phase of division. This accounts for the adverse effects of these drugs: anaemia, catharsis and alopecia among others (Buss and Waigh, 1995; Sengupta, 1995; Garth and Milles, 1991). Thus, the studies aimed at improving antineoplastic agents mainly focused on the search for more selective drugs.

1.9 Mechanisms of action of podophyllotoxin

Lignans inhibit the polymerization of tubulin and DNA topoisomerase II (Imbert 1998; Buss and Waigh, 1995). Studies on Structure-Activity Relationships (SAR) have

shown that podophyllotoxin like compounds preferentially inhibit tubulin polymerization which leads to arrest of the cell cycle in the metaphase (Margolis and Wilson, 1978). It is a competitive inhibitor of colchicine binding to tubulin (Cortese et al., 1977). Its affinity is double than that of colchicine. Furthermore, colchicine binds to tubulin almost irreversibly while podophyllotoxin derivatives do so reversibly which makes them less toxic and hence more useful in the field of cancer therapy. However, etoposide-like compounds are potent irreversible inhibitors of DNA topoisomerase II and their action is based on the formation of a nucleic acid-drug-enzyme complex which induces single and double stranded DNA breaks as the initial step in a series of biochemical transformations that eventually lead to cell death (Margolis and Wilson, 1978).

1.9.1 Computational strategy for virtual screening of potent lead molecules

It is widely appreciated that advancement in the biological component of drug development has catalyzed a shift in the strategies and tactics that underlie the drug discovery process. New information has evolved to describe disease states at the molecular rather than organism level which in turn presents those involved in drug development with a large array of well-defined targets. Additionally, economic factors are driving the need for a shorter lead-to-drug development time. A number of methodologies have evolved to integrate the higher degree of molecular information, number of new targets and need for efficiency. This integration has been most widely implemented in the coupling of high throughput screening (HTS) with high-output chemical synthesis (Hoever and Zbinden, 2004). HTS relies on the development of efficient and reliable assays to permit the evaluation of a large number of compounds against a target in a rapid and often automated manner. The large volume of HTS data is modeled in order to assess structure-activity relationships but problems arise when these models suffer from distortion by false positives. Combinatorial Chemistry, the synthesis of a very large number of compounds using a single scaffold and a diverse array of reactants, has also made an attempt to address the need for a large number of new drug leads (Lavastre et al., 2004). This methodology is severely hindered by the labor-intensive and cost effective measures required for its preparation and purification in such a large number of compounds. Virtual screening, using a computational approach to assess the interaction of an *in silico* library of small molecules and the structure

of a target macromolecule, has arisen as an alternative method for the rapid identification of new drug leads. A great deal of effort has been made to create reliable and efficient software that evaluates the highly complex enthalpic and entropic nature of the interaction between small molecules and their macromolecular receptors. A typical virtual library screening (VLS) approach involves several stages (Figure 1.4) including parallel efforts that involve small molecule and macromolecule preparation. Various stages of the VLS methodology described in Figure 1.4 have been previously reviewed (Lyne, 2002; Oprea and Matter, 2004; Shoichet et al., 2002) and provide an excellent resource for detailed analyses of many of the components of this process.

High throughput screening has been routinely used for the identification of small molecule drugs of a specific target. One of the most widely used virtual screening approaches in rational drug design is Quantitative Structure Activity Relationship (QSAR) (Selassie et al., 2002). The QSAR is based on statistical analysis of the relationship between certain biological activities of a chemical against quantitative attributes of the structure of the chemical. The resultant statistical model may be used to predict the activity of an unknown chemical by its quantitative attributes calculated from its structure. The molecular modeling techniques: molecular docking and rescoring using Prime MMGB/SA were widely used for virtual screening of inhibitors for energetically favorable interaction with receptor and the energy score has been used to build models for prediction of the inhibitory activity (pIC_{50}) (Podlipnika and Bernardib, 2007). Further, the newly developed structure-based linear interaction energy method implementing a surface generalized Born (SGB-LIE) (Zhou et al 2001) continuum model for solvation could be used to build a binding affinity model for estimating the free energy of binding a diverse set of inhibitors. The LIE method has been applied on a number of protein-ligand systems with promising results, (Tominaga and Jorgensen, 2004; Leiros et al., 2004; Ostrovsky et al., 2003) producing small errors in the order of 1 kcal/mol for free energy prediction.

By generating QSAR, Docking, MM-GB/SA and SGB-LIE models of a sufficient pool of potential analogues of podophyllotoxin, these virtual screening methods can be

extended to facilitate the search for the potential drugs with low toxicity and better biological activity against cancer.

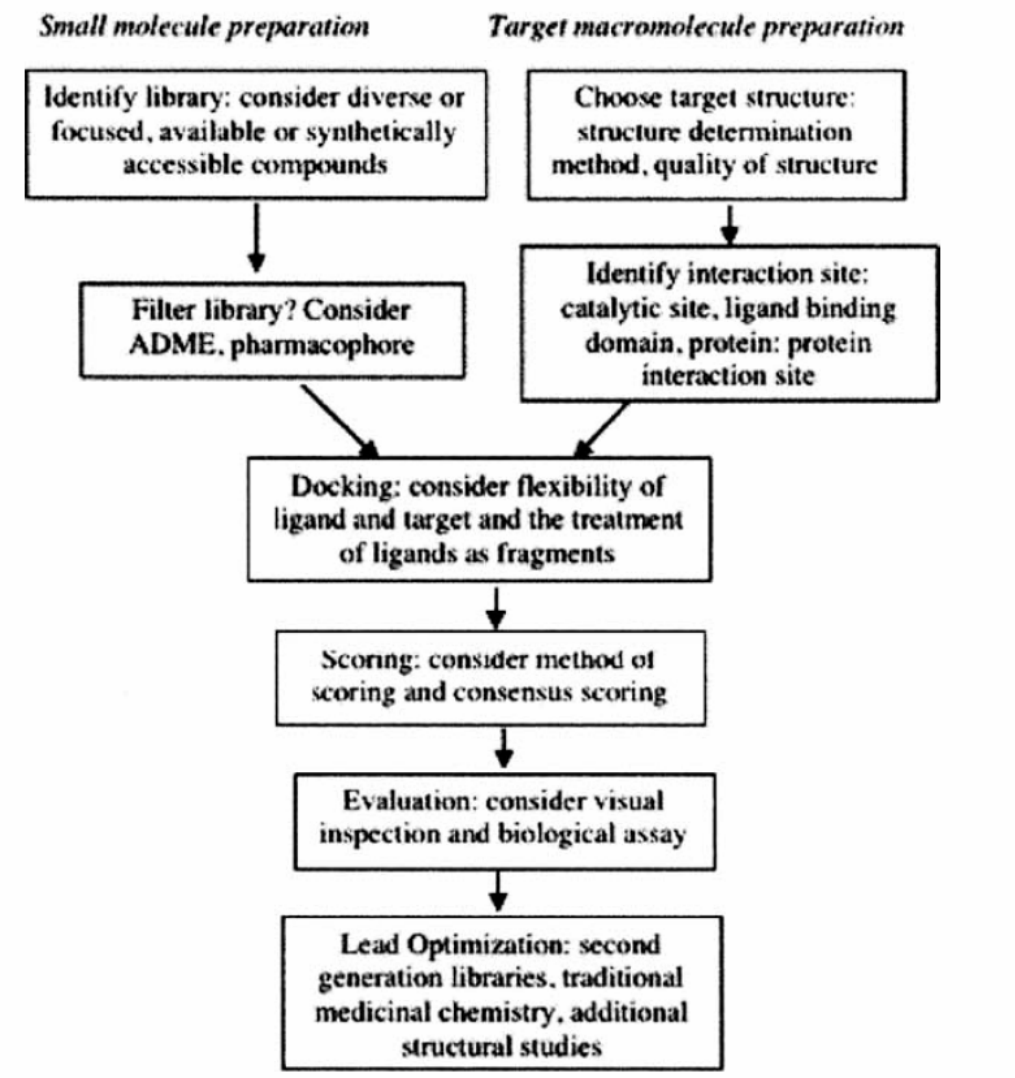


Figure 1.4. A typical virtual library scheme (VLS). Stages include both small molecule library preparation (choice of library, considerations for filtering) and target preparation (choice of structure of target and identification of binding site). In VLS, the library is docked into the target, scored and evaluated. Any possible leads are optimized in later stages.

1.9.2 Predicted models of structure activity relationship of podophyllotoxin

Efforts at improving their clinical efficacy further by overcoming the drug resistance, myelosuppresion and poor bioavailability problems (Witterland, 1996; Jardine, 1980) associated with podophyllotoxin continues to be challenging. Consequently, the number of analogues of podophyllotoxin has increased considerably. In this regard it is necessary to investigate and prepare new, more potent and less toxic analogues, that is, with better therapeutic indices. On this basis, studies on lignans have been addressed to modify the lactone moiety in the scaffold structure of podophyllotoxin and prepare analogues for better therapeutics and low toxicity. In many an experimental study during the past decade, a large number of podophyllotoxin analogues have been designed, synthesized and evaluated. These include modification of the A ring (Cho et al., 1996), B ring (Thurston et al., 1986), C ring (Beers et al., 1988; Laatsch et al., 1996), D ring (Wang et al., 1993) and E ring (Gordaliza et al., 2000) in the scaffold structure of podophyllotoxin. These studies provided important insights into development of the structure-activity relationship models between analogues. This has resulted in a paradigm shift in identifying such problems early in the drug discovery processes.

In silico approaches of preparation of combinatorial library and their computational screening, the rational approaches can be used for achieving the similar goal with less time and low cost with better therapeutic indices. Combinations of the computational methods such as docking with other methods, MD simulation, free binding energy calculation, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analyses (CoMSIA), ADME screening, predicting biological activity and toxicity analyses will provide a lot of insights into the biological system in rational drug designing. These computational techniques can be used for the study of the interaction of enzymes with active agents and screening of the lead drug candidates. This knowledge provides researchers a unique chance to learn the functioning of these enzymes in biological conditions and finally helps them in designing a suitable agent to affect the function of an enzyme. This will decrease the amount of experimental work required to synthesize and test a large number of possible candidates. With the help of a variety of computational methods, traditional drug development has greatly benefited from the computational research.

Expenses and the development period for a new drug have been reduced considerably as computational methods can dramatically decrease the number of candidates of a drug which need to be synthesized and tested.

Docking, modeling, molecular simulation, QSAR, virtual screening, free energy calculations and data mining methods have been used directly in rational drug discovery projects to speed up development and help find good agents. These methods produce a lot of information on a variety of drug related research. They benefit basic scientific activities as well as industrious efforts. But most of these computational tools have their own limitations and they need further development on some basic, methodological, and application problems. A lot of applications have demonstrated that if a proper approach is chosen on a specific research, good results can be produced to solve targeted problems (Titmuss et al., 1999; Barreca et al., 1999).

In this work we used computational methods to explore the binding structures, binding affinity and inhibition mechanism of active ligands in their corresponding receptors. Tubulin/podophyllotoxin bio-system was used in the work. By the application of computational techniques, we tried to learn more about the bio-system, especially the tubulin and its mechanism of inhibition that produces information for a researcher to understand their biological functions affected by inhibitors. Further, we tried to assess how a ligand affects its receptor and what type of ligand will better inhibit tubulin. Also we tried to develop approaches to calculate the activity of a set of ligands by applying the ways of free energy of binding (FEB) and quantitative structure-activity relationship (QSAR). These are convenient approaches which can be used to normal set of compounds to benefit ligand activity evaluations in a rational drug design.

In most QSAR approaches such as CoMFA and CoMSIA, the QSAR models are built on a variety of calculated properties of ligands alone (Klebe et al., 1994; Buolamwini et al., 2002). They are ligand-based computer-aided drug design approaches. On the other hand, in molecular modeling and docking, ligands are modeled into a given active site. The shape of an active site of a receptor is the key element used to design a new ligand. These

structure based methods focus on the steric and energetic fitting of ligand into a corresponding active site. In the work we combine ligand-based approach with receptor-based approach in building QSAR models by applying the conformational search and alignment from a flexible docking simulation into the PHASE (Schrodinger) method. The binding structure of a ligand in its receptor is important to understand the interaction between ligand and receptor. It is also a basis for many other studies, such as binding affinity calculation, new ligand improvement to fit a binding site better and MD simulation to explore the ligand effect on receptor dynamic properties. Docking is one of the ways to explore binding structures of a ligand in a receptor. In this work we used flexible docking simulation (Glide SP and XP) to predict binding structures of ligands in their receptors.

In this work we studied binding of podophyllotoxin analogues in tubulin: the binding structures and modes, the activity of structural specialties, interaction with tubulin and activation mechanism. Podophyllotoxin binds irreversibly to the tubulin heterodimer at a site near the interface between α- and β– tubulin (colchicine binding site) and thereby inhibits microtubule assembly (Cortese et al., 1977). Since the number of analogues of podophyllotoxin increased considerably in order to obtain better therapeutic use and low toxicity, the insights on the inhibition mechanism of podophyllotoxin will benefit the development of new, highly effective inhibitors of tubulin. Several classes of podophyllotoxin analogues have been proposed and tested. Learning the bound structures of these inhibitors in tubulin will help understand the interaction and binding affinity with tubulin. The activity model produced a way to evaluate the activity of interested inhibitors. Also the QSAR of a set of inhibitors was built based on the obtained binding structures. The knowledge on the inhibition mechanism will help researchers to design new inhibitors with minimum structural requirement which can effectively inhibit tubulin with minimum or nil toxicity. Crystal structure of tubulin/podophyllotoxin complex is available now. This provides us with an opportunity to explore the binding structures and binding affinity of other analogues. Docking and free energy of binding will be used to study podophyllotoxin analogues in the tubulin. These computations will provide structural and energetic information of these analogues in tubulin which will also be used to explain some

experimental results and their physicochemical properties. The information from this work will provide help to other researchers who try to develop new potent podophyllotoxin agents. Meanwhile, we try to develop an approach to calculate and evaluate the binding affinity of ligand in its receptor that is used to predict the activity of ligands in a reasonable computer requirement for a normal set of molecules. This approach will benefit activity calculation of a large set of ligands to a known structure receptor in computer-aided rational drug design.

Each of these eight pieces of work has distinct characteristics. At the same time they are related to one another. To clearly and coherently demonstrate the goal, results and conclusion of each piece of work, we have arranged each work chapter wise in a publishing format. The format will benefit readers to understand the idea of development, conclusion, coherence and full significance as each chapter will be a full manuscript for background to conclusion at publication stage. A final summary will link the seven parts together and give a general conclusion of the whole work.

CHAPTER 2

Assessment of genetic diversity among *Podophyllum hexandrum* genotypes of the North-western Himalayan region for podophylltoxin production using RAPDs molecular marker

---

Abstract

Podophyllum hexandrum (Indian Mayapple) is an important medicinal plant valued all over the world. Genetic diversity among 28 populations of *P. hexandrum* distributed in 11 geographical regions from Himachal Pradesh (a part of the northwestern Himalayan) was analyzed using RAPD markers for the purpose of conservation planning. The genetic diversity was high measured by percentage of polymorphic bands (PPB = 92.37%) and Shanon Information index (I = 0.5014) among the studied populations. The mean coefficient of gene differentiation (Gst) was 0.6933, indicating that 33.77% of the genetic diversity resided within the populations. Analysis of molecular variance indicated that the source of variation among groups was 53% and among populations within groups was 47%. An overall value of mean estimated number of gene flow (Nm = 0.110) indicated that there was limited gene flow among the populations. The existence of variation among 28 populations as observed through podophyllotoxin content was proved to be coupled with geographical altitude (R^2= 0.861) and local ecological conditions (temperature, rainfall, humidity, soil pH etc) but not on genetic diversity (R^2 = 0.050). Based on the observed genetic variations among the populations of *Podophyllum* study, we recommend for *in situ* conservation and germplasm collection expeditions in future conservation plans.

2. 1 Introduction

The quality and quantity of medicinal plants are serious issues for the pharmaceutical and dietary supplement industry. Traditionally this material has been harvested from the wild. Increasing public demand for these products is creating a serious environmental problem as demand is outpacing the supply and endangering the survival of many of these species in the world including *Podophyllum hexandrum*.

The Himalayan region is home to numerous highly valued medicinal plants including *Podophyllum hexandrum* Royle (Berberidaceae) which is a herbaceous, rhizomatous species of great medicinal importance, now endangered in India (Mayapple). It has long been used by the Himalayan natives and the American Indians (Anon et al., 1970). The rhizome of several *Podophyllum* species has been found to be the source of podophyllotoxin lignan that has important biological activity blocking mitosis (Loike et al., 1976a) and its use as the starting compound of the precursor of semi synthetic chemotherapeutic drugs such as etoposide, teniposide and etophos (Stahelin et al., 1991; Imbert, 1998). The Indian species *P. hexandrum* Royle contains three times more podophyllotoxin than its American counterpart *P. peltatum* (Fay et al., 1985). The rhizomes are being indiscriminately harvested in large quantity from the wild to meet the ever increasing demand for the crude drug. As a result of this and a lack of organized cultivation, *P. hexandrum* has been reported as a threatened species from the Himalayan region. The population size of *P. hexandrum* is very low (40-700 plants per location) and is declining every year. Some of the populations in certain pockets have virtually disappeared owing to anthropogenic activities and overexploitation (Bhadula et al., 1996). It is distributed in restricted pockets throughout the alpine Himalayan region (Nayar et al 1990). In the natural habitat, seed germination and seedling establishment are very poor and propagation is mostly through rhizomes. Since the species is already endangered and exploitation of its underground parts continues to exceed the rate of natural regeneration, it needs immediate attention for conservation. Studies of its population biology and genetic diversity are important for successful development of conservation strategies.

Considerable variations in morphological characters such as plant height, leaf characteristics, fruit weight, seed weight and color have been reported in *P. hexandrum* from the interior Himalayas (Bhadula et al., 1996; Airi et al., 1997; Purohit et al., 1998). At least four distinct morphological variants, i.e. with 1, 2, 3 and 4 leaves have been reported (Purohit et al., 1998). Podophyllotoxin content varies greatly between populations of *P. hexandrum* from the northwestern Himalayas (Sharma et al., 2000). Therefore, it would be inefficient, environmentally destructive and economically unsound to randomly harvest the May apple. For the rapid and economic development of *P. hexandrum* as an alternative crop, it would be essential to evaluate the natural populations of *P. hexandrum* for podophyllotoxin content and agronomic traits. For evaluation purposes, determination of relationship between and within populations cannot be made by just chemotyping the populations as the production of podophyllotoxin can be influenced by environmental factors like altitude, temperature, rainfall, relative humidity and soil pH (Alam et al., 2009). Polypeptide profiles and esterase isozyme analyses have indicated the existence of high inter and intra population variations in *P.* hexandrum (Bhadula et al., 1996). However, the protein markers are influenced by the stages of plant growth as well as environmental factors and hence may give erroneous results. DNA markers such as RFLPs (Botstein et al., 1980) and RAPD (Williams et al., 1990; Kazan et al., 1993), on the other hand are quite stable and highly polymorphic in nature. RFLP markers are less polymorphic, more expensive and laborious, compared with RAPD. Genetic diversity in *P. hexandrum* has been studied from specimens collected from two forest divisions of Himachal Pradesh (Chamba and Kullu) using RAPD markers (Sharma et al., 2000b). However, for the successful establishment of *P. hexandrum* as an alternative crop for the production of podophyllotoxin, genetic diversity of *Podophyllum hexandrum* needs to be understood for better identification and commercial level propagation of useful secondary compounds from the Himalayan region.

The aim of the present study of genetic diversity and population structure as well as their relationship with podophyllotoxin content in the diminishing wild populations of *P. hexandrum* in the northwestern Himalayas, Himachal Pradesh, India, is to ensure appropriate methods of conservation management for the remaining populations with great genetic variations.

2.2 Materials and Methods

2.2.1 Plant materials

Twenty-eight accessions of *P. hexandrum* were collected for RAPD studies and podophyllotoxin estimation from 28 sites covering 11 forest divisions in the interior range of the northwestern Himalayan region, Himachal Pradesh, India. *P. heaxandrum* populations from Himachal Pradesh exhibited variation in the number of leaves per plant and leaf shape etc (Figure 2.1). These results support the findings of morphological variants on the basis of the presence of 1, 2, 3 and 4 leaves from the Garhwal region of the western Himalayas (Bhadula et al 1996). We recorded morphological variations in 280 plants (Table-2.1). One representative plant from each site (considered as genotype) was used for molecular characterization. Young leaves were frozen in liquid nitrogen and stored at -80 ^{0}C prior to DNA isolation. The roots of the uprooted plants were trimmed and washed with running tap water to remove dirt soil followed by washing with double distilled water. The washed roots were then dried separately at 60 ^{0}C for 24 hrs in an oven and used for podophyllotoxin estimations.

2.2.2 Isolation of DNA

Total genomic DNA was extracted from frozen leaves by the CTAB method (Saghai-Maroof et al., 1984). Samples of 500 mg were ground to powder in liquid nitrogen using a mortar and a pestle. The powders were transferred to a 30 ml sterile Falcon tube with 12.5 ml of CTAB buffer. The extraction buffer consisted of 2% (w/v) CTAB (Cetyl trimethyl ammonium bromide, sigma), 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 9.5 and 0.2 % (V/V) β-mercaptoethanol. After incubating the homogenate for 1 hour at 65 ^{0}C an equal volume of chloroform was added and centrifuged at 10,000 rpm for 20 min. DNA was precipitated with 1/10 volume (ml) of 3 M sodium acetate and an equal volume of isopropanol followed by centrifugation at 10,000 rpm for 10 minutes. RNA was removed by RNase treatment. DNA was quantified by comparing with uncut λ DNA on the agarose gel diluted to 12.5ng.μl^{-1} and used in PCR.

Figure 2.1. Germplasm of *P. hexandrum* collected from different altitudes and their morphological variations (leaf shape).

Table 2.1. Twenty eight populations of *Podophyllum hexandrum* collected from different forest divisions, their podophyllotoxin content and polymorphic features.

Name of Forest Division	Sampling site	No. of leaves	Age (yr)	Altitude (m)	Podophyllotoxin content [% dry weight] (mean ± S.E.)	Total number of bands	No. of polymorphic bands	% of polymorphic bands (PPB)
Parvati	R/4 Kasol)	1-2	1	1570	3.567 ± 0.747	56	46	82.14
	Twin Multivora	3-4	3	1300	4.753 ± 0.796	57	47	82.45
	Anganoala	1-3	2	1300	3.020 ± 0.524	56	46	82.14
Kullu	Brundhar	1-3	2	1916	4.077 ± 0.270	79	69	87.34
	Gulaba	1-3	2	2895	5.943 ± 0.591	77	66	85.71
	Chander Khani	1-3	2	3352	8.033 ± 0.454	83	73	87.95
	Kaned Nry	1-2	1	2150	4.657 ± 0.850	80	70	87.50
	Sanghar Nry	1-3	2	2100	4.173 ± 0.276	80	70	87.50
Dodrakwar	Madhvi Thach	1-2	1	3048	6.207 ± 0.743	90	80	88.88
	Kala Pani	2-3	2	2743	5.800 ± 0.212	85	75	88.23
Seraj	Sojha Nry	2-3	2	2667	6.607 ± 0.348	97	87	89.69
	Jalora C-3(b)	2-3	2	2473	6.790 ± 0.855	90	80	95.55
Churah	DPF-D-1892-C1	1-3	2	3750	8.487 ± 0.565	67	57	85.07
	DPF-D-791-C1	3-1	2	2700	5.753 ± 0.411	68	58	85.29
Lahaul	Myar Valley	2-3	2	4300	9.533 ± 0.484	96	86	89.58
	Nayan ghar	2-4	3	4300	8.857 ± 0.427	88	78	88.63
Palampur	Bada Bangal	1-3	2	2895	7.097 ± 0.797	90	80	88.88
	Chota Bangal	2-3	2	2700	6.573 ± 0.827	83	73	87.95
	IHBT	2-3	2	2800	5.183 ± 0.780	81	71	87.65
Rampur	Bander Thach	1-2	1	2895	6.773 ± 0.640	88	78	88.63
	Saropa Nry	2-3	2	2499	6.097 ± 0.942	83	73	87.45
Kinnaur	Nichar Nry	1-3	2	2190	4.760 ± 0.291	91	81	89.01
	Rango(N-C-8)	1-3	2	2710	5.797 ± 0.552	83	73	87.95
Pangi	Sach Range	3-4	3	2712	6.133 ± 0.216	80	71	87.65
	Killer Range	2-4	3	2850	5.967 ± 0.692	78	68	87.17
	Purthi Range	2-3	2	2900	6.233 ± 0.790	83	73	87.95
Bharmaur	Ghoei DPF	1-3	2	2080	5.700 ± 0.692	83	73	87.95
	Samara RF	2-4	3	2590	6.030 ± 0.825	85	75	88.23

2.2.3 RAPD amplification

Out of twenty, nineteen random decamer primers (Operon Tech USA) A, B, C and D series were used individually for RAPD analysis. DNA was amplified following the protocol of Kazan et al., (1993). All 19 primers were previously reported in the study of Genetic Diversity among 16 promising cultivars of Ginger (Nayak et al., 2005). Amplification reactions were performed in volumes of 25 µl containing 10 mM Tris HCl, pH 9.0, 1.5 mM $MgCl_2$, 50 mM Tris HCl, pH 9.0, 1.5 mM $MgCl_2$, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 25 ng template DNA and 0.5 unit of Taq polymerase (Sigma). DNA amplification was performed using a Gene Cycler (BioRAD USA). The first cycle consisted of denaturation of template DNA at 94 ^{0}C for 5 min., primer annealing at 37 ^{0}C for 1 min. and primer extension at 72 ^{0}C for 2 min. In the next 42 cycles the period of denaturation was reduced to 1 min. at 92 ^{0}C while the primer annealing and primer extension time remained the same as in the first cycle. The last cycle consisted of only primer extension (72 ^{0}C) for 7 min.

The amplification products were resolved in 1.5% agarose gel (0.5 X TBE buffer) containing 40 mM Tris base, 20 mM sodium acetate, 20 mM EDTA and glacial acetic acid to pH 7.2 followed by ethidium bromide staining and visualization in UV light for photography which have been performed three times and were found reproducible (Sambrook et al., 1989).

2.2.4 Data Collection and analysis

The relationship among genomic DNA was assessed by comparing RAPD fragments separated according to their size and the presence/absence of shared fragments. The banding patterns obtained from RAPD were scored as present (1) or absent (0). Pairwise distance matrix was calculated using the Jaccard similarity coefficient (Sneath et al., 1973). The similarity values were used to generate a dendrogram via the unweighted pair group method with arithmetic average (UPGMA). POPGENE32 software was used to calculate Nei's unbiased genetic distance between the different populations using all RAPD markers inclusive of monomorphic markers. Nei's unbiased genetic distance is an accurate estimate of the number of gene differences per locus when populations are small. Population

diversity (Hs) and total gene diversity (Ht) (Nei, 1973) were calculated within 28 populations and within 11 major groups (as per their collection site) by POPGENE software. Genetic diversity within and among populations was measured by the percentage of polymorphic bands (PPB). Estimate of gene flow (Nm) was calculated by the gene differentiation (Gst) using (0.25(1-Gst)/Gst). In order to describe genetic variability among the populations, the non-parametric analysis of molecular variance (AMOVA) (Excoffier et al., 1992) program version 1.5 was used where the variation component was partitioned among populations, among populations within regions and among regions. The input files for AMOVA were prepared by using AMOVA-PREP (Miller, 1998), version-1.01.

2.2.5 Resolving power

The resolving power (Rp) (Prevost et al., 1999) of a primer was calculated using Rp = $\sum$ *IB*, where *IB* (band informative) takes the value of 1-[2x (0.5 – *P*)], *P* being the proportion of the 28 genotypes (*P. hexandrum* populations analyzed) that contains the band. The resolving power (Rp) of the 19 RAPD primers ranged from 2.0 for primer OPC15 to 16.571 for OPA18. Three RAPD primers (OPA02, OPA18, and OPD13) possess the highest Rp values (14.286, 16.571, and 13.786, respectively) and each one can able to distinguish all 28 *Podophyllum* cultivars.

2.2.6 Extraction and determination of podophyllotoxin

Dried roots were ground to powder in a mortar with a pestle. Podophyllotoxin was extracted following the established procedure of Broomhead et al., 1990. Fifty milligrams of root powder was suspended in 20 ml of ethanol and continuously stirred at 60 ^{0}C for 20 min. using a magnetic stirrer. The extract was filtered through Whatman filter paper No.1. The second, third and fourth extractions of the same samples were done with 10 ml ethanol for 10 min. under the same condition. All the extracts were pooled and ethanol was evaporated to dryness in a water bath shaker at 60 ^{0}C. The resultant residue was dissolved in 10 ml acetonitrile (HPLC grade) and filtered with 0.22 μm durapore membrane filter (Millipore) for HPLC analysis.

HPLC analysis was carried out using a Nova Pack C18 cartridge column (250 x 4.6 mm) in Water's HPLC system. Acetonitrile : water : methanol (37:58:5) was used as a

mobile phase with a flow rate of 1.0 ml. min^{-1}. Crude extract (20 µl) was used for injection into the HPLC system. Podophyllotoxin was detected at 230 nm (490 E multi wavelength Detector, Waters). Podophyllotoxin (0.1 $g.l^{-1}$; Sigma, P-4405) was used as a standard for calculating podophyllotoxin content in the samples on the basis of peak heights. All the experiments on extraction of podophyllotoxin and HPLC analyses were repeated three times.

2.3 Results and Discussion

2.3.1 Morphological markers

Podophyllum hexandrum populations collected from different forest divisions of Himachal Pradesh exhibited variation in number of leaves per plant, leaf shape etc. The number of leaves were seen to be 1, 2, 3 & 4 and varied among the plant samples of different age groups (Table 2.1). The results supported the findings of morphological variants on the basis of the presence of 1, 2, 3 and 4 leaves per plant from the Garhwal western Himalayas (Bhadula et al., 1996). The morphological variations in 280 plants from 28 sites covering 11 forest divisions were recorded. In general 39.5% plants had single leaf, 30% had two leaves, 20% had three leaves and 10.5% had four leaves. The single-leaf plant did not bear fruit.

2.3.2 RAPD marker size and Patterns

The RAPD technique has been successfully used in a variety of taxonomic and genetic diversity studies (Nayar et al., 1990) and it was found suitable for our use with *Podophyllum* populations because of its ability to generate reproducibly polymorphic markers. A total of 28 plant samples were fingerprinted using 19 RAPD makers. A total of 131 (an average of 6.89 bands per primer) RAPD loci was scored out of which 121 (92.37%) were polymorphic and only 10 (7.63%) were monomorphic bands (Table 2.2). The number of amplification fragments produced per primer as well as their size ranged from 250 bp to 3,100 bp (Figure 2.2) which is analytically appropriate and in conformity with those recorded with certain other plants examined analogously (Airi et al., 1997). The observed high proportion of polymorphic loci suggests that there is a high degree of genetic

Table 2.2. Nineteen primers used to amplify all DNA samples collected from 28 populations of *Podophyllum hexandrum* with the number of bands generated by each primer.

Primer	Nucleotide sequence	G+C content (%)	Total no. of loci	Polymorphic Loci	Monomorphic Loci	Percentage of polymorphic Loci	Total no. of bands amplified	Resolving power
OPA01.	5'CAGGCCCTTC3'	70	3	3	0	100	52	3.714
OPA02.	5'TGCCGACCTG3'	70	10	10	0	100	200	14.286
OPA04.	5'AATCGGGCTG3'	60	6	5	1	83.3	112	8.0
OPA08.	5'GTGACGTAGG3'	60	7	7	0	100	112	8.0
OPA11.	5'CAATCGCCGT3'	60	7	7	0	100	132	9.429
OPA13.	5'CAGCACCCAC3'	70	3	3	0	100	34	2.429
OPA18.	5'AGGTGACCTG3'	60	11	9	2	90	232	16.571
OPB11.	5'GTAGACCCGT3'	60	9	7	2	77.8	148	10.571
OPB15.	5'GGAGGGTGTT3'	60	6	3	3	50	138	9.857
OPB18.	5'CCACAGCAGT3'	60	9	9	0	100	148	10.571
OPB19.	5'ACCCCCGAAG3'	70	10	10	0	100	134	9.571
OPC08.	5'TGGACCGGTG3'	70	5	5	0	100	97	6.929
OPC12.	5'TGTCATCCCC3'	60	6	6	0	100	83	5.929
OPC15.	5'GACGGATCAG3'	60	4	4	0	100	28	2.0
OPC16.	5'CACACTCCAG3'	60	7	7	0	100	88	6.286
OPD05.	5'TGAGCGGACA3'	60	4	3	1	75	71	5.071
OPD08.	5'GTGTGCCCCA3'	70	8	7	1	87.5	132	9.429
OPD11.	5'AGCGCCATTG3'	60	7	7	0	100	123	8.786
OPD13.	5'GGGGTGACGA3'	70	9	9	0	100	193	13.786
Total			**131**	**121**	**10**		**2257**	

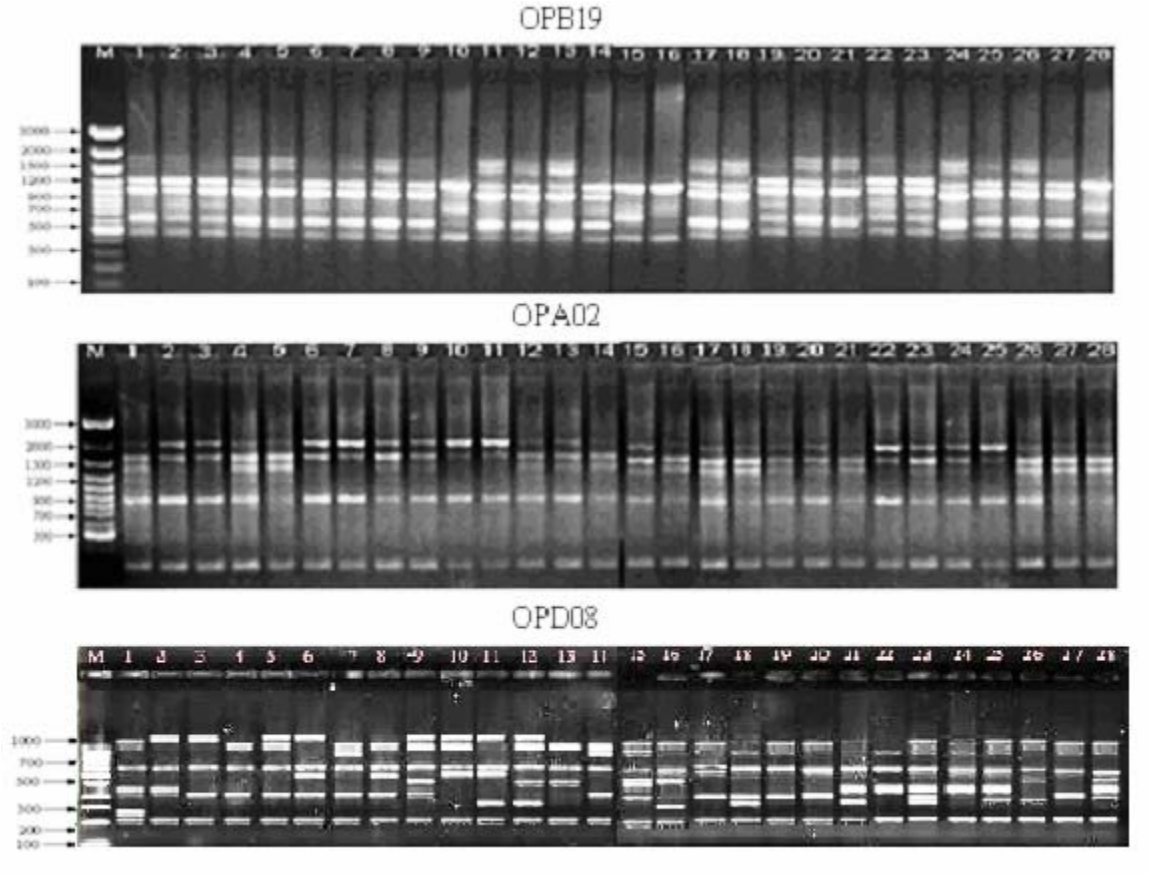


Figure 2.2. RAPD amplification products obtained from the 28 genotypes of *P. hexandrum* studied. 1. R/4, Kasol; 2. Twin Multivora; 3. Anganoala; 4. Brundhar; 5. Gulaba; 6. ChanderKhani; 7. Kaned Nursery; 8. Sanghar Nursery; 9. Madhvi Thach; 10. Kala Pani; 11. Jalora Pass (Sojha Nursery); 12. Jalora c-30(b); 13. DPF (D-1892-C1); 14. DPF (D-791-C1); 15. Myar Valley; 16. Nayan Ghar; 17. Bada Bhangal; 18. Chota Bangal; 19. IHBT; 20. Bander Thach; 21. Saropa Nursery; 22. Nichar Nursery; 23. Rango (NC-8); 24. Sach Range; 25. Killer Range; 26. Purthi Range; 27. Ghoei DPF; 28. Samara RF. M = the size of molecular markers in base pairs using λ DNA.

variations in the *Podophyllum* population. Three to ten types of amplification fragments (monomorphic + polymorphic) were produced by each primer on different populations. The resolving power of the 19 RAPD primers ranged from 2.0 for primer OPC15 to a maximum of 16.571 for primer OPA18. In addition to its high resolving power, RAPD primer OPA18 has the ability to distinguish all 28 *Podophyllum* populations.

2.3.3 Podophyllotoxin content

Podophyllotoxin content was extracted and analyzed in triplicate from 28 populations of *P. hexandrum* distributed into 11 forest divisions at different altitudes. It was found that the podophyllotoxin content in the root of the plants obtained from the Lahaul forest division (at an altitude of 4300 m) was high (8.857 to 9.533% on dry weight basis) compared to the root samples collected from other forest divisions. The lowest values obtained were samples from Parvati Forest Division (at an altitude of 1300 m) 3.020 to 4.753% on dry weight basis (Table 2.1). For populations in the same as well as other forest divisions, the podophyllotoxin content increased with the increase in altitude (Table 2.1). Podophyllotoxin content was correlated with altitude (Fig. 2.3a) and environmental variables like temperature, rainfall, relative humidity and soil pH but not with genetic diversity (Fig. 2.3b). The results have been supported by the study done by Sharma et al., 2000). They reported considerable reduction in the podophyllotoxin content in the roots of plants collected from higher altitude after growing them at lower altitude.

(a) (b)

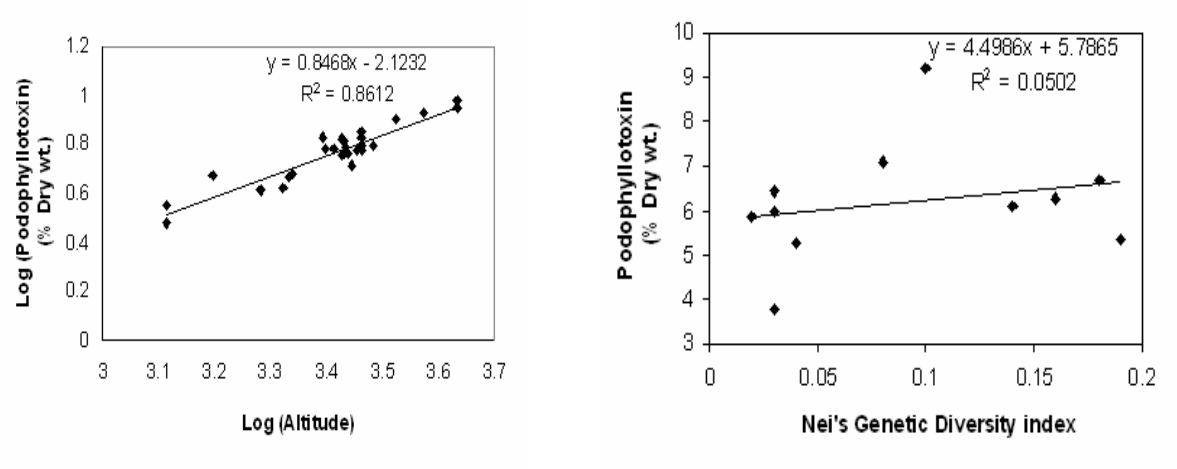


Figure 2.3. Regression analysis based on (a) Log_{10} Podophyllotoxin content and Log_{10} M (altitude) between 28 populations of *P. hexandrum* and (b) Podophyllotoxin content and Nei's genetic diversity index of the *P. hexandrum* populations among 11 regions, suggest altitude affects podophyllotoxin content, not genetics.

2.3.4 Phylogenetic analysis

Based on RAPD marker, the similarity index values ranged from 0.61 to 0.96. These values were used to construct a dendrogram using unweighted pair group method with arithmetic average (UPGMA). Populations from 11 forest divisions were clustered into region-specific groups with the exception of Kullu forest division (Fig. 2.4). All the 28 *Podophyllum hexandrum* populations were distributed into 12 main clusters (C1-C12). Cluster C1 represented Parvati forest division with 3 different populations namely R/4 Kasol, TwinMultivora and Anganola. Populations from Kullu forest division were distributed among two clusters- C2 & C3. The cluster C2 has populations from Brundhar, Gulaba and C3 has populations from Sanghar nursery, Kaned nursery and Chanderkhani.

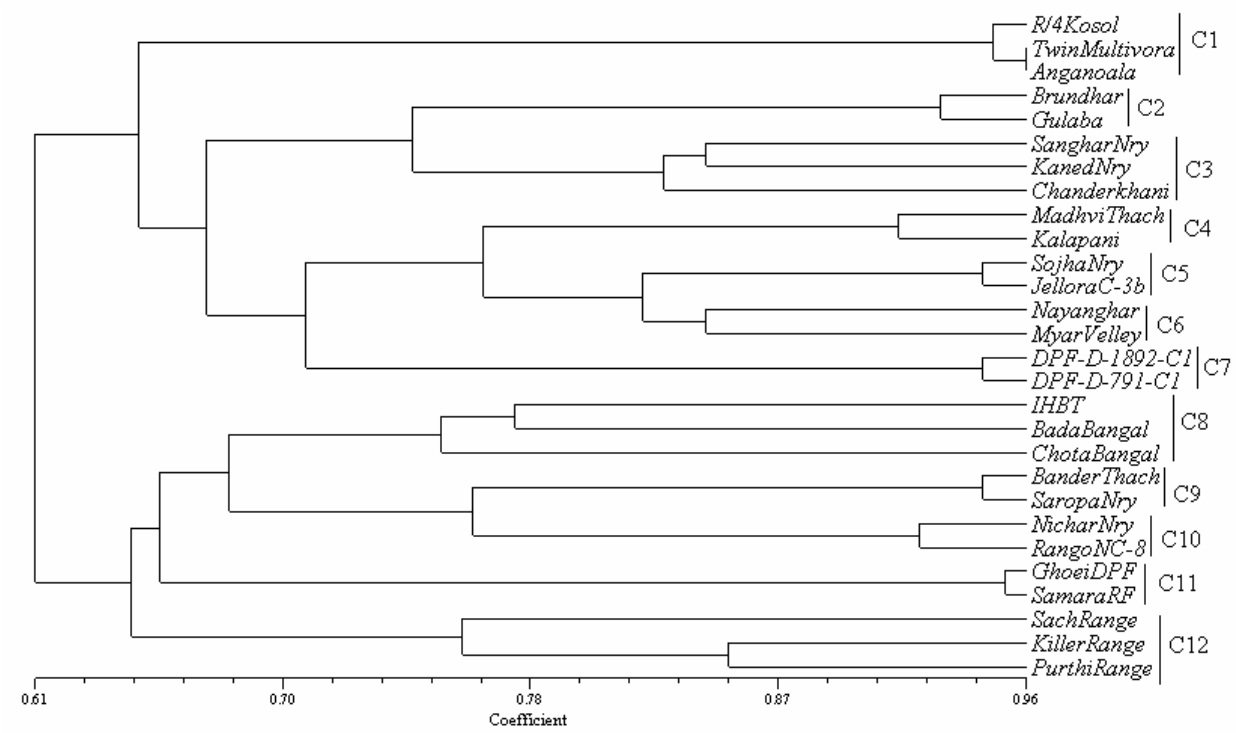


Figure 2.4. Dendrogram illustrating genetic relationships among 28 populations generated by UPGMA cluster analysis calculated from 2257 RAPD bands produced by 19 primers.

The cluster C4 (Dodrakwar forest division) has 2 populations- MadhviThach and Kalapani. Similarly cluster C5 and cluster C6 comprise 2 populations each from forest division Seraj (Sojha nursery, Jellora Pass) and Lahaul (Nayanghar, Myarvelley) respectively. The Churah division (cluster C7) has 2 populations- DPF-D-1892-C1 and DPF-D-791-C1. The cluster

C8 (Palampur forest division) has 3 populations- IHBT, Bada Bangal and Chota Bangal. The cluster C9 (Rampur division) has 2 populations- BanderThach and Saropa nursery. The cluster C10 representing Kinnaur forest division has 2 populations- Nichar nursery and Rango-NC-8. The cluster C11 (Bharmaur forest division) has 2 populations- Ghoei DPF and Samara RF. The cluster C12 (Pangi forest division) has 3 populations- Sach Range, Killer Range and Purthi Range. The results indicate high genetic diversity in *P. hexandrum* from Himachal Pradesh.

2.3.5 Genetic diversity analysis

A relatively high genetic variation was detected among the *Podophyllum* populations. The minimum value of genetic similarity is 0.4733 (genetic distance 0.7481) between Sach Range and Rajgiri Range and the maximum value of genetic similarity is 0.9618 (genetic distance 0.03889) between R/4 Kasol and Rajgiri. Wide genetic diversity between populations of *P. hexandrum* was evident from the high number of polymorphic markers. The minimum number of polymorphic markers varies between 46 (in R/4 Kasol and Rajgiri) and a maximum of 87 (in Nayanghar) (Table 2.1). The total number of polymorphic loci is 121, thereby, giving an estimate of profound (>92.37%) polymorphism. Polymorphisms also vary substantially within the discrete groups of plants and the minimum was found to be 20.0% (Kullu) whereas the maximum was 50.0% (Dodrakwar, Churah, Lahaul, Rampur, Kinnaur and Bharmaur). AMOVA analysis enabled a partitioning of the overall RAPD variations among groups (with respect to their regions) and among populations within groups (Table 2.3).

Table 2.3. Summary of analysis of molecular variance (AMOVA) based on RAPD genotypes of *Podophyllum hexandrum* (levels of significance are based on 1000 iteration steps, d.f.: degree of freedom; S.S.D.: sum of square deviation; P-value: probability of null distribution).

Source of variance	d.f	S.S.D.	Variance component	Percentage	P-value
Among forest division	10	429.43	12.64	53	< 0.001
Among populations within forest division	17	189.97	11.18	47	< 0.001
Total	**27**	**619.39**	**23.81**		

More than half of the total variations in the studied populations: 53% and 47% could be accounted for by difference among the forest divisions and between populations within a forest division, respectively. All components of molecular variations were significant ($P < 0.001$). The gene diversity computed among different groups of populations was recorded in between 0.00229 – 0.1918. The effective number of alleles was the minimum for Bharmaur (1.0458) and the maximum for Seraj (1.3588) across the populations collected from different forest divisions. The same order of genetic heterogeneity was discerned through Shannon's information index which varied from 0.0317 (Bharmaur) to 0.2779 (Kullu). On an all genotype basis, the observed number of alleles was 1.9237 and the effective number of alleles was found to be 1.5835 per locus. Similarly, the total gene diversity (Ht) among populations was 0.3377 and within populations (Hs) was 0.1036. Shannon's information index was 0.5014 and estimated gene flow was found to be 0.2212 among the 28 *P. hexandrum* populations (Table 2.4 & 2.5).

The present study and similar studies on ginger (Nayak et al., 2005), lotus (Campose et al., 1994), sweet potato (Cannoly et al., 1994) and *Andrographic paniculata* (Padmesh et al., 1999) suggest that RAPD is more appropriate for analysis of genetic variability in loosely related genotypes. Moreover, the RAPD markers used here were able to differentiate *P. hexandrum* populations collected from 11 forest divisions into 12 distinct region specific clusters. The study also indicates that *P. hexandrum* populations in the northwestern Himalayan region are genetically highly diverse. The high genetic variations in *P. hexandrum* may be attributed partially to the cross pollinated nature of *P. hexandrum*. AMOVA revealed that there was significant variation arising from habitat-correlated genetic difference (47%). It suggests that the effects of gene flow, genetic drift and local ecological conditions (altitude, temperature, rainfall, humidity, soil, pH etc) also play an important role in the variation of the genetic structure of *P. hexandrum* populations. Considering the high genetic differentiation among the wild populations of *P. hexandrum*, preservation of only a few populations may not adequately protect the genetic variation within the species in the Himalayan region. At present, the rate of propagation of *P. hexandrum* is far less than the rate of its exploitation. This species or at least a large part of its genetic diversity may be lost in the near future owing to its importance and over exploitation as a medicinal plant if

Table 2.4. Summary of genetic variation statistics for all loci of RAPD among the *Podophyllum hexandrum* populations with respect to their distributions among eleven forest divisions.

Sampling sites	Sample size	Observed no. of alleles (mean ± SD)	Effective no. of alleles (mean ± SD)	Nei's gene diversity (mean ± SD)	Shannon's information index (mean ± SD)	Ht (mean ± SD)	Average number of polymorphic bands	Percentage of polymorphic bands (PPB)
Parvati	3	1.07 (0.25)	1.06 (0.20)	0.03 (0.11)	0.04 (0.16)	0.03 (0.01)	46.33	33.33
Kullu	5	1.46 (0.50)	1.34 (0.40)	0.19 (0.22)	0.28 (0.31)	0.19 (0.05)	71.80	20.00
Dodrakwar	2	1.07 (0.25)	1.07 (0.25)	0.03 (0.13)	0.05 (0.18)	0.03 (0.02)	67.50	50.00
Seraj	2	1.36 (0.48)	1.36 (0.48)	0.18 (0.24)	0.25 (0.33)	0.18 (0.06)	76.50	49.67
Churah	2	1.15 (0.36)	1.15 (0.36)	0.08 (0.18)	0.11 (0.25)	0.08 (0.03)	70.00	50.00
Lahual	2	1.20 (0.40)	1.20 (0.40)	0.10 (0.20)	0.14 (0.28)	0.102 (0.04)	81.00	50.00
Palampur	3	1.36 (0.48)	1.29 (0.39)	0.16 (0.21)	0.23 (0.31)	0.16 (0.05)	74.66	33.33
Rampur	2	1.05 (0.23)	1.05 (0.23)	0.03 (0.11)	0.04 (0.16)	0.03 (0.01)	75.50	50.00
Kinnaur	2	1.08 (0.27)	1.08 (0.27)	0.04 (0.13)	0.05 (0.18)	0.04 (0.02)	77.00	50.00
Pangi	3	1.31 (0.47)	1.25 (0.37)	0.14 (0.21)	0.20 (0.30)	0.14 (0.04)	70.66	33.33
Bharmaur	2	1.05 (0.21)	1.05 (0.21)	0.02 (0.11)	0.03 (0.15)	0.02 (0.01)	74.00	50.00

Table-2.5. Genetic variability across all the populations of *Podophyllum hexandrum.*

Observed no. of alleles	Effective no. of alleles	Nei's gene diversity	Shannon's Information index	Ht	Hs	Gst	Estimate of gene flow (0.5(1-Gst)/Gst	Number of polymorphic alleles	Percent of polymorphic alleles
1.94 (0.27)	1.58 (0.32)	0.34 (0.15)	0.50 (0.20)	0.34 (0.02)	0.10 (0.01)	0.69	0.22	121	92.37

appropriate conservation measures are not adopted. Since a single or even a few plants will not represent the whole genetic variability in *P. hexandrum,* there appears to be a need for maintenance of sufficiently large populations in natural habitats for conservation of its genetic diversity and avoidance of genetic erosion.

CHAPTER 3

Characterization of genetic structure of *Podophyllum hexandrum* populations using ISSR-PCR markers and its relationship with podophyllotoxin content.

Abstract

To obtain accurate estimates of genetic structure for the purpose of conservation planning for wild Indian Mayapple (*Podophyllum hexandrum*) in the Northwestern region of Himalayas, Himachal Pradesh, genetic diversity among and within 28 populations were analyzed. Eleven microsatellite DNA markers were used to quantify the genetic structure. Out of 68 ISSR loci tested, 88.23 % were polymorphic. The genetic diversity was found to be high (percentage of polymorphic bands, PPB = 83.82 %; Shannon's information index, I = 0.441) at the population level but low within individual studied populations (PPB = 15.05%; Shannon's information index I = 0.0989). The mean coefficient of gene differentiation (Gst) was 0.6296, indicating that 29.44 % of the genetic diversity resided within the population. Analysis of molecular variance (AMOVA) indicated that 62% of the genetic diversity among the studied populations was attributed to geographical location while 38% was attributed to differences in their habitats. An overall value of mean estimated number of gene flow (Nm = 0.147) indicated that there was limited gene flow among the sampled populations. The existing variation in podophyllotoxin content among the populations was proved to be coupled with geographical altitude (R^2= 0.861) but not with genetic variations (R^2 = 0.146). Hence, we conclude that any further cultivation of this species demands optimization of environmental factors in order to increase the rate of podophyllotoxin production from collected populations.

3.1 Introduction

Over the years, the ability of a population to respond adaptively to environmental changes depends on the level of genetic variability or diversity it contains (Ayala & Kiger 1984). During the process of evolution, genetic differentiation by natural selection to facilitate reproductive isolation involves the presupposition of the origin of geographic races, subspecies and species (Stebbins, 1999). A species without enough genetic diversity is thought to be unable to cope with changing environments or evolving competitors and parasites (Schaal et al., 1991). Therefore, studies of genetic diversity and the structure of population within a species may not only illustrate the evolutionary process and mechanism but also can provide information useful for biological conservation of *P. hexandrum* (Berberidaceae), an endangered medicinal herb that grows in the wild in the interior Himalayan ranges of India.

The Himalayan region is home to numerous highly valued medicinal plants including *Podophyllum hexandrum* Royle (Indian Mayapple). It is recognized for its anticancer properties. The rhizomes and roots of *P. hexandrum* contain anti-tumor lignans such as podophyllotoxin, 4'-demethyl podophyllotoxin and podophyllotoxin 4-O-glucoside (Tyler et al., 1988; Broomhead & Dewick, 1990). Among these lignans, podophyllotoxin is the most important for its use in the synthesis of anti-cancer drugs etoposide, teniposide and etophos (Issell et al., 1984). These compounds have been used for the treatment of lung and testicular cancers as well as certain leukemia (Stahelin and Wartburg, 1991; Imbert, 1998). In addition, podophyllotoxin is also the precursor to a new derivative CPH 82 that is being tested for rheumatoid arthritis and other derivatives for the treatment of psoriasis and malaria (Leander and Rosen, 1988; Lerndal & Svensson, 2000). Total synthesis of podophyllotoxin is an expensive process and availability of the compound from natural resources is an important issue for pharmaceutical companies that manufacture these drugs (Canel et al., 2001). Podophyllotoxin content in the Himalayan may- apple is high (4.3 %) compared with other species of *Podophyllum*, notably *P. peltatum* (0.25%), the most common species in the American subcontinent (Jackson and Dewick, 1984).

The population size of *P. hexandrum* in the Himalayan region is very low (40-700 plants per location) and is declining every year. Some of the populations in certain pockets have virtually disappeared owing to anthropogenic activities and overexploitation (Bhadula et al., 1996). Thus, there is a need to conserve genetic diversity of this prized medicinal plant which may become extinct if its reckless exploitation continues. In dwindling populations the size of the surviving population greatly affects genetic diversity (Gao, 2005). Therefore, arresting the decline of *P. hexandrum* population in the wild and studying the genetic structure of the remaining populations is of critical importance. Traditionally, for commercial purposes and germplasm conservation, a large number of *P. hexandrum* populations are collected from different sites and cultured in nurseries. This method of operation leads to gene flow from the wild to the introduced populations. Considerable variation in morphological characters such as plant height, leaf characteristics, fruit weight, seed weight, color, biochemical characters such as podophylloresin and podophyllotoxin content in rhizomes have been reported in *P. hexandrum* plants from the Garhwal Himalayas (Bhadula et al., 1996; Airi et al., 1997; Purohit et al., 1999). This as well as the RAPD study of *P. hexandrum* indicates the existence of high inter- and intra-population variations (Sharma et al., 2000). However, this study is restricted to only two geographical locations (two districts of Himachal Pradesh, India). Moreover, the impact of geographical distance/altitude with the genetic variation as well as its relationship with the podophyllotoxin content of *P. hexandrum* populations have not been reported so far which we feel is very important for conservation. This needs the study of genetic variation of *P. hexandrum* populations with wide geographical coverage. To our knowledge, no report has been made on the genetic diversity, population structure and gene flow among the populations of *P. hexandrum* in the Himalayan region with high resolution molecular markers like ISSR.

Inter Simple Sequence Repeats (ISSR) amplifies inter-microsatellite sequences at multiple loci throughout the genome (Salimath et al., 1995; Li & Xia, 2005). An ISSR molecular marker technique permits the detection of polymorphism in microsatellites and inter-micorsatellite loci without previous knowledge of DNA sequences (Zietkiewicz et al., 1994). Furthermore, they are highly reproducible due to their primer length and high

stringency achieved by the annealing temperature. This technique has been widely used to investigate genetic diversity (Reddy and Nagaraju, 1999; Li and Xia, 2005; Chen et al., 2005) because of its advantages in overcoming the limitations of allozyme and RAPD techniques (Wolfe et al., 1998; Ratnaparkhe et al., 1998; Esselman et al., 1999).

In this study, we investigated the genetic diversity and population structure as well as their relationship with podophyllotoxin content in the diminishing wild populations of *P. hexandrum* in the northwestern Himalays, Himachal Pradesh, India, with the aim to provide insight to facilitate conservation management of the remaining populations. Appropriate conservation management should be adopted including *in situ* conservation and germplasm collection from the remaining populations with the greatest genetic variation.

3.2 Materials and Methods

3.2.1 Plant materials

The plant material used in the study of genetic diversity and population structure was obtained from 28 populations (28 sites), covering 11 geographical locations (Forest Divisions) with altitudes ranging from 1300–4300m in the northwestern Himalayan region, Himachal Pradesh, India (Table-3.1). Each population consists of about 7-8 plants with different age groups (1^{st}, 2^{nd}, 3^{rd} and 4^{th} year). Only 3^{rd} year plants were used for extraction of podophyllotoxin. The interval between samples was 2-5m. The pair wise distance between populations and between forest divisions was 0.5 – 32Km and 10 – 400Km respectively. The morphological feature of each plant sample as well as the environmental factors for each sampling site was documented. The variations in environmental factors among different sites were represented in Figure 3.1.

Table 3.1. Twenty eight populations of *Podophyllum hexandrum* collected from different sites at different altitudes covering eleven forest divisions and their podophyllotoxin content.

Forest Division	Sampling site	Altitude (m)	Podophyllotoxin content [% dry weight] (mean ± S.E.)
Parvati	R/4 Kasol	1570	3.567 ± 0.747
	Twin Multivora	1300	4.753 ± 0.796
	Anganoala	1300	3.020 ± 0.524
Kullu	Brundhar	1916	4.077 ± 0.270
	Gulaba	2895	5.943 ± 0.591
	ChanderKhani	3352	8.033 ± 0.454
	Kaned Nry	2150	4.657 ± 0.850
	Sanghar Nry	2100	4.173 ± 0.276
Dodrakwar	Madhvi Thach	3048	6.207 ± 0.743
	Kala Pani	2743	5.800 ± 0.212
Seraj	Sojha Nry	2667	6.607 ± 0.348
	Jalora-C-3b	2473	6.790 ± 0.855
Churah	DPF-D-1892-C1	3750	8.487 ± 0.565
	DPF-D-791-C1	2700	5.753 ± 0.411
Lahaul	Myar Valley	4300	9.533 ± 0.484
	Nayan Ghar	4300	8.857 ± 0.427
Palampur	Bada Bangal	2895	7.097 ± 0.797
	Chota Bangal	2700	6.573 ± 0.827
	IHBT	2800	5.183 ± 0.780
Rampur	Bander Thach	2895	6.773 ± 0.640
	Saropa Nry	2499	6.097 ± 0.942
Kinnaur	Nichar Nry	2190	4.760 ± 0.291
	Rango-NC-8	2710	5.797 ± 0.552
Pangi	Sach Range	2712	6.133 ± 0.216
	Killer Range	2850	5.967 ± 0.692
	Purthi Range	2900	6.233 ± 0.790
Bharmaur	Ghoei DPF	2080	5.700 ± 0.692
	Samara RF	2590	6.030 ± 0.825

*Podophyllotoxin content varied significantly among populations, $F=17.22$, $P< 0.001$; as well as between regions, $F = 3.70$, $P < 0.009$.

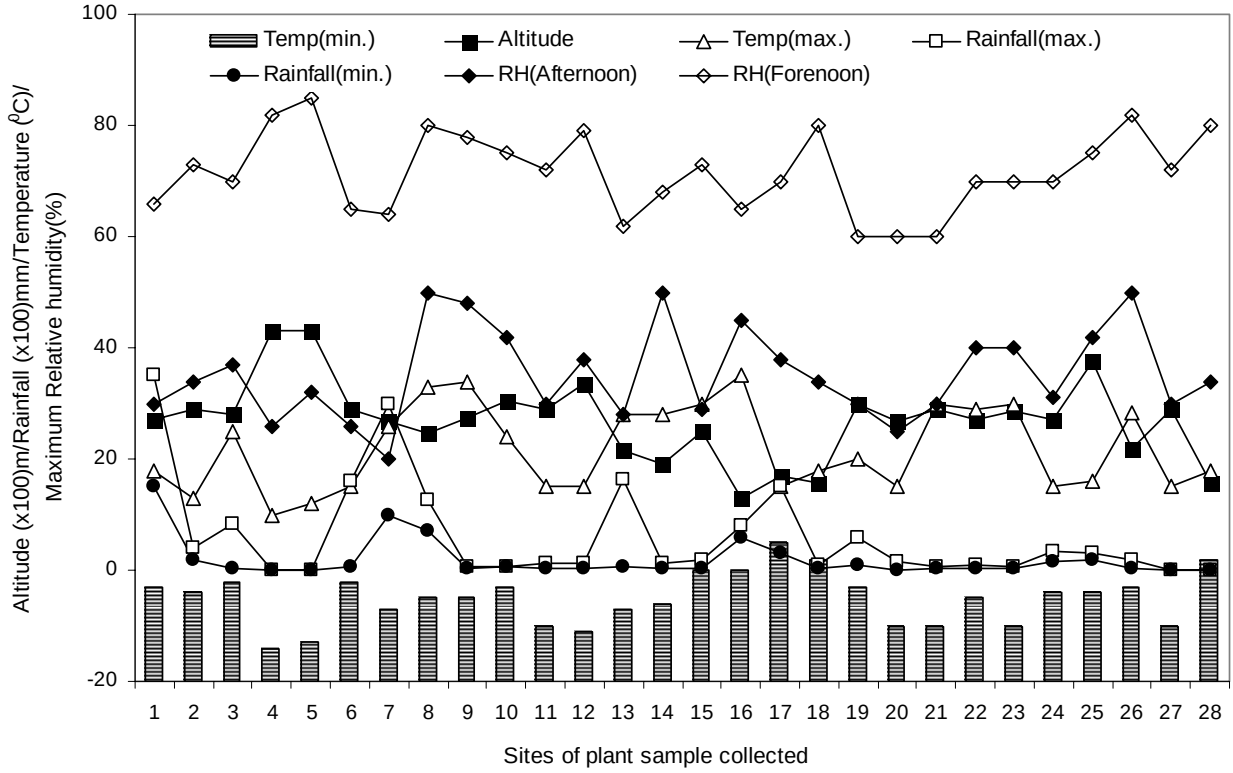


Figure 3.1. The meteorological observation during the season of June to July, 2006 (harvesting period of plant samples). The collection sites are in serial order as mentioned in Table 3.1.

