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

Synopsis

of

Ph. D Thesis

Submitted by

Md. Afroz Alam

Enrollment Number: 051756

Jaypee University of Information Technology
Waknaghat, P.O. Dumehar Bani, Kandaghat, Solan-173215, H.P.

April 2009

Introduction

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 [1]. The rhizome yields a very valuable drug 'podophyllotoxin' which has a great demand in pharmaceutical industries throughout the world [2]. In addition, the drug obtained from Himalayan mayapple contains resin of superior quality in comparison to American mayapple [3]. These compounds have been used for the treatment of lung and testicular cancers as well as certain types of leukemia [4]. Besides, podophyllotoxin is also the precursor to semi synthetic chemotherapeutic drugs such as etoposide, teniposide and etophos [4]. The Indian species *P. hexandrum* contains three times more podophyllotoxin than its American counterpart *P. peltatum* [5]. So *Podophyllum hexandrum* is being uprooted unscientifically for commercial trade. As a result of this and a lack of organized cultivation, *P. hexandrum* has been reported as a threatened species in the Himalayan region [6]. No major steps have been taken for commercial cultivation of this very useful plant. 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. Moreover podophyllotoxin content varies greatly between populations of *P. hexandrum* from the northwestern Himalayas [7]. Therefore, it would be inefficient, environmentally destructive and economically unsound to randomly harvest May apple from the wild. 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. Understandings of its population biology and genetic diversity as well as its relationship with podophyllotoxin content are hence essential. DNA-based molecular-marker techniques such as RAPD [8] ISSR [9] and AFLP [10] have been proved powerful in genetic diversity estimations [11]. However, the impact of environmental factors with the yield of podophyllotoxin cannot be ruled out as the plant is very adaptable to a wide range of environmental conditions. A number of investigations have demonstrated that the quality and quantity of several secondary metabolites have a close relationship with plant habitats [12]. Hence, it demands management of soil nutrients and optimization of climatic factors for the successful conservation of *P. hexandrum*.

Thus genetic characterization of populations of *P. hexandrum* based on DNA profiling as well as assessment of impact of genetic and environmental factors on podophyllotoxin yield are essential for better identification and ascertaining of individuals that could be used for establishing cultivars for commercial level propagation in the Himalayan region.

Podophyllotoxin 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 [13]. Further improvement in their clinical efficacy by overcoming the drug resistance, myelosuppresion and poor bioavailability problems associated with podophyllotoxin [14], 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 studies provided important insights into the development of the structure-activity relationship models between analogues.

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 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.

Objectives

- To ascertain genetic diversity and population structure based on DNA profiling techniques such as RAPD, ISSR and AFLP as well as their relationship with podophyllotoxin content in the diminishing wild populations of *P. hexandrum* in the northwestern Himalayas, Himachal Pradesh, India.
- To establish the relationship between the environmental factors and podophyllotoxin yield if any and to develop prediction model for selection of conservation area for commercial cultivation of *P. hexandrum* populations.
- Development of computational screening models based on physicochemical properties, Docking, MM-GB/SA and SGB-LIE to facilitate the search for the potential drugs with low toxicity and better biological activity based on podophyllotoxin skeleton.

Outline of the works

Each of these seven 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 from background to conclusion at publication stage.

Study of genetic variation among *Podophyllum hexandrum* populations using RAPDs, ISSRs and AFLPs DNA marker

Plant materials

Twenty-eight populations of *P. hexandrum* were collected from 28 sites covering 11 forest divisions in the interior range of northwestern Himalayan region, Himachal Pradesh, India (Table 1). Young leaves were frozen in liquid nitrogen and stored at -80 ^{0}C prior to DNA isolation. The roots were trimmed from uprooted plants and dried at 60 ^{0}C for 24 hrs in an oven and used for podophyllotoxin estimations.

Table 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

Extraction and quantification of genomic DNA

Total genomic DNA was extracted from frozen leaves by the CTAB method [15]. RNA was removed by RNase treatment. DNA was quantified by using Nanodrop and then comparing with uncut λ DNA on the 1% agrose gel. The DNA sample was diluted to 12.5ng.μl^{-1} and used for PCR amplification.

Extraction and quantification of podophyllotoxin

Podophyllotoxin was extracted following the established procedure [2] from the 28 populations of *P. hexandrum*. HPLC analysis was carried out using a Nova Pack C18 cartridge column (250 x 4.6 mm) in Water's 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 area and peak heights.

1. Genetic characterization using RAPD markers

Methods

Nineteen random decamer primers (Operon Tech USA) were used for RAPD analysis (Table 2). DNA was amplified following the protocol [16]. 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 and 0.5x TBE-buffer followed by ethidium bromide staining and visualization in UV light for photography. Molecular sizes were estimated using DNA markers (Sigma).

Data analysis

The relationship among genomic DNA was assessed by comparing RAPD fragments separated according to their size and the presence/absence of shared fragments. Pairwise distance matrix was calculated using the Jaccard similarity coefficient [17]. The similarity values were used to generate a dendrogram via the unweighted pair group method with arithmetic average (UPGMA). Population diversity (Hs) and total gene diversity (Ht) [18] were calculated among populations within forest division and among 11 forest divisions by POPGENE software. The non-parametric analysis of molecular variance (AMOVA) was done using GenAlex.

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

Results and discussion

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 bands (Table 2). The number of amplification fragments produced per primer as well as their size range from 250 bp to 3,100 bp (Figure 1) which is analytically appropriate and in conformity with those recorded with certain other plants examined analogously [19]. The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variation in the *Podophyllum* population (Table 2). All the 28 *Podophyllum hexandrum* populations were distributed into 12 main clusters (C1-C12) with similarity value ranging from 0.61 to 0.96. Populations from 11 forest divisions were clustered into region-specific groups with the exception of Kullu forest division. The results indicate high genetic diversity in *P. hexandrum* from Himachal Pradesh. The analysis of molecular variance revealed 53% of variations among the 11 forest divisions and 47% of variation among populations within a forest division. 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.

