Original Article

Forward Modeling of the Coumarin Antifungals; SPR/SAR Based Perspective

Saeed Soltani, Shima Dianat, and Soroush Sardari*

Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran

Abstract

Although, coumarins are a group of compounds which are naturally found in some plants, they can be synthetically produced as well. Because of their diverse derivatives, origin and properties most of them can be used for medicinal purposes. For example, they can be used against fungal diseases or in studying structure and biological properties of antifungal agents to discover new compounds with the similar activity. A Structure Property/Activity Relationship (SAR) can be utilized in prediction of biological activity of desired molecules.

In order to represent a relationship between the physicochemical properties of coumarin compounds and their biological activities, 68 coumarins and coumarin derivatives with already reported antifungal activities were selected and eleven attributes were generated. The descriptors were used to perform artificial neural network (ANN) and to build a model for predicting effectiveness of the new ones. The correlation coefficient between the experimental and the predicted MIC values pertaining to all the coumarins was 0.984. This study paves the way for further researches about antifungal activity of coumarins, and offers a powerful tool in modeling and prediction of their bioactivities.

Avicenna J Med Biotech 2009; 1(2): 95-103

Keywords: Antifungal activity, Coumarin, Modeling, Neural network

*** Corresponding author:**
Soroush Sardari, Ph.D., Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, No. 69, Pasteur Ave, Tehran, Iran, 13164
Tel: +98 21 66405535
Fax: +98 21 66465132
E-mail:
ssardari@hotmail.com;
sardari@pasteur.ac.ir
Received: 4 Mar 2009
Accepted: 25 Jul 2009

Introduction

During the last two decades, human fungal infections have increased among immune compromised individuals [1]. *Candida albicans* (*C. albicans*) is the major agent of candidosis in humans [2] which is the commonest invasive fungal infection in patients with malignant haematological disease and in bone marrow transplant recipients [3]. One common cause of mortality among hospitalized patients is nosocomial infection due to opportunistic fungal pathogens [4]. The development of azole-based antifungal drugs has revolutionized the treatment of many fungal infections, but therapy may still necessitate application of the highly toxic drug amphotericin B or a combination of drugs. Due to rapid emergence of resistance in fungal pathogens to the conventional drugs, discovery of new potent antifungal compounds is necessary. Plant extracts containing coumarin derivatives demonstrate antifungal activity [5] and some synthetic coumarin derivatives are also active against the yeast *C. albicans* [6]. Coumarin is a benzopyrone and a naturally occurring constituent of many plants and essential oils, including tonka beans, sweet clover, woodru, oil of cassia and lavender [7]. The presence of phenolic,

hydroxy and carboxylic acid groups on the coumarin nucleus has been considered necessary for antimicrobial activity [8]. The coumarins are extremely variable in structure and due to the various types of substitutions in the basic structural form their biological activity is influenced [9]. As a result, a lot of biological parameters should be evaluated to increase our understanding of the mechanisms by which these coumarins act and a careful structure-property/activity-relationship study of coumarins should be conducted.

The so called "Cheminformatics" was introduced to the common use. It is often described as part of the analytical chemistry that by making use of mathematics, probability theory, mathematical statistics, as well as the decision-making theory and computer techniques, has been applied to a diverse range of problems in the field of chemistry [10]. By combining together the elements of informatics and chem.-ical analysis, cheminformatics appeared to be particularly useful in the professional work of pharmacists. It is concerned with the search for new chemical compounds as potential drugs, clinical analysis of these compounds, optimization of drug formulation, evaluation of its quality as well as leading to recognition of complicated processes in which the drug substances are involved in a human organism [11].

Among the multivariate analyses used in the cheminformatics, the principal component analysis (PCA), cluster analysis (CA) and artificial neural networks (ANNs) have been the most widely used methods [12]. Their valuable features are that they can present the correct interpretation of the measured data and obtain the maximum useful information from them [13]. A feed-forward Multi-layer Perceptron (MLP) neural network is the most commonly used paradigm in medicinal chemistry. They usually consist of an input layer, one output layer and one or two hidden or middle layer(s). All units in one layer are connected to all the units in the next layers [14]. The signals flow from the first input layer forward through hidden nodes, where a weighed sum of inputs is computed and passed through activation function and the result is finally presented to the output layer. This process is called "feed-forward" [15].

A proper weight setