ORIGINAL RESEARCH

Modeling of thermodynamic and physico-chemical properties of coumarins bioactivity against *Candida albicans* using a Levenberg–Marquardt neural network

Seyyedeh Soghra Mousavi[1]
Hanieh Bokharaie[2]
Shadi Rahimi[3]
Sima Azadi Soror[4]
Mehrdad Hamidi[5]

[1]Department of Biotechnology, School of Pharmacy, Zanjan University of Medical Science, Zanjan, Iran; [2]Genetic Group, Biology Department, Faculty of Basic Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran; [3]Department of Biology, Faculty of Science, Tarbiat Moallem University, Tehran, Iran: [4]Plant Protection Department, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran; [5]Department of Pharmaceutics, School of Pharmacy, Zanjan University of Medical Science, Zanjan, Iran

Abstract: In recent years, due to vital need for novel fungicidal agents, investigation on natural antifungal resources has been increased. The special features exhibited by neural network classifiers make them suitable for handling complex problems like analyzing different properties of candidate compounds in computer-aided drug design. In this study, by using a Levenberg–Marquardt (LM) neural network (the fastest of the training algorithms), the relation between some important thermodynamic and physico-chemical properties of coumarin compounds and their biological activities (tested against *Candida albicans*) has been evaluated. A set of already reported antifungal bioactive coumarin and some well-known physical descriptors have been selected and using LM training algorithm the best architecture of neural model has been designed for forecasting the new bioactive compounds.

Keywords: Levenberg–Marquardt algorithm, coumarin, neural network

Introduction

Human fungal infections have increased at an alarming rate for the last 20 years, mainly among immunocompromised individuals.[1] Although it appears to be a great array of antifungal drugs, there is at present a quest for new generations of antifungal compounds due to the low efficacy, side effects, or resistance associated with the existing drugs.[2,3] However, there are only a limited number of known clinically available antifungal reagents, including amphoterlcin B, ketoconazole, fluor, oconazole, and itraconazole. These antifungal drugs have disadvantages including high toxicity, ineffective towards some fungi, and low bioavailability, thus they are not able to meet fully the needs of the patients.[4] Coumarin compounds are naturally occurring constituent of many plants and essential oil which comprise a chromenone ring, often a chromen-2-one or chromen-4-one ring.[5] Selected coumarins are known to have antifungal activity. For example, Sardari et al[6] described how a limited number of coumarins are active against *C albicans*, *Cryptococcus neoformans*, *Saccharomyces cerevisiae*, and *Aspergillus niger*. Several different parameters should be evaluated to design a new coumarin antifungal.[7,8] Recently, artificial neural networks (ANNs) have most widely been used in drug design. They usually consist of 3 or 4 input layers, 1 output layer, and 1 or 2 hidden ones.[9] In pattern classification, understanding the class boundaries by the classifier needs a training phase with a training algorithm.[10] Gradient-based training algorithms, like back-propagation, are not efficient due to the fact that the gradient vanishes in the solution.[11] Hessian-based algorithms allow

Correspondence: Seyyedeh Soghra Mousavi
Department of Biotechnology, School of Pharmacy, Zanjan University of Medical Science, PO Box 4513956184, Zanjan, Iran
Tel +98-241 4273635
Fax +98-241 4273639
Email s-mousavi@hotmail.com

59

the network to learn more subtle features of a complicated mapping.[12] The training process converges quickly as the solution is approached, because the Hessian does not vanish in the solution. The Levenberg–Marquardt (LM) algorithm is basically a Hessian-based algorithm for nonlinear least squares optimization.[13,14] For neural network training, the objective function is the error function of the type

