

Evaluation of binding interaction of coumarin antifungals to bilipid membrane using Dock scoring function and Levenberg-Marquardt neural network

Seyedeh Soghra Mousavi^{1*}, Hanieh Bokharaie², Seyed Reza Mirhafez³, Seyed Mostafa Alavi³, Hiva Safari⁴, Zahra Mirzazadeh⁵ and Mehrdad Hamidi⁶

¹Department of Biotechnology, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran

²Genetic Group, Biology Department, Faculty of Basic Sciences, Science and Research Branch, Islamic Azad University, PO Box 14515/755, Tehran, Iran

³Departments of Cellular and Molecular Biology, School of Biology, College of Science, Tehran University, Tehran, Iran

⁴Department of Biotechnology, Tehran Payam e Noor University, Tehran, Iran

⁵Department of Natural Recourses and Fisheries, Azad University of Savadkoh, Mazndaran, Iran

⁶Departments of Pharmaceutics, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran

Received 18 January 2011; revised 5 October 2011; accepted 10 November 2011

Recently, investigation on natural antifungal resources has been increased due to vital need for brand new fungicidal agents. Neural network programs have lots of special features that make them suitable for handling complex problems like analyzing different properties of candidate compounds in computer-aided drug design. In the present study, Levenberg-Marquardt neural network (the fastest of the training algorithms) was used, and the relationship between some important thermodynamic and physico-chemical properties of coumarin compounds and their biological activities (measured by binding interaction energy to bilayer membrane of eukaryote cells) were evaluated. A set of already reported antifungal bioactive coumarin and some well known physical descriptors were selected and, by using Levenberg-Marquardt training algorithm, the best architecture of neural model was designed for predicting the effects of new bioactive compounds. Results revealed that the best architecture according to the term of calculation cycles and considering the correlating behaviour and output cycles of calculation was 19-7-6-1. In addition, the results revealed that the most sensitive input are Log P and molar refractivity. Descriptors, viz., surface tension, energy of LUMO and energy of HOMO were the most important inputs. The correlation coefficient between the observed and the interaction energy values was 0.9132. The study also showed that Dock scoring function can be used for modeling of coumarins antifungal bioactivity.

Keywords: Antifungal compounds, Dock scoring function, Levenberg-Marquardt neural network, interaction energy

Introduction

During the last 20 years, human fungal infections have increased at an alarming rate, mainly among immunocompromised individuals¹. Apparently it appears that there are great array of antifungal drugs, but the quest for new generation of antifungal compounds is still continuing due to the low efficacy, side effects or resistance associated to the existing drugs². At present, there are only a limited number of known clinically available antifungal agents like amphotericin B, ketoconazole, fluconazole and itraconazole that could be chosen for treatment³. However, these antifungal drugs have disadvantages

including high toxicity, ineffectiveness towards some fungi and low bioavailability. So they do not meet the needs of patients completely⁴.

Coumarin compounds are naturally occurring constituent of many plants, which contain a chromenone ring, often a chromen-2-1 or chromen-4-1 ring⁵. Some coumarins are known to have antifungal activity⁶. In one of the study, it has been shown that a limited number of coumarins are active against several fungi, viz., *Candida albicans*, *Cryptococcus neoformans*, *Saccharomyces cerevisiae* and *Aspergillus niger*⁷. However, several parameters are to be considered in designing a new coumarin antifungal^{8,9}. Recently, artificial neural networks (ANNs) have been used most widely in drug design. They usually consist of three or four input layers, one

*Author for correspondence:

Tel: (+98-241) 4273635; Fax: (+98-241) 4273639

E-mail: s-mousavi@hotmail.com

output layer and one or two hidden ones¹⁰. In pattern classification, understanding the class boundaries by the classifier requires a training phase with a training algorithm¹¹. Gradient-based training algorithms like back-propagation are not efficient due to the fact that the gradient vanishes in the solution¹². Using Hessian-based algorithms makes possible the network to learn more subtle features of a complicated mapping¹³. The training process converges quickly as the solution is approached, because the Hessian does not vanish at the solution¹⁴. The LM algorithm is basically a Hessian-based algorithm for nonlinear least squares optimization. Since they can find the complex relationship between predictor variables (inputs) and predicted variables (output), LM algorithm trained ANNs have received increased attention in drug discovery¹⁵. According to previous studies, interaction energy between eukaryote membrane and binding ligand could be considered as bioactivity of that ligand, because it has been proven that bioactivity and interaction energy with biomembrane are closely related and also change and behave in a same way¹⁶.