About 5g of young leaves from each representative plant samples were obtained and placed in a zip-lock plastic bag containing silica gel which speeded up the drying process. Each sample does not necessarily denote a separate genetic individual. The samples were stored at -80 ^{0}C until use.

3.2.2 Extraction and quantification of podophyllotoxin

Ethanolic extract of podophyllotoxin following the procedure of Broomhead et al., (1990) was measured using HPLC analysis as described in chapter 2. All the experiments on extraction of podophyllotoxin and HPLC analysis were repeated three times.

3.2.3 Genomic DNA extraction

Genomic DNA of every representative plant samples was isolated using CTAB method (Saghai-Maroof et al., 1984). DNA was quantified by comparison with known concentration lambda DNA following electrophoresis on a 1% agarose gel.

3.2.3 Evaluation of primers

A total of 30 ISSR primers were screened with 10 plant samples. After assessing the effects of Mg^{+2} concentration, template DNA concentration and temperature during the annealing stage of the amplification, 11 primers which produced clear and reproducible fragments were selected for further analysis. The sequences of these ISSR primers listed in Table 3.2 were commercially synthesized by Sigma Inc.

3.2.4 ISSR amplification

The PCR amplification was performed in a 25 µl reaction volume containing 100 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$, 10 mM each of DNTP, 0.4 µM of primer,0.01% gelatin,1 unit of Taq polymerase and 25 ng of genomic DNA. Initial denaturation for 5 min at 94 ^{0}C was followed by 40 cycles of 1 minute at 94 ^{0}C, 1 min. at specific annealing temperature, 2 min. at 72 ^{0}C and a 10 min. final extension step at 72 ^{0}C. The annealing temperature for each primer is mentioned in Table 3.2. Amplification products were electrophoresed on 2.0 % agarose gels run at constant voltage and 1X TBE for approximately 2 hrs, visualized by staining with ethidium bromide and photographed under ultraviolet light (using Gel Doc, Biorad). Molecular weights were estimated using DNA markers (Sigma). Gel-Pro analyzer version 3-1 software was used to score ISSR profile.

3.2.5 Resolving power

According to Prevost & Wilkinson (1999), the resolving power (Rp) of a primer is calculated by $Rp = \sum IB$, where *IB* (band informativeness) takes the value of 1-[2x (0.5 – *P*)], *P* being the proportion of the 28 *Podophyllum* population containing the band.

3.2.6 Data analysis

ISSR amplified fragments were scored for band presence (1) or absence (0) and a binary qualitative data matrix was constructed. Data analyses were performed using the NTSYS pc version 2.0 computer package program (Rohlf, 1992). Pair wise distance matrix was calculated using the Jaccard similarity coefficient (Sneath and Sokal, 1973) and was used to generate a dendrogram via the un-weighted pair group method with arithmetic average (UPGMA). Mantel tests (Mantel, 1967) were performed using Arlequin 3.11 to analyze the effects of geographical distance on genetic variation. Regression analysis was done to study the impact of altitude on genetic variation and podophyllotoxin content using MINITAB statistical package. The genetic diversity within and among populations was measured by the percentage of polymorphic bands (PPB). The data matrix of ISSR was also used for assessment of genetic structure, genetic differentiation, gene flow and diversity. Nei's (1978) genetic distances (D) between different geographical populations were calculated using ARELQUIN 3.11. Measurement of diversity including gene diversity (H), observed number of alleles (Ne), gene flow and Shannon's information index (I) were estimated by POPGEN 1.32 software. In order to describe genetic structure and variability among the populations, the non-parametric analysis of molecular variance (AMOVA) program version 1.5 was used (Excoffier et al., 1992) where the variation component was partitioned among individuals within populations, among populations within regions and among regions. The input files for AMOVA were prepared by using AMOVA-PREP version-1.01 (Miller, 1998).

Table 3.2. List of primers used for ISSR amplification, GC content, annealing temperature (Tm), total number of loci, the level of polymorphism, size range of fragments and resolving power (Y = C, T; R = A,G).

Primer	Primer Sequence (5'~3')	GC (%)	Annealing Temperature (^{0}C)	Total number of loci	Number of polymorphic loci	Percentage of polymorphic loci	Total number of fragments amplified	Resolving power
P02	5'AGAGAGAGAGAGAGAGT3'	47.06	45	5	1	20	131	9.357
P08	5'TGTGTGTGTGTGTGTGA3'	47.06	55	3	2	66.7	57	4.071
P10	5'AGAGAGAGAGAGAGAGYT3'	44.44	45	3	2	66.7	63	4.5
P 13	5'CTCTCTCTCTCTCTCTRA3'	44.44	45	5	4	80	80	5.714
P 16	5'CCGCCGCCGCCGCCGCCG3'	100.00	50	5	4	80	65	4.642
P 17	5'GGCGGCGGCGGCGGCGGC3'	100.00	50	8	8	100	146	10.428
P 21	5'CTTCACTTCACTTCA3'	40.00	45	7	6	85.7	155	11.071
P 22	5'TAGATCTGATATCTGAATTCCC3'	36.36	55	12	12	100	177	12.642
P 23	5'AGAGTTGGTAGCTCTTGATC 3'	45.00	55	6	6	100	133	9.50
P 24	5'CATGGTGTTGGTCATTGTTCCA3'	45.45	50	5	3	60	98	7.0
P 25	5'ACTTCCCCACAGGTTAACACA3'	47.62	50	9	9	100	198	14.142
	Total			**68**	57		**1303**	

3.3 Results

3.3.1 Podophyllotoxin content

Podophyllotoxin from rhizomes was extracted and analyzed in triplicate from 28 populations of *P. hexandrum* distributed into 11 forest divisions at different altitudes. It was found that the podophyllotoxin content in the root of the plants obtained from the Lahaul forest division was comparatively more (8.857 to 9.533 % on dry weight basis) than that of the root samples collected from other forest divisions with the minimum from Parvati (3.020 to 4.753 % on dry weight basis) (Table 3.1). For populations in the same forest division as well as among the forest divisions, the podophyllotoxin content increased with the increase in altitude (Table 3.1).

3.3.2 ISSR polymorphism

A total of 68 ISSR loci were detected. Out of the 68 loci surveyed, 60 were polymorphic (83.82%). Out of the 11 primers used in the study; P02, P08, P13 and P16 revealed 8 monomorphic loci existed in all of the 28 populations. The high reproducibility of ISSR markers may be due to the use of longer primers and higher annealing temperature than those used for RAPD. The annealing temperature in this study ranged from 45 ^{0}C to 55 ^{0}C. The primers used for ISSR amplification, GC content, annealing temperature (Tm), total number of loci, the level of polymorphism; resolving power and size of fragments were shown in Table 3.2. The resolving power of the 11 primers used in the study ranged from 4.071 for primer P08 to a maximum of 14.142 for primer P25. Figure 3.2 shows the amplification profiles obtained with four of the primers (P21, P27, P25 and P22) those are used to amplify segments of genomic DNA from all 28 populations. Most of the ISSR products were in the range of 250bp to 1350bp.

3.3.3 Genetic diversity and differentiation

Among the 28 populations, the mean coefficient of gene differentiation (Gst) was 0.6296, indicating 29.40% of the total genetic diversity within the populations. Based on the Gst value, the mean estimated number of gene flow (Nm) between populations was found to be 0.147 (Table 3.3)

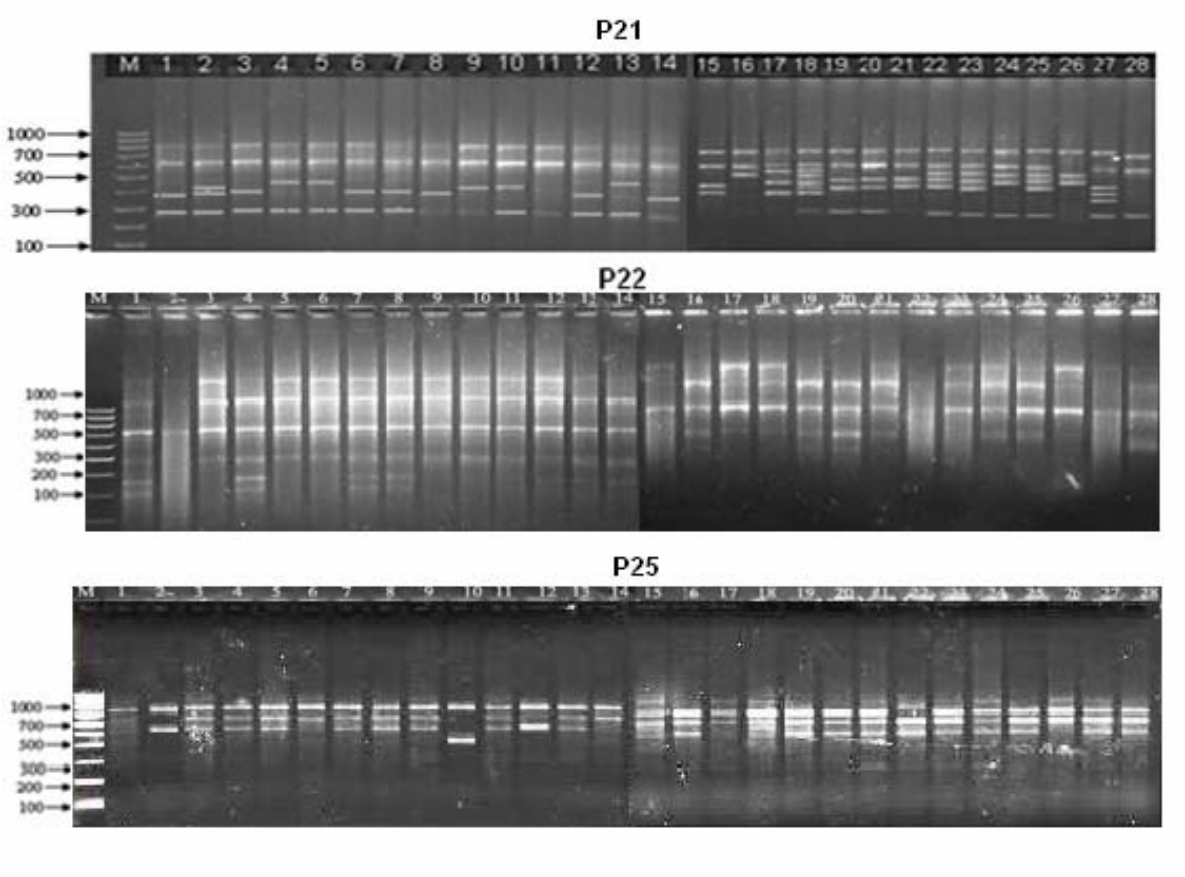


Figure 3.2. ISSR amplification products obtained from the 28 populations of *Podophyllum hexandrum*: 1. R/4 Kasol; 2. Twin Multivora; 3. Anganoala; 4. Brundhar; 5. Gulaba; 6. Chanderkhani; 7. Kaned Nursery; 8. Sanghar Nursery; 9. Madhvi Thach; 10. Kala Pani; 11. Sojha Nursery; 12. Jalora c-30(b); 13. DPF (D-1892-C1); 14. DPF (D-791-C1); 15. Myar Valley; 16. Nayangarh; 17. Bada Bangal; 18. Chota Bangal; 19. IHBT; 20. Bander Thach; 21. Saropa Nursery; 22. Nichar Nursery; 23. Rango (N-C-8); 24. Sach Range; 25. Killer Range; 26. Purthi Range; 27. Ghoei DPF; 28. Samara RF; M = M = the size of molecular markers in base pairs using λ DNA.

The study also revealed significant genetic variation among the populations distributed region wise. An average of 15.05% ranging from 5.88% to 27.94% of the loci obtained were polymorphic (Table 3.4). The mean observed number of alleles (Na) varied form 1.0588 ± 0.2370 (Kinnaur division) to 1.2794 ± 0.4520 (Kullu division) whereas the effective number of alleles (Ne) were less than those of Na for every forest division and ranged from 1.0588 ± 0.2370 (Kinnaur) to 1.2206 ± 0.4177 (Seraj).

Table 3.3. Overall genetic variability across all the populations of *Podophyllum hexandrum*

Sample size	Observed number of alleles	Effective number of alleles	Nei's gene diversity	Shannon's information index	Ht	Hs	Gst	Estimate of gene flow	Percentage of polymorphic loci (%)
28	1.838 (0.371)	1.497 (0.337)	0.294 (0.173)	0.441 (0.239)	0.294 (0.030)	0.109 (0.007)	0.6296	0.147	83.82

Table 3.4. Summary of genetic variation statistics for all loci of ISSR among the *Podophyllum hexandrum* populations with respect to their distributions among eleven forest divisions.

Forest divisions	Sample size	Na	Ne	H	I	Ht	Number of polymorphic loci	Percentage of ploymorphic loci (%)
Parvati	3	1.1765 (0.3841)	1.1412 (0.3072)	0.0784 (0.1707)	0.1123 (0.2445)	0.0784 (0.0291)	12	17.65
Kullu	5	1.2794 (0.4520)	1.1781 (0.3107)	0.1103 (0.1762)	0.1576 (0.2588)	0.1059 (0.0310)	19	27.94
Dodrakwar	2	1.1324 (0.3414)	1.1324 (0.3414)	0.0662 (0.1707)	0.0917 (0.2366)	0.0662 (0.0291)	9	13.24
Seraj	2	1.2206 (0.4177)	1.2206 (0.4177)	0.1059 (0.2089)	0.1529 (0.2895)	0.1103 (0.0436)	15	22.06
Churah	2	1.1471 (0.3568)	1.1471 (0.3568)	0.0735 (0.1784)	0.1019 (0.2473)	0.0735 (0.0318)	10	14.17
Lahaul	2	1.0882 (0.2857)	1.0882 (0.2857)	0.0441 (0.1429)	0.0612 (0.1981)	0.0441 (0.0204)	6	8.82
Palampur	3	1.1912 (0.3962)	1.1529 (0.3169)	0.0850 (0.1761)	0.1217 (0.2522)	0.0850 (0.0310)	13	19.12
Rampur	2	1.1324 (0.3414)	1.1324 (0.3414)	0.0662 (0.1707)	0.0917 (0.2366)	0.0662 (0.0291)	9	13.24
Kinnaur	2	1.0588 (0.2370)	1.0588 (0.2370)	0.0294 (0.1185)	0.0408 (0.1643)	0.0294 (0.0140)	4	5.88
Pangi	3	1.1176 (0.3246)	1.0941 (0.2597)	0.0523 (0.1443)	0.0749 (0.2066)	0.523 (0.0208)	8	11.76
Bharmaur	2	1.1176 (0.3246)	1.1176 (0.3246)	0.0588 (0.1623)	0.0815 (0.2250)	0.0588 (0.0263)	8	11.76
Mean		1.1510	1.1330	0.0700	0.0989	0.1128	10.27	15.05

Na = Observed number of alleles; Ne = Effective number of alleles; H = Nei's gene diversity; I = Shannon's Information index; Mean = Na, Ne, H and I of all over loci of 28 populations; Ht = Total gene diversity; Hs = Population diversity; Gst = Gene differentiation; Nm = Number of gene flow.

The mean Nei's genetic diversity (H) ranged from 0.0294 at Kinnaur to 0.1103 at Kullu forest division which showed overall 2.04% to 11.03% heterozygosis in the population of *P. hexandrum* distributed among different forest divisions. Shannon's indices (I) ranged from 0.0408 to 0.1576 with an average of 0.0989 among the forest divisions. The results reveal that out of the 11 forest divisions, the populations belonging to Kullu division exhibited the greatest variability (PPB=27.94%), I = 15.76%) (Table 3.4).

3.3.4 AMOVA analysis

The analysis of molecular variance indicated that over half of the total variation in the studied populations (62%) could be accounted for by differences among the 11 forest divisions with a further 38% being accounted for by the variation among populations within a forest division (Table 3.5). All components of molecular variance were significant ($P < 0.001$).

Table 3.5. Summary of nested analysis of molecular variance (AMOVA) based on ISSR genotypes of *Podophyllum hexandrum* (levels of significance are based on 1000 iteration steps, d.f.: degree of freedom; S.S.D.: sum of square deviation; P-value: probability of null distribution).

Source of variation	d.f.	S.S.D.	Variance component	Percentage	P-value
Among regions	10	209.750	6.693	62	< 0.001
Among populations within regions	17	70.500	4.147	38	< 0.001
Total	**27**	**280.25**	**10.84**		

3.3.5 Cluster analysis

Cluster analysis of ISSR data based on Jaccard similarity matrix among the 28 populations with respect to their geographical location (Forest Divisions) generated a dendrogram with 11 clusters (C1-C11) (Figure 3.3). For this analysis one representative individual from each population was taken into consideration. All the populations in each region clustered together having similarity co-efficient values ranging from 0.57 to 0.96. Cluster C1 represented parvati forest division with 3 different populations R/4 kasol, Anganoala and Twin Multivora. Cluster C2 (Kullu forest division) had 5 populations

Brundhar, Gulaba, Sanghar Nry, Chander Khani and Kaned Nry. Cluster C3 represented Rampur forest division which had 2 populations Bander Thach and Saropa Nry. Cluster C4 represented Lahaul forest division with 2 populations Nayanghar and Myar Valley. Cluster C5 had 2 populations (Kinnaur forest division) of Nichar Nry and Rango (N-C-8). Cluster C6 (Pangi division) had 3 populations Sach Range, Killer Range and Purthi Range. Cluster 7 (Bharmaur forest division) had 2 populations Ghoi DPF and Samara RF. Cluster C8 represented Palampur forest division which had 3 populations IHBT, Bada Bangal and Chota Bangal. Similarly, 2 populations Madhvi Thach and Kala Pani, each from forest division Dodrakwar (Cluster C9). Cluster C10 (Churah forest division) and Cluster C11 (Seraj Forest division) had 2 populations from each froest division namely DPF-D-1892-C1, DPF-D-791-C1 and Sojha Nry, Jellora C-3b respectively. The result indicates high genetic diversity in *P. hexandrum* populations from Himachal Pradesh.

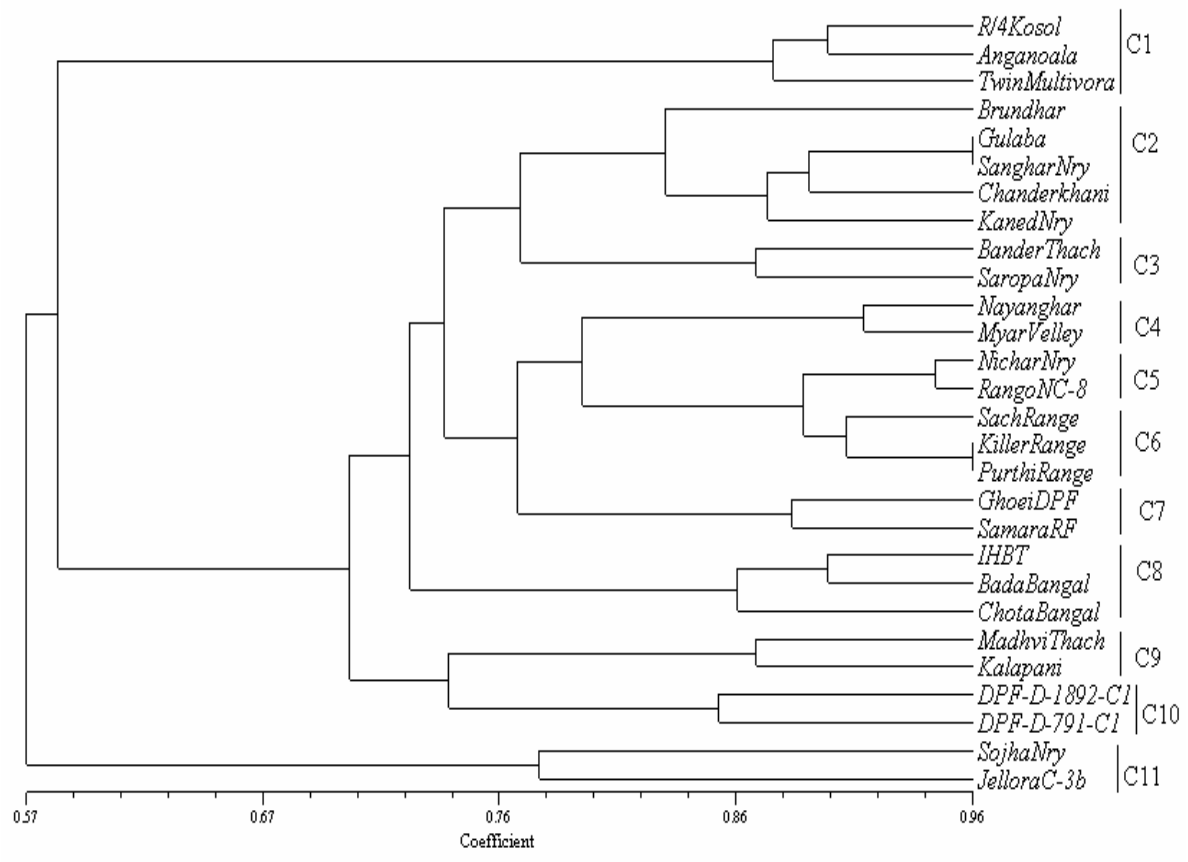


Figure 3.3. Dendrogram illustrating genetic relationships among 28 populations in population diversity study generated by UPGMA cluster analysis calculated from 1303 ISSR bands produced by 11 primers.

3.4 Discussion

ISSR–PCR has been successfully employed to reveal genetic variations in potatoes (*Solancium tuberosum*) (Bornet et al., 2002), wild emmer wheat (*Toiticum dicoccoides*) (Fahima et al., 2002), *Oryza officinalis* (Gao, 2005), mangrove populations (Jian et al., 2004), etc. to characterize genomic diversity. As the primers can anchor both at their 3'end or 5'end, the annealing of the primer only at the 3'end or 5'end of the microsatellite motif could avoid internal priming and smear formation. The anchor also allows only a subset of the targeted inter-repeat regions to be amplified, thereby reducing a high number of PCR products expected from the priming of dinucleotide inter-repeat regions to a set of about 10-50 easily resolvable bands. Pattern complexity can be tailored by applying different primer lengths and sequences (Zietkiewioz et al., 1994). Based on its unique characters, ISSR technique can detect more genetic loci than isozyme and has higher stability than RAPD. Our work is the first application of this method to the wild population of *P. hexandrum*. The experimental results of this study provide evidence of the reliability and usefulness of ISSR markers to assess the high genetic diversity within and between *P.hexandrum* populations from the northwestern Himalayas (Himachal Pradesh).

In this study, high levels of genetic diversity (PPB = 83.82 %, I = 0.4413) were found at the species' level in the studied populations of *P. hexandrum*. However, low levels of genetic diversity (PPB = 15.05%, I = 0.0989) occurred within populations. In general, dispersal resulting in colonization and gene flow into existing populations is very important for both the persistent and genetic success of a species (Hamrick and Godt, 1996). In population genetics, a value of gene flow (Nm) < 1.0 (less than one migrant per generation into a population) or equivalently, a value of gene differentiation (Gst) > 0.25 is generally regarded as the threshold quantity beyond which significant population differentiation occurs (Slatin, 1987). The high Gst value (0.6296) and the low Nm value (0..147) both indicated rapid genetic differentiation among the 28 populations, especially among the regions. About 38 % of the genetic variation in the samples can be attributed to variation among populations. Populations from the same forest division have been clustered together (Fig. 3.3). This indicates that gene flow in the studied populations of *P. hexandrum* occurred mainly within the same forest division rather than among divisions. Partitioning of diversity

is mainly influenced by the system of reproduction. The high genetic variation in *P. hexandrum* may be attributed partly to the cross-pollinated nature of *P. hexandrum.* Instead, the resulted genetic diversity may also be due to clonal propagation of *P. hexandrum.* Although clonal propagation contributes towards genetic uniformity within each population, Hangelbroek et al. (2002) reported that clonal plant species can have high levels of genetic variation in some cases. Similar reports have been made in ISSR studies of populations of *Ceriops tangal* in Thailand and China (Ge and Sun, 2001) and *Heritiera littoralis* in china and Australia (Jian et al., 2004). In these two studies, the estimated value of Gst was 0.529 and 0.426 respectively. Higher Gst value indicates a lower value of gene flow (Nm) among populations and higher genetic differentiation in populations. In our study, genetic diversity values among populations belonging to different regions vary with a more pronounced geographical separation of the populations concerned, implying isolation by altitude.

AMOVA revealed that there was significant variation arising from habitat-correlated genetic difference (38%) suggesting that besides the effects of gene flow and genetic drifts, local ecological conditions (altitude, distance, temperature, rainfall, humidity, soil, pH etc) also played an important role in the variation of the genetic structure in the studied populations of *P. hexandrum*. Considering the high genetic differences among the wild populations of *P. hexandrum*, preservation of only a few populations may not adequately protect the genetic variation within the species in the Himalayan region. At present, the rate of propagation of *P. heaxandrum* is far less than the rate of its exploitation. This species or at least a large part of its genetic diversity may be lost in the near future owing to its importance and overexploitation as a medicinal plant if appropriate conservation measures are not adopted. Since a single or even a few plants cannot represent the whole genetic variability in *P. hexandrum,* there appears to be a need for maintenance of a sufficiently large population in natural habitats to conserve genetic diversity in *P. hexandrum* and avoid genetic erosion.

The low level of genetic diversity and low gene flow among populations detected in this study point towards the possibility of instances of a single isolated population possessing unique genotypes not found in other populations. It is, therefore, imperative for

conservation planners to design suitable conservation strategies and ensure that separate populations are targeted for conservation rather than conserving a few selected populations.

For conservation purposes, it is very important to assess the potentiality of *P. hexandrum* populations for podophyllotoxin production. Total synthesis of podophyllotoxin is an expensive process and availability of the compound from the natural resources is an important issue for pharmaceutical companies that manufacture anticancer drugs (Canel et al., 2000). The existing variations in podophyllotoxin content among the populations were proved to be coupled with altitude (Fig. 3.4a) and environmental variables but not with genetic diversity (Fig. 3.4b). This result is very much supported by the studies done by Sharma et al., (2000). They reported considerable reduction in the podophyllotoxin content in the roots of plants collected from a higher altitude and after being grown at a lower altitude. Thus, the study demands the optimization of environmental factors in order to increase the rate of production of podophyllotoxin from collected populations.

(a) **(b)**

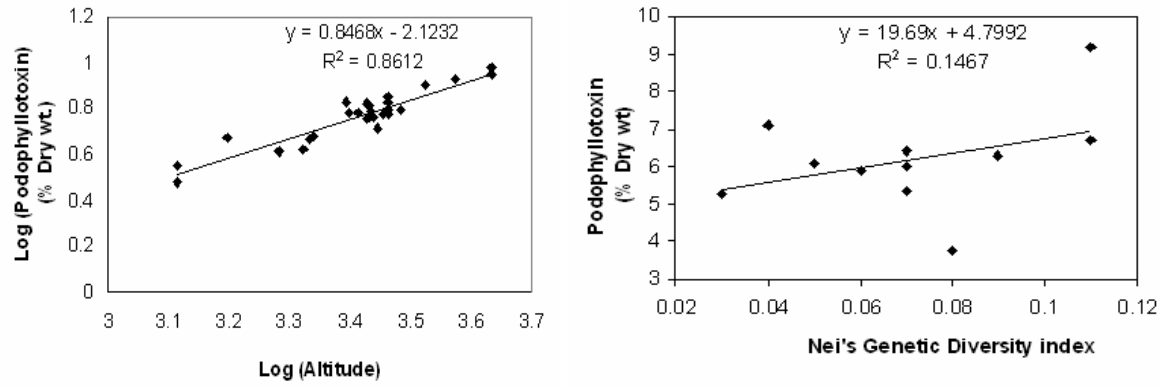


Figure 3.4. Regression analysis based on (a) $\log_{10}$ M (geographical altitude) and Log_{10} (podophyllotoxin content) between 28 populations; (b) Nei's genetic diversity values and podophyllotoxin content between 11 populations (forest division wise) of *P. hexandrum*. There is a significant correlation of t-test of regression coefficient ($t = 286.83$, $p < 0.001$) between Log M (altitude) and Log (podophyllotoxin content) and between Nei's genetic diversity values and podophyllotoxin content ($t = -13.65$, $p < 0.001$).

Ex situ conservation may also be appropriate since low genetic diversity within a population means total genetic diversity in a population may be adequately captured in only

a few transplants from the wild which would not be the case for species with high levels of genetic diversity within a population. It would be beneficial to find ways to strengthen the gene flow among populations to maintain the natural genetic variation within populations of *P. hexandrum*. In this study, genetic diversity of the populations within a region appeared to have increased minimally with higher altitude. This phenomenon has not been observed among plants from different forest divisions. Similar findings were made by Fahima et al., (2002) who reported that microsatellite polymorphisms in natural populations of wild emmer wheat were best explained by variation of altitude and temperature in August. The observed increase in genetic diversity as well as podophyllotoxin content with increase in altitude is an interesting phenomenon that requires further research.

CHAPTER 4

Congruence of RAPD and ISSR markers for evaluation of genomic relationship among 28 populations of *Podophyllum hexandrum* from Himachal Pradesh, India

Abstract

Twenty eight populations of *Podophyllum hexandrum* were selected to study genetic relationship using RAPD and ISSR markers from the northwestern Himalayas, Himachal Pradesh, India. Nineteen RAPD primers and 11 ISSR primers amplified a total of 131 and 68 scoreable bands of which 92.37% and 83.82 % were polymorphic. The mean coefficient of gene differentiation (Gst) was 0.6933 and 0.6296 indicating that 33.77 % and 29.44 % of the genetic variation resided within the populations. Estimated value of gene flow using RAPD (Nm = 0.11059), ISSR (Nm = 0.1470) individually and in combination of RAPD+ISSR (Nm = 0.1211) markers indicated that there was limited gene flow among the populations. The dendrogram obtained from UPGMA analysis revealed grouping of populations with respect to their forest division except with the Kullu forest division. The existence variation among 28 populations based on the percentage of polymorphic bands (PPB) was proved to be coupled with geographical altitude (r = 0.474). The genetic similarity matrices generated by ISSR and RAPD markers were highly correlated (r = 0.721 at P = 0.001) and showed similar estimation between two systems. Both the markers were equally useful in providing some understanding about the genetic relationship of different populations in *Podophyllum.*

4.1 Introduction

The Himalayan region is home to numerous highly valued medicinal plants including *Podophyllum hexandrum* Royle (Berberidaceae) also known as Indian Mayapple. It is distributed in very restricted pockets in the Himalayan zone at altitudes ranging from 1300 to 4300 m (Nayar *et al*., 1990). *P. hexandrum* is a herbaceous plant and has an extensive rhizogenous system that allows it to spread and survive as established colony. It is recognized for its anti-cancer properties. The rhizomes and roots of *P. hexandrum* contain anti-tumor lignans such as podophyllotoxin, 4'-dimethyl podophyllotoxin and podophyllotoxin 4-o-glucoside (Imbert, 1998). Among the plethora of physiological activities and potential medicinal and agricultural applications, the antineoplastic and antiviral properties of podophyllotoxin congeners and their derivatives are arguably the most important from a pharmacological perspective. Semisynthetic derivatives of epipodophyllotoxin, e.g. etoposide (VP-16) (Allevi et al., 1993), etopophos (Schacter, 1996) and teniposide (VM-26) are effective agents in the treatment of lung cancer, a variety of leukemia and other solid tumors (Van Uden et al., 1989). Growing demand in the world for anti-cancer drugs adds much to the importance of Podophyllotoxin and the *Podophyllum* species. Presently, *P. hexandrum* (an Indian species) and *P. peltatum* (an American species) are the commercial sources of podophyllotoxin for the pharmaceutical industry. However, the Indian species *P. hexandrum* Royal contains three times more podophyllotoxin than the American species *P. peltatum* (Fay & Ziegler, 1985) adding further to the importance of *P. hexandrum.* However, the annual supply is at present estimated at 50 – 80 tonnes while the demand is more than 100 tonnes. To meet the ever increasing demand for the crude drug, the rhizomes of *P. hexandrum* are being indiscriminately collected in large quantities. As a result of this and a lack of organized cultivation, *P. hexandrum* has been reported as a threatened species from the Himalayan region. The population size of *P. hexandrum* is very low (40-700 plants per location) and is declining rapidly every year. Some of the populations in certain pockets have virtually disappeared owing to anthropogenic activities and overexploitation (Bhadula et al., 1996). In the natural habitat, seed germination and seedling establishment are very poor and propagation is mostly through rhizomes. Since the species is already endangered and exploitation of its underground parts continues to exceed the rate of natural regeneration, it needs immediate attention for conservation. In this regard,

studies of its population biology and genetic diversity are important for successful development of conservation strategies.

Considerable variations in morphological characters such as plant height, leaf characteristics, fruit weight, seed weight and color etc. have been reported in *P. hexandrum* from the interior Himalayas (Bhadula et al., 1996; Airi et al., 1997; Purohit et al., 1999). Therefore, it would be inefficient, environmentally destructive and economically unsound to randomly harvest *P. hexandrum* from the wild. Polypeptide profiles and esterase isozyme analysis have indicated the existence of high inter- and intra-population variations in *P. hexandrum* (Bhadula et al., 1996). However, the protein markers are influenced by the stages of plant growth as well as environmental factors and hence may give erroneous results. DNA markers such as ISSR (Salimath et al., 1995) and RAPD (Williams et al., 1990; Kazan et al., 1993) on the other hand are quite stable and highly polymorphic in nature. To our knowledge, no report has been prepared on the genetic diversity, population structure and gene flow among the populations of *P. hexandrum* in the Himalayan region with high resolution molecular markers like ISSR and RAPD for combined analysis.

Inter Simple Sequence Repeats (ISSR) amplifies inter-microsatellite sequences at multiple loci throughout the genome (Salimath et al., 1995; Li & Xia, 2005). An ISSR molecular marker technique permits the detection of polymorphism in microsatellites and inter-micorsatellite loci without previous knowledge of DNA sequences (Zietkiewicz et al., 1994). Furthermore, they are highly reproducible due to their primer length and to the high stringency achieved by the annealing temperature. This technique has been widely used to investigate genetic diversity and population genetic structure (Reddy & Nagaraju, 1999; Li & Xia 2005; Chen et al., 2005) because of its advantages in overcoming limitations of allozyme and RAPD techniques (Wolfe et al., 1998; Ratnaparkhe et al., 1998; Esselman et al., 1999).

In this study we investigated the genomic relationship among the wild populations of *P. hexandrum* and their relationship with geographical altitude in the northwestern Himalays, Himachal Pradesh, India, with the aim to facilitate conservation management of

the remaining populations. Appropriate conservation management could be adopted including in situ conservation and germplasm collection from those of the remaining populations with the greatest genetic variations.

4.2 Materials and Methods

4.2.1 Plant material

Twenty-eight populations of *P. hexandrum* were collected from 28 sites covering 11 forest divisions with altitudes ranging from 1300m to 4300m from the interior range of the northwestern Himalayan region, Himachal Pradesh, India (Table 4.1). One forest division has 2-5 selected sampling sites. From each site representative plant samples (7-8 plants) of different age groups (1^{st}, 2^{nd}, 3^{rd} and 4^{th} year) were collected (one population). Fresh leaves (about 5g) from these plants were harvested, mixed together and placed in a zip-lock plastic bag containing silica gel which speeded up the drying process. The pair wise distance between populations within a forest division was 0.5 – 32 km whereas the pair wise distance between forest divisions was 10 – 400 km. The samples were brought to the laboratory and stored at -80 ^{0}C prior to DNA isolation.

4.2.2 Isolation of DNA

Total genomic DNA was extracted from frozen leaves by the CTAB method (Saghai-Maroof et al., 1984). Samples of 500 mg were ground to powder in liquid nitrogen using a mortar and pestle. The powders were transferred to a 30 ml sterile Falcon tube with 12.5 ml of CTAB buffer. The extraction buffer consisted of 2% (w/v) CTAB (Cetyl trimethyl ammonium bromide, sigma), 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 9.5 and 0.2% (v/v) β-mercaptoehanol. After incubating the homogenate for 1 hour at 65 ^{0}C an equal volume of chloroform was added and centrifuged at 15,000 rpm at 10 ^{0}C for 20 min. DNA was precipitated with 1/10 volume (ml) of 3M sodium acetate and an equal volume of isopropanol followed by centrifugation at 10,000 rpm for 10 minutes. RNA was removed by RNase treatment. DNA was quantified by comparing with known quantity of uncut λ DNA on the agarose gel, diluted to 12.5ng.μl^{-1} and used in PCR.

Table 4.1. Twenty eight populations of *Podophyllum hexandrum* collected from different sites, at different altitude covering eleven forest divisions and their polymorphic features using RAPD, ISSR, RAPD+ISSR markers.

Forest Division	Sampling site	Altitude (m)	Total Number of bands			% of Polymorphic Band (PPB)		
			RAPD	ISSR	RAPD +ISSR	RAPD	ISSR	RAPD+ISSR
Parvati	R/4 Kasol	1570	56	38	94	82.14	71.05	77.65
	Twin Multivora	1300	57	36	93	82.45	69.44	77.41
	Anganoala	1300	56	37	93	82.14	70.27	77.41
Kullu	Brundhar	1916	79	58	137	88.88	81.03	91.97
	Gulaba	2895	76	48	124	88.23	77.08	83.06
	Sanghar Nry	2100	80	51	131	85.07	78.43	83.96
	Kaned Nry	2150	80	53	133	85.29	79.24	84.21
	ChanderKhani	3352	83	52	135	89.69	78.84	84.44
Dodrakwar	Madhvi Thach	3048	90	50	140	95.55	78.00	85.00
	Kala Pani	2743	85	51	136	89.58	78.43	84.55
Seraj	Sojha Nry	2667	97	42	139	88.63	73.80	84.89
	Jalora C-3b	2473	90	41	131	87.34	73.17	83.96
Churah	DPF-D-1892-C1	3750	67	45	112	85.71	75.55	81.25
	DPF-D-791-C1	2700	68	47	115	87.95	76.59	81.73
Lahaul	Nayanghar	4300	88	48	136	87.50	77.08	84.55
	Myar Valley	4300	96	50	146	87.50	78.00	85.61
Palampur	IHBT	2800	81	42	123	88.88	73.80	82.92
	Bada Bangal	2895	90	45	135	87.95	75.55	84.44
	Chota Bangal	2700	83	43	126	87.65	74.41	83.33
Rampur	Bander Thach	2895	88	50	138	88.63	78.00	84.78
	Saropa Nry	2499	83	49	132	87.45	77.55	84.09
Kinnaur	Nichar Nry	2190	91	48	139	89.01	77.08	84.89
	Rango (N-C-8)	2710	83	50	133	87.95	78.00	84.21
Pangi	Sach Range	2712	81	48	129	87.65	77.08	83.72
	Killer Range	2850	78	45	123	87.17	75.55	82.92
	Purthi Range	2900	83	46	129	87.95	76.08	83.72
Bharmaur	Ghoei DPF	2680	83	45	128	87.95	75.55	83.59
	Samara RF	2590	85	45	130	88.23	75.55	83.84
Total			**2257**	**1303**	**3560**	**87.50**	**76.07**	**83.50**

4.2.3 RAPD amplification

A total of 20 RAPD primers (Operon Tech USA) were initially screened and out of them 19 random decamer primers A, B, C and D series (Table 4.2) which produced clear reproducible fragments were selected for further analysis. DNA was amplified following the protocol of Williams et al. (1990). Amplification reactions were performed in volumes of 25 µl containing 10 mM Tris HCl, pH 9.0, 1.5 mM $MgCl_2$, 50 mM Tris HCl, pH 9.0, 1.5 mM

$MgCl_2$, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 25 ng template DNA and 0.5 unit of Taq polymerase (Sigma). DNA amplification was performed using a Gene Cycler (BioRAD USA). The First cycle consisted of denaturation of template DNA at 94 ^{0}C for 5 min., primer annealing at 37 ^{0}C for 1 min. and primer extension at 72 ^{0}C for 2 min. In the next 42 cycles the period of denaturation was reduced to 1 min. at 92 ^{0}C while the primer annealing and primer extension time remained the same as in the first cycle. The last cycle considered of only primer extension (72 ^{0}C) for 7 min. The PCR products were stored at 4 ^{0}C before analysis.

4.2.4 ISSR amplification

A total of 30 ISSR primers (commercially synthesized from Sigma Incorporation) were screened with 10 plant samples. After assessing the effects of Mg^{+2}concentration, template DNA concentration and temperature during the annealing stage of the amplification, 11 primers which produced clear and reproducible fragments were selected for further analysis. The sequences of these ISSR primers are listed in Table 4.2. The PCR amplification was performed in a 25 µl reaction volume containing 100 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$, 10 mM each of DNTP, 0.4 µM of primer, 0.01 % gelatin,1 unit of Taq polymerase and 25 ng of genomic DNA. Initial denaturation for 5 min at 94 ^{0}C was followed by 40 cycles of 1 minute at 94 ^{0}C, 1 min. at specific annealing temperature, 2 min. at 72 ^{0}C and a 10 min. final extension step at 72 ^{0}C. The annealing temperature in this study ranged from 45 ^{0}C to 55 ^{0}C.

4.2.5 Agarose gel electrophoresis

Amplification products were electrophoresed on 1.5% agarose gels run at constant voltage (50 V) in 1X TBE for approximately 2 hrs, visualized by staining with ethidium bromide (@0.5 µg ml^{-1}) and a total of 2.5 µl loading buffer (1.0 X TBE, 50 % glycerol, 0.25 % xylene cyanol) was added to each reaction before electrophoresis. After electrophoresis the gels were observed under a UV-transilluminator and documented in Gel-Doc 2000 (Bio-Rad). Gel-Pro analyzer version 3-1 software was used to score ISSR profile. The

reproducibility of the amplification was confirmed by repeating each experiment three times.

Table 4.2. RAPD and ISSR primers used, total number of recorded markers for each primer and their percentage of polymorphic band along with resolving power of DNA samples collected from 28 populations of *Podophyllum hexandrum*.

Primer type	Nucleotide sequence	No. of recorded bands	Percentage of polymorphic band (PPB)	Resolving power
RAPD				
OPA01.	5'CAGGCCCTTC3'	3	100	3.714
OPA02.	5'TGCCGACCTG3'	10	100	14.286
OPA04.	5'AATCGGGCTG3'	6	83.3	8.0
OPA08.	5'GTGACGTAGG3'	7	100	8.0
OPA11.	5'CAATCGCCGT3'	7	100	9.429
OPA13.	5'CAGCACCCAC3'	3	100	2.429
OPA18.	5'AGGTGACCTG3'	11	90	16.571
OPB11.	5'GTAGACCCGT3'	9	77.8	10.571
OPB15.	5'GGAGGGTGTT3'	6	50	9.857
OPB18.	5'CCACAGCAGT3'	9	100	10.571
OPB19.	5'ACCCCCGAAG3'	10	100	9.571
OPC08.	5'TGGACCGGTG3'	5	100	6.929
OPC12.	5'TGTCATCCCC3'	6	100	5.929
OPC15.	5'GACGGATCAG3'	4	100	2.0
OPC16.	5'CACACTCCAG3'	7	100	6.286
OPD05.	5'TGAGCGGACA3'	4	75	5.071
OPD08.	5'GTGTGCCCCA3'	8	87.5	9.429
OPD11.	5'AGCGCCATTG3'	7	100	8.786
OPD13.	5'GGGGTGACGA3'	9	100	13.786
ISSR				
P02	5'AGAGAGAGAGAGAGAGT3'	5	20	9.358
P08	5'TGTGTGTGTGTGTGTGA3'	3	66.7	4.071
P10	5'AGAGAGAGAGAGAGAGYT3'	3	66.7	4.500
P 13	5'CTCTCTCTCTCTCTCTRA3'	5	80	5.714
P 16	5'CCGCCGCCGCCGCCGCCG3'	5	80	4.642
P 17	5'GGCGGCGGCGGCGGCGGC3'	8	100	10.428
P 21	5'CTTCACTTCACTTCA3'	7	85.7	11.071
P 22	5'TAGATCTGATATCTGAATTCCC3'	12	100	12.642
P 23	5'AGAGTTGGTAGCTCTTGATC 3'	6	100	9.500
P 24	5'CATGGTGTTGGTCATTGTTCCA3'	5	60	7.000
P 25	5'ACTTCCCCACAGGTTAACACA3'	9	100	14.142

4.2.6 Data analysis

RAPD and ISSR amplified fragments were scored for band presence (1) or absence (0) and a binary qualitative data matrix was constructed. Data analyses were performed

using the NTSYS pc version 2.0 computer package program (Rohlf, 1992). Pair wise distance matrix was calculated using the Jaccard similarity coefficient (Sneath & Sokal, 1973). The similarity values were used to generate a dendrogram via the un-weighted pair group method with arithmetic average (UPGMA). The data matrix of RAPD and ISSR were also used for assessment of genetic structure, genetic differentiation, gene flow and diversity. Measurement of diversity including gene diversity (H), gene flow and Shannon's information index (I) were estimated by POPGEN 1.32 software (Table 4.3).

4.2.7 Resolving power

According to Prevost & Wilkinson (1999) the resolving power (Rp) of a primer is: $Rp = \sum IB$, where *IB* (band informativeness) takes the value of: 1-[2x (0.5 – *P*)], *P* being the proportion of the 28 genotype (*P. hexandrum* populations analyzed) containing the band.

4.3 Results

4.3.1 RAPD band patterns

The 19 RAPD markers used in the study generated a total of 131 bands (an average of 6.89 bands per primer) out of which 121 (92.37 %) were polymorphic and only 10 (7.63 %) were monomorphic bands. The size of amplified fragments produced ranged from 250 bp to 3100 bp. The number of polymorphic bands varied in between 46 (in R/4 Kasol) to a maximum of 87 (in Sojha nursery) (Table 4.1). The total number of polymorphic alleles were 121, thereby giving an estimate of profound (> 92 %) polymorphism. Also polymorphism differed substantially within the discrete groups of plants with an average of 42.69 % and was found to be between 20.0 % (Kullu) and 50.0 % (Dodrakwar, Churah, Lahaul, Rampur, Kinnaur and Bahrmaur) (Table 4.3). The mean coefficient of gene differentiation (Gst) was 0.6933 (Table 4.4) indicating 33.77 % of the total genetic diversity within the populations. The gene flow (Nm) calculation based on the Gst value (Slatkin & Barton, 1989) between populations was found to be 0.11059 (Table 4.4).

Table 4.3. Summary of genetic variation and polymorphic features estimated using RAPD, ISSR and RAPD+ISSR markers among the *Podophyllum hexandrum* populations with respect to their distributions among eleven forest divisions.

Forest Division	Nei's Genetic Diversity (H) (mean ± SD)			Shannon's information index (I) (mean ± SD)			Percentage of polymorphic loci		
	RAPD	ISSR	RAPD+ISSR	RAPD	ISSR	RAPD+ISSR	RAPD	ISSR	RAPD+ISSR
Parvati	0.0305 (0.1129)	0.0784 (0.1707)	0.0469 (0.1369)	0.0437 (0.1616)	0.1123 (0.2445)	0.0672 (0.1961)	33.33	17.65	10.55
Dodrakwar	0.0344 (0.1270)	0.1029 (0.2037)	0.0578 (0.1603)	0.0476 (0.1760)	0.1427 (0.2824)	0.0801 (0.2222)	50.00	20.59	11.56
Churah	0.0763 (0.1805)	0.0588 (0.1623)	0.0704 (0.1743)	0.1058 (0.2503)	0.0815 (0.2250)	0.0957 (0.2416)	50.00	11.76	14.07
Seraj	0.1794 (0.2407)	0.1176 (0.2137)	0.1583 (0.2332)	0.2487 (0.3337)	0.1631 (0.2962)	0.2194 (0.3232)	49.67	23.53	31.66
Lahual	0.0992 (0.2002)	0.1985 (0.2465)	0.1332 (0.2216)	0.1376 (0.2775)	0.2752 (0.3417)	0.1846 (0.3072)	50.00	39.71	26.63
Kullu	0.1918 (0.2160)	0.2282 (0.2018)	0.2042 (0.2114)	0.2779 (0.3088)	0.3375 (0.2920)	0.2983 (0.3038)	20.00	58.82	50.25
Palampur	0.1595 (0.2140)	0.0850 (0.1761)	0.1340 (0.2045)	0.2284 (0.3065)	0.1217 (0.2522)	0.1919 (0.2928)	33.33	19.12	30.15
Rampur	0.0267 (0.1129)	0.0662 (0.1707)	0.0402 (0.1363)	0.0370 (0.1565)	0.0917 (0.2366)	0.0557 (0.1890)	50.00	13.24	08.04
Kinnaur	0.0382 (0.1333)	0.0294 (0.1185)	0.0352 (0.1282)	0.0529 (0.1848)	0.0408 (0.1643)	0.0488 (0.1777)	50.00	5.88	07.04
Pangi	0.1391 (0.2069)	0.0523 (0.1443)	0.1094 (0.1920)	0.1992 (0.2963)	0.0749 (0.2066)	0.1567 (0.2749)	33.33	11.76	24.62
Bharmaur	0.0229 (0.1049)	0.0588 (0.1623)	0.0352 (0.1282)	0.0317 (0.1455)	0.0815 (0.2250)	0.0488 (0.1777)	50.00	11.76	07.04
Mean	**0.090727**	**0.128227**	**0.093164**	**0.128227**	**0.138445**	**0.131564**	**42.69**	**21.25**	**40.09**

Table 4.4. Genetic variability estimated among 28 populations of *Podophyllum hexandrum.*

Nei's gene diversity	Shannon's Information index	Hs	Gst	Estimate of gene flow (Nm) (0.25(1-Gst)/Gst	Number of polymorphic alleles	% of polymorphic alleles
RAPD						
0.3377 (0.1545)	0.5014 (0.2036)	0.1036 (0.0056)	0.6933	0.11059	121	92.37
ISSR						
0.2944 (0.1731)	0.4413 (0.2391)	0.1090 (0.0076)	0.6296	0.1470	57	83.82
RAPD+ISSR						
0.3229 (0.1620)	0.4809 (0.2176)	0.1054 (0.0062)	0.6735	0.1211	178	89.45

Hs = Population diversity; Gst = Gene differentiation.

The resolving power (Rp) of the 19 RAPD primers ranged from 2.0 (for primer OPC15) to 16.571 (for OPA18). Three RAPD primers (OPA02, OPA18, OPD13) possess the differentiate all the 28 populations of *P. hexandrum* collected from the wild (Table 4.2).

4.3.2 ISSR band patterns

Eleven ISSR primers used in the study generated a total of 68 ISSR bands (an average of 6.18 bands per primer) out of which 57 were polymorphic (83.82%). The amplified PCR fragments ranged from 250 bp to 1350 bp. Out of these 11 primers; P02, P08, P13, P16, P21 and P24 revealed 11 monomorphic bands existing in all the 28 populations. The high reproducibility of ISSR markers may be due to the use of longer primers and higher annealing temperature than those used for RAPD. The annealing temperature in this study ranged from 45 ^{0}C to 55 ^{0}C. The number of polymorphic bands varied between 25 (in Twin Multivora) and 47 (in Brundhar) (Table 4.1). The total number of polymorphic alleles is 57, thereby giving an estimate of > 83% polymorphism. It differed substantially within the discrete groups of plants with an average of 21.25% and was found to be in between 5.88% (Kinnaur forest division) and 58.82% (Bahrmaur forest division). The mean coefficient of gene differentiation (Gst) was 0.6296, indicating

29.44% of the total genetic diversity within the populations. The total gene flow (Nm) between populations was found to be 0.1470 (Table 4.4).

The resolving power (Rp) of the 11 ISSR primers ranged from 4.071 (for primer P08) to 12.642 (for primer P22). Three ISSR primers (P21, P22, P25) possess the highest Rp values (11.071, 12.642 and 14.142 respectively) and each one is able to differentiate all the 28 populations of *Podophyllum* collected from the wild (Table 4.2).

4.3.3 Dendrogram analysis

Genetic similarity was calculated from the Jaccard similarity index value for all the 28 accessions of *Podophyllum hexandrum* considering ISSR and RAPD markers individually as well as together. Based on the RAPD marker alone, the similarity index values ranged from 0.61 to 0.96. These values were used to construct a dendrogram using UPGMA method. Populations from 11 forest divisions were clustered into region-specific groups with the exception of the Kullu forest division.

Similarly, based on the ISSR markers alone, the similarity index values ranged from 0.57 to 0.96. All the populations were grouped under 12 main clusters. The populations collected from the Kullu forest division were distributed in 2 clusters like that of RAPD markers. The similarity index values did not differ significantly for ISSR and RAPD systems.

Based on the combined markers the similarity index value ranged from 0.59 to 0.94 and has been used to construct the dendrogram (Figure 4.1). All the 28 *Podophyllum hexandrum* populations were distributed into 12 main clusters (Ia-Im). Cluster Ia represents Parvati Forest Division with 3 different populations- R/4 Kasol, TwinMultivora and Anganola. Populations from Kullu Forest Division were distributed in two clusters; Ib & Ic. Ib has the populations- Brundhar and Gulaba and Ic has the populations- Snaghar nursery, Kaned nursery and Chanderkhani.

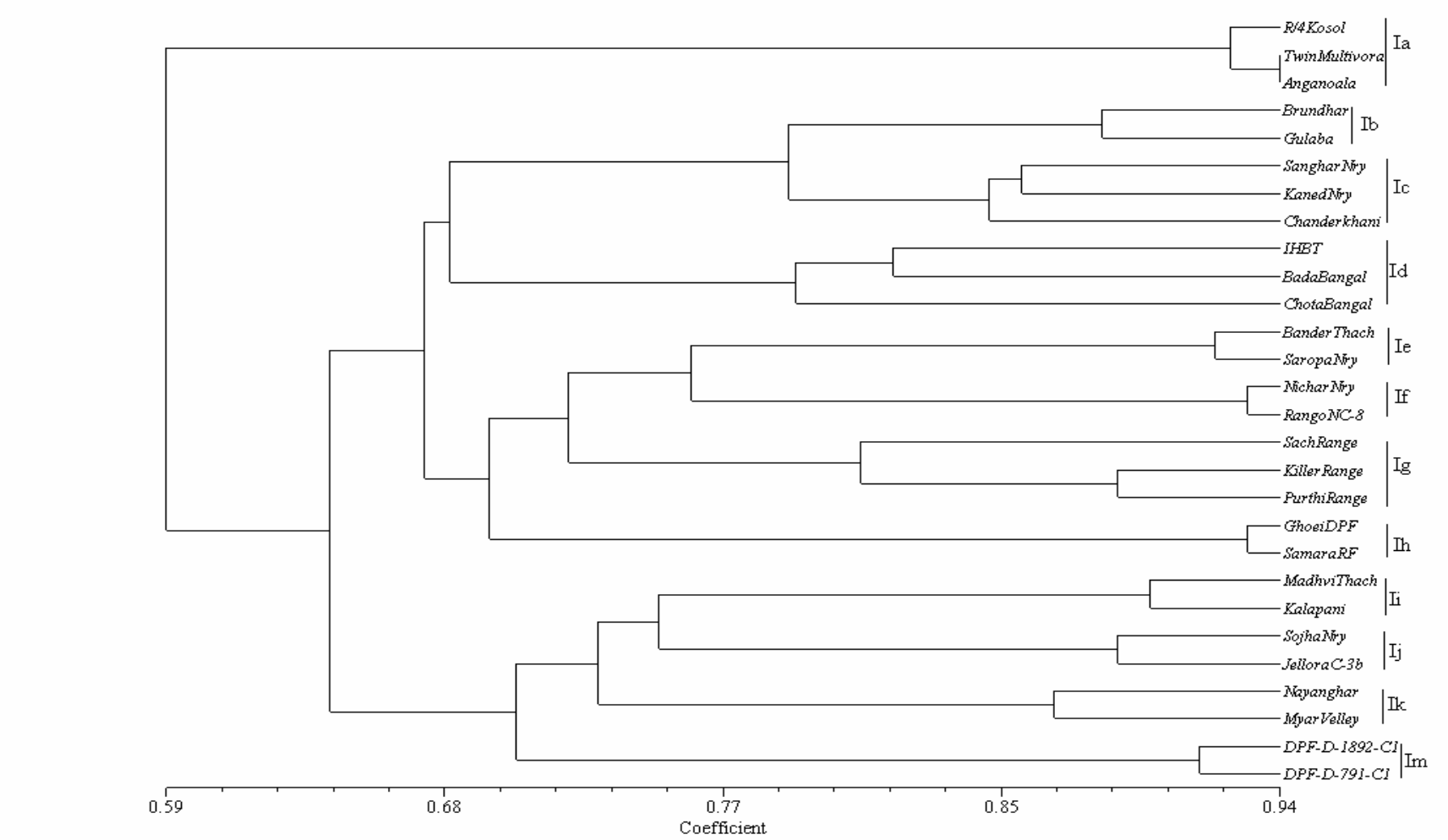


Figure 4.1. Dendrogram illustrating genetic relationships among 28 populations generated by UPGMA cluster analysis calculated from 3560 RAPD+ISSR bands produced by 30 primers.

Cluster Id (Palampur forest division) has 3 populations- IHBT, Bada Bangal and Chota Bangal. Cluster Ie (Rampur division) has 2 populations- BanderThach and Saropa nursery. Cluster If representing Kinnaur Forest Division has 2 populations- Nichar nursery and Rango NC-8. Cluster Ig (Pangi Forest Division) has 3 populations- Sach Range, Killer Range and Purthi Range. Cluster Ih (Bharmaur Forest Division) has 2 populations- Ghoei DPF and Samara RF. Cluster Ii (Dodrakwar Forest Division) has 2 populations- MadhviThach and Kalapani. Similarly Seraj (Cluster Ij), Lahaul (cluster Ik) and Churah Forest Divisions have 2 populations each- Sojha nursery & Jellora Pass; Nayanghar & Myarvelley and DPF-D-1892-C1 & DPF-D-791-C1 respectively. The results indicate high genetic diversity in *P. hexandrum* from Himachal Pradesh.

4.3.4 Comparison of genetic relationship in *P. hexandrum*

Data collected with ISSR and RAPD markers were used to compare genetic similarities between various populations of *P. hexandrum.* The comparison of similarity matrices obtained a value of r = 0.721, at P = 0.001 which indicated a good correlation between data generated by both the systems (Figure 4.2). Further, the existing variations among 28 populations as observed through percentage of polymorphic band (PPB) were proved to be coupled with geographical altitude (r = 0.474) (Figure 4.3).

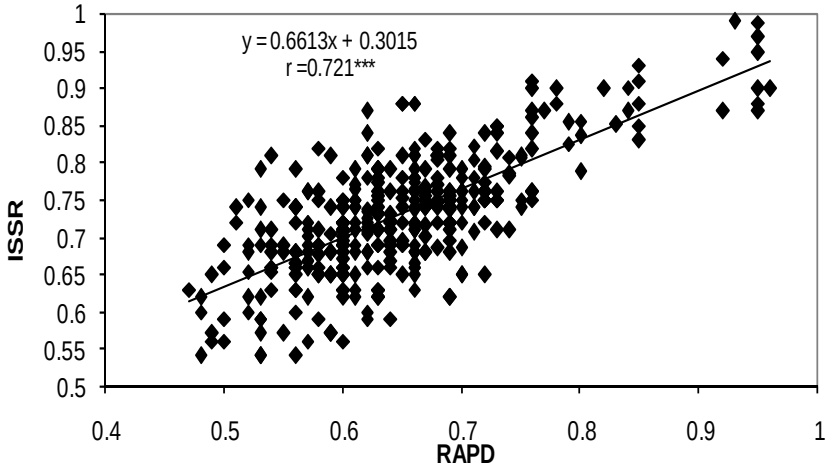


Figure 4.2. Regression analysis of similarity matrices obtained by RAPD and ISSR markers in *Podophyllum hexandrum* populations. The symbol ***, indicates the value is significant at p = 0.001 level of significance.

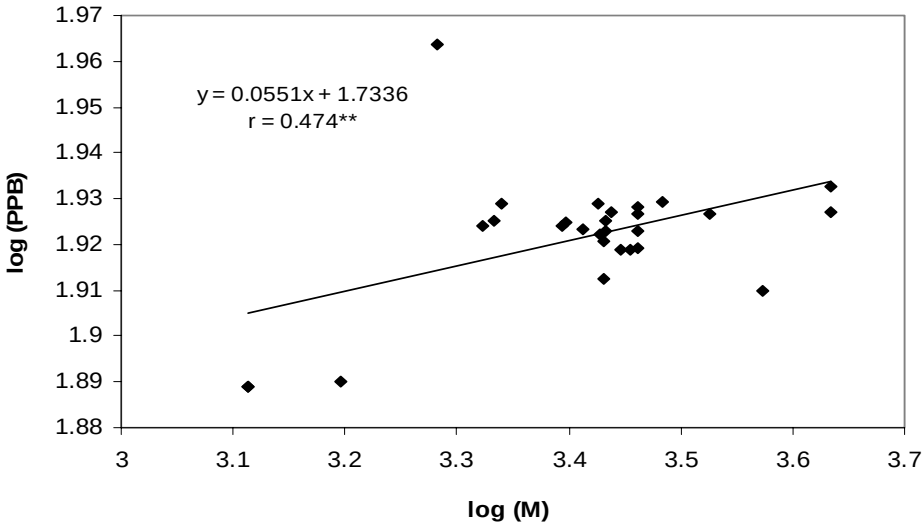


Figure 4.3. Regression analysis based on $\log_{10}$ M (geographical altitude) and Log PPB (percentage of polymorphic band) between 28 populations. The symbol **, indicates the value is significant at p = 0.01 level of significance.