$$e(x) = \frac{1}{2} \sum_{k=0}^{p-1} \sum_{l=0}^{n_0-1} (d_{kl} - a_{kl})$$

A unit in 1 layer is connected to all units in the next,[13] where the a_{kl} is the actual output at the output neuron l for the input k, the d_{kl} is the desired output at the output neuron l for the input k, p is the total number of training patterns, n_0 represents the total number of neurons in the output layer of the network, and x represents the weights and biases of the network.[14] Because they can find the complex relationship between predictor variables (inputs) and predicted variables (output), LM algorithm trained ANN has received growing attention in drug discovery. In this study, first several LM neural networks are built for a set of thermodynamic and physico-chemical properties of antifungal coumarins. After that, the best architecture in terms of the least error and cycle of calculation is selected. Eventually, this neural model is used to calculate the correlation coefficient between thermodynamic and physico-chemical properties and bioactivity of antifungal coumarins (tested against *Candida albicans*) and the role of these properties in bioactivity of coumarins antifungal is 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).[15–18] Geometry optimization is carried out by using the semiempirical PM3 method,[19] implemented in (HyperCube, Inc. Gainesville, FL, USA)™ program package.[20] All of these descriptor are generated by different application such as ACDLAB (11.02 release 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 highest occupied molecular orbital (HOMO), and lowest unoccupied molecular orbital (LUMO) are extracted from MOPAC 93 data files.

The dataset is composed of some coumarins and coumarin derivatives that have previously shown antifungal

Table 1 Some thermodynamic and physico-chemical descriptors for all congeners
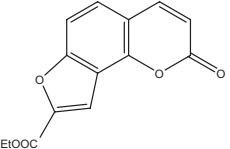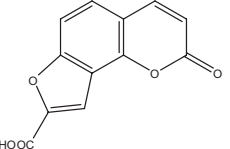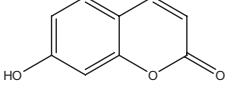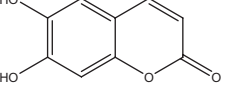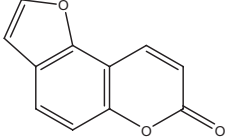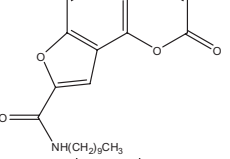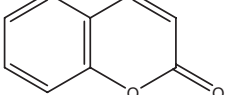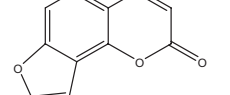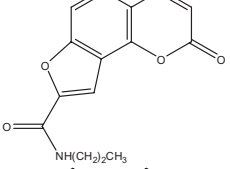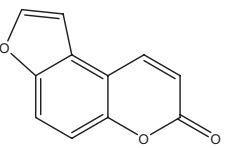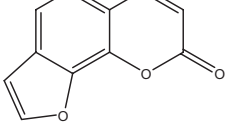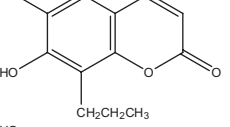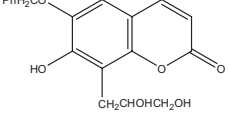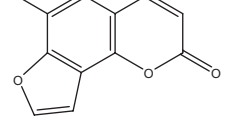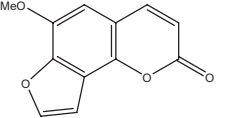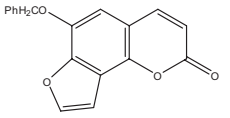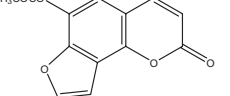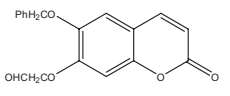
Descriptor	Parameter	References
1	Vicinal carbon atoms substitution pattern	8
2	Sasvol (solvent-accessible volume)	8
3	Vdwvol (van der Waals volume)	8
4	Symmetry of molecule	8
5	Maxq+ (the largest positive charge over the atoms in molecules)	8
6	Vapor pressure	8
7	Energy of HUMO (highest occupied molecular orbital)	8
8	Energy of LUMO (lowest unoccupied molecular orbital)	8
9	Molecular mass (Da)	8
10	Dipole moment of the molecule	15
11	Density (g/cm)	15
12	Retention time	15
13	Heat capacity	16
14	Standard enthalpy of formation	16
15	Specific polarizability of molecule	17
16	Molar refractivity (cm³)	17
17	Molar volume (cm³)	18
18	Log P	18
19	Surface tension (dyne/cm)	18