The objective of the present study was to build a LM neural network for a set of thermodynamic and physico-chemical properties of antifungal coumarins. For this reason, the best architecture according to the least error and cycle of calculation was selected

and correlation coefficient between thermodynamic and physico-chemical properties, and bioactivity of antifungal coumarins (tested against *Candida albicans* bilayer membrane) were calculated. Finally the role of these properties in bioactivity of coumarins antifungal was discussed.

Material and Methods

In the first step, some thermodynamic and physico-chemical descriptors for all congeners were computed or taken from the literature (Table 1)^{9,17-20}. Geometry optimization was carried out by using the semiempirical PM3 method²¹ implemented in HyperChemTM program package (HyperChemTM, 1996). All these descriptors were generated by different applications, such as, ACDLAB (11.02 released on 21 May, 2008), HyperChem (8.0.2) and MOPAC 93, together with the help of references (Table 1). For example, the basic thermodynamic properties, such as, standard enthalpy of formation, standard free enthalpy of formation, molar entropies, heat capacities, energies of HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) are extracted from MOPAC 93 data files. Also membrane model (Fig. 1) was drawn using ACDLAB (11.02), HyperChem (8.0.2) and ChemDraw (1998, V2). Drawn membrane is a fluid phospholipid bilayer embedded glycoproteins. The phospholipid bilayer is so arranged that the polar

Table 1—List of Descriptors used in this study

Descriptor	Parameter	References
1	Vicinal carbon atoms substitution pattern	9
2	Sasvol (Solvent-Accessible volume)	9
3	Vdvwol (Van der Waals Volume)	9
4	Symmetry of molecule	9
5	Maxq+ (the largest positive charge over the atoms in molecules)	9
6	Vapour pressure	9
7	Energy of HUMO (highest occupied molecular orbital)	9
8	Energy of LUMO (lowest unoccupied molecular orbital)	9
9	Molecular mass (Da)	9
10	Dipole moment of the molecule	17
11	Density (g/cm)	18
12	Retention time	18
13	Heat capacity	19
14	Standard enthalpy of formation	19
15	Specific polarizability of molecule	20
16	Molar refractivity (cm ³)	20
17	Molar volume (cm ³)	20
18	Log P	20
19	Surface tension (dyne/cm)	20

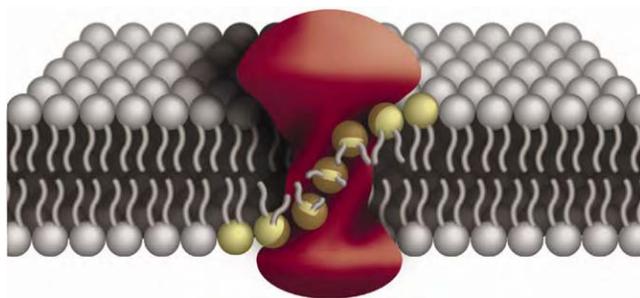


Fig 1—Diagram of membrane used in docking drawn by a set of mentioned software

ends of the molecules (the phosphate and glycerol portion of the phospholipid that is soluble in water) form the outermost and innermost surface of the membrane, while the non-polar ends (the fatty acid portion of the phospholipid that is insoluble in water) form the center of the membrane. For calculating the best distance, this membrane was optimized by using semiempirical methods applied in HyperChem software and the best optimized one was chosen as final model.

Docking calculation was carried out using Hex docking software (4.5, 2006). The grid resolution was 0.4 and augment root node was applied false. A single (rigid body) molecular mechanics energy calculated for each docking solution (IE Energies), and Newton-like energy minimization (MM Minimization) applied to each docking ligand. These energies were calculated using “soft” Lennard-Jones and hydrogen bond potentials, adapted from the OPLS force-field parameters, along with an explicit charge-charge electrostatic contribution. When docking complexes where conformational changes are known to be small, this gives an effective way to prune many “false-positive” orientations and to enhance the energy of the “right answer”. Interaction energies were used as bioactivity of coumarin compound in bilipid membrane of *Candida albicans* and they are shown in Table 2.