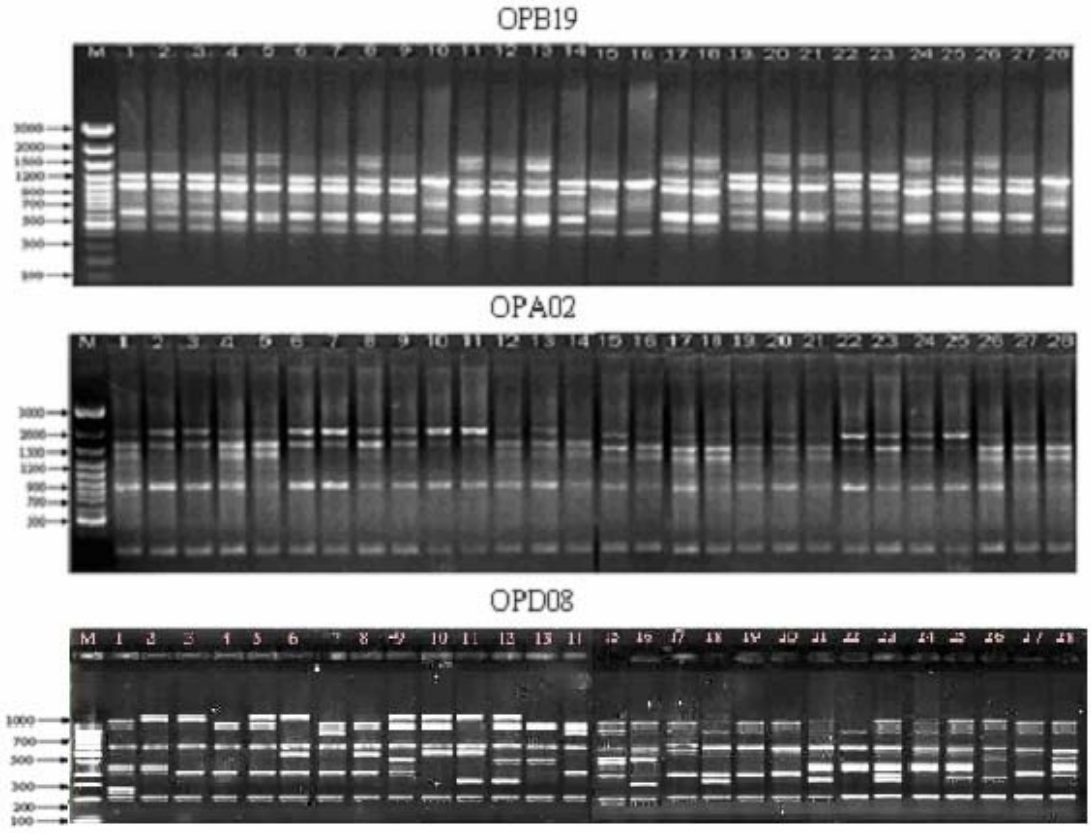


Figure 1. RAPD amplification products obtained from the 28 genotypes of *P. hexandrum* studied. 1. R/4 Kasol 2. Twin Multivora 3. Anganoala 4. Brundhar 5. Gulaba 6. ChanderKhani 7. Kaned Nry 8. Sanghar Nry 9. Madhvi Thach; 10. Kala Pani 11. Sojha Nursery 12. Jalora-C-3b 13. DPF-D-1892-C1 14. DPF-D-791-C1 15. Myar Valley 16. Nayan Ghar 17. Bada Bangal 18. Chota Bangal 19. IHBT 20. Bander Thach 21. Saropa Nry 22. Nichar Nry 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.

2. Genetic characterization using ISSR markers

Methods

Eleven ISSR primers (after being screened out of 30 primers) were selected (Table 3) and were commercially synthesized by Sigma Inc. 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 mentioned in Table 3, 2 min. at 72 ^{0}C and a 10 min. final extension step at 72 ^{0}C. Amplification products were electrophoresed on 1.5% agarose gel and 0.5X TBE buffer, visualized by staining with ethidium bromide and photographed under ultraviolet light (using Gel Doc, Biorad). Molecular sizes were estimated using λ DNA markers (Sigma).

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

Results and discussion

Eleven ISSR primers used in the study generated a total of 68 ISSR loci (an average of 6.18 loci per primer) out of which 57 were polymorphic (83.82%). The amplified PCR fragment size ranged from 250 bp to 1350 bp (Figure 2). 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% (Bahrmaur 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.

The analysis of molecular variance revealed 62% of variations among the 11 forest divisions and 38% of variation among populations within a forest division. 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). All the populations in each region clustered together having similarity co-efficient values ranging from 0.57 to 0.96. The result indicates high genetic diversity in *P. hexandrum* populations from Himachal Pradesh.