activity. Bioactivities of these compounds are screened by the well dilution method and have been taken by literature search (Table 2).[6,21–26] One major problem is the reporting of antifungal activity in 2 different forms of 50% inhibitory concentration (IC$_{50}$) and minimal inhibitory concentration (MIC). By multiplying the IC$_{50}$ values by 2,[17] we obtain a close equivalent of MIC level; hence, our dataset becomes uniform, because this calculated number is approximately equal to MIC. We used antifungal screening results of isolates of *C albicans* for the simulation of their bioactivity. The Error Back Propagation (EBP) algorithm has been a significant improvement in neural network research, but it has a weak convergence rate.[27,28] Many efforts have been made to speed up EBP algorithm.[29,30] All of these methods lead to little acceptable results. The LM algorithm ensues from development of EBP algorithm dependent methods. It gives a good exchange between the speed of the Newton algorithm and the stability of the steepest descent method [31] that are 2 basic theorems of LM algorithm. In this paper, a feed-forward neural network with LM algorithms applied for modeling the bioactivity of coumarins antifungal. A standard feed-forward network, with LM algorithms and with 1–3 hidden layer architecture, was chosen. For solving the problem of over-fitting, the number of neurons was kept at a minimum. However, the optimum architecture with target error less than 0.02% was created with

variation in the total number of nodes and hidden layers. This neural model is built with NeuroSolutions (NeuroDimension, Inc. Gainesville, FL, USA) (version: v5.07, 2008). For validation of our model and to analyze the influence of inherent randomness on the prediction stability, 10 repetitions of the complete validation process with different random seeds
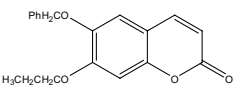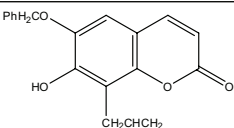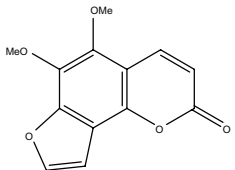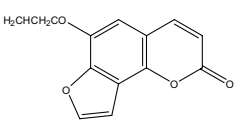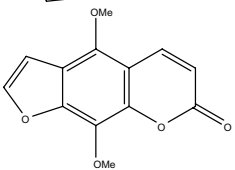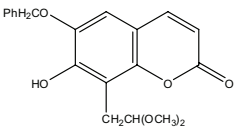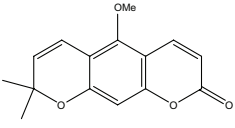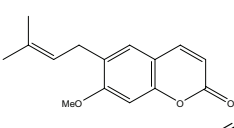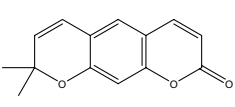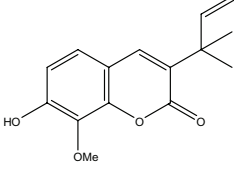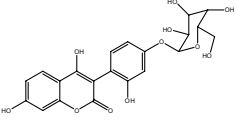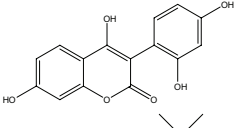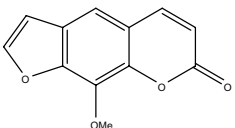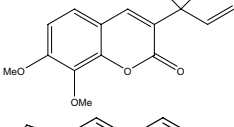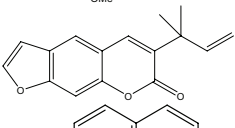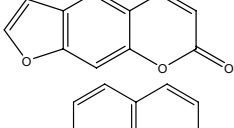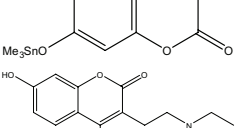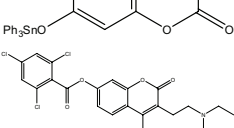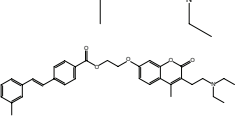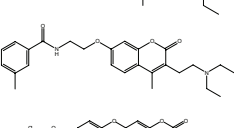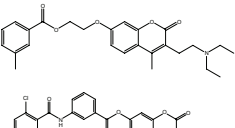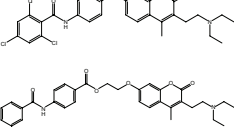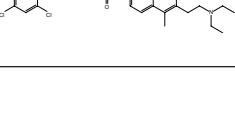
were made in all cases (Y-scrambling test). Accuracy has been selected for evaluation of predictive performance of a single validation process, whereas a coefficient of correlation of accuracies obtained across 10 repetitions is established as a measure of learning stability. Also cross-validation was applied by leave-n-out method.