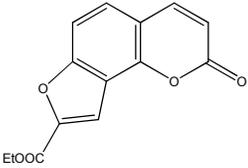
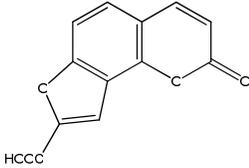
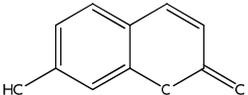
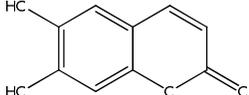
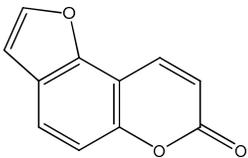
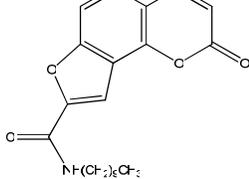
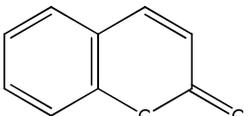
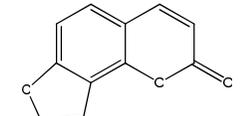
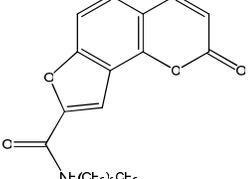
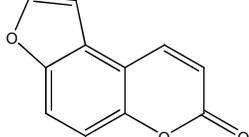
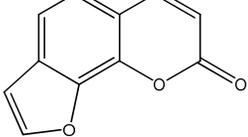
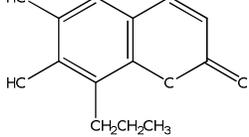
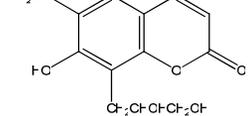
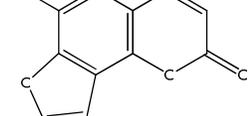
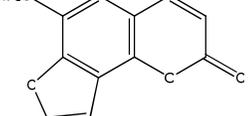
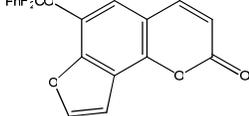
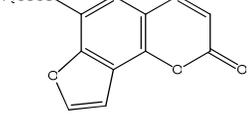
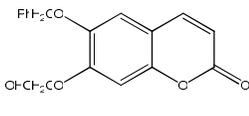
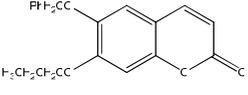
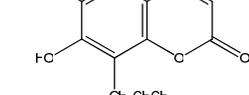
The dataset was composed of the coumarins and coumarin derivatives whose antifungal activities have been previously proved^{8,22-25}. The Error Back Propagation (EBP) algorithm has been a significant improvement in neural network research, but it has a weak convergence rate²⁶. Many efforts have been made to speed up EBP algorithm^{27,28}. However, all of these methods led to little acceptable results. The LM algorithm arises from the development of EBP algorithm dependent methods. It gives a good mix between the speed of Newton algorithm and

the stability of the steepest descent method²⁹ as both are the basic theorems of LM algorithm. In the present study, a feed forward neural network with LM algorithms was applied for modeling the bioactivity of coumarins antifungal. A standard feed-forward network with LM algorithms and 1 to 3 hidden layers architecture were chosen. For solving the problem of over-fitting, the number of neurons was kept at minimum³⁰. However, the optimum architecture with target error less than 0.01% was created with variation in the total number of nodes and hidden layers. This neural model was established based on NeuroSolutions (version v5.07, 2008). For validation of our model, 10 repetitions of the complete validation process with different random seeds were made in all the cases (Y-scrambling test), in order to analyze the influence of inherent randomness on the prediction stability. Accuracy has been selected for evaluation of predictive performance of a single validation process, while a coefficient of correlation (CO) of accuracies obtained across 10 repetitions and this was established as a measure of learning stability. Also cross-validation was applied by leave-n-out method.

Results

The computed basic physico-chemical and thermodynamic descriptors for coumarins are presented in Table 1. These are the most general descriptors that are picked up from the past QSAR researches and cover the binding properties of a ligand to a receptor. Some of these descriptors like molar refractivity, Log P, surface tension, density and molecular mass, mostly present the physical properties of docked ligand, while the rest cover the chemical behaviour of molecules. Various architectures of neural network, i.e., Y-scrambling R^2 , validation set error and calculation cycles, are shown in Table 3. Y-Scrambling R^2 indicates that the architecture is not over fitted by free weight in the network. There are significant differences between various architectures, which vary from 0.323 to 0.982. Previous results have shown that more hidden layers essentially do not lead to decreasing Y-Scrambling R^2 . Calculation cycles present the speed of network and deeply depend on the architecture of the network. More hidden layers increase the calculation cycles, but the balance between the number of hidden layers and free weight in network usually determine the number of calculation cycles. In the present study, LM trained ANN was used to build a neural model for

