



# INTERNATIONAL JOURNAL OF PHARMACEUTICAL RESEARCH AND DEVELOPMENT (IJPRD)

Platform for Pharmaceutical Researches & Innovative Ideas  
www.ijprd.com

## IMPROVING BINDING AFFINITIES OF CONTINUUM SOLVENT ENERGY MODEL OF PODOPHYLLOTOXIN ANALOGUES USING GA, RBFNN FOR PREDICTION OF CYTOTOXIC ACTIVITY

Md. Afroz Alam<sup>1\*</sup>,

Anurupa Devi. Y<sup>1</sup>, Thalitha Jane. D<sup>1</sup>, Thavavel. V<sup>2</sup>

<sup>1</sup>Department of Bioinformatics, Karunya University, Karunya Nagar, Coimbatore, 641114, Tamil Nadu, India.

<sup>2</sup>Department of Computer Application, Karunya University, Karunya Nagar, Coimbatore, 641114, Tamil Nadu, India.

### ABSTRACT

Podophyllotoxin and its congeners have great 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  $\alpha$ - and  $\beta$ -tubulin interface and interfere with tubulin polymerization. The genetic algorithm and RBFNN were used to obtain the binding free energy values which are then compared with binding free energy values obtained by LIE 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. On the basis of this insight, 15 analogues of varying ring modifications were taken and then validated with their experimental potencies of tubulin polymerization inhibition. An accurate estimation of binding free energies of these analogues using the proposed method provides a means to screen the compound libraries for generating more potent and specific inhibitors of tubulin by testing rationally designed lead compounds based on podophyllotoxin derivatives. Considering this as our objective we have constructed the RBFNN prediction model that revealed better prediction of activity on the basis of squared correlation coefficient between experimental and predicted activity for the test training set ( $R^2 = 0.9757$ ), test set ( $R^2 = 0.9986$ ) and validation set ( $R^2 = 0.8122$ ) compounds is significant. High accuracy of prediction model than GA and LIE model with respect to the experimental activity establishes the fact that the use of RBFNN as an efficient prediction model for analysing drug activity in the process of drug development.

### Correspondence to Author



**Md. Afroz Alam**

Assistant Professor (SG),  
Department of Bioinformatics,  
Karunya University, Karunya  
Nagar, Coimbatore, 641114,  
Tamil Nadu, India.

### Email

maarkzin@gmail.com

### Key Words

Genetic Algorithm, RBFNN, LIE model, Podophyllotoxin, Tubulin.

## INTRODUCTION

Internationally, there are 10 million new cancer cases and approximately 7 million cancer-related deaths reported each year, making cancer research a top priority<sup>1</sup>. Many cancer cells have defects in their progression through the cell cycle or their regulation of cell death. Although the goal of cell division is for each daughter cell to inherit one and only one copy of each intact chromosome, defects in this process can lead to aneuploidy, genetic instability and, ultimately, tumourigenesis<sup>2</sup>. Many tumour cells are insensitive to signals that normally promote cell adherence, differentiation and apoptosis. In particular, a distinct feature of tumours is a lack of regulation through the cell cycle, resulting in uncontrolled proliferation<sup>3</sup>. Various mechanisms involved in the pathophysiology of human cancers have been exploited as pharmacological targets, and numerous drugs have been discovered or designed to defend the cancer. It is a challenge, however, that many cancer cells develop resistance to such compounds following long term treatment, and thus there arose a need to develop more drugs that can be used for cancer chemotherapy<sup>4</sup>. In this perspective the role of natural products in drug discovery and development, especially for the anticancer drugs, with 67% of anticancer drugs being either natural products or natural product derivatives or natural product mimics became the limelight and a lot of research has been done on them<sup>5</sup>. Podophyllotoxin is one such natural compound that represents perhaps the most significant addition to the pharmacopoeia of cancer chemotherapeutic agents since the last decade<sup>6</sup>.

### Anti cancerous agents and Microtubules

The 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<sup>6</sup>. Microtubules are hollow tubes consisting of  $\alpha$ - and  $\beta$ -tubulin heterodimers that polymerise parallel to a cylindrical axis<sup>7</sup>. Pharmacological studies have revealed that microtubules play crucial roles in the regulation of the mitotic apparatus and the disruption of microtubules can induce cell-cycle arrest in the M phase, formation of abnormal mitotic spindles and final

triggering of signals for apoptosis<sup>8</sup>. Their importance in mitosis and cell division makes microtubules an important and interesting target for anticancer drugs. The discovery that the cytotoxic activity of various compounds is through interference with the mitotic spindle apparatus has attracted much attention within the last four decades, and microtubules have become an attractive pharmacological target for anticancer drug discovery<sup>9</sup>. Currently there are three well established drug binding sites on tubulin, the vinca domain, the taxane site and the colchicine site. The vinca domain is located adjacent to the exchangeable GTP binding site in tubulin at the plus end interface<sup>10</sup>. The taxane site resides in a deep hydrophobic pocket at the lateral interface between adjacent protofilaments, within the lumen of the microtubule<sup>11</sup>. Thirdly, the colchicine site is located at the intra- dimer interface between  $\alpha$ -tubulin and  $\beta$ -tubulin<sup>12</sup>. Out of them Microtubule destabilising agents which bind to the colchicine binding site and Podophyllotoxin which competes with the colchicine binding site have attained much focus<sup>6</sup>. Structure activity relationships (SAR) have shown that podophyllotoxin analogues preferentially inhibit tubulin polymerization, leading to the arrest of the cell cycle in the metaphase<sup>13</sup>.

### Mechanism of action of Podophyllotoxin

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. 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<sup>14</sup>. 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<sup>15</sup>. Podophyllotoxin binds to  $\beta$ -tubulin at its interface with  $\alpha$ -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  $\alpha$ - and  $\beta$ -tubulin complexed with podophyllotoxin (PDB\_ID:1SA1), showing that podophyllotoxin also binds at the colchicine site<sup>16</sup>.

Although colchicine site agents share a general toxicity, the promise to discover therapeutically useful analogues has fuelled 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 in determining the essential features for activity. A rational approach for the discovery of a pharmaceutically acceptable, economically viable activity model is important to develop and test large number of these analogues. Previously the magnitude of free energy changes upon binding of podophyllotoxin 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<sup>17</sup>. However its clinical application in the treatment of cancer has been limited by severe toxic side effects<sup>18</sup>. 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<sup>6</sup>. As a result the dependence on the computational models has been increased for virtual screening and prediction of biological activities of podophyllotoxins. Since, Hansch's seminal work on quantitative structure activity relationship (QSAR) analysis<sup>19</sup>, many different QSAR methods including 2D and 3D QSAR approaches have been developed<sup>20</sup>. In contrast to classical statistical methods such as regression analysis or partial least squares analysis (PLS), genetic algorithms and artificial neural networks were used to enable the investigation of complex, nonlinear relationships<sup>21</sup>. Presently, we employed genetic algorithm (GA) and Radial basis function neural network (RBFNN) to predict the cytotoxic activity of podophyllotoxin analogues, and the results of the prediction by GA, RBFNN and LIE approach are compared with those obtained by in vitro

experiments in order to select the most accurate method.

## MATERIALS AND METHOD

### The LIE model for prediction

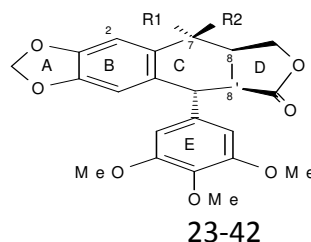
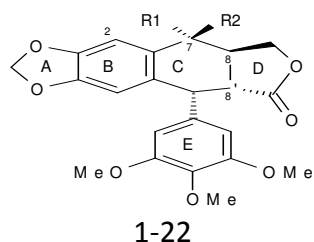
Followed by the preparation of the receptor using PPrep and ligands using Lig Prep<sup>17</sup>, ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated using grid receptor generation program. The ligands were docked initially by the "standard precision" method and further refined by "extra precision" Glide algorithm<sup>17</sup>. The docked complex corresponding to each analogue was transported to the Liaison package for subsequent SGB-LIE calculations. All the calculations of the SGB-LIE method performed in the Liaison package using Schrodinger Inc. 2008 for 112 analogues of podophyllotoxin were taken from<sup>17</sup>. The Liaison module performs LIE calculations in the OPLS force field with a residue-based cut off of 15Å°. The average LIE energy terms were used for building binding affinity model and free energy estimation for podophyllotoxin analogues. This approach 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<sup>17</sup>. 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.