Table 2 Structure and bioactivity of studied coumarins. The observed MICs and structures of coumarin compounds are derived from mentioned references

Number	Compound	MIC observed, µg/mL	Number	Compound	MIC observed, µg/mL
1		1,000	2		2,000
3		1,000	4		1,000
5		250	6		250
7		1,000	8		62.5
9		250	10		250
11		250	12		250
13		1,000	14		250
15		500	16		500
17		250	18		1,000
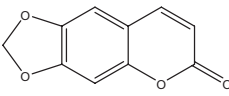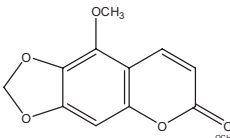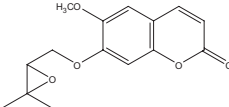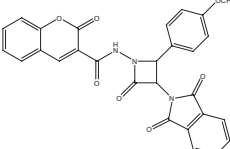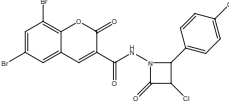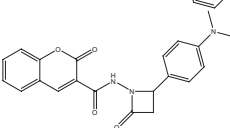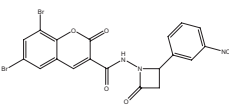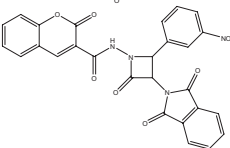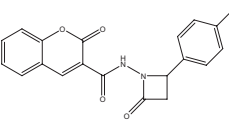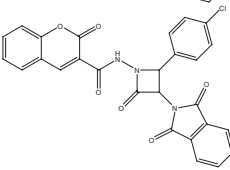
(Continued)

Table 2 *(Continued)*

Number	Compound	MIC observed, µg/mL	Number	Compound	MIC observed, µg/mL
19		1,000	20		1,000
21		500	22		500
23		80	24		1,000
25		70	26		64
27		25	28		93.75
29		512	30		64
31		42.65	32		78.75
33		16.65	34		31.4
35		5	36		25
37		500	38		15.6
39		15.6	40		125
41		31.3	42		7.8
43		15.6	44		7.8

(Continued)

Table 2 *(Continued)*

Number	Compound	MIC observed, μg/mL	Number	Compound	MIC observed, μg/mL
45		250	46		250
47		250	48		2,150
49		1,979	50		3,321
51		3,478	52		2,035
53		3,752	54		2,705

Abbreviation: MIC, minimal inhibitory concentration

Result

The computed basic physico-chemical and thermodynamic descriptors for coumarins presented in Table 1. Various architectures of neural network are shown in Table 3. In this study, LM trained ANN has been used to build a neural model for prediction of leading antifungal coumarins. The best architecture, according to the term of calculation cycles and considering the correlating behavior and output cycles of calculation was 19-8-1. ANNs are used to modeling 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 sensitivity analysis concerns methods to analyze these relationships. Perturbations of neural networks are caused by machine imprecision, and they can

Table 3 Different plan of some applied networks by focus on errors

HL	Design	Y-scrambling R^2	Validation set error	Calculation cycles	Training set error
1	19-3-1	0.672	0.007787	668	0.009987
1	19-6-1	0.761	0.009988	564	0.009941
1	19-8-1	0.763	0.009277	456	0.009781
1	19-11-1	0.553	0.009900	787	0.009887
1	19-15-1	0.221	0.009856	623	0.009924
2	19-6-10-1	0.447	0.009677	980	0.00996
2	19-10-11-1	0.664	0.09967	1345	0.09999
2	19-13-12-1	0.774	0.9779	2321	0.09999
3	19-8-3-5-1	0.441	0.09959	1255	0.08812
3	19-7-8-6-1	0.364	0.08999	876	0.06812

Abbreviation: HL, hidden layers.

be simulated by embedding disturbances in the original inputs or connection weights, allowing us to study the characteristics of a function under small perturbations of its parameters. Sensitivity analysis is a measure of how the outputs change when the inputs are changed. The result of this analysis could help to predict the bioactivity of new antifungal coumarins. Result shows that the most sensitive input are 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. Descriptors energy of LUMO, energy of HOMO, and surface tension were the most important inputs. The least important descriptor was determined as the density. The correlation coefficient between the observed and the predicted MIC value was 0.9266. Predicted activity varies from 22.55878 to 2010.87537 (Figure 1). Y-scrambling result showed that the classification accuracy for randomized datasets was significantly lower than for the original datasets (data not shown). The highest error is observed for compound 11, 34, and 43. Cross-validation is done by leave-some-out (some = 4) validating method. Validation showed that average of absolute errors was 0.029.

Discussion

The development of a new drug is still a challenging, time-consuming, and cost-intensive process. 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, the ANNs enable the investigation of complex nonlinear relationships. Therefore, neural networks are ideally suited to be used in drug design and Quantitative structure-activity relationship

Figure 1 Plot of predicted activity vs observed one.

(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 course of training, such a network of neurons "learns" by changing the weights of its neurons. Two different learning methods can be distinguished:[32] supervised and unsupervised learning. When learning is unsupervised, the neural network is provided with the input patterns. After some iteration, it should be settled to a stable state. The goal of supervised learning methods is to find a model that associates correctly the inputs (representation of the objects) with the targets (representation of the responses). The targets serve not only as a criterion for how well the system has been trained, but also they influence the correction of each weight. Also the best-known example of a neural network training algorithm is back-propagation, it is the easiest algorithm to understand, it is also a good choice if the dataset is very large, contains a great deal of redundant data and finally it still has advantages in some circumstances, but modern second-order algorithms such as conjugate gradient descent and LM are substantially faster (eg, an order of magnitude faster). LM is typically the fastest of the training algorithms and performs calculations using the entire dataset that might improve the performance of network.[33,34]