Table 2—Coumarin compounds and their calculated interaction energies

No.	Compound	Interaction energy (kcal/mol)	No.	Compound	Interaction energy (kcal/mol)
1		-4.34	2		-3.55
3		-4.56	4		-4.67
5		-3.87	6		-4.45
7		-5.78	8		-3.45
9		-4.66	10		-3.97
11		-5.42	12		-3.56
13		-3.34	14		-4.00
15		-4.12	16		-5.11
17		-3.77	18		-6.56
19		-4.54	20		-4.78

21		-4.87	22		-4.98
23		-3.78	24		-4.83
25		-2.55	26		-6.40
27		-3.45	28		-4.11
29		-4.99	30		-4.63
31		-4.28	32		-4.08
33		-4.55	34		-4.06
35		-5.21	36		-4.09
37		-3.44	38		-4.79
39		-2.97	40		-4.33
41		-3.99	42		-3.46
43		-4.73	44		-4.67
45		-3.90	46		-5.67

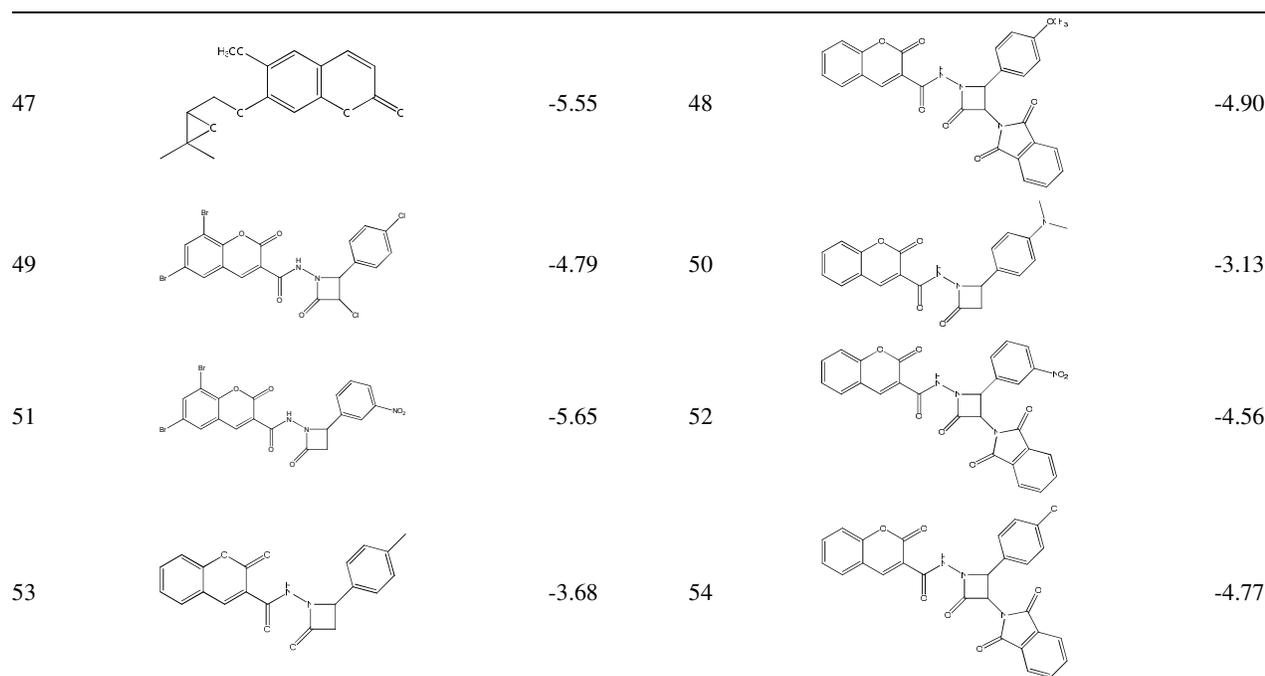


Table 3—Different structures of some applied networks

HL*	Design	Y-Scrambling R ²	Validation Set Error	Calculation Cycles
1	19-4-1	0.982	0.006543	789
1	19-6-1	0.678	0.008965	897
1	19-8-1	0.721	0.009821	467
1	19-11-1	0.467	0.009432	989
1	19-15-1	0.323	0.009121	897
2	19-7-6-1	0.789	0.009033	511
2	19-10-11-1	0.690	0.09777	1254
2	19-13-12-1	0.756	0.09543	1546
3	19-8-3-5-1	0.896	0.09859	1764
3	19-7-8-6-1	0.564	0.09123	1357

*HL: Hidden Layers

the prediction of leading antifungal coumarins. The best architecture, according to the term of calculation cycles and considering the correlating behaviour and output cycles of calculation was 19-7-6-1. Artificial neural networks are used to model systems that receive inputs and produce outputs. The relationships between the inputs, outputs and the representation parameters are critical issues in the design of a good model for bioactive compounds, and methods concerning sensitivity analysis in analyzing these relationships. Perturbations of neural networks are caused by machine imprecision, and they can be simulated by embedding disturbances in the original inputs or connection weights, allowing us to study the characteristics of a function under small perturbations

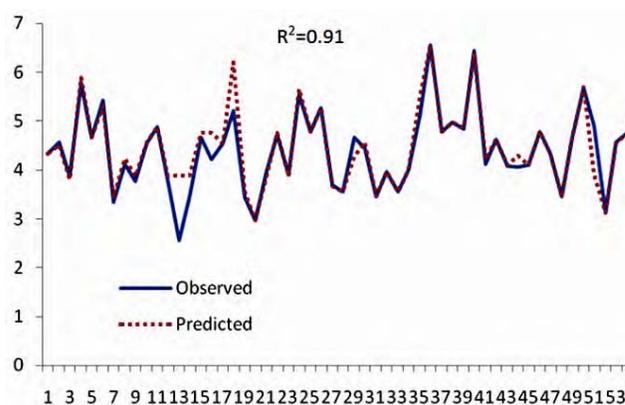


Fig. 2—Plot of predicted activity versus observed one

of its parameters. Sensitivity analysis is a measure of how the outputs change when the inputs are changed. The present results revealed that the most sensitive inputs were Log P and molar refractivity. The input importance shows the relative importance of each input column. The importance is the sum of the absolute weights of the connections from the input node to all the nodes in the first hidden layer. In the present case, descriptors surface tension, energy of LUMO and energy of HOMO were the most important inputs. The correlation coefficient between the observed and the interaction energy values was found to be 0.9132 (Fig. 2). Predicted activity varied from -3.11 to -6.74 . Y-scrambling result showed that the classification accuracy for randomized datasets

were significantly lower than the original datasets (data not shown). The highest error was observed for compound 12, 27, and 33. Cross validation was done by leave-some-out (some=4) validating method. The validation showed that average of absolute errors was 0.017.

Discussion

The development of a new drug is still a challenging, time-consuming, and cost-intensive process. However, computational methods can be used to assist and speed up the drug discovery process. In contrast to classical statistical methods, such as, regression analysis or partial least squares analysis (PLS), the artificial neural networks make the investigation of complex nonlinear relationships possible. In other words, neural networks are ideally suited for drug design and QSAR. They consist of many basic units, called artificial neurons (or simply neurons), which perform identical tasks. A neuron collects a series of input signals and transforms them into the output signal *via* a transfer function. In the training course, such a network of neurons 'learns' by changing the weights of its neurons. LM is typically the fastest of the training algorithms that apply the back propagation rule and performs calculations using the entire data set that might improve the performance of network^{31,32}. In the present study, LM training algorithm was used to investigate the relationship between antifungal activity score data for a dataset of coumarin antifungals with the thermodynamic and physico-chemical descriptors. In a previous study, it has been shown that neural network could be used to correlate various parameters³³. The present descriptors were based on the molecular structures of compounds. Among the architectures constructed, the best ANN architecture was found to be 19-7-6-1, which is in accordance with the previous study³⁴. The statistical criteria of different architecture are shown in Table 3. The quiet low error for the training and validation set indicates that training and validation are absolutely successful. Thermodynamic and physico-chemical descriptors play a crucial role in the interaction of candidate compounds with their specific receptors (biological membrane). The results have shown that descriptors LUMO (lowest unoccupied molecular orbital) and HOMO (Highest occupied molecular orbital energy) and surface tension were the most important among all descriptors. Falandysz *et al*¹⁸ have also found the similar results for chloronaphthalenes compounds. Surface tension of