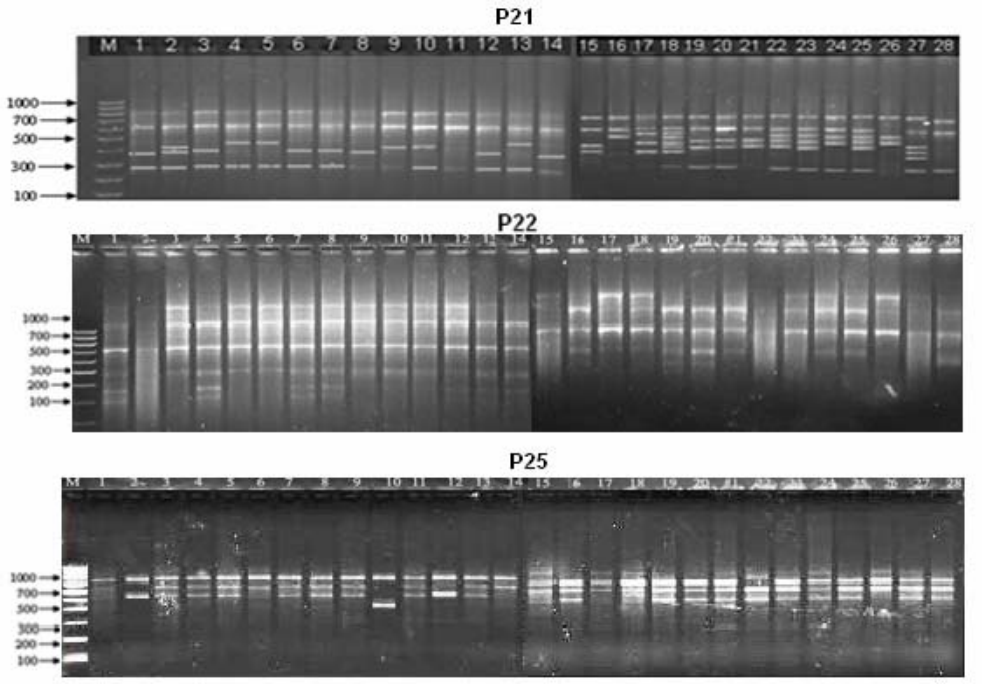


Figure 2. ISSR amplification products obtained from the 28 populations of *Podophyllum hexandrum*.

3. Genetic characterization using AFLP markers

Methods

AFLP was performed as described by Vos et al. [10] and was conducted using the Small Genome Primer Kit AFLP System II (Invitrogen life technology) and visualized with the Polyacrylamide Gel Electrophoresis (PAGE) system. 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 *Eco*RI and *Mse*I oligonucleotide adapters using 1 µl (1 units/µl) of T4 DNA ligase at 22 ^{0}C for 2 hrs. The pre-amplification reactions were performed by following PCR 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. Primer labeling is performed by phosphorylating the 5' end of the *EcoR* I primers with γ^{-32}P dATP (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 accordingly [20]. The reaction products were then size-fractioned on 6% denaturing polyacrylamide gel on a DNA sequencing apparatus. Electrophoresis was carried out for 4.0 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.

Results and discussion

The AFLP technique is highly sophisticated and reproducible due to its stringent amplification procedure [21] and has previously been used successfully in a variety of taxonomic and genetic diversity studies in other species. 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 4). The mean polymorphism rate was 69.60% observed among primer pairs). The AFLP banding pattern of 28 *P. hexandrum* genotypes is shown in the Figure 3 obtained with one primer combination. The fragment size ranged from 20 bp to 750 bp. The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variation among the *Podophyllum* genotypes. Based on AFLP marker, the similarity index values ranged from 0.59 to 0.90. These values were used to construct a dendrogram based on UPGMA method. All the 28 *Podophyllum hexandrum* populations were distributed into 14 main clusters (C1-C14). Populations from 11 forest divisions were clustered into region-specific groups with the exception of Kullu forest division which is distribute in 4 different clusters C5, C7, C10 and C14. Wide genetic diversity between populations of *P. hexandrum* was evident from the high number of polymorphic markers. The total number of polymorphic loci is 466, thereby, giving an estimate of profound (>84.40%) polymorphism.

The analysis of molecular variance revealed 64% of variations among the 11 forest divisions and 36% of variation among populations within a forest division. The gene diversity computed among different groups of populations was recorded in between 0.05 (Bharmaur) – 0.26 (Kullu). 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 0.13 among the 28 *P. hexandrum* populations. 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.

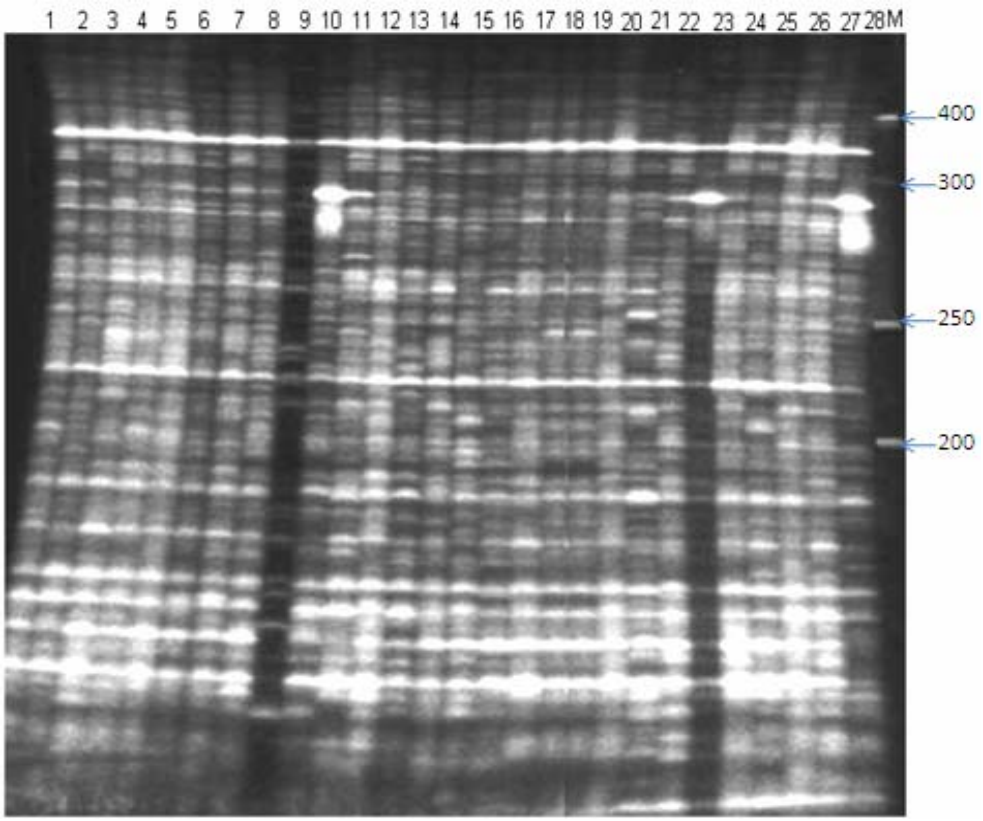


Figure 3. AFLP amplification products obtained from the 28 genotypes of *P. hexandrum* studied using 13 *Eco*RI/*Mse*I primer combinations.