In this study, LM training algorithm is applied to discover the relationship between antifungal activity score data for a dataset of coumarins antifungal with the thermodynamic and physico-chemical descriptors. Descriptors are derived from molecular structure. Among the architectures constructed, the best ANN architecture is 19-8-1. Table 3 shows the statistical criteria of different architecture. The quiet low error for the training and validation set indicate 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 (eg, biological membrane). The results have shown that descriptors LUMO and HOMO energy are the most important among all descriptors. LUMO is the lowest energy level in the molecule that contains no electrons. When a molecule acts as a Lewis acid (an electron-pair acceptor) in bond formation, incoming electron pairs are received in its LUMO. Molecules with low-lying LUMOs are more able to accept electrons than those with high LUMOs; thus, the LUMO descriptor should measure the electrophilicity of a molecule. HOMO is the highest energy level in the molecule that contains electrons. It is crucially important in governing molecular reactiv-
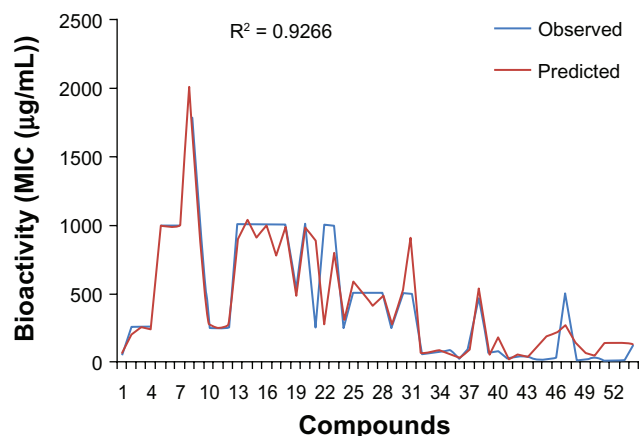
ity and properties. When a molecule acts as a Lewis base (an electron-pair donor) in bond formation, the electrons are supplied by the molecule's HOMO. Molecules with high HOMOs are more able to donate their electrons and are hence relatively reactive compared with molecules with low-lying HOMOs; hence, the HOMO descriptor should measure the nucleophilicity of a molecule. Both descriptors strongly define how a compound could interact with a receptor. The third important descriptor is surface tension. This is in accordance with previous studies.[17] The most sensitive is Log P and molar refractivity. Log P (the octanol/water partition coefficient) and molar refractivity are molecular descriptors that can be used to relate chemical structure to observed chemical behavior. Log P is related to the hydrophobic character of the molecule. The molecular refractivity index of a substituent is a combined measure of its size and polarizability. Because of their flexibility, supervised neural network have found a great application in drug design, for example a network with LM learning algorithm can be employed for the following applications in drug design: analysis of multidimensional data, classification and prediction of biological activity and ADME-Tox (absorption, distribution, metabolism, excretion, and toxicity) properties, lead discovery, comparison of compound libraries and analysis of the similarity. Unfortunately LM has some important limitations, specifically it can be only used on single output networks, and be used with the sum squared error function, and has memory requirements proportional to W^2 (where W is the number of weights in the network; this makes it impractical for reasonably big networks). LM training algorithms seem to be very prone to stick in local minima in the early phases.[35] Modification in LM algorithm to decrease these limits may increase its application in drug design.

Disclosure

The authors report no conflicts of interest for this work.

References

1. Denning DW, Evans EGV, Kibbler CC, et al. Guidelines for the investigation of invasive fungal infections in haematological malignancy and solid organ transplantation. *Eur J Clin Microbiol Infect Dis*. 1997; 16(6):424–436.
2. Portillo A, Vila R, Freixa B, Adzet T, Cãnigueral S. Antifungal activity of Paraguayan plants used in traditional medicine. *J Ethnopharmacol*. 2001;76(1):93–98.
3. Micheal AP. Epidemiology of fungal infections: the promise of molecular typing. *Clin Infect Dis*. 1995;20(2):1535–1539.
4. Coleman DC, Rinaldi MG, Haynes KA, et al. Importance of Candida species other than Candida albicans as opportunistic pathogens. *Med Mycol*. 1998;36 Suppl 1:56–65.