4.3.5 Heterozygosity and molecular variance

Heterozygosity and molecular variance were calculated using RAPD and ISSR markers individually as well as together. Nei's gene diversity (H) value calculated for RAPD, ISSR and RAPD+ISSR were 0.3377 ± 0.1545, 0.2944 ± 0.1731 and 0.3229 ± 0.1620 respectively which showed overall 9.07% to 12.82% heterozygosity among the populations of *P. hexandrum*. Similarly, the Shannon's information indices (I) were 0.5014 ± 0.2036, 0.4413 ± 0.2391 and 0.4809 ± 0.2176 for RAPD, ISSR and RAPD+ISSR markers respectively showed gene diversity measurement with an average of 0.4745 among the forest divisions (Table 4.4).

4.4 Discussion

Considering the high genetic differences among the wild populations of *P. hexandrum*, conservation of only a few populations may not adequately protect the genetic variations within the species in the Himalayan region. At present, the rate of propagation of *P. hexandrum* is far less than the rate of its exploitation. This species or at least a large part of its genetic diversity may be lost in the near future owing to its importance and its exploitation as a medicinal plant if appropriate measures for conservation are not adopted.

Since no single or even a few plants represent the whole genetic variability of *P. hexandrum*, there appears to be a need to maintain sufficiently large populations in natural habitats to conserve the genetic diversity and avoid genetic erosion.

4.4.1 Analysis of polymorphic feature

Twenty eight accessions of *Podophyllum hexandrum* were fingerprinted by using 19 RAPD and 11 ISSR markers. 19 RAPD markers produced 2257 bands on an average of 118.78 bands per primer and 11 ISSR markers generated 1303 bands on an average of 118.45 bands per primer. A high percentage of polymorphism was detected in all the populations of *P. hexandrum* which were examined. The high proportion of polymorphic loci suggests that there is a high degree of genetic variation among the *Podophyllum* populations. The genetic variation estimated by RAPD markers is PPB = 92.37 % and I = 0.5014 whereas for ISSR markers it is 83.82 % and I = 0.4413 at the population level. In this study RAPD was found to be more efficient than the ISSR as it detected 92.37% polymorphic DNA markers in *P. hexandrum* as compared with 83.82% for ISSR. This is in contrast with the result obtained for several other plant species like wheat (Nagaoka et al., 1997) and vigna (Ajibade et al., 2000). This may be due to less number of ISSR markers used in the study compared to RAPDs. It may also be due to the higher annealing temperature used for ISSR markers in comparison to RAPDs and thus the chances of PPB are high in RAPDs. The 19 RAPD and 11 ISSR primers in the present study yielded 178 polymorphic markers that unambiguously discriminated 28 genotypes into 12 clusters. Geographically isolated population accumulates genetic differences as they adapt to different environments. Genetic variations among elite genotypes of *Podophyllum hexandrum* based on RAPD and ISSR analysis could be useful in selecting parents to be crossed for generating appropriate populations intended for both genome mapping and breeding purposes. In general, dispersal resulting in colonization and gene flow into existing populations is very important for both persistence and genetic success of a species (Hamrick & Godt, 1996). In population genetics, a value of gene flow (Nm) < 1.0 (less than one migrant per generation into a population) or equivalently, a value of gene differentiation (Gst) > 0.25 is generally regarded as the threshold quantity beyond which significant population differentiation occurs (Slatin, 1987).

The average Resolving Power (Rp) of ISSR primers (8.48) was higher than RAPD primers (6.89). There is seemingly a linear relationship between the Rp of each primer and the number of recorded markers. This relationship was stronger for RAPDs (r = 0.81) than for ISSRs (r = 0.42). Prevost & Wilkinson (1999) have studied the nature of this relationship using a total of 371 hypothetical primers producing 8, 10 or 12 band position (r = 0.98). However, they have found a seemingly linear relationship between the Rp of real ISSR primers and the number of genotype of potato cultivars identified (r = 0.80). In our case we have observed a greater correlation using RAPDs than ISSRs primers probably due to the low number of cultivars analyzed.

4.4.2 Estimation of genetic relationships

Cluster analysis by using RAPD, ISSR and combination of RAPD+ISSR revealed distribution of different populations with respect to their forest divisions. This indicated that gene flow in the studied populations of *P. hexandrum* occurred mainly within the same forest division rather than between divisions. Moreover, RAPD and ISSR markers used here were able to differentiate *P. hexandrum* populations collected from 11 forest divisions into 12 distinct region specific clusters except the Kullu forest division (Figure 4.1). The study also indicates that *P. hexandrum* populations in the northwestern Himalayan region are genetically highly diverse. The high genetic variations in *P. hexandrum* may be attributed partially to the cross pollinated nature of *P. hexandrum*. Partitioning of diversity is mainly influenced by the system of reproduction. Instead, the resultant genetic diversity may also be due to clonal propagation of *P. hexandrum.* Although clonal propagation contributes towards genetic uniforming within each population, Hangelbroek et al., (2002) reported that clonal plant species can have high levels of genetic variation in some cases. The low gene flow among populations detected in this study point towards the possibility of instances of single isolated populations possessing unique genotypes not found in other populations. It is, therefore, imperative for conservation planners in designing conservation strategies for wild populations of *P. hexandrum* to ensure that many separate populations are targeted for conservation rather than conserving a few selected populations.

The correlation between RAPD and ISSR Jaccard's similarity coefficient value (r = 0.721) indicated a good correlation between data generated by both the systems (Figure 4.2). The observed increase in genetic variations among the populations was coupled with an increase in altitude (r = 0.474) (Figure 4.3). A similar finding was made by Fahima et al., (2002) who reported that microsatellite polymorphisms in natural populations of wild emmer wheat were best explained by variation of altitude and temperature in August. The primers with poly (GC)n and poly (GA)n motifs produced more polymorphism than any other motif. A somewhat similar result was also reported by Ajibade et al., (2000) where they found that the primer containing the CT repeats was one of those which did not give interpretable phenotype when analyzed while primers with GA and CA repeats revealed polymorphism in the genus Vigna.

Based on polymorphic feature, genetic diversity, genetic similarity and gene flow among the populations of *Podophyllum* based on RAPD and ISSR study, we conclude that future conservation plans for this species should be specifically designed to include representative populations with the highest genetic variation for both *in situ* conservation and germplasm collection expeditions.

CHAPTER 5

Amplified fragment length polymorphism (AFLP) analysis of genetic variation in *Podophyllum hexandrum*

Abstract

Podophyllum hexandrum is an important endangered Indian medicinal plant valued all over the world for its anticancer properties. Despite limited knowledge of the levels of genetic diversity and relatedness, their cultivation as a source of podophyllotoxin is widespread. In order to facilitate reasoned scientific decisions on its management and conservation and prepare for selective breeding programme, genetic analysis of 28 populations was performed using amplified fragment length polymorphism (AFLP) markers. A 13 pairs of AFLP primers (*Eco*RI/*Mse*I) generated a total of 551 polymorphic loci out of which 466 (84.40%) were polymorphic loci. The mean coefficient of gene differentiation (Gst) was 0.51, indicating that 26% of the genetic diversity resided within the population. Analysis of molecular variance (AMOVA) indicated that 64% of the genetic diversity among the studied populations was attributed to geographical location while 36% was attributed to differences in their habitats. An overall value of mean estimated number of gene flow (Nm = 0.24) indicated that there was limited gene flow among the sampled populations. The high levels of population differentiation detected suggest that provenance source is an important factor in the conservation and exploitation of *P. hexandrum* genetic resources.

5.1 Introduction

Podophyllum hexandrum Royle commonly called the Himalayan Mayapple, is an endangered perennial herb belonging to the family Berberidaceae and is distributed in the north-western Himalayan region: India, China, Nepal, Pakistan, Afghanistan and other areas (Chatterjee, 1952; Fu, 1992; Ying, 1979). In India, *P. hexandrum* mainly distributed in the wild to very restricted pockets throughout the Alpine Himalayan region (Nayar et al., 1990) at altitude of 1300 m – 4300 m. It has long been used by the Himalayan natives and the American Indian (Anon, 1970) for the treatment of certain types of cancers. The rhizomes of several *Podophyllum* species have been found to be the source of podophyllotoxin lignan that has important biological activity blocking mitosis (Loike and Horwitz 1976a; Loike *et al*. 1978) and its use as the starting compound of semi synthetic chemotherapeutic drugs such as etoposide, teniposide and etophos (Stahelin and Wartburg, 1991; Imbert, 1998). *P. hexandrum* was used in folk medicine by the local people (Li, 1975) and recognized for its anti-cancer properties (Canel et al., 2000; Issell et al., 1984). The total synthesis of podophyllotoxin is complicated due to the presence of four chiral centers, a rigid *trans*-lactone and an axial 7-aryl substituent (Alam et al., 2008). Hence, *P. hexandrum* (an Indian species) and *P. peltatum* (an American species) are presently the commercial source of podophyllotoxin for the pharmaceutical industry. However, the yield of podophyllotoxin from *P. peltatum* is low (~ 0.25% based on dry weight) in comparision to *P. hexandrum*, which contains ~ 4% of podophyllotoxin by dry weight. The rhizomes are being indiscriminately harvested in large quantities from the wild to meet the ever increasing demand for the crude drug. As a result of this and a lack of organized cultivation, *P. hexandrum* has been reported as a threatened species from the Himalayan region, the population size of *P. hexandrum* is very low (40-700 plants per location) and is declining each year. Presently, the size of the wild population has been declining rapidly in the north-western region of the Himalayas, owing to habitat fragmentation, over-exploitation, long dormancy and low rate of natural regeneration. The demand for the compound continues to increase and thus encourages domestication and conservation of *P. hexandrum* in the Himalayan region.

A thorough understanding of the level and distribution of genetic variation in *P. hexandrum* populations is essential for the conservation and management policies (Fritsch and Rieseberg, 1996). PCR-based DNA fingerprinting techniques such as random amplified polymorphic DNA analysis (RAPD) and amplified fragment length polymorphism (AFLP) represent a very informative and cost-effective approach for assessing genetic diversity of a wide range of organisms (Vos et al., 1995; Williams et al., 1990). The usefulness of RAPD for genetic characterization of *P. hexandrum* populations has been confirmed in several studies (Sharma et al., 2000). AFLP fingerprinting combine's universal applicability with high power of discrimination and reproducibility (Janssen et al., 1996) and has been shown to be very efficient in detecting polymorphism in species where little variation could previously be found by RFLP analysis. AFLP markers are reliable for the assessment of genetic variation among and within populations (Curtis and Taylor, 2003; Folkertsma et al., 1996; Jiang et al., 2004; Lamote et al., 2002; Loh et al., 2000; Vos et al., 1995); AFLPs do not require any prior knowledge of a species genetics, which is an especially useful feature (Ajmone-Marsan et al., 1998; De Knijiff et al., 2001; Han et al., 2000). The utility, repeatability and efficiency of the AFLP technique lead to the broader application of this technique. To our knowledge no AFLP analysis has been done to analyze the genetic variation in *Podophyllum hexandrum* population. Our objectives in this study were to examine the levels of genetic differentiation between natural populations of *P. hexandrum* to define genetically distinct units for conservation purposes, and to quantify levels of genetic diversity within populations to optimize sampling strategies for the efficient maintenance of variability in the species using AFLP.

5.2 Materials and Methods

5.2.1 Sampling of plants

Twenty-eight populations of *P. hexandrum* were collected from 28 sites covering 11 forest divisions with an altitude ranging from 1300m to 4300m from the interior range of northwestern Himalayan region, Himachal Pradesh, India (Table 5.1). One forest division has 2-5 selected sampling sites. From each site representative plant samples (7-8 plants) of different age groups (1^{st}, 2^{nd}, 3^{rd} and 4^{th} year) were collected (one population). Fresh leaves (about 5g) from these plants were harvested, mixed together and placed in a zip-lock plastic

bag containing silica gel which speeded up the drying process. The pair wise distance between populations within a forest division was 0.5 – 32 km, whereas, the pair wise distance between forest divisions was 10 – 400 km. The samples were brought to laboratory and stored at -80 ^{0}C prior to DNA isolation.

Table 5.1. Twenty eight populations of *Podophyllum hexandrum* collected from different forest divisions and their podophyllotoxin content.

Forest Division	Sampling site	Altitude (m)	Podophyllotoxin content [% dry weight] (mean ± S.E.)
Parvati	R/4 Kasol	1570	3.567 ± 0.747
	Twin Multivora	1300	4.753 ± 0.796
	Anganoala	1300	3.020 ± 0.524
Kullu	Brundhar	1916	4.077 ± 0.270
	Gulaba	2895	5.943 ± 0.591
	ChanderKhani	3352	8.033 ± 0.454
	Kaned Nry	2150	4.657 ± 0.850
	Sanghar Nry	2100	4.173 ± 0.276
Dodrakwar	Madhvi Thach	3048	6.207 ± 0.743
	Kala Pani	2743	5.800 ± 0.212
Seraj	Sojha Nry	2667	6.607 ± 0.348
	Jalora-C-3b	2473	6.790 ± 0.855
Churah	DPF-D-1892-C1	3750	8.487 ± 0.565
	DPF-D-791-C1	2700	5.753 ± 0.411
Lahaul	Myar Valley	4300	9.533 ± 0.484
	Nayan Ghar	4300	8.857 ± 0.427
Palampur	Bada Bangal	2895	7.097 ± 0.797
	Chota Bangal	2700	6.573 ± 0.827
	IHBT	2800	5.183 ± 0.780
Rampur	Bander Thach	2895	6.773 ± 0.640
	Saropa Nry	2499	6.097 ± 0.942
Kinnaur	Nichar Nry	2190	4.760 ± 0.291
	Rango-NC-8	2710	5.797 ± 0.552
Pangi	Sach Range	2712	6.133 ± 0.216
	Killer Range	2850	5.967 ± 0.692
	Purthi Range	2900	6.233 ± 0.790
Bharmaur	Ghoei DPF	2080	5.700 ± 0.692
	Samara RF	2590	6.030 ± 0.825

5.2.2. Isolation of DNA

Total genomic DNA was extracted from frozen leaves by the CTAB method (Saghai-Maroof et al., 1984). Samples of 500 mg were ground to powder in liquid nitrogen using a mortar and a pestle. The powders were transferred to a 30 ml sterile Falcon tube with 12.5 ml of CTAB buffer. The extraction buffer consisted of 2% (w/v) CTAB (Cetyl trimethyl ammonium bromide, sigma), 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 9.5 and 0.2 % (V/V) β-mercaptoethanol. After incubating the homogenate for 1 hour at 65 ^{0}C an equal volume of chloroform was added and centrifuged at 10,000 rpm for 20 min. DNA was precipitated with 1/10 volume (ml) of 3 M sodium acetate and an equal volume of isopropanol followed by centrifugation at 10,000 rpm for 10 minutes. RNA was removed by RNase treatment. DNA was quantified by comparing with known quantity of uncut λ DNA on the agarose gel and Nano Drops (BioRAD), diluted to 150 ng. μl^{-1} and used in AFLP-PCR.

5.2.3 Extraction and quantification of podophyllotoxin

Ethanolic extract of podophyllotoxin following the procedure of Broomhead et al., (1990) was measured using HPLC analysis as described in chapter 2. All the experiments on extraction of podophyllotoxin and HPLC analysis were repeated three times.

5.2.4 Amplified fragment length polymorphisms (AFLP)

AFLP was performed as described by Vos et al. (1995) and was conducted using the Small Genome Primer Kit AFLP System II (Invitrogen life technology) and visualized with the Polyacrylamide Gel Electrophoresis (PAGE) system. DNA fragments were amplified using the procedure by Vos et al. (1995) modified as follows. Template DNA (150 ng) was digested by 1 µl mixture of *Eco*R I/*MSe* I (1.25 units/µl) at 37 ^{0}C for 2h and at 70^{0}C for 15min to inactivate the restriction endonuclease and ligated to commercial *EcoR* I and *Mse* I oligonucleotide adapters using 1 µl (1 units/µl) of T4 DNA ligase at 22 ^{0}C for 2 hrs. The adapter-ligated DNA is diluted 1:10 ratio and was amplified using a mixture of 5 µl of DNA from the ligation reaction (diluted), 40 µl of preamp primer mix II, 5 µl of 10 X PCR buffer, and 1 µl of *Taq* polymerase (3 units/ µl). The pre-amplification reactions were performed on a Eppendorf Gradient Thermal cycler, using the following cycling parameters: 20 cycles at

94 ^{0}C for 30 s, 56 ^{0}C for 60 s, and 72 ^{0}C for 60s and temperature is 4 ^{0}C. Preamplification products were diluted as 1:50 ratio. Primer labeling is performed by phosphorylating the 5' end of the *EcoR* I primers with γ^{-32}P ATP (5000 Ci/mmol) and T4 kinase at 37 ^{0}C for 1h and heat inactivate the enzyme at 70 ^{0}C for 10 min. Selective amplification was performed using reaction mix composed of 2.5 µl of diluted DNA from preamp, 4.5 µl of *MSe* I primer (6.7 ng/ µl), 0.5 µl of labeled *EcoR* I primer (27.8 ng/ µl), 2.5 µl of 10X PCR buffer, 1.5 µl of distilled water, and 1 µl of *Taq* polymerase (3 units/ µl).

The selective amplification PCRs were performed by another touchdown program as follows: One cycle at 94 ^{0}C for 30 s, 65 ^{0}C for 30 s, and 72 ^{0}C for 60 s lower the annealing temperature each cycle 0.7 ^{0}C during 12 cycles. This gives the touchdown phase of 13 cycles. After completing the touchdown phase of 13 cycles, continued with 23 more cycles at 94 ^{0}C for 30 s, 56 ^{0}C for 30 s and 72 ^{0}C for 60 s. Both pre- and selective amplification conditions were modified according to Myburg et al. (2000). An aliquot of 2 µl of selective amplification product was mixed with 2 µl of formamide dye, denatured for 3 min at 95 ^{0}C, and chilled on ice immediately. The reaction products were then size-fractioned on 6% denaturing polyacrylamide gel on a DNA sequencing apparatus. Electrophoresis was carried out for 3.5 hr in 1X TBE buffer at 1000 V. Gel photograph has been developed by X-ray film and also seen by using Gel Doc phosphor measure system. The resulting banding pattern was analyzed manually.

5.2.5 Data analysis

A total of 13 *EcoR* I + 2 bases/*Mse*I +3 bases AFLP has been used in this study (Table 5.2). Each sample was scored as '1' if a fragment was present and '0' for absence. The numbers of polymorphic fragments produced by each primer set are listed in Table 5.2. Data were imported into the multivariate data analysis program, NTSYSpc, (version 2.1, ExeterSoftware, Setauket, NY). The module SimQual was used to generate a similarity matrix based on simple matching coefficients and the module SAHN performed sequential, agglomerative, hierarchical and non-overlapping clustering by unweighted pair-group method, arithmetic average (UPGMA). The module Tree Plot was used to convert the data generated by SAHN into a dendrogram. Pairwise distance matrix was calculated using the

Jaccard similarity coefficient (Sneath et al., 1973). Support for clusters was evaluated by bootstrapping analysis (Felsenstein 1995). One thousand permutation data sets were generated by resampling with replacement of characters within the combined 1/0 data matrix. POPGENE32 software was used to calculate Nei's unbiased genetic distance between the different populations using all AFLP markers inclusive of monomorphic markers. Nei's unbiased genetic distance is an accurate estimate of the number of gene differences per locus when populations are small. Population diversity (Hs) and total gene diversity (Ht) (Nei, 1973) were calculated within 28 populations and within 11 major groups (as per their collection site) by POPGENE software. Genetic diversity within and among populations was measured by the percentage of polymorphic bands (PPB). Estimate of gene flow (Nm) was calculated by the gene differentiation (Gst) using (0.25(1-Gst)/Gst). In order to describe genetic variability among the populations, the non-parametric analysis of molecular variance (AMOVA) (Excoffier et al., 1992) program version 1.5 was used where the variation component was partitioned among populations, among populations within regions and among regions. The input files for AMOVA were prepared by using AMOVA-PREP (Miller, 1998), version-1.01. A three-dimensional (3D) scatter plot (Principle component analysis, PCA) was generated using NTSYS. PCA is a method used for reducing the number of variables in a complex set of data. Multidimensional scaling was used to represent the relationships among 28 genotypes.

5.2.6 Resolving power

According to Prevost & Wilkinson (1999) the resolving power (Rp) of a primer is: $Rp = \sum IB$, where *IB* (band informativeness) takes the value of: 1-[2x (0.5 – *P*)], *P* being the proportion of the 28 genotype (*P. hexandrum* populations analyzed) containing the band.

5.3 Results and Discussion

5.3.1 AFLP marker size and Patterns

The AFLP technique is highly sophisticated and reproducible due to its stringent amplification procedure (Folkertsma et al., 1996; Brown 1996) and has previously been used successfully in a variety of taxonomic and genetic diversity studies in other species (Maughan et al., 1996; Sharma et al., 1996; Paul et al., 1997; Arens et al., 1998). It was

found suitable for our use with *Podophyllum* populations because of its ability to generate reproducibly polymorphic markers. In this investigation all the 28 genotypes were fingerprinted using 13 primer combinations. The 13 AFLP primer pairs used in this study generated a total of 7192 bands (an average of 553.23 bands per primer) out of which 4616 (mean = 355.08 per pair) were polymorphic across 28 regionally adapted *P. hexandrum* genotypes, along with a total of 2576 monomorphic bands (Table 5.2). The mean polymorphism rate was 69.60% observed among primer pairs. On the basis of these data and DNA marker analyses in other crop species (Maughan et al., 1996; Powell et al., 1996; VanToai et al., 1997), AFLPs offer superior efficiency in terms of polymorphism rate when compared with other DNA marker systems. The resolving power (Rp) of the 13 AFLP primers ranged from 17.071 (for primer combination E+TC/M+CTC) to 83.929 (for primer combination E+TC/M+CTG) which differentiate all the 28 populations of *Podophyllum* collected from the wild (Table 5.2).

Table 5.2. AFLP markers obtained from 13 primer combinations among 28 *Podophyllum* genotypes.

Pimer name (*Eco*RI/*Mse*I)	**Total Loci**	**Monomorphic Loci**	**Polymorphic Loci**	**Polymorphic Loci(%)**	**Number of fragments amplified**	**Resolving Power**
E+AG/M+CAG	52	2	50	96.15	727	51.929
E+TC/M+CTG	78	4	74	94.87	1175	83.929
E+TC/M+CAT	32	3	29	90.63	442	31.571
E+TC/M+ATG	26	0	26	100	287	20.500
E+AT/M+CAA	40	4	36	90	536	38.286
E+AA/M+CTC	24	1	23	95.83	279	19.929
E+TC/M+CTC	28	0	28	100	239	17.071
E+TG/M+CTG	53	8	45	84.91	605	43.214
E+AT/M+CAT	40	13	27	67.5	500	35.714
E+TA/M+CTC	27	3	24	88.89	442	31.571
E+TG/M+CAT	37	4	33	89.19	376	26.857
E+TG/M+CTC	50	12	38	76	580	41.429
E+AT/M+CTG	64	31	33	51.56	1004	71.714
Total	**551**	**85**	**466**		**7192**	

The AFLP banding pattern of 28 *P. hexandrum* genotypes is shown in the Figure 5.1(a-b) obtained with two primer combinations. The fragment size ranged from 20 bp to 750 bp

(Figure 5.1a-b). The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variation among the *Podophyllum* genotypes.

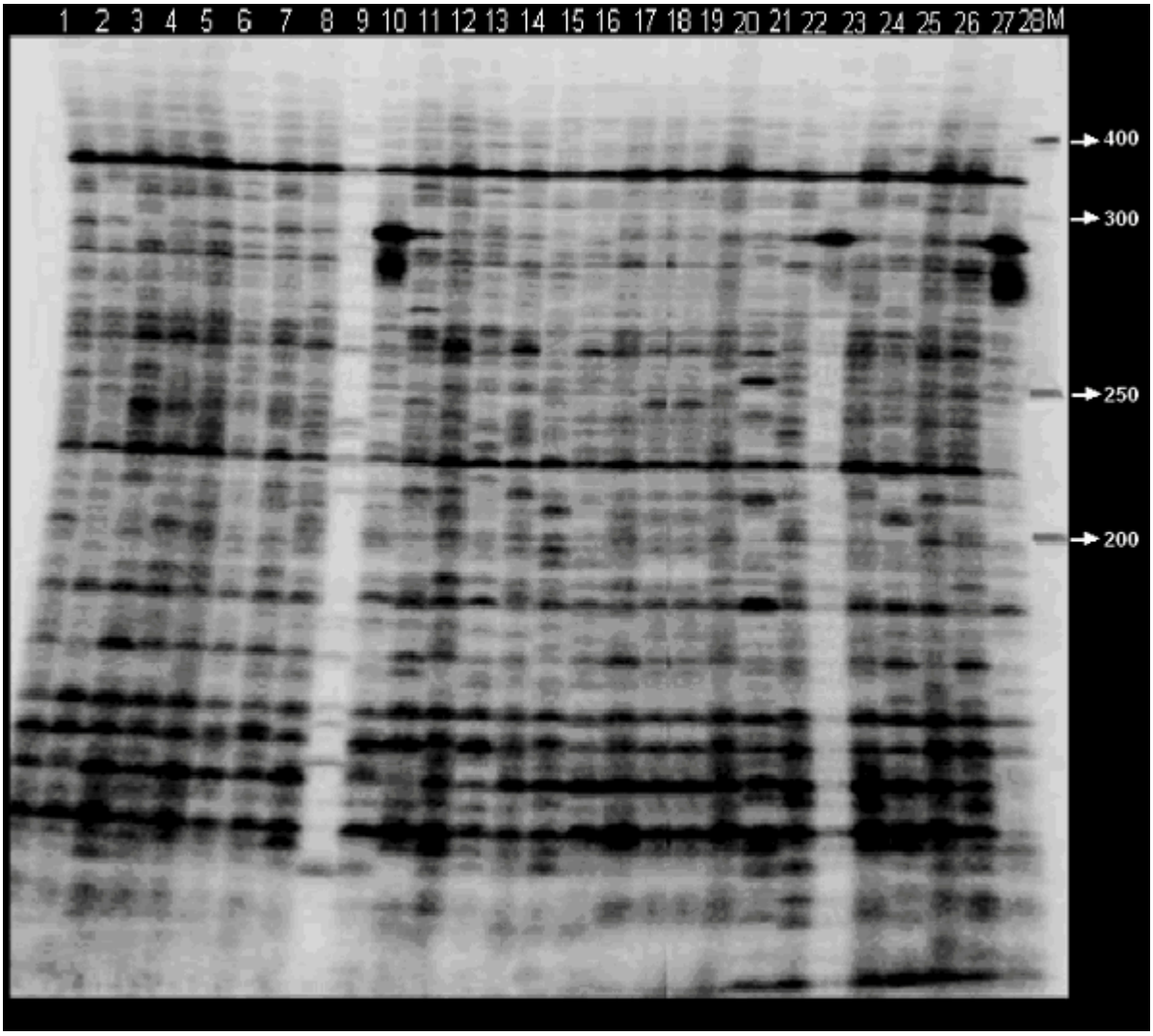


Figure 5.1(a). AFLP amplification products obtained from the 28 genotypes of *P. hexandrum* studied using E+TC/M+CTG primer combination. 1. R/4 Kasol; 2. Twin Multivora; 3. Anganoala; 4. Brundhar; 5. Gulaba; 6. ChanderKhani; 7. Kaned Nursery; 8. Sanghar Nursery; 9. Madhvi Thach; 10. Kala Pani; 11. Jalora Pass (Sojha Nursery); 12. Jalora c-30(b); 13. DPF (D-1892-C1); 14. DPF (D-791-C1); 15. Myar Valley; 16. Nayan Ghar; 17. Bada Bhangal; 18. Chota Bangal; 19. IHBT; 20. Bander Thach; 21. Saropa Nursery; 22. Nichar Nursery; 23. Rango (NC-8); 24. Sach Range; 25. Killer Range; 26. Purthi Range; 27. Ghoei DPF; 28. Samara RF. M = the size of molecular markers in base pairs using λ DNA.

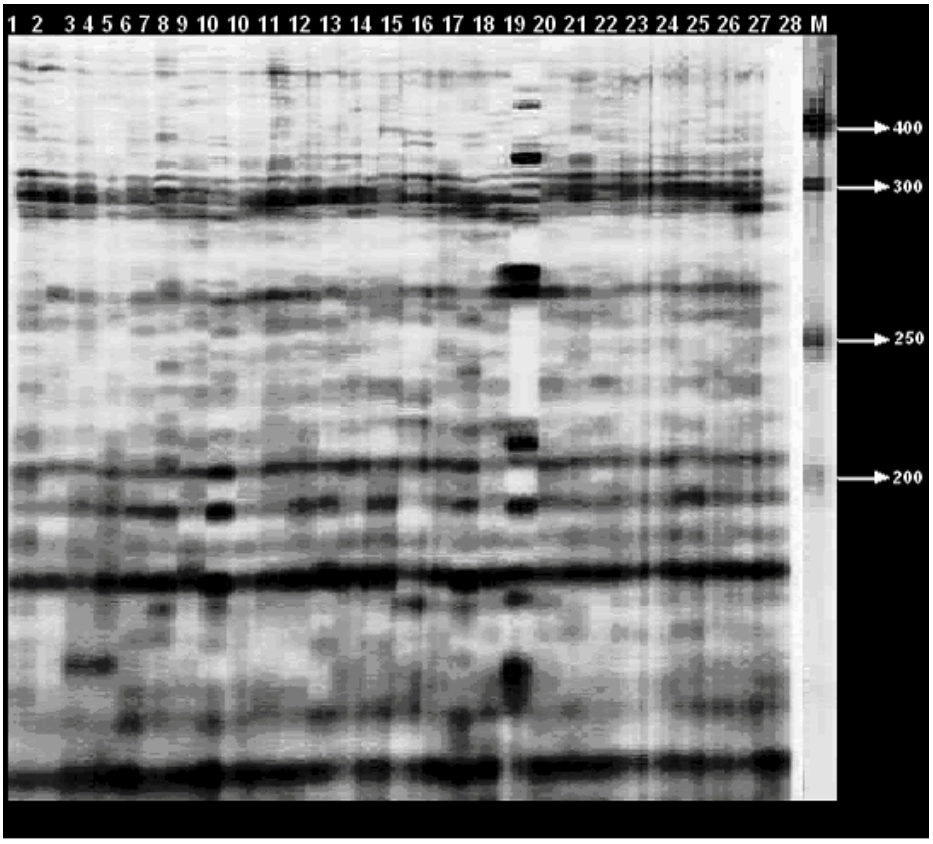


Figure 5.1(b). AFLP amplification products obtained from the 28 genotypes of *P. hexandrum* studied E+AG/M+CAG primer combination. 1. R/4, Kasol; 2. Twin Multivora; 3. Anganoala; 4. Brundhar; 5. Gulaba; 6. ChanderKhani; 7. Kaned Nursery; 8. Sanghar Nursery; 9. Madhvi Thach; 10. Kala Pani; 11. Sojha Nursery; 12. Jalora c-30(b); 13. DPF (D-1892-C1); 14. DPF (D-791-C1); 15. Myar Valley; 16. Nayan Ghar; 17. Bada Bhangal; 18. Chota Bangal; 19. IHBT; 20. Bander Thach; 21. Saropa Nursery; 22. Nichar Nursery; 23. Rango (NC-8); 24. Sach Range; 25. Killer Range; 26. Purthi Range; 27. Ghoei DPF; 28. Samara RF. M = the size of molecular markers in base pairs using λ DNA.

5.3.2 Podophyllotoxin content

Podophyllotoxin content was extracted and analyzed in triplicate from 28 populations of *P. hexandrum* distributed into 11 forest divisions at different altitudes. It was found that the podophyllotoxin content in the root of the plants obtained from the Lahaul forest division (at an altitude of 4300 m) was high (8.857 to 9.533% on dry weight basis) compared to the root samples collected from other forest divisions. The lowest values obtained were samples from Parvati Forest Division (at an altitude of 1300 m) 3.02 to 4.75% on dry weight basis (Table 5.1).

For populations in the same as well as other forest divisions, the podophyllotoxin content increased with the increase in altitude (Table 5.1). Figure 5.2 shows the correlation between the podophyllotoxin content and Nei's genetic diversity and is revealed from the plot that podophyllotoxin content is not related with genetic differentiation. However, it is related with environmental variables like temperature, rainfall, relative humidity and soil pH as discussed in chapter 6.

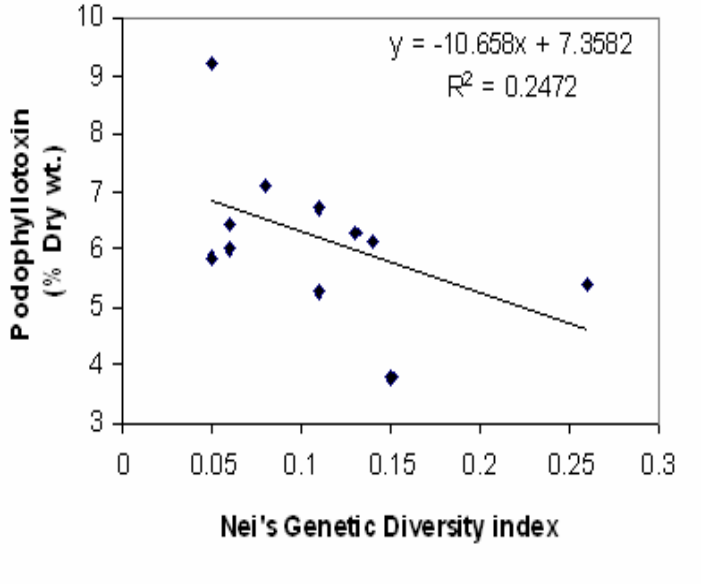


Figure 5.2. Regression analysis between podophyllotoxin content and Nei's genetic diversity index of the *P. hexandrum* populations among 11 regions, suggested podophyllotoxin content is not related with genetic differentiation.

5.3.3 Phylogenetic analysis

Based on AFLP marker, the similarity index values ranged from 0.59 to 0.90. These values were used to construct a dendrogram using unweighted pair group method with arithmetic average (UPGMA). Populations from 11 forest divisions were clustered into region-specific groups with the exception of Kullu forest division (Figure 5.3). All the 28 *Podophyllum hexandrum* populations were distributed into 14 main clusters (C1-C14). Cluster C1 represented Parvati forest division with 3 different populations namely R/4 Kasol, TwinMultivora and Anganola. The cluster C2 (Palampur forest division) and C3 (Churah) has 3 populations each- IHBT, Bada Bangal, Chota Bangal and (DPF-D-1892-C1 and DPF-D-791-C1 respectively. Genotypes from Pangi forest division were distributed within one cluster – C4 and having Sach Range, Killer Range, Purthi Range population.

Genotypes of Kullu forest division scattered with 4 clusters C5, C7, C10, and C14 having Brundhar, Sanghar Nry, Gulaba, Kaned Nry and Chander Khani repectively. The cluster C6 (Rampur forest division) and C8 (Dodrakwar forest division) having genotypes Bander Thach, Saropa Nry and Madhvi Thach, Kala Pani respectively. The clusters C9, C11, C12 and C13 comprise 8 populations each from forest division Bharmaur (Samara RF and Ghoi DPF), Seraj (Sojha nursery and Jellora C-3b), Kinnaur (Rango-NC-8 and Nichar nursery) and Lahaul (Myar Valley and Nayagarh) respectively. The results indicate high genetic diversity in *P. hexandrum* from Himachal Pradesh. Figure 5.3 shows the strict consensus tree and the internal support from the bootstrap analysis. The two methods gave very similar topologies. Most of the clusters revealed bootstrap support values > 50%, while the rest of the dendrograms showed poor bootstrap support (lower than 50%) and were unresolved.

5.3.4 Genetic diversity analysis

Wide genetic diversity between populations of *P. hexandrum* was evident from the high number of polymorphic markers. The minimum number of polymorphic varies between 57 (Lahual and Bharmaur forest division) and a maximum of 379 (Kullu forest division) including all the 13 primer pairs in combination. Polymorphisms also vary substantially within the discrete groups of plants and the minimum was found to be 10.34% (Lahual and Bharmaur forest divisions) whereas the maximum was 68.78% (Kullu forest division). The total number of polymorphic loci is 466 thereby, giving an estimate of profound (>84.40%) polymorphism. AMOVA analysis was conducted to apportion variation into (1) among groups (with respect to their forest divisions) and (2) among population within groups (Table 5.3). More than half of the total variations in the studied populations: 64% and 36% could be accounted for by difference among the forest divisions and between populations within a forest division, respectively. All components of molecular variations were significant ($P < 0.001$). The gene diversity computed among different groups of populations was recorded in between 0.05 (Bharmaur) – 0.26 (Kullu).

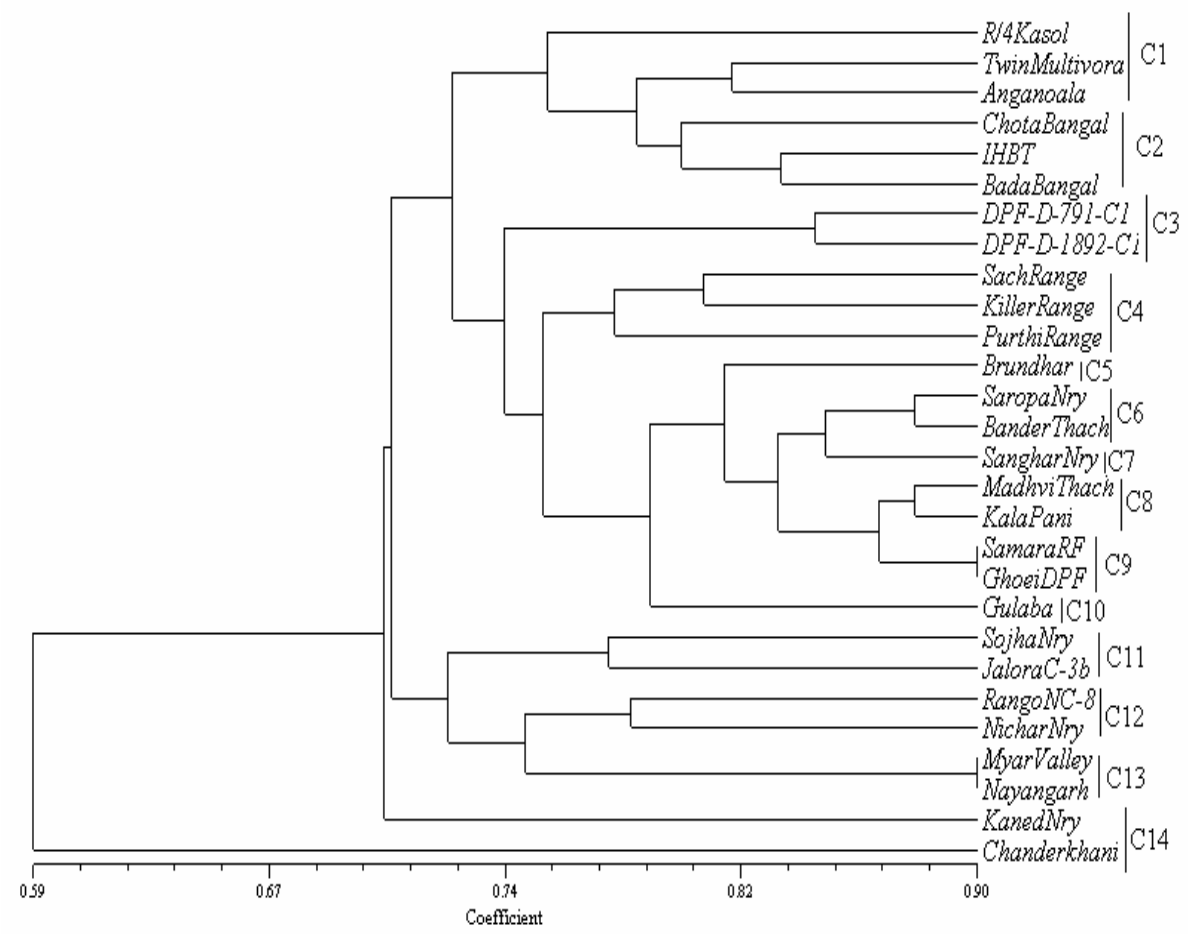


Figure 5.3. Dendrogram illustrating genetic relationships among 28 populations generated by UPGMA cluster analysis based on AFLP markers. The values in the node indicate the bootstrap values after 1000 permutations.

Table 5.3. Analysis of molecular variance (AMOVA) (levels of significance are based on 1000 iteration steps, d.f.: degree of freedom; S.S.D.: sum of square deviation; P-value: probability of null distribution).

Source of variance	d.f	S.S.D.	Variance component	Percentage	P-value
Among forest division	10	454.41	18.64	64	< 0.001
Among genotypes within forest division	17	167.86	9.87	36	< 0.001
Total	**27**	**622.27**	**28.51**		

The effective number of alleles was the minimum for Bharmaur (1.10) and Lahual and the maximum for Kullu (1.43) across the populations collected from different forest divisions. The same order of genetic heterogeneity was discerned through Shannon's information

index which varied from 0.07 (Bharmaur) and Lahaul to 0.39 (Kullu). On an all genotype basis, the observed number of alleles was 1.84 and the effective number of alleles was found to be 1.44 per locus. Similarly, the total gene diversity (Ht) among populations was 0.26 and within populations (Hs) was 0.13. Shannon's information index was 0.41 and estimated gene flow was found to be 0.24 among the 28 *P. hexandrum* populations (Table 5.4 & 5.5).

In the present study, we have used AFLPs to study genetic diversity wihin and between 28 geographically isolated populations of *podophyllum hexandrum*. Significant levels of population differentiation were found based on AFLP markers. The study indicates that *P. hexandrum* populations in the northwestern Himalayan region are genetically highly diverse. The high genetic variations in *P. hexandrum* may be attributed partially to the cross pollinated nature of *P. hexandrum*. The dendrogram and AMOVA analyses all indicated a pattern of connectivity among populations that related to the geographical distance separating them in accordance with Wright's (1931) model of isolation by distance. Although the phyllogram reconstruction resulted in a topology that reflected the geographical distribution of populations, in a few places it was only weakly supported by the bootstrap value. AMOVA analyses revealed that there was significant variation arising from habitat-correlated genetic difference (36%). Such genetic differentiation generally results from stochastic events, such as genetic drift and local selection, exacerbated by the diminishied exchange of individuals among populations (gene flow) (Wright 1931). Considering the high genetic differentiation among the wild populations of *P. hexandrum*, preservation of only a few populations may not adequately protect the genetic variation within the species in the Himalayan region. At present, the rate of propagation of *P. hexandrum* is far less than the rate of its exploitation. This species or at least a large part of its genetic diversity may be lost in the near future owing to its importance and over exploitation as a medicinal plant if appropriate conservation measures are not adopted. Since a single or even a few plants will not represent the whole genetic variability in *P. hexandrum,* there appears to be a need for maintenance of sufficiently large populations in natural habitats for conservation of its genetic diversity and avoidance of genetic erosion.

Table 5.4. Summary of genetic variation statistics for all loci of AFLPs among the *Podophyllum hexandrum* populations with respect to their distributions among eleven forest divisions.

Populations sampled from	Sample size	Observed no. of alleles (Na) (mean ± SD)	Effective no. of alleles (Ne) (mean± SD)	Nei's gene diversity (H) (mean ± SD)	Shannon's information index (I) (mean ± SD)	Ht (mean ± SD)	Number of polymorphic loci	Percentage of polymorphic loci
Parvati	3	1.34 ± 0.47	1.27 ± 0.37	0.15 ± 0.21	0.22 ± 0.30	0.15 ± 0.04	187	33.94
Kullu	5	1.69 ± 0.46	1.43 ± 0.35	0.26 ± 0.19	0.39 ± 0.27	0.26 ± 0.03	379	68.78
Dodrakwar	2	1.12 ± 0.33	1.12 ± 0.33	0.06 ± 0.16	0.09 ± 0.22	0.06 ± 0.03	68	12.34
Seraj	2	1.22 ± 0.42	1.22 ± 0.42	0.11 ± 0.21	0.15 ± 0.29	0.11 ± 0.04	123	22.32
Churah	2	1.16 ± 0.36	1.16 ± 0.36	0.08 ± 0.18	0.11 ± 0.25	0.08 ± 0.03	86	15.61
Lahual	2	1.10 ± 0.30	1.10 ± 0.30	0.05 ± 0.15	0.07 ± 0.21	0.05 ± 0.02	57	10.34
Palampur	3	1.28 ± 0.45	1.23 ± 0.36	0.13 ± 0.20	0.18 ± 0.29	0.13 ± 0.04	156	28.31
Rampur	2	1.12 ± 0.33	1.12 ± 0.33	0.06 ± 0.16	0.09 ± 0.23	0.06 ± 0.02	68	12.34
Kinnaur	2	1.22 ± 0.41	1.22 ± 0.41	0.11 ± 0.21	0.15 ± 0.29	0.11 ± 0.04	119	21.60
Pangi	3	1.32 ± 0.47	1.25 ± 0.37	0.14 ± 0.21	0.20 ± 0.30	0.14 ± 0.04	175	31.76
Bharmaur	2	1.10 ± 0.30	1.10 ± 0.30	0.05 ± 0.15	0.07 ± 0.21	0.05 ± 0.02	57	10.34

Table-5.5. Genetic variability across all the populations of *Podophyllum hexandrum.*

No. of alleles (Na)	**Effective no. of alleles (Ne)**	**Nei's gene diversity (H)**	**Shannon's Information index (I)**	**Ht**	**Hs**	**Gst**	**Estimate of gene flow(Nm) 0.25(1-Gst)/Gst**	**Total no. of polymorphic loci**	**% PPL**
1.84 (0.36)	1.44 (0.33)	0.26 (0.17)	0.41 (0.23)	0.26 (0.03)	0.13 (0.01)	0.51	0.24	466	84.40

For conservation aspects, obtaining quick, accurate estimates of the distribution of genetic variation in a cost-effective manner is particularly important. To date, most of the studies of this nature have employed RAPDs. In the present study we have used AFLPs to detect variability in common genotypes, although a parallel study on a subset of the same materials using RAPDs and ISSRs have shown that information on diversity levels and relationships between populations is congruent between all the three marker systems. The AFLP technology is extremely robust and proficient at revealing intra-population diversity and estimating genetic distance between individuals and populations (Travis et al., 1996; Arens et al., 1998; Winfield et al., 1998). Furthermore, three times the number of data points (amplification products) were generated with AFLPs compared to RAPDs over an equivalent period of time. These factors contribute to the conclusion that AFLPs provide a cost-effective procedure to monitor the extent and distribution of diversity in *Podophyllum hexandrum.*

CHAPTER 6

Impact of Soil Nutrient and Environmental Factors on Podophyllotoxin Content among 28 *Podophyllum hexandrum* Populations of the North-western Himalayan Region Using Linear and Non-linear Approach

Abstract

Podophyllotoxin is the active ingredient in the rhizome of an endangered Indian medicinal herb, *Podophyllum hexandrum*. Podophyllotoxin content in the *P. hexandrum* differs greatly in different natural habitats. In order to facilitate reasoned scientific decisions on its domestication, conservation and sustainable utilization, the effects of soil nutrients and environmental factors on podophyllotoxin content in the rhizome of *P. hexandrum* were for the first time investigated in the north-western Himalayan region, Himachal Pradesh, India. The podophyllotoxin content reached higher than 6.62% of root dry weight when soil pH value was about 4.82, soil organic carbon was higher than 3.23% and nitrogen content was higher than 2.7% of soil dry weight. However, soil available with phosphorous content higher than 0.419% and potassium content higher than 1.56% resulted in low podophyllotoxin content. The strong and linear relationship detected between podophyllotoxin as well as soil nutrients, environmental factors and altitude suggested that further optimization of these factors are very important in the conservation and exploitation of *P. hexandrum.* In this regard the prediction model like artificial neural network (ANN) and multiple linear regression (MLR) developed in this study to map the effect of these factors on podophyllotoxin yield will be of great help. The ANN prediction model revealed better prediction of yield (r^2 = 0.9905) than MLR prediction model (r^2 = 0.9302). Lower level of root mean square error (RMSE) for ANN model (0.0399) than MLR model (0.2939) with respect to the experimental measurement establishes the ANN method as an efficient tool for optimization of soil nutrients and climatic factors for podophyllotoxin yield.

6.1 Introduction

The Himalayan region is home to numerous highly valued medicinal herbs including *Podophyllyum hexandrum* Royle (Berberidaceae) also known as Indian Mayapple that has the endangered status in India. It is distributed in very restricted pockets in the Himalayan zone at altitudes ranging from 1300 to 4300 m from the sea level. It is recognized for its anti-cancer properties. The rhizomes and roots of *P. hexandrum* contain anti-tumor lignans such as podophyllotoxin, 4'-dimethyl podophyllotoxin and podophyllotoxin 4-o-glucoside (Imbert, 1998) which has long been used by the Himalayan natives and the American Indians as a canthartic cholagog. Among the plethora of physiological activities and potential medicinal and agricultural applications, the antineoplastic and antiviral properties of podophyllotoxin congeners and their derivatives are arguably the most important from the pharmacological point of view. Semisynthetic derivatives of epipodophyllotoxin, e.g. etoposide (VP-16) (Allevi et al., 1993), etopophos (Schacter, 1996) and teniposide (VM-26) are effective agents in the treatment of lung cancer, a variety of leukemia and other solid tumors (Van Uden et al., 1989). Growing demand in the world for anti-cancer drugs has added much to the importance of podophyllotoxin.

The total synthesis of podophyllotoxin is complicated due to the presence of four chiral centers, a rigid *trans*-lactone and an axial 7-aryl substituent (Gordaliza et al., 2004). Hence, *P. hexandrum* (an Indian species) and *P. peltatum* (an American species) are presently the commercial source of podophyllotoxin for the pharmaceutical industry. However, the yield of podophyllotoxin from *P. peltatum* is low (~ 0.25% based on dry weight) in comparision to *P. hexandrum* which contains ~ 4% of podophyllotoxin by dry weight (Jackson and Dewick, 1984). The demand for the compound continues to increase and thus encourages domestication and conservation of *P. hexandrum* in the Himalayan region. In an attempt at ex situ conservation Sharma et al., (2000), collected root samples of *P. hexandrum* from Jalori Pass and Khajjiar (from high altitude) and grew at Palampur (low altitude). This led to reduction in podophyllotoxin content in the plant sample. Conservation strategies adopted at Y.S. Parmar, University of Agricultural Science, Nauni, Himachal Pradesh, India, further resulted in reduction in podophyllotoxin content. The plant secondary metabolism is a rather complex physiological process and is affected by many

environmental factors. A number of investigations have demonstrated that the quality and quantity of several secondary metabolites have close relationship with plant habitats (Endress, 1994). Soil is essential for the growth and metabolism of plants as it provides nutrients (nitrogen (N), phosphorus (P), potassium (K), sulphur (S)) and metal elements. The shift of nutrition supplies in soil definitely leads to the alteration of both primary and secondary metabolism and consequently results in changes in productivity of the secondary metabolites. However, in addition to soil factors the effect of climatic factors on podophyllotoxin content at natural habitats can not be ruled out. This is obvious as the plant is quite adaptable to a wide range of environmental conditions. It can survive under varying growing conditions and adapt well from the extreme low winter temperature of the northern climates to the high summer temperatures and at altitudes ranging from 1300 to 4300 m. Therefore, planting of *P. hexandrum* plants at lower altitudes may not be useful because of the low amount of podophyllotoxin. Further, we have been able to prove that the podophyllotoxin content is not related to genomic variation (Alam et al., 2008). Hence, it demands management of soil nutrients and optimization of climatic factors for the successful conservation of *P. hexandrum*.

Statistical methods such as artificial neural network (ANN) and multiple linear regressions (MLR) are very useful in this respect. MLR has been used to explain the spatial variations in soil nutrients and its impact on crop yield at field scale (Sudduth et al., 1996). However, MLR requires a normal distribution of the input variables which is not always the case (Atkinson et al., 1997), the non-linear predictors such as ANNs have been used to solve various problems in agriculture. For example, Sudduth et al., (1998) successfully predicted corn yield with back propagation neural network models based on soil texture, topography, pH and some soil nutrient elements. This prediction model was superior to those of the nonparametric statistical benchmark methods. To the best of our knowledge no work has been reported to have mapped the yield of podophyllotoxin from *P. hexandrum* populations with respect to soil nutrients and climatic factors.

Artificial neural networks can be used to develop empirically based agronomic models. The ANN structure is based on the human brain's biological neural processes.

Interrelationships of correlated variables that symbolically represent the interconnected processing neurons or nodes of the human brain are used to develop models. ANNs find relationships by observing a large number of input and output examples to develop a formula that can be used for predictions (Pachepsky et al., 1996). Non-linear relationships overlooked by other methods can be determined with a little a priori knowledge of the functional relationship (Elizondo et al., 1994). A minimum of three layers is required in an ANN: the input, hidden and output layers. The input and output layers contain nodes that correspond to input and output variables respectively. Data move between layers across weighted connections. A node accepts data from the previous layer and calculates a weighted sum of all its inputs, t:

$$t_i = \sum_{j=1}^{n} w_{ij} x_j \qquad (1)$$

where n is the number of inputs, w is the weight of the connection between node i and j and x is the input from node j. A transfer function is then applied to the weighted value, t, to calculate the node output, $o_{i..}$

$$o_i = f(t_i) \qquad (2)$$

The most commonly used transfer function is a sigmoidal function for the hidden and output layers and a linear transfer function is commonly used for the input layer. The number of hidden nodes determines the number of connections between inputs and outputs and may vary depending on the specific problem under study.

The objective of present study is to examine the contents of soil nutrient factors such as soil organic matter, pH value, total N, total P and total K as well as different climatic factors like high temperature, low temperature, rain fall, relative humidity and altitudes to analyze their relationships with podophyllotoxin contents of *P. hexandrum* populations in the Himalayan region. The results obtained and the prediction model developed would guide the soil management and optimization of environmental and geographical factors for the domestication, conservation and sustainable utilization of *P. hexandrum* at commercial scale from the Himalayan region. The effects of soil nutrient and environmental factors on the podophyllotoxin production in the root of *P. hexandrum* will be investigated for the first time.

6.2 Materials and Methods

6.2.1 Sample stations and plant materials

Populations of *Podophyllum hexandrum* were sampled from 28 sites covering 11 forest divisions at different altitudes (1300 – 4300m) from the north-western Himalayas, Himachal Pradesh, India in July and August, 2006. From each site representative plant samples were collected in triplicate. The interval among replicates was 2-5 m. Each patch was considered one accession and coded according to the site of collection. A total of 8-10 plants were collected form each sampling site. The pair wise distance between populations within a forest division was 0.5 – 32 Km whereas the pair wise distance between forest divisions was 10 – 400 Km (Figure 6.1).

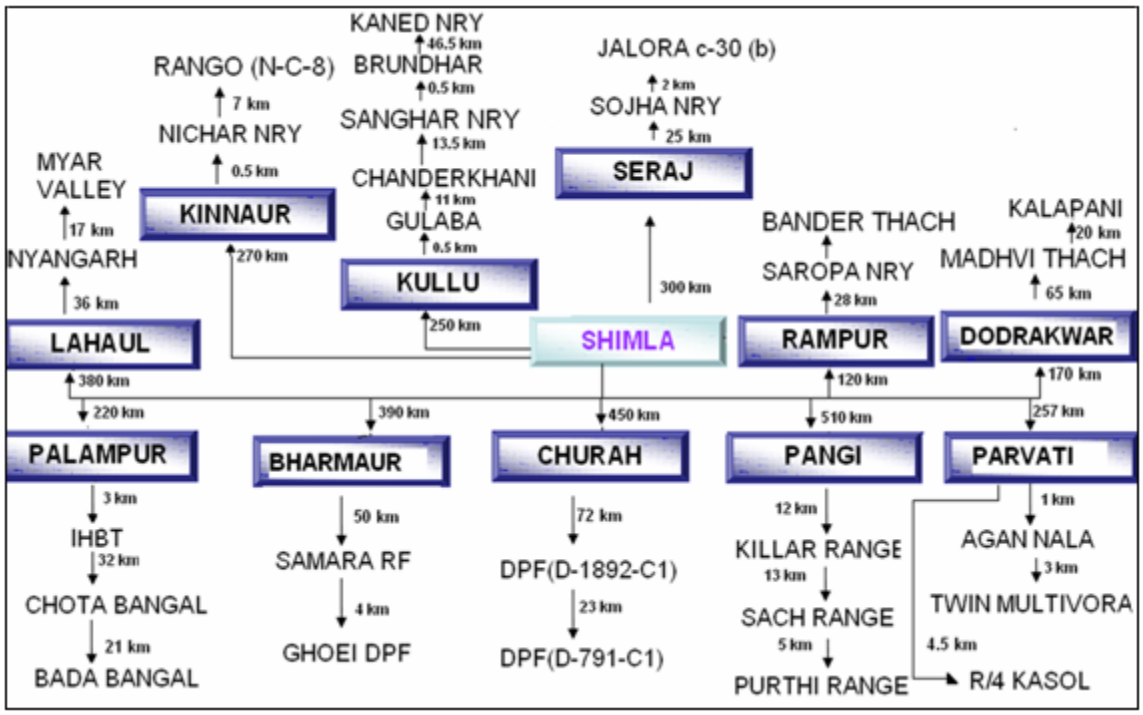


Figure 6.1. The difference in distance (Km) between sampling sites of respective forest divisions which is measured from a centre point Shimla.

The fresh roots of the uprooted plants were trimmed and washed with running tap water to remove the soil particles followed by washing with double distilled water. The washed roots were then dried separately at 60 ^{0}C for 24 hrs in an oven and used for podophyllotoxin estimations. To diminish the effects of the age differences on podophyllotoxin content, only 3 year old plants were selected.

Soil samples (root layer) from each sampling site were collected in triplicate after uprooting the plants and brought to the laboratory for analysis of soil nutrients. The soils in the Himalayan region originated from weathering of sedimentary rocks technically known as podzol. The soil is mostly rocky mixed with humus. However, the texture of the soil may be sandy loam, mica shift, loamy or rock clay state. Since the sampling was done from the wild there was no human intervention. The forest types were generally subalpine to alpine. Therefore, temperate conifers and temperate broad leaf at altitudes between 1300m and 3000m were found. However, the forest types like subalpine to alpine pasture and temperate conifers were found above 3000m altitude revealing wide variation in topography and geographical variation in the Himalayan region. From each sampling site the monthly average data on environmental factors such as temperature, humidity and rain fall were taken from January to December, 2006 and were properly documented.

6.2.2 Extraction of podophyllotoxin and quantification

Dried roots were ground to powder in a mortar with a pestle. Podophyllotoxin was extracted following the procedure of Broomhead et al., (1990) and estimated using HPLC. All the experiments on extraction of podophyllotoxin and HPLC analysis were repeated three times.

6.2.3 Quantitative analysis of soil nutrition

The soil samples were collected in triplicate from each site. About 200g of soil was collected from the root level and was air-dried to a constant weight and then sieved through a 2 mm-mesh. The fine soil (particles <2 mm) was used for nutrient analysis. Soil water pH was determined by dissolving 5g of air-dried soil sample into 5ml of SMP buffer (12.9 mM paranitrophenol, 15.4 mM of K_2CrO_4, 0.361 M of $CaCl_2.2H_2O$, 12.6 mM of $Ca(OAc)_2$, adjusting to pH 7.50 with 15 % NaOH) and measured pH value with pH meter (Mehlich, 1976; Soil and Plant Analysis Council, 1992). The soil organic matter was determined by measuring organic carbon content according to the wet-oxidation procedure described by Mebius (1960). Total nitrogen was estimated by Kjeldahl digestion through steam distillation. The resulting ammonium was converted into boric acid and titrated with 0.10 N or 0.02 N HCL to pH 4.6 using automatic titrator. This method measures both organic and

inorganic forms of nitrogen which were reported as dry weight percent (Rump and Krist, 1992; Kimble et al., 1993). Total phosphorus (organic & inorganic) was determined by perchloric acid digestion (Olsen and Sommers, 1982). Total potassium was analyzed by sodium hydroxide digestion and estimated by atomic absorption spectrometry.

6.2.4 Statistical analysis

The correlation and regression analysis between the podophyllotoxin content and soil nutrients, environmental factors and altitude were examined by using MINITAB statistical package.

6.2.5 Neural network data-mapping model development

Neural networks are known as useful tools for pattern recognition, identification and classification. A neural network model can determine the input–output relationship for a complicated system based on the strength of their interconnection presented in a set of sample data (Howard and Mark, 2000). Such a model can provide data approximation and signal-filtering functions beyond optimal linear techniques (Clifford and Lau, 1992). Therefore, neural-network models provide more robust results for complicated system analysis than conventional mathematical models. In this study, a back propagation neural-network model was created using Stuttgard Neural Network Simulator package (SNNS version 4.2) and trained using the environmental factors and soil nutrition parameters as the inputs and the measured corresponding podophyllotoxin reading as the output. The topological structure of this neural network model consisted of 13 input neurons in the input layer and one output neuron in the output layer to match the 13-1 input–output pattern of the training data set. One hidden layer with eight neurons was the optimal topology for the neural-network model determined by a trial-and-error method (Figure 6.2).

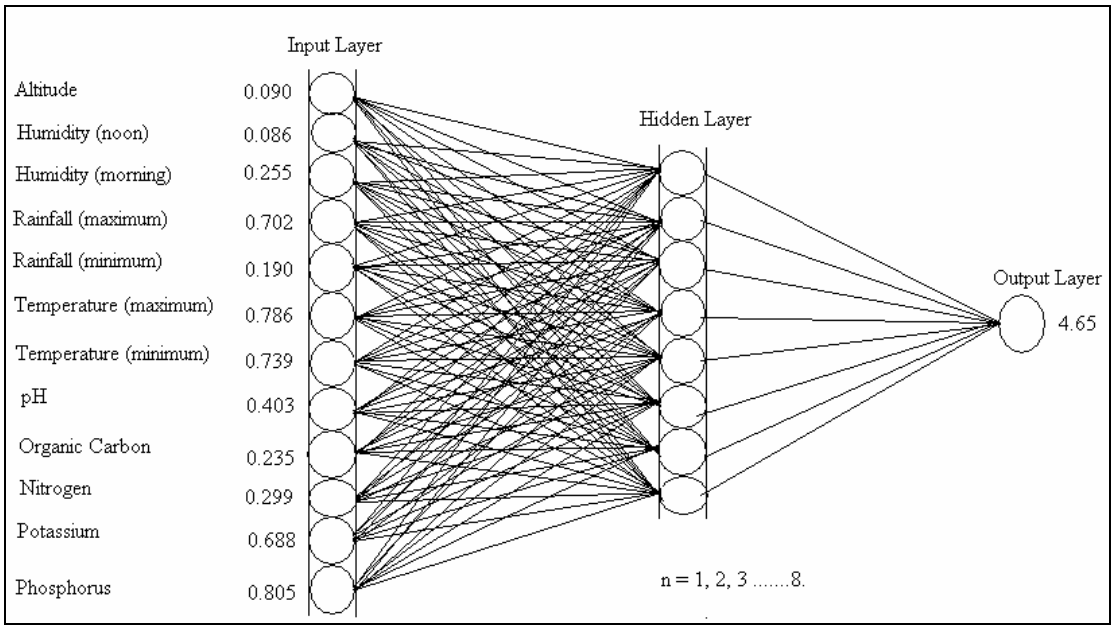


Figure 6.2. Layers and connection of a feed-forward back propagating artificial neural network.