Table 4. AFLP markers obtained from 12 primer combinations among 28 *Podophyllum* genotypes.

Pimer name (*Eco*RI/*Mse*I)	Total Loci	Monomorphic Loci	Polymorphic Loci	Polymorphic Loci(%)	Total no. of fragments	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 existing variations in podophyllotoxin content among the populations were proved to be coupled with altitude but not with genetic diversity (Figure 4). Thus the study demands the optimization of environmental factors in order to increase the rate of production of podophyllotoxin from collected populations.

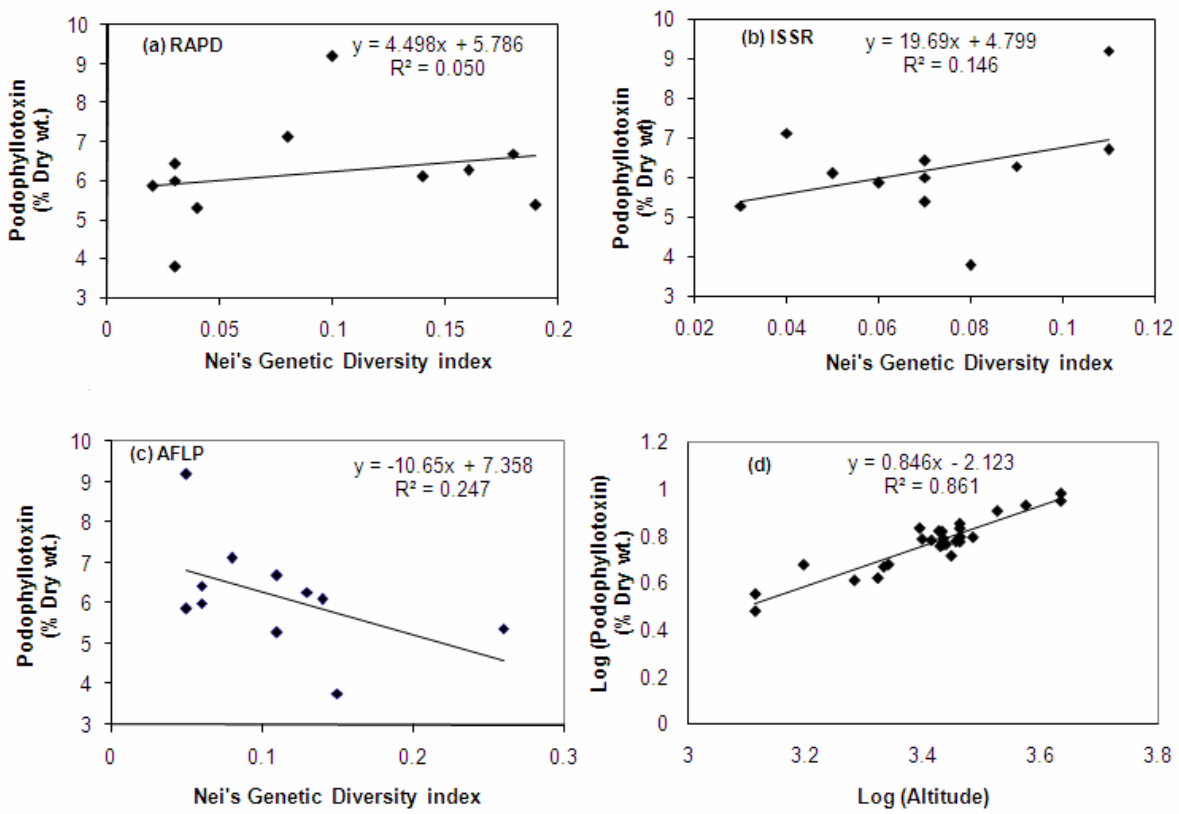


Figure 4. Regression analysis based on (a), (b) & (c) Nei's genetic diversity values of RAPD, ISSR & AFLP marker and podophyllotoxin content between 11 populations (forest division wise) of *P. hexandrum*. (d) Log_{10} M (geographical altitude) and Log_{10} (podophyllotoxin content) between 28 populations.

4. Impact of soil nutrient and environmental factors on podophyllotoxin content

Methodology

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. About 200g of soil was collected from the root level (in triplicate) 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. The correlation and regression analysis between the podophyllotoxin content and soil nutrients, environmental factors and altitude were examined by using MINITAB statistical package. 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.

Results and discussion

Podophyllotoxin content in the *P. hexandrum* differs greatly in different natural habitats. 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 and environmental factors 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) (Figure 5) and multiple linear regressions (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 ($r^2 = 0.991$) than MLR prediction model ($r^2 = 0.930$). Low root mean square error (RMSE) for ANN model (0.040) is less than MLR model (0.294) (Figure 6) 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. Both the ANN and MLR 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 aspects.

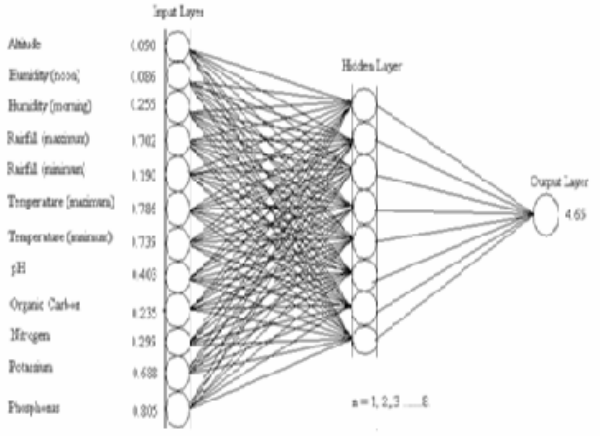


Figure 5. Layers and connection of a feed forward back propagating ANN.