5. Urdangarín C, Regente MC, Jorrín J, La Canal L. Sunflower coumarin phytoalexins inhibit the growth of the virulent pathogen Sclerotinia sclerotiorum. *J Phytopathol*. 1999;147(7):441–443.
6. Sardari S, Mori Y, Horita K, et al. Synthesis and antifungal activity of coumarins and angular furanocoumarins. *Bioorg Med Chem*. 1999; 7(9):1933–1940.
7. Daoubi M, Dura-Patro R, Hmamouchi M. 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*. 2004;60(9):927–932.
8. Spowage BM, Bruce C, Hirst JD. Interpretable correlation descriptors for quantitative structure-activity relationships. *J Cheminform*. 2009; 1(22):1–13.
9. Asadi R, Mustapha N, Sulaiman N, Shiri N. New supervised multilayer feed forward neural network model to accelerate classification with high accuracy. *Eur J Sci Res*. 2009;33(1):163–178.
10. Castillo PA, Carpio J, Merelo JJ, et al. Evolving multilayer perceptions. *Neural Process Lett*. 2000;12(2):115–127.
11. Bishop C. *Neural Networks for Pattern Recognition*. UK: Oxford University Press; 1995.
12. Levenberg K. A method for the solution of certain nonlinear problems in least squares. *Quart Appl Math*. 1944;2(3):164–168.
13. Marquardt DW. An algorithm for least squares estimation of nonlinear parameters. *SIAM J*. 1963;11(2):431–441.
14. Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. *IEEE Trans NeuNets*. 1994;5(6):989–993.
15. Falandysz J, Raciniwska M, Blazejowski J, Golaszewska I, Skurski P. Thermodynamic and physicochemical properties vis. some ecotoxicological features of polychlorinated naphthalenes. Dioxin'98. Proceedings of 18th Symposium on Halogenated Organic Pollutants; Aug 17–21. Finland, Stockholm; 1998.
16. Lei YD, Wania F, Shiu WY. Vapour pressures of the polychlorinated naphthalenes. *J Chem Eng Data*. 1999;44(3):5777–5780.
17. Soltani S, Dianat S, Sardari S. Forward modeling of the coumarin antifungals; SAR based perspective. *AJMB*. 2009;1(2):95–103.
18. Soltani S, Keymanesh K, Sardari S. Evaluation of structural features of membrane active antifungal peptides using artificial neural network. *J Biol Sci*. 2008;8(5):435–447.
19. Stewart JJP. Optimalization of parameters for semiempirical methods. II Applications. *J Comp Chem*. 1989;10(4):221–227.
20. HyperChem™. Release 5.0 for Windows, Reference Manual. Hypercube, Inc, Publication HC50-00-01-00; 1996.
21. Godoya MFP, Victora SR, Bellini AM, et al. Inhibition of the symbiotic fungus of leaf-cutting ants by coumarins. *J Braz Chem Soc*. 2005; 16(6):669–672.
22. Benharref A, Collado IG. 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*. 2004;60(9):927–932.
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: synthesis, structural and biological studies. *J Organomet Chem*. 2005;690(7):134–144.
24. Mouri T, Yano T, Kochi SI, Ando T, Hori M. Synthesis and antifungal activity of new 3, 4,7-trisubstituted coumarins. *J Pestic Sci*. 2005; 30(3):209–213.
25. Stein AC, Alvarez S, Avancini C, Zacchino S, Poser GV. Antifungal activity of some coumarins obtained from species of Pterocaulon (Asteraceae). *J Ethnopharm*. 2006;107(11):95–98.
26. Giri S, Sharan P, Nizamuddin. Syntheses of some l-(substituted coumarin-3-carboxamido)-3-substituted-4-aryl-2-azetidinones as potential antifungal agents. *Agric Biol Chem*. 1989;53(4): 1153–1155.
27. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. *Nature*. 1986;323:533–536.
28. Whitbrock M, Zagha M. An Implementation of backpropagation learning on GF11, a large SIMD parallel computer. *Parallel Comput*. 1990; 14(3):317–327.

29. Hagan MT, Menhaj MB. Training feed forward network with the Marquardt algorithm. *IEEE Trans Neural Net*. 1994;5(6):989–993.

30. Solla SA, Levin E, Fleisher M. Accelerated learning in layered neural networks. *Complex Sys*. 1988;2(5):625–639.

31. Bello MG. Enhanced training algorithms, and integrated training/architecture selection for multi layer perceptron networks. *IEEE Trans Neural Net*. 1992;3(7):864–875.

32. Shah S, Meka PF. A fast, local algorithm for training feed forward neural networks. Proceedings of International Joint Conference on Neural Networks, San Diego, CA; 3, 1990. p. 41–46.

33. DiMasi J. The value of improving the productivity of the drug development process: faster times and better decisions. *Pharmacoeconomics*. 2002;20 Suppl 3:1–10.

34. Lu C, Shi BX. Circuit design of on-chip back propagation learning neural network with programmable neuron characteristics. *Chin J Semiconductors*. 2000;21(6):1164–1169.

35. Lampton M. Damping-undamping strategies for the Levenberg–Marquardt nonlinear least squares method. *Comp Phys J*. 1997;11(1):110–115.