The evaluation criterion for determining the optimal topology was the best correlation value of the training set. The neural-network model was trained in an iterative training process using the obtained training set as follows:

D_T,i = {0.090 0.086 0.255 0.702 0.190 0.786 0.739 0.403 0.235 0.299 0.688 0.805 0.266}

The first value referred to altitude, the next six numbers are the climatic factors and the next five values belong to soil nutrition parameters and the last number is the corresponding average podophyllotoxin content measured from the roots collected from the corresponding site. To avoid possible bias, the order of input–output data pair in a training data set was randomized before the training process. During the training process, the back propagation (BP) training algorithm compares the estimated output value with the target value (namely the measured value). Then it tunes weighting values connecting all the neurons to minimize the difference between the estimated and the target values until the error is smaller than a predefined level or the number of the iteration reaches a preset maximum number. The constructed model was trained with the input data for an epoch of 10,000 with 0.1 learning rate. After completing the training process all weighting indices that describe the interconnection strengths between neighboring neurons are fixed and the neural-network

model will then be capable of mapping input variables to an estimated output promptly and accurately.

The neural-network model developed here was applied to sigmoid transfer function to compute the strength of interconnection between each pair of neurons. The input variables in this model were normalized basing on their possible ranges to avoid data saturation with the help of the following equation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (3)$$

where x, x_{min}, x_{max} and x_{norm} are the real-valued input variables, the minimum and maximum possible value of the input variables and their normalized value respectively. The output from this neural-network model is an indexed value that corresponds to the input variables. To get the real-valued output, the indexed output value needs to be denormalized according to the following equation:

$$y = y_{norm}\,(y_{max} - y_{min}) + y_{min} \qquad (4)$$

where y, y_{min}, y_{max} and y_{norm} are the real-valued output variable, the minimum and maximum possible value of the real-valued output and the indexed output value from the neural-network model.

6.3 Results

6.3.1 Podophyllotoxin content in the root of *P. hexandrum*

Podophyllotoxin content was extracted and analyzed in triplicate from 28 populations of *P. hexandrum*. This investigation guides selection of soil type, altitude height and environmental factors for the cultivation of *P. hexandrum* to improve the podophyllotoxin content. It was found that the podophyllotoxin content in the root of plants obtained from Lahaul forest division was comparatively more (8.857 to 9.533 % on dry weight basis) than that in the root samples collected from other forest divisions with a minimum from Parvati (3.020 to 4.753% on dry weight basis (Table 6.1). The variation in podophyllotoxin content was found to be significant among the populations ($F = 17.22$, $P < 0.001$) as well as among the forest divisions ($F = 3.70$, $P < 0.009$).

Table 6.1. Twenty eight populations of *Podophyllum hexandrum* collected from different sites at different altitudes covering eleven forest divisions and their podophyllotoxin content.

Name of Forest Division	Sampling site	Altitude (m)	*Podophyllotoxin (% dry weight) (Mean ± sd)	Predicted value using ANN	Predicted value using MLR
Parvati	Twin Multivora	1300.00	3.567 ± 0.747	4.65	4.286
	R/4,Kasol(C-II-a-Nry)	1570.00	4.753 ± 0.796	3.595	3.495
	Anganoala (R/9) Rajgiri	1300.00	3.020 ± 0.524	3.054	3.547
Dodrakwar	Madhvi Thach	3048.00	6.207 ± 0.743	4.315	4.209
	Kala Pani	2743.20	5.800 ± 0.212	5.864	5.587
Churah	DPF(D-1892-C1) (Chaoundi)	3750.00	8.487 ± 0.565	4.03	4.357
	DPF(D-791-C1	2700.00	5.753 ± 0.411	4.643	4.569
Seraj	Jalora Pass (Sojha Nry)	2667.00	6.607 ± 0.348	7.975	7.944
	Jalora c-30(b)	2473.20	6.790 ± 0.855	6.395	6.802
Lahaul	Myar Valley	4300.00	9.533 ± 0.484	5.625	5.860
	Nayan ghar	4300.00	8.857 ± 0.427	6.615	6.559
Kullu	Brundhar	1916.00	4.077 ± 0.270	6.692	5.786
	Gulaba	2895.00	5.943 ± 0.591	8.348	8.343
	ChanderKhani	3352.80	8.033 ± 0.454	5.713	6.209
	Kaned Nry	2150.00	4.657 ± 0.850	8.913	9.133
	Sanghar Nry	2100.00	4.173 ± 0.276	9.511	9.427
Palampur	Bada Bangal	2895.00	7.097 ± 0.797	5.055	5.930
	Chota Bangal	2700.00	6.573 ± 0.827	7.115	6.655
	IHBT	2800.00	5.183 ± 0.780	6.627	6.541
Rampur	Bander Thach	2895.00	6.773 ± 0.640	6.348	5.853
	Saropa Nry	2499.40	6.097 ± 0.942	5.864	5.893
Kinnaur	Nichar Nry	2190.00	4.760 ± 0.291	4.568	4.897
	Rango (N-C-8)	2710.00	5.797 ± 0.552	5.648	5.842
Pangi	Sach Range	2712.70	6.133 ± 0.216	6.307	5.967
	Killer Range	2850.00	5.967 ± 0.692	5.769	6.221
	Purthi Range	2900.00	6.233 ± 0.790	6.314	6.470
Bharmaur	Ghoei DPF	2080.00	5.700 ± 0.692	5.686	6.040
	Samara RF	2590.80	6.030 ± 0.825	6.243	6.133

*Podophyllotoxin content varied significantly among 28 populations, $F=17.22$, $P< 0.001$; as well as between 11 forest divisions, $F = 3.70$, $P < 0.009$.

6.3.2 Effect of Altitude

All the 28 sites chosen for sampling of *P. hexandrum* populations were at different geographical locations with altitude ranging from a minimum of 1300 m (Parvati forest division) to a maximum of 4300 m (Lahaul forest division) (Table 6.1). The podophyllotoxin content in the root sampled from these sites increased progressively from

low altitude to high altitude. Figure 6.3 shows the linear regression analysis between the altitude and podophyllotoxin content including all the sampling sites. The respective correlation coefficient (r) was 0.928 and reached statistical significance level ($P < 0.001$) which is a symbol of accelerated correlation and indicates that podophyllotoxin production has been significantly favored at increased altitude.

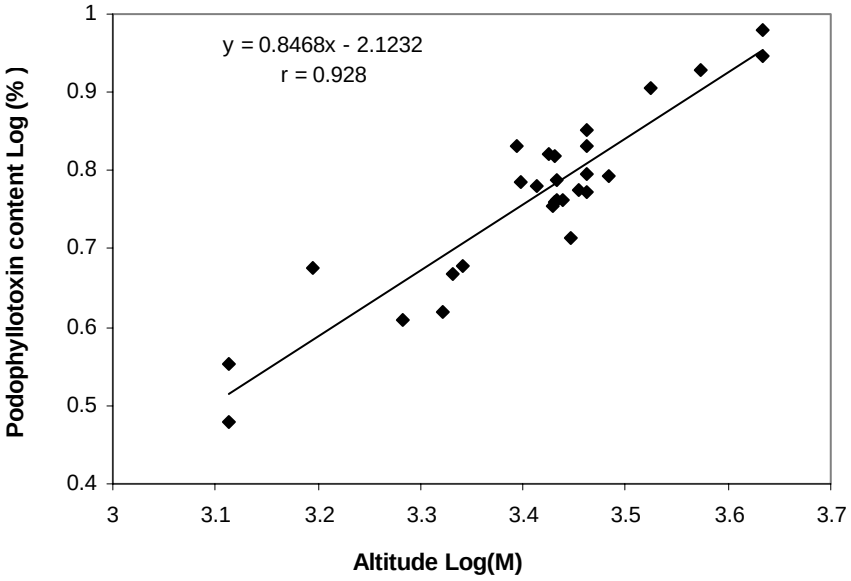


Figure 6.3. Regression analysis based on Log_{10} Podophyllotoxin content and Log_{10} M (altitude) between 28 populations of *P. hexandrum*

6.3.3 Effect of environmental factors on podophyllotoxin content

The environmental factors recorded during the course of the experiment are given in Figure 6.4. It revealed a wide range in climatic factors like temperature (minimum & maximum), rainfall (minimum & maximum) and relative humidity (forenoon & afternoon) among the different sites from where the samples were collected. At these sites the minimum temperature ranges from 2 ^{0}C to -10 ^{0}C, maximum temperature ranges from 12 ^{0}C to 35 ^{0}C, rainfall varies from a minimum of 0.00 cm to a maximum of 400 cm whereas relative humidity varies from 30% to 90% in the forenoon and 60% to 20% in the afternoon.

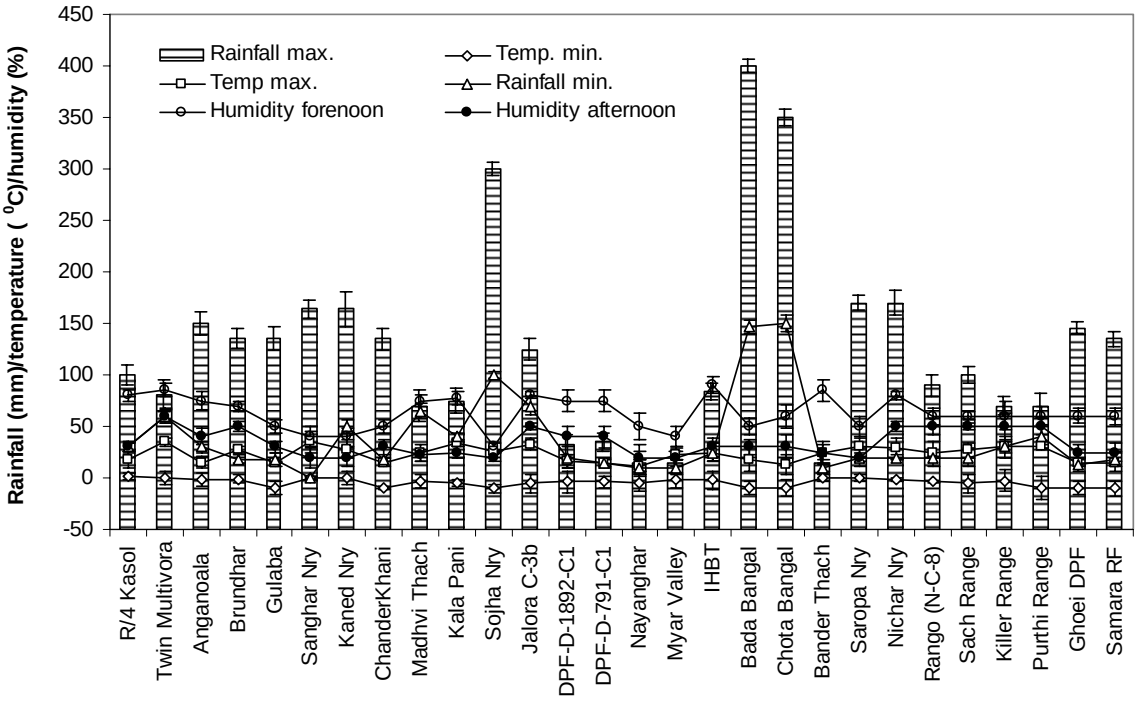


Figure 6.4. The meteorological observations have been made during the course of experiment with respect to the site of collection of *Podophyllum hexandrum* populations. The values are the average of data collected in each site from January to December 2006.

It was seen that the variation in podophyllotoxin content in the root of *Podophyllum* is highly dependent on these climatic factors. The variation in podophyllotoxin content was seen to be related positively to humidity; r = 0.825 (at afternoon) and r = 0.844 (at forenoon) and it reached statistical significance level $P < 0.001$ (Figure 6.5a and 6.5b). The correlation coefficient between podophyllotoxin content was -0.595 (significant at $P < 0.01$) with maximum rainfall and 0.717 (significant at $P < 0.001$) with minimum rainfall (Figure 6.5c, 6.5d). The linear correlation coefficient (r) was -0.720 for maximum temperature (significant at $P < 0.001$) and -0.635 ($P < 0.001$) for minimum temperature and are negatively correlated with podophyllotoxin content (Figure 6.5e, 6.5f).

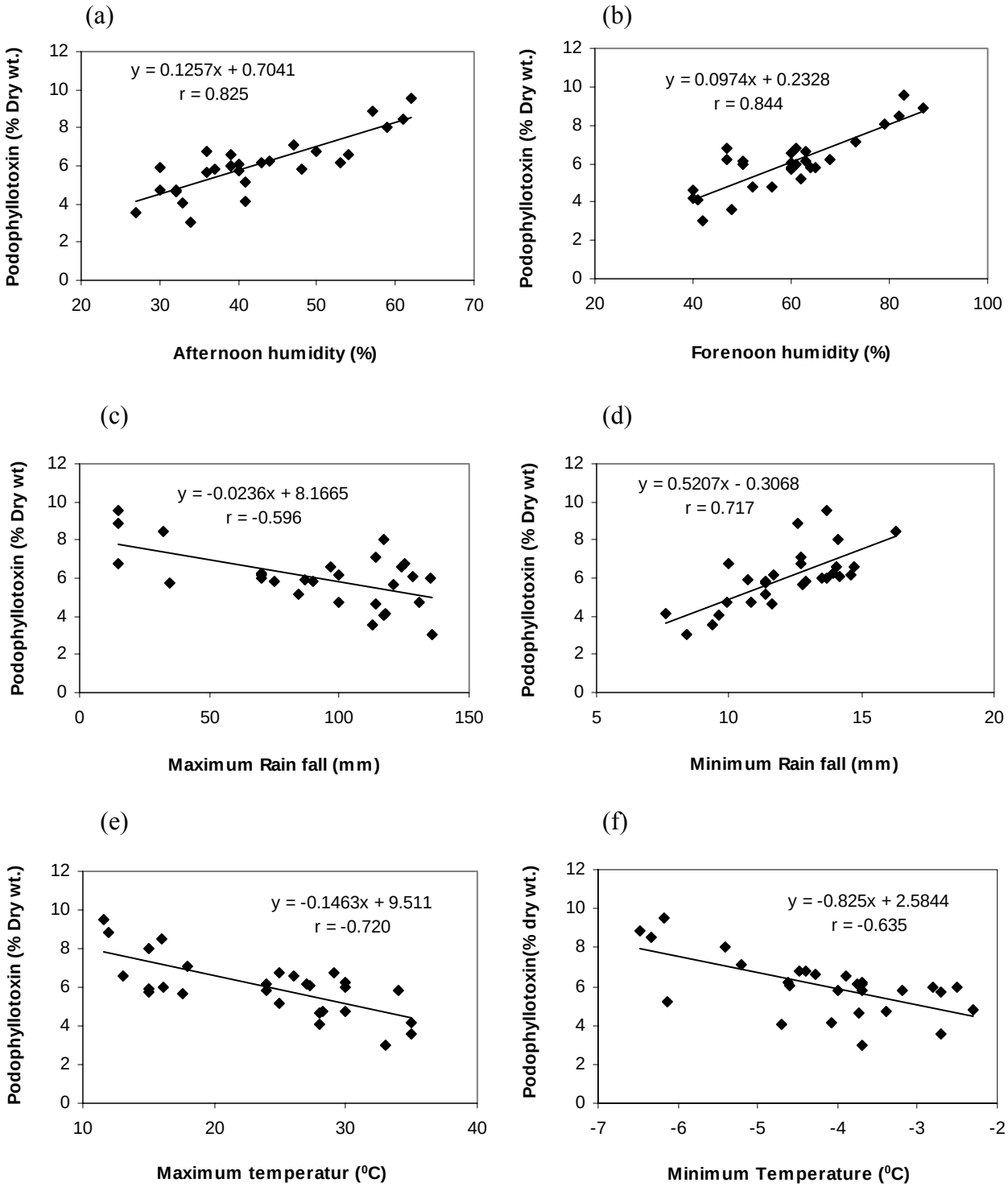


Figure 6.5. (a-f). The relationship between environmental variables humidity % (afternoon and forenoon), rainfall (maximum and minimum), temperature ^{0}C (maximum and minimum).

6.3.4 Effect of soil organic carbon (C)

According to our investigation, soil organic carbon content of respective sites mainly ranged between 2.26 % (Anganoala) and 8.07 % (Myar Valley). Figure 6.6a & 6.6b show the statistical analysis results of the linear regression between soil organic carbon and podophyllotoxin contents in the root of *Podophyllum hexandrum* of all populations between the altitudes of 1300 m to 2700 m (low altitude) and 2710 m to 4300 m (high altitude). In both the groups the correlation coefficients (r) were larger than 0.660 and are statistically significant at $P < 0.001$ (Figure 6.6a and 6.6b). The podophyllotoxin content in the root reached 5.182 % on an average in the soil organic carbon content of 3.23 % (on average). However, increase in the organic carbon with respect to altitudes higher than 3.23 % revealed an increase in podophyllotoxin content upto 6.86 % (on average). The results demonstrated that high soil organic carbon significantly favored podophyllotoxin production in the root of *P. hexandrum* found at the altitude higher than 2700 m in Himachal Pradesh.

6.3.5 Effect of soil pH

Figure 6c and 6d show that the soil water pH value at the altitudes above 1300 m in Himachal Pradesh ranges from 4.04 to 6.92. Therefore, *P. hexandrum* grew well in acidic soil condition. The statistical analysis demonstrated that podophyllotoxin contents in the root of *P. hexandrum* grown at varying soil pH ($pH < 5.5$ and $pH > 5.5$) values had significantly negative linear correlationship in both the population groups. The correlation coefficient (r) in both the groups was larger than -0.475 and reached statistical significance levels ($P < 0.05$) (Figure 6.6c and 6.6d). Podophyllotoxin contents reached about 6.62 % (on average) when the soil pH value was 4.82 (on average) whereas, the podophyllotoxin content decreased to 0.93 % when the soil pH value was higher than 5.5. This revealed that the acidic pH in the soil favored the podophyllotoxin accumulation in the roots of *P. hexandrum*.

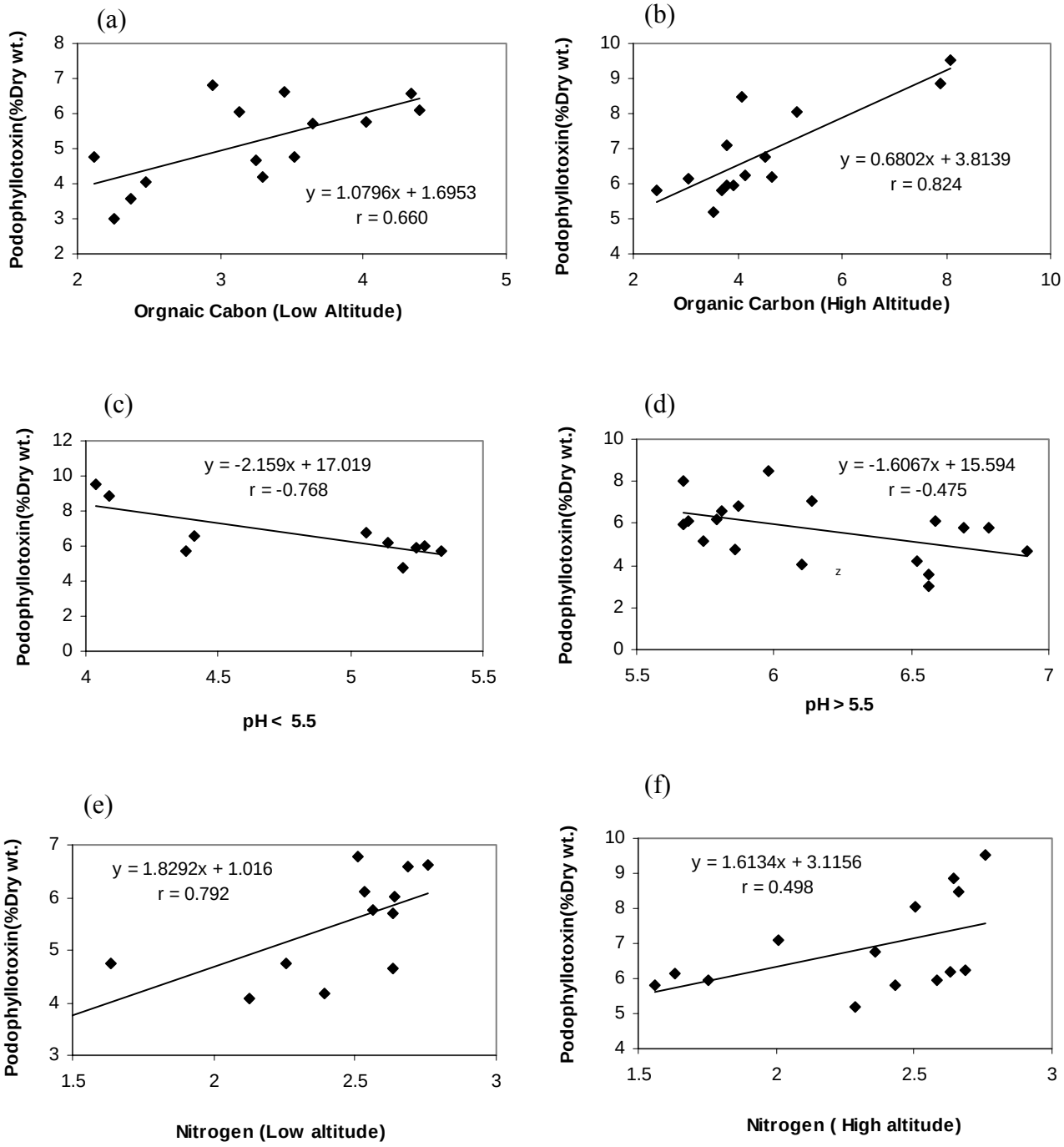


Figure 6.6. (a-f) The relationship between soil organic carbon, pH, nitrogen contents and podophyllotoxin contents in the root of *Podophyllum hexandrum* at different altitude in northwestern Himalayan region. Each pattern was classified according to low and high altitude.

6.3.6 Effects of soil nitrogen (N)

In Himachal Pradesh from the altitude of 1300 m to 4300 m the total soil nitrogen contents range between 1.25 % and 4.16 %. At the altitude above 1300 m, the accumulation of podophyllotoxin in the root increased with the increase of total nitrogen content in the soil. The linear correlation coefficients (r) were 0.792 (among populations at altitude 1300 m to 2700 m) and 0.498 (among populations at 2710 m to 4300 m) and both reached the statistical significance levels ($P < 0.001$and $P < 0.05$) respectively (Figure 6.6e and 6.6f). The podophyllotoxin content reached 6.86% (on average) when the soil nitrogen content was higher than 2.27%.

6.3.7 Effect of phosphorus (P)

Among 28 sampling sites, total phosphorus content in the soil ranged from 0.106 % to 0.24 %. Figure 6.7a and 6.7b shows the statistical results of the relationships between soil phosphorus contents and podophyllotoxin contents in roots of *P. hexandrum.* The linear correlation coefficients (r) were -0.725 and -0.916 between both the groups of populations at altitudes of 1300 m – 2700 m and 2710 m – 4300 m and reached significant levels of the negative linear relationship ($P < 0.001$). The results demonstrated that with the increase of soil phosphorous content above 0.149 % inhibited podophyllotoxin accumulation in the root.

6.3.8 Effects of soil potassium (K)

Total soil potassium contents range between 1.08 % and of 2.22 % (Figure 6.7c and 6.7d). The correlation coefficients (r) of the negative linear relationship between podophyllotoxin content and potassium content is -0.709 and -0.758 at the altitudes of 1300 m to 2700 m and 2710 m to 4300 m respectively and reached a significant level ($P < 0.001$). This indicates that total soil potassium content also had a significant relationship with the podophyllotoxin production in the root of *P. hexandrum.*

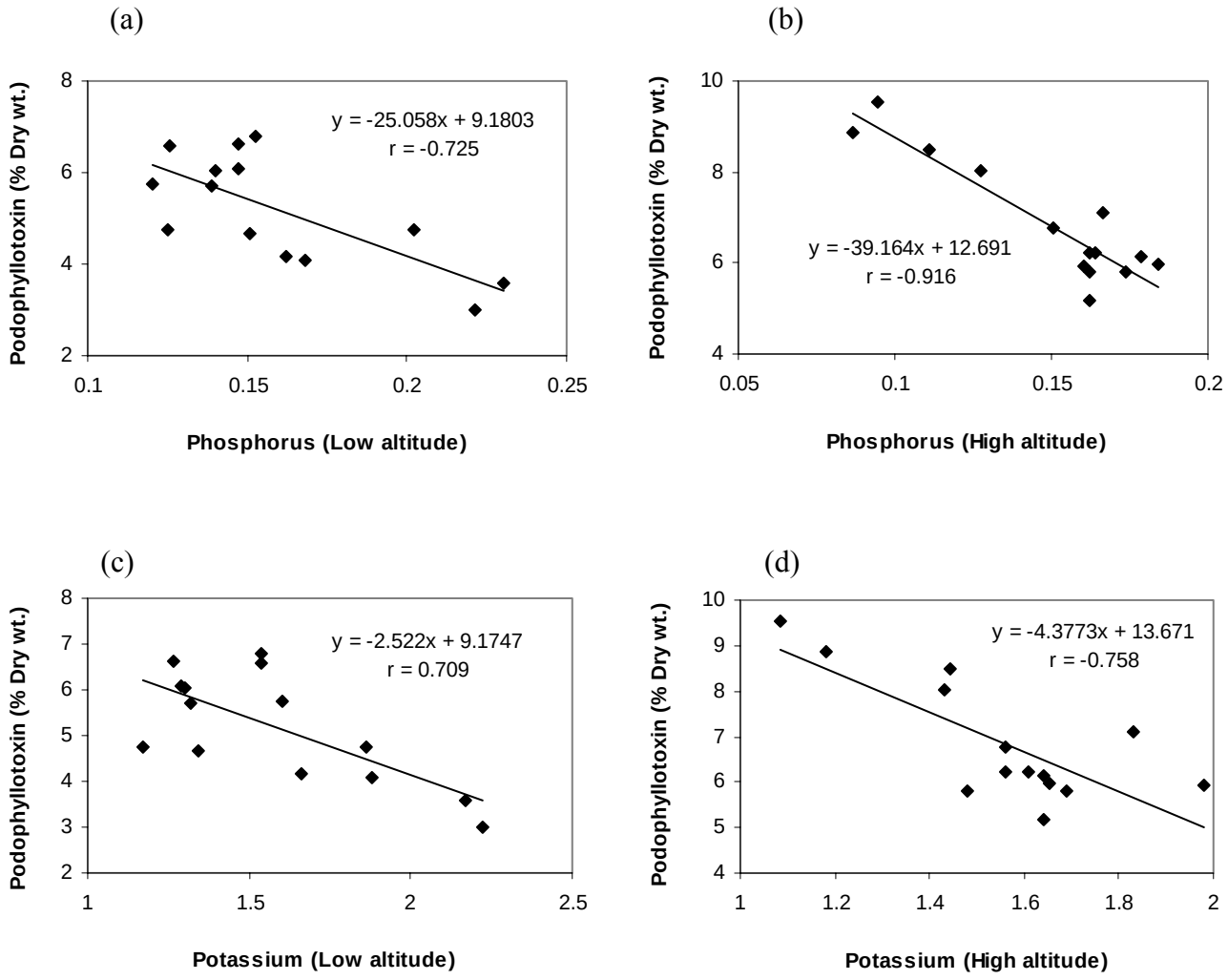


Figure 6.7. (a-d) The relationship between phosphorus, potassium contents and podophyllotoxin contents in the root of *Podophyllum hexandrum* at different altitudes in the northwestern Himalayan region. Each pattern was classified according to low and high altitude.

6.3.9 Performance measure of ANN and MLR model

The prediction results of the ANN model determined the prediction phase which is represented in Table 6.1. It revealed better prediction of yield ($r^2 = 0.9905$) in comparison to the experimental measurement and the Root Mean Square Error (RMSE) is very low (0.0399). To assess the performance of the neural-network model more thoroughly, a comparative study between the neural-network model and a baseline regression model was made. The baseline model was a best-fit regression model obtained from similar parameters that has been used for ANN prediction. Regression analysis of the best-fit model estimated values resulted in r^2 of 0.9302. Comparing the r^2 values from estimated podophyllotoxin

values using both the neural-network model and the best-fit regression model including the exponential model, the neural network model showed more strong correlation than the other prediction models (Figure 6.8). Similarly, the RMS error was 0.2939 from the 'best-fit' model in comparison to 0.0399 from the neural-network model. Comparing the results (the r^2 value and RMS error) it is verified that the neural-network model can provide more accurate estimations of podophyllotoxin values than a 'best-fit' regression model. More importantly, a well trained neural-network model can be implemented in real-time to estimate the required podophyllotoxin content in terms of changed environmental and soil factors with a minimal computational load.

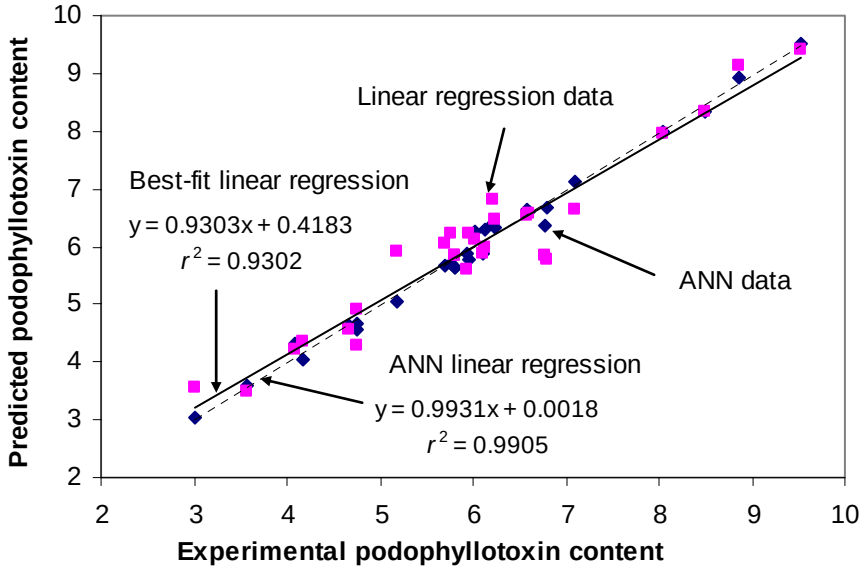


Figure 6.8. Comparison of estimated podophyllotoxin content using an artificial neural-network model (---, ANN) and a 'best-fit' regression model (—, MLR). Here r^2 represent coefficient of determination.

6.4 Discussion

The existing variations in podophyllotoxin content were proved to be coupled with altitude, environmental variables and soil nutritional factors. These results have been very much supported by the studies done by Sharma et al., (2000). They reported considerable reduction in the podophyllotoxin content in the roots of plants collected from a higher altitude when grown at a lower altitude. The altitude ranges from 1300 m (Parvati forest division) to a maximum of 4300 m (Lahaul forest division). As a result, there is a wide

variation in climatic factors among the sampling sites and thus there is variation in podophyllotoxin content among the populations used in the study.

This is the first report demonstrating that high podophyllotoxin production in the root of *Podophyllum hexandrum* has close relationship with the high soil carbon, low soil pH value, high soil nitrogen, low concentration of soil phosphorus and potassium as well as environmental variables. Among these soil pH, soil organic matter and soil nitrogen are most significantly correlated with the podophyllotoxin production. Our work demonstrated that the soil in the northwestern Himalayan region, Himchal Pradesh, is acidic and *P. hexandrum* has not only adapted well to it but it also produces more podophyllotoxin in the acidic soil. Other researches also reported that pH value influenced secondary metabolites production. Cell culture of *Lupinus polyphyllus* increased alkaloids production when pH value decreased from 5.5 to 3.5 in the culture medium (Endress, 1994). Hydrogen ions in soil change the membrane permeability of the root cell so that soil pH affects the growth of plant directly. It also affects the uptake of the soil nutrients by plants indirectly (Endress, 1994). Therefore, the high podophyllotoxin yield in acidic soil may be caused by the influence of the acidic soil on the availability and uptake of soil elements such as N, P and K (Bhojwani and Razdan, 1996).

Soil organic matter provides plants with NPK and essential metal co-factors for metabolism. High soil organic matter content can uniformly supply the nutrition to plants, guarantee the plants a good growth and metabolic status and enhance the resistance of the plants to stresses. All these are the bases of secondary metabolism. Our work demonstrated that high soil organic matter favors the high podophyllotoxin yield of *P. hexandrum.* Other soil factor significantly affecting the podophyllotoxin production is Nitrogen (N). Soil N has close positive relationship with plant growth and metabolism because the N is a structural component of amino acids from which proteins are synthesized and wide metabolisms take place. Nitrogen is needed in the production of phenylalanine which is the starting point of the general phenylpropanoid pathway leading to podophyllotoxin production. Nitrogen forms such as organic N nutrients might be more important to high podophyllotoxin yield than the inorganic ones. Therefore, how N nutrient and which N forms take part in the

biosynthesis process of podophyllotoxin is the essential work for improving podophyllotoxin yield in the future study.

High phosphorus and potassium content in soil inhibited the podophyllotoxin yield. The reduced phosphate content in the culture medium was also reported to increase secondary metabolite accumulation in other plant species (Knobloch and Berlin, 1983; Endress, 1994). The level of exchangeable K in the forest conditions is rarely a problem for plant growth (Li and Huang 1989; Xi 1994). But certain species are sensitive to it (Bhojwani and Razdan, 1996). The mechanism of P and K affecting the podophyllotoxin yield in the cellular bioprocess remains unclear. It might be due to the fact that both elements regulate the activity of certain enzymes involved in podophyllotoxin biosynthesis or in N metabolism or in the cells' energy level such as ATP during these processes (Endress, 1994).

Many Indian medicinal herbs have higher medicinal productivity in their original habitat than in cultivated lands. Soil nutrient characters and environmental factors similar to original habitats must be most suitable for the active compound production. According to the results in this research, podophyllotoxin production of *Podophyllum hexandrum* or other bioactive compound production from herbs cultivated in farms can be improved through the soil management to mimic the soil condition similar to their original ones. The first approach is to measure the key soil nutrient factors such as pH value, organic matter content and N content as well as P and K contents when choosing a farm for the cultivation of *P. hexandrum.* The second approach is the balanced fertilization. It is important and necessary to increase and maintain the soil organic matter and N content at high levels during the period of the plants' growth and harvesting season. Organic N fertilizers are suggested to be used for the cultivation of *Podophyllum hexandrum*.

Soil matrix is a complex organic ecosystem with complex interaction between nutrition and microorganisms. Furthermore, plant secondary metabolism itself is also a complex physiological process. The secondary metabolite production is influenced by the plant's own physiological age and status and other environmental factors. Therefore, the effects of soil on podophyllotoxin production of *P. hexandrum* are far complicated beyond

those mentioned above. Impact of these topological and climatic factors with podophyllotoxin content is not studied so far. However, it is assumed that the variation in podophyllotoxin content is dependent on these factors which call for further research. The prediction model like ANN and MLR developed in this study to map the effects of these factors on podophyllotoxin yield will be helpful up to a certain extent for conservation of the plant. The results showed that using a combination of topographic soil and environmental data, we were able to successfully predict podophyllotoxin yield with ANN and MLR**.** Both the models could provide useful information regarding selection of sites, optimization of soil and environmental factors in order to increase the yield of podophyllotoxin and thus are very important before planning any conservation strategy.

CHAPTER 7

Computational and molecular modeling evaluation of the cytotoxic activity of podophyllotoxin analogues

Abstract

Podophyllotoxin and its structural derivatives, a class of tubulin polymerization inhibitors, have been the objective of numerous studies to prepare better and safer anti-cancer drugs. A library of podophyllotoxin analogues has been designed consisting of 154 analogues. Their molecular interactions and binding affinities with tubulin protein (PDB ID: 1SA1) have been studied using the docking-molecular mechanics based on generalized Born/surface area (MM-GB/SA) solvation model. Quantitative structure activity relationships were developed between the cytotoxic activity (pIC_{50}) of these compounds and molecular descriptors like docking score and binding free energy. For both the cases the r^2 was in the range of 0.642-0.728 indicating good data fit and r^2_{cv} was in the range of 0.631-0.719 indicating that the predictive capabilities of the models were acceptable. In addition, a linear correlation was observed between the predicted and experimented pIC_{50} for the validation data set with correlation coefficient r^2 of 0.806 and 0.887, suggesting that the docked structure orientation and the interaction energies are reasonable. Low levels of root mean square error for the majority of inhibitors establish the docking and prime/MM-GBSA based prediction model as an efficient tool for generating more potent and specific inhibitors of tubulin protein by testing rationally designed lead compounds based on podophyllotoxin derivatization.

7.1 Introduction

Normal cell division, intracellular transport, cellular motility, cell signaling and maintenance of cell shape are all dependent on highly regulated dynamic instability process of the tubulin/microtubule system. Microtubules are hollow tubes consisting of α- and β-tubulin heterodimers that polymerize parallel to a cylindrical axis. Mitotic microtubules are very dynamic structures, switching between growing and shortening states, a process known as dynamic instability. Drugs that inhibit tubulin polymerization/depolymerization are commonly used as chemotherapeutic agents for a variety of cancer, as well as for probing microtubule dynamics in cellular and biochemical processes. Well-known examples are vinblastine, vincristine and paclitaxel. However, the mechanism of action of these microtubule poisons with tubulin is different. For example, paclitaxel, vinca alkaloids, colchicinoids and dolastatin appear to bind different sites on the tubulin α-β heterodimer (Hamel, 1996).

Podophyllotoxin is clinically effective anti-cancer agent that represents perhaps the most significant addition to the pharmacopoeia of cancer chemotherapeutic agents in the last decade (Brewer et al., 1979). Prompted by the clinical successes of the podophyllotoxin, significant efforts have been focused on identifying new analogues that have a similar mechanism of action yet superior properties such as low or nil toxic side effects and better oral availability (Jardine, 1980; Keller-Juslen et al., 1971; Weiss et al., 1975). A consistent number of structural modifications have been introduced in the original structure of podophyllotoxin in order to overcome the side effects associated with its utilization as anti-cancer drug. The study and assessment of these have permitted the clinical development and their usage in the treatment of different types of cancer.

Since the discovery of the therapeutic properties of podophyllotoxin, new findings related to its activities, its mechanism of action and pharmacological properties have been unveiled. Structure-Activity Relationships (SAR), have shown that podophyllotoxin analogues preferentially inhibit tubulin polymerization, which leads to arrest of the cell cycle in the metaphase (Snyder et al., 1976; Margolis et al., 1978). The first substance known to have this activity was colchicine, which binds to a specific site on the protein,

known as the colchicine site. Different derivatives of podophyllotoxin have demonstrated to bind to the same site, as shown by the fact that podophyllotoxin has been reported to compete with colchicine for the binding site in tubulin (Cortese et al., 1977), its affinity is more than double to that of colchicine. Furthermore, colchicine binds to tubulin almost irreversibly whereas podophyllotoxin derivatives do so reversibly, which makes them less toxic and more useful in the field of cancer therapy (David-Pfeuty et al., 1979; Lin et al., 1981). The discovery of new natural and semisynthetic compounds of podophyllotoxin being cytotoxic by interference with tubulin have attracted much attention in the last several years. The microtubule complex has thus proven to be a compelling target for the development of anti-cancer therapeutic agents (Loehrer, 1991). The crystal structures show that both colchicine and podophyllotoxin bind at the interface between α and β subunit of tubulin protein (Figure 7.1).

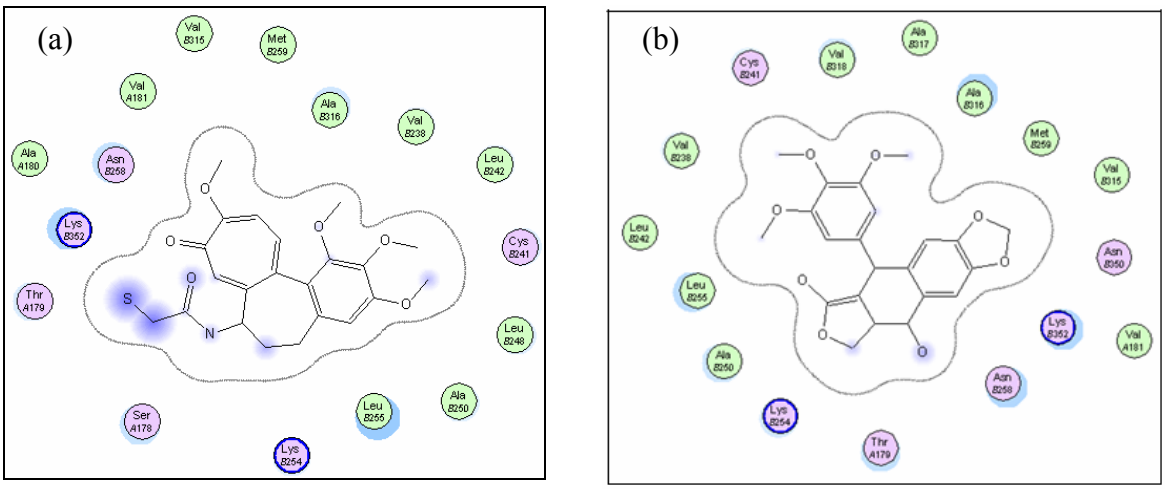


Figure 7.1. Diagram showing (a) Colchicine binding site (PDB ID: 1SA0) and (b) Podophyllotoxin binding site (PDB ID: 1SA1).

Crystal structures of tubulin with colchine and podophyllotoxin have been reported. The crystal structures show that all colchicines and podophyllotoxin bind at the interface between α and β subunit of tubulin protein. The binding site of tubulin-podophyllotoxin complex (PDB ID: 1SA1, R = 4.20 Å) consists of amino acids such as; Serα178, Thr α179, Ala α180, Val α181, Val β238, Cys β241, Leu β242, Ala β250, Lys β254, Leu β255, Asn β258, Met β259, Val β315, Ala β316, Ala β317, Val β318, Asn β350, Lys β352, and Ileu

β378. Similar amino acid residues were also found in the binding site of tubulin-colchicine complex (PDB ID: 1SA0, R = 3.58 Å). Although the overall shape of the podophyllotoxin binding site is relatively the same in comparison to colchicine binding site, there are subtle differences among them (root mean square deviation (RMSD) = 1.29 Å). The hydrophobic center that is located in the middle of trimethoxyphenyl moiety of podophyllotoxin is surrounded by Leu β242, Ala β250, Leu β255, Ala β316, Val β318 and Ile β378 residues.

The great diversity of the podophyllotoxin analogues, the huge number of assays carried out on them, and the different mechanisms of action observed in different series make it difficult to clearly define the minimum structural requirements necessary for their biological activity. Additionally, the results available have been obtained by different authors; at different times using different technologies on very diverse types of tumors or cultures of neo-plastic cells. For all these reasons, greater systematization would be required to obtain definitive conclusions. The mechanism of action of any drug is very important in drug development. Generally, the drug compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug–receptor interactions are required for high activity. Understanding the nature of these interactions is very significant and theoretical calculations, in particular the molecular docking method, seem to be a proper tool for gaining such understanding. The docking results obtained will give information on how the chemical structure of the drug should be modified to achieve suitable interactions and for the rapid prediction and virtual prescreening of anti-tumor activity.

Given the mechanism of action of podophyllotoxin analogues two accepted mechanisms have been proposed. One consists of the inhibition of tubulin polymerization and the second accepted mechanism consisting of the irreversible inhibition of DNA-topoisomerase II (Hamel et al., 1996; David-Pfeuty et al., 1979; Sun et al., 1993). Structure-Activity Relationships (SAR) studies reported earlier have shown that podophyllotoxin like compounds preferentially inhibit tubulin polymerization, which leads to arrest of the cell cycle in the metaphase (Snyder et al., 1976; Margolis et al., 1978). However, etoposide like compounds are potent irreversible inhibitors of DNA topoisomerase II and their action is based on the formation of a nucleic acid-drug-enzyme complex, which includes single and

double stranded DNA breaks that eventually lead to cell death (Margolis et al., 1978; Yamashita et al., 1991).

In this work we created a virtual library of podophyllotoxin analogues which were collected from different sources and screened them for tubulin binding. Further, prediction models for predicting the cytotoxic activity of these compounds were developed based on binding interaction with tubulin as descriptor. This prediction model was used for predicting the cytotoxic activity of newly developed analogues. We have used the molecular modeling techniques (molecular docking and rescoring using Prime/MM-GBSA) to find the series of podophyllotoxin analogues that should be modified for energetically favorable interaction with tubulin and for better cytotoxic activity.

7.2 Materials and methods

7.2.1 Preparation of protein

The X-ray structure of the complex between podophyllotoxin and tubulin protein (PDB ID: 1SA1) has been used as initial structure in the preparation of podophyllotoxin binding site. After manual inspection and cleaning of structure we retained a complex consisting of protein chains α & β and podophyllotoxin ligand. Hydrogens were added to the model automatically via the Maestro interface (Schrodinger 2007) leaving no lone pair and using an explicit all-atom model. All the water molecules were removed form the complex. The multi step Schrodinger's protein preparation tool (PPrep) has been used for final preparation of protein. PPrep neutralized side chains that are not close to the binding cavity and do not participate in salt bridges (Schrodinger 2007). This step is then followed by restrained minimization of co-crystallized complex, which reorients side chain hydroxyl groups and alleviates potential steric clashes. Progressively weaker restraints (tethering force constants 3, 1, 0.3, 0.1) were applied to non-hydrogen atoms only. The complex obtained was minimized using OPLS-2005 force field with Polack-Ribiere Conjugate Gradient (PRCG) algorithm (Polak et al., 1969). The minimization was stopped either after 5000 steps or after the energy gradient converged below 0.05 kcal/mol.

7.2.2 Virtual library design

The virtual library of podophyllotoxin analogues contains 154 compounds divided into 5 sub libraries. All these compounds are taken from various sources belonging to different ring modifications.

*Sublib-**I*** commonly known as tetralinelactones consists of 52 compounds (1-52) (Table 7.1a). These molecules were rationally designed as functional mimics of natural podophyllotoxin with the goal of simplifying the chemical synthesis and improving the cytotoxic activity. Structural modifications of varying radicals were mainly introduced at position 7 in podophyllotoxin scaffold. Reports have been made of compounds with oxygenated substituents in the form of ethers, esters and diverse nitrogen radicals (San Feliciano et al., 1989; San Feliciano et al., 1990; San Feliciano et al., 1993; Gordaliza et al., 1994; Castro et al., 1994; Gordaliza et al., 1995; Miguel del Corral et al., 1995; Doré et al., 1996; Gordaliza et al., 1996; Gordaliza et al., 1995).

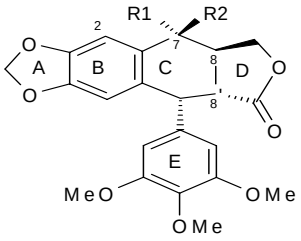
*Sublib-**II*** contains compounds (53- 97) (Table 7.1b) commonly known as non-lactonic tetralines. Structural modifications in this group include the opening of the lactone ring (D-ring) in podophyllotoxin scaffold, to give rise to compounds with different degrees of oxidation at positions C-9 and C-9' (San Feliciano et al., 1993; Gordaliza et al., 1994; Castro et al., 1994; Gordaliza et al., 1995; Doré et al., 1996; Gordaliza et al., 1996). In general these molecules lack lactone rings.

*Sublib-**III*** also includes a group of lignans (98-120) (Table 7.1c) that have heterocyclic rings fused to the cyclolignan skeleton. This group is commonly called as pyrazolignans (San Feliciano et al., 1993; Gordaliza et al., 1994; Dore et al., 1996; Gordaliza et al., 1996; Gordaliza et al., 1995) and isoxazolignans (Gordaliza et al., 1996; Gordaliza et al., 1995; Gordaliza et al., 1996; Miguel del Corral et al., 1997) and they were obtained by reacting podophyllotoxine with differently substituted hydrazines and hydroxylamines.

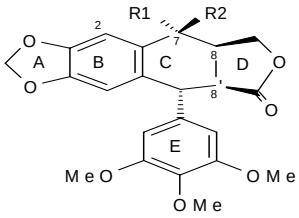
*Sublib-**IV*** includes the compounds (121-126) (Table 7.1d) commonly called lactonic and non-lactonic naphthalene. These molecules were obtained by structural modification of C and D- rings and have proportionally much lower activity (San Feliciano et al., 1993; Gordaliza et al., 1994; Doré et al., 1996).

*Sublib-**V*** contains compounds (127-154) (Table 7.1e) commonly known as aza-podophyllotoxin analogues. The preparation of this group of compounds requires selective chemical manipulation of the two aromatic rings (B and E-rings) of the podophyllotoxin scaffold. These molecules are readily prepared from anilines, benzaldehydes and tetronic acid or 2, 3-cyclopentanedione in good to excellent yield and have also shown better cytotoxic activity (Hitotsuyanagi et al., 2000).

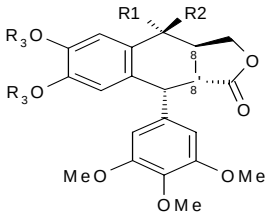
The virtual library of podophyllotoxins was built from the scaffolds by different ring modification and substitution of functional groups as mentioned in Table 1(a-e). We used ISIS Draw 2.3 software for sketching structures and converting them to their 3D representation by using ChemSketch 3D viewer of ACDLABS 8.0. LigPrep (Schrodinger 2007)was used for final preparation of ligands from libraries for docking. LigPrep is a utility of Schrodinger software suit that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for tatutomers and steric isomers and perform a geometry minimization of ligands. The ligands were minimized by means of Molecular Mechanics Force Fields (OPLS-2005) with default setting.



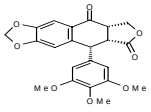
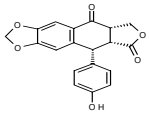
1-22



23-42

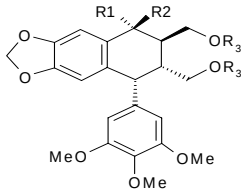


45-52

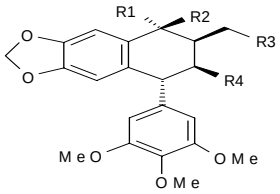
Analogue	R1	R2	Expt.IC50	Analogue	R1	R2	Expt.IC50
1	OH	H	0.012	22	=N-OMe		0.2
2	H	H	0.010	23	H	H	0.10
3	H	H(2-OMe)	0.01	24	H	H(2-OMe)	0.23
4	OH	H(2-OMe)	-	25	OH	H	6.0
5	OH	H(4'-OH	0.027	26	OH	H(2-OMe)	-
6	OAc	H	0.625	27	OAc	H	0.55
7	OAc	H	-	28	OAc	H(2-OMe)	1.02
8	OMe	H	0.06	29	OMe	H	0.12
9	H	OH	0.06	30	H	OH	-
10	H	Ac	0.05	31	H	OH(2-OMe)	0.11
11	H	OMe	0.06	32	H	OAc	0.44
12	H	Cl	0.6	33	H	OAc(2-OMe)	0.51
13	Cl	H	0.6	34	H	OMe	0.12
14	=O		1.8	35	H	Cl	-
15	H	Br	-	36	Cl	H	-
16	Br	H	-	37	H	H $\Delta^{8(8')}$	-
17	H	H(4'-OH)	-	38	H	H Δ^{7}	0.013
18	H	H(4'-OAc)	-	39	=O	12.0	
19	H	OAc(4'-OAc)	-	40	=N-OH		2.3
20	= N-OH		2.3	41	N-OAc		-
21	=N-OAc		2.1	42	=N-OMe		2.3
43			-	44			-

Analogue	R1	R2	R3	Expt.IC50	Analogue	R1	R2	R3	Expt.IC50
45	H	H	H	-	49	H	OH	H	-
46	H	H	Ac	-	50	H	OAc	Ac	-
47	OH	H	H	-	51	H	H	H $\Delta^{8(8')}$	-
48	OAc	H	Ac	-	52	H	H	Ac $\Delta^{8(8')}$	-

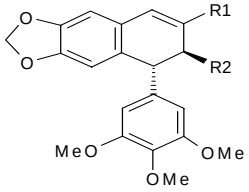
Table 7.1 (a). Podophyllotoxin derivatives (Tetralactones) with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.



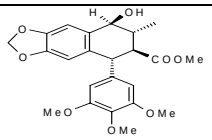
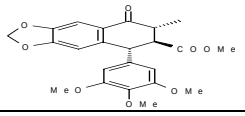
53 - 59.



62-81



82 - 92.

Analogue	R1	R2	R3	Expt.IC50	Analogue	Structure	Expt.IC50
53	OH	H	H	1.2	60		23.3
54	H	OH	H	12.0			
55	H	OAc	Ac	-			
56	H	OMe	H	11.6			
57	H	OMe	Ac	9.7			
58	OMe	H	H	-	61		3.5
59	OMe	H	Ac	9.7			

Analogue	R1	R2	R3	R4	Expt.IC50	Analogue	R1	R2	R3	R4	Expt.IC50
62	H	H	OH	COOMe	0.058	71	H	OMe	OH	CH_2OH	11.6
63	H	H	OAc	COOMe	0.21	72	H	OMe	OAc	CH_2OAc	9.7
64	H	H	OAc	CH_2OAc	5.14	73	H	OH	OH	CH_2OH	47.9
65	OH	H	OH	CH_2OH	23.9	74	H	OH	OH	COOMe	1.1
66	OH	H	OH	COOMe	0.22	75	=O		OH	COOMe	5.63
67	OAc	H	OAc	CH_2OAc	7.4	76	=O		OAc	COOMe	0.20
68	OAc	H	OAc	COOMe	1.1	77	=N-OH		OAc	COOMe	2.0
69	OMe	H	OH	CH_2OH	23.2	78	H	H	CHO	COOMe	2.34
70	OMe	H	OAc	CH_2OAc	19.4	79	H	H	=N-OMe	COOMe	2.30
80	H	H	=N-OMe	COOMe	10.94	81	H	H	=N-allyl	COOMe	2.5

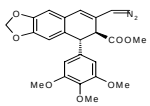
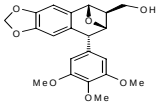
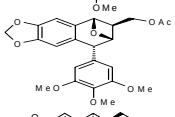
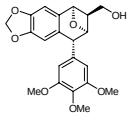
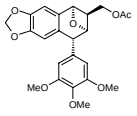
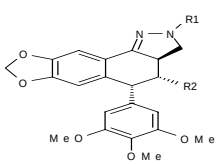
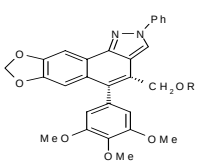
Analogue	R1	R2	Expt.IC50	Analogue	R1	R2	Expt.IC50
82	CH_2OH	COOMe	0.02	89	CH=N-OH	COOMe	2.27
83	CHO	CH_2OH	0.25	90	CH=N-OMe	COOMe	0.22
84	CHO	COOMe	0.23	91		COOMe	0.20
85	CH=N-NH2	COOMe	0.57	92		CH_2OH	1.00
86	CH=N-NH-CH_2CF_3	COOMe	0.48	93			0.57
87	CH=N-NH-Ph	COOMe	1.94	94			6.25
88	CH=N-NH-Ph	CH_2OH	1.02	95			5.66
96			-	97			-

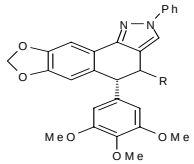
Table 7.1. (b). Podophyllotoxin derivatives (Nonlactonic tetralines) with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.



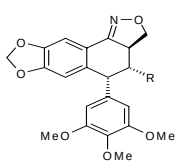
98 - 109.



110-111



112-114

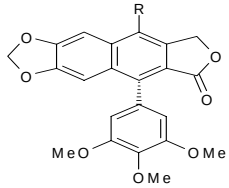


115-120

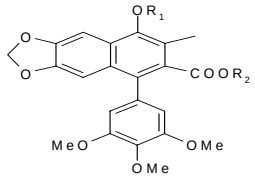
Analogue	R1	R2	Expt.IC50	Analogue	R1	R2	Expt.IC50
98	Ph	COOH	1.9	104	p-BrPh	COOH	-
99	Ph	COOMe	1.00	105	p-MePh	COOMe	1.00
100	Ph	CH_2OH	4.1	106	Me	COOH	-
101	Ph	CH_2OAc	4.7	107	Me	COOMe	5.6
102	m-NO_2Ph	COOH	-	108	CONH2COOH	COOH	-
103	m-NO_2Ph	COOMe	4.5	109	$COCH_3COOMe$	COOMe	21

Analogue	R	Expt.IC50	Analogue	R	Expt.IC50
110	H	10	116	COOMe	23
111	Ac	-	117	COOMe(4'-OH)	12
112	CHO	21	118	CH_2OH	2.6
113	CH_2OH	-	119	CH_2O	2.4
114	CH_2Ac	2.2	120	CHO	-
115	COOH	2.2			

Table 7.1(c) Podophyllotoxin derivatives (Pyrazolignans and isoxazolignan) with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.



121-122



123-126

Analogue	R		Expt.IC50	Analogue	R1	R2	Expt.IC50
121	H		5.1	124	Ac	H	5.90
122	OAc		44.25	125	Ac	Me	16.59
123	H	Me	12.20	126	H	OMe	2.15

Table 7.1(d) Podophyllotoxin derivatives (lactones and non-lactonic naphthalene) with cytotoxic activities against P-388 cell line used in the work.

Modification 1 **Modification 2**

Substitution of B & E ring at 1 and 2 analogues:

Modification 1			Analog	Modification 2		
B Ring	E Ring	Expt. IC50		B Ring	E Ring	Expt. IC50
I	VII	100	141	I	VII	0.0018
II	VII	80	142	II	VII	0.0017
III	VII	100	143	III	VII	4.9
III	VIII	39	144	III	VIII	0.76
III	XII	2.0	145	III	XII	0.77
IV	VII	29	146	IV	VII	2.6
V	VII	100	147	V	VII	0.0041
VI	VII	63	148	VI	VII	0.92
I	VIII	40	149	I	VIII	0.048
I	IX	100	150	I	IX	0.0053
I	X	100	151	I	X	0.13
I	XI	60	152	I	XI	0.0053
I	XII	100	153	I	XII	0.030
I	VII	71	154	I	VII	0.028

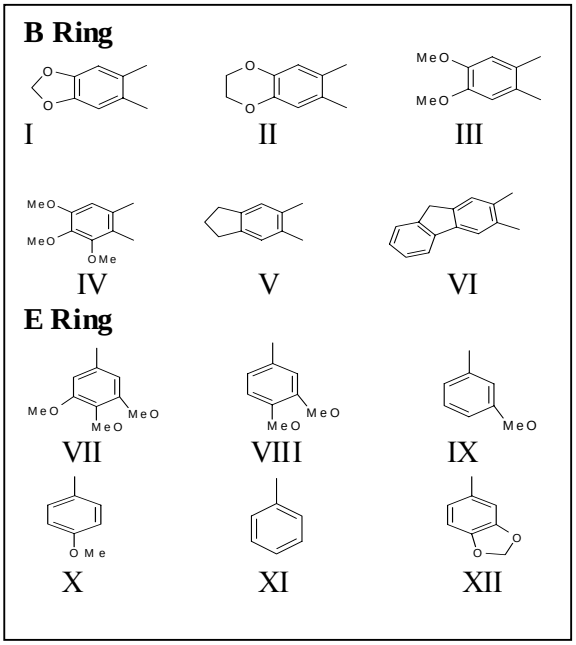


Table 7.1(e). Aza-podophyllotoxin derivatives with cytotoxic activities against P-388 cell line used in the work.

7.2.3 Docking procedure

The Schrodinger Glide program version 4.0 has been used for docking (Friesner et al., 2004; Halgren, 2004). The best 10 poses and corresponding scores have been evaluated using Glide in single precision mode (Glide SP) for each ligand from the virtual library of podophyllotoxin. For each screened ligand, the pose with the lowest Glide SP score has been taken as the input for the Glide calculation in extra precision mode (Glide XP). To soften the potential for non-polar parts of the receptor, we scaled van der Waals radii of receptor atoms by 1.00 with partial atomic charge 0.25.

7.2.4 Rescoring using Prime/MM-GBSA approach

For each ligand, the pose with the lowest Glide score was rescored using Prime/MM-GBSA approach (Lyne et al., 2006). This approach is used to predict the free energy of binding for set of ligands to receptor. The docked poses were minimized using the local optimization feature in Prime and the energies of complex were calculated using the OPLS-

AA force field and generalized-Born/surface area (GBSA) continuum solvent model. The binding free energy (ΔG_{bind}) is then estimated using equation:

$$\Delta G_{bind} = E_{R:L} - (E_R + E_L) + \Delta G_{solv} + \Delta G_{SA} \quad (1)$$

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

In order to explore the reliability of the proposed models we used the cross validation method. The cross validation analysis performed by using the leave one out (LOO) method in which one compound removed from the data set and its activity predicted using the model derived from the rest of the data points. Prediction error sum of squares (PRESS) is a standard index to measure the accuracy of a modeling method based on the cross validation technique. The r^2_{cv} was calculated based on the *PRESS* and *SSY* (Sum of squares of deviations of the experimental values from their mean) using following formula.

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

Where y_{exp}, y_{pred} and $\bar{y}$ are the predicted, observed and mean values of the cytotoxic activities of the podophyllotoxin analogues.

7.3 Results and Discussions

7.3.1 Molecular docking of podophyllotoxin and its analogues

The original crystal structure of tubuline-podophyllotoxin complex (PDB ID: 1SA1) was used to validate the Glide-XP docking protocol. This was done by moving the co-crystallized podophyllotoxin ligand outside of active site and then docking it back into the active site. The top 10 configurations after docking were taken into consideration to validate the result (Table 7.2). The RMSD was calculated for each configuration in comparison to the co-crystallized podophyllotoxin and the value was found to be in between 0.02-0.85 Å.

Table 7.2. The RMSD and docking score from the docking simulation of 10 lowest configurations of co-crystal podophyllotoxin in Tubulin protein (ISA1).

Configuration	Glide Score	[a]ΔG_{score}	[b]RMSD (Å)	[c]RMSD (Å)
1	-10.26	0	0.85	0.60
2	-10.20	-0.06	0.02	0.86
3	-9.80	-0.46	0.68	1.33
4	-9.72	-0.54	0.57	1.26
5	-9.50	-0.76	0.04	0.67
6	-9.25	-1.01	0.04	0.67
7	-8.78	-1.48	0.80	0.59
8	-8.47	-1.79	0.13	1.02
9	-7.87	-2.39	0.03	0.79
10	-7.72	-2.54	0.07	0.90

[a]$\Delta G_{score} = E_i - E_{lowest;}$ [b]RMSD = RMSD between docked and crystallographic podophyllotoxin structure; [c]RMSD = RMSD between docked poses corresponding to each configuration.

Whereas the RMSD value calculated out of 10 accepted poses for each configuration was found in between 0.59 – 1.33 Å. This revealed that the docked configurations have similar binding positions and orientations within the binding site and are similar to the crystal structure. The best docked structures, which is the configuration with the lowest Glide score is compared with the crystal structure as shown in Figure 7.2(a-b). These docking results illustrate that the best-docked podophyllotoxin complex agrees well with its crystal structure and that Glide (XP)-docking protocol successfully reproduces the crystal tubulin-podophyllotoxin complex.

Glide 4.0 in XP mode have been used to dock the library (I-V) into the podophyllotoxin binding site of tubulin. The docking score (G-score) has been used for screening the virtual library of podophyllotoxin analogues. The binding modes of five superimposed ligands from each class within podophyllotoxin binding site are given in Figure 7.3(a-e) respectively. In this figure we can observe that all the ligands are well fitted to the defined binding pocket. All the 154 podophyllotoxin analogues were found to be good binder with tubulin. Table 7.3 reveals the distribution of Glide Score of various sublibraries (I-V) of podophyllotoxin analogues. We can observe that the most potent podophyllotoxin analogues were found among the tetralinelactones ligands (Sublib-I) followed by non-lactonic tetralines, aza-podophyllotoxin, pyrazolignans & isoxazolignans and lactonic & non-lactonic naphthalene podophyllotoxin derivatives in sequential order.

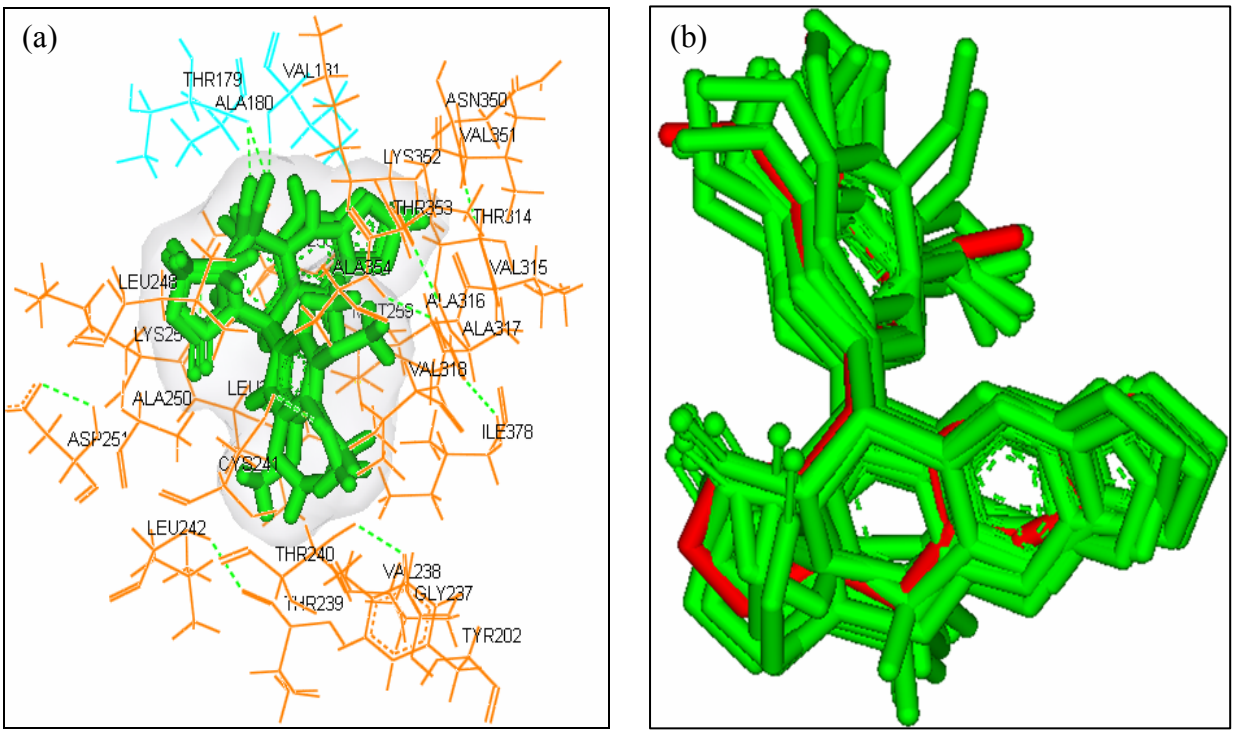


Figure 7.2. Superposition of the docked configurations of co-crystallized podophyllotoxin: (a) with binding site and (b) only the superposed structure (red one represents the X-ray podophyllotoxin structure). RMSD (heavy atoms) = 0.02 to 0.85 Å.

For each ligand in the virtual library, the pose with the lowest Glide score was rescored using Prime/MM-GBSA approach. This approach is used to predict the binding free energy (ΔG_{bind}) for set of ligands to receptor. Table 7.4 reveals the distribution of ΔG_{bind} of various sublibraries (I-V) of podophyllotoxin analogues. It can be seen that the most potent podophyllotoxin analogues were found among the tetralinelactone ligands (Sublib-I) followed by nonlactonic tetralines, aza-podophyllotoxin, pyrazolignans & isoxazolignans and lactonic & non-lactonic naphthalene podophyllotoxin derivatives in sequential order. The majority (83.12%) of the ligands from sublibraries (I-V) have ΔG_{bind} between -30.0 and -15.0 kcal/mol. On the contrary only 16.88% of ligands have ΔG_{bind} higher than -15.0 kcal/mol. So we can conclude that the structural derivatives of podophyllotoxin used in the study is far more focused to tubulin binding site based on both Glide score and Prime/MM-GBSA approaches.

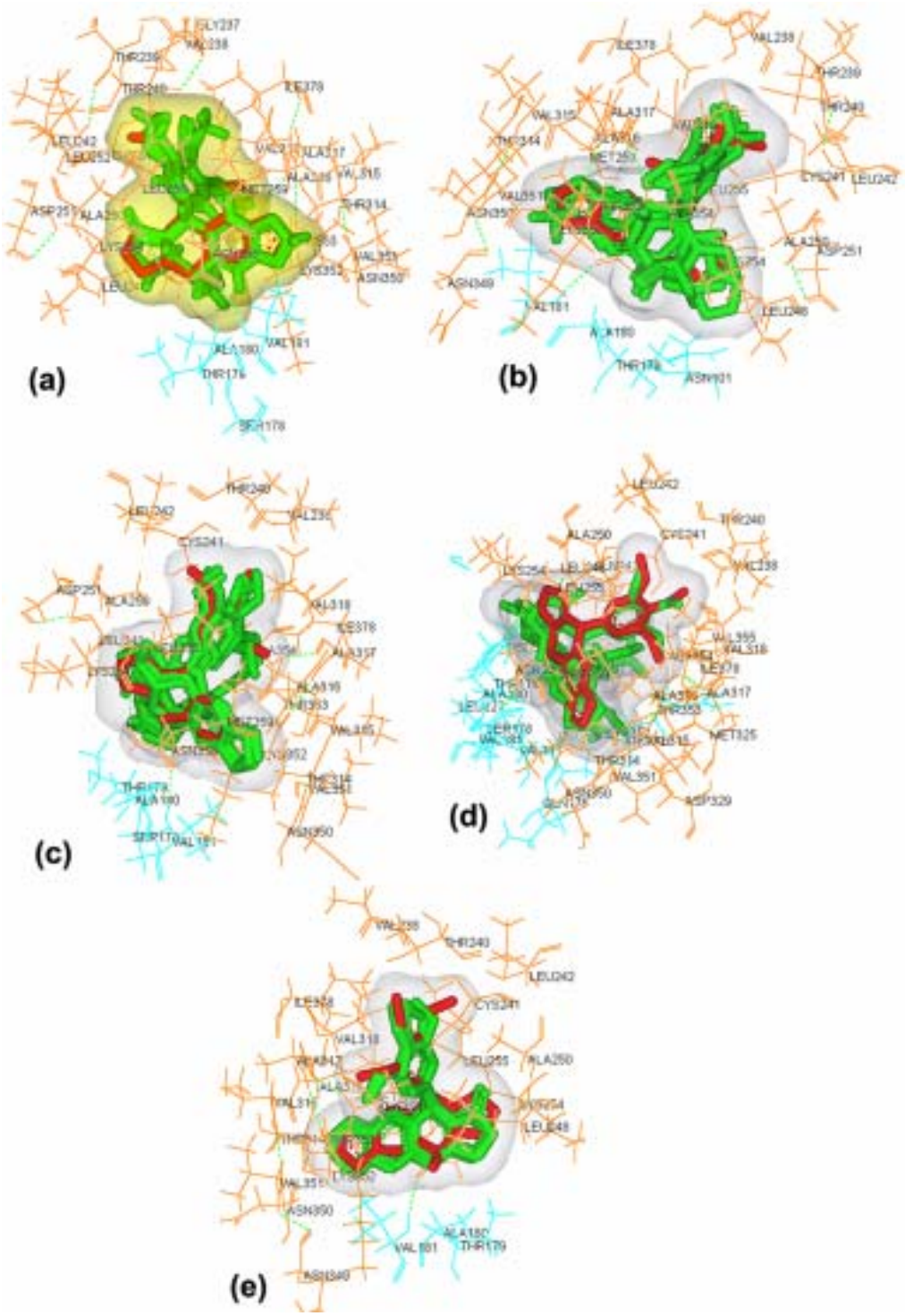


Figure 7.3. (a-e). Superposition of podophyllotoxin analogues (5 analogues) belonging to (a) Tetraline lactones, (b) Non-lactonic tetralines, (c) Pyrazoline and isoxazoline derivatives, (d) Lactonic and non-lactonic naphthaline and (e) Aza-podophyllotoxin derivatives within binding site of Tubuline. The majority (85.71%) of the ligands from sublibraries (I-V) have Glide score between -10.0 and -8.0. On the contrary only 14.29% of ligands have Glide score higher than -8.0.

Table 7.3: Glide score distribution in sublibraries of podophyllotoxin analogues.

Glide XP	I	II	III	IV	V	Total
G Score <- 11	0	0	0	0	1	1
-10.0 > G Score ≥ -11.0	12	11	2	0	3	28
-9.0 > G Score ≥ -10.0	34	26	9	0	5	74
-8.0 > G Score ≥ -9.0	5	7	7	2	6	27
-7.0 > G Score ≥ -8.0	0	0	3	3	9	15
-6.0 > G Score ≥ -7.0	0	0	2	1	4	7
Total	52	44	23	6	28	

Table 7.4: Distribution of binding free energy (ΔG_{bind}) in sublibraries of podophyllotoxin analogues.

Prime	I	II	III	IV	V	Total
$\Delta G_{bind} < -30$	3	0	0	0	1	4
$-25 > \Delta G_{bind} \geq -30$	10	2	0	0	7	19
$-20 > \Delta G_{bind} \geq -25$	7	4	4	1	2	18
$-15 > \Delta G_{bind} \geq -20$	12	10	9	3	8	42
$-10 > \Delta G_{bind} \geq -15$	8	13	7	1	9	38
$-5 > \Delta G_{bind} \geq -10$	7	9	2	0	0	18
$0 > \Delta G_{bind} \geq -5$	5	3	0	0	0	8
Total	52	41	22	4	27	

7.3.2 Building models for prediction of pIC_{50} using Glide score and Prime/MM-GBSA

We selected some ligands with known cytotoxic activity (pIC_{50}) from virtual library (I-V). Cytotoxic activity of these podophyllotoxin analogues (Sublib I-V) was generally evaluated using P-388 leukemia cells and were collected from different sources (Hitotsuyanagi et al., 2000; Haar et al., 1996; Gordaliza et al., 2000) and included in Table 1a-e. It has been seen that the dihydroquinoline analogues (127-154) belonging to aza-podophyllotoxin (sublib-V) have significantly better activitives (pIC_{50} in the range of -2.00 μM to 2.77 μM) compared to the other sublibraries (I-IV). On the contrary, the tetralinelactones (1-52 analogues; sublib-I) showed comparatively better activity of pIC_{50} value ranging from -1.08 μM to 2.00 μM than non-lactonic tetralines (sublib-II). The pyrazolignans & isoxazolignans, lactonic & non-lactonic naphthalene generally showed very weak or no activity. The mode of action of podophyllotoxin structural derivatives is reported to be due to inhibition of microtubule assembly through binding to tubulin (Hamel et al., 1996; Brewer et al., 1979). Thus, in this study we have taken tubulin protein as the molecular target and built prediction model for prediction of cytotoxic activity by

considering the Glide score and ΔG_{bind} as descriptors. The equation (3) of the model and the corresponding statistics are shown below:

$$pIC_{50} = - 8.725(\pm 0.644) – 0.938(\pm 0.592)* \text{G-score} \qquad (3)$$

$$(N = 120, r^2 = 0.642, s = 0.692, F = 211.86, r^2_{cv} = 0.631, PRESS = 58.349)$$

The root mean square error (RMSE) between the experimental pIC_{50} values and the predicted pIC_{50} values obtained by the regression model was 0.626 μM, which is an indicator of the robustness of the fit and suggested that the calculated pIC_{50} based on Glide score is reliable. The quality of the fit can also be judged by the value of the squared correlation coefficient (r^2), which was 0.642 for the data set. Figure 7.3 graphically shows the quality of fit. The statistical significance of the prediction model is evaluated by the correlation coefficient r^2, standard error s, F-test value, leave-one-out cross-validation coefficient r^2_{cv} and predictive error sum of squares PRESS. The regression model developed in this study is statistically ($r^2_{cv} = 0.631$, $r^2 = 0.642$, F = 211.86) best fitted and consequently used for prediction of cytotoxic activities (pIC_{50}) of the podophyllotoxin analogues as reported in Table 7.5 (a-e). The average root mean square error between predicted and experimental pIC_{50} values was 0.838 μM using leave-one-out cross validation technique which further revealed the reliability of the model for prediction of cytotoxicity.

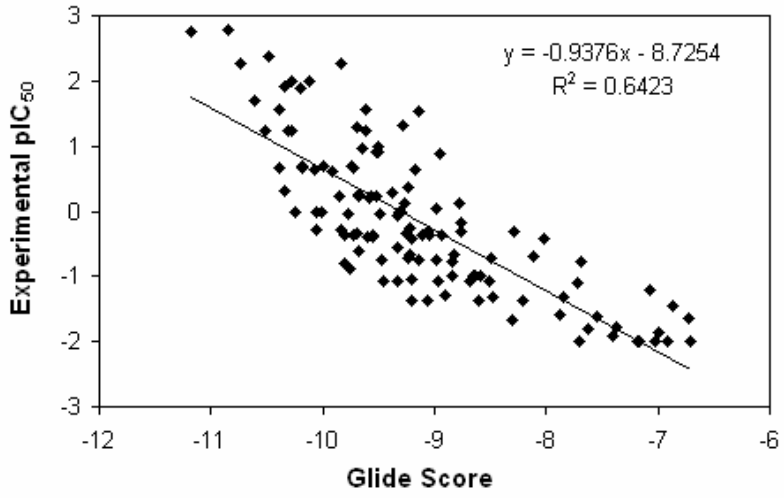


Figure 7.4. Models for predicting cytotoxic activity (pIC_{50}) of the podophyllotoxin analogues based on Glide score.

Table 7.5. (a). Predicted cytotoxic activities of Tetralinelactones podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity for selected analogues.

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})	Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})
1	-10.34	-30.82	1.92	0.97	1.80	27	-9.67	-17.31	0.26	0.34	-0.13
2	-10.12	-32.00	2.00	0.76	1.97	28	-9.31	-15.33	-0.01	0.01	-0.41
3	-10.27	-31.27	2.00	0.91	1.87	29	-9.53	-18.28	0.92	0.21	0.01
4	-9.19	-2.41	-	-0.11	-2.26	30	-10.29	-26.96	-	0.92	1.25
5	-10.38	-26.93	1.57	1.01	1.25	31	-9.65	-21.00	0.96	0.32	0.40
6	-9.59	-19.70	0.20	0.27	0.21	32	-9.23	-14.44	0.36	-0.07	-0.54
7	-9.21	-21.20	-	-0.09	0.43	33	-9.37	-15.07	0.29	0.06	-0.45
8	-10.32	-27.45	1.22	0.95	1.32	34	-9.50	-19.51	0.92	0.19	0.19
9	-10.29	-26.97	1.22	0.92	1.25	35	-9.53	-26.13	-	0.21	1.13
10	-9.70	-27.86	1.30	0.37	1.38	36	-9.50	-27.53	-	0.19	1.33
11	-9.62	-27.35	1.22	0.29	1.31	37	-9.41	-23.36	-	0.10	0.74
12	-9.52	-18.36	0.22	0.20	0.02	38	-10.20	-29.07	1.89	0.84	1.55
13	-9.57	-17.24	0.22	0.24	-0.14	39	-8.68	-3.35	-1.08	-0.59	-2.12
14	-9.22	-10.42	-0.26	-0.08	-1.11	40	-9.05	-12.05	-0.36	-0.24	-0.88
15	-9.43	-4.85	-	0.12	-1.91	41	-10.48	-12.44	-	1.10	-0.82
16	-9.34	-7.34	-	0.03	-1.55	42	-9.11	-2.69	-0.36	-0.18	-2.22
17	-9.54	-21.20	-	0.22	0.43	43	-9.54	-12.68	-	0.22	-0.79
18	-9.73	-26.74	-	0.40	1.22	44	-9.56	-18.70	-	0.24	0.07
19	-10.01	-22.89	-	0.66	0.67	45	-9.56	-14.03	-	0.24	-0.60
20	-8.94	-3.87	-0.36	-0.34	-2.05	46	-9.89	-8.92	-	0.55	-1.33
21	-8.77	-9.15	-0.32	-0.51	-1.30	47	-8.97	-10.19	-	-0.31	-1.15
22	-9.74	-19.00	0.70	0.41	0.11	48	-10.53	-20.85	-	1.15	0.38
23	-9.51	-23.13	1.00	0.19	0.70	49	-9.33	-9.28	-	0.02	-1.28
24	-9.18	-19.22	0.64	-0.12	0.14	50	-9.29	-12.43	-	-0.02	-0.83
25	-8.84	-6.00	-0.78	-0.44	-1.75	51	-9.33	-16.24	-	0.02	-0.28
26	-9.08	-7.24	-	-0.22	-1.57	52	-10.81	-8.44	-	1.41	-1.40

Expt., experimental values; Pred., predicted values; $pIC_{50} = - \log_{10} IC_{50}$; [a] based on equation (3) and [b]as per equation (4).

Table 7.5. (b). Predicted cytotoxic activities of Nonlactonic tetralinelactones podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity for selected analogues.

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})	Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})
53	-9.33	-19.72	-0.08	0.02	0.22	76	-9.99	-7.44	0.70	0.64	-1.54
54	-9.34	-6.60	-1.08	0.03	-1.66	77	-9.06	-18.77	-0.30	-0.23	0.08
55	-10.15	-10.12	-	0.79	-1.16	78	-9.56	-18.58	-0.37	0.24	0.05
56	-8.97	-5.00	-1.06	-0.32	-1.89	79	-9.80	-16.42	-0.36	0.46	-0.26
57	-10.43	-3.07	-0.99	1.05	-2.16	80	-9.20	-20.09	-1.04	-0.10	0.27
58	-9.73	-1.65	-	0.40	-2.37	81	-9.60	-11.89	-0.40	0.27	-0.90
59	-8.85	-5.47	-0.99	-0.43	-1.82	82	-10.61	-11.74	1.70	1.22	-0.92
60	-9.20	-8.21	-1.37	-0.10	-1.43	83	-9.92	-29.53	0.60	0.57	1.62
61	-9.33	-3.74	-0.54	0.02	-2.07	84	-10.08	-19.15	0.64	0.73	0.13
62	-10.51	-15.51	1.24	1.13	-0.39	85	-9.86	-20.75	0.24	0.52	0.36
63	-10.19	-27.26	0.68	0.83	1.29	86	-10.34	-12.83	0.32	0.97	-0.77
64	-9.24	-20.00	-0.71	-0.06	0.26	87	-10.06	-25.00	-0.29	0.71	0.97
65	-8.60	-13.11	-1.38	-0.66	-0.73	88	-10.06	-12.69	-0.01	0.71	-0.79
66	-9.72	-8.56	0.66	0.39	-1.38	89	-9.72	-14.66	-0.36	0.39	-0.51
67	-9.77	-17.47	-0.87	0.43	-0.11	90	-10.39	-14.06	0.66	1.02	-0.59
68	-9.49	-15.16	-0.04	0.17	-0.44	91	-10.18	-23.40	0.70	0.82	0.74
69	-9.07	-10.38	-1.37	-0.22	-1.12	92	-9.30	-19.72	0.00	0.00	0.22
70	-8.91	-6.89	-1.29	-0.37	-1.62	93	-9.68	-17.67	0.24	0.35	-0.08
71	-9.46	-11.48	-1.06	0.15	-0.96	94	-9.80	-21.70	-0.80	0.46	0.50
72	-8.58	-12.83	-0.99	-0.68	-0.77	95	-9.48	-9.09	-0.75	0.16	-1.30
73	-8.30	-9.98	-1.68	-0.94	-1.18	96	-9.39	-14.60	-	0.08	-0.52
74	-9.78	-6.13	-0.04	0.44	-1.73	97	-9.48	-13.98	-	0.16	-0.60
75	-8.99	-15.85	-0.75	-0.30	-0.34						

Expt., experimental values; Pred., predicted values; $pIC_{50} = - \log_{10} IC_{50}$; [a] based on equation (3) and [b]as per equation (4).

Table 7.5. (c). Predicted cytotoxic activities of Pyrazolignans and Isoxazolignans podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity for selected analogues.

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})	Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})
98	-9.83	-19.79	-0.28	0.49	0.23	110	-8.65	-13.10	-1.00	-0.62	-0.73
99	-10.01	-23.89	0.00	0.66	0.81	111	-6.41	-10.39	-	-2.71	-1.12
100	-9.68	-18.77	-0.61	0.35	0.08	112	-8.48	-11.44	-1.32	-0.78	-0.97
101	-8.82	-15.37	-0.67	-0.46	-0.41	113	-7.90	-15.68	-	-1.32	-0.36
102	-8.97	-8.05	-	-0.31	-1.45	114	-9.69	-21.66	-0.34	0.36	0.49
103	-9.21	-19.24	-0.65	-0.09	0.15	115	-9.26	-18.67	-0.34	-0.04	0.07
104	-8.69	-20.62	-	-0.57	0.34	116	-8.21	-11.70	-1.36	-1.03	-0.93
105	-10.25	-22.50	0.00	0.88	0.61	117	-8.51	-13.72	-1.08	-0.75	-0.64
106	-7.37	-10.61	-	-1.81	-1.09	118	-9.21	-18.62	-0.41	-0.09	0.06
107	-9.14	-14.83	-0.75	-0.15	-0.48	119	-9.56	-18.00	-0.38	0.23	-0.03
108	-9.63	-19.87	-	0.31	0.24	120	-6.22	-6.29	-	-2.89	-1.70
109	-7.85	-10.88	-1.32	-1.37	-1.05						

Expt., experimental values; Pred., predicted values; $pIC_{50} = - \log_{10} IC_{50}$; [a] based on equation (3) and [b]as per equation (4).

Table 7.5. (d). Predicted cytotoxic activities of lactonic and non-lactonic naphthalene podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity for selected analogues.

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})
121	-8.50	-16.07	-0.71	-0.75	-0.30
122	-6.73	-12.90	-1.65	-2.41	-0.76
123	-7.72	-16.23	-1.09	-1.49	-0.28
124	-7.68	-15.94	-0.77	-1.52	-0.32
125	-7.07	-14.72	-1.22	-2.09	-0.50
126	-8.76	-20.35	-0.18	-0.51	0.30

Expt., experimental values; Pred., predicted values; $pIC_{50} = - \log_{10} IC_{50}$; [a] based on equation (3) and [b]as per equation (4).

Table 7.5. (e). Predicted cytotoxic activities of Aza-podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor and experimental activity for selected analogues.

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})	Ligand	Glide Score	ΔG_{bind} (kcal/mol)	Expt. pIC_{50}	[a]Pred. pIC_{50} (Gscore)	[b]Pred. pIC_{50} (ΔG_{bind})
127	-7.18	-13.22	-2.00	-1.99	-0.71	141	-11.18	-27.83	2.74	1.76	1.38
128	-7.41	-15.09	-1.90	-1.78	-0.45	142	-10.85	-29.75	2.77	1.45	1.65
129	-7.71	-11.53	-2.00	-1.50	-0.95	143	-8.11	-17.20	-0.69	-1.12	-0.14
130	-7.88	-13.64	-1.59	-1.34	-0.65	144	-9.28	-18.56	0.12	-0.03	0.05
131	-8.29	-19.66	-0.30	-0.95	0.21	145	-8.78	-22.20	0.11	-0.49	0.57
132	-6.87	-13.81	-1.46	-2.29	-0.63	146	-8.02	-18.22	-0.41	-1.21	0.00
133	-6.71	-10.85	-2.00	-2.43	-1.05	147	-10.48	-28.47	2.39	1.10	1.47
134	-7.63	-12.37	-1.80	-1.57	-0.83	148	-8.99	-19.49	0.04	-0.30	0.18
135	-7.54	-15.25	-1.60	-1.66	-0.42	149	-9.28	-28.47	1.32	-0.02	1.47
136	-7.17	-14.15	-2.00	-2.00	-0.58	150	-9.83	-30.30	2.28	0.49	1.73
137	-7.03	-10.93	-2.00	-2.13	-1.04	151	-8.95	-24.99	0.89	-0.34	0.97
138	-7.38	-16.19	-1.78	-1.81	-0.29	152	-10.74	-27.80	2.28	1.34	1.37
139	-6.92	-10.68	-2.00	-2.24	-1.08	153	-9.14	-29.99	1.52	-0.16	1.68
140	-6.99	-12.37	-1.85	-2.17	-0.83	154	-9.62	-25.29	1.55	0.29	1.01

Expt., experimental values; Pred., predicted values; $pIC_{50} = - \log_{10} IC_{50}$; [a] based on equation (3) and [b]as per equation (4).

We have used Prime/MM-GBSA protocol for rescoring Glide XP poses of the podophyllotoxin analogues. From the results collected in Table 7.5 (a-e) we didn't find correlation between Glide score and ΔG_{bind} energy ($r^2 = 0.3175$) (Figure 7.5). However, we did find a better correlation between ΔG_{bind} energy and experimental pIC_{50} ($r^2 = 0.7285$) (Figure 7.6).

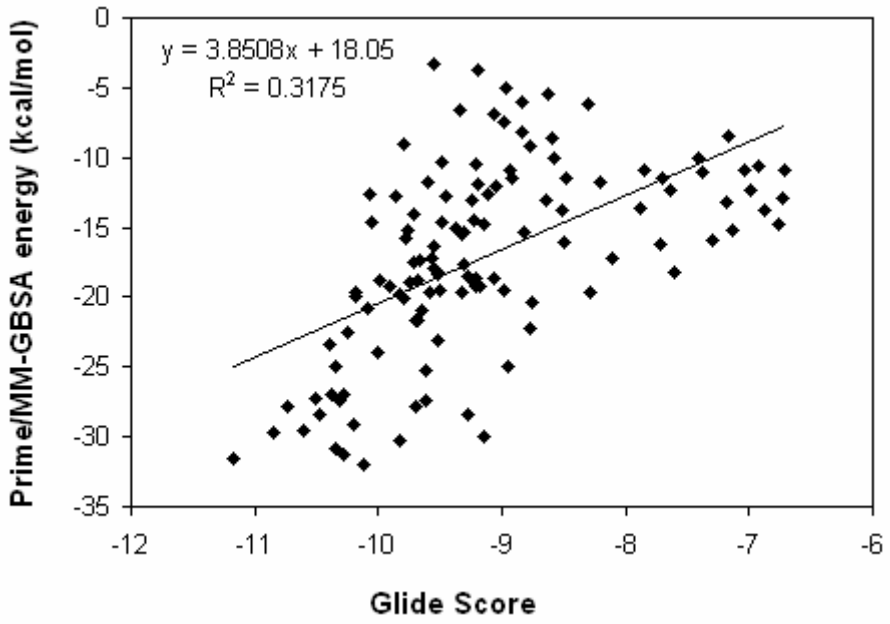


Figure 7.5. Relationship between Glide score and Prime/MM-GBSA energy.

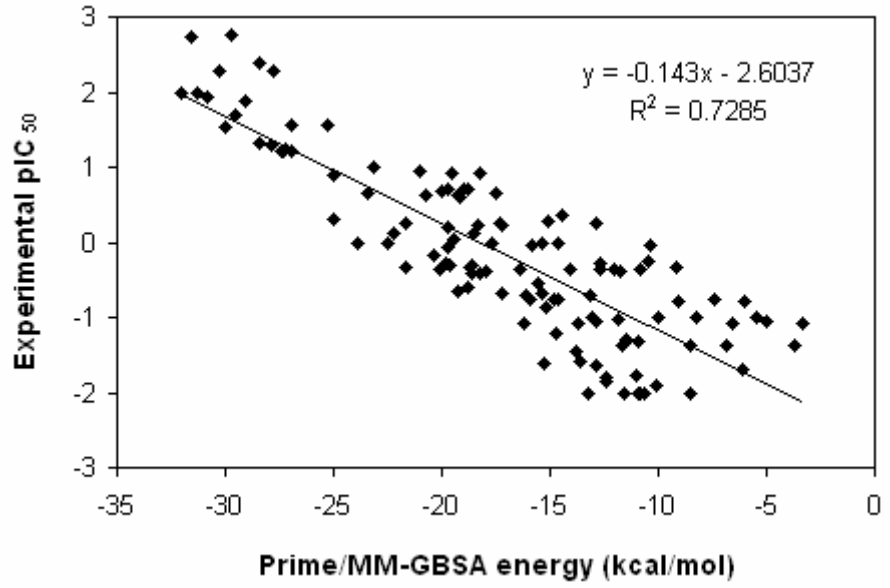


Figure 7.6. Models for predicting cytotoxic activity (pIC_{50}) of the podophyllotoxin analogues based on Prime/MM-GBSA energy (ΔG_{bind}).

Rescoring using Prime/MM-GBSA leads to minor changes of the ligand conformations (due to energy minimization of the ligand in receptor's environment) and consequent stabilization of receptor and ligand complex. A linear regression model for prediction of predicted pIC_{50} of cytotoxicity has been developed by considering some analogues with known pIC_{50}. In this model we have taken ΔG_{bind} energy as a descriptor. The equation (4) of the model and the corresponding statistics are shown below:

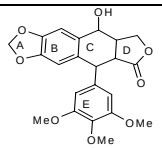
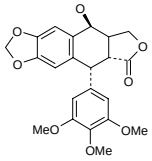
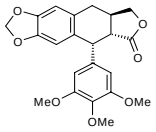
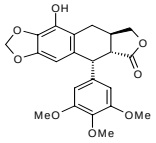
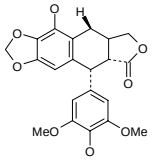
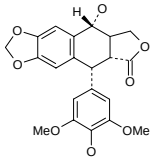
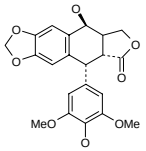
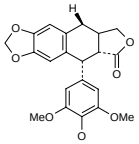
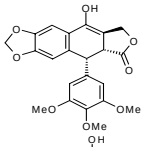
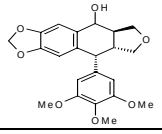
$$pIC_{50} = -2.604(\pm 0.008) - 0.143(\pm 0.148)^* \Delta G_{bind} \qquad (4)$$

$$(N = 120, r^2 = 0.728, s = 0.603, F = 316.58, r^2_{cv} = 0.719, PRESS = 44.415)$$

The statistical significance of the prediction model is evaluated by the correlation coefficient r^2, standard error s, F-test value, leave-one-out cross-validation coefficient r^2_{cv} and predictive error sum of squares PRESS. The regression model developed based on ΔG_{bind} energy is statistically (r^2_{cv} = 0.719, r^2 = 0.728, F = 316.58) best fitted and consequently used for prediction of cytotoxic activities (pIC_{50}) of the podophyllotoxin analogues as reported in Table 5(a-e). The average root mean square error between predicted and experimental pIC_{50} values was 0.770 μM by using leave-one-out cross validation technique which further revealed the reliability of the model for prediction of cytotoxicity. However, we may observe that model using ΔG_{bind} descriptors are better for predicting cytotoxicity (pIC_{50}) with a root mean square error of 0.575 μM than model using Glide score as a descriptor.

To judge the accuracy of the prediction models developed based on Glide score and ΔG_{bind} energy for predicting tubulin polymerization inhibition potencies, we have taken a separate data set called as validation test consisting of 16 compounds (Table 7.6). Their potencies and chemical structures were obtained from literature (Haar et al., 1996; Loike et al., 1978). Experimentally determined relative potencies of the drugs based on in vitro study are also provided in order to evaluate the accuracy of predictions. For all compounds, both the prediction models (equations 3 & 4) produce exactly the same trend for relative potencies, even though the exact magnitudes of these values do not match very well (Table 7.7).

Table 7.6. The experimental IC_{50} value for in vitro tubulin polymerization inhibition by podophyllotoxin analogues.

Analogue	Name	Structure	IC_{50} (μM)
1	Podophyllotoxin		0.6
2	Epipodophyllotoxin		5.0
3	Deoxypodophyllotoxin		0.5
4	β-Peltatin		0.7
5	α-Peltatin		0.5
6	4'-Demethylpodophyllotoxin		0.5
7	4'-Demethylepipodophyllotoxin		2.0
8	4'-Demethyldeoxypodophyllotoxin		0.2
9	Dehydropodophyllotoxin		25
10	Anhydropodophyllol		1.0

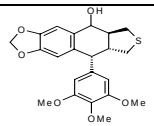
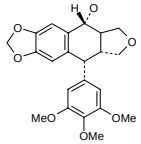
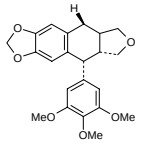
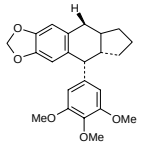
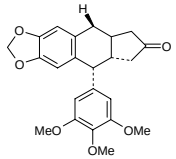
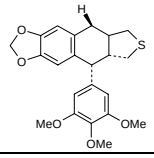
11	Podophyllotoxin cyclic sulfide		10
12	Podophyllotoxin-cyclic ether		1.0
13	Deoxypodophyllotoxin-cyclic ether		0.8
14	Deoxypodophyllotoxin-cyclopentane		5.0
15	Deoxypodophyllotoxin-cyclopentanone		5.0
16	Deoxypodophyllotoxin-cyclic sulfide		10

Table 7.7. Predicted inhibition of in vitro microtubule assembly by podophyllotoxin analogues using Glide score (XP) and Prime/MM-GBSA energy as a descriptor of podophyllotoxin analogues (16 compounds).

Ligand	Glide Score	ΔG_{bind} (kcal/mol)	[a]Ext. pIC_{50}	[b]Pred. pIC_{50}	[c]Pred. pIC_{50}
Podophyllotoxin	-9.54	-19.62	0.22	0.22	0.20
Epipodophyllotoxin	-9.03	-14.37	-0.70	-0.26	-0.55
Deoxypodophyllotoxin	-9.44	-21.15	0.30	0.12	0.42
β-Peltatin	-8.88	-19.34	0.15	-0.40	0.16
α-Peltatin	-9.09	-20.96	0.30	-0.20	0.39
4'-Demethylpodophyllotoxin	-9.70	-19.70	0.30	0.37	0.21
4'-Demethylepipodophyllotoxin	-8.75	-17.46	-0.30	-0.52	-0.11
4'-Demethyldeoxypodophyllotoxin	-9.54	-22.64	0.70	0.21	0.63
Dehydropodophyllotoxin	-7.74	-10.93	-1.40	-1.47	-1.04
Anhydropodophyllol	-9.35	-17.70	0.00	0.04	-0.07
Podophyllotoxin cyclic sulfide	-8.10	-14.24	-1.00	-1.13	-0.57
Podophyllotoxin-cyclic ether	-9.53	-16.63	0.00	0.21	-0.23
Deoxypodophyllotoxin-cyclic ether	-9.35	-22.28	0.10	0.04	0.58
Deoxypodophyllotoxin-cyclopentane	-8.70	-14.26	-0.70	-0.57	-0.56
Deoxypodophyllotoxin-cyclopentanone	-8.62	-16.48	-0.70	-0.64	-0.25
Deoxypodophyllotoxin-cyclic sulfide	-8.27	-13.47	-1.00	-0.97	-0.68

[a]Expt., experimental values; $pIC_{50} = - \log_{10} IC_{50}$

[b]Pred. pIC_{50}; predicted pIC_{50} based on Glide score and was calculated using equation (3).

[c]Pred. pIC_{50}; predicted pIC_{50} based on Prime energy and was calculated using equation (4).

The overall RMSE between the experimental and predicted pIC_{50} value was 0.27 μM and 0.25 μM respectively by using Glide score and ΔG_{bind} energy, which means that the docking and Pime/MM-GBSA modeling was able to predict the cytotoxic activity of 16 podophyllotoxin analogues more reliably. Figure 7.7a & 7.7b graphically shows the quality of fit for the validation set.

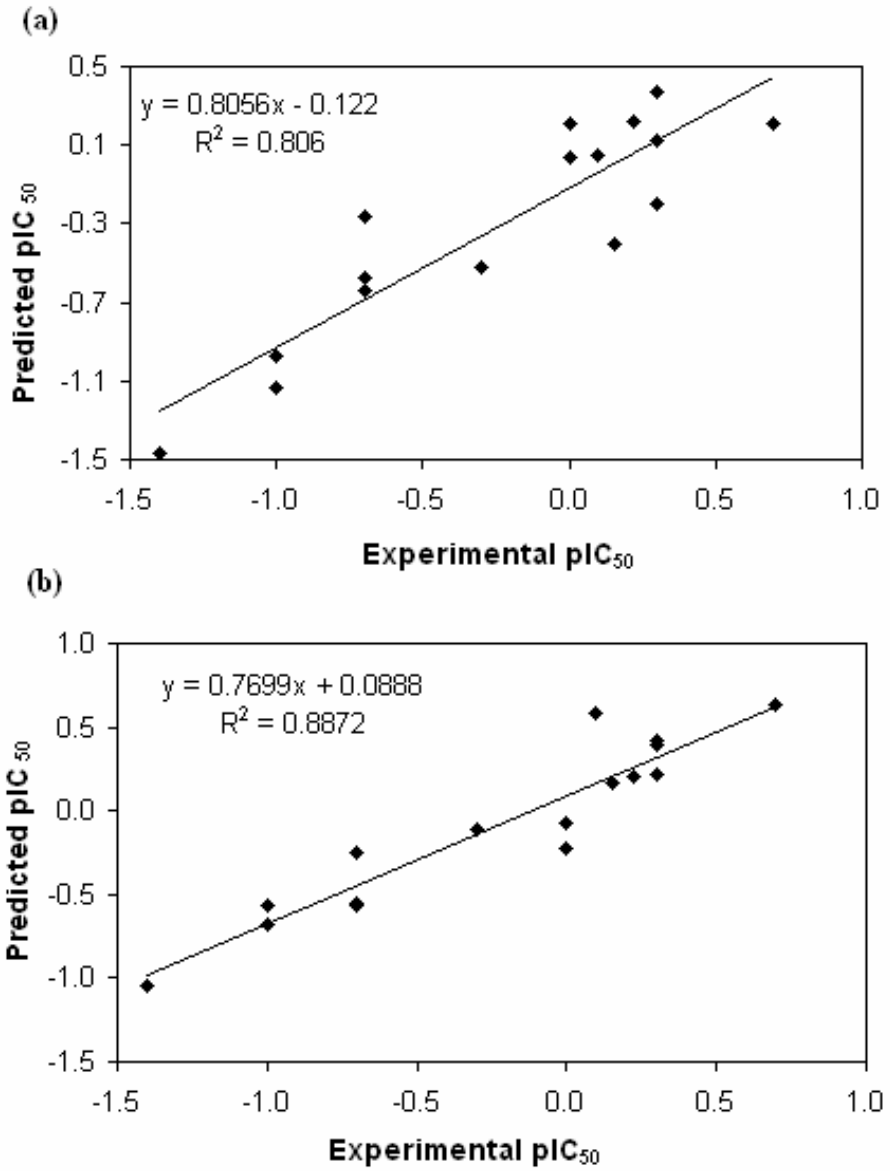


Figure 7.7. (a-b). Relationship between experimental and predicted pIC50 values of the validation set (16 compounds) using (a) Glide score and (b) Prime/MM-GBSA energy.

7.4 Biological significance

The podophyllotoxin structural derivatives available till now belong either to only A, B, C, D, E or combination of ring modifications resulting in wide range of cytotoxicity activity. The modifications involving the C and D rings (lactone ring) gave the most promising results on activity (Gordliza et al., 2000). Reports have been made of *cis* and *trans*-lactones isomers, appearing either naturally or synthesized by transformations and inter-conversions (San Feliciano et al., 1989; San Feliciano et al., 1993; San Feliciano et al., 1991). Furthermore, within the lactonic group those compounds displaying a *trans*-junction between the tetraline and lactone fragments were more potent than their *cis*-analogues (San Feliciano et al., 1993; Gordaliza et al., 1994; Dore et al., 1996). Based on docking study and rescoring using Prime/MM-GBSA we have seen that the *trans*-lactones have more negative Glide score and ΔG_{bind} energy value in comparison to *cis*-lactones. The observed differences in ΔG_{bind} energy and Glide score between *trans*- and *cis*-lactones could be explained in terms of their conformation and spatial arrangement of the lactone ring in relation to the other fused rings. Whereas, in the *trans*-lactones the four rings are almost co-planar, in the main conformer of *cis*-derivatives the lactone adopts an almost perpendicular disposition with respect to the other three rings. Both Glide score and ΔG_{bind} energy revealed that the lactonic group of podophyllotoxin ligands binds to tubulin protein with high affinity and showed better activity (Table 5a). In general the non-lactonic group of podophyllotoxin analogues is less potent as antitumor agents. The Glide score and ΔG_{bind} energy of these compounds have been seen to be less negative in comparison to lactonic tetralines leading to less potent interaction with tubulin protein and thus lower activity (Table 5b). Thus the presence of lactone moiety is very essential for better activity. However, aldehydes at position 9 are more potent than alcohols at this position. This is in accordance with the previous suggestion that an electrophilic group at this position is critical for the possible interaction with the biomolecules. Other transformations on the lactone ring include its reduction leading again to much less cytotoxic compounds and acetylation of the hydroxyl groups does not modify potency (San Feliciano et al., 1993; Brewer et al., 1979; Dore et al., 1996). The predicted pIC_{50} values of the analogues calculated based on Glide score and ΔG_{bind} energy value as descriptors are very close to experimental pIC_{50} revealing good prediction models. The pyrazoline and isoxazoline derivatives were less potent than podophyllotoxin as

cytotoxicity. Computational techniques based on Glide docking and rescoring using Prime/MM-GBSA also revealed similar conclusion. The pyrazoline derivatives tested so far showed cytotoxicity activity to two and three orders of magnitude lower than those of podophyllotoxin, thus confirming that the presence of the lactone moiety as a prominent requirement for high cytotoxic activity to be achieved. The lactonic and non-lactonic naphthaline group of derivatives has proportionately much lower activity (San Feliciano et al., 1993; Gordaliza et al., 1994; Dore et al., 1996) in comparison to other groups. The Glide score and ΔG_{bind} energy value of this group of compounds are not so good and thus seem to interact with the tubulin protein less efficiently. Aza-podophyllotoxin analogues were generally prepared by selective chemical modification of the two aromatic rings (ring B and E) of natural podophyllotoxin (Hitotsuyanagi et al., 2000). The Glide score among the ligands of these libraries vary in between −11.18 and -6.71 and the overall difference is also very small -5.07. It revealed that these entire ligands bind in tubulin protein with high affinity showed activity (pIC_{50}) in between -2.00 μM and 2.77 μM. A few aza-podophyllotoxin analogues were proved to be more than twice as cytotoxic as natural podophyllotoxin (Hitotsuyanagi et al., 2000). Among aza-podophyllotoxin derivatives the best activity, was found for structure 142 (pIC_{50} = 2.77 μM) with better Glide score (-10.85) within the library. Similarly, the ΔG_{bind} energy among the ligands of these libraries varies in between −30.30 to -10.68 kcal/mol and the overall difference is also very small (-19.31 kcal/mol). It revealed that all these ligands bind to tubulin protein with high affinity showed activity (pIC_{50}) in between -2.00 μM and 2.77 μM. Since these groups of analogues bind with tubulin and inhibit microtubule polymerization, the relationship obtained between Glide score and cytotoxic activity is more probable. Moreover, the linear regression model for prediction of predicted pIC_{50} of cytotoxicity developed by considering Glide score and ΔG_{bind} energy as descriptors (equations 3 & 4) seems to be accurate. Reasonably, good agreement between predicted and experimental pIC_{50} is found suggesting that the calculated pIC_{50} based on Glide score and ΔG_{bind} energy are robust and accurate.

7.5 Conclusion

We have compiled a virtual library of podophyllotoxin analogues built through structural modification of scaffold structure of natural podophyllotoxin. Docking and rescoring have been done using Prime/MM-GBSA in the work to get insights into ligand:tubulin interactions and corresponding cytotoxic activity of podophyllotoxin analogues. In the docking simulations, the flexible docking reproduced the binding structure of crystal structures well. These experiments verified the docking protocol adopted in the work. Also the docking simulations of structurally similar inhibitors showed that the docking simulation could dock inhibitors into a receptor comparable to the crystal structure complex with podophyllotoxin. Several sets of podophyllotoxin analogues have been studied in the docking simulations. Results showed that these analogues bind in a very similar mode. The magnitude of the binding affinity can be a key factor that decides the activeness of an individual inhibitor. An energetic evaluation of the binding affinity will provide a way to estimate the activity of inhibitors. In any binding energy calculation, the correct binding structure of each ligand has to be determined first prior to binding energy estimation. Only the binding structure of podophyllotoxin with tubulin is available. But the binding structures of the analogues of podophyllotoxin are not available. We use flexible docking to determine the binding structure of the podophyllotoxin analogues with tubulin protein. Very similar binding structures were obtained for a set of analogues. This makes a credible prediction model of the cytotoxic activity (pIC_{50}) calculation possible. The calculated Glide score and binding free energy value of a set of structural analogues demonstrates excellent linear correlation to the experimental cytotoxic activity. These models could be useful to predict the range of activities for new podophyllotoxin analogues. We also found that refinement of poses and consequent rescoring using Prime/MM-GBSA lead to better predictivity of pIC_{50}. The information that we have expressed in this study may lead to the designing (synthesis) of more potent podophyllotoxin derivatives for inhibition of microtubule polymerization.

CHAPTER 8

Application of linear interaction energy method for binding affinity calculations of podophyllotoxin analogues with tubulin using continuum solvent model and prediction of cytotoxic activity

Abstract

Podophyllotoxin and its analogues have important therapeutic value in the treatment of cancer, due to their ability to induce apoptosis in cancer cells in a proliferation-independent manner. These ligands bind to colchicine binding site of tubulin near the α- and β-tubulin interface and interfere with tubulin polymerization. The binding free energies of podophyllotoxin-based inhibitors of tubulin were computed using a linear interaction energy (LIE) method with a surface generalized Born (SGB) continuum solvation model. A training set of 76 podophyllotoxin analogues was used to build a binding affinity model for estimating the free energy of binding for 36 inhibitors (test set) with diverse structural modifications. The average root mean square error (RMSE) between the experimental and predicted binding free energy values was 0.56 kcal/mol which is comparable to the level of accuracy achieved by the most accurate methods, such as free energy perturbation (FEP) or thermodynamic integration (TI). The squared correlation coefficient between experimental and SGB-LIE estimates for the free energy for the test set compounds is also significant (R^2 = 0.733). On the basis of the analysis of the binding energy, we propose that the three-dimensional conformation of the A, B, C and D rings is important for interaction with tubulin. On the basis of this insight, 12 analogues of varying ring modification were taken, tested with LIE methodology and then validated with their experimental potencies of tubulin polymerization inhibition. Low levels of RMSE for the majority of inhibitors establish the structure-based LIE method as an efficient tool for generating more potent and specific inhibitors of tubulin by testing rationally designed lead compounds based on podophyllotoxin derivatization.

8.1 Introduction

Microtubules are involved in a wide range of cellular functions and are critical to the life cycle of the cell. Composed of alternating α- and β-protofilaments, microtubules are highly dynamic macromolecular assemblies that are organized in a polar, spatial and temporal cell cycle specific manner. The organization is regulated by numerous factors including the intrinsic ability of microtubule subunits, tubulin heterodimers, to form non-equilibrium, dynamic polymers. The α- and β-tubulin rapidly assemble and disassemble to meet the cell's needs (Downing et al., 1998; Downing et al., 1998). Since inhibition of tubulin polymerization or blockage of microtubule disassembly increases the number of cells in metaphase arrest, microtubules are attractive molecular targets for anticancer therapeutics. Small molecules have been shown to bind at four major drug binding sites on tubulin: the vinca, taxane, colchicine and peloruside A (Jordan et al., 1998; Cragg et al., 2004; Huzil et al., 2008).

Among the plethora of physiological activities and potential medicinal and agricultural applications, the antineoplastic and antiviral properties of podophyllotoxin congeners and their derivatives are arguably the most eminent from a pharmacological perspective. Podophyllotoxin is an antitumor lignan mainly found in the plants *Podophyllum hexandrum* and *Podophyllum peltatum*. Since the discovery of the therapeutic properties of podophyllotoxin, new findings related to its activities, its mechanism of action and pharmacological properties have been unveiled. Structure-activity Relationships (SAR) have shown that podophyllotoxin analogues preferentially inhibit tubulin polymerization, which leads to arrest of the cell cycle in the metaphase (Snyder et al., 1976; Margolis et al., 1978). Different derivatives of podophyllotoxin have been demonstrated to bind to the colchicine site, as shown by the fact that podophyllotoxin has been reported to compete with colchicine for the binding site in tubulin (Cortese et al., 1977) and its affinity is double than that of colchicine. These compounds including colchicine affect cancer and normal cells alike and lead to the appearance of adverse side effects (Ayres et al., 1990). Following binding of podophyllotoxin, the GTP hydrolyzing capacity of tubulin is inhibited, but colchicine stimulates an assembly-independent GTPase activity directed at the exchangeable site-bound GTP (Lin et al., 1981). Podophyllotoxin binds to β-tubulin at its interface with α-tubulin

resulting in inhibition of tubulin polymerization. This binding mode was recently confirmed by the determination of a 4.20 Å X-ray structure of α- and β-tubulin complexed with podophyllotoxin (PDB_ID:1SA1), showing that podophyllotoxin also binds at the colchicine site (Ravelli et al., 2004).

While podophyllotoxin has played a central role in elucidating the physical properties and biological functions of tubulin and microtubules, its high toxicity has limited its therapeutic application (Schilstra et al., 1989). Although colchicine site agents share a general toxicity, the promise to discover therapeutically useful analogues has fueled continued research. Over the years, a large number of natural and synthetic analogues of podophyllotoxin have been identified as colchicine site inhibitors. Since a wide variety of molecular scaffolds are available for optimization, this diversity presents a significant challenge to determining the essential features for activity. A rational approach for the discovery of a pharmaceutically acceptable, economically viable activity model awaits development of a predictive quantitative structure-activity relationship. With the advent of parallel synthesis methods and technology, we might expect the number of podophyllotoxin analogues to be tested to grow dramatically. Combinatorial methods could also be envisioned as a semi-rational approach to this discovery strategy. One method of orchestrating these strategies is to make use of linear interaction energy (LIE) models for the rapid prediction and virtual prescreening of cytotoxic activity. The linear interaction energy approximation is a way of combining molecular mechanics calculations with experimental data to build a model scoring function for the evaluation of ligand-protein binding free energies. The LIE method (Åqvist et al., 1994) is a semi-empirical model that has become widely used to predict protein-ligand binding affinities. In LIE, the free ligand in water and the solvated protein-ligand complex are simulated and from these two calculations the ligand surrounding electrostatic and van der Waals (vdw) energies are collected. The binding free energy is then evaluated as proposed by Åqvist et al (1994). A continuum solvation model was developed based on the proposed LIE method by adding continuum electrostatic ligand-water interaction energies by using an equivalent form of equation (Carlsson et al., 2006). However, the proposed generalized Born (GB)-LIE method overestimates the change in solvation energy and this is caused by consistent

underestimation of the effective Born radii in the protein-ligand complex (Carlsson et al., 2006). To further assess the usefulness of continuum models for estimating binding free energies, more accurate GB models should be carried out. The LIE method has been applied on a number of protein-ligand systems with promising results producing small errors on the order of 1 kcal/mol for free energy prediction (Zhou et al., 2001). This approach could then be applied to larger sets of inhibitors and contribute to fast and efficient ligand design. At present, a linear interaction energy method for rational design of podophyllotoxin analogues for tubulin polymerization inhibition has not been determined.

The availability of structural information on tubulin facilitates understanding the structure-activity relationships (SAR) for tubulin polymerization inhibition. In this study, we have applied a structure-based linear interaction energy method implementing a surface generalized Born (SGB) continuum model for solvation to build a binding affinity model for estimating the binding free energy for a diverse set of podophyllotoxin analogues with tubulin. The magnitude of free energy changes upon binding of inhibitors to tubulin directly correlates with the experimental potency of these inhibitors; hence, fast and accurate estimation of binding free energies provides a means to screen the compound libraries for lead optimization and for generating more potent and specific inhibitors of tubulin by testing rationally designed lead compounds based on podophyllotoxin derivatization.

8.2 Materials and methods

8.2.1 LIE Methodology

The LIE method employs experimental data on binding free energy values for a set of ligands (referred as training set) to estimate the binding affinities for a set of novel compounds. The method is based on the linear response approximation (LRA), which dictates that binding free energy of a protein-ligand system is a function of polar and non-polar energy components that scale linearly with the electrostatic and van der Waals interactions between a ligand and its environment. The free energy of binding (FEB) for the complex is derived from considering only two states: (1) free ligand in the solvent and (2) ligand bound to the solvated protein. The conformational changes and entropic effects pertaining to unbound receptor are taken into account implicitly and only interactions

between the ligand and either the protein or solvent are computed during molecular mechanics calculations. Among the various formulations of the LIE methodology developed in the past, the SGB-LIE method (Zhou et al., 2001) has been shown to be 1 order of magnitude faster than the methods based on explicit solvent with the same order of accuracy. In the LIE method,

$$\Delta G_{bind} = \alpha \langle \Delta U_{ele} \rangle + \beta \langle \Delta U_{vdw} \rangle + \gamma \langle \Delta_{SASA} \rangle \qquad (1)$$

where $\langle \Delta U_{ele} \rangle$ *and* $\langle \Delta U_{vdw} \rangle$ denotes the average change in the electrostatic and van der Waals interaction energy of the ligand in the free and bound states, respectively and $\langle \Delta_{SASA} \rangle$ is the change in the solvent-accessible surface area (SASA) of the ligand. The α, β, and γ terms are adjustable parameters that need to be determined by fitting the experimental data on the training set compounds. The SGB-LIE method also offers better accuracy in treating the long-range electrostatic interactions. However, the SGB-LIE method used in this studied is based on the original formulation proposed by Jorgensen and implemented in *Liaison* (Schrödinger, Inc. Portland, OR) using the OPLS-2005 force field. A novel feature of Liaison is that the simulation takes place in implicit (continuum) rather than explicit solvent—hence the name Liaison, for Linear Interaction Approximation in Implicit Solvation. The explicit-solvent version of the methodology was first suggested by Aqvist (Hansson et al.,1995), based on approximating the charging integral in the free-energy-perturbation formula with a mean-value approach, in which the integral is represented as half the sum of the values at the endpoints, namely the free and bound states of the ligand. The empirical relationship used by Liaison is shown below:

$$\Delta G_{bind} = \alpha (\langle U^b_{ele} \rangle - \langle U^f_{ele} \rangle) + \beta (\langle U^b_{vdw} \rangle - \langle U^f_{vdw} \rangle) + \gamma (\langle U^b_{cav} \rangle - \langle U^f_{cav} \rangle) \qquad (2)$$

Here < > represents the ensemble average, *b* represents the bound form of the ligand, *f* represents the free form of the ligand, and α, β and γ are the coefficients. U_{ele}, U_{vdw} and U_{cav} are the electrostatic, van der Waals and cavity energy terms in the SGB continuum solvent model. The cavity energy term, U_{cav}, is proportional to the exposed surface area of the ligand. Thus, the difference: $\langle U^b_{cav} \rangle - \langle U^f_{cav} \rangle$ measures the surface area lost by contact with the receptor. The energy terms involved can be computed using energy minimization, molecular dynamics, or Monte Carlo calculations. In the SGB model of solvation, there is no explicit van der Waals or electrostatic interaction between the solute and solvent. The contribution for net free energy of solvation comes from two energy terms, namely, reaction

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

8.2.2 Computational details

Preparation of receptor and ligands was done using the Schrödinger package from Schrödinger Inc (Portland, 2004). All the calculations for the SGB-LIE method were performed in the Liaison package from Schrödinger Inc (Liaison et al., 2005). The Liaison module performs LIE calculations in the OPLS force field with a residue-based cutoff of 15Å. The OPLS force field was also used for charge assignment and all energy calculations.

8.2.3 Receptor preparation

The X-ray structure of the complex between podophyllotoxin and tubulin protein (PDB_ID: 1SA1) has been used as initial structure in the preparation of podophyllotoxin binding site. After manual inspection and cleaning of structure we retained a complex composed of protein chains α and β and podophyllotoxin ligand. Hydrogen were added to the model automatically via the Maestro interface (Schrodinger L. L. C. 2007) leaving no lone pair and using an explicit all-atom model. All the water molecules were removed from the complex. The multi step Schrödinger's Protein preparation tool (PPrep) has been used for final preparation of protein. PPrep neutralizes side chains that are not close to the binding cavity and do not participate in salt bridges (Schrodinger L. L. C. 2007). This step is then followed by restrained minimization of co-crystallized complex, which reorients side chain hydroxyl groups and alleviates potential steric clashes. Progressively weaker restraints (tethering force constants 3, 1, 0.3, 0.1) were applied to nonhydrogen atoms only. The complex structure was energy minimized using OPLS_2005 force field and the conjugate gradient algorithm, keeping all atoms except hydrogen fixed. The minimization was stopped

either after 1000 steps or after the energy gradient converged below 0.01 kcal/mol. The energy-minimized receptor structure was subsequently used for docking of podophyllotoxin analogues and SGB-LIE calculations.

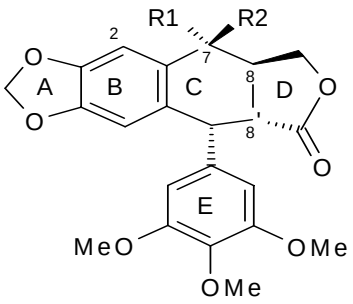
8.2.4 Preparation of ligands

Podophyllotoxin is well known for its antitumor activity. However, the clinical application of it and its analogues in the treatment of cancer has been limited by severe toxic side effects during administration of the drugs (Jardine et al., 1980; Weiss et al., 1975). With a view to achieving greater therapeutic efficiency many podophyllotoxin analogues have been isolated and via molecular manipulation, a large number of semisynthetic derivatives have been synthesized. However, new findings related to their activities, mechanism of action and pharmacological properties have been unexplored. A total of 112 podophyllotoxin analogues were used in the study and were taken from various sources belonging to different ring modifications. For better interpretation all these compounds were divided into following 4 sublibraries.

*Sublib-**I*** commonly known as tetralinelactones consist of 29 compounds (1-29) (Table 8.1a). These molecules were rationally designed as functional mimics of natural podophyllotoxin with the goal of simplifying the chemical synthesis and improving the cytotoxic activity. Structural modification mainly introduced varying radicals at position 7 in podophyllotoxin scaffold. Reports have been made of compounds with oxygenated substituents in the form of ethers, esters and diverse nitrogen radicals (San Feliciano et al., 1989; San Feliciano et al., 1993; Miguel et al., 1995; Doré et al., 1996; Gordaliza et al., 1996).

*Sublib-**II*** contains compounds (30-70) (Table 8.1b) known as non-lactonic tetralines. Structural modifications in this group include the opening of the lactone ring (D-ring) in podophyllotoxin scaffold, to give rise to compounds with different degrees of oxidation at positions C-9 and C-9' (San Feliciano et al., 1993; Miguel et al., 1995; Doré et al., 1996; Gordaliza et al., 1996). In general these molecules lacking a lactone ring.

Table 8.1. (a). Podophyllotoxin derivatives (Tetraline lactones) with cytotoxic activities against P-388 cell line used in the work.

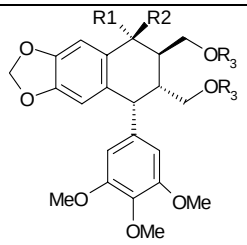


Analogue	R1	R2	Expt.IC_{50}	Analogue	R1	R2	Expt.IC_{50}
1	OH	H	0.012	16	H	H	0.10
2	H	H	0.010	17	H	H(2-OMe)	0.23
3	H	H(2-OMe)	0.01	18	OH	H	6.0
4	OH	H(4'-OH	0.027	19	OAc	H	0.55
5	OAc	H	0.625	20	OAc	H(2-OMe)	1.02
6	OMe	H	0.06	21	OMe	H	0.12
7	H	OH	0.06	22	H	OH(2-OMe)	0.11
8	H	Ac	0.05	23	H	OAc	0.44
9	H	OMe	0.06	24	H	OAc(2-OMe)	0.51
10	H	Cl	0.6	25	H	OMe	0.12
11	Cl	H	0.6	26	H	H Δ^7	0.013
12	=O		1.8	27	=O		12.0
13	= N-OH		2.3	28	=N-OH		2.3
14	=N-OAc		2.1	29	=N-OMe		2.3
15	=N-OMe		0.2				

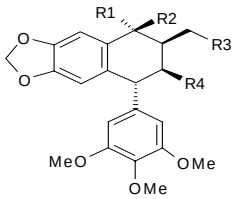
*Sublib-**III*** also includes a group of lignans (71-84) (Table 8.1c) that have heterocyclic rings fused to the cyclolignan skeleton. This group is commonly called as pyrazolignans (San Feliciano et al., 1993; Doré et al., 1996; Gordaliza et al., 1995) and isoxazolignans (Gordaliza et al., 1996; Gordaliza et al., 2001) and they were obtained by reacting podophyllotoxin with differently substituted hydrazines and hydroxylamines.