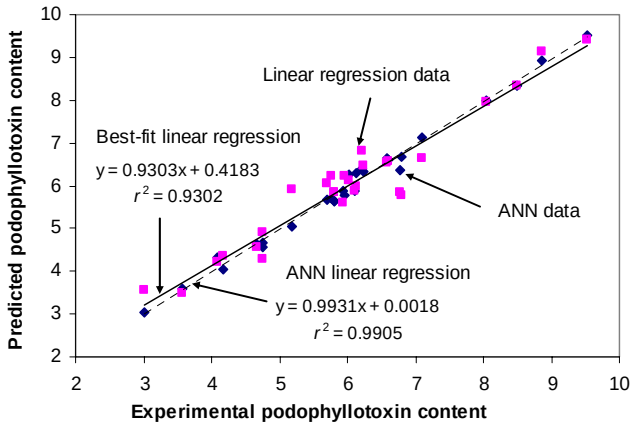


Figure 6. Comparison of estimated podophyllotoxin content using an artificial neural-network model (--,ANN) and a 'best-fit' regression model (—, MLR).

Development of computational screening methods of podophyllotoxin analogues

The following computational screening models have been developed to speed up development and help find good agents in rational drug discovery projects.

(a) Docking and molecular mechanics based on generalized Born/surface area (MM-GB/SA) solvation model.

(b) A linear interaction energy (LIE) method with a surface generalized Born (SGB) continuum solvation model.

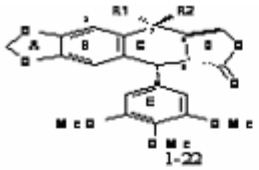
(c) A Quantitative structure-activity relationship (QSAR) model based on molecular descriptors.

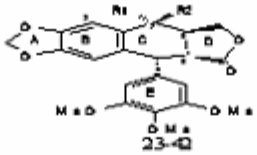
These methods produce a lot of information on a variety of drug related research. They benefit basic scientific activities as well as industrious efforts. The information from this work will provide help to other researchers who try to develop new potent podophyllotoxin agents.

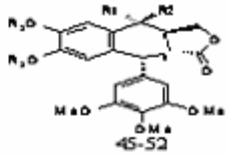
A library of podophyllotoxin analogues

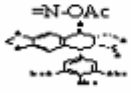
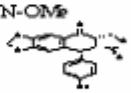
The library of podophyllotoxin analogues contains 154 compounds divided into 5 sublibraries: tetralinelactones consists of 52 (1-52) compounds [22] (Table 5); pyrazolignans and isoxazolignans [23] (98-120) (Table 6); non-lactonic tetralines [24] (53-97) (Table 7); lactonic and non-lactonic naphthalene [25] (121-126) (Table 8) and aza-podophyllotoxin analogues [26] (127-154) (Table 9). All these compounds were taken from various sources and they belong to different ring modifications. The library of podophyllotoxin analogues was built from the scaffolds by different ring modifications and substitutions of functional groups as mentioned in Table 5-9. We have 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 [27] was used for final

preparation of ligands. The ligands were minimized by means of Molecular Mechanics Force Fields (OPLS-2005).


1-22

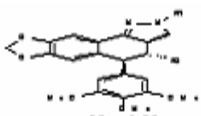

23-42

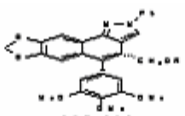

45-52

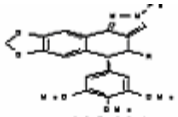
Analogue	R1	R2	Expt IC50	Analogue	R1	R2	Expt IC50
1	OH	H	0.012	22		=N-OMe	0.2
2	H	H	0.010	23	H	H	0.10
3	H	H(2-OMe)	0.01	24	H	H(2-OMe)	0.23
4	OH	H(2-OMe)	-	25	OH	H	6.0
5	OH	H(4'-OH	0.027	26	OH	H(2-OMe)	-
6	OAc	H	0.625	27	OAc	H	0.55
7	OAc	H	-	28	OAc	H(2-OMe)	1.02
8	OMe	H	0.06	29	OMe	H	0.12
9	H	OH	0.06	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$\Delta^{8(87)}$	-
17	H	H(4'-OH)	-	38	H	HΔ^{7}	0.013
18	H	H(4'-OAc)	-	39		=O 120	
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			-

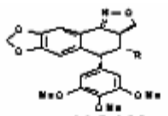
Analogue	R1	R2	R3	Expt IC50	Analogue	R1	R2	R3	Expt IC50
45	H	H	H	-	49	H	OH	H	-
46	H	H	Ac	-	50	H	OAc	Ac	-
47	OH	H	H	-	51	H	H	H$\Delta^{8(87)}$	-
48	OAc	H	Ac	-	52	H	H	Ac$\Delta^{8(87)}$	-

Table 5. Podophyllotoxin derivatives (Tetralactones) [22] 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

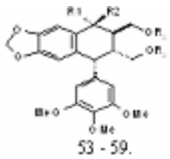

112-114


115-120

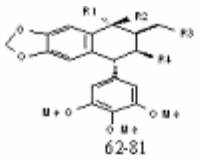
Analogue	R1	R2	Expt.IC50	Analogue	R1	R2	Expt.IC50
98	Ph	COOH	1.9	104	p-BrPh	COOH	-
99	Ph	COOMe	1.00	105	p-MePh	COOMe	1.00
100	Ph	CH_2OH	4.1	106	Me	COOH	-
101	Ph	CH_2OAc	4.7	107	Me	COOMe	5.6
102	m-NO_2Ph	COOH	-	108	$CONH_2COOH$	COOH	-
103	m-NO_2Ph	COOMe	4.5	109	$COCH_3COOMe$	COOMe	21

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_3O	2.4
114	CH_2Ac	2.2	120	CHO	-
115	COOH	2.2			

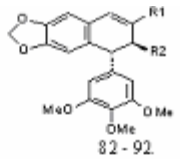
Table 6. **Podophyllotoxin derivatives (Pyrazolignans and isoxazolignan) [23] with cytotoxic activities against P-388 cell line as well as new proposed structural derivatives with unknown cytotoxic activity used in the work.**



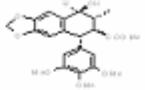
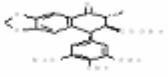
53 - 59.