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

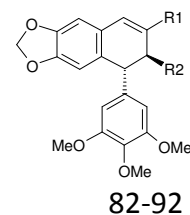
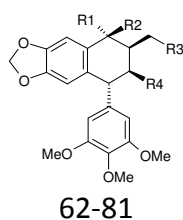
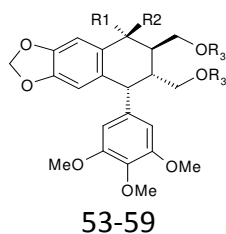
Here  $\langle \rangle$  represents the ensemble average,  $b$  represents the bound form of the ligand,  $f$  represents the free form of the ligand, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are the coefficients.  $U_{ele}$ ,  $U_{vdw}$  and  $U_{cav}$  are the electrostatic, van der Waals and cavity energy terms in the SGB continuum solvent model. Presently, total of 112 podophyllotoxin analogues were used in the study divided in to four sub libraries namely **Sub lib-I**: Commonly known as tetraline

lactones (1-29) (Table 1a) that are rationally designed with the goal of simplifying the chemical synthesis and improving the cytotoxic activity. **Sub lib-II:** Contains compounds (30-70) (Table 1b) known as nonlactonic tetralines (lack lactone rings). **Sub lib- III:** Includes a **Table 1a** Podophyllotoxin derivatives (Tetraline lactones) with cytotoxic activities against P-388 cell line grouped as Sublib-1<sup>17</sup>.

group of lignans (71–84) (Table 1c) that have heterocyclic rings fused to the cyclolignan skeleton. **Sub lib-IV:** Contains 28 compounds (85–112) (Table 1d) commonly known as aza podophyllotoxin analogues<sup>17</sup>.





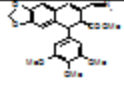
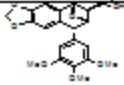
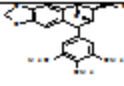
Analogue	R1	R2	Expt.IC <sub>50</sub>	Analogue	R1	R2	Expt.IC <sub>50</sub>
1	OH	H	0.012	16	H	H	0.10
2	H	H	0.010	17	H	H(2-OMe)	0.23
3	H	H(2-OMe)	0.01	18	OH	H	6.0
4	OH	H(4'-OH)	0.027	19	OAc	H	0.55
5	OAc	H	0.625	20	OAc	H(2-OMe)	1.02
6	OMe	H	0.06	21	OMe	H	0.12
7	H	OH	0.06	22	H	OH(2-OMe)	0.11
8	H	Ac	0.05	23	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 <sup>7</sup>	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				

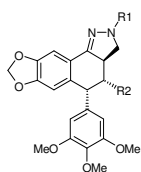
**Table 1b** Podophyllotoxin derivatives (Nonlactonic tetralines) with cytotoxic activities against P-388 cell line grouped as sublib-2<sup>17</sup>.

Analogue	R1	R2	R3	Expt.IC <sub>50</sub>	Analogue	Structure	Expt.IC <sub>50</sub>
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

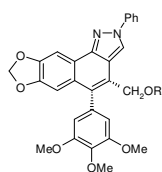
Analogue	R1	R2	R3	R4	Expt.IC <sub>50</sub>	Analogue	R1	R2	R3	R4	Expt.IC <sub>50</sub>
37	H	H	OH	COOMe	0.058	47	H	OMe	OAc	CH <sub>2</sub> OAc	9.7
38	H	H	OAc	COOMe	0.21	48	H	OH	OH	CH <sub>2</sub> OH	47.9
39	H	H	OAc	CH <sub>2</sub> OAc	5.14	49	H	OH	OH	COOMe	1.1
40	OH	H	OH	CH <sub>2</sub> OH	23.9	50	=O		OH	COOMe	5.63
41	OH	H	OH	COOMe	0.22	51	=O		OAc	COOMe	0.20
42	OAc	H	OAc	CH <sub>2</sub> OAc	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 <sub>2</sub> OH	23.2	54	H	H	=N-OMe	COOMe	2.30
45	OMe	H	OAc	CH <sub>2</sub> OAc	19.4	55	H	H	=N-OMe	COOMe	10.94
46	H	OMe	OH	CH <sub>2</sub> OH	11.6	56	H	H	=N-allyl	COOMe	2.5

**Table 1c** Podophyllotoxin derivatives (Pyrazolignans and isoxazolignan) with cytotoxic activities against P-388 cell line grouped as sublib-3<sup>17</sup>.

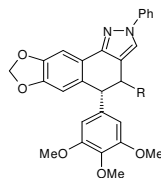
Analogue	R1	R2	Expt.IC <sub>50</sub>	Analogu e	R1	R2	Expt.IC <sub>50</sub>
57	CH <sub>2</sub> OH	COOMe	0.02	64	CH=N-OH	COOMe	2.27
58	CHO	CH <sub>2</sub> OH	0.25				
59	CHO	COOMe	0.23	65	CH=N-OMe	COOMe	0.22
60	CH=N-NH <sub>2</sub>	COOMe	0.57	66		COOMe	0.20
61	CH=N-NH-CH <sub>2</sub> CF <sub>3</sub>	COOMe	0.48	67		CH <sub>2</sub> OH	1.00
62	CH=N-NH-Ph	COOMe	1.94	68			0.57
63	CH=N-NH-Ph	CH <sub>2</sub> OH	1.02				
69			6.25				
				70			5.66



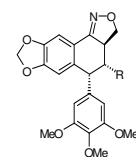
98 - 109



110-111



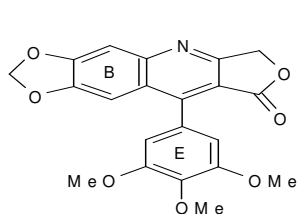
112-114



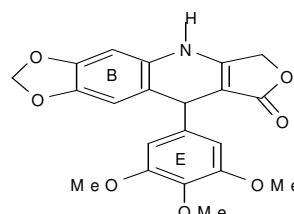
115-120

Analogue	R1	R2	Expt.IC <sub>50</sub>	Analogue	R1	R2	Expt.IC <sub>50</sub>
71	Ph	COOH	1.9	74	m-NO <sub>2</sub> Ph	COOMe	4.5
72	Ph	CH <sub>2</sub> OH	4.1	75	Me	COOMe	5.6
73	Ph	CH <sub>2</sub> OAc	4.7	76	COCH <sub>3</sub> COOMe	COOMe	21

Analogue	R	Expt.IC <sub>50</sub>	Analogue	R	Expt.IC <sub>50</sub>
77	H	10	81	COOMe	23
78	CHO	21	82	COOMe(4'-OH)	12
79	CH <sub>2</sub> Ac	2.2	83	CH <sub>2</sub> OH	2.6
80	COOH	2.2	84	CHO	2.4

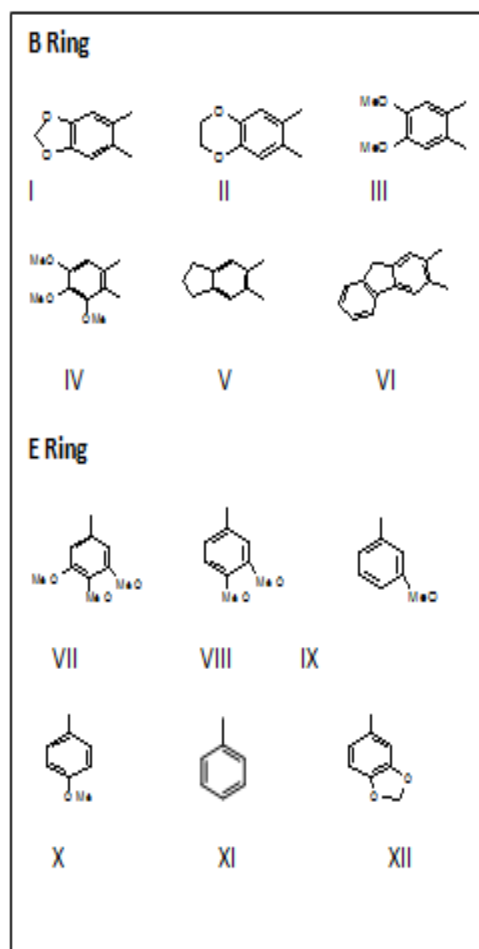
**Table 1d** Aza-podophyllotoxin derivatives with cytotoxic activities against P-388 cell line grouped as sublib-4<sup>17</sup>.