*Sublib-**IV*** contains 28 compounds (85-112) (Table 8.1d) commonly known as aza-podophyllotoxin analogues. The preparation of this group of compounds requires selective chemical manipulation of the two aromatic rings (B and E-rings) of the podophyllotoxin scaffold. These molecules are readily prepared from anilines, benzaldehydes and tetronic acid or 2, 3-cyclopentanedione in good to excellent yield and have also shown better cytotoxic activity (Hitotsuyanagi et al., 2000).

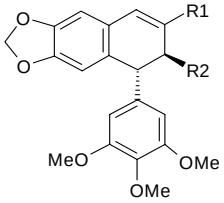
Table 8.1. (b). Podophyllotoxin derivatives (Nonlactonic tetralines) with cytotoxic activities against P-388 cell line used in the work.



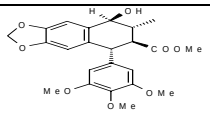
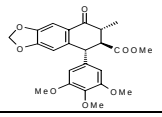
30-34



37-56



57-67

Analogue	R1	R2	R3	Expt.IC_{50}	Analogue	Structure	Expt.IC_{50}
30	OH	H	H	1.2	35		23.3
31	H	OH	H	12.0			
32	H	OMe	H	11.6			
33	H	OMe	Ac	9.7			
34	OMe	H	Ac	9.7	36		3.5

Analogue	R1	R2	R3	R4	Expt.IC_{50}	Analogue	R1	R2	R3	R4	Expt.IC_{50}
37	H	H	OH	COOMe	0.058	47	H	OMe	OAc	CH_2OAc	9.7
38	H	H	OAc	COOMe	0.21	48	H	OH	OH	CH_2OH	47.9
39	H	H	OAc	CH_2OAc	5.14	49	H	OH	OH	COOMe	1.1
40	OH	H	OH	CH_2OH	23.9	50	=O		OH	COOMe	5.63
41	OH	H	OH	COOMe	0.22	51	=O		OAc	COOMe	0.20
42	OAc	H	OAc	CH_2OAc	7.4	52	=N-OH		OAc	COOMe	2.0
43	OAc	H	OAc	COOMe	1.1	53	H	H	CHO	COOMe	2.34
44	OMe	H	OH	CH_2OH	23.2	54	H	H	=N-OMe	COOMe	2.30
45	OMe	H	OAc	CH_2OAc	19.4	55	H	H	=N-OMe	COOMe	10.94
46	H	OMe	OH	CH_2OH	11.6	56	H	H	=N-allyl	COOMe	2.5

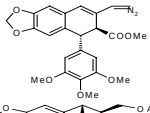
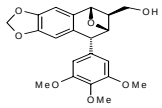
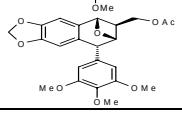
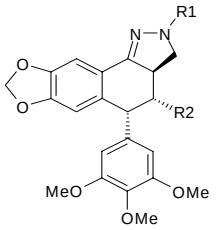
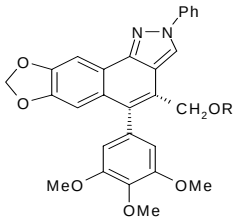
Analogue	R1	R2	Expt.IC_{50}	Analogue	R1	R2	Expt.IC_{50}
57	CH_2OH	COOMe	0.02	64	CH=N-OH	COOMe	2.27
58	CHO	CH_2OH	0.25				
59	CHO	COOMe	0.23	65	CH=N-OMe	COOMe	0.22
60	CH=N-NH2	COOMe	0.57	66		COOMe	0.20
61	CH=N-NH-CH_2CF_3	COOMe	0.48	67		CH_2OH	1.00
62	CH=N-NH-Ph	COOMe	1.94	68			0.57
63	CH=N-NH-Ph	CH_2OH	1.02				
69			6.25				
				70			5.66

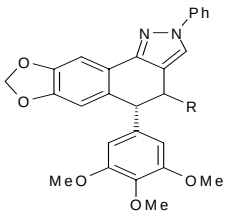
Table 8.1. (c). Podophyllotoxin derivatives (Pyrazolignans and isoxazolignan) with cytotoxic activities against P-388 cell line used in the work.



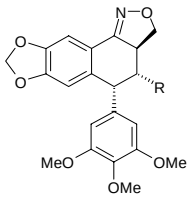
71-76



77



78-79

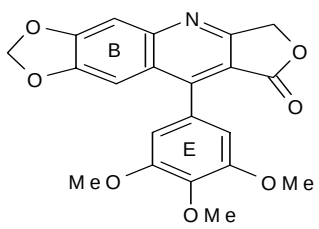


80-84

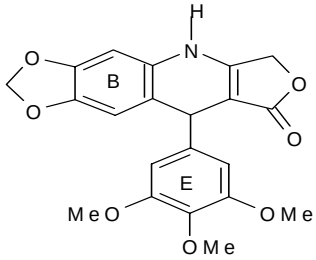
Analogue	R1	R2	Expt.IC_{50}	Analogue	R1	R2	Expt.IC_{50}
71	Ph	COOH	1.9	74	m-NO_2Ph	COOMe	4.5
72	Ph	CH_2OH	4.1	75	Me	COOMe	5.6
73	Ph	CH_2OAc	4.7	76	$COCH_3COOMe$	COOMe	21

Analogue	R	Expt.IC_{50}	Analogue	R	Expt.IC_{50}
77	H	10	81	COOMe	23
78	CHO	21	82	COOMe(4'-OH)	12
79	CH_2Ac	2.2	83	CH_2OH	2.6
80	COOH	2.2	84	CHO	2.4

Table 8.1. (d). Aza-podophyllotoxin derivatives with cytotoxic activities against P-388 cell line used in the work.



Modification 1

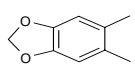


Modification 2

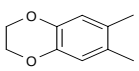
Substitution of B & E ring at 1 and 2 modifications:

Modification 1				Modification 2			
Analogue	B Ring	E Ring	Expt. IC_{50}	Analogue	B Ring	E Ring	Expt. IC_{50}
85	I	VII	100	99	I	VII	0.0018
86	II	VII	80	100	II	VII	0.0017
87	III	VII	100	101	III	VII	4.9
88	III	VIII	39	102	III	VIII	0.76
89	III	XII	2.0	103	III	XII	0.77
90	IV	VII	29	104	IV	VII	2.6
91	V	VII	100	105	V	VII	0.0041
92	VI	VII	63	106	VI	VII	0.92
93	I	VIII	40	107	I	VIII	0.048
94	I	IX	100	108	I	IX	0.0053
95	I	X	100	109	I	X	0.13
96	I	XI	60	110	I	XI	0.0053
97	I	XII	100	111	I	XII	0.030
98	I	VII	71	112	I	VII	0.028

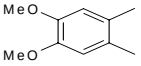
B Ring



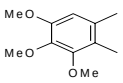
I



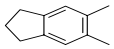
II



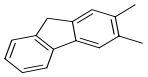
III



IV

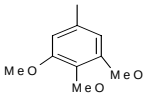


V

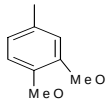


VI

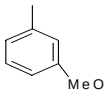
E Ring



VII



VIII



IX

X

XI

XII

All these podophyllotoxin analogues were built from the scaffold by different ring modification and substitution of functional groups as mentioned in Table 8.1(a-d). We used ISIS Draw 2.3 software for sketching structures and converting them to their 3D representation by using ChemSketch 3D viewer of ACDLABS 8.0. LigPrep (Schrodinger L. L. C. 2007) was used for final preparation of ligands from libraries. LigPrep is a utility of Schrödinger software suit that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for tatutomers and steric isomers and performing a geometry minimization of ligands. The ligands were minimized by means of Molecular Mechanics Force Fields (MMFFs) with default setting. Each of these compounds had associated in vitro cytotoxicity values (IC_{50} values reported in μM) against cell line P388. Studied on in vitro cytotoxicity of podophyllotoxin and its analogues were reported mostly on P388 cell line. The reason being due to its resistance to anticancer drug vinorelbine (Marty et al., 2001). P388 is a murine leukemia cell line. Out of the seven β-tubulin isotype classes; class I was the major β-tubulin isotype (60–72%), followed by class III (11.3-11.7 %) while β-tubulin classes IVa + IVb were the least abundant (1.2–1.7%) of total β-tubulin in P388 cell line (Aggarwal et al., 2008).

8.2.5 Docking of the ligands

All the ligands were docked to the tubulin receptor using Glide version 4.0. After ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated using grid-receptor generation program, using van der Waals scaling of the receptor at 0.4. The default size was used for the bounding and enclosing boxes was generated at the centroid of the tubulin binding site by selecting the bound podophyllotoxin ligand. The ligands were docked initially using the “standard precision” method and further refined using “xtra precision” Glide algorithm. For the ligand docking stage, van der Waals scaling of the ligand was set at 0.5. Of the 50,000 poses that were sampled, 4,000 were taken through minimization (conjugate gradients 1,000) and the 30 structures having the lowest energy conformations were further evaluated for the favorable Glide docking score. A single best conformation for each ligand was considered for further analysis.

8.2.6 LIE Calculations

The docked complex corresponding to each analogue was transported to the Liasion package for subsequent SGB-LIE calculations. Sampling technique such as molecular dynamics (MD) has been used for LIE conformation space sampling in the present work. The system was initially heated to 300 K for 5 ps and then subjected to a MD simulation for 25 ps. A residue-based cutoff of 12 Å was set for the non-bonding interactions. The non-bonded pair list was updated every 10 fs. The time integration step of 1.0 fs and sampling LIE energies every 10 steps was used. During the MD simulations, all the residues of the receptor beyond 12 Å from the bound ligand were frozen. Similarly, the average LIE energies for the ligand were obtained using the OPLS-2005 force field. The average LIE energy terms were used for building binding affinity model and free energy estimation for podophyllotoxin analogues. The α, β and γ LIE fitting parameters were determined based on Gaussian elimination method using Matlab 6.5 as described by Thomas & Finny (Thomas et al., 2001) and by fitting the experimental data on the training set compounds.

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

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

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

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

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

Where y_{pred}, $y_{observed}$ and y_{mean} are the predicted, observed and mean values of the cytotoxic activities of the podophyllotoxin analogues and PRESS is the sum of the predictive sum of squares. The predictive ability of the models is expressed by the r^2 predictive value, which is analogous to cross-validated r^2 (q^2).

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

8.3 Results and Discussions

The original crystal structure of tubulin-podophyllotoxin complex (PDB ID: 1SA1) was used to validate the Glide-XP docking protocol. This was done by moving the co-crystallized podophyllotoxin ligand outside of active site and then docking it back into the active site. The top 10 configurations after docking were taken into consideration to validate the result (Table 8.2). The RMSD was calculated for each configuration in comparison to the co-crystallized podophyllotoxin and the value was found to be in between 0.02-0.85 Å.

Table 8.2. The RMSD and docking score from the docking simulation of 10 lowest configurations of co-crystal podophyllotoxin with tubulin (ISA1).

Configuration	Glide Score	[a]ΔG_{score}	[b]RMSD (Å)	[c]RMSD (Å)
1	-10.26	0	0.85	0.60
2	-10.20	-0.06	0.02	0.86
3	-9.80	-0.46	0.68	1.33
4	-9.72	-0.54	0.57	1.26
5	-9.50	-0.76	0.04	0.67
6	-9.25	-1.01	0.04	0.67
7	-8.78	-1.48	0.80	0.59
8	-8.47	-1.79	0.13	1.02
9	-7.87	-2.39	0.03	0.79
10	-7.72	-2.54	0.07	0.90

[a]$\Delta G_{score} = E_i - E_{lowest;}$ [b]RMSD = RMSD between docked and crystallographic podophyllotoxin structure; [c]RMSD = RMSD between docked poses corresponding to each configuration.

Whereas the RMSD value calculated out of 10 accepted poses for each configuration was found in between 0.59–1.33 Å. This revealed that the docked configurations have similar binding positions and orientations within the binding site and are similar to the crystal structure. The best docked structure, which is the configuration with the lowest Glide score is compared with the crystal structure and is shown in Figure 8.1. These docking results illustrate that the best-docked podophyllotoxin complex agrees well with its crystal structure and that Glide (XP)-docking protocol successfully reproduces the crystal tubulin-podophyllotoxin complex.

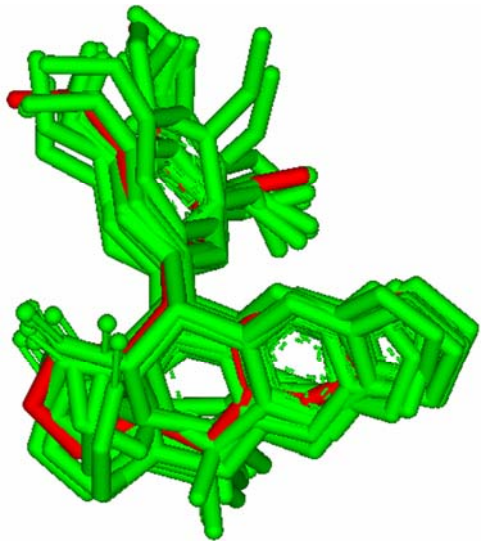


Figure 8.1. Superposition of all the docked configurations of podophyllotoxin on crystal structure (red-stick). RMSD (heavy atoms) = 0.02 to 0.85 Å.

We have applied the SGB-LIE method to a training set of 76 podophyllotoxin analogues to build a binding affinity model that was then used to compute the free energy of binding and predicted pIC_{50} for a test set of 36 analogues. Further the SGB-LIE model developed was validated using 12 new podophyllotoxin analogues for which the experimental tubulin polymerization inhibition was known. The training set for building the binding affinity model was comprised of four subsets of podophyllotoxin analogues as mentioned in Table 8.1(a-d). For all the four subsets included in the training set the experimental IC_{50} values against the cell lines P388 are available. With the wide range of difference between the IC_{50} values and the large diversity in the structures, the combined set of 76 ligands is ideal to be

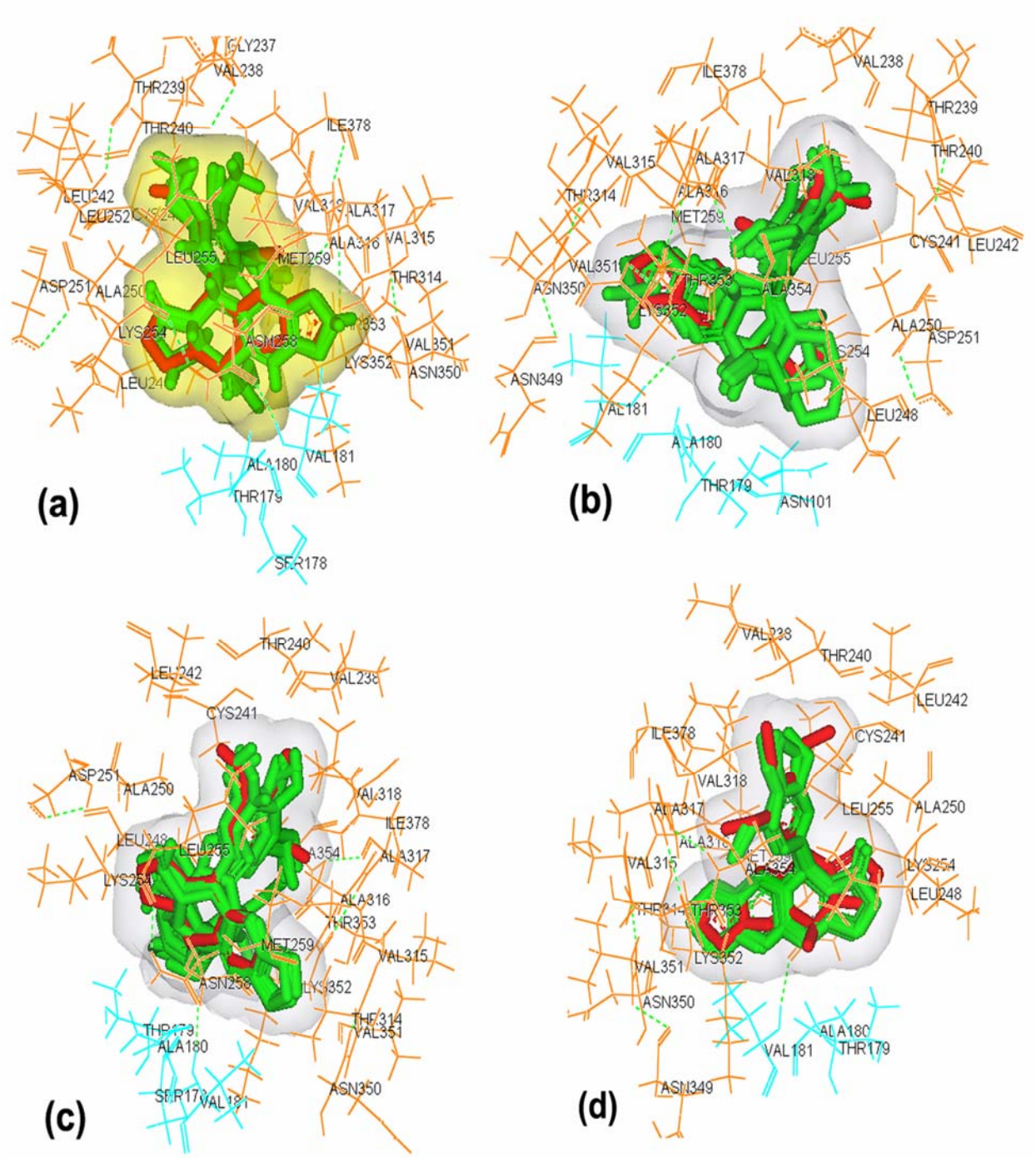


Figure 8.2(a-d). Superposition of podophyllotoxin analogues (5 analogues) belonging to (a) Tetraline lactones, (b) Non-lactonic tetralines, (c) Pyrazoline and isoxazoline derivatives and (d) Aza-podophyllotoxin derivatives within binding site of tubulin along with the co-crystal podophyllotoxin (red color).

considered as a training set, as the set does not suffer from bias, due to the similarity of the structures. Also, the training set containing 76 analogues contains enough data points not to suffer from over parameterization by the LIE model. Training set compounds were docked into the colchicine binding site of tubulin protein and the SGB-LIE calculations were performed using the Liaison module. The simulations were performed both for the ligand-free and ligand-bound state. The various interaction energy terms described in the methods were collected and are presented in Table 8.3 (a-d). The largest contribution for the binding energy comes from the van der Waals (vdw) interactions. This is obvious as the podophyllotoxin analogues used in the study are mostly lipophilic molecules that interact favorably with a binding cavity lined with hydrophobic residues. The hydrophobic center that is located in the middle of trimethoxyphenyl moiety of podophyllotoxin is surrounded by Leu β242, Ala β250, Leu β255, Ala β316, Val β318 and Ile β378 residues (Fig 8.2 a-e). The cavity energy term in the bound state is smaller (1.45 kcal/mol to 2.07 kcal/mol) than in the free state (3.18 kcal/mol to 7.62 kcal/mol) for all the compounds, as there is less energy penalty for creating a cavity in solvent when part of the ligand is buried into the hydrophobic binding site. The reaction field energy term in the free state lies in a very narrow range (-22.41 kcal/mol to -28.73 kcal/mol) for all compounds, but it varies in a wide range in the bound state (-6.37 kcal/mol to -31.33 kcal/mol) as the solvent accessible surface area varies with ligand structure in the bound form. The energy values in Table 8.3 (a-d) were used to fit equation 2 using the Gaussian elimination method. The values obtained for the three fitting parameters, α, β and γ are -0.141, -0.093 and -1.071, respectively. The large value of the cavity energy term signifies the fact that binding is largely driven by the ligand's ability to bury itself in the binding cavity, which is understandable given that most of the ligands are highly hydrophobic in nature. Even though the R value is low, vdw interactions contribute significantly toward the free energy of binding due to the large magnitude of the vdw interaction term. In Table 8.3 (a-d), the experimental free energy values obtained from the $RTIC_{50}$ and the free energy values estimated using SGB-LIE fitting parameters are presented. The root mean square error (RMSE) between the experimental values and the values obtained by the fit was 0.48 kcal/mol, which is an indicator of the robustness of the fit.

Table 8.3. (a). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the first Training subset inhibitors (Tetralinelactone podophyllotoxin analogues) using SGB-LIE method.

Ligand	$<U_{ele}>$[1] kcal/mol	$<U_{vdw}>$[1] kcal/mol	$<U_{cav}>$[1] kcal/mol	$pIC_{50,\ expt}$[2]	$\Delta G_{bind,expt}$[3] kcal/mol	$\Delta G_{bind,LIE}$[4] kcal/mol	$pIC_{50,\ pred}$[5]
1	11.7	-41.4	3.8	1.921	-2.6	-1.9	1.413
2	12.4	-42.0	4.3	2.002	-2.7	-2.5	1.810
3	12.0	-44.7	4.2	2.002	-2.7	-2.0	1.483
5	10.5	-52.5	3.8	0.198	-0.3	-0.7	0.542
7	10.5	-48.5	4.1	1.217	-1.7	-1.3	0.977
8	13.7	-43.6	3.8	1.298	-1.8	-2.0	1.471
9	9.0	-35.9	2.4	1.217	-1.7	-0.5	0.363
11	11.4	-47.4	3.8	0.220	-0.3	-1.3	0.960
13	10.4	-49.2	2.1	-0.359	0.5	0.8	-0.581
15	13.9	-57.4	3.8	0.697	-0.9	-0.7	0.509
16	10.4	-53.4	4.2	0.997	-1.4	-1.0	0.755
18	10.6	-63.8	2.3	-0.777	1.1	1.9	-1.391
19	12.9	-58.9	4.2	0.257	-0.3	-0.8	0.618
21	11.5	-53.7	3.5	0.917	-1.2	-0.4	0.269
23	11.0	-57.8	4.0	0.359	-0.5	-0.5	0.355
25	11.5	-54.2	3.7	0.917	-1.2	-0.6	0.449
26	11.4	-49.4	3.6	1.892	-2.6	-0.8	0.623
28	10.5	-56.2	2.9	-0.359	0.5	0.6	-0.479
29	7.1	-58.6	3.4	-0.359	0.5	0.8	-0.593

[1] $<U_{ele}>$, $<U_{vdw}>$ and $<U_{cav}>$ energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment. [2] pIC_{50} refers to the experimental predicted cytotoxic activity using P388 cell line and is calculated as $pIC_{50} = -\log IC_{50}$. [3] $\Delta G_{bind,expt}$ refers to free energy of binding for tubulin inhibition and is computed using the relationship: $\Delta G_{binding} \approx -2.303\ RTpIC_{50,expt}$, where 298 K is used in the work for temperature T. [4] $\Delta G_{bind,LIE}$ refer to the absolute free energy values obtained using SGB-LIE method. [5] $pIC_{50,\ pred}$ refers to predicted cytotoxic activity of ligands and is estimated using the relationship: $pIC_{50pred} = -(\Delta G_{bind,LIE} / 2.303\ RT)$.

Table 8.3. (b). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the second Training subset inhibitors (Nonlactonic tetralines podophyllotoxin analogues) using SGB-LIE method.

Ligand	$<U_{ele}>^1$ kcal/mol	$<U_{vdw}>^1$ kcal/mol	$<U_{cav}>^1$ kcal/mol	$pIC_{50,\ expt}$ [2]	$\Delta G_{bind,expt}$ [3] kcal/mol	$\Delta G_{bind,LIE}$ [4] kcal/mol	$pIC_{50,\ pred}$ [5]
30	8.3	-44.4	2.8	-0.161	0.2	-0.0	0.022
32	6.8	-54.0	2.1	-1.151	1.6	1.8	-1.335
33	9.0	-61.8	3.1	-0.836	1.1	1.2	-0.854
34	7.9	-52.6	2.2	-0.953	1.3	1.4	-1.040
36	11.8	-57.2	2.8	-0.616	0.8	0.7	-0.482
37	14.3	-46.7	4.4	1.012	-1.4	-2.4	1.744
38	11.8	-42.4	3.7	0.719	-1.0	-1.8	1.288
40	10.3	-51.0	1.7	-1.012	1.4	1.4	-1.053
41	11.4	-44.8	4.3	0.924	-1.3	-2.1	1.548
42	6.6	-51.4	2.2	-0.968	1.3	1.5	-1.097
44	9.2	-48.0	1.8	-0.851	1.2	1.2	-0.867
45	7.6	-48.5	1.8	-0.990	1.3	1.5	-1.098
46	8.0	-57.0	2.2	-1.181	1.6	1.8	-1.340
48	9.2	-56.8	1.7	-1.364	1.9	2.1	-1.545
49	11.8	-50.7	3.3	0.015	-0.0	-0.5	0.356
51	12.5	-46.8	4.0	0.653	-0.9	-1.7	1.221
52	8.3	-47.0	2.8	-0.257	0.3	0.1	-0.100
54	11.0	-55.5	2.8	-0.557	0.8	0.6	-0.426
55	11.4	-59.4	2.1	-1.181	1.6	1.7	-1.242
56	12.3	-54.1	2.2	-0.763	1.0	0.9	-0.663
57	14.0	-51.9	4.7	1.489	-2.0	-2.2	1.632
59	14.5	-52.1	4.4	0.763	-1.0	-1.9	1.428
60	13.3	-46.4	3.3	0.323	-0.4	-1.1	0.808
62	11.1	-59.2	2.8	-0.733	1.0	0.9	-0.656
64	12.1	-54.2	2.7	-0.477	0.6	0.4	-0.290
65	14.8	-51.8	4.8	1.034	-1.4	-2.4	1.789
66	13.2	-53.7	4.2	0.462	-0.6	-1.3	0.989
68	14.8	-56.2	4.3	0.528	-0.7	-1.5	1.122
69	12.7	-59.5	2.6	-0.777	1.1	0.9	-0.669
70	6.9	-55.4	2.4	-1.012	1.4	1.6	-1.143

Table 8.3. (c). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the third Training subset inhibitors (Pyrazolignans and Isoxazolignans podophyllotoxin analogues) using SGB-LIE method.

Ligand	$<U_{ele}>^1$ kcal/mol	$<U_{vdw}>^1$ kcal/mol	$<U_{cav}>^1$ kcal/mol	$pIC_{50,\ expt}{}^2$	$\Delta G_{bind,expt}{}^3$ kcal/mol	$\Delta G_{bind,LIE}{}^4$ kcal/mol	$pIC_{50,\ pred}{}^5$
71	11.8	-47.1	2.4	-0.565	0.8	0.2	-0.117
73	5.6	-50.8	2.7	-0.726	1.0	1.0	-0.725
74	9.7	-53.0	3.2	-0.660	0.9	0.1	-0.095
75	9.5	-55.3	3.4	-0.726	1.0	0.1	-0.089
77	11.2	-65.4	3.4	-0.909	1.2	0.8	-0.592
78	5.8	-50.6	2.1	-1.012	1.4	1.6	-1.198
80	4.4	-44.5	2.7	-0.623	0.8	0.6	-0.477
82	9.6	-63.2	2.8	-0.924	1.3	1.5	-1.112
84	10.8	-47.3	2.3	-0.653	0.9	0.4	-0.328

Table 8.3. (d). Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the fourth Training subset inhibitors (Aza-podophyllotoxin analogues) using SGB-LIE method.

Ligand	$<U_{ele}>^1$ kcal/mol	$<U_{vdw}>^1$ kcal/mol	$<U_{cav}>^1$ kcal/mol	$pIC_{50,\ expt}{}^2$	$\Delta G_{bind,expt}{}^3$ kcal/mol	$\Delta G_{bind,LIE}{}^4$ kcal/mol	$pIC_{50,\ pred}{}^5$
85	6.1	-60.6	3.0	-1.951	2.7	1.6	-1.150
87	6.7	-54.6	2.3	-2.017	2.7	1.6	-1.202
89	4.9	-48.2	3.3	-0.521	0.7	0.2	-0.164
91	5.1	-53.8	2.5	-1.936	2.6	1.6	-1.174
92	8.9	-54.0	2.2	-1.826	2.5	1.4	-1.023
94	7.7	-48.6	2.4	-1.239	1.7	0.9	-0.632
95	6.3	-50.3	2.8	-1.085	1.5	0.7	-0.542
97	4.4	-66.5	3.6	-2.061	2.8	1.7	-1.245
99	12.9	-41.8	4.3	2.420	-3.3	-2.6	1.895
100	11.9	-36.8	5.0	2.706	-3.7	-3.6	2.629
102	5.2	-52.3	4.5	0.345	-0.5	-0.7	0.486
103	1.8	-53.9	4.9	0.271	-0.4	-0.5	0.368
105	9.8	-45.9	4.6	1.635	-2.2	-2.0	1.494
107	8.2	-57.1	5.1	0.939	-1.3	-1.3	0.987
108	8.6	-37.0	4.6	2.405	-3.3	-2.7	2.008
110	9.4	-42.4	4.9	2.303	-3.1	-2.7	1.961
111	8.0	-48.9	5.6	2.185	-3.0	-2.5	1.870
112	8.5	-46.5	4.9	1.782	-2.4	-2.1	1.577

The quality of the fit can also be judged by the value of the squared correlation coefficient (r^2), which was 0.871 for the training set. Figure 8.3 graphically shows the quality of fit.

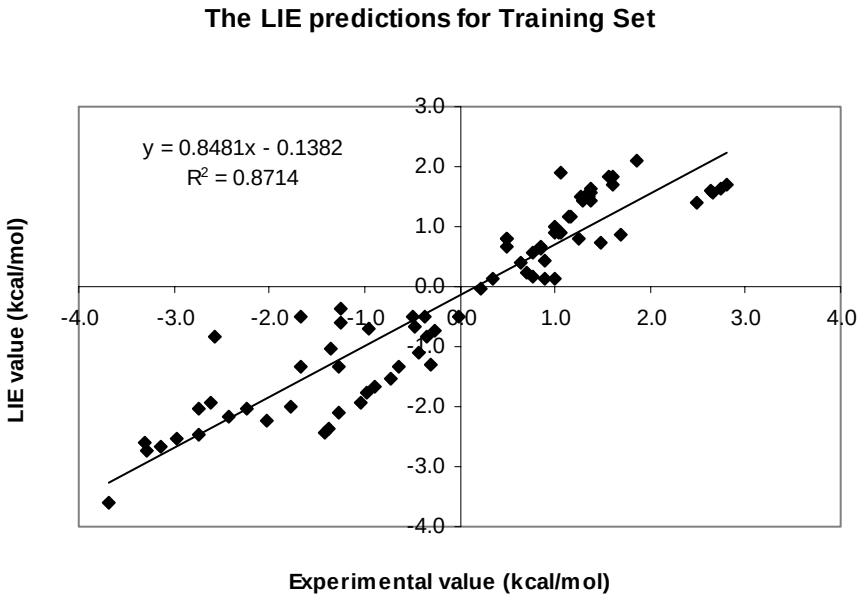


Figure 8.3. Free energy values estimated by the SGB-LIE method for 76 podophyllotoxin analogues comprising the training set plotted against corresponding experimental data. The RMS error is 0.481 kcal/mol between the two data sets for 76 ligands studied here.

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

$$\Delta G_{bind} = (-0.141)\langle U_{ele}\rangle + (-0.093)\langle U_{vdw}\rangle + (-1.07)\langle U_{cav}\rangle \quad (3)$$

(n = 76, r^2 = 0.871, r^2_{pred} = 0.864, s = 0.598, F = 166.8, P = 0.0001, q^2 = 0.865, PRESS = 28.05)

The SGB-LIE model developed in this study is statistically (q^2 = 0.865, r^2 = 0.871, F = 166.77) best fitted and consequently used for prediction of cytotoxic activities (pIC_{50}) of training and test sets of molecules as reported in Table 8.3 (a-d) & 8.4. The predicted activity calculated from free energy of binding is satisfactory with small deviation compared with experimental activity of training and test sets of molecules. The calculated free energy of binding (FEB) represents the experimental activity well.

Table 8.4. Average electrostatic (ele), van der Waals (vdw) and cavity (cav) energy terms as well as binding affinity model calculations for the Test set using SGB-LIE method.

Ligand	$\langle U_{ele} \rangle$[1] kcal/mol	$\langle U_{vdw} \rangle$[1] kcal/mol	$\langle U_{cav} \rangle$[1] kcal/mol	$pIC_{50,\ expt}$[2]	$\Delta G_{bind,expt}$[3] kcal/mol	$\Delta G_{bind,LIE}$[4] kcal/mol	$pIC_{50,\ pred}$[5]
4	11.0	-49.5	3.8	1.569	-2.1	-1.1	0.797
6	9.2	-44.9	3.3	1.217	-1.7	-0.7	0.512
10	10.9	-53.3	3.7	0.220	-0.3	-0.6	0.454
12	13.7	-53.7	2.4	-0.257	0.3	0.5	-0.366
14	11.6	-49.2	2.3	-0.323	0.4	0.4	-0.315
17	11.9	-46.7	3.0	0.638	-0.9	-0.6	0.447
20	10.6	-47.2	2.5	-0.007	0.0	0.2	-0.169
22	10.2	-52.2	3.5	0.961	-1.3	-0.3	0.238
24	13.8	-50.9	4.0	0.293	-0.4	-1.5	1.097
27	12.4	-62.7	3.3	-1.078	1.5	0.6	-0.408
31	10.3	-54.4	2.0	-1.049	1.4	1.5	-1.095
35	9.4	-54.1	2.2	-0.961	1.3	1.4	-1.009
39	10.3	-61.7	3.1	-0.763	1.0	1.0	-0.717
43	13.8	-53.0	3.4	0.066	-0.1	-0.6	0.477
47	10.0	-53.3	2.1	-0.939	1.3	1.3	-0.966
50	10.3	-46.9	1.5	-0.939	1.3	1.3	-0.948
53	12.4	-50.4	2.3	-0.535	0.7	0.5	-0.363
58	14.3	-49.0	3.9	0.587	-0.8	-1.6	1.187
61	13.5	-58.8	4.2	0.227	-0.3	-0.9	0.689
63	13.5	-58.5	3.8	0.037	-0.0	-0.6	0.428
67	12.4	-56.5	3.3	-0.257	0.3	-0.0	0.013
72	1.8	-42.8	2.5	-0.623	0.8	1.1	-0.790
76	10.6	-63.8	1.9	-0.983	1.3	2.4	-1.777
79	7.8	-49.3	2.7	-0.689	0.9	0.6	-0.446
81	9.1	-60.4	2.0	-1.071	1.5	2.1	-1.545
83	10.4	-44.8	2.4	-0.601	0.8	0.1	-0.073
86	5.2	-63.6	3.0	-2.288	3.1	1.9	-1.405
88	7.4	-59.0	2.8	-1.819	2.5	1.4	-1.030
90	5.9	-56.2	3.3	-1.232	1.7	0.9	-0.640
93	2.8	-52.7	2.8	-1.731	2.4	1.5	-1.073
96	3.6	-49.5	2.2	-1.987	2.7	1.7	-1.251
98	2.7	-47.8	2.3	-1.870	2.5	1.6	-1.187
101	2.0	-61.1	4.6	-0.733	1.0	0.5	-0.337
104	2.1	-46.0	3.5	-0.403	0.5	0.2	-0.137
106	7.7	-52.1	2.7	-0.051	0.1	0.9	-0.630
109	6.3	-52.0	5.1	1.159	-1.6	-1.5	1.099

[1] $\langle U_{ele} \rangle$, $\langle U_{vdw} \rangle$ and $\langle U_{cav} \rangle$ energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment. [2] pIC_{50} refers to the experimental predicted cytotoxic activity using P388 cell line and is calculated as $pIC_{50} = -\log IC_{50}$. [3] $\Delta G_{bind,expt}$ refers to free energy of binding for tubulin inhibition and is computed using the relationship: $\Delta G_{binding} \approx -2.303\ RTpIC_{50,expt}$, where 298 K is used in the work for temperature T. [4] $\Delta G_{bind,LIE}$ refer to the absolute free energy values obtained using SGB-LIE method. [5] $pIC_{50,\ pred}$ refers to predicted cytotoxic activity of ligands and is estimated using the relationship: $pIC_{50pred} = -(\Delta G_{bind,LIE} / 2.303\ RT)$.

Theoretically, FEB can be partitioned into several components: vdw, electrostatic and solvent accessible surface area (SASA) (Åqvist et al., 1994). In this study the SASA energy term has been replaced by the cavity energy term as proposed by Zhou et al., 2001.

Satisfied with the robustness of the binding affinity model developed using the training set, we applied the LIE model to the podophyllotoxin analogues comprising the test set. The test set includes 36 compounds categorized into four subgroups as mentioned above in Table 8.1 (a-d). The analogues comprising the test set were also obtained from different sources (Hitotsuyanagi et al., 2006; Marty et al., 2001). Since the experimental values of IC_{50} for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the SGB-LIE method for new ligands. Table 8.4 presents the free energy values estimated for the 36 test compounds for which experimental IC_{50} values were available to enable the accuracy check. The free energy values were estimated based on optimized SGB-LIE parameters α, β and γ from the training set. The overall RMSE between the experimental and predicted free energy of binding values was 0.56 kcal/mol which is comparable to the level of accuracy achieved by the most accurate methods such as free energy perturbation. The squared correlation coefficient between experimental and SGB-LIE estimates for the free energy for the test set compounds is also significant ($R^2 = 0.733$). The estimated free energy values for the test set ligands are plotted against the experimental data in Figure 8.4. There is a close match between the experimental and LIE free energy values of the ligands in the test set. The predicted cytotoxic activity estimated based on LIE free energy is also very close to experimental cytotoxic activity for the test set (Table 8.4).

To evaluate the accuracy of the SGB-LIE estimation for tubulin polymerization inhibition potencies, we have taken a separate data set called as validation set consisting of 12 analogues of podophyllotoxin (Table 8.4). Colchicine and its two structural derivatives were also taken in the validation set (Table 8.4) in view of that these compounds also binds to tubulin in the same binding site. Their experimental activity and chemical structures were obtained from literature (Haar et al., 1996; Lokie et al., 1978).

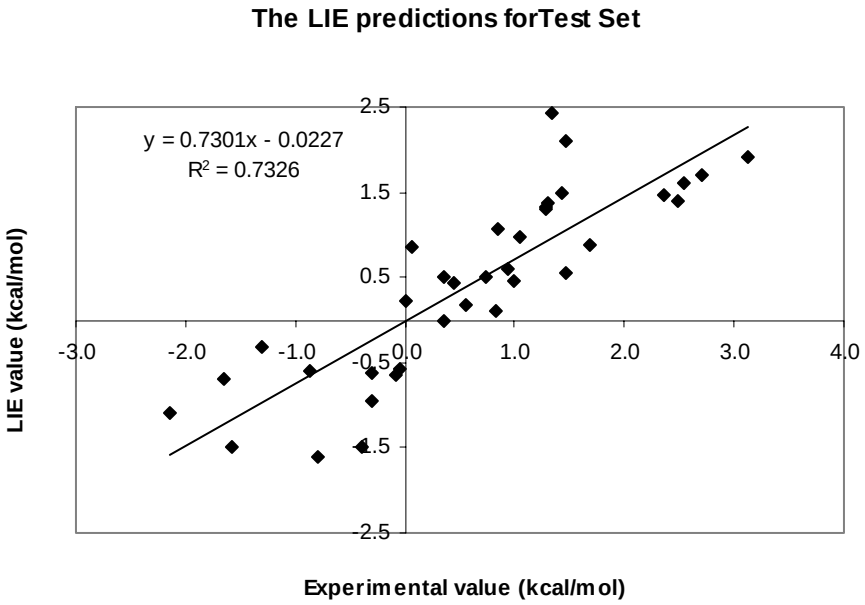
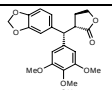
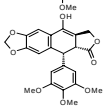
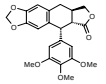
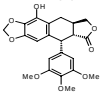
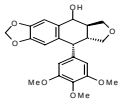
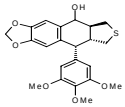
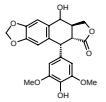
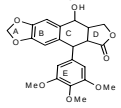
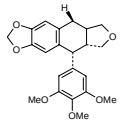
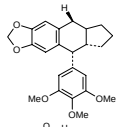
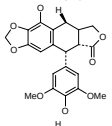
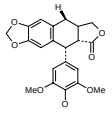
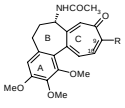
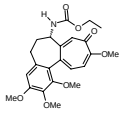


Figure 8.4. Free energy values estimated by the SGB-LIE method for 36 podophyllotoxin analogues comprising the Test set plotted against corresponding experimental data. The RMS error is 0.561 kcal/mol between the two data sets for 76 ligands studied here.

The experimental activity (IC_{50} value) of these compounds obtained from in vitro study of tubulin polymerization inhibition (TPI). For all the compounds excluding colchicine and its two derivatives, SGB-LIE predictions produce exactly the same trend for tubulin polymerization inhibition, even though the exact magnitudes of these values do not match very well to experimental values (Table 8.5). Podophyllotoxin competitively inhibit the binding of colchicine to tubulin (Hastie et al., 1991), implying that it bind to tubulin at the same site. The structural feature of podophyllotoxin that share with colchicine is the trimethoxyphenyl moiety. For colchicine and podophyllotoxin, it has been suggested that the binding sites for the two drugs do not completely overlap, with the trimethoxyphenyl rings of the agents binding in the same site on the tubulin heterodimer (Andreu et al., 1982; Andreu et al., 1982). Harr et al., (1996) suggested that the trimethoxyphenyl rings of the two drugs were situated in different regions of space, nearly orthogonal to each other. This revealed that these rings may bind to different regions of tubulin at the colchicine binding

site. The RMSE between the experimental and predicted binding free energy was 1.32 kcal/mol.

Table 8.5. The validation set along with their experimental activity expressed as the IC_{50} value for tubulin polymerization inhibition (TPI).

Ligand	Name	Structure	Experimental activity
1	G4		>50 M
2	Dehydropodophyllotoxin		>25 μM
3	Deoxypodophyllotoxin		0.5 μM
4	β-Peltatin		0.7 μM
5	Anhydropodophyllol		1.0 μM
6	Podophyllotoxin cyclic sulfide		10 μM
7	4'-Demethylpodophyllotoxin		0.5 μM
8	Podophyllotoxin		0.6 μM
9	Deoxypodophyllotoxin cyclic ether		0.8 μM
10	Deoxypodophyllotoxin cyclopentane		5.0 μM
11	α-Peltatine		0.5 μM
12	4'-Demethyldeoxypodophyllotoxin		0.2 μM
13	Colchicine		2.4 μM
14	3-(Ethoxycarbonyl)-3-demethylthiocolchicine		1.4 μM

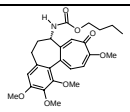
15	3-(Butoxycarbonyl)-3-demethylthiocolchicine		1.4 μM

Table 8.6. LIE fitting, free energy values (ΔG_{bind}, kcal/mol) and predicted potencies (pIC_{50}), obtained from the SGB-LIE method and experimental data for the validation set.

Ligand	$<U_{ele}>$[1] kcal/mol	$<U_{vdw}>$[1] kcal/mol	$<U_{cav}>$[1] kcal/mol	$TPI_{, expt}$[2] (pIC_{50} value)	$\Delta G_{bind,expt}$[3] kcal/mol	$\Delta G_{bind,LIE}$[4] kcal/mol	$TPI_{, pred}$[5] (pIC_{50} value)
1	13.4	-46.3	4.3	-7.699	10.5	-2.2	1.588
2	14.0	-56.3	4.1	-1.398	1.9	1.2	-0.905
3	11.8	-60.1	3.3	0.301	-0.4	-0.1	0.072
4	12.1	-61.2	3.6	0.155	-0.2	0.1	-0.078
5	10.8	-58.5	3.5	0.000	0.0	0.2	-0.117
6	8.3	-47.3	3.5	-1.000	1.4	0.9	-0.638
7	11.8	-63.7	3.5	0.301	-0.4	-0.0	0.007
8	11.6	-56.2	3.2	0.222	-0.3	-0.3	0.241
9	10.8	-57.4	4.0	0.097	-0.1	-0.4	0.292
10	10.7	-55.7	2.5	-0.699	0.9	1.0	-0.711
11	11.5	-60.7	4.0	0.301	-0.4	-0.2	0.185
12	14.1	-63.4	4.2	0.699	-0.9	-0.6	0.429
13	13.6	-53.4	6.0	-0.380	0.5	-3.4	2.474
14	15.1	-67.1	6.0	-0.146	0.2	-2.4	1.751
15	8.1	-56.3	6.1	-0.146	0.2	-2.5	1.809

[1] $<U_{ele}>$, $<U_{vdw}>$ and $<U_{cav}>$ energy terms represents the ensemble average of the energy terms calculated as the difference between bound and free state of ligands and its environment. [2] pIC_{50} refers to the experimental predicted IC_{50} value for TPI and is calculated as $pIC_{50} = -\log IC_{50}$. [3] $\Delta G_{bind,expt}$ refers to binding free energy for tubulin-analogue interaction and is computed using the relationship: $\Delta G_{binding} \approx -2.303\ RTpIC_{50,expt}$, where 298 K is used in the work for temperature T. [4] $\Delta G_{bind,LIE}$ refer to the absolute binding free energy values obtained using SGB-LIE method. [5] $pIC_{50,\ pred}$ refers to predicted IC_{50} value for TPI based on SGB-LIE method and is estimated using the relationship: $pIC_{50pred} = -(\Delta G_{bind,LIE} / 2.303\ RT)$.

For compound G4 the RMSE is more than 1.29 kcal/mol. Excluding G4 from the data set the RMSE for the rest of the 11 compounds is 0.29 kcal/mol, which means that the SGB-LIE modeling was able to predict the binding free energy of the 11 compounds within 0.29

kcal/mol, which is comparable to the level of accuracy achieved by the most accurate methods, such as free energy perturbation.

8.4 Conclusion

We have demonstrated that the SGB-LIE method can be applied to estimate the binding free energy with a high level of accuracy for a diverse set of podophyllotoxin analogues with tubulin. The magnitude of free energy changes upon binding of these analogues to tubulin have directly correlated with the experimental potency of these inhibitors. Despite the limitation imposed by the insufficient sampling inherent in the energy minimization protocol, the method has reproduced experimental data with reasonably small error for the majority of podophyllotoxin analogues. Using LIE methodology, we have been able to verify the experimental observation that derivatized podophyllotoxin compounds, with their C-ring removed (as in G4) or was unsaturated (as in dehydropodophyllotoxin), have inhibition potencies reduced. When the C-ring's substituent was removed as in deoxypodophyllotoxin or substituted to B-ring as in β-peltatin, the resulting analogues were still a potent inhibitor. This indicated that the three-dimensional conformation of the C-ring and the resulting conformational influence on the D-ring is important for interaction with tubulin. This concurs with the finding that sterioisomers like epipodophyllotoxin are much less potent. The decreased potency of lactone D-ring analogues was also usually predicted by SGB-LIE model. Few analogues with modifications on the E-ring have been tested in vitro for TPI. Removal of the 4'-methyl to give the phenol results in a small increase in potency. An increase in potency is also seen when the C-ring hydroxyl is moved to ring B: α-peltatin is slightly more potent than β-peltatin. The influences of these structural modifications were correctly predicted by SGB-LIE model developed in the study. However, the SGB-LIE predictions could not produce exactly the same trend of tubulin polymerization inhibition for the colchicine and two of its structural derivatives. This is obvious as the mode of interaction of colchicine is different at the colchicine binding site of tubulin than that of podophyllotoxin. It was suggested that the trimethoxyphenyl rings of the two drugs were bind to different regions of tubulin at the colchicine binding site. Podophyllotoxin is well known for its antitumor activity. It has better tubulin polymerization inhibition in comparison to colchicine. However, the clinical application of it and its analogues in the treatment of cancer has been limited by severe toxic side effects during

administration of the drugs. With a view to achieving greater therapeutic efficiency many podophyllotoxin analogues have been isolated and via molecular manipulation, a large number of semisynthetic derivatives have been synthesized. However, new findings related to their activities, mechanism of action and pharmacological properties have been unexplored. The interaction of colchicine with tubulin is 'irreversible' and temperature-sensitive. Podophyllotoxin binds faster than colchicine and the binding is reversible and less temperature-sensitive which makes them more useful in the field of cancer therapy. The temporal and reversible binding of podophyllotoxin with tubulin overcomes the problem of inhibiting the cell multiplication of normal cell. Most of the toxic effects of the podophyllotoxin and its derivatives are due to their scant selectivity between cancer and normal cells.

Moreover, the SGB-LIE method is able to predict the binding free energy and cytotoxic activity of rationally designed podophyllotoxin congeners with relative success. The difference in the exact magnitudes of estimated vs. experimental free energy of binding for compounds in the training set, test set and validation set may be due to the limitations imposed by inadequate sampling and force field parametrization. In addition, the calculation of absolute binding free energy from experimental IC_{50} values for cytotoxicity obtained from the in vitro cell line is only an approximation. In practical the IC_{50} value of a drug molecule is dependent upon a number of factors including solubility, membrane permeability, p-glycoprotein activity against the compound, etc. However, the SGB-LIE model developed is able to predict the binding energy of the validation set quite accurately in comparison to the binding kinetics in vitro. This may be the fact that tubulin is the most potential target for podophyllotoxin. Further, the strong relationship between the experimental and predicted FEB could be established by in vitro studies of all these podophyllotoxin analogues with isolated tubulin. A detailed study on the SARs for podophyllotoxin analogues can throw light on the moieties and functional groups important in determining the inhibition potency. The close estimation of inhibition potencies of a wide range of structural derivatives for podophyllotoxin establishes the SGB-LIE methodology as an efficient tool for screening novel compounds with very different structures. The mechanism of action of any drug is very important in drug development. Generally, the drug

compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug–receptor interactions are required for high activity. Understanding the nature of these interactions is very significant and theoretical calculations, in particular the SGB-LIE method, seem to be a proper tool for gaining such understanding. The results obtain will give information on how the chemical structure of the drug should be modified to achieve suitable interactions and for the rapid prediction and virtual prescreening of anti-tumor activity. This will lead to new proposals regarding possible improvements to the therapeutic indices of podophyllotoxins. Compared to the empirical methods, such as scoring function approaches, the LIE method is more accurate due to the semiempirical approach adopted in which experimental data are used to build the binding affinity model. The SGB-LIE method seems promising when compared to the FEP or TI methods in achieving comparable accuracy with must faster speed even for structurally very different ligands.

CHAPTER 9

Quantitative structure-activity relationship (QSAR) of the podophyllotoxin: the development of predictive in vitro cytotoxic activity models.

Abstract

Podophyllotoxin is a unique lignan occurring as a constituent of *Podophyllum* species. Because of the effectiveness of podophyllotoxin in the treatment of various types of cancer, development of useful semisynthetic drugs from podophyllotoxin derivatives is prompt. However, recent reports of immunosupression and lacking of specificity to cancerous cells with podophyllotoxin derivatives have spawned a renewed effort to develop nontoxic analogues of podophyllotoxin. Continuous efforts to develop more potent, less toxic and better oral availability, lot of podophyllotoxin analogues have been synthesized which provides the pre-requisite for developing of QSAR model to screen the potent analogues and predict the cytotoxic activity. A QSAR model has been generated with a data set of 119 podophyllotoxin analogues with known cytotoxic activity. Several types of descriptors including topological, spatila, thermodynamics, information content, leadlikeness and E-state indices have been used to derive a quantitative relationship between cytotoxic activity and structural properties. The robustness of the QSAR models was characterized by the values of the internal leave one out cross-validation r^2 (q^2) for the training set and external predictive r^2 for the test set. Statistically significant model (r^2 = 0.909; q^2 = 0.888) was obtained with the descriptors like shadow areas, dipole moment, strain energy, E-state indices, superdelocalizibility and partial positive surface area. QSAR model developed in this study shall aid further design of novel potent podophyllotoxin derivatives.

9.1 Introduction

The aryltetralin lactone (-)-podophyllotoxin occupies a unique position among lignan natural products since its glucopyranoside derivative was recognized as a potent anti-tumor factor (Jardine, 1980). This discovery entails a particularly fascinating account, involving a multitude of investigations conducted over a period of more than a century (Stahelin and von Wartburg, 1991). Among the plethora of physiological activities and potential medicinal and agricultural applications, the anti-neoplastic and antiviral properties of podophyllotoxin congeners and their derivatives are arguably the most eminent from a pharmacological perspective. Nevertheless, podophyllotoxin including the most successful chemotherapy drugs have undesirable side effects that limit their utility. Their drawback is that when these drugs are given systemically, they bind tubulin indiscriminately, leading to the destruction of both cancerous and healthy cells, the consequence of which is the presence of serious side effects in all known cancer chemotherapy applications. Thus, the studies aim at improving anti-neoplastic agents, mainly focused on the search for more selective drugs.

After the discovery of the therapeutic properties of podophyllotoxin, new findings related to its activities, its mechanism of action and pharmacological properties have been unveiled. Structure-Activity Relationships (SAR), have shown that podophyllotoxin analogues preferentially inhibit tubulin polymerization, which leads to the arrest of cell cycle in the metaphase (Snyder et al., 1976; Margolis et al., 1978). Different derivatives of podophyllotoxin have demonstrated to bind to the same site, as shown by the fact that podophyllotoxin has been reported to compete with colchicine for the binding site in tubulin (Cortese et al., 1977), its affinity is more than double to that of colchicine. Furthermore, colchicine binds to tubulin almost irreversibly; podophyllotoxin derivatives do so reversibly, which make them less toxic and hence more useful in the field of cancer therapy. Following binding of podophyllotoxin, the GTP hydrolyzing capacity of tubulin is inhibited, but colchicine stimulates an assembly-independent GTPase activity directed at the exchangeable site-bound GTP (Lin et al., 1981). Podophyllotoxin binds to β-tubulin at its interface with α-tubulin resulting in inhibition of tubulin polymerization. This binding mode was recently

confirmed by the determination of a 4.20 Å X-ray structure of αβ-tubulin complexed with podophyllotoxin (PDB_ID: 1SA1), showing that podophyllotoxin also binds at the colchicine site (Ravelli et al., 2004).

Prompted by the clinical successes of the podophyllotoxin, significant efforts have been focused on identifying new analogues that have a similar mechanism of action yet with superior properties such as low or nil toxic side effects and better oral availability. These SARs refered to etoposide, have been reviewed by Damayanthy and Lown (1996). Modification of the A ring (Cho et al., 1996, Bertounesque et al., 1996) gave compounds that showed significant activity but less than that of etoposide, whereas modification of the B ring (Thurston et al., 1986) resulted in a loss of activity. Modification of the C ring (King et al., 1946, Beers et al., 1988, Hitotsuyanagi et al., 1994, Hitotsuyanagi et al., 1995) by aromatization or expansion gave compounds less potent than podophyllotoxin. One of the modifications of the D ring (Wang et al., 1993, Kadow et al., 1989) produced GP-11, which is almost equipotent with etoposide (Wang et al., 1993). E ring oxygenation did not affect DNA cleavage. It has also been observed that the free rotation of the E ring is necessary for anti-tumor activity (Hitotsuyanagi et al., 1995, Stahelin et al., 1973, Berkowitz et al., 1996, Ayres et al., 1982, Visser et al., 1989, Tian et al., 1996). The C-7 substituted aglycones (Levy et al., 1983, Utsugi et al., 1996, Pagani et al., 1996) and the aza analogues (Vander Eycken et al., 1989, Tomoika et al., 1993, Pearce et al., 1990, Lionard et al., 1993) have a significant place in these recent developments. The substitution of a glycosidic moiety with aryl or alkyl amines produced enhanced activity, e.g.TOP-53 (Kitamura et al., 1997, Mross et al., 1996). The high selectivity of TOP-53 has been attributed to its distribution into the lung and its persistence. Modification in the sugar ring resulted in the development of the agent NK-611 which is currently undergoing clinical trials (Utsugi et al., 1996, Pagani et al., 1996, Daley et al., 1998). Also many podophyllotoxin analogues have been isolated and via molecular manipulation, a large number of semisynthetic derivatives have been obtained. The study and assessment of these have permitted the clinical development and their usage in the treatment of different types of cancer. A rational approach for the discovery of a pharmaceutically acceptable, economically viable, podophyllotoxin-based anticancer drug awaits development of a global mechanism of action model for organic cyclolignans and/or

a predictive quantitative structure-activity relationship (QSAR) model. With the advent of parallel synthesis methods and technology, we might expect the number of anticancer podophyllotoxin derivatives to be tested to achieve dramatic growth. One method of orchestrating these strategies is to make use of QSAR models for the rapid prediction and virtual prescreening of cytotoxic activity. Based on the activity data generated from the above wet lab studies we have developed a QSAR model for determining the cytotoxic activity of podophyllotoxin analogues.

Traditional QSAR studies have been used since early 1970s to predict activities of untested molecules. The pharmacophore based, quantum mechanics (QM) based and physicochemical based 3D-QSAR have been employed to build QSAR models for a wide range of applications and have shown good predictivity (Pasha et al., 2008). With a wide range of molecular structures and their complementary activities, it has been assumed that the most important criterion for a systematic study of 3D-QSAR has been satisfied. Although comparative molecular field analyses (CoMFA) are statistically excellent and offer good predictive performance, they are inherently limited to the need to align with the database molecules correctly within 3D space. The determination of the 'active' conformation that each compound will retain is a critical issue due to unavailability of X-ray structure. We should have some knowledge or hypothesis regarding active conformations of the molecules under study as a prerequisite for structural alignment. Hence, the developed models based on CoMFA may not suit to drug design, because of a false conformational hypothesis. However, we were motivated to explore possible alternatives that would use alignment free descriptors derived from 2D or 3D molecular topology and thus alleviate frequent ambiguity of structural alignment typical of 3D QSAR methods. Accordingly in this QSAR study, we have applied E-state, electronic, structural, topological quantum mechanics and physicochemical based descriptors, which can be calculated without structural alignments. Initial structure-activity findings have already been reported (Gordaliza et al., 2000; Harr et al., 1996), but a sufficient variety of analogues are now available to permit a more systematic analysis of the structural features of the podophyllotoxin molecule required for its interaction with purified tubulin.

Quantitative structure-activity relationship (QSAR) is one of the most important methods in chemometrics, which give information that is useful for drug design and medicinal chemistry (Marder et al., 2001, Tuppurainen et al., 1999). A QSAR equation is a mathematical equation that correlates the biological activity to a wide variety of physical or chemical parameters (Hansch et al., 2001, Livingstone et al., 2000). There are many examples available in literature in which QSAR models have been used successfully for the screening of compounds for biological activity (Shi et al., 1998, Oloff et al., 2005, Menese-Marcel et al., 2005, Santana et al., 2006). The model developed in the present study is the first of its kind for cytotoxic activity prediction of podophyllotoxin congeners and is of high statistical quality. The behavior of QSAR model is examined with a variety of statistical parameters and the contribution of various descriptors are analyzed. The methodology used in the present study is in line with that has been used by Deswal and Roy (2006) for the development of thrombin inhibitors.

9.2 Materials and methods

9.2.1 Data set

A total of 119 podophyllotoxin analogues were used in the study and were taken from various sources belonging to different ring modifications. For better interpretation all these compounds were divided into following five sublibraries. These molecules were divided into 81 molecules in training set and 38 molecules in test set.

*Sublib-**I*** commonly known as tetralinelactones consists of 29 compounds (1-29) (Table 9.1a). These molecules were rationally designed as functional mimics of natural podophyllotoxin with the goal of simplifying the chemical synthesis and improving the cytotoxic activity. Structural modifications are mainly introduced at varying radicals at position 7 in podophyllotoxin scaffold. Reports have been made on compounds with oxygenated substituents in the form of ethers, esters and diverse nitrogen radicals (San Feliciano et al., 1989; San Feliciano et al., 1993; Miguel et al., 1995; Doré et al., 1996; Gordaliza et al., 1996).

*Sublib-**II*** contains compounds (30-70) (Table 9.1b) known as non-lactonic tetralines. Structural modifications in this group include the opening of the lactone ring (D-ring) in podophyllotoxin scaffold, to give rise to compounds with different degrees of oxidation at positions C-9 and C-9' (San Feliciano et al., 1993; Miguel et al., 1995; Doré et al., 1996; Gordaliza et al., 1996). In general these molecules lack any lactone rings.

*Sublib-**III*** also includes a group of lignans (71-85) (Table 9.1c) that have heterocyclic rings fused to the cyclolignan skeleton. This group is commonly called as pyrazolignans (San Feliciano et al., 1993; Doré et al., 1996; Gordaliza et al., 1995) and isoxazolignans (Gordaliza et al., 1996; Gordaliza et al., 2001) and they were obtained by reacting podophyllotoxin with differently substituted hydrazines and hydroxylamines.

*Sublib-**IV*** includes the compounds (86-91) (Table 9.1d) commonly called lactonic and non-lactonic naphthalene. These molecules were obtained by structural modification of C and D-rings and have proportionally much lower activity (San Feliciano et al., 1993; Gordaliza et al., 1994; Doré et al., 1996).

*Sublib-**V*** contains compounds (92-119) (Table 9.1e) commonly known as aza-podophyllotoxin analogues. The preparation of this group of compounds requires selective chemical manipulation of the two aromatic rings (B and E-rings) of the podophyllotoxin scaffold. These molecules are readily prepared from anilines, benzaldehydes and tetronic acid or 2, 3-cyclopentanedione in good to excellent yield and have also shown better cytotoxic activity (Hitotsuyanagi et al., 2000).

To validate the 3D QSAR model for tubulin polymerization inhibition, a validation set containing 16 tubuline polymerization inhibitors (113-128) (Table 9.1f) of different activity classes were analyzed. The experimental potencies and chemical structures were obtained from literature (Haar et al., 1996; Lokie et al., 1978).

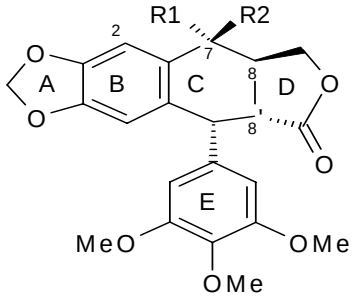
All these podophyllotoxin analogues were built from the scaffolds by different ring modification and substitution of functional groups as mentioned in Table 9.1(a-f). We used

ISIS Draw 2.3 software for sketching structures and converting them to their 3D representation by using ChemSketch 3D viewer of ACDLABS 8.0. LigPrep (Schrodinger L. L. C. 2007) was used for final preparation of ligands from libraries. LigPrep is a utility of Schrödinger software suit that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for tatutomers and steric isomers and performing a geometry minimization of ligands. The ligands were energy minimized using Macromodel module of Schrodinger with default parameters and applying molecular mechanics force fields (MMFFs). Truncated Newton Conjugate Gradient (TNCG) minimization method was used with 500 iterations and convergence threshold of 0.05kJ/mol. Each of these compounds had associated in vitro cytotoxicity values (IC_{50} values reported in μM) against cell line P388.