the compound decides how easily the compound moves itself around and into the inner diameter. In fact, lower surface tension energy promotes coalescence. LUMO is the lowest energy level in the molecule that contains no electrons. Molecules with low-lying LUMOs are more capable to accept electrons than those with high LUMOs. Thus, LUMO descriptor measures the electrophilicity of a molecule. HOMO (highest occupied molecular orbital) is the highest energy level in the molecule that contains electrons. It is crucially important in governing molecular reactivity and properties. Thus, HOMO descriptor measures the nucleophilicity of a molecule. Both descriptors strongly define how a compound could interact with membrane. These findings are in accordance with previous studies³⁵. The most sensitive descriptors were found to be Log P and molar refractivity. Log P (the octanol/water partition coefficient) and molar refractivity can be used to relate chemical structure to observed chemical behavior. Log P is related to the hydrophobic character of the molecule, while the molecular refractivity index of a substituent is a combined measure of its size and polarizability.

Conclusion

Coumarins are a diverse group of compounds that are found in plants and some of them have remarkable antifungal activity. In order to develop new antifungal medicines, there is a great trend to predict antifungal activity of candidate compounds using neural networks. In the present study, ANN based models helped to understand better the capabilities of the coumarin compounds as new antifungal agents. The bioactive compounds selected in the present study had a wide range of structures and chemical properties in order to make the outcome of the study more reliable and authentic, which could further help in predicting behaviour of similar compounds. The study revealed those physico-chemical properties of coumarins that have the most important impact on the efficiency of antifungal activity. This can help to use coumarins in the antifungal drug design studies in future.

References

- 1 Bakheshwain S, El Khizzi N, Al Rashed A M, Al Ajlan A & Parvez S, Isolation of opportunistic fungi from dermatophytic samples, *Asian J Dermatol*, 3 (2011) 13-19.
- 2 Stein A C, Alvarez S, Avancini C, Zacchino S & Von Poser G, Antifungal activity of some coumarins obtained from species of *Pterocaulon* (Asteraceae), *J Ethnopharmacol*, 107 (2006) 95-98.