Modification 1



Modification 2

Modification 1				Modification 2			
Analog	B Ring	E Ring	Expt. IC <sub>50</sub>	Analog	B Ring	E Ring	Expt. IC <sub>50</sub>
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

**prediction model**

This section of the present work illustrates how GA can be applied to optimize the optimal energy parameters namely, electrostatic, van der Waals and cavity energy with respect to *in vitro* cytotoxicity values (IC<sub>50</sub>) obtained against cell line P388<sup>17</sup>. All the energy parameters were normalized by standard deviation normalization method and were used for GA optimization by following procedure.

A) Encoding: The GA programmed for the present work operates on a population of candidate solutions encoded ( $\vartheta_1, \vartheta_2, \vartheta_3$ ). Each variable in the parameter set is encoded as a binary string and all are concatenated to form a chromosome. Chromosome representation is a three-element vector containing  $\vartheta_1, \vartheta_2, \vartheta_3$  parameters in binary genetic coding. The precision of binary representation is four bits per parameter (i.e. [1 0 0 1; 0

1 1 1; 0 1 1 1]). It begins with a randomly constructed population of initial guesses.

B) Fitness function: A fitness function assessing the performance for each chromosome must be designed before searching for the optimal values for energy parameters. Several measurement indicators have been proposed and employed to evaluate the prediction accuracy of models such as mean absolute errors (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) in prediction problems. To compare the results achieved by the present model with the

experimental value, this study employed regression analysis.

C) Selection and Crossover: Our selection technique employs a “roulette wheel” mechanism to probabilistically select individuals based on their performance. The selected candidates were allowed to participate in an arithmetic crossover, the procedure that recombines promising candidates in order to create the next generation.

Settings for GA Operators is tabulated below :

Population size	50	Mutation	Uniform
Evolution generation	50	$p_c$	0.6
Selection	Roulette wheel	$p_m$	0.01
Crossover	One-point	Elitism	Yes

### Radial basis function (RBF) neural network

RBF network is an extremely powerful neural network, which uses a cluster algorithm in the first step (unsupervised training) and calculates the approximation in the second step (supervised training)<sup>22</sup>. All the standard deviation normalized energy parameter namely; electrostatic, van der Waals and cavity energy were used as an input for RBF neural network model. RBF neural network includes an input layer which serves only to distribute the input data among the hidden neurons. A single hidden layer consists of the hidden neurons which are expressed as non-linear transfer functions and are derived from Gaussian bell curves. An output layer, which outputs neurons in turn have a linear transfer function and makes it possible to calculate the optimum weights associated with these neurons. The RBF network receives a k dimensional input vector X and outputs a scalar value using the general formula:

$$y = f(x) = w_0 + \sum_{i=1}^N w_i \varphi_i(X)$$

Where,  $w_0$  is a bias and  $w_i$  (1, 2 . . . n) are weight values, N represents the number of nodes in the hidden layer,  $\varphi_i(X)$  is a Gaussian function with the centre u and radius  $\sigma$ , i.e.,

$$\varphi_i(X) = \exp\left(\frac{-\|X - u_i\|^2}{\sigma_i^2}\right)$$

### RBF neural network design for cytotoxicity prediction

The RBF neural network has two operating modes, namely training and testing. Training an RBF network involves determining the number of hidden RBFs, the corresponding centres and the output layer weight matrix with criterion to minimize the mean squared errors (MSE) defined as  $E = \frac{1}{N} \sum_{i=1}^N [y(k) - \hat{y}(k)]^2$ .

a) Determination of number of hidden RBF units: The number of hidden nodes was determined experimentally in this research. Various numbers of hidden nodes were used, and the network with the best predictive capability was selected.

b) Determination of centre of the RBF unit: The k-means clustering is the most common algorithm which was used in this work. Its implementation process is shown briefly as follows:

Step 1: Initialization. Choose random values for the initial centers.

Step 2: Allocate each point to nearest center. Cluster m' includes all the points  $x_n$ ,

$$\text{where } \|X_n - C_{m'}\| = \min_m \|X_n - C_m\|$$

Step 3: Update each center. The new center value is the mean value of all the points in one cluster.

Step 4: Repeat step 2 until centres do not move.

c) Determination of the width of the RBF unit: After the centers of the RBF units have been established, the width of each RBF unit can be determined. The width of an RBF unit is selected as the root mean-square distance to the nearest  $p$ th RBF node, where  $p$  is the value of the overlapping parameter for the nearest neighbor. For the  $i$ th node, the  $\sigma_i$  is given by the following equation:

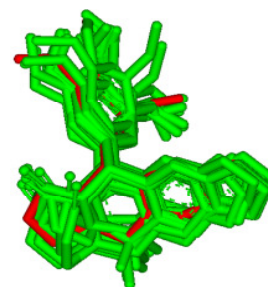
$$\sigma_i = \sqrt{\frac{1}{p} \sum_{k=1}^p \|u_k - u_i\|^2}$$

A typical value for  $P$  is suggested as 2

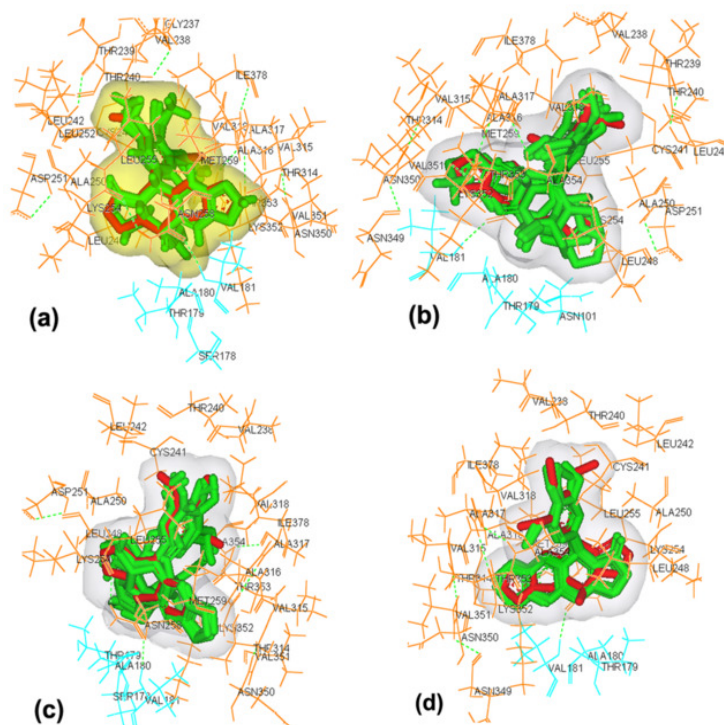
d) Determination of the output layer weights: Once the centers and the widths have been chosen, a supervised learning algorithm is applied to train the weights between the hidden and the output layer nodes. The output layer weights are determined using the least mean squares (LMS) algorithm in the present work.

## RESULTS

The original crystal structure of tubulin–podophyllotoxin complex PDB\_ID:1SA1 was used to validate Glide XP docking protocol. Fig.1 represents the Superposition of all the docked configurations of podophyllotoxin on crystal structure (red-stick). Fig.2(a–d) represents the Superposition of podophyllotoxin analogues (five analogues) belonging to (a) tetraline lactones, (b) non-lactonic tetralines, (c) pyrazoline and isoxazoline derivatives and (d) aza-derivatives within binding site of tubulin along with the co-crystal podophyllotoxin (red color). The table 2 represents the validation set along with their experimental activity expressed as the IC50 value for tubulin polymerization inhibition (TPI).

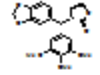
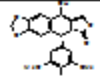
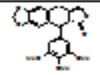
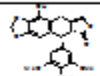
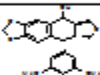
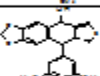
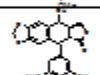
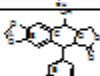
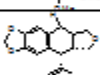
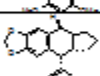
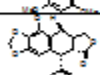
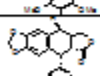

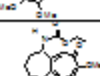
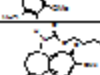


**Fig1.** Superposition of all the docked configurations of podophyllotoxin on crystal structure (red-stick). RMSD (heavy atoms) = 0.02–0.85Å<sup>17</sup>.



**Fig2. (a–d).** Superposition of podophyllotoxin analogues (five 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)<sup>17</sup>.