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

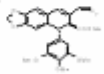
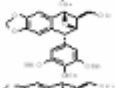
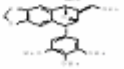
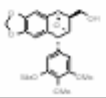
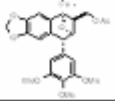
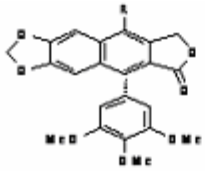
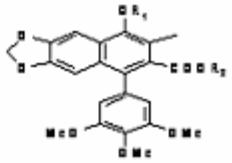
Analogue	R1	R2	Expt.IC50	Analogue	R1	R2	Expt.IC50
82	CH_2OH	COOMe	0.02	89	CH=N-OH	COOMe	2.27
83	CHO	CH_2OH	0.25	90	CH=N-OMe	COOMe	0.22
84	CHO	COOMe	0.23	91		COOMe	0.20
85	CH=N-NH2	COOMe	0.57	92		CH_2OH	1.00
86	CH=N-NH-CH_2CF_3	COOMe	0.48	93			0.57
87	CH=N-NH-Ph	COOMe	1.94	94			6.25
88	CH=N-NH-Ph	CH_2OH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

Table 7. Podophyllotoxin derivatives (Nonlactonic tetralines) [24] 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 8. Podophyllotoxin derivatives (lactones and non-lactonic naphthalene) [25] **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		
B Ring	E Ring	Expt. IC50	Analog	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

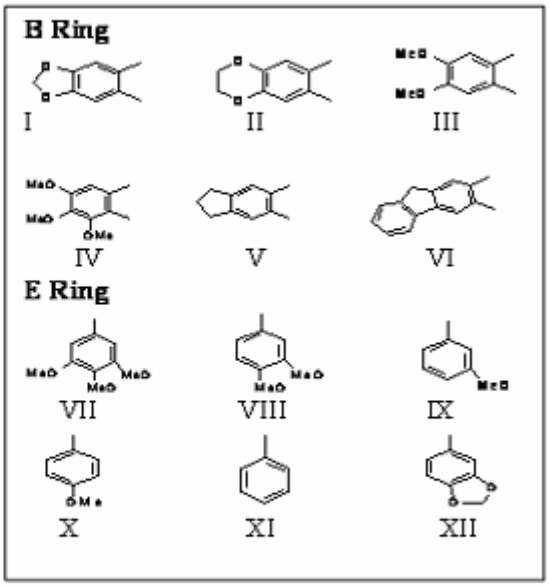


Table 9. Aza-podophyllotoxin derivatives [26] with cytotoxic activities against P-388 cell line used in the work.

5. Docking and MM-GB/SA solvation model

Methodology

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 [27] 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. The complex obtained was minimized using OPLS-2005 force field with Polack-Ribiere Conjugate Gradient (PRCG) algorithm. The minimization was stopped either after 5000 steps or after the energy gradient converged below 0.05 kcal/mol.

The Schrodinger Glide program version 4.0 has been used for docking [28]. The best 10 poses and corresponding scores have been evaluated using Glide in single precision mode (Glide SP). The pose with the lowest Glide SP score has been taken as the input for the Glide calculation in extra precision mode (Glide XP). For each ligand, the pose with the lowest Glide score was rescored using Prime MM-GB/SA approach [29]. 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 (GB/SA) continuum solvent model. The binding free energy (ΔG_{bind}) is then estimated using equation:

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

where $E_{R:L}$ is energy of the complex, $E_R + E_L$ is sum of the energies of the ligand and unliganded receptor, the outcome of the use of OPLS-AA force field, ΔG_{solv} (ΔG_{SA}) is the difference between GB/SA solvatation energy (surface area energy) of complex and sum of the corresponding energies for the ligand and unliganded protein.

Results and discussion

In the docking simulations, the flexible docking reproduced well the binding structure of crystal structures and thus verified the docking protocol adapted in the work. All the 154 podophyllotoxin analogues were found to be good binders with tubulin. The following quantitative structure activity relationships (QSAR) models were developed between the cytotoxic activity (pIC_{50}) of these compounds and molecular descriptors like docking score and binding free energy.

$$pIC_{50} = - 8.725(\pm 0.644) - 0.938(\pm 0.592)* \text{ G-score}$$
$$(N = 120, r^2 = 0.642, s = 0.692, F = 211.86, r^2_{cv} = 0.631, PRESS = 58.349)$$

$$pIC_{50} = -2.604(\pm 0.008) - 0.143(\pm 0.148)* \Delta G_{bind}$$
$$(N = 120, r^2 = 0.728, s = 0.603, F = 316.58, r^2_{cv} = 0.719, PRESS = 44.415)$$

For both the cases the r^2 was in the range of 0.642 to 0.728 indicating good data fit and r^2_{cv} was in the range of 0.631 to 0.719 indicating that the predictive capabilities of the models were acceptable. In addition, a linear correlation was observed between the calculated binding free energy and pIC_{50} for the validation set with correlation coefficient (r^2) in the range of 0.806 to 0.887, suggesting that the docked structure orientation and the interaction energies are reasonable. Low level of root means square error for the majority of inhibitors which establish the docking and prime MMGBSA 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 derivatives.

6. A LIE-SGB continuum solvation model

Methodology

The structure-based linear interaction energy method implementing a surface generalized Born (SGB-LIE) [30] 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. The docked complex corresponding to each analogue was transported to the LIAISON package for subsequent SGB-LIE calculations. The molecular dynamics (MD) technique has been used for LIE conformation space sampling. The system was initially heated to 300 K in 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 using the following equation.

$$\Delta G_{bind} = \alpha\,(\langle U^{b}_{ele}\rangle - \langle U^{f}_{ele}\rangle) + \beta\,(\langle U^{b}_{vdw}\rangle - \langle U^{f}_{vdw}\rangle) + \gamma\,(\langle U^{b}_{cav}\rangle - \langle U^{f}_{cav}\rangle)$$

where bracketed terms represent the ensemble average of the energy terms such as van der Waals (*Uvdw*), electrostatic (*Uele*) and cavity (*Ucav*) energy. All the terms are evaluated for interaction between ligand, both in the free (f) and bound (b) state and its environment. The α, β and γ are LIE fitting parameters. The values obtained for the three fitting parameters, α, β and γ, are -0.141, -0.0926, and -1.07, respectively.