9.2.2 Descriptor calculation

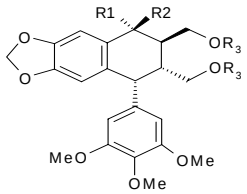
E-state indices (Gregorio et al., 1998), log P (Meylan and Howard, 1995), Superpendentic index (Gupta et al., 1999), structural (Liu et al., 1998), symmetrical, topological, lead likeness (Lipinski et al., 2001), electronic Wang-Ford atomic charge (Santana et al., 2006) and extended Huckel partial charge (Deswal et al., 1996; Eliopoulus et al., 1996, Brenwald et al., 1998) functions), bulk, moments, orbital energies, molecular connectivity indexes (6), gravitational indexes (7), hydrophobicity (Livingstone et al., 2000), steric (Shi et al., 1998, Oloff et al., 2005) and thermodynamic factors (Meneses-Marcel et al., 2005) and topological descriptors were calculated using ADME Model Builder software package (version 4.5). The Superpendentic index is computed from the pendent matrix. These descriptors help differentiate the molecules mostly according to their size, degree of branching, flexibility and overall shape. Some of the descriptors included in the study are listed and described in Table 9.2.

Table 9.1. (a). Podophyllotoxin derivatives (Tetraline lactones) with cytotoxic activities against P-388 cell line used in the work.

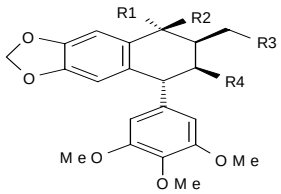


Analogue	R1	R2	IC_{50}(μM)	Analogue	R1	R2	IC_{50}(μM)
1	OH	H	0.012	15		=N-OMe	0.2
2	H	H	0.010	16	H	H	0.10
3	H	H(2-OMe)	0.01	17	H	H(2-OMe)	0.23
4	OH	H(4'-OH	0.027	18	OH	H	6.0
5	OAc	H	0.625	19	OAc	H	0.55
6	OMe	H	0.06	20	OAc	H(2-OMe)	1.02
7	H	OH	0.06	21	OMe	H	0.12
8	H	Ac	0.05	22	H	OH(2-OMe)	0.11
9	H	OMe	0.06	23	H	OAc	0.44
10	H	Cl	0.6	24	H	OAc(2-OMe)	0.51
11	Cl	H	0.6	25	H	OMe	0.12
12		=O	1.8	26	H	H Δ^7	0.013
13		= N-OH	2.3	27		=O	12.0
14		=N-OAc	2.1	28		=N-OH	2.3
				29		=N-OMe	2.3

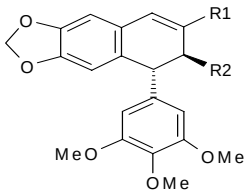
Table 9.1. (b). Podophyllotoxin derivatives (Nonlactonic tetralines) with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.



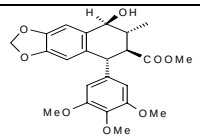
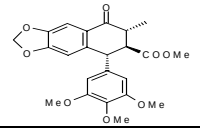
53 - 59.



62-81



82 - 92.

Analogue	R1	R2	R3	IC_{50}(μM)	Analogue	Structure	IC_{50}(μM)
30	OH	H	H	1.2	35		23.3
31	H	OH	H	12.0			
32	H	OMe	H	11.6			
33	H	OMe	Ac	9.7			
34	OMe	H	Ac	9.7			
					36		3.5

Analogue	R1	R2	R3	R4	IC_{50}(μM)	Analogue	R1	R2	R3	R4	IC_{50}(μM)
37	H	H	OH	COOMe	0.058	47	H	OMe	OAc	CH_2OAc	9.7
38	H	H	OAc	COOMe	0.21	48	H	OH	OH	CH_2OH	47.9
39	H	H	OAc	CH_2OAc	5.14	49	H	OH	OH	COOMe	1.1
40	OH	H	OH	CH_2OH	23.9	50	=O		OH	COOMe	5.63
41	OH	H	OH	COOMe	0.22	51	=O		OAc	COOMe	0.20
42	OAc	H	OAc	CH_2OAc	7.4	52	=N-OH		OAc	COOMe	2.0
43	OAc	H	OAc	COOMe	1.1	53	H	H	CHO	COOMe	2.34
44	OMe	H	OH	CH_2OH	23.2	54	H	H	=N-OMe	COOMe	2.30
45	OMe	H	OAc	CH_2OAc	19.4	55	H	H	=N-OMe	COOMe	10.94
46	H	OMe	OH	CH_2OH	11.6	56	H	H	=N-allyl	COOMe	2.5

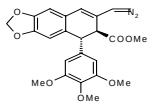
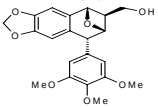
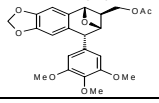
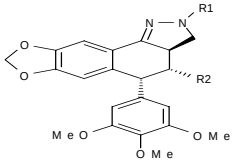
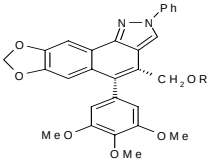
Analogue	R1	R2	IC_{50}(μM)	Analogue	R1	R2	IC_{50}(μM)
57	CH_2OH	COOMe	0.02	64	CH=N-OH	COOMe	2.27
58	CHO	CH_2OH	0.25	65	CH=N-OMe	COOMe	0.22
59	CHO	COOMe	0.23	66		COOMe	0.20
60	CH=N-NH2	COOMe	0.57	67		CH_2OH	1.00
61	CH=N-NH-CH_2CF_3	COOMe	0.48	68			0.57
62	CH=N-NH-Ph	COOMe	1.94	69			6.25
63	CH=N-NH-Ph	CH_2OH	1.02	70			5.66

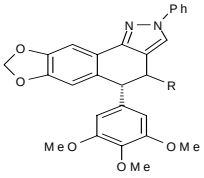
Table 9.1 (c). Podophyllotoxin derivatives (Pyrazolignans and isoxazolignan) with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.



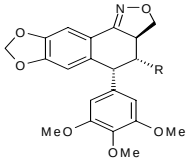
98 - 109.



110-111



112-114

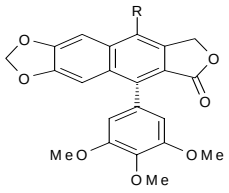


115-120

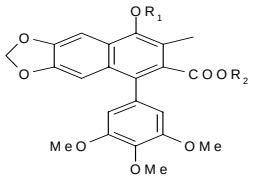
Analogue	R1	R2	IC_{50}(μM)	Analogue	R1	R2	IC_{50}(μM)
71	Ph	COOH	1.9	75	m-NO_2Ph	COOMe	4.5
72	Ph	COOMe	1.00	76	p-MePh	COOMe	1.00
73	Ph	CH_2OH	4.1	77	Me	COOMe	5.6
74	Ph	CH_2OAc	4.7				

Analogue	R	IC_{50}(μM)	Analogue	R	IC_{50}(μM)
78	H	10	82	COOMe	23
79	CHO	21	83	COOMe(4'-OH)	12
80	CH_2Ac	2.2	84	CH_2OH	2.6
81	COOH	2.2	85	CH_2O	2.4

Table 9.1. (d). Podophyllotoxin derivatives (lactones and non-lactonic naphthalene) with cytotoxic activities against P-388 cell line used in the work.



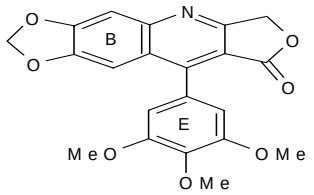
121-122



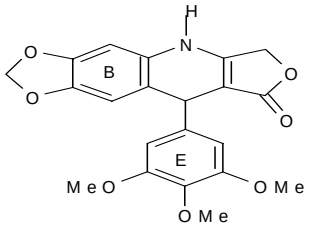
123-126

Analogue	R		IC_{50}(μM)	Analogue	R1	R2	IC_{50} (μM)
86	H		5.1	89	Ac	H	5.90
87	OAc		44.25	90	Ac	Me	16.59
88	H	Me	12.20	91	H	OMe	2.15

Table 9.1. (e). Aza-podophyllotoxin derivatives with cytotoxic activities against P-388 cell line used in the work.



Modification 1

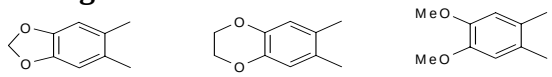


Modification 2

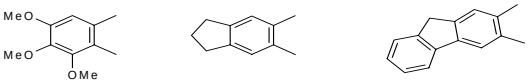
Substitution of B & E ring at 1 and 2 analogues:

Modification 1				Modification 2			
Analogue	B Ring	E Ring	IC_{50}(μM)	Analog	B Ring	E Ring	IC_{50}(μM)
92	I	VII	100	106	I	VII	0.0018
93	II	VII	80	107	II	VII	0.0017
94	III	VII	100	108	III	VII	4.9
95	III	VIII	39	109	III	VIII	0.76
96	III	XII	2.0	110	III	XII	0.77
97	IV	VII	29	111	IV	VII	2.6
98	V	VII	100	112	V	VII	0.0041
99	VI	VII	63	113	VI	VII	0.92
100	I	VIII	40	114	I	VIII	0.048
101	I	IX	100	115	I	IX	0.0053
102	I	X	100	116	I	X	0.13
103	I	XI	60	117	I	XI	0.0053
104	I	XII	100	118	I	XII	0.030
105	I	VII	71	119	I	VII	0.028

B Ring

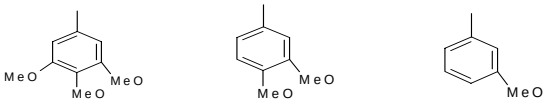


I II III



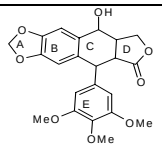
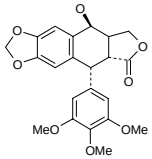
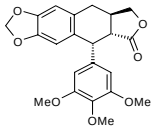
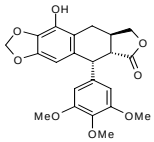
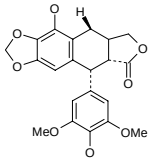
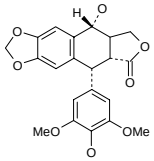
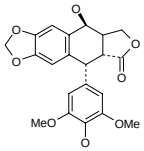
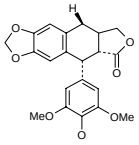
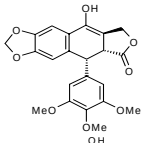
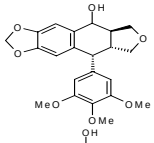
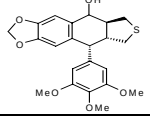
IV V VI

E Ring



VII VIII IX X XI XII

Table 9.1. (f). The experimental IC_{50} value for in vitro tubulin polymerization inhibition by podophyllotoxin analogues (validation set).

Analogue	Name	Structure	IC_{50} (μM)
1	Podophyllotoxin		0.6
2	Epipodophyllotoxin		5.0
3	Deoxypodophyllotoxin		0.5
4	β-Peltatin		0.7
5	α-Peltatin		0.5
6	4'-Demethylpodophyllotoxin		0.5
7	4'-Demethylepipodophyllotoxin		2.0
8	4'-Demethyldeoxypodophyllotoxin		0.2
9	Dehydropodophyllotoxin		25
10	Anhydropodophyllol		1.0
11	Podophyllotoxin cyclic sulfide		10

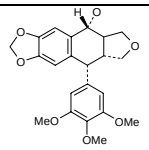
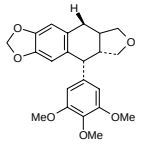
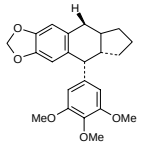
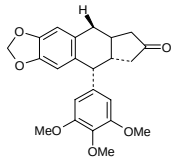
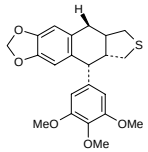
12	Podophyllotoxin-cyclic ether		1.0
13	Deoxypodophyllotoxin-cyclic ether		0.8
14	Deoxypodophyllotoxin-cyclopentane		5.0
15	Deoxypodophyllotoxin-cyclopentanone		5.0
16	Deoxypodophyllotoxin-cyclic sulfide		10

Table 9.2. List of descriptors used in the study.

Type	Descriptors
E-state indices	Electro-topological-state indices
Electronic	Partial positive surface area (AM1), partial negative surface area (AM1), relative positive charge (AM1), relative negative charge (AM1), relative positive charged surface area (AM1), relative negative charged surface area (AM1), weighted positive charged partial SA (AM1), weighted negative charged partial SA (AM1), fractional negative charged partial SA (AM1), fractional positive charged partial SA (AM1), Huckel molecular orbital indices: minimum autopolarizability value, maximum autopolarizability value, minimum bond order, maximum bond order, minimum electron density value, maximum electron density value, highest occupied molecular orbital, lowest unoccupied molecular orbital, maximum free valence value, minimum nucleophilic superdelocalizability, maximum nucleophilic superdelocalizability, min. free radical superdelocalizability, max. free radical superdelocalizability, heat of formation, dipole moments, energy of the highest occupied orbital, energy of the lowest unoccupied orbital, electronegativity, hardness, mean partial charge on H atoms, most negative partial charge on H atom, most positive partial charge on H atom, most negative partial charge on C atom, mean partial charge on C atoms, most positive partial charge on O atom, mean partial charge on O atoms, most negative partial charge on heteroatom, mean partial charge on heteroatoms, most positive partial on heteroatom.
Information content	Information of atomic composition index, superpendentivity index, superpendentivity index Carbon only.
Spatial	Radius of gyration, Jurs descriptors, shadow indices, area, density, Length-to-breath ratios.
Structural	Topological symmetry, geometrical symmetry, combined symmetry, conformational flexibility indices, molecular distance edge descriptors, moment of inertia indices, geometric moment indices, number of single bonds, number of aromatic bonds.
Thermodynamic	Average energy resulting from all group energies, bond strain energy of molecule, angle strain energy of molecule, non-bonded strain energy of molecule, torsional strain energy of molecule, other strain energy of molecule, total strain energy of molecule.
Leadlikeness	Calculate of LogP (Meylan, Howard), calculate of LogS, calculate of LogP(Moriguchi, Hirono).
Topological	Wiener index, Kier and Hall molecular connectivity indices, path count and length descriptors, topological polar surface area (TPSA), Balban indices.

9.2.3 Regression analysis

The total number of descriptors calculated was 372. A systematic search was performed to determine significant descriptors. Some of the descriptors were rejected because they contained a value of zero for all the compounds. In order to minimize the effect of colinearity and to avoid redundancy correlation matrix developed with a cut off value of 0.6 and the variables physically removed from the analysis which show exact linear dependencies between subsets of the variables and multi colinearity (high multiple correlations between subsets of the variables). From descriptors thus remained, the set of descriptors that would give the statistically best QSAR models were selected from the large pool using a Genetic function approach. The genetic algorithm starts with the creation of a population of randomly generated parameter sets. The usage probability of a given parameter from active set is 0.5 in any of the initial population sets. The sets are then compared according to their objective functions. The form of objective function favors sets that have the r^2 as high as possible, while minimizing the number of parameters used as descriptors. The higher the score the higher the probability of a given set will be used for the creation of the next generation of sets. Creation of a consecutive generation involves crossovers between set contents, as well as mutations. The parameters set used for genetic algorithm includes: mutation 0.1, crossover 0.9, population 300, number of generations 1000, R^2 floor limit 50% and objective function was R^2/N_par. The algorithm runs until the desired number of generations is reached. Equations were developed between the observed activity and the descriptors. The best model was selected based on the r^2, r^2_{adj}, F-ratio and $q^2.r^2$ is an indication of the model data fit.

9.2.4 Validation test

The predictive capability of the equation (q^2) is determined using leave-one-out cross validation method. The relation for q^2 is as shown below.

$$q^2 = 1 - \frac{PRESS}{TOTAL} = 1 - \frac{\sum_{i=1}^{n}\left(y_{\exp} - y_{pred}\right)^2}{\sum_{i=1}^{n}\left(y_{\exp} - \overline{y}\right)^2}$$

Where, y_{pred}, $y_{\exp}$ and $\overline{y}$ are the predicted, experimental and mean values of activity, respectively. A large F indicates that the model fit is not a chance occurrence. r^2 and r^2_{adj}

above a value of 0.6 indicate good model fit while q^2 above 0.55 indicates good predictive capability for the model. Further, statistical significance of the relationship between the cytotoxic activity and chemical structure descriptors was obtained by randomization process. The test set was done by repeatedly permuting the activity values of the data set and using the permuted values to generate QSAR models and then comparing the resulting scores with the score of the original QSAR model generated from non-randomized activity values. If the original QSAR model is statistically significant, its score should be significantly better than that of permuted data (Deswal et al., 2006). The randomization test was performed at 90%, 95% and 99% confidence intervals. The higher the confidence level, the more conveniently randomization tests are run. In this direction, nine trials were run at 90% confidence level, 19 trials at 95%, 49 trials at 98% and 99 trials at 99% confidence level.

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

It has been shown that a high value of statistical characteristics need not be the proof of a highly predictive model (Golbraikh et al., 2002; Roy et al., 2007). Hence, in order to evaluate the predictive ability of our QSAR model, we used the method described by Golbraikh et al., (2002) and Roy et al., 2007). The values correlation coefficient of predicted and actual activities and correlation coefficient for regressions through the origin (predicted vs. observed activities and vice versa) were calculated using the regression of analysis Toolpak option of excel sheet and other parameters were calculated as reported by the above authors (Golbraikh et al., 2002; Roy et al., 2007). To arrive at the predictive r^2 (r^2_{pred}) the following equation was used (Roy et al., 2007).

$$r^2_{pred} = 1 - \frac{\sum \left(Y_{pred_{Test}} - Y_{Test} \right)^2}{\sum \left(Y_{Test} - \overline{Y}_{Training} \right)^2}$$

Where $Y_{pred_{Test}}$ and Y_{Test} are the predicted and observed activity values, respectively, of the Test set compounds and $\overline{Y}_{Training}$ is the mean activity value of the Training set. Further evaluation of the predictive ability of the model was done by determining the value of rm^2 by the following equation (Roy et al., 2007):

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

Where r^2 is the square correlation coefficient between observed and predicted values and r_0^2 is the squared correlation coefficient between observed and predicted values without intercept. The values of *k and k'*, slopes of the regression line of the predicted activity vs. actual activity and vice versa, were calculated using the following equations (Golbraikh et al., 2002):

$$k = \frac{\sum y_i \widetilde{y_i}}{\sum \widetilde{y_i}^2} \text{ and } k' = \frac{\sum y_i \widetilde{y_i}}{\sum y_i^2}$$

where $\widetilde{y_i}$ and *yi* are the predicted and actual activities, respectively.

9.3 Results and Discussion

The 119 active compounds considered as potential of P388 cell inhibition were segregated into 81 training and 38 test sets. The experimental IC_{50} values against the cell line P388 for these compounds set are available. With the wide range of difference between the IC_{50} values and the large diversity in the structures, the combined data set of 81 molecules and 38 molecules are ideal to be considered as training and test set, as both the sets do not suffer from bias, due to the similarity of the structures. The various molecular descriptors (372 in total) as described in Table 9.2 were calculated initially. By applying missing value test, zero test and correlation test with cutoff value of 0.6 we have discarded the most likely parameters that resulted in 117 parameters. Furthermore, parameters were discarded by applying genetic algorithm and finally selected 13 parameters for development of QSAR model. Taking a brute force approach, we increased the number of parameters in the QSAR equation one by one and evaluated the effect of addition of new term on the statistical quality of the model. As the correlation coefficient, r^2 can be easily increased by the number of terms in the QSAR equation; we took the cross-validation correlation

coefficient, q^2, as the limiting factor for a number of descriptors to be used in the final model. It was observed that the q^2 value increases till the number of descriptors in the equation reached up to 12 as shown in Table 9.3. When number of descriptors in the equation was 13, there was a decrease in q^2 value of the model. So the number of descriptors was restricted to 12 in the final QSAR model. The best significant relationship for the inhibition of P388 cell-line has been deduced to be

pIC_{50} = - 1.99 + 3.85 V5CH - 1.86 SNMN + 0.0625 SRMX - 8.49 SHDW4 + 11.6 SHDW5 - 0.088 L/B2 - 0.00401 STRA1 - 0.198 GEOM4 + 0.0683 ELOW1+ 0.108 DIP 0.0408 DIPZ - 0.0565 RPCS (1)

(N = 81; r^2 = 0.817; s = 0.354; PRESS = 11.759; r^2_{adj} = 0.781; q^2 = 0.714; F = 22.39)

where, V5CH, SNMN, SRMX, STRA1, L/B2, GEOM4, RPCS and ELOW1 signify 5th order chain molecular connectivity valence, minimum nucleophilic superdelocalizability, maximum free radical superdelocalizability, angle strain energy of molecules, length to breadth ratio, mass weighted length to width ratio, relative positive charge surface area and difference between minimum and maximum E-state values, respectively; SHDW4 and SHDW5 indicate shadow area 4 and 5; DIP and DIPZ signify dipole moment and dipole moment Z. It was found that the compound numbers **15, 96** and **111** were outliers with prediction error in between 0.72 to 0.91. The quality of the above QSAR model has been improved further by removing these compounds.

pIC_{50} = - 1.39 + 4.26 V5CH - 2.87 SNMN + 0.0508 SRMX - 7.65 SHDW4 + 11.3 SHDW5 - 0.299 L/B2 - 0.00545 STRA1 - 0.226 GEOM4 + 0.0905 ELOW1 + 0.0694 DIP + 0.0128 DIPZ - 0.160 RPCS_AM1 (2)

(N = 78; r^2 = 0.909; s = 0.236; PRESS = 4.840; r^2_{adj} = 0.872; q^2 = 0.888; F = 24.28)

where N is the number of compounds in the training set, r^2 is the squared correlation coefficient, s is the estimated standard deviation about the regression line, r^2_{adj} is the square of adjusted correlation coefficient for degree of freedom, F is the measure of variance which compares two models differing by one or more variables to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the F-test is

above a threshold value and q^2 is the square of the correlation coefficient of the cross-validation.

Table 9.3. Statistical assessment of QSAR equations with varying number of descriptors.

No. of descriptors	QSAR equation	r^2	press	q^2
1	pIC_{50} = - 5.57 + 10.7 SHDW5	0.40	29.27	0.32
2	pIC_{50} = - 3.30 - 6.10 SHDW4 + 12.1 SHDW5	0.45	25.65	0.41
3	pIC_{50} = - 3.69 - 5.96 SHDW4 + 11.9 SHDW5 + 0.0975 DIP	0.56	21.38	0.51
4	pIC_{50} = - 3.83 + 2.77 V5CH - 6.21 SHDW4 + 12.0 SHDW5 + 0.104 DIP	0.58	20.54	0.52
5	pIC_{50} = - 3.13 + 4.01 V5CH - 2.16 SNMN - 6.23 SHDW4 + 11.6 SHDW5 + 0.0888 DIP	0.65	17.85	0.59
6	pIC_{50} = - 2.94 + 3.79 V5CH - 2.05 SNMN - 6.35 SHDW4 + 11.0 SHDW5 + 0.118 DIP + 0.0403 DIPZ	0.67	16.78	0.61
7	pIC_{50} = - 3.16 + 3.68 V5CH - 1.94 SNMN - 6.35 SHDW4 + 11.0 SHDW5 + 0.0617 ELOW1 + 0.113 DIP + 0.0383 DIPZ	0.69	15.70	0.64
8	pIC_{50} = - 2.65 + 3.89 V5CH - 1.94 SNMN - 6.37 SHDW4 + 11.1 SHDW5 - 0.399 L/B2 + 0.0585 ELOW1 + 0.112 DIP + 0.0380 DIPZ	0.71	13.82	0.68
9	pIC_{50} = - 2.33 + 3.91 V5CH - 2.02 SNMN - 6.86 SHDW4 + 11.1 SHDW5 - 0.405 L/B2 - 0.00481 STRA1 + 0.0556 ELOW1 + 0.115 DIP + 0.0422 DIPZ	0.72	12.46	0.71
10	pIC_{50} = - 2.17 + 3.86 V5CH - 2.02 SNMN - 6.91 SHDW4 + 10.9 SHDW5 - 0.410 L/B2 - 0.00500 STRA1 + 0.0571 ELOW1 + 0.109 DIP + 0.0401 DIPZ - 0.0358 RPCS	0.73	11.92	0.72
11	pIC_{50} = - 2.29 + 3.81 V5CH - 2.03 SNMN + 0.0545 SRMX - 7.00 SHDW4 + 11.3 SHDW5 - 0.491 L/B2 - 0.00523 STRA1 + 0.0613 ELOW1 + 0.106 DIP + 0.0364 DIPZ - 0.0534 RPCS	0.79	10.71	0.75
12	pIC_{50} = - 1.39 + 4.26 V5CH - 2.87 SNMN + 0.0508 SRMX - 7.65 SHDW4 + 11.3 SHDW5 - 0.299 L/B2 - 0.00545 STRA1 - 0.226 GEOM4 + 0.0905 ELOW1 + 0.0694 DIP + 0.0128 DIPZ - 0.160 RPCS	0.91	4.84	0.89
13	pIC_{50} = 0.052 + 3.70 V5CH - 2.26 SNMN + 0.0519 SRMX - 0.00360 SHDW3 - 5.90 SHDW4 + 7.85 SHDW5 - 0.805 L/B2 - 0.00462 STRA1 - 0.0808 GEOM4 + 0.120 ELOW1 + 0.0389 DIP + 0.0495 DIPZ - 0.137 RPCS	0.92	9.37	0.78

The QSAR model developed in this study is statistically (r^2 = 0.909, q^2 = 0.888, F = 24.28) best fitted and consequently used for prediction of cell inhibition (pIC_{50}) of training and test sets of molecules as reported in Table 9.4 and Table 9.5.

Table 9.4. Observed and predicted inhibitory activity to P388 cell line of Training set of podophyllotoxin derivatives.

Compound	P388 cell line inhibition (pIC_{50})			Compound	P388 cell line inhibition (pIC_{50})		
No.	Observed	Predicted	Residual	No.	Observed	Predicted	Residual
2	2.00	2.24	0.24	62	-0.29	-0.16	0.13
3	2.00	1.77	0.23	63	-0.01	-0.16	0.15
6	1.22	0.84	0.38	64	-0.36	-0.46	0.10
7	1.22	0.66	0.56	66	0.70	0.90	0.20
8	1.30	1.22	0.08	67	0.00	-0.12	0.12
9	1.22	0.89	0.33	68	0.24	0.32	0.08
11	0.26	0.15	0.11	69	-0.80	-0.68	0.12
12	0.27	0.38	0.11	71	-0.28	-0.23	0.05
15	0.70	-0.02	0.72	72	0.00	0.40	0.40
16	1.00	0.46	0.54	73	-0.61	-0.77	0.16
17	0.64	1.04	0.40	74	-0.67	-0.68	0.01
19	0.26	0.62	0.36	76	0.00	-0.07	0.07
21	0.92	0.64	0.28	77	-0.75	-1.03	0.28
22	0.96	1.27	0.31	78	-1.00	-0.82	0.18
23	0.36	0.74	0.38	79	-1.32	-1.29	0.03
24	0.29	-0.24	0.53	80	-0.34	-0.73	0.39
27	-1.08	-1.25	0.17	81	-0.34	-0.07	0.27
28	-0.36	-0.32	0.04	87	-1.65	-1.24	0.41
30	-0.08	0.04	0.12	90	-1.22	-1.43	0.21
32	-1.06	-0.92	0.14	91	-0.18	-0.36	0.18
33	-0.99	-0.67	0.32	93	-1.90	-1.80	0.10
34	-0.99	-1.44	0.45	92	-2.00	-1.70	0.30
37	1.24	1.07	0.18	96	-0.30	-1.02	0.72
38	0.68	0.91	0.23	97	-2.00	-1.97	0.03
44	-1.37	-1.71	0.34	100	-1.60	-1.39	0.22
45	-1.29	-1.08	0.21	103	-1.78	-1.71	0.07
47	-0.99	-0.93	0.06	104	-2.00	-1.54	0.46
49	-0.04	-0.30	0.26	105	-1.85	-1.77	0.08
51	0.70	0.29	0.41	106	2.74	2.17	0.57
52	-0.30	-0.02	0.28	108	-0.69	-0.34	0.35
53	-0.37	-0.39	0.02	111	-0.41	0.50	0.91
55	-1.04	-0.73	0.31	113	0.04	0.40	0.36
56	-0.40	-0.43	0.03	114	1.32	0.79	0.53
57	1.70	1.80	0.10	115	2.28	2.18	0.10
59	0.64	0.43	0.21	116	0.89	1.22	0.33
61	0.32	0.34	0.02				

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

Table 9.5. Observed and predicted inhibitory activity to P388 cell line of Test set of podophyllotoxin derivatives.

Compound	P388 cell line inhibition (pIC_{50})			Compound	P388 cell line inhibition (pIC_{50})		
No.	Observed	Predicted	Residual	No.	Observed	Predicted	Residual
1	1.92	0.97	0.95	46	-1.06	-1.21	0.15
4	1.57	0.31	1.26	48	-1.68	-1.80	0.12
5	0.20	0.44	0.24	50	-0.75	-0.68	0.07
10	0.22	0.37	0.15	54	-0.36	-0.47	0.11
13	-0.36	-0.30	0.06	58	0.60	0.73	0.13
14	-0.32	-0.11	0.21	60	0.24	-0.05	0.29
18	-0.78	0.24	1.02	65	0.66	0.96	0.30
20	-0.01	0.11	0.12	70	-0.75	-0.92	0.17
25	0.92	1.76	0.84	75	-0.65	-0.57	0.08
26	1.89	1.91	0.02	82	-1.36	-1.60	0.24
28	-0.36	0.02	0.38	83	-1.08	-1.11	0.03
31	-1.08	-0.92	0.16	84	-0.41	-0.53	0.12
35	-1.37	-0.78	0.59	85	-0.38	-0.10	0.28
36	-0.54	-1.12	0.58	86	-0.75	-0.78	0.03
39	-0.71	-0.69	0.03	88	-1.09	-1.08	0.01
40	-1.38	-1.09	0.29	89	-0.77	-0.43	0.34
41	0.66	0.80	0.14	94	-2.00	-1.96	0.04
42	-0.87	-0.82	0.05	95	-1.59	-1.52	0.08
43	-0.04	0.21	0.25				

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

The quality of the prediction models for the training compounds before and after removal of outliers have been shown in Figure 9.1a & 9.1b. The regression coefficient (r^2) and the cross-validation coefficient (q^2) of the QSAR model were 0.909 and 0.888, respectively which revealed good predictive capabilities as shown by the leave-one-out method. The standard error of estimate for the model was 0.236, which is an indicator of the robustness of the fit and suggested that the predicted pIC_{50} based on equation (2) is reliable. The prediction residuals for the training set before and after removal of outliers have been shown in Figure 9.2a & 9.2b. Further, statistical significance of the relationship between the inhibitory activity and chemical structure descriptors was demonstrated by randomization procedure. Based on the randomization test, we found that the r^2 value of the original model was much higher than any of the trials using permuted data, showing thereby that the model developed is statistically significant and robust. The results of randomization test at various confidence levels are shown in Table 9.6.

Table 9.6. Results of randomization test performed to check the validation of model

Confidence level	90%	95%	98%	99%
Total trials	9	19	48	99
R from non-random	**0.909**	**0.909**	**0.909**	**0.909**
Random r > non-random	0	0	0	0
Random r < non-random	9	19	48	99
Mean value of r from random trials	**0.431**	**0.386**	**0.403**	**0.394**
Standard deviation of random trials	0.086	0.093	0.057	0.083

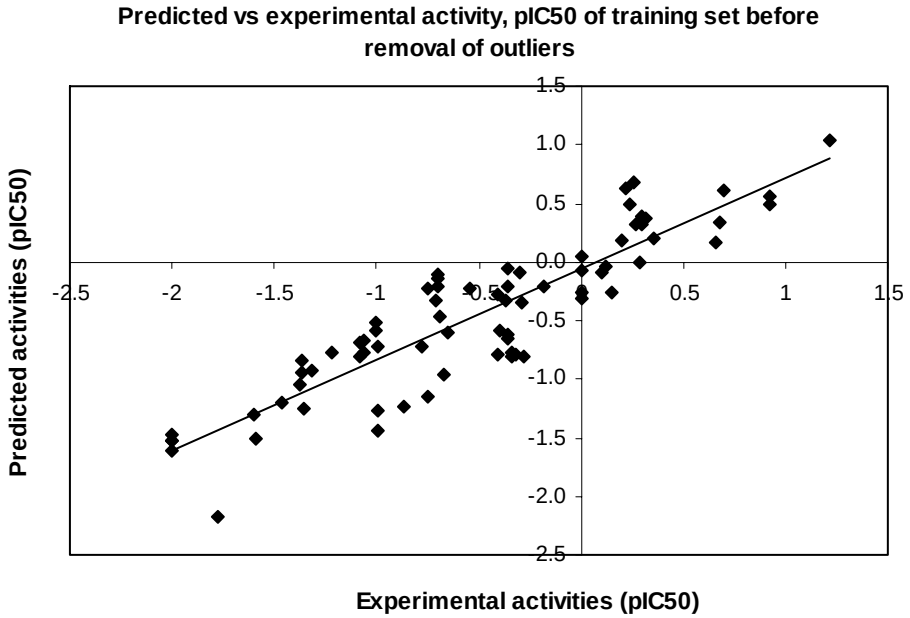


Figure 9.1. (a). Relationship between predicted and experimental activities as per equation (1) before removal of outliers.

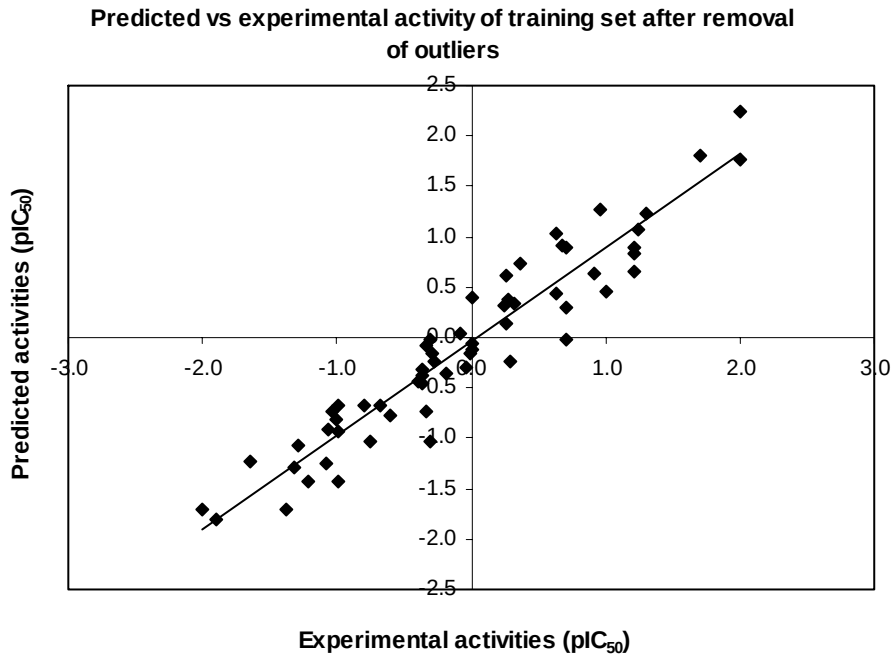


Figure 9.1. (b). Relationship between predicted and experimental activities as per equation (2) after removal of outliers.

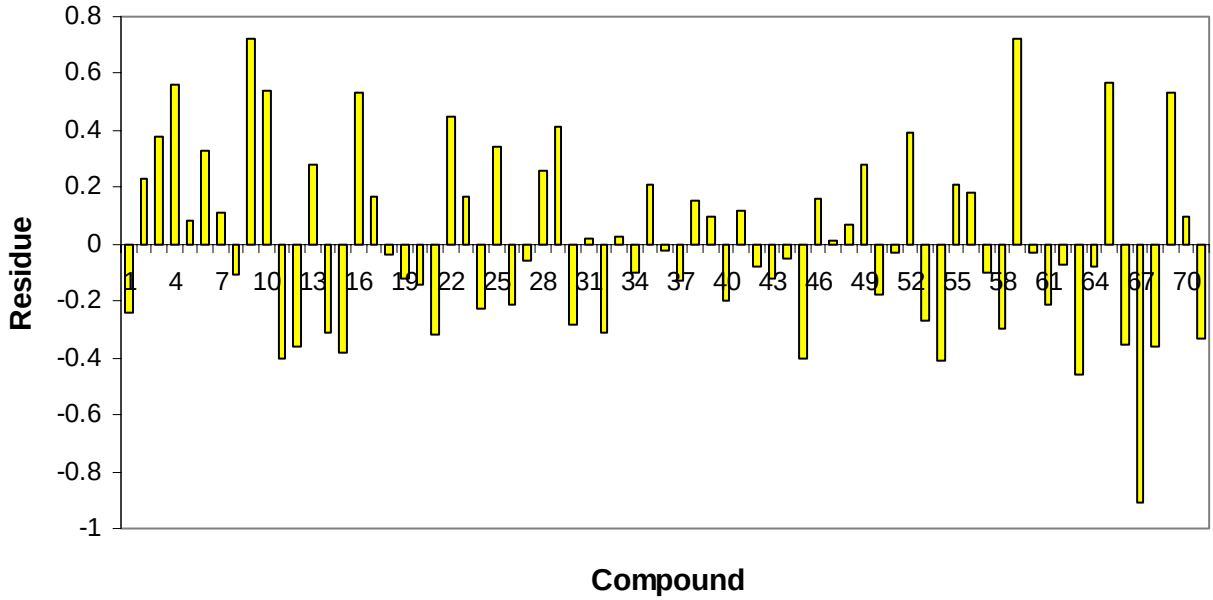


Figure 9.2. (a). The residuals between experimental activities and predicted activities from the QSAR models before removing the outliers in the training set.

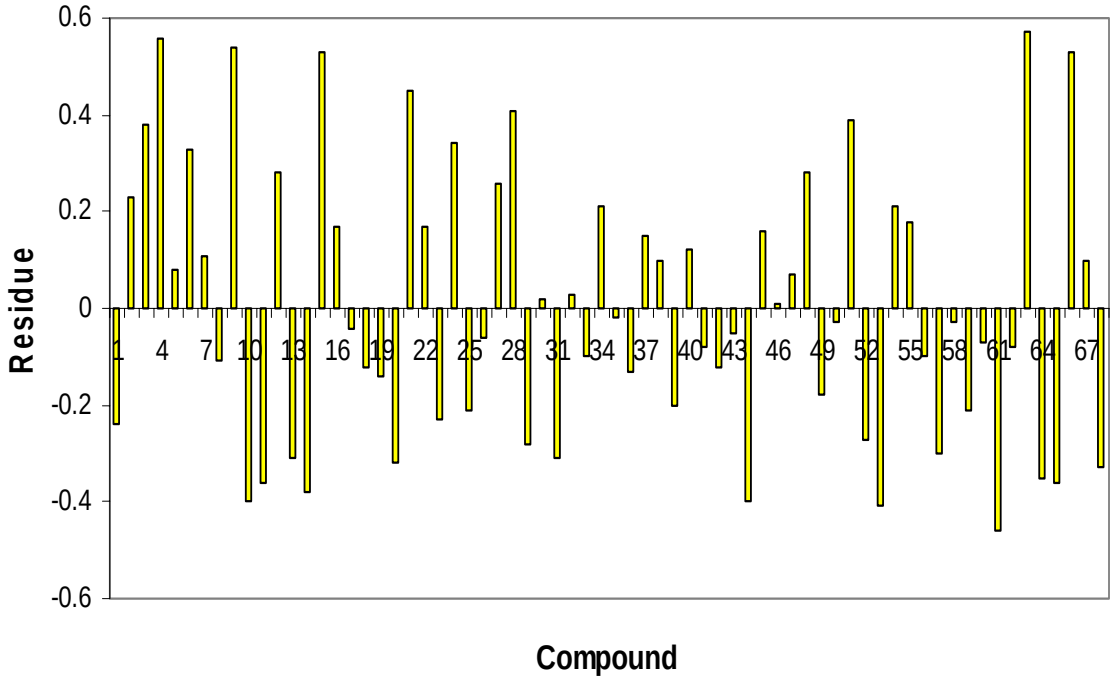


Figure 9.2. (b). The residuals between the experimental activities and predicted activities from the QSAR models after removing the outliers in the training set.

Table 9.7. Correlation matrix of the descriptors used in the QSAR model.

	V5CH	SNMN	SRMX	SHDW4	SHDW5	L/B2	STRA1	GEOM4	ELOW1	DIP	DIPZ	RPCS_AM1
VSCH	1.0											
SNMN	0.205	1.0										
SRMX	0.069	0.095	1.0									
SHDW4	0.023	0.036	0.017	1.0								
SHDW5	-0.056	-0.099	-0.170	0.332	1.0							
L/B2	0.056	0.058	0.076	0.044	0.035	1.0						
STRA1	0.066	-0.092	0.037	-0.256	-0.041	-0.081	1.0					
GEOM4	0.073	0.008	0.014	-0.122	-0.201	0.336	0.110	1.0				
ELOW1	0.007	-0.147	-0.082	0.025	0.013	-0.666	-0.053	0.398	1.0			
DIP	-0.135	-0.186	-0.088	-0.047	0.069	0.065	-0.008	-0.062	0.046	1.0		
DIPZ	0.140	0.069	0.065	0.086	0.138	-0.067	0.164	-0.053	-0.036	-0.494	1.0	
RPCS_AM1	-0.045	0.155	0.188	-0.067	-0.221	0.025	-0.091	-0.141	-0.027	-0.208	-0.015	1.0

The inter-correlation of the descriptors used in the final model (Eq. 2) was very low which is in conformity to the study that for a statistically significant model, it is necessary that the descriptors involved in the equation should not be inter-correlated with each other (Deswal et al., 2006). The correlation matrix for the used descriptors is shown in Table 9.7. To further check the inter-correlation of descriptors variance inflation factor (VIF) analysis was performed. In this model, the VIF values of these descriptors are 1.091 (V5CH), 1.143 (SNMN), 1.091 (SRMX), 1.236 (SHDW4), 1.337 (SHDW5), 1.302 (L/B2), 1.188 (STRA1), 1.727 (GEOM4), 1.377 (ELOW1), 1.527 (DIP), 1.524 (DIPZ) and 1.252 (RPCS_AM1). Based on VIF analysis it has been found that the descriptors used in the final model have very low inter-correlation.

Satisfied with the robustness of the QSAR model developed using training set, we applied the QSAR model to the podophyllotoxin analogues comprising the test set. As the experimental values of IC_{50} for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the QSAR model for new ligands. Table 9.5 represents the predicted pIC_{50} values of the test set based on equation (2). The overall root mean square error (RMSE) between the experimental and predicted pIC_{50} values was 0.265 which revealed good predictability. The squared correlation coefficient between experimental and predicted pIC_{50} values for the test set is also significant (r^2 = 0.824). Removing the outliers such as compound number **4** and **18** with prediction error > 1.0 leads to further increase in the correlation coefficient to the highest accuracy level (r^2 = 0.927). The figure 9.3a and 9.3b show the quality of the fit. The prediction residuals between the experimental and predicted pIC_{50} values for the test set before and after removal of outliers have been represented in Figure 9.4a & 9.4b. It clearly shows that the error rates are almost negligible.

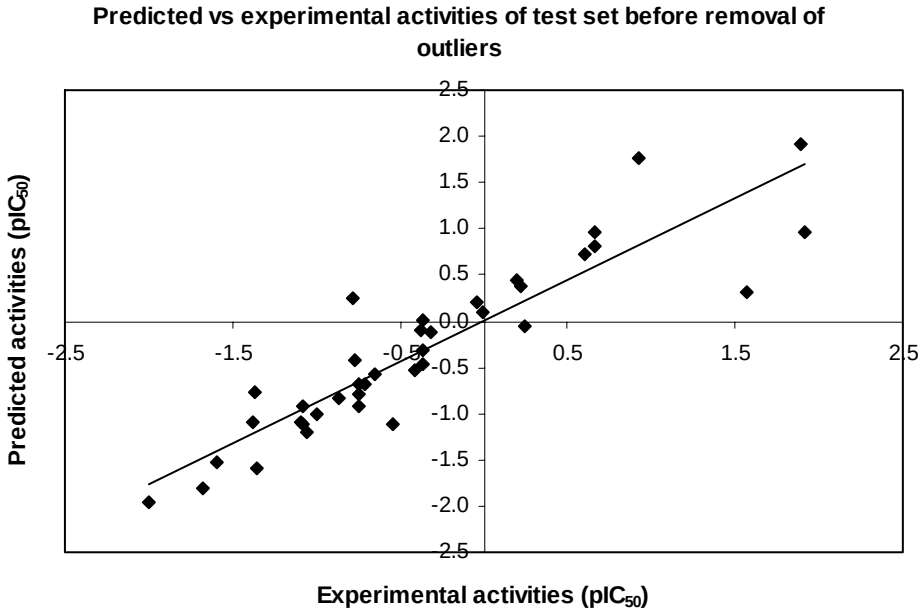


Figure 9.3. (a). Relationship between predicted and experimental activities as per equation (2) before removal of outliers.

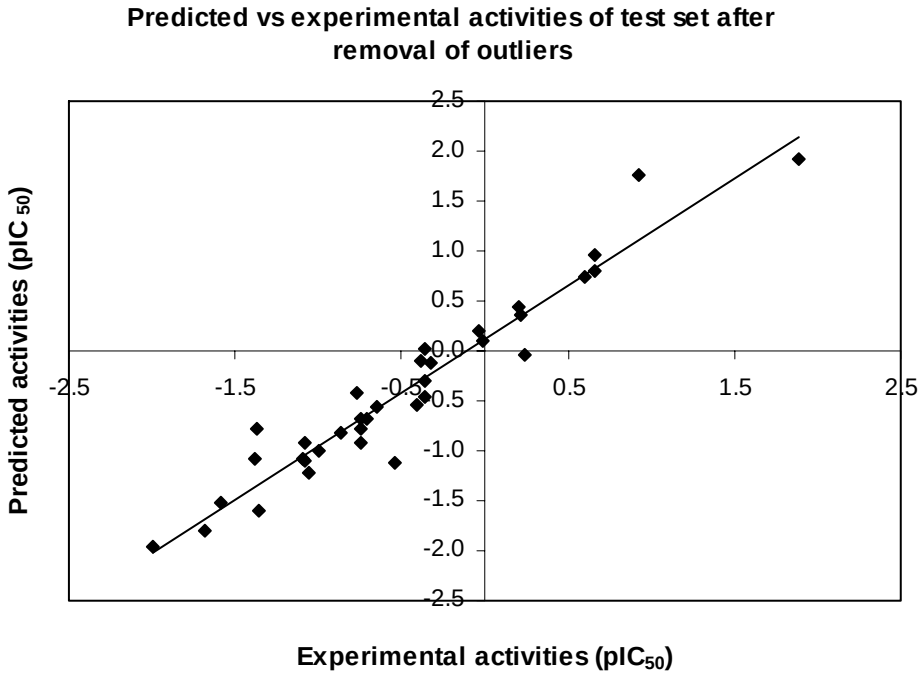


Figure 9.3. (b). Relationship between predicted and experimental activities as per equation (2) after removal of outliers.

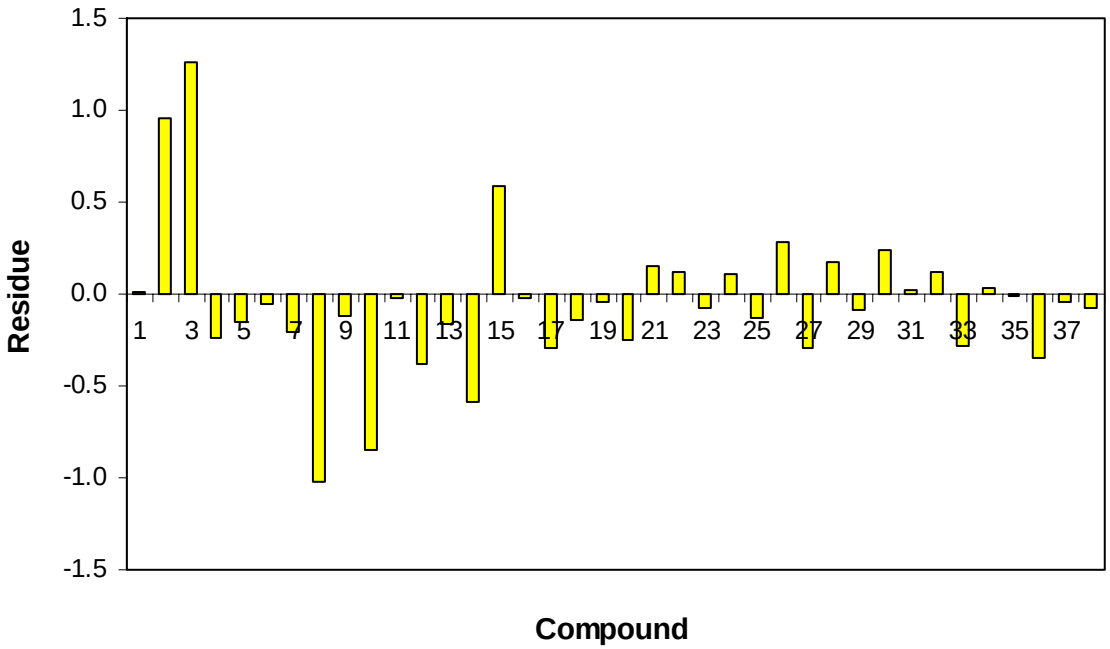


Figure 9.4. (a). The residuals between experimental activities and predicted activities from the QSAR models before removing the outliers in the test set.

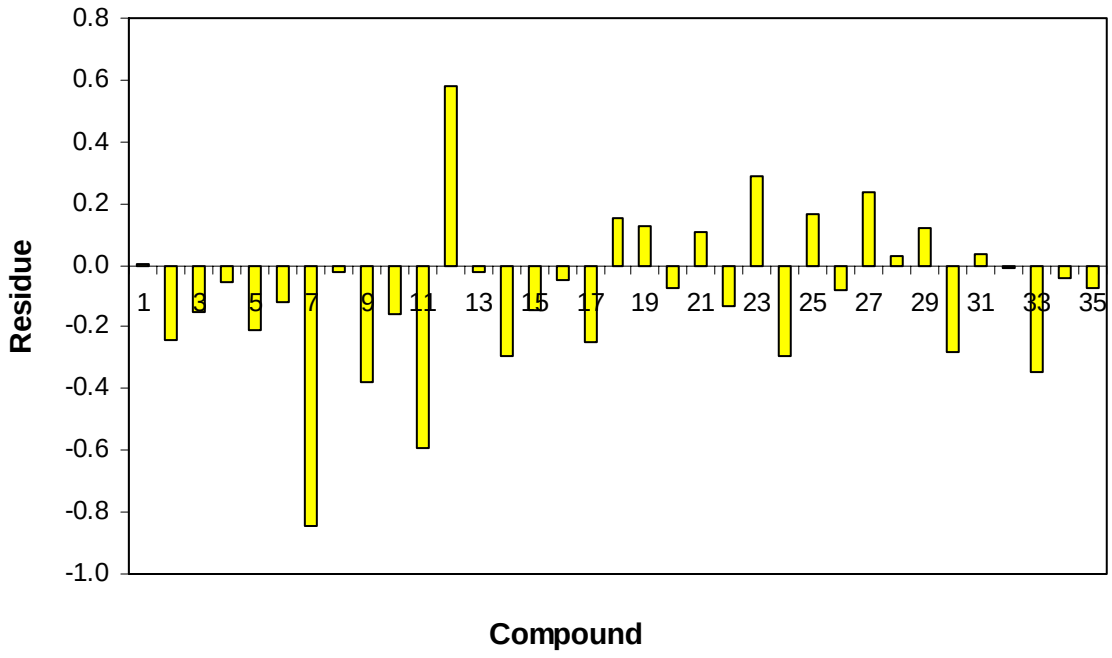


Figure 9.4. (b). The residuals between experimental activities and predicted activities from the QSAR models after removing the outliers in the test set.

The estimated correlation coefficient between experimental and predicted pIC_{50} values with intercept (r^2) and without intercept (r_0^2) are 0.927 and 0.868 respectively. The value of $[(r^2 - r_0^2)/r^2]$ = (0.927 – 0.868)/0.927 = 0.063, which is less than 0.1 (stipulated value) (Golbraikh et al., 2002). Also the values of k and k' were 0.925 and 1.004, which are well within the specified range of 0.85 and 1.15. Being the value of q^2 = 0.888, the model corroborates with the criteria for a QSAR model to be highly predictive (Golbraikh et al., 2002). Also the value of r^2_{pred} = 0.918 and rm^2 = 0.702 were found to be in the acceptable range (Roy et al., 2007), thereby indicating the good external predictability of the QSAR model.

To evaluate the accuracy of the QSAR model for tubulin polymerization inhibition potencies, we have taken a separate data set called validation set consisting of 16 analogues of podophyllotoxin (Table 9.7). Their experimental activity and chemical structures were obtained from literature (Haar et al., 1996; Lokie et al., 1978). The experimental activity (IC_{50} value) of these compounds obtained from in vitro study of tubulin polymerization inhibition (TPI). For all the compounds QSAR predictions produce exactly the same trend for tubulin polymerization inhibition, even though the exact magnitudes of these values do not match very well to experimental values (Table 9.8). Podophyllotoxin competitively inhibit the binding of colchicine to tubulin (Hastie et al., 1991), implying that it binds to tubulin at the same site. The structural feature of podophyllotoxin that shares with colchicine is the trimethoxyphenyl moiety. For colchicine and podophyllotoxin, it has been suggested that the binding sites for the two drugs do not completely overlap, with the trimethoxyphenyl rings of the agents binding in the same site on the tubulin heterodimer (Andreu et al., 1982). Harr et al., (1996) suggested that the trimethoxyphenyl rings of the two drugs were situated in different regions of space, nearly orthogonal to each other. This revealed that these rings may bind to different regions of tubulin at the colchicine binding site. The RMSE between the experimental and predicted TPI was 0.295. Figure 9.5 shows the quality of fit between the experimental and predicted tubulin polymerization inhibition of the validation set.

Table 9.8. Observed and predicted inhibitory activity to Tubulin polymerization of validation set of podophyllotoxin derivatives.

Compound No.	Compound name	Tubulin polymerization inhibition (pIC_{50})		
		Observed	Predicted	Residual
1	Podophyllotoxin	0.22	0.73	0.51
2	Epipodophyllotoxin	-0.70	-0.14	0.56
3	Deoxypodophyllotoxin	0.30	0.55	0.25
4	β-Peltatin	0.15	-0.04	0.19
5	α-Peltatin	0.30	0.33	0.03
6	4'-Demethylpodophyllotoxin	0.30	0.36	0.06
7	4'-Demethylepipodophyllotoxin	-0.30	0.34	0.64
8	4'-Demethyldeoxypodophyllotoxin	0.70	0.80	0.10
9	Dehydropodophyllotoxin	-1.40	-1.13	0.27
10	Anhydropodophyllol	0.00	-0.27	0.27
11	Podophyllotoxin cyclic sulfide	-1.00	-0.90	0.10
12	Podophyllotoxin-cyclic ether	0.00	-0.37	0.37
13	Deoxypodophyllotoxin-cyclic ether	0.10	0.12	0.02
14	Deoxypodophyllotoxin-cyclopentane	-0.70	-0.70	0.00
15	Deoxypodophyllotoxin-cyclopentanone	-0.70	-0.66	0.04
16	Deoxypodophyllotoxin-cyclic sulfide	-1.00	-1.01	0.01

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

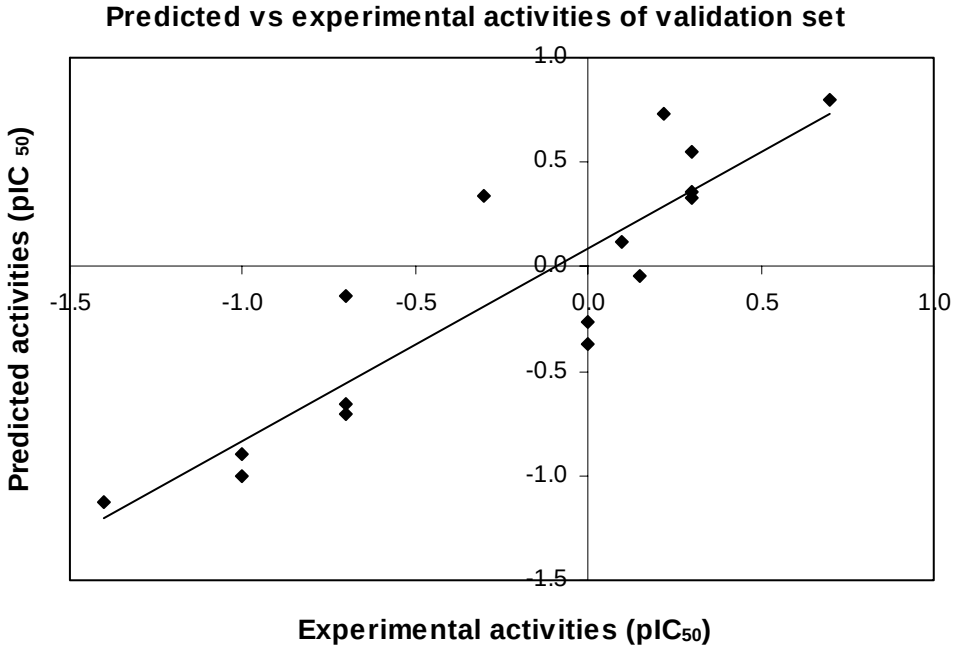


Figure 9.5. Relationship between predicted and experimental activities as per equation (2).

Based on the QSAR model, it is observed that one of the important parameters that contribute to the potential activity of podophyllotoxin is SHDW5, shadow area 5. On the contrary, it is observed that the shadow area 4 has a negative effect on the activity, as defined by the negative value of the descriptor, SHDW4. Other parameters that also contribute to the activity as per Eq. (2) are (a) 5^{th} order chain molecular connectivity valence (V5CH) (b) maximum free radical superdelocalizability (SRMX) (c) difference between minimum and maximum E-state values (ELOW1) (d) dipole moment (DIP), which have positive effect on the activity i.e. the activity increases with the increase of the values of these parameters. Whereas the descriptors such as (a) minimum nucleophilic superdelocalizability (SNMN) (b) length to breadth ratio (L/B2) (c) angle strain energy of molecule (STRA1) (d) mass weighted length to breadth ratio (GEOM4) and (e) relative positive charge surface area (RPCS), were negatively related to the activity.

9.4 Conclusion

In this study, we used a more systematic way of variable selection in order of missing value test → zero test → simple correlation test → multicollinearity test → genetic algorithm to obtain QSAR models for 119 podophyllotoxin derivatives. Using a combination of topological, electro-topological-state indices, electronic and thermodynamic descriptors of chemical structures, we have built several robust QSAR models with high values of q^2 (for training sets) and predictive r^2 (for test set). The high predictive ability of the models allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates. Therefore, these models should facilitate the rational design of novel derivatives, guide the design of focused libraries based on the podophyllotoxin skeleton and facilitate the search for related structures with similar biological activity from large databases.

CONCLUSION

In the present study high genetic variations among the *Podophyllum hexandrum* populations from the Northwestern Himalayas (Himachal Pradesh) have been investigated using RAPD, ISSR and AFLP based DNA profiling. Parallel studies on a subset of the same material using RAPD, ISSR and AFLP have shown that information on genetic diversity and relationships between populations is congruent among all the three molecular markers. The high genetic variations in *P. hexandrum* may be attributed partly to the cross-pollinated nature or clonal propagation of *P. hexandrum.* The low level of genetic diversity within the population and low gene flow among populations detected in this study point towards the possibility of a single isolated population possessing unique genotypes not found in other populations. Based on polymorphic features, genetic diversity, genetic similarity and gene flow among the populations of *Podophyllum* based on RAPD, ISSR and AFLP study, we have come to the conclusion that for both *in situ* conservation and germplasm collection expeditions for this species should be specifically designed to include representative populations with the highest genetic variations. For commercial level cultivation of this important plant as a source of podophyllotoxin, it is necessary to select the appropriate sites and hence both the ANN and MLR models developed in this study could provide useful information.

We have compiled a virtual library of podophyllotoxin analogues built through structural modification of scaffold structure of natural podophyllotoxin. Docking and rescoring have been done using Prime/MM-GBSA in the work to get insights into ligand:tubulin interactions and corresponding cytotoxic activity of podophyllotoxin analogues. Several sets of podophyllotoxin analogues have been studied in the docking simulations. Results showed that these analogues bind in a very similar mode. This suggests that they interact with the enzyme in a very similar way. This makes a credible prediction model of the cytotoxic activity (pIC_{50}) calculation possible. The calculated Glide score and binding free energy value of a set of structural analogues demonstrate excellent linear correlation to the experimental cytotoxic activity. These models could be useful to predict the range of activities for new podophyllotoxin analogues. We also found that refinement of poses and consequent rescoring using Prime/MM-GBSA led to better predictability of pIC_{50}.

The information that we have expressed in this study may lead to the designing (synthesis) of more potent podophyllotoxin derivatives for inhibition of microtubule polymerization.

We have demonstrated that the SGB-LIE method can be applied to estimate the binding free energy with a high level of accuracy for a diverse set of podophyllotoxin analogues with tubulin. The magnitude of free energy changes upon binding of these analogues to tubulin has been directly correlated with the experimental potency of these inhibitors. The method has reproduced experimental data with reasonably small error for the majority of podophyllotoxin analogues. Using LIE methodology, we have been able to verify the experimental observation of a few podophyllotoxin analogues, quite accurately in comparison to the binding kinetics in vitro. The close estimation of inhibition potencies of a wide range of structural derivatives for podophyllotoxin establishes the SGB-LIE methodology as an efficient tool for screening novel compounds with very different structures. The results obtain will give information on how the chemical structure of the drug should be modified to achieve suitable interactions and for the rapid prediction and virtual prescreening of anti-tumor activity. This will lead to new proposals regarding possible improvements to the therapeutic indices of podophyllotoxins. Compared to the empirical methods, such as scoring function approaches, the LIE method is more accurate due to the semiempirical approach adopted in which experimental data are used to build the binding affinity model.