**Table 2.** The validation set along with their experimental activity expressed as the IC<sub>50</sub> value for tubulin polymerization inhibition (TPI)<sup>17</sup>.

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 sulphide		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	β-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 3.** Comparison of cytotoxic activities of podophyllotoxin analogues obtained by LIE, GA and RBFNN with the experimental activity for the training set.

Ligand	<sup>1</sup> ⟨Uele⟩	<sup>1</sup> ele Std Dev	<sup>3</sup> ⟨ele⟩ opt <sub>GA</sub>	<sup>1</sup> ⟨Uvdw⟩	<sup>1</sup> vdw Std Dev	<sup>3</sup> ⟨vdw⟩ opt <sub>GA</sub>	<sup>1</sup> ⟨Ucav⟩	<sup>2</sup> ⟨Ucav⟩ Std Dev	<sup>3</sup> ⟨Cav⟩ opt <sub>GA</sub>	<sup>4</sup> pIC <sub>50</sub> , <sub>Expt</sub>	<sup>5</sup> pIC <sub>50</sub> , <sub>LIE</sub>	<sup>6</sup> pIC <sub>50</sub> , <sub>GA</sub>	<sup>7</sup> pIC <sub>50</sub> , <sub>RBFNN</sub>
1	11.7	8.2	8.0	-41.4	-33.5	-32.5	3.8	0.4	0.4	1.9	1.4	1.9	1.9
2	12.4	8.9	8.6	-42.0	-34.1	-38.3	4.3	0.9	0.6	2.0	1.8	2.0	2.0
3	12	8.5	8.3	-44.7	-36.8	-36.7	4.2	0.8	0.6	2.0	1.5	2.0	2.0
5	10.5	7.0	6.1	-52.5	-44.6	-47.6	3.8	0.4	0.2	0.2	0.5	0.2	0.2
7	10.5	7.0	6.4	-48.5	-40.6	-39.1	4.1	0.7	0.7	1.2	1.0	1.2	1.2
8	13.7	10.2	6.9	-43.6	-35.7	-34.8	3.8	0.4	0.4	1.3	1.5	1.3	1.3
9	9	5.5	6.0	-35.9	-28.0	-25.0	2.4	-1.0	0.0	1.2	0.4	1.1	1.2
11	11.4	7.9	6.5	-47.4	-39.5	-43.9	3.8	0.4	0.1	0.2	1.0	0.5	0.2
13	10.4	6.9	6.5	-49.2	-41.3	-38.0	2.1	-1.3	-1.2	-0.4	-0.6	-0.4	-0.4
15	13.9	10.4	9.0	-57.4	-49.5	-55.0	3.8	0.4	0.1	0.7	0.5	0.9	0.7
16	10.4	6.9	6.9	-53.4	-45.5	-41.1	4.2	0.8	0.4	1.0	0.8	1.0	1.0
18	10.6	7.1	7.1	-63.8	-55.9	-54.1	2.3	-1.1	-1.0	-0.8	-1.4	-0.8	-0.8
19	12.9	9.4	8.0	-58.9	-51.0	-54.7	4.2	0.8	0.3	0.3	0.6	0.7	0.3
21	11.5	8.0	7.2	-53.7	-45.8	-41.4	3.5	0.1	0.1	0.9	0.3	0.9	0.9
23	11	7.5	6.5	-57.8	-49.9	-46.9	4	0.6	0.2	0.4	0.4	0.4	0.4
25	11.5	8.0	8.0	-54.2	-46.3	-47.7	3.7	0.3	0.1	0.9	0.4	0.9	0.9
26	11.4	7.9	8.9	-49.4	-41.5	-36.6	3.6	0.2	0.3	1.9	0.6	1.9	1.9
28	10.5	7.0	5.7	-56.2	-48.3	-46.2	2.9	-0.5	-0.4	-0.4	-0.5	-0.4	-0.4
29	7.1	3.6	4.3	-58.6	-50.7	-45.4	3.4	0.0	0.2	-0.4	-0.6	-0.4	-0.4
30	8.3	4.8	4.9	-44.4	-36.5	-32.2	2.8	-0.6	-0.6	-0.2	0.0	-0.2	-0.2
32	6.8	3.3	4.4	-54	-46.1	-45.9	2.1	-1.3	-0.7	-1.2	-1.3	-1.2	-1.2
33	9	5.5	5.2	-61.8	-53.9	-48.3	3.1	-0.3	-0.5	-0.8	-0.9	-0.8	-0.8
34	7.9	4.4	5.0	-52.6	-44.7	-44.1	2.2	-1.2	-0.9	-1.0	-1.0	-1.0	-1.0
36	11.8	8.3	7.3	-57.2	-49.3	-54.7	2.8	-0.6	-0.8	-0.6	-0.5	-0.8	-0.6
37	14.3	10.8	7.0	-46.7	-38.8	-38.6	4.4	1.0	0.2	1.0	1.7	1.0	1.0
38	11.8	8.3	7.5	-42.4	-34.5	-39.4	3.7	0.3	0.0	0.7	1.3	1.0	0.7
40	10.3	6.8	6.8	-51	-43.1	-39.2	1.7	-1.7	-2.0	-1.0	-1.1	-1.0	-1.0
41	11.4	7.9	7.0	-44.8	-36.9	-39.1	4.3	0.9	0.4	0.9	1.5	1.2	0.9
42	6.6	3.1	4.5	-51.4	-43.5	-41.4	2.2	-1.2	-0.8	-1.0	-1.1	-1.0	-1.0
44	9.2	5.7	6.2	-48	-40.1	-42.1	1.8	-1.6	-1.4	-0.9	-0.9	-0.9	-0.9
45	7.6	4.1	4.7	-48.5	-40.6	-36.9	1.8	-1.6	-1.1	-1.0	-1.1	-1.0	-1.0
46	8	4.5	5.2	-57	-49.1	-47.0	2.2	-1.2	-1.0	-1.2	-1.3	-1.2	-1.2
48	9.2	5.7	4.5	-56.8	-48.9	-45.9	1.7	-1.7	-1.0	-1.4	-1.5	-1.4	-1.4
49	11.8	8.3	7.6	-50.7	-42.8	-46.6	3.3	-0.1	-0.4	0.0	0.4	0.3	0.0
51	12.5	9.0	8.5	-46.8	-38.9	-43.5	4	0.6	0.2	0.7	1.2	1.3	0.7
52	8.3	4.8	5.0	-47	-39.1	-41.5	2.8	-0.6	-0.2	-0.3	-0.1	-0.3	-0.3
54	11	7.5	5.1	-55.5	-47.6	-41.0	2.8	-0.6	-0.7	-0.6	-0.4	-0.6	-0.6
55	11.4	7.9	6.9	-59.4	-51.5	-54.2	2.1	-1.3	-1.4	-1.2	-1.2	-1.2	-1.2
56	12.3	8.8	7.5	-54.1	-46.2	-48.6	2.2	-1.2	-1.5	-0.8	-0.7	-0.8	-0.8
57	14	10.5	8.0	-51.9	-44.0	-49.2	4.7	1.3	1.0	1.5	1.6	1.6	1.5
59	14.5	11.0	9.0	-52.1	-44.2	-50.0	4.4	1.0	0.5	0.8	1.4	1.5	0.8
60	13.3	9.8	8.0	-46.4	-38.5	-43.9	3.3	-0.1	-0.4	0.3	0.8	0.6	0.3
62	11.1	7.6	4.4	-59.2	-51.3	-47.0	2.8	-0.6	-0.1	-0.7	-0.7	-0.7	-0.7
64	12.1	8.6	5.4	-54.2	-46.3	-44.2	2.7	-0.7	-0.5	-0.5	-0.3	-0.5	-0.5
65	14.8	11.3	7.0	-51.8	-43.9	-50.0	4.8	1.4	1.0	1.0	1.8	1.1	1.0
66	13.2	9.7	7.0	-53.7	-45.8	-49.3	4.2	0.8	0.2	0.5	1.0	0.5	0.5
68	14.8	11.3	10.0	-56.2	-48.3	-55.5	4.3	0.9	0.3	0.5	1.1	1.4	0.5
69	12.7	9.2	6.3	-59.5	-51.6	-51.4	2.6	-0.8	-0.8	-0.8	-0.7	-0.8	-0.8
70	6.9	3.4	4.2	-55.4	-47.5	-43.7	2.4	-1.0	-0.6	-1.0	-1.1	-1.0	-1.0