Results and discussion

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). The root mean square error (RMSE) between the experimental ΔG values and the values obtained by the SGB-LIE fit was 0.48 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 0.871 for the training set. 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.

$$\Delta G = (-0.141)\langle Uele\rangle + (-0.093)\langle Uvdw\rangle + (-1.07)\langle Ucav\rangle$$

($N = 76$, $r^2 = 0.8714$, $r^2_{pred} = 0.864$, $s = 0.598$, $F = 166.8$, $P = 0.0001$, $q^2 = 0.865$, $PRESS = 28.05$)

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 free energy values for the test set were estimated based on optimized SGB-LIE parameters α, β and γ from the training set. There is a close match between the experimental and LIE free energy values of the ligands in the test set. Figure 7a & 7b graphically show the quality of fit for the training and test set.

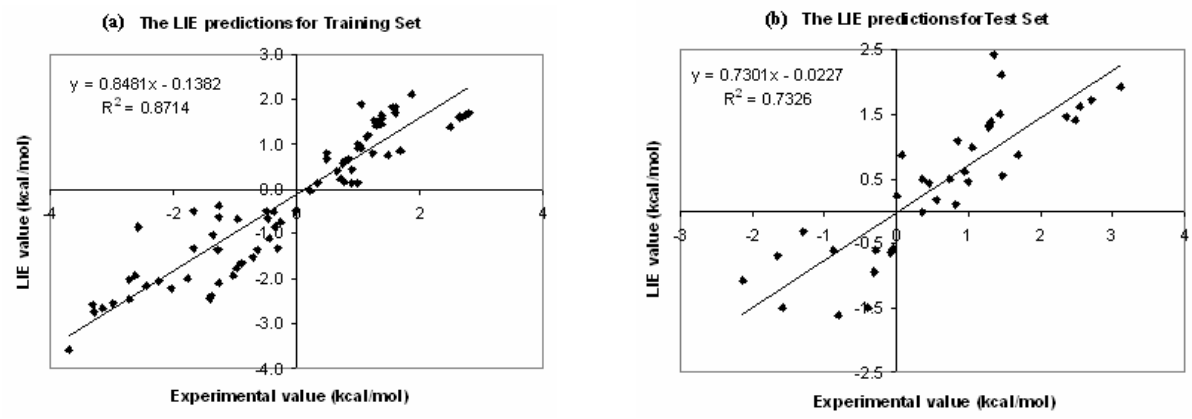


Figure 7. Free energy values estimated by the SGB-LIE method for (a) 76 podophyllotoxin analogues comprising the Training set plotted against corresponding experimental data; the RMS error is 0.48 kcal/mol and (b) for 36 podophyllotoxin analogues comprising the Test set plotted against corresponding experimental data; the RMS error is 0.56 kcal/mol.

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 with a root-mean-square error (RMSE) of 0.56 kcal/mol with respect to experimental data. Low levels of RMSE for the majority of analogues 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.

7. Computational screening of podophyllotoxin analogues based on QSAR model

Dataset : A total of 119 podophyllotoxin analogues were used in the study and were taken from various sources belonging to different ring modifications. These molecules were divided into 81 molecules in training set and 38 molecules in test set.

Descriptor calculation: A set of 372 molecular descriptors belonging to E-state indices log P structural, symmetrical, topological, lead likeness, electronic, extended Huckel partial charge molecular moments, orbital energies, molecular connectivity indexes, gravitational indexes, hydrophobicity, steric, thermodynamic factors and topological descriptors were calculated using ADME Model Builder software package. We used a more systematic way of variable selection in

order of missing value test → zero test → simple correlation test → genetic algorithm to obtain a set of good descriptors. QSAR equations were developed between the observed biological activity and the descriptors. 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 y are the predicted, experimental and mean values of activity, respectively. Evaluation of the predictive ability of the model was done by determining the value of rm^2 by the following equation [31].

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

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 best significant relationship for the inhibition of P388 cell-line has been deduced to be

pIC50 = - 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

(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, 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.

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. The quality of the prediction models for the training compounds before and after removal of outliers have been shown in Figure 8. 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 QSAR equation is reliable.

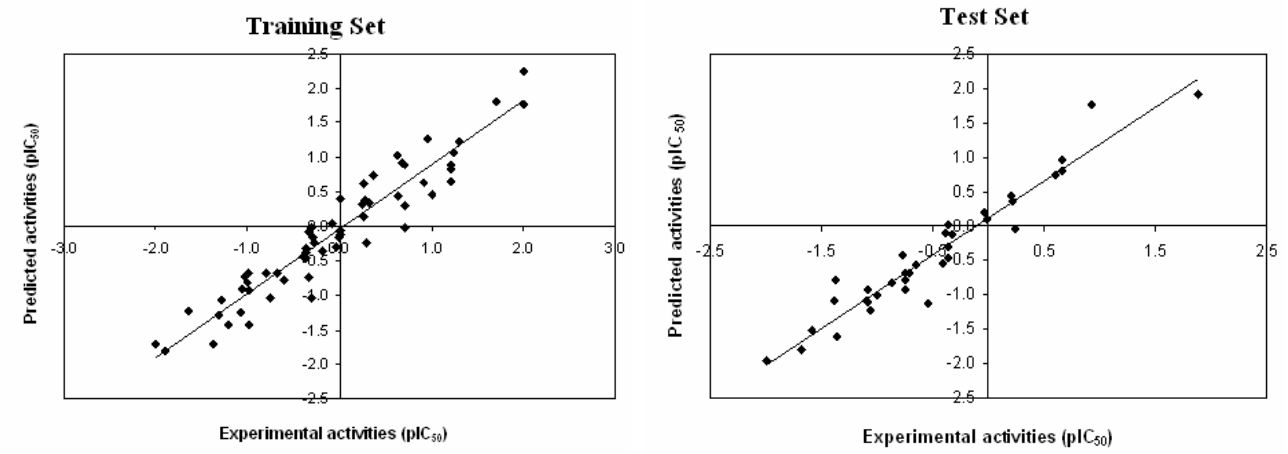


Figure 8. Relationship between predicted and experimental activities as per QSAR equation.