In most methods to build a 3-D QSAR, the selection of meaningful molecular descriptors are key factors to determine the success of a model. In this study we have used a more systematic ways such as: missing value test → zero test → simple correlation test → genetic algorithm to obtain the meaningful descriptors leading to QSAR model development. The high predictability of the model developed here in this study allows virtual screening of chemical databases or virtual libraries determined by either synthetic feasibility or commercial availability of starting materials to prioritize the synthesis of most promising candidates. Therefore, these models should facilitate the rational design of novel derivatives, design of focused libraries based on the podophyllotoxin skeleton and the search for related structures with similar biological activity from large databases.

The satisfactory results obtained for virtual screening of podophyllotoxin analogues and prediction of cytotoxic activity based on screening methodology of Docking-MM-GBSA, SGB-LIE and QSAR will help design new generation inhibitors. In the work we have used different computational approaches to explore the binding interaction with tubulin polymerization inhibition potency and cytotoxic activity of podophyllotoxin derivatives. We hope that knowledge and insight on the screening models learnt from the work will help a lot in the battle against cancer and benefit humanity at large.

Bibliography

Adams RP & Demeke T (1993). Systematic relationship in *Juniperus* based on random amplified polymorphic DNAs (RAPDs). *Taxon* 42: 553-571.

Aggarwal A, Kruczynski A, Frankfurter A, Correia JJ & Lobert S (2008). Murine leukemia P388 vinorelbine-resistant cell lines are sensitive to vinflunine. *Invest. New Drugs*, 26: 319–330.

Airi S, Rawal RS, Dhar U & Purohit AN (1997). Population studies on *Podophyllum hexandrum* Royal – a dwindling medical plant of the Himalaya. *Plant Genet. Resour. Newsl.* 110: 29-34.

Aitkin SA, Tinker NA, Mather DE & Fortin MG (1994). A method for detecting DNA polymorphism in large populations. *Genome* 37: 506–508.

Ajani JA, Mansfield PF, & Dumas P (1999). Oral etoposide for patients with metastatic gastric adenocarcinoma. *Cancer J. Sci. Am.* 5:112–114.

Ajibade SR, Weeden NF & Chite SM, (2000). Inter simple sequence repeat analysis of genetic relationship in the genus *Vigna. Euphytica* 111: 47-55.

Ajmone Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M (1998). Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor. Appl. Genet.* 96: 219-227.

Alam MA & Naik PK (2009). Impact of soil nutrients and environmental factors on podophyllotoxin content among 28 *Podophyllum hexandrum* Populations of Northwestern Himalayan Region using linear and non-linear approach. *Comm. in Soil Sci. Plant Anal.* (in press).

Alam MA, Pallavi G, Gulati AK, Mishra GP & Naik PK (2008). Characterization of genetic structure of *Podophyllum hexandrum* populations–an endangered medicinal herb of Northwestern Himalaya using ISSR-PCR markers and its relatedness with podophyllotoxin content. *Afr. J. Biotech.* 7 (8): 1028-1040.

Allevi P, Anastasia M, Ciureda P, Bigatti E & Macdonald P (1993). Stereoselective glucosidation of *Podophyllum* lignans. A new simple synthesis of etoposide. *J. Org. Chem.* 58: 4175-4178.

Andreu JM & Timasheff SN (1982). Conformational states of tubulin liganded to colchicine, tropolone methyl ether, and podophyllotoxin. *Biochem.* 21: 6465-6476.

Andreu JM & Timasheff SN (1982). Interaction of tubulin with single ring analogues of colchicine. Biochemistry, 21(3), 534-543.Atken SG & Gardiner SE (1991) SDS-PAGE of seed proteins in *Festuca* (poaceae): taxonomic implications. *Can. J. Bot.* 69: 1215-1432.

Anon (1970). *Podophyllum hexandrum*. The wealth of India: A Dictionary of Indian Raw Material and industrial products: Raw Materials. Vol. VIII: Ph.Re. Publication and Information Directorate. CSIR, New Delhi. Pp. 170-174.

Aqvist J, Medina C & Samuelsson JE (1994). A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Prot. Eng.* 7: 385-391.

Arens P, Coops H, Jansen J & Vosman B (1998). Molecular genetic analysis of black polar (*Populus nigra* L.) along Dutch rivers. *Mol. Ecol.* 7: 11-18.

Atkinson P & Tatnall A (1997). Neural networks in remote sensing. *Int. Jour. Rem. Sen.* 18 (4): 699-709.

Ayala FJ & Kiger JA (1984). *Modern Genetics*, 2nd ed Benjamin/Cummings, Menlo Park USA.

Ayres DC & Lim CK (1982). Modification of the Pendant Ring of Podophyllotoxin. *Cancer Chemother. Pharmacol.* 7:99-101.

Ayres DC & Loike JD (1990). Lignans: Chemical, Biological and Clinical Properties, *Cambridge University Press*, Cambridge, (chapters 3 and 4).

Badhwar RL & Sharma BK (1963). A note on germination of *Podophyllum* seeds. *Indian For.* 89: 445-447.

Barreca ML, Carotti A, Carrieri A, Chimirri A, Monforte AM, Calace MP & Rao A (1999). Comparative molecular field analysis (CoMFA) and docking studies of nonnucleoside HIV-1 RT inhibitors. *Bioorg. Med. Chem.* 7: 2283-2292.

Beers SA, ImaKura Y, Dai HJ, Li DH, Cheng YC & Lee KH (1988). Anti-AIDS agents. 29[1]. Anti-HIV activity of modified podophyllotoxin derivatives *J. Natl. Prod.* 51: 901.

Berkowitz DB, Maeng JH, Dantzig AH, Shepard RL & Nomnan BH (1996). Chemoenzymatic and Ring E-Modular Approach to the (-)-Podophyllotoxin Skeleton. Synthesis of 3′,4′,5′-Tridemethoxy-(-)-podophyllotoxin. *J. Am. Chem. Soc.* 118: 9426.

Bertounesque E, Imbert T & Monneret C (1996). Synthesis of podophyllotoxin A-ring pyridazine analogue. *Tetrahedron* 52: 14236.

Bhadula SK, Singh A, Lata H, Kunyal CP, & Purohit AN (1996). Genetic resources of *Podophyllum hexandrum* Royal, an endangered medicinal species from Garhwal Himalaya, India. *Int. plant Gen. Resour Newsl.* 106: 26-29.

Bhojwani SS & Razdan MK (1996). Plant Tissue Culture: The Theory and Practice, *Elsevier*, Amsterdam. Revised ed. pp. 48-49: 246-268.

Bjorneboe O, Moen F, Nygaard H, Haavik TK & Svensson B (1998). CPH-82 (Reumacon), versus auranofin (Ridaura): a 36-week study of their respective onset of action rates in RA. *Scand. J. Rheumatol.* 27: 26–31.

Blair MW, Panaud O & McCouch SR (1999). Inter-simple sequence repeats (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice *(Oryza sativa* L.). *Theo. Appl. Genet.* 98: 780–792.

Bornet B & Branchard M (2001). Non-anchored inter-simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting. *Pl. Mol. Bio. Rep.* 22: 427–432.

Bornet B, Goraguer F, Joly G & Branchard M (2002). Genetic diversity in European and Arentinian cultivated potatoes (*Solanum tuberosum*) subsp. *tuberosum* detected by inter-simple sequence repeats (ISSRs). *Genome* 45: 48-484.

Botstein D, White RL, Skolink MH & Davies RW (1980). Construction of a genetic map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* 32: 314-331.

Brenwald NP, Gill M J & Wise R (1998). Prevalence of a Putative Efflux Mechanism among Fluoroquinolone-Resistant Clinical Isolates of *Streptococcus pneumoniae*, *Antimicrob. Agents Chemother*. 42: 2032-2035.

Brewer CF, Loik JD, Horwitz SB, Sternlicht H & Gensler WJ (1979). Conformational analysis of podophyllotoxin and its congeners. Structure--activity relationship in microtubule assembly. *J. Med. Chem* 22: 215.

Broomhead AJ & Dewick PM (1990). Tumor inhibitory aryltralin lignans in *Podophyllum versipelle, Diphyllcia cymosa* and *Diphyllicia grayi*. *Phytochem.* 29: 3831-3837.

Brown JKM (1996). The choice of molecular marker methods for population genetic studies of plant pathogens. *New Phyto.* 133:183-195.

Buolamwini JK, & Assefa H (2002). CoMFA 3D-QSAR Analysis of HIV-1 RT Non-nucleoside inhibitors, TIBO Derivatives based on Docking conformation and alignment. *J. Med. Chem.* 45: 841-852.

Bush EJ, & Jones DW (1995). Asymmetric total synthesis of (-)-podophyllotoxin. *J. Chem. Soc. Perkin Trans.* 1: 1489–1492.

Buss AD & Waigh RD (1995). In *Natural Products as Leads for new Pharmaceuticals*. Wolff, M.E., Ed.; John Wiley & Sons, Inc: New York, Charper 24.

Caetano-Anolles G, Bassam BJ, & Bresshoff PM (1991). DNA amplification fingerprinting: A strategy for genome analysis. *Plant Mol. Biol. Rptr.* 9(4): 294-307

Campose LP, Raelson JV & Grant WI (1994). Genome elationship among lotus species based on random amplified polymorphic DNA (RAPD). *Theor. Appl. Genet.* 88: 317- 422.

Canel C, Dayan FE, Ganzera M, Rimando A, Burandt C, Khan I. & Moraes R. M. (2001). Increased yield of podophyllotoxin from leaves of *Podophyllum peltatum* L. by in situ conversion of podophyllotoxin 4-O-β-D glucopyranoside. *Planta Medica* 67: 1-3.

Canel C, Moraes RM, Dayan FE & Ferreira D (2000). Molecules of interest "Podophyllotoxin". *Phytochem.* 54: 115-120.

Cannoly AG, Godwin ID, Cooper M & Delacy IH (1994). Interpretation of randomly amplified polymorphic DNA marker data for fingerprinting sweet potato. (*Inomoea batatas* L.). *Theor. Appl. Genet.* 88: 332-336.

Carlsson J, Ander M, Nervall M, Aqvist J (2006) Continuum Solvation Models in the Linear Interaction Energy Method. *J. Phys. Chem.* 110 (24):12034-12041.

Castro MA, Gordaliza M, Miguel del Corral JM & San Feliciano A (1994). Preparation of triols and ethers relaterd to *popodhyllotoxin. Org. Prep. Proced. Int.* 26: 539.

Charters YM, Robertson A, Wilkinson MJ & Ramsay G (1996). PCR analysis of oilseed rape cultivars (*Brassica napus* L) using 5'-anchored simple sequence repeat (SSR) primers. *Theor. Appl. Genet.* 92: 442-447.

Chatterjee R (1952). Indian Podophyllum. *Econ. Bot.* (Indian). 6:342–354.

Chen JM, Liu X, Gituru WR, Wang JY & Wang QF (2005). Genetic variation within the endangered quillwort *Isotoma petraea* (Lobeliaceae). *Mol. Ecol.* 8: 775-789.

Cho SJ, Kashiwada Y, Bastow KF, Cheng YC & Lee KH (1996). Antitumor agents 163 three-dimensional quantitative structure-activity relationship study of 4'-O-Demethylepipodophyllotoxin Analogs Using the Modified CoMFA/q^2-GRS Approach. *J. Med. Chem.* 39: 1396-1405.

Chowdhury MA, Vandenberg B & Warkentin T (2002). Cultivar identification and genetic relationship among selected breeding lines and cultivars in chickpea (*Cicer arietinum* L.). *Euphytica* 127: 317–325.

Clifford G & Lau Y (1992). Neural Networks: Theoretical Foundations and Analysis. *IEEE, New York, NY*.

Comes HP & Abbott RJ (1999). Reticulate evolution in the Mediterranean species complex of *Senecio* sect. *Senecio*: uniting phylogenetic and population-level approaches, in

Molecular systematics and plant evolution, (eds. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornal), *Taylor and Francis, London.*171-198.

Cortese F, Bhattacharyya B & Wolff P (1977). Podophyllotoxin as a probe for the colchicine binding site of tubulin. *J. Biol. Chem.* 252: 1134.

Cragg GM & Newman DJ (2004). A tale of two tumor targets: topoisomerase I and tubuline: the Wall and Wani contribution to cancer chemotherapy. *J. Nat. Prod.* 67: 259-296.

Crochemore M L, Huyghe C, Kerlan MC, Durand F & Julier B (1996). Partitioning and distribution of RAPD variation in a set of population of the *Medicago sativa* complex. *Agronomie (Paris)* 16: 421-432.

Culley TM & Wolfe AD (2001). Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. *Heredity* 86(5): 545-556.

Curtis JMR & Taylor EB (2003). The genetic structure of coastal giant salamanders (*Dicamptodon tenebrosus*) in a managed forest. *Biol. Conserv.* 115:45–54.

Dahlberg JA, Zhang X, Hart GE & Mullet JE (2002). Comparative Assessment of Variation among Sorghum Germplasm Accessions Using Seed Morphology and RAPD Measurements. *Crop Sci.* 42: 291-296.

Daley L, Guminski Y, Demerseman P, Kruczynski A, Etievant C, Imbert T, Hill BT & Monneret C (1998). Synthesis and antitumor activity of new glycosides of epipodophyllotoxin, analogues of etoposide, and NK 611. *J. Med. Chem*. 41:4475-4485.

Damayanthi Y & Lown JW (1998). Structural analogs of tylophora alkaloids may not be functional analogs. *Curr. Med. Chem.* 5: 205.

David-Pfeuty T, Simon C & Pantaloni D (1979). Effect of antimitotic drugs on tubulin GTPase activity and self-assembly. *J Biol Chem.* 254(22):11696–11702.

De Knijff P, Denkers F, van Swelm ND & Kuiper M (2001). Genetic affinities within the herring gull *Larus argentatus* assemblage revealed by AFLP genotyping. *J. Mol. Evol.* 52:85–93.

de Kroon H, Whigham DF & Watson MA (1991). Developmental ecology of mayapple: Effects of rhizome severing, fertilization and timing of shoot senescence. *Func. Ecol.* 5: 360–368.

Debener T, Janakiram T & Mattiesch L (2000). Sports and seedlings of rose varieties analyzed with molecular markers. *Plant. Breed.* 119:71–74.

Demeke T & Adams RP (1994). The use of RAPD to determine germplasm collection strategies in the African species *Phytplaca dodecandra* (Phytolaccaceae), in *Conservation of plant genes II: Utilization of ancient and modern DNA,* (eds. R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. (Adams), Missouri Botanical Garden, Missouri, pp. 131-140.

Deswal S & Roy N (2006). Quantitative structure activity relationship studies of aryl heterocycle-based thrombin inhibitors, *J. Med. Chem*. 41(11):1339-1346.

Dore JC, Viel C, Pageot N, Gordaliza M, Castro MA, Miguel del Corral JM. & San Feliciano A (1996). Multivariate analysis approach to antineoplastic and antiviral structure-activity relationships to a series of podophyllotoxins *J. Pharm. Belg.* 51: 9.

Downing KH & Nogales E (1998). New insight into microtubules structure and function from the atomic model of tubuline. *Eur. Biophys. J.* 27: 431-436.

Downing KH & Nogales E (1998). Tubuline structure: insight into microtubule properties and functions. *Curr. Opin. Struct. Biol.* 8:785-791.

Drummond RSM, Keeling DJ, Richardson TE, Gardner RC & Wright SD (2000). Genetic analysis and conservation of 31 surviving individuals of a rare New Zealand tree, *Metrosideros bartlettii* (Myrtaceae). *Mol. Ecol.* 9:1149–1157.

Ehrlich HA, Gelfand D & Sninsky JJ (1991). Recent advances in the polymerase chain reaction. *Science* 252:1643-1650.

Ekstrom KK, Hoffma T, Linne, B, Eriksoon & Glimelius B (1998). Single-dose etoposide in advanced pancreatic and biliary cancer; a phase II study. *Oncol. Rpt.* 5:931–934.

Eliopoulos GM & Moellering RC Jr. (1996). Antimicrobial combinations, p. 330–396. *In* V. Lorian (ed.), Antibiotics in laboratory medicine, 4th ed. Williams & Wilkins, Baltimore, MD.

Elizondo DA, McClendon RW & Hoogenboom G (1994). Neural network models for predicting flowering and physiological maturity of soybean. *Transactions of the ASAE* 37: 981-988.

Endress R (1994). Soil nutrient factors related to salidroside production of *Rhodiola sachalinensis* distributed in Chang Bai Mountain. *Plant Cell Biotech*. Pp.187-242.

Escaravage N, Questiau S, Pornon A, Doche B & Taberlet P (1998). Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP makers. *Mol. Ecol.* 7:975–982.

Esselman EJ, Li JQ, Crawford DJ, Windus JL & Wolfe AD (1999). Clonal diversity in the rare *Calamagrostis porteri* spp. Insperata (Poaceae): comparative results for

allozymes and random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) markers. *Mol. Ecol.* 8:443-451.

Excoffier L, Smouse PE & Quattro JM (1992). Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479- 491.

Fahima T, Roder MS, Wendehake K, Kirzhner VM & Nevo E (2002). Microsatellite polymorphism in natural population of wild emmer wheat, *Trititum dicoccoides* in Israel. *Theor. Appl. Gent.* 104:17-29.

Fay DA & Ziegler HW (1985). Botanical source differentiation of *Podophyllum* resin by high performance liquid chromatography. *J liq Chromatogr.* 8:1501-1506.

Felsenstein J (1995). Phylogenies from restriction sites, a maximum likelihood approach. *Evolution* 46: 557-574.

Fernández ME, Figueiras AM & Benito C (2002). The use of ISSR and RAPD markers for detecting DNA polymorphism, genotype identification and genetic diversity among barley cultivars with known origin. *Theor. Appl. Genet.* 104: 845–851.

FirstDiscovery2.7, 2.7 ed.; Schrodinger Inc.: Portland, 2004.

Folkertsma RT, Rouppe van der Voot JNAM, de Groot KE, van Zandvoort PM, Schots A, Gommers FJ, Helder & Bakker J (1996). Gene pool similarities of potato cyst nematode populations assessed by AFLP analysis. *Mole. Plant Micr. Inter.*. 9: 47-54.

Forster J & Knaak C (1995). Estimation of the genetic distance of 21 winter rapeseed varities by RAPD analysis in comparison to RFLP results. Proc. 9th Int. rape seed Congr. Cambridge, UK., 1184-1186.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P & Shenkin PS (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47:1739-49.

Fritsch PW & Rieseberg LH (1996). The use of random amplified polymorphic DNA (RAPD) in conservation genetics. *In* T. Smith and B. Wayne (eds.), *Mol. Genet. App. Cons.* pp. 54–73. Oxford Univ. Press, New York.

Fu L-G (1992). *Plant Red Book of China: Rare Threatened Plants*, Science Press, Beijing, pp. 184–185.

Gao LZ (2005). Microsatellite variation within and among population of *Oryza officinalis* (Poaceae), an endangered wild rice from China. *Mol. Ecol.* 14:4287-4297.

Gardiner SE, Forde MB & Slack CR (1986). Grass cultivar identification by sodium dodecylsulphate polyacrylamide gel elecgrophoresis. *Newz. J. Agri. Res.*. 29: 93-206.

Garth P & Milles PH (1991). *The toxicity of anticancer drugs*, Pergamon Press: New York.

Ge XJ & Sun M (1999). Reproductive biology and genetic diversity of a cryptoviviparous mangrove *Aegiceras corniculatum* (Myrsinaceae) using allozyme and intersimple sequence repeat (ISSR) analysis. *Mol. Ecol.* 8(12): 2061-2069.

Ge XJ & Sun M (2001). Population genetic structure of *Ceriops tagal* (Rhizophoraceae) in Thailand and China. *Wet. Ecol. Manag.* 9:203-209.

Geber MA, de Kroon H & Watson MA (1997). Organ preformation in mayapple as a mechanism for historical effects on demography. *J. Ecol.* 85:211–223.

Gimenes MA, Lopes CR, Galgaro ML, Valls JFM & Kochert G (2000). Genetic variation and phylogenetic relationships based on RAPD analysis in section *Caulorrhizae*, genus *Arachis* (Leguminosae). *Euphytica* 116:187-195.

Golbraikh A & Tropsha A (2002). Beware of q^2. *J. Mol. Graph. Model.* 20(4): 269-276.

Gonzalez G, Aleman S & Infante D (2003). Asexual genetic variability in *Agave fourcroydes* II: selection among individuals in clonally propagated population. *Plant. Sci.* 165:595–601.

Gordaliza M, Castro MA, García-Grávalos MD, Ruiz-Lázaro P, Miguel del Corral JM & San Feliciano A (1994). Synthesis and evaluation of pyrazolignans. A new class of cytotoxic agents, *Arch. Pharm. (Weinheim)* 327:175.

Gordaliza M, Castro MA, Miguel del Corral JM & San Feliciano A (2000). Antitumor properties of podophyllotoxin and related compound. *Curr. Pharm. Des.* 6:1811-1839.

Gordaliza M, Castro MA, San Feliciano A, Miguel del Corral JM, López-Vazquez ML & Faircloth GT (2001). Cytotoxic cyclolignans related to podophyllotoxin. *Patent EP* 711765 A1.

Gordaliza M, Faircloth GT, Castro M A, Miguel del Corral JM, López-Vázquez ML & San Feliciano A (1996). Immunosuppressive cyclolignans. *J. Med. Chem.* 39:2865.

Gordaliza M, Garcia PA, del Corral JM, Castro MA & Gomez Zurita MA (2004). Podophyllotoxin: distribution, sources, applications and newcytotoxic derivatives. *Toxicon* 44: 441–459.

Gordaliza M, Miguel del Corral JM, Castro MA, García-Grávalos MD, Broughton H & San Feliciano A (2001). Cytotoxic cyclolignans related to podophyllotoxin. *IL. Farmaco.* 56(4): 297-304.

Gordaliza M, Miguel del Corral JM, Castro MA, López-Vázquez ML, García PA, Feliciano, AS & García-Grávalos MD (1995). Selective cytotoxic cyclolignans. *Bioorg. Med. Chem. Lett.* 5: 2465-2468.

Gordaliza M, Miguel del Corral JM, Castro MA, López-Vázquez ML, San Feliciano A, García-Grávalos MD & Carpy A (1995). Synthesis and evaluation of pyrazolignans. A new class of cytotoxic agents. *Bioorg. Med. Chem.* 3:1203-1210.

Gregorio de C, Kier LB & Hall LH (1998). QSAR modeling with the electrotopological state indices: Corticosteroids, *J. Comput. Aid. Mol. Des.*12:557-561.

Gupta R & Sethi KL (1983). Conservation of medicinal plant resources in Himalayan region. In: Jain SK, Mehra KL (Eds.), *Conservation of Tropical Plant Resources. Botanical Survey of India.* Howrah, pp. 101-107.

Gupta S, Singh M & Madan AK (1999). Superpendentic index: a novel topological descriptor for prediction of biological activity, *J. Chem. Inf. Comput. Sci.*, 39: 272-277.

Haar ET, Rosenkranz HS, Hamel E & Day BW (1996). Computational and Molecular Modeling Evaluation of the Structural Basis for Tubuline Polymerization Inhibition by Colchincine Site Agents. *Bioorg. Med. Chem.* 10: 1659-1671.

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT & Banks JL **(**2004**).** Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening *J. Med. Chem.* 47:1750-1759.

Hamel E (1996). Antimitotic natural products and their interactions with tubulin. *Med. Res. Rev.* 16: 207-231.

Hamrick JL & Godt MJW (1996). Conservation genetics of endemic plant species. Pp. 281-304. In Avise, J. C. and J. L. Hamrick (eds). *Cons. Genet.*, Chapman and Hall, New York.

Han TH, Jeu MD, Eck HV & Jacobsen E (2000). Genetic diversity of Chilean and Brazilian *Alstroemeria* species assessed by AFLP anaslysis. *Heredity*. 84: 564-569.

Hangelbroek HH, Ouborg NJ, Santamaria L & Schwenk K (2002). Clonal diversity and structure within a population of pondweed *Potamogeton pectinatus* foraged by Bewick's Swans. *Mol. Ecol.* 11:2137-2150.

Hansch C, Kurup A, Garg R & Gao H (2001). Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms. *Chem. Rev.* 101: 619-672.

Hansson T & Aqvist J (1995). Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Prot. Eng***.** 8:1137-1145.

Hastie SB (1991). Interactions of colchicine with tubulin. *Pharmac. Ther.* 51: 377-401.

Hess J, Kadereit JW & Vargas P (2000). The colonization hystory of *Olea europea* L. in Macaronesia based on internal transcribed spacer 1(ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD), and inter-simple sequence repeats (ISSR). *Mol. Ecol*. 9: 857–868.

Hitotsuyanagi Y, Fukuyo M, Kyoko T, Kobayashi M, Ozeki A, Itokawa H & Takeya K (2000). 4-Aza-2, 3-dehydro-4-deoxypodophyllotoxin: Simple Aza-podophyllotoxin Analogues Possessing Potent Cytotoxicity. *Bioorg. Med. Chem. Lett.* 10: 315-317.

Hitotsuyanagi Y, Ichihara Y, Takeya K & Itokawa H (1994). Synthesis of 4-Oxa-2-azapodophyllotoxin, a novel analog of the antitumour lignin podophyllotoxin. *Tetrahedron Lett.* 35:9401-9402.

Hitotsuyanagi Y, Kobayeshi M, Takeya K & Itokowa H. (1995). Synthesis of 4-thia-2-azapodophyllotoxin, a new analogue of the antitumour lignan podophyllotoxin. *J. Chem. Soc. Perkin Trans.* 1:1387.

Hoever M & Zbinden P (2004). The evolution of microarrayed compound screening. *Drug Discov. Today* 9(8): 358-65.

Holm B, Sehested M & Jesen PB (1998). Improved targeting of brain tumors using dexrazoxane rescue of topoisomerase II combined with supra-lethal doses of etoposide and teniposide. *Clin. Cancer Res.* 4:1367–1373.

Horwtiz SB & Loike JD (1977). A comparison of the mechanism of action of VP 16-213 and podophyllotoxin. *Lloydia* 40: 82–89.

Howard D & Mark B (2000). *Neural Network Toolbox*. Version 4. The MathWorks, Inc. Natick, MA, USA.

Huang JC & Sun M (2000). Genetic diversity and relationships of sweet potato and its relatives in *Ipomoea* series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theo. App. Genet.* 100: 1050–1060.

Huzil JT, Chik JK, Slysz GW, Freedman H, Tuszynski J, Taylor RE, Sackett DL & Schriemer DC, (2008) A Unique Mode of Microtubule Stabilization Induced by Peloruside A. *J. Mol. Biol*. 378: 1016–1030.

Imbert TF (1998). Discovery of podophyllotoxins. *Biochimie* 80: 207-22.

Iruela M, Rubio J, Cubero JI, Gil J & Millan T (2002). Phylogenetic analysis in the genus *Cicer* and cultivated chickpea using RAPD and ISSR markers. *Theo. App. Genet*. 104: 643-651.

Issell BF, Muggia FM & Carter SK (1984). Etoposide (VP-16)- Current Status and New Developments. Academic Press, Orlando, FL, USA.

Jackson DE & Dewick PM (1984). Aryltetralin lignans from *Podophyllum hexandrum* and *Podophyllum peltatum* (isolated from the roots). *Phytochem.* 23: 1147-1152.

Jackson DE & Dewick PM (1984). Biosynthesis of *Podophyllum* lignans-II. Interconversion of aryltetralin lignan in *Podophyllum hexandrum*. *Phytochem.* 23: 1039-1042.

Jaiswal M, Khadikar PV, Scozzafava A & Supuran CT (2004). Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides, *Bioorg. Med. Chem. Lett.* 14: 3283-3290.

Janssen P, Coopman R, Huys G, Swings J, Bleeker M, Vos P, Zabeau M & Kersters K (1996). Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiol.* 142:1881-1893.

Jardine I (1980). In *Anticancer Agents based on Natural Product Models*, Cassady JM, Douras JD, Ed., Academic Press: New York, Charpter 9.

Jardine I (1980). Podophyllotoxins. In: Anticancer Agents Based on Natural Product Models. Academic, New York, pp. 319±351.

Jian SG, Tang T, Zhong Y & Shi SH (2004). Variation in inter-simple sequence repeat (ISSR) in mangrove and non-mangrove population of *Heritiera littoralis* (Sterculiaceae) from China and Australia. *Aquat. Bot.* 79: 75-86.

Jiang CS, Jia HS, Ma XR, Zou DM & Zhang YZ (2004). AFLP analysis of genetic variability among *Stylosanthes guianensis* accessions resistant and susceptible to the Stylo Anthracnose. *Acta Bot. Sin.* 46:480–488.

Jordan A, Hadfield JA, Lawrence NJ & McGown AT (1998). Tubuline as a target of anticancer drugs: agents which interact with the mitotic spindle. *Med. Res. Rev.* 18: 259-296.

Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK & Brar DS (2000). Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theo. App. Genet.* 100: 1311–1320.

Kadow JF, Vyas DM & Doyle DM (1989). Synthesis of etoposide Lactam via Mitsunobu reaction sequence. *Tetrahedron Lett*., 30: 3299-3302.

Kafkas S & Perl-Treves R (2002). Inter-specific relationships in the genus *Pistacia* L. (Anacardiaceae) based on RAPD fingerprints. *Hort. Sci.* 37: 168-171.

Kantety RV, Zeng XP, Bennetzen JL & Zehr BE (1995). Assessment of genetic diversity in Dent and Popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. *Mol. Breed.* 1: 365–373.

Kaplan IW (1942). *Codylomata acuminata*. *New Orleans Med. Surg. J.* 94: 388.

Karp A, Kresovich S, Bhat KV, Ayad WG & Hodgkin T (1997). Molecular tools in plant genetic resources conservation: a guide to the technologies. *In: IPGRI Tech. Bull.* 2: 1-47.

Kazan K, Manners JM & Cameron DF (1993). Genetic variation in agronomically important species of *Stylosanthes* determine using random amplified polymorphic DNA markers. *Theo. Appl. Genet.* 85: 882-288.

Keiper FJ & McConchie R (2000). An analysis of genetic variation in natural populations of Sticherus flabellatus [R. Br. (St John)] using amplified fragment length polymorphism (AFLP) markers. *Mol. Ecol.* 9(5): 571-81.

Keller-Juslen C, Kuhn M, von Wartburg A & Stahelin H (1971). Synthesis and antimitotic activity of glycosidic lignan derivatives related to podophyllotoxin. *J. Med. Chem.* 14: 936.

Kimble JM, Knox EG & Holzhey CS (1993). Soil survey laboratory methods for characterizing physical and chemical properties and mineralogy of soils. *In Application of Agriculture Analysis in Environmental Studies, eds.* K. B. Hoddinott and A. O. O'Shay, 23-31. American Society for Testing and Materials, Philadelphia.

King JL, Sullivan M (1946) The similarity of the effects of podophyllin and colchicines and their use in the treatment of *Codylomata acuminata*. *Science* 104: 244-245.

Kitamura R, Bandoh T, Tsuda M & Satoh T (1997). Determination of a new podophyllotoxin derivative, TOP-53, and its metabolite in rat plasma and urine by high performance liquid chromatophy which electrochemical detection. *J. Chromat. B: Biomed. Appl.* 690: 283.

Klebe G, Abraham U (1994). Mietzner T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity. *J. Med. Chem.* 37: 4130-4146.

Knobloch KH & Berlin J (1983). Influence of phosphate on the formation of Indole alkloids and phenolic compounds in cell suspension cultures of *Catharanthus roseus*. I. Comparision of enzyme activities and product accumulation. *Plant Cell Tissue Org. Cult.* 2:333-341.

Koller B, Lehmann A, McDermott JM & Gessler C (1993). Identification of apple cultivars using RAPD markers. *Theo. Appl. Genet.* 85: 901-904.

Kresovich S, Williams JGK, McFerson JR, Routman EJ & Schaal BA (1992). Characterisation of genetic identities and relationship of *Brassica oleracea* L. via a random amplified polymorphic DNA assay. *Theo. Appl. Genet.* 85: 190-196.

Kumar R, Singh PK, Arora R, Sharma A & Prasad J (2005). Radioprotection by *Podophyllum hexandrum* in the liver of mice: A mechanistic approach. *Env. Toxicol. and Pharmacol.* 20: 326–334.

Laatsch H, Ernst BP & Noltemeyer M (1996). Synthesis of Sterically Fixed Podophyllotoxin. *Liebigs Ann.* 731-737.

Lakshmi M, Railakshmi S, Parani M, Anuratha CS & Parida A (1997). Use of molecular markers in assessing interspecific genetic variability in the mangrove species *Acanthus illicifolius* Linn. (Acanthaceae). *Theo. Appl. Genet.* 94: 1121-1127.

Lamote V, Rold´an-Ruiz I, Coart E, De Loose M & Van Bockstaele E (2002). A study of genetic variation in *Iris pseudacorus* populations using amplified fragment length polymorphisms (AFLPs). *Aqu. Bot.* 73: 19–31.

Landa K, Benner B, Watson MA & Gartner J (1992). Physiological integration for carbon in mayapple (*Podophyllum peltatum*), a clonal perennial herb. OIKOS 63:348–356.

Lavastre O, Bonnette F & Gallard L (2004). Parallel and combinatorial approaches for synthesis of ligands. *Curr. Op. Chem. Biol.* 8: 311-318.

Laverty TM & Plowright RC (1988). Fruit and seed set in mayapple (*Podophyllum peltatum*): Influence of intraspecific factors and local enhancement near *Pedicularis canadensis*. *Can. J. Bot.* 66: 173–178.

Leander K & Rosen B (1988). Medicinal uses for podophyllotoxin. *U.S patent* 4: 788, 216.

Leiros HKS, Brandsdal BO, Andersen OA Os V, Leiros I, Helland R, Otlewski J, Willassen NP, & Smalas AO (2004). Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. *Protein Sci.* 13: 1056-1070.

Lerndal T & Svensson B (2000). A clinical study of CPH 82 vs. methotrexate in early rheumatoid arthritis. *Rheumatology* (Oxford) 39: 316.

Levy RK, Hall IH & Lee KH (1983). In Anticancer Agents Based on Natural Products Models; Cassady. *J. Pharm. Sci*. *72*:1158.

Lewontin RC (1973). Population Genetics. *Annu. Rev. Genet.*.7:1-17.

Li DR & Huang ZYU (1989). *Manual of Practical Soil Fertilizers*. China Agriculture Technology Press, Beijing. pp. 28-43.

Li F & Xia N (2005). Population structure and genetic diversity of an endangered species, *Glyptostrobus pensilis* (Cupressaceae). *Bot. Bull. Sinica* 46: 155-162.

Li GM (1975). Introduction to a medicine plant, *Sinopodophyllum emodi* Wall. var. chinense Sprague. *J. Bot.* 2: 28.

Liaison, version 4.0, Schrödinger, LLC, New York, NY, 2005

Lin CM & Hamel E (1981). Effects of inhibitors of tubulin polymerization on GTP hydrolysis. *J. Biol. Chem.* 256: 9242.

Lionard P, Quron JC & Husson HP (1993). Asymmetric Synthesis. XXVIII. Hydroxylated Benzoquinolizidine Analogues of Podophyllotoxin via the CN(R,S) Method. *Tetrahedron* 49: 3995-4006.

Lipinski A, Lombardo F, Dominy B & Feeney P (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Reviews* 46(1-3): 3-26.

Liu SS, Cai CX & Li Z (1998). Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance edge (MDE) vector λ, *J. Chem. Inf. Comput. Sci.* 38: 387-394.

Livingstone DJ (2000). The Characterization of Chemical Structures Using Molecular Properties. A Survey, *J. Chem. Inf. Comput. Sci.* 40: 195-209.

Loehrer PJ. Sr (1991). Etoposide therapy for testicular cancer. *Cancer* 67: 220.

Loh JP, Kiew R, Set O, Gan LH & Gan YY (2000). Study of genetic variation and relationships within the bamboo subtribe Bambusinae using amplified fragment length polymorphism. *Ann. Bot.* 85: 607–612.

Loike JD & Horwitz SB (1976a). Effects of podophyllotoxin and VP16-213 on microtubule assembly in vitro and nucleoside transport in HeLa cells, *Biochem.* 15: 5435-5442.

Loike JD, Brewer CF, Sternlicht H, Gensler WJ & Horwitz SB (1978). Structure-activity study of the inhibition of microtubules assembly in vitro by podophyllotoxin and its congeners. *Cancer Res.* 38: 2688.

Lu JJ, Knox MR, Ambrose MJ, Brown JKM & Ellis THN (1996). Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR based methods. *Theo. App. Genet.* 93: 1103-1111.

Lyne PD (2002). Structure-based virtual screening: An overview. *Drug Disc. Today* 7: 1047-1055.

Lyne PD, Lamb ML & Saeh JC (2006). *Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring*. *J. Med. Chem.* 49: 4805-4808.

Mantel NA (1967). The detection of disease clustering and generalized regression approach. *Cancer Res.* 27: 209-220.

Marder VJ(2001). Thrombolytic therapy. *Blood Rev*. 15: 143-157.

Margolis RL & Wilson L (1978). Opposite end assembly and disassembly of microtubules at steady state in vitro. *Cell* 13: 1.

Marty M, Fumoleau P, Adenis A, Rousseau Y, Merrouche Y, Robinet G, Senac I & Puozzo C (2001). Oral vinorelbine pharmacokinetics and absolute bioavailability study in patients with solid tumors. *Ann. Oncol.* 12(11): 1643–1649.

Maughan PJ, Maroof MAS, Buss GR & Huesti GM (1996). Amplified fragment length polymorphism (AFLP) in soyabeen: species diversity, inheritance, and near-isogenic line analysis, *Theo. App. Genet.* 93: 392-401.

Mauria S, Singh NN, Mukherjee AK & Bhat KV (2000). Isozyme characterization of Indian maize inbreds. *Euphytica*. 112: 253-259.

Mebius LJ (1960). A rapid method for the determination of organic carbon in soil. *Anal. Chem. Acta* 22: 120-124.

Mehlich A (1976). New buffer pH method for rapid estimation of exchangeable acidity and lime requirement of soils. *Comm. Soil Sci. Plant Anal.*7: 637-652.

Meneses-Marcel Y, Marrero- Ponce Y, Machado-Tugores A, Monterro-Torres D M, Pereira JA, Escario JJ, Nogel-Ruiz C, Ochao VJ, Aran AR, Martinez-Fernadez RN & Garcia Sanchez (2005). A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: Outcomes of in silico studies supported by experimental results, *Bioorg. Med. Chem. Lett.* 15: 3838-3843.

Metais I, Aubry C, Hamon B & Jalouzot R (2000). Description and analysis of genetic diversity between commercial bean lines (*Phaseolus vulgaris* L.). *Theo. App. Genet.* 101: 1207–1214.

Meylan WM & Howard PH (1995). Atom/fragment contribution method for estimating octanolwater partition coefficients. *J. Pharm. Sci.* 84(1): 83-92.

Miguel del Corral JM, Gordaliza M, Castro MA, García-Grávalos MD, Broughton H & San Feliciano A (1997). Bioactive Isoxazoline and Oxime Derivatives from. 7-Ketolignans.*Tetrahedron* 53: 6555-6564.

Miguel del Corral JM, Gordaliza M, Castro MA, Morales LJ, López JL & San Feliciano A (1995). Methyl ethers of podophyllotoxin-related cyclolignans. *J. Nat. Prod.* 58: 870.

Miller MP (1998). AMOVA-PREP 1.01: A Program for the Preparation of the AMOVA Input Files from Dominant-Marker Raw Data. Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ.

Minocha A & Long BH (1984). Inhibition of the DNA catenation activity of type II topoisomerase by VP 16-213, VM – 26. *Biochem. Biophys. Res. Commun.* 122:165–170.

Moraes-Cerdeira RM, Burandt CL Jr, Bastos JK, Nanayakkara NPD & McChesney JD (1998). In vitro propagation of *Podophyllum peltatum*. *Pl. Medica* 64:42–46.

Morgante M & Olivieri AM (1993). PCR-amplified microsatelites as markers in plant genetics. *Plant J.* 3:175-182.

Mross K, Huttmann A, Herbst K, Hanauske AR, Schilling T, Manegald C, Burk K, & Hossfield DK (1996). Pharmacokinetics and pharmacodynamics of the new podophyllotoxin derivative NK 611. A study by the AIO groups PHASE-I and APOH. *Cancer Chemother. Pharmacol.* 38:217-224.

Mulcahy DL, Cresti M, Linskens HF, Intrieri C, Silvestroni O, Vignani R & Pancaldi M (1995). DNA fingerprinting of Italian grape varieties: a test of reliability in RAPDs. Adv. *Hort. Sci.* 9: 185–187.

Myburg AA, O'Malley D, Sederoff RR & Whetten R (2000). Highthroughput multiplexed AFLP analysis of interspecific hybrids of *Eucalyptus* trees species. Plant & Animal Genome VIII conference, San Diego, CA, p 544.

Nagaoka T & Ogihara Y (1997). Applicability of inter-simple sequence repeat polymorphism in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theo. Appl. Genet.* 94: 597-602.

Nayak S, Naik PK, Acharya L, Mukhaerjee AK, Panda PC & Das P (2005). Assesment of Genetic Diversity among 16 promising cultivars of Ginger using cytological and molecular markers. *Science Asia.* 60: 485-492.

Nayar MP & Sastry APK (1990). A simple microanalytical technique for determination of podophyllotoxin in *Podophyllum hexandrum* roots by quantitative RP–HPLC and RP–HPTLC. *Data Book of Indian Plants,* Botanical Survey of India, Kolkata. P. 272-273.

Nei M (1973). Analyses of gene diversity in subdivided populations, *Proc. Natl. Acad. Sci. USA* 70: 3321-3323.

Nei M (1978). Estimation of average heterozygodity and gentic distance from a small number of individuals. *Genetics* 89: 83-590.

Oloff S, Mailman RB & Trospha A (2005). Application of Validated QSAR Models of D_1 Dopaminergic Antagonists for Database Mining, *J. Med. Chem*. 48: 7322-7332.

Olsen SR & Sommers LE (1982). Phosphorus. In *Methods of soil analysis, eds.* A. L. Page, R. H. Millerand and D. R. Keeney, Part 2, vol.9, second ed. pp. 403-429.

Oprea T & Matter H (2004). Integrating virtual screening in lead discovery. *Curr. Op. Chem. Biol.* 8:349-358.

Ostrovsky D, Udier-Blagovic M & Jorgensen WL (2003). Analyses of activity for factor Xa inhibitors based on Monte Carlo simulations. *J. Med. Chem.* 46: 5691-5699.

Pachepsky Ya A, Timlin D & Varallyay G (1996). Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Sci. Soc. of American J.* 60: 727-733.

Padmesh P, Sabu KK, Seeni & Pushpaangadan S (1999). The use of RAPD in assessing genetic variability in Andrographis paniculata Nees, a hepatoprorective drug, *Curr. Sci.* 76: 833-835.

Pagani O, Zucchetti M, Sessa C, De Jong J, D' Incalci M, De Fusco M, Kaeser-Froehlich A, Hanauske A & Cavalla F (1996). Clinical and Pharmacokinetic Study of Oral NK611, a New Podophyllotoxin Derivatives. *Cancer Chemother. Pharmacol.* 38: 541-547.

Palacious C, Kresovich S & GonzalezCandelas F (1999). A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). *Mol. Ecol.* 8: 645-657.

Parentoni SN, Magalhaes JV, Pacheco CAP, Santos MX, Abadie T, Gama EEG, Guimaraes PEO, Meirelles WF, Lopes MA, Vasconcelos MJV & Paiva E (2001). Heterotic groups based on yield-specific combining ability data and phylogenetic relationship determined by RAPD markers for 28 tropical maize open pollinated varieties. *Euphytica*121: 197-208.

Pasha FA, Chung HW, Cho SJ & Kang SB (2008). 3D-quantitative structure activity analysis and quantum chemical analysis of pyrido-di-indoles," *Inter. J. Quan. Chem.* 108(2): 391–400.

Paul S, Wachira FN, Powell W & Waugh R (1997). Diversity and genetic differentiation among populations of India and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. *Theo. Appl. Genet.* 94: 255-263.

Pearce H, Bach N, Cramer T, Danks M, Grindey G, Katterjohn C, Rinzel S & Beck W (1990). Synthesis and cellular pharmacology of ±4'-demethyl- l-O-[4,6-O-(ethylidene)-beta-D-glucopyranosyl]-2'-azapodophyllotoxin. *Proc. Am. Assoc. Cancer Res.* 31: 441.

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G & Motto M (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theo. Appl. Genet.* 97: 1248–1255.

Podlipnika C & Bernardib A (2007). Design of a focused virtual library to explore cholera toxin B-site. *Acta Chim. Slov.* 07 (54): 425–436.

Polak E, & Ribiere G (1969). *Revue Francaise Inf. Rech. Oper., Serie Rouge. 16-R1,* 35-43.
Polanco C & Ruiz ML (2002). AFLP analysis of somaclonal variation in *Arabidopsis thaliana* regenerated plants. *Plant Sci.* 162: 817–824.

Policansky D (1983). Patches, clones, and self-fertility of mayapples (*Podophyllum peltatum* L.). *Rhodora* 85: 253–256.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S & Rafalski A (1996). The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225–238.

Prabhu RR & Gresshoff PM (1994). Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. *Plant Mol. Biol.* 26: 105–116.

Prevost A & Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theo. Appl. Genet.* 98: 107-112.

Puecher DI, Ibanez MA & Di Renzo MA (1996). Classification and diversity values of seventeen cultivars of *Eragrostis curvula*. *Seed Sci. Tech.* 24: 139-149.

Pugh N, Khan I, Moraes RM & Pasco D (2001). Podophyllotoxin lignans enhace Il- 1b but suppress TNF-a mRNA expression in LPS-treated monocytes. *Immunopharmacol. Immunotoxicol.* 23: 83–95.

Purohit AN, Lata H, Nautiyal S & Purohit MC (1998). Some characteristics of four morphological variants of *Podophyllum hexandrum* Royle, *Plant Genet. Resour. Newsl,* 114: 51-52.

Purohit MC, Bahuguna R, Maithani UC, Purohit AN & Rawat MSM (1999). Variation in podophylloresin and podophyllotoxin contents in different population of *Podophyllum hexandrum*. *Curr. Sci.* 77: 1078-1080.

Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP & Devarumath RM (2001). RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification and phylogenetic relationships in peanut (*Archis hypogea*) cultivars and wild species. *Genome.* 44: 763-772.

Ratnaparkhe MB, Gupta VS, Murthy MRV & Ranjekar PK (1995). Genetic fingerprinting of pigeon pea (*Cajanus cajan* (L.) Millsp.) and its wild relatives using RAPD markers. *Theo. Appl. Genet.* 91: 893-898.

Ratnaparkhe MB, Tekeoglu M, Muehlbauer FJ (1998). Intersimple-sequence-repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene cluster. *Theo. Appl. Genet.* 97: 515-519.

Ravelli RBG, Gigant B, Curmi PA, Jourdain I, Lachkar S, Sobel A, & Knossow M (2004). Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. (2004). *Nature* 428: 198–202.

Ray Choudhury P, Kohli S, Srinivasan K, Mohapatra T & Sharma R P (2001). Identification and classification of aromatic rices based on DNA fingerprinting. *Euphytica*.118: 243-251.

Reddy KD & Nagaraju A (1999). Genetic characterization of the silkworm *Bombyx mori* by simple sequence repeat (SSR)- anchored PCR. *Heredity* 83: 681-687.

Rieseberg LH (1996). Homology among RAPD fragments in interspecific comparisons. *Mol. Ecol.* 5: 99-105.

Rohlf FJ (1992). NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System, Version 2.0. State University of New York, Stony Brook., NY.

Roy PP & Roy K (2008). On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* 27: 302-313.

Rump HH & Krist H (1992). *Laboratory manual for the examination of water, Waste Water and Soil.* VCH, Weinheim. second ed. pp. 11-113.

Saghai-Maroof MA, Soliman KM, Jorgensen RA & Allard RW (1984). Ribosomal spacer length in barley: Mendelian inheritance, Chromosomal location and population dynamics. *Proc. Natl. Acad. Sci. (USA)* 81: 8104-8118.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB & Erlich HA (1988). Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* 239: 487–491.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA & Arnheim N (1985). Enzymatic amplification of *β*-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350–1354.

Sale MM, Potts BM, West AK & Reid JB (1996). Molecular differentiation within and between *Eucalyptus risdonii, E. amygdalina* and their hybrids using RAPD markers. *Aust. J. Bot.* 44: 559-569.

Salimath SS, De Oliveira AC & Godwin ID (1995). Assesment of genomic origins and genetic diversity in the genus Eleusin with DNA markers. *Genome* 38: 757-763.

Sambrook J, Fritsch EF & Maniatis T (1989). Molecular Cloning. Cold Spring Harbor Laboratory Press, New York.

San Feliciano A, Caballero Reneta, Del Rey B & Sancho I (1991). Diterpene acids from *Juniperus communis* var. *hemisphaerica. Phytochem.* 30: 3134–3136.

San Feliciano A, Gordaliza M, Miguel del Corral JM, Castro MA, Garcia- Gravalos MD & Ruiz Lazaro P. (1993). Antineoplastic and Antiviral Activities of Some Cyclolignans. *Planta Med.*, 59: 246-249.

San Feliciano A, Miguel del Corral JM, Gordaliza M & Castro A (1990). Lignan from *Juniperous Sabina. Phytochem.* 29: 1335-1338.

San Feliciano A, Miguel del Corral JM, Gordaliza M & Castro MA (1989). Acetylated lignans form *Juniperus sabina. Phytochem.* 28: 659-660.

Santana L, Uriarte H, Gonzalez-Diaz H, Zagotto R, Soto-Otero E & Mendez-Alvarez J, (2006). A QSAR Model for in Silico Screening of MAO-A Inhibitors. Prediction, Synthesis, and Biological Assay of Novel Coumarins. *J. Med. Chem.* 49:1149-1156.

Schaal BA, Leverich WJ & Rogstad SH (1991). Comparison of methods for assessing genetic variation in plant conservation biology. In: Falk, D.A., Holsinger, K.E. (Eds.), Genetics and Conservation of Rare Plants. Oxford University Press, New York pp, 123-134.

Schaal BA, O'Kane-Jr SL & Rogstad SH (1991). DNA variation in plant populations. *Trend. Eco. Evol.* 6: 329-333.

Schacter L (1996). Etoposide phosphate: what, why, where, and how? *Seminars in Oncology*. 23: 1-7.

Schilstra JM, Martin RS & Bayley MP (1989). The effect of Podophyllotoxin on Microtubule Dynamics. *J. Bio. Chem.* 264: 8827-8834.

Schrodinger LLC. , http://www.schrodinger.com, (accessed: 24. 04.2007).

Selassie CD, Mekapati SB & Verma RP (2002). QSAR: Then and Now *Curr. Top. Med. Chem.* 23: 1357-1379.

Sengupta SK (1995). *Cancer Chemotherapeutic Agents,* ed. by Foye WO De, American Chemical Society: Washington DC, Charper 5. p. 205-217.

Shapiro S & Guggenheim B (1998). Inhibition of Oral Bacteria by Phenolic Compounds. Part 1. QSAR Analysis using Molecular Connectivity. *Quant. Struct-Act. Relat*. 17: 327-337.

Sharma KD, Singh BM, Sharma TR, Katoch M & Guleria S (2000b). Molecular analysis of variability in *Podophyllum hexandrum* Royle- an endangered medicinal herb of northwestern Himalaya. *Plant Genet. Reso. Newsl.* 124: 57-61.

Sharma SK, Knox MR and Ellis, THN (1996). AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPDanalysis. *Theo. Appl. Genet.* 93(5-6): 751-758.

Sharma TR, Singh BM, Sharma NR & Chauhan RS (2000a). Identification of high podophyllotoxin producing biotypes of Podophyllum hexandrum form north-western Himalaya, *J. Plan. Biochem. & Biotech.* 9: 49-51.

Shi LM, Fan Y, Myers TG & Paul JN (1998). Mining the NCI Anticancer Drug Discovery Databases: Genetic Function Approximation for the QSAR Study of Anticancer Ellipticine Analogues. *J. Chem. Inf. Comput. Sci*. 38: 189-199.

Shoichet B, McGovern S, Wei B & Irwin J (2002). Lead discovery using molecular docking. *Curr. Op. Chem. Biol.* 6: 439-446.

Slatin M (1987). Gene flow and geographic structure of natural populations. *Science (Washington, DC)* 236: 787-792.

Slatkin M & Barton NH (1989). A comparison of three indirect methods for estimating the average level of gene flow. *Evolution* 43: 1349–1368.

Sneath PHA & Sokal K (1973). Numerical Taxonomy. pp. 100-308.

Snyder JA & McIntosh RJ (1976). Biochemistry and physiology of microtubules. *Ann. Rev. Biochem*. 45: 699.

Soil and Plant Analysis Council (1992). *Handbook on reference methods for soil analysis,* Athens, GA, USA.

Stahelin HF & Wartburg AV (1991). The chemical and biological route form podophyllotoxin glucoside to etoposide: Ninth Cain Memorail Award Lecture. *Cancer. Res.* 51: 5-15.

Stähelin, HF (1973) Activity of a new glycosidic lignan derivative (VP 16-213) related to podophyllotoxin in experimental tumors. *Eur. J. Cancer*. 9: 215-221.

Stebbins GL (1999). A brief summary of my ideas on evolution. *Am. J. Bot.* 86:1207-1208.

Stewart, CN & Porter DM (1995). RAPD profiling in biological conservation: an application to estimating clonal variation in rare and endangered *Iliamna* in Virginia. *Bio.l Cons.* 74: 135-142.

Sudduth K, Fraisse C, Drummond S & Kitchen N (1998). Integrating spatial data collection, modeling and analysis for precision agriculture, *First Int. Conf. on Geospatial Information in Agriculture and Forstery*, Lake Buena Vista, Florida. 1-3, *Juin.* 2: 66-173.

Sudduth K, Fraisse C, Drummond S & Kitchen N (1996). Analysis of spatial factors influencing crop yield, Procedings of 3. *Int. Conf. on Precision Agriculture*. 129-140. Madison, WI.

Sun L, McPhail AT, Hamel E, Lin CM, Hastie SB, Chang JJ & Lee KH (1993). Antitumor agents. 139. Synthesis and biological evaluation of thiocolchicine analogs 5,6-dihydro-6(S)-(acyloxy)- and 5,6-dihydro-6(S)-[(aroyloxy)methyl]-1,2,3-trimethoxy-9-(methylthio)-8H- cyclohepta[a]naphthalen-8-ones as novel cytotoxic and antimitotic agents. *J.Med. Chem.* 36: 544.

Suyama Y, Obayashi K & Hayashi I (2000). Clonal structure in a dwarf bamboo (Sasa senanensis) population inferred from amplified fragment length polymorphism (AFLP) fingerprints. *Mol. Ecol.* 9(7): 901-6.

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nuc. Acids Res.* 17: 6463-6471.

Thomas GB & Finny RL (2001). In Calculus and Analytic Geometry, 9th Edn., Addison-Wesley.

Thurston LS, Irie H, Tani S, Han FS, Liu ZC, Cheng YC & Lee KH (1986). Antitumour agents. 78. Inhibition of human DNA topoisomerase II by podophyllotoxin and alpha-peltatin analogs. *J. Med. Chem.,* 29: 1547- 1550.

Tian X, Yang MG, Chen YZ (1996). Anticancer Drugs (VI)-Synthesis and activity of Spin Labeled Derivatives of 4-Amino-4-deoxy-4'- demethylepipodophyllotoxin. *Chem. Res. Chin. Univ*. 12: 304-308.

Titmuss SJ, Keller PA & Griffith R (1999). Docking experiments in the flexible nonnucleoside inhibitor binding pocket of HIV-1 resease transcriptase. *Bioorg. Med. Chem.* 7: 1163-1170.

Tominaga Y & Jorgensen WL (2004). General model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* 47: 2534-2549.

Tomioka K, Kubota Y & Koga K (1993). Design, synthesis, and antitumour activity-absolute configuration relationships of podophyllotoxin aza- analogues. *Tetrahedron*, 49: 1891-1900.

Travis SE, Maschinski J & Keim P (1996). Analysis of genetic variation in *Astragalus cremnophylax var. cremophylax,* a critically endangered plant, using AFLP markers. *Mol. Ecol.* 5: 735-745.

Tuppurainen K (1999). Frontier orbital energies, hydrophobicity and steric factors as physical qsar descriptors of molecular mutagenicity. A review with a case study: MX compounds, *Chemosphere.* 38: 3015-3030.

Tyler VE, Brady LR & Robbers JE (1988). Lea & Febiger, Philadelphia. *Pharmacology 9th edn.*

Utsugi T, Shibata J, Sugimoto Y, Aoyagi K, Wierzba K, Kobunai T, Terada T, Ohhara T, Tsuruo T & Yamada Y (1996). Antitumor activity of a novel podophyllotoxin derivative (TOP-53) against lung cancer and lung metastatic cancer. *Cancer Res.* 56: 2809–2814.

Van der Eycken J, Bosmans JP, Van Haver D, Vandewalle M, Hulkenberg A, Veerman W & Nieuwenhuizen R (1989). The synthesis of 4- desoxy-2-azapodophyllotoxins. *Tetrahedron Lett.,* 30: 3873-3876.

Van Uden W, Pras N, Visser JF & Malingre Th. M (1989). Detection and identification of podophyllotoxin produced by cell cultures derived from *Podophyllum hexandrum* Royle. *Plant Cell Rep*. 8: 165-168.

VanToai TT, Peng J & St. Martin SK (1997). Using AFLP markers to determine the genomic contribution of parents to populations. *Crop Sci.* 37: 1370–1373.

Vasconcelos MJV, de Barros EG, Moreira MA & Vieira C (1996). Genetic diversity of the common bean *Phaseolus vulgaris* L. determined by DNA-based molecular markers. *Braz. J. Genet.* 19: 447-451.

Visser GW, Bijma AT, Dijksman JA, Herscheid JD & Int J (1989). Radiation Applied and Instr. Part A, *Appl. Rad. Isotop.* 40: 47.

Vos PR, Hogers M, Bleeker M, van de Lee Reijans T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M & Zabeau M (1995). AFLP: a new concept for DNA fingerprinting. *Nuc. Acids Res*. 23: 4407–4414.

Vosman B, Visser D, van der Voort JR, Smulders MJM & van Eeuwijk F (2004). The establishment of 'essential derivative' among rose varieties using AFLP. *Theor. Appl. Genet.* 109: 1718–1725.

Wang JZ, Tian X & Tsumura H (1993). Antitumour activity of a new low immunosuppressive derivative of podophyllotoxin (GP-11) and its mechanisms. *Anti-Cancer Drug Des.* 8:193-202.

Weiss SG, Tin-Wa M, Perdue RE Jr, & Farnsworth NR (1975). Potential anticancer agents II: antitumor and cytotoxic lignans from Linum album (Linaceae). *J. Pharm. Sci*. 64: 95.

Welsh J & McClelland M (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18: 17213-17218.

Welsh J, Honeycutt RJ, McClelland M. & Sobral BWS (1991). Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). *Theo. App. Genet.*. 82: 473-476.

Williams CE & St. Clair DA (1993). Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicum esculentum*. *Genome* 36: 619-830.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA & Tingey SV (1990). DNA Polymorphisms amplified by arbitrary primers and useful as genetic markers. *Nucl. Acids Res.* 18: 6531-6535.

Winfield MO, Arnold GM & Cooper F (1998). A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the Upper Severn Area of the UK using AFLP markers. *Mol. Ecol.* 7: 3-10.

Witterland AH, Koks CH & Beijnen JH (1996). Etoposide phosphate, the water soluble prodrug of etoposide. *Phar. World Sci.* 18: 163.

Wolfe AD, Xiang QY & Kephart SR (1998). Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Mol. Ecol* 7: 1107-1126.

Wolfe AD, Xiang QY & Kephart SR (1998). Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proc. Natl. Acad. Sci.* USA 95: 5112–5115.

Wright S (1931). Evolution in Mendelian populations. *Genetics*. 16: 97–159.

Xi ZB (1994). Chemical Fertilizers. pp 24-321. Beijing: Science Press.

Yamashita Y, Kawada SZ, Fujii N, & Nakano H (1991). Induction of mam malian topoisomerase I and II mediated DNA cleavage by saintopin, a new antitumor agent from fungus. *Biochem.* 30: 5838-5845.

Yang W, de Oliveria AC, Godwin I, Schertz KF & Bennetzen JL (1996). Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Sci.* 36:1669-1676.

Ying TS (1979). On *Dysosma*Woodson and *Sinopodophyllum* Ying, gen. nov. of the Berberidaceae. *Acta Phytoxon. Sin.* 17: 15–23.

Zabeau M & Vos P (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. Publication 0 534 858 A1, bulletin 93/13. *European Patent Office*, Munich, Germany.

Zabeau M (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. *Euro. Patent Appl. No.* 0534858 A1.

Zamora R, Jamilena M, Ruiz Rejon M & Blanca G (1996). Two new species of the carnivorous genus *Pinguicula,* (Lentibulariaceae) from Mediterranean habitats. *Plant Sys. Evol.* 200: 41-60.

Zhou RH, Friesner RA, Ghosh A, Rizzo RC, Jorgensen WL & Levy RM (2001). New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B.* 105: 10388-10397.

Zietkiewicz E, Rafalski A & Labuda D (1994).Genome fingerprinting by simple sequence repeat (SSR)- anchored polymerase chain reaction amplification. *Genomics* 20:176-183.

List of Publications

1. Alam MA, Naik PK (2009). Molecular modelling evaluation of the cytotoxic activity of podophyllotoxin analogues. *J Comput Aided Mol Des.* 23: 209-225.

2. Alam MA, Naik PK (2009). Applying Linear Interaction energy method for Binding affinity calculations of podophyllotoxin analogues with Tubulin using continuum solvent model and prediction of cytotoxic activity. *J. Mol. Graph. Model* doi:10.1016/j.jmgm.2009.02.003 (Article in press).

3. Alam MA, Mishra GP, Naik PK (2009). Congruence of RAPD and ISSR markers for evaluation of genomic relationship among 28 populations of *Podophyllum hexandrum* from Himachal Pradesh. *Turk. J. Bot*. 33 (2009) 1-12.

4. lam MA, Naik PK (2008). Impact of soil nutrients and environmental factors on podophyllotoxin content among 28 *Podophyllum hexandrum* Populations of Northwestern Himalayan Region using linear and non-linear approach. *Comm. Soil Sci. Plant Anal.* (Article in press).

5. Alam MA, Pallavi G, Gulati AK, Mishra GP, Naik PK (2008). Characterization of genetic structure of *Podophyllum hexandrum* populations–an endangered medicinal herb of Northwestern Himalaya using ISSR-PCR markers and its relatedness with podophyllotoxin content. *African journal of Biotechnology.* 7(8): 1028-1040.

6. Alam MA, Pallavi G, Gulati AK, Mishra GP, Naik PK (2008). Assessment of Genetic Diversity among 28 *Podophyllum hexandrum* populations of Northwestern Himalayan Region using RAPD Markers. *Indian Journal of Biotechnology* (Article in Press).