71	11.8	8.3	7.0	-47.1	-39.2	-45.0	2.4	-1.0	-1.3	-0.6	-0.1	-0.6	-0.6
73	5.6	2.1	3.0	-50.8	-42.9	-37.5	2.7	-0.7	-0.3	-0.7	-0.7	-0.9	-0.7
74	9.7	6.2	5.4	-53	-45.1	-48.8	3.2	-0.2	-0.4	-0.7	-0.1	-0.7	-0.7
75	9.5	6.0	5.3	-55.3	-47.4	-49.2	3.4	0.0	-0.4	-0.7	-0.1	-0.7	-0.7
77	11.2	7.7	4.2	-65.4	-57.5	-57.5	3.4	0.0	0.3	-0.9	-0.6	-0.9	-0.9
78	5.8	2.3	3.5	-50.6	-42.7	-38.0	2.1	-1.3	-1.0	-1.0	-1.2	-1.4	-1.0
80	4.4	0.9	3.0	-44.5	-36.6	-30.3	2.7	-0.7	-0.3	-0.6	-0.5	-0.6	-0.6
82	9.6	6.1	6.2	-63.2	-55.3	-58.3	2.8	-0.6	-0.5	-0.9	-1.1	-0.9	-0.9
84	10.8	7.3	5.6	-47.3	-39.4	-35.6	2.3	-1.1	-1.3	-0.7	-0.3	-0.7	-0.7
85	6.1	2.6	3.5	-60.6	-52.7	-55.0	3	-0.4	-0.7	-2.0	-1.2	-2.0	-2.0
87	6.7	3.2	2.8	-54.6	-46.7	-43.8	2.3	-1.1	-1.1	-2.0	-1.2	-2.0	-2.0
89	4.9	1.4	2.5	-48.2	-40.3	-38.0	3.3	-0.1	0.4	-0.5	-0.2	-0.5	-0.5
91	5.1	1.6	2.6	-53.8	-45.9	-49.7	2.5	-0.9	-0.5	-1.9	-1.2	-1.9	-1.9
92	8.9	5.4	4.8	-54	-46.1	-49.9	2.2	-1.2	-1.4	-1.8	-1.0	-1.8	-1.8
94	7.7	4.2	4.8	-48.6	-40.7	-38.9	2.4	-1.0	-1.3	-1.2	-0.6	-1.2	-1.2
95	6.3	2.8	2.9	-50.3	-42.4	-41.2	2.8	-0.6	-0.2	-1.1	-0.5	-1.1	-1.1
97	4.4	0.9	0.9	-66.5	-58.6	-57.5	3.6	0.2	0.4	-2.1	-1.2	-2.1	-2.1
99	12.9	9.4	8.9	-41.8	-33.9	-32.7	4.3	0.9	0.6	2.4	1.9	2.4	2.4
100	11.9	8.4	7.7	-36.8	-28.9	-34.2	5	1.6	1.6	2.7	2.6	2.7	2.7
102	5.2	1.7	2.4	-52.3	-44.4	-39.9	4.5	1.1	1.5	0.3	0.5	0.3	0.3
103	1.8	-1.7	0.8	-53.9	-46.0	-44.3	4.9	1.5	1.8	0.3	0.4	-0.3	0.3
105	9.8	6.3	6.8	-45.9	-38.0	-35.5	4.6	1.2	0.8	1.6	1.5	1.6	1.6
107	8.2	4.7	4.8	-57.1	-49.2	-52.4	5.1	1.7	1.8	0.9	1.0	0.9	0.9
108	8.6	5.1	6.1	-37	-29.1	-26.6	4.6	1.2	1.5	2.4	2.0	2.4	2.4
110	9.4	5.9	5.9	-42.4	-34.5	-33.1	4.9	1.5	1.8	2.3	2.0	2.3	2.3
111	8	4.5	5.4	-48.9	-41.0	-36.4	5.6	2.2	2.1	2.2	1.9	2.2	2.2
112	8.5	5.0	7.1	-46.5	-38.6	-39.4	4.9	1.5	1.1	1.8	1.6	1.8	1.8

<sup>1</sup>⟨Uele⟩ std Dev, <sup>1</sup>⟨Uvdw⟩ std Dev, <sup>1</sup>⟨Ucav⟩ std Dev refers to the normalized data. <sup>2</sup>(ele) opt<sub>GA</sub>, <sup>2</sup>(vdw) opt<sub>GA</sub>, <sup>2</sup>(Cav) Opt<sub>GA</sub> refer to the optimized electrostatic van der Waals and Cavity energy parameters with GA <sup>3</sup>pIC<sub>50,Expt</sub> Refers to the experimental biological activity of the podophyllotoxin analogues. <sup>4</sup>pIC<sub>50,LIE</sub> Refers to the free

energy of the podophyllotoxin analogues calculated using LIE approach<sup>17</sup>. <sup>5</sup>pIC<sub>50,GA</sub> Refers to the Biological activity of the podophyllotoxins calculated using GA. <sup>6</sup>pIC<sub>50,ANN</sub> Refers to the Biological activity of the podophyllotoxins predicted by ANN.

**Table 4.** Comparison of cytotoxic activities of podophyllotoxin analogues obtained by LIE, GA and RBFNN with the experimental activity for the test set.

Ligand	<sup>1</sup> ⟨Uele⟩	<sup>1</sup> ele Std Dev	<sup>2</sup> (ele) opt <sub>GA</sub>	<sup>1</sup> ⟨Uvdw⟩	<sup>1</sup> vdw Std Dev	<sup>2</sup> (vdw) opt <sub>GA</sub>	<sup>1</sup> ⟨Ucav⟩	<sup>2</sup> ⟨Ucav⟩ Std Dev	<sup>2</sup> (Cav) opt <sub>GA</sub>	<sup>3</sup> pIC <sub>50,Ext</sub>	<sup>4</sup> pIC <sub>50,LIE</sub>	<sup>5</sup> pIC <sub>50,GA</sub>	<sup>6</sup> pIC <sub>50,RBFNN</sub>
4	11	8.6	8.9	-49.5	-40.3	-40.6	3.8	0.2	0.1	1.6	0.8	1.6	1.6
6	9.2	6.8	6.7	-44.9	-35.7	-35.4	3.3	-0.3	0.4	1.2	0.5	1.2	1.0
10	10.9	8.5	6.6	-53.3	-44.1	-48.5	3.7	0.1	0.0	0.2	0.5	0.2	0.1
12	13.7	11.3	9.0	-53.7	-44.5	-49.9	2.4	-1.2	-1.5	-0.3	-0.4	-0.3	-0.4
14	11.6	9.2	7.0	-49.2	-40.0	-35.0	2.3	-1.3	-1.5	-0.3	-0.3	-0.3	-0.3
17	11.9	9.5	8.6	-46.7	-37.5	-41.7	3	-0.6	-0.8	0.6	0.4	0.6	0.3
20	10.6	8.2	7.6	-47.2	-38.0	-41.5	2.5	-1.1	-1.1	0.0	-0.2	0.0	-0.1
22	10.2	7.8	7.8	-52.2	-43.0	-45.7	3.5	-0.1	0.2	1.0	0.2	1.0	0.9
24	13.8	11.4	6.5	-50.9	-41.7	-45.0	4	0.4	0.1	0.3	1.1	0.4	0.3