Being the value of $q^2 = 0.888$, the model corroborates with the criteria for a QSAR model to be highly predictive [32]. Also the value of $R^2_{pred} = 0.918$ and $rm^2 = 0.702$ were found to be in the acceptable range [31] thereby indicating the good external predictability of the QSAR model.

The high predictive ability of the computational models developed 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, 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

- High genetic variation among the populations of *P. hexandrum* was revealed using RAPD, ISSR and AFLP molecular markers.
- The yield of podophyllotoxin is not related to genetic variations but infact related to environmental factors.
- Significant correlation co-efficient and low level of root mean square error, established the docking and prime MM-GB/SA 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 derivatives.

- Thus fast and accurate estimation of binding free energies based on LIE-SGB 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.
- 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).
- 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.

List of Publications

1. Afroz Alam M, Naik PK (2009). Molecular modelling evaluation of the cytotoxic activity of podophyllotoxin analogues. *J Comput Aided Mol Des.* 23: 209-225.

2. Afroz Alam M, 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.

3. Afroz Alam M, 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: 1-12.

4. Afroz Alam M, 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. (Accepted in Journal: *Comm. Soil Sci. Plant Anal.).* Article in press (Publisher: Taylor & Francis).

5. Afroz Alam M, 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. Afroz Alam M, 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 (Publisher: CSIR, India).

References

1. Anon (1970). Raw Materials. Vol. VIII: Ph.Re. Publication and Information Directorate. CSIR, New Delhi, 170-174.
2. Broomhead AJ and Dewick PM (1990). *Phytochemistry* 29: 3831-3837.
3. Jackson DE & Dewick PM (1984). *Phytochemistry* 23:1147-1152.

4. Imbert TF (1998) Discovery of podophyllotoxins. Biochimie 80: 207-22.
5. Fay DA & Ziegler HW (1985). *J. liq. Chromatogr,* 8: 1501-1506.
6. Gupta R, Sethi KL (1983). In: Jain, S.K., Mehra, K.L. (Eds.). *Botanical Survey of India, Howrah.* pp. 101-107.
7. Sharma TR, Singh BM, Sharma N R & Chauhan RS (2000). *J. Plan Biochem. & Biotech.* 9 49-51.
8. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA & Tingey SV (1990). *Nucl.Acids Res.* 18: 6531-6535.
9. Morgante M & Olivieri AM (1993). *Plant J.* 3: 175-182.
10. 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. *Nucleic Acids Res* 23: 4407–4414.
11. Lu JJ, Knox MR, Ambrose MJ, Brown JKM & Ellis THN (1996). *Theoretical and Applied Genetics* 93:1103-1111.
12. Endress R (1994). *Plant Cell Biotechnology*. Pp.187-242.
13. Weiss SG, Tin-Wa M, Perdue RE Jr, Farnsworth NR (1975). *J. Pharm. Sci*. 64: 95.
14. Witterland AH, Koks CH & Beijnen JH (1996). Etoposide phosphate, the water soluble prodrug of etoposide. *Pharmacy World Sci.* 18:163.
15. Saghai-Maroof MA, Soliman KM, Jorgensen RA & Allard RW (1984) *Proc. Natl. Acad. Sci. (USA)* 81 8104-8118.
16. Kazan K, Manners JM, Cameron DF (1993). *Theor. Appl. Genet.* 85: 882-288.
17. Sneath PHA & Sokal K (1973). *WHF, San Francisco*. 100-308.
18. Nei M (1978). *Genetics* 89:83-590.
19. Airi S, Rawal RS, Dhar U & Purohit AN (1997). *Plant Genet. Resour. Newsl*. 110: 29-34.
20. Myburg AA, O'Malley D, Sederoff RR & Whetten R (2000). *Plant & Animal Genome* VIII conference, San Diego, CA, p 544].
21. Brown JKM (1996). *New Phytologist* 133**:**183-195.
22. Gordaliza M, Castro MA, Miguel del Corral JM & San Feliciano A (2000). Antitumor properties of podophyllotoxin and related compound. *Current Pharmaceutical Design* 6:1811-1839.
23. San Feliciano A, Gordaliza M, Miguel del Corral JM, Castro MA, Garcı´a-Gra´valos MD, Ruiz-La´zaro P (1993) Planta Med 59:246. doi:10.1055/s-2006-959660
24. Castro MA, Gordaliza M, Miguel del Corral JM, San Feliciano A (1994) Org Prep Proced Int 26:539.
25. San Feliciano A, Gordaliza M, Miguel del Corral JM, Castro MA, Garcı´a-Gra´valos MD, Ruiz-La´zaro P (1993) Planta Med 59:246.doi:10.1055/s-2006-959660
26. 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. *Bioorganic & Medicinal Chemistry Letters* 10: 315-317.
27. Schrodinger LLC. , http://www.schrodinger.com, (accessed: 24. 04.2007).
28. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL **(**2004**).** *J. Med. Chem.* 47:1750-1759.
29. Lyne PD, Lamb ML, Saeh JC (2006). *J. Med. Chem.* 49: 4805-4808.
30. Zhou R, Frienser RA, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM (2001). *J. Phys.Chem. B.*105: 10388-10397.
31. Roy PP & Roy K (2008). *QSAR Comb Sci*, 27:302-313.
32. Golbraikh A & Tropsha A (2002). *J Mol Graph Model*. 20(4):269-276.