- 3 Shathele M S & Fadlemlula A, *In vitro* effectiveness of some antifungal drugs in treatment of *Trichophyton verrucosum*, dermatophytic fungi, *Asian J Anim Vet Adv*, 5 (2010) 180-192.
- 4 Coleman D C, Rinaldi M G, Haynes K A, Rex J H, Summerbell R C *et al*, Importance of *Candida* species other than *Candida albicans* as opportunistic pathogens, *Med Mycol*, 36 (1998) 56-65.
- 5 El-Seedi H R, Antimicrobial Arylcoumarins from *Asphodelus microcarpus*, *J Nat Prod*, 70 (2007) 118-120.
- 6 Razavi S M, Plant coumarins as allelopathic agents, *Int J Biol Chem*, 5 (2011) 86-90.
- 7 Sardari S, Mori Y, Horita K, Micetich R G, Nishibe S *et al*, Synthesis and antifungal activity of coumarins and angular furanocoumarins, *Bioorg Med Chem*, 7 (1999) 1933-1940.
- 8 Daoubi M, Duran-Patron R, Hmamouchi M, Hernandez-Galan R, Benharref A *et al*, Screening study for potential lead compounds for natural product-based fungicides: I. Synthesis and *in vitro* evaluation of coumarins against *Botrytis cinerea*, *Pest Manag Sci*, 60 (2004) 927-932.
- 9 Spowage B M, Bruce C L & Hirst J D, Interpretable correlation descriptors for quantitative structure-activity relationships, *J Cheminform*, 1 (2009) 1-13.
- 10 Asadi R, Mustapha N, Sulaiman N & Shiri N, New supervised multi layer feed forward neural network model to accelerate classification with high accuracy, *Eur J Sci Res*, 33 (2009) 163-178.
- 11 Gonzalez-Diaz H, Vilar S, Santana L & Uriarte E, Medicinal chemistry and bioinformatics—Current trends in drugs discovery with networks topological indices, *Curr Top Med Chem*, 7 (2007) 1015-1029.
- 12 Bishop C M, *Neural networks for pattern recognition* (Clarendon Press, Oxford, UK) 1995, 245-270.
- 13 Levenberg K, A method for the solution of certain problems in least squares, *Quart Appl Math*, 2 (1944) 164-168.
- 14 Zjavka E, Differential polynomial neural network, *J Artif Intell*, 4 (2011) 89-99.
- 15 Mullai P & Rene E R, A simple multilayer neural network for prediction bacterial xylanase production by *Bacillus* sp. using *Avena sativa* as substrate, *Biotechnology*, 7 (2008) 499-504.
- 16 Wesolowski M & Konieczynski P, Thermoanalytical, chemical and principal component analysis of plant drugs, *Int J Pharm*, 262 (2003) 29-37.
- 17 Falandysz J, Raciniwska M, Blazejowski J, Golaszewska I & Skurski P, Thermodynamic and physicochemical properties of some ecotoxicological features of polychlorinated naphthalenes, paper presented in Dioxin'98, 18th Symposium on Halogenated Organic Pollutants, Finland, Stockholm, 17-21 August, 1998.
- 18 Falandysz J, Puzyn T, Szymanowska B, Kawano M, Markuszewski M *et al*, Thermodynamic and physicochemical descriptors of chloronaphthalenes: An attempt to select features explaining environmental behaviour and specific toxic effects of these compounds, *Pol J Environ Stud*, 10 (2001) 217-235.
- 19 Lei Y D, Wania F & Shiu W Y, Vapor pressures of the polychlorinated naphthalenes, *J Chem Eng Data*, 44 (1999) 577-582.
- 20 Bhatia M S, Ingale K B, Choudhari P B, Bhatia N M & Sawant R L, Application of quantum and physicochemical molecular descriptors utilizing principal components to study mode of anticoagulant activity of pyridyl chromen-2-one derivatives, *Bioorg Med Chem*, 17 (2009) 1654-1662.
- 21 Stewart J J P, Optimization of parameters for semiempirical methods II. Applications, *J Comp Chem*, 10 (1989) 221-264.
- 22 Godoya M F P, Victor S R, Bellini A M, Guerreiro G, Rocha W C *et al*, Inhibition of the symbiotic fungus of leaf-cutting ants by coumarins, *J Braz Chem Soc*, 16 (2005) 669-672.
- 23 Nath M, Jairath R, Eng G, Song X & Kumar A, Triorganotin (IV) derivatives of umbelliferone (7-hydroxycoumarin) and their adducts with 1,10-phenanthroline: A synthesis, structural and biological studies, *J Organomet Chem*, 690 (2005) 134-144.
- 24 Mouri T, Yano T, Kochi S, Ando T & Hori M, Synthesis and antifungal activity of new 3,4,7-trisubstituted coumarins, *J Pestic Sci*, 30 (2005) 209-213.
- 25 Giri S, Sharan P & Nizamuddin, Syntheses of some 1-(substituted coumarin-3-carboxamido)-3-substituted-4-aryl-2-azetidinones as potential antifungal agents, *Agric Biol Chem*, 53 (1989) 1153-1155.
- 26 Witbrocka M & Zagha M, An implementation of backpropagation learning on GF11, a large SIMD parallel computer, *Parallel Comput*, 14 (1990) 329-346.
- 27 Hagan M T & Menhaj M B, Training feed forward networks with the Marquardt algorithm, *IEEE Trans Neural Netw*, 5 (1994) 989-993.
- 28 Solla S A, Levin E & Fleisher M, Accelerated learning in layered neural networks, *Complex Syst*, 2 (1988) 625-640.
- 29 Bello M G, Enhanced training algorithms, and integrated training/architecture selection for multilayer perceptron networks, *IEEE Trans Neural Netw*, 3 (1992) 864-875.
- 30 Hsieh K L, Employing artificial neural network into achieving parameter optimization of multi response problem with different importance degree consideration, *Inform Technol J*, 9 (2010) 918-926.
- 31 DiMasi J A, The value of improving the productivity of the drug development process: Faster times and better decisions, *Pharmacoconomics*, 20 (2002) 1-10.
- 32 Lu C & Shi B X, Circuit design of on-chip back propagation learning neural network with programmable neuron characteristics, *Chinese J Semicond*, 21 (2000) 1164-1169.
- 33 Khanale P B, Recognition of Marathi numbers using artificial neural networks, *J Artif Intell*, 3 (2010) 135-140.
- 34 Tambunan U S F & Parikesit A A, *In silico* design of drugs and vaccines for Dengue disease, *Trends Bioinform*, 7 (2010) 56-65.
- 35 Soltani S, Dianat S & Sardari S, Forward modeling of the coumarin antifungals: SPR/SAR based perspective, *Avicenna J Med Biotechnol*, 1 (2009) 95-103.