27	12.4	10.0	6.3	-62.7	-53.5	-56.9	3.3	-0.3	-0.9	-1.1	-0.4	-1.1	-1.1
31	10.3	7.9	7.7	-54.4	-45.2	-45.3	2	-1.6	-2.0	-1.0	-1.1	-1.0	-1.1
35	9.4	7.0	7.4	-54.1	-44.9	-48.4	2.2	-1.4	-1.7	-1.0	-1.0	-1.0	-1.1
39	10.3	7.9	6.6	-61.7	-52.5	-55.6	3.1	-0.5	-0.7	-0.8	-0.7	-0.8	-0.8
43	13.8	11.4	8	-53	-43.8	-48.4	3.4	-0.2	-0.7	0.1	0.5	0.1	0.0
47	10	7.6	7.1	-53.3	-44.1	-40.7	2.1	-1.5	-1.9	-0.9	-1.0	-0.9	-0.9
50	10.3	7.9	7.5	-46.9	-37.7	-41.6	1.5	-2.1	-2.0	-0.9	-0.9	-0.9	-1.0
53	12.4	10.0	8.0	-50.4	-41.2	-45.0	2.3	-1.3	-1.6	-0.5	-0.4	-0.5	-0.6
58	14.3	11.9	7.6	-49	-39.8	-44.5	3.9	0.3	-0.2	0.6	1.2	0.6	0.5
61	13.5	11.1	7	-58.8	-49.6	-53.1	4.2	0.6	0.1	0.2	0.7	0.2	0.1
63	13.5	11.1	7	-58.5	-49.3	-54.2	3.8	0.2	0.0	0.0	0.4	0.0	0.0
67	12.4	10.0	7.5	-56.5	-47.3	-54.2	3.3	-0.3	-0.6	-0.3	0.0	-0.3	-0.4
72	1.8	-0.6	3.4	-42.8	-33.6	-30.1	2.5	-1.1	-0.5	-0.6	-0.8	-0.6	-0.5
76	10.6	8.2	7.1	-63.8	-54.6	-51.0	1.9	-1.7	-1.4	-1.0	-1.8	-1.0	-1.1
79	7.8	5.4	6.3	-49.3	-40.1	-38.0	2.7	-0.9	-1.4	-0.7	-0.4	-0.7	-0.7
81	9.1	6.7	6.4	-60.4	-51.2	-45.7	2	-1.6	-1.5	-1.1	-1.5	-1.1	-1.1
83	10.4	8.0	6.9	-44.8	-35.6	-36.6	2.4	-1.2	-1.7	-0.6	-0.1	-0.6	-0.6
86	5.2	2.8	1.5	-63.6	-54.4	-56.0	3	-0.6	-0.2	-2.3	-1.4	-2.3	-2.2
88	7.4	5.0	4.2	-59	-49.8	-51.8	2.8	-0.8	-1.0	-1.8	-1.0	-1.8	-1.8
90	5.9	3.5	4.1	-56.2	-47.0	-48.7	3.3	-0.3	-0.5	-1.2	-0.6	-1.2	-1.1
93	2.8	0.4	1.9	-52.7	-43.5	-39.0	2.8	-0.8	-0.6	-1.7	-1.1	-1.7	-1.6
96	3.6	1.2	2.0	-49.5	-40.3	-37.0	2.2	-1.4	-1.1	-2.0	-1.3	-2.0	-1.9
98	2.7	0.3	1.6	-47.8	-38.6	-33.0	2.3	-1.3	-1.0	-1.9	-1.2	-1.9	-1.8
101	2	-0.4	1.6	-61.1	-51.9	-48.7	4.6	1.0	1.2	-0.7	-0.3	-0.7	-0.5
104	2.1	-0.3	3.0	-46	-36.8	-34.0	3.5	-0.1	0.1	-0.4	-0.1	-0.4	-0.3
106	7.7	5.3	5.6	-52.1	-42.9	-42.2	2.7	-0.9	-0.2	-0.1	-0.6	-0.1	-0.1
109	6.3	3.9	3.8	-52	-42.8	-41.3	5.1	1.5	2.0	1.2	1.1	1.2	1.2

<sup>1</sup>⟨Uele⟩ std Dev, <sup>1</sup>⟨Uvdw std Dev, <sup>1</sup>⟨Ucav⟩ std Dev refers to the normalized data. <sup>2</sup>(ele) opt<sub>GA</sub>, <sup>2</sup>(vdw) opt<sub>GA</sub>, <sup>2</sup>(Cav) Opt<sub>GA</sub> refer to the optimized electrostatic van der Waals and Cavity energy parameters with GA <sup>3</sup>pIC50<sub>Expt</sub> Refers to the experimental biological activity of the podophyllotoxin analogues. <sup>4</sup>pIC50<sub>LIE</sub> Refers to the free

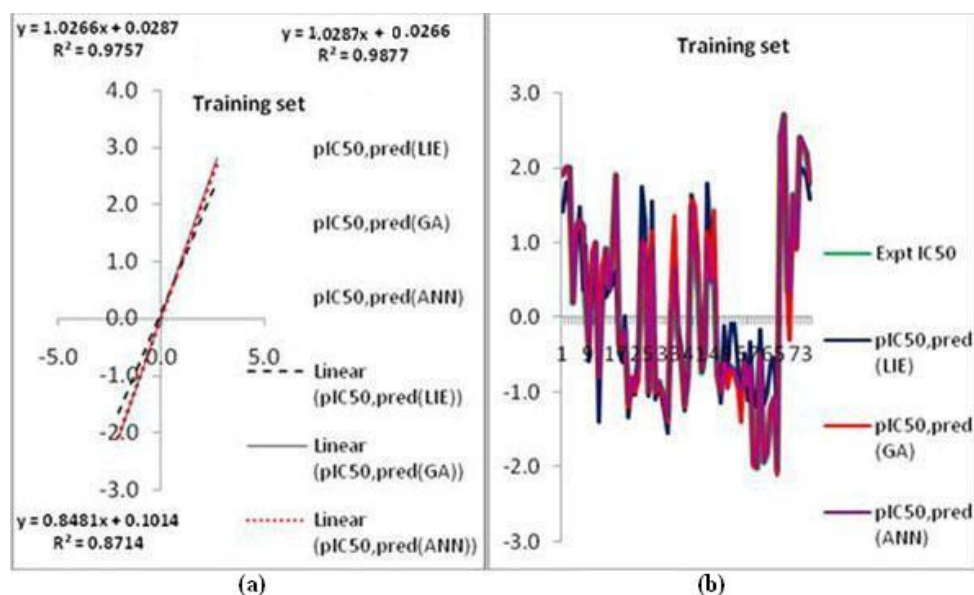
energy of the podophyllotoxin analogues calculated using LIE approach<sup>17</sup>. <sup>5</sup>pIC50<sub>GA</sub> Refers to the Biological activity of the podophyllotoxins calculated using GA. <sup>6</sup>pIC50<sub>ANN</sub> Refers to the Biological activity of the podophyllotoxins predicted by ANN.

**Table 5.** Comparison of cytotoxic activities of podophyllotoxin analogues obtained by LIE, GA and RBFNN with the experimental activity for validation set

Ligand	<sup>1</sup> ⟨Uele⟩	<sup>1</sup> ele Std Dev	<sup>3</sup> (ele) opt <sub>GA</sub>	<sup>1</sup> ⟨Uvdw⟩	<sup>1</sup> vdw Std Dev	<sup>3</sup> (vdw) opt <sub>GA</sub>	<sup>1</sup> ⟨Ucav⟩	<sup>2</sup> ⟨Ucav⟩ Std Dev	<sup>3</sup> (Cav) opt <sub>GA</sub>	<sup>4</sup> pIC <sub>50,Expt</sub>	<sup>5</sup> pIC <sub>50,LIE</sub>	<sup>6</sup> pIC <sub>50,GA</sub>	<sup>7</sup> pIC <sub>50,RBFNN</sub>
1	13.4	7.5	8.0	-46.3	-36.0	-35.0	4.3	0.2	0.3	-7.7	1.6	1.7	-0.3
2	14	8.1	3.5	-56.3	-46.2	-62.4	4.1	0.4	0.5	-1.3	-0.9	-1.3	-1.1
3	11.8	5.9	7.6	-60.1	-50.0	-47.5	3.3	-0.4	-0.4	0.3	0.1	0.3	-0.5
4	12.1	6.2	6.6	-61.2	-51.1	-49.9	3.6	-0.1	0.1	0.2	-0.1	0.2	-0.6
5	10.8	4.9	7.1	-58.5	-48.4	-57.4	3.5	-0.2	0.1	0.0	-0.1	0.0	-1.1
6	8.3	2.4	5.1	-47.3	-37.2	-49.7	3.5	-0.2	-0.6	-1.0	-0.6	-1.0	-1.1
7	11.8	5.9	9.6	-63.7	-53.6	-51.7	3.5	-0.2	-1.0	0.3	0.0	0.3	-0.5
8	11.6	5.7	8.1	-56.2	-46.1	-49.9	3.2	-0.5	-0.5	0.2	0.2	0.2	-0.5
9	10.8	4.9	5.1	-57.4	-47.3	-50.1	4	0.3	0.6	0.1	0.3	0.1	-1.1
10	10.7	4.8	6.3	-55.7	-45.6	-43.5	2.5	-1.2	-1.2	-0.7	-0.7	-0.7	-1.0
11	11.5	5.6	7.0	-60.7	-50.6	-53.7	4	0.3	0.2	0.3	0.2	0.3	-0.6
12	14.1	8.2	10.0	-63.4	-53.3	-66.6	4.2	0.5	0.1	0.7	0.4	0.7	-0.5
13	13.6	7.1	5.3	-53.4	-43.1	-50.1	6	1.9	0.1	-0.3	2.5	-0.3	-1.1
14	15.1	9.2	7.9	-67.1	-57.0	-68.9	6	2.3	0.3	-0.1	1.8	-0.1	-1.1
15	8.1	2.2	6.1	-56.3	-46.2	-61.2	6.1	2.4	0.6	-0.1	1.8	-0.1	-1.1

<sup>1</sup>⟨Uele⟩ std Dev, <sup>1</sup>⟨Uvdw⟩ std Dev, <sup>1</sup>⟨Ucav⟩ std Dev refers to the normalized data. <sup>2</sup>(ele) opt<sub>GA</sub>, <sup>2</sup>(vdw) opt<sub>GA</sub>, <sup>2</sup>(Cav) Opt<sub>GA</sub> refer to the optimized electrostatic van der Waals and Cavity energy parameters with GA. <sup>3</sup>pIC<sub>50,Expt</sub> Refers to the experimental biological activity of the podophyllotoxin analogues. <sup>4</sup>pIC<sub>50,LIE</sub> Refers to the free energy of the podophyllotoxin analogues calculated using LIE approach<sup>17</sup>. <sup>5</sup>pIC<sub>50,GA</sub> Refers to the Biological activity of the podophyllotoxins calculated using GA. <sup>6</sup>pIC<sub>50,ANN</sub> Refers to the Biological activity of the podophyllotoxins predicted by ANN. The quality of the fit can be judged by the value of the squared correlation coefficient R<sup>2</sup> which was 0.9757 of

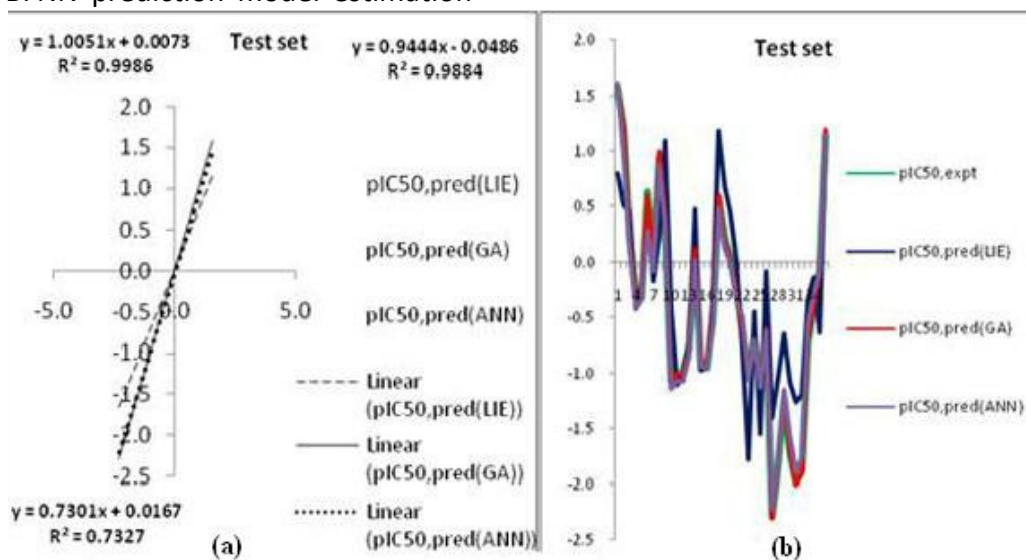
GA method and 0.9877 of RBFNN method (Figure 3) with respect to 0.8714 of LIE approach which graphically represents the robustness of the binding affinity model developed using the training set. We applied the Prediction model (Constructed using RBFNN with the GA optimized data) to the podophyllotoxin analogues comprising the test set. The test set includes 36 compounds categorized into four subgroups. The analogues comprising the test set were also obtained from the same source<sup>17</sup>. Since the experimental values of IC<sub>50</sub> for these inhibitors are already available, this set of molecules provides an excellent data set for testing the prediction power of the present model for new ligands.



**Fig.3.** Showing the squared correlation coefficients (a) and patterns of activity values of GA, RBFNN and LIE method for 76 podophyllotoxin analogues comprising the training set plotted against corresponding experimental data (b).

The squared correlation coefficient of GA and RBFNN estimates for the free energy for the test set compounds is  $R^2 = 0.9986$  of GA,  $R^2 = 0.9884$  of RBFNN method with respect to LIE approach ( $R^2 = 0.7337$ ) in Figure 4 which graphically shows the quality of fit to evaluate the accuracy of the GA/RBFNN prediction model estimation

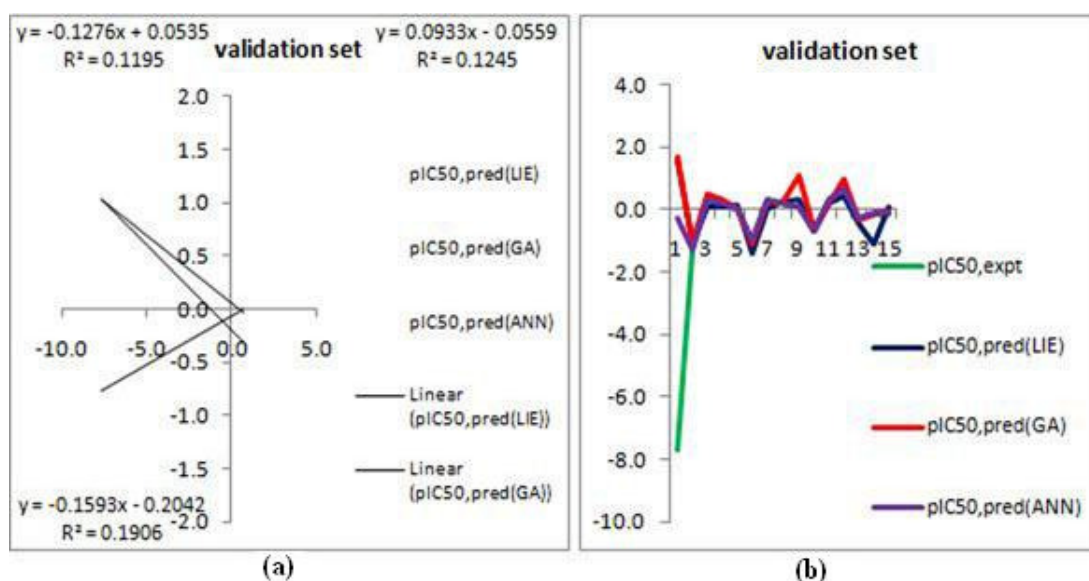
for tubulin polymerization inhibition potencies. we have taken a separate data set called as validation set consisting of 15 analogues of podophyllotoxin with the known  $IC_{50}$  values (Table 2).The same model has been used to calculate the  $pIC_{50}$  values of these analogues.



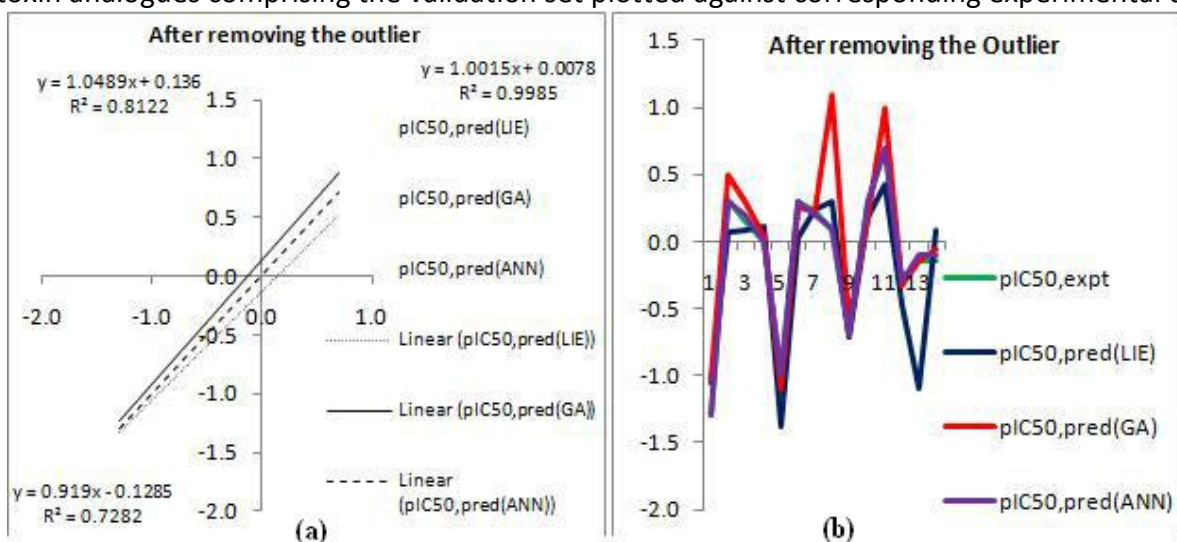
**Fig.4.** Showing the squared correlation coefficients (a) and patterns of activity values of GA, RBFNN and LIE method for 36 podophyllotoxin analogues comprising the test set plotted against corresponding experimental data (b).

In order to explore the reliability of the present method the G4 podophyllotoxin present as a first analogue in table 2 representing the validation set is removed

through Leave one out (LOO) and the  $R^2$  values for the three sets of data obtained through LIE, GA, and RBFNN with respect to the experimental data set are calculated.



**Fig.5.** Showing the squared correlation coefficients (a) and patterns of activity values of GA, RBFNN and LIE method for 15 podophyllotoxin analogues comprising the validation set plotted against corresponding experimental data (b).



**Fig.6.** Showing the squared correlation coefficients (a) and patterns of activity values of GA, RBFNN and LIE method for 14 podophyllotoxin analogues after removing the outlier comprising the validation set plotted against corresponding experimental data (b)

## DISCUSSION

Neural networks and genetic algorithms are versatile methods for a variety of tasks in rational drug design, including analysis of structure–activity data, establishment of quantitative structure–activity relationships. Hybrid algorithms that combine GAs and neural networks appear in the literature as promising methods for strengthening the impact of computational methods in drug design<sup>21</sup>. However, the application of neural networks and GAs requires some essential knowledge of these methods before it can be properly

employed. One of the studies says more potent dihydrofolate reductase inhibitors by solving the neural network inversion problem for QSAR with a GA<sup>23</sup>. They built a QSAR model based on a dataset of phenyl substituted diaminodihydrotriazines, with a small fully-connected feed-forward/back-propagation neural network, and determined the maximum activity on the structure– activity surface using a GA. Owing to the molecular representation of the compounds (related to the physicochemical properties of the substituents on the phenyl ring), the GA enables prediction of the

required molecular properties for higher activity molecules. A variety of the application fields for self organizing neural networks in drug design, including the mapping of a molecular surface into a 2D plane<sup>25</sup>. And finally, Schneider, (2000) have reached the conclusion that neural networks and GA are valuable tools for drug design and have the potential for further development.

## CONCLUSION

Speeding up drug discovery and development is of central interest in all pharmaceutical companies. Applying the mathematical and computational approaches in drug discovery is one of the finest methods which provide results with good accuracy in less time. In this work, the three methods for predicting the biological activity of the podophyllotoxin analogues are compared. The pIC<sub>50</sub> calculated based upon LIE method is compared with the results obtained through the GA and RBF neural network. The optimized data is used to build a prediction model and the results were compared with the Experimental IC<sub>50</sub> values. The results obtained have proved that the Genetic Algorithm and Neural networks 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 prediction activity based upon cytotoxic activity of these analogues to tubulin have directly correlated with the experimental potency of these inhibitors by using LIE, GA, and RBF neural network which can be seen through regression correlation coefficient ( $R^2$ ) value and quality of fit. Among these methods RBF neural network give the most prominent result which overlaid to the exact experimental activity. These methods have reproduced experimental data with reasonably small error for the majority of podophyllotoxin analogues.

## ACKNOWLEDGEMENT

We are thankful to Dr. Patrik Gomez and Dr. Janet Vaniella for providing a very nice working atmosphere and resources.

## REFERENCES

1. Shibuya K Mathers CD Boschi-Pinto C Lopez AD Murray CJ, Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease. *Cancer*, 26, 2002, 2–37.
2. Nasmyth K, Segregating sister genomes: the molecular biology of chromosome separation. *Science*, 297, 2002, 559–565,.
3. Stewart ZA Matthew D Westfall, Jennifer A, Cell-cycle dysregulation and anticancer therapy. *Tren. Pharmacol. Sci*, 24, 2003, 139–145.
4. Drukman S, Kavallaris M, Microtubule alterations and resistance to tubulin-binding agents. *Int. J. Oncol*, 21, 2002, 621–628.
5. Newman JD Cragg GM, Snader KM, Natural products as sources of new drugs over the period 1981. *J. Nat. Prod*, 66, 2002, 1022-1037.
6. Alam MA, Naik PK, Molecular modelling evaluation of the cytotoxic activity of podophyllotoxin analogues. *J Comp. Aid. Mo.l Des*, 23, 2008, 209-225.
7. Amos LA, Focusing-in on microtubules. *Curr. Opin. Struct. Biol.* 10, 2000, 236-241.
8. Wang LG, Liu XM, Kries W, Budman DR, The effect of antimicrotubule agents on signal transduction pathways of apoptosis: a review. *Can. Chemo. Pharmaco*, 44, 199, 355-36.
9. Shi Q, Chen K, Morrisnatschke SL, Lee KH, Recent progress in the development of tubulin inhibitors as antimitotic antitumor agents. *Curr. Pharm.Des*, 4: 1998, 219-248.
10. Jordan MA, Margolis RL, Himes RH, Wilson L, Identification of distinct class of vinblastine binding sites on Microtubules. *J. Mol. Biol*, 187, 1986, 61-73.
11. Nogales E, Wolf SG, Downing KH, Structure of the alpha beta tubulin dimer by electron crystallography. *Nature*, 391, 1998, 199- 203.
12. Uppuluri S, Knipling L, Sackett DL, Wolff J, Localization of the colchicine-binding site of tubulin. *Proc. Natl. Acad. Sci. USA*, 90, 1993, 11598-11602.

13. Snyder JA, McIntosh RJ, Biochemistry and physiology of microtubules. *Annu Rev Biochem*, 45, 1976, 699-720.
14. Cortese F, Bhattacharyya B, Wolff P, Podophyllotoxin as a probe for the colchicines binding site of tubulin. *J. Biol. Chem*, 252, 1977, 1134–1140.
15. Lin CM, Hamel E, Effects of inhibitors of tubulin polymerization on GTP hydrolysis *J. Biol. Chem*, 256, 1981, 9242–9245.
16. Ravelli RB, Gigant B, Curmi PA, Jourdain I, Lachkar S, Sobel A & Knossow M: Insight into tubulin regulation from a complex with colchicine and a stathminlike domain. *Nature*, 428, 2004, 198-202.
17. Alam MA, Naik PK, Applying linear interaction energy method for binding affinity calculations of podophyllotoxin analogues with tubulin using continuum solvent model and prediction of cytotoxic activity. *Jour. Mol. Graph and Mod*, 27, 2009, 930–943.
18. Buss AD, Waigh RD, Natural Products as Leads for new Pharmaceuticals. Wolff, M.E., Ed.; John Wiley & Sons, Inc: New York, Chapter 24.
19. Fujita T, Iwasa J, Hansch C, A new substituent constant,  $\pi$ , derived from partition coefficients. *J. Am. Chem. Soc*, 86, 1964, 5175–5180.
20. Hansch C, Hoekman D, Gao H, Comparative QSAR: toward a deeper understanding of chemobiological interactions. *Chem. Rev*, 96, 1996, 1045–1075.
21. Lothar T, Johann G, Neural networks and genetic algorithms in drug design. *Drug disc.tod*, 6, 2001, 102-107.
22. Aboozar K, Hamid M, Prediction of solubility of gases in polystyrene by Adaptive Neuro-Fuzzy Inference System and Radial Basis Function Neural Network. *Exp Sys with Appl*, 37, 2010, 3070–3074.
23. Burden FR, Winkler DA, Use of automatic relevance determination in QSAR studies using Bayesian neural networks. *J. Chem. Inf. Comput. Sci*, 40, 2000 1423–1430.
24. Schneider G, Neural networks are useful tools for drug design. *Neural Networks*, 13, 2000, 15–16.
25. Anzali S, Johann G, Ulrike H, Jaroslaw P, Jens S, Andreas T, The use of self-organizing neural networks in drug design. *3D QSAR in Drug Des*, 2: 1998, 273–299.

\*\*\*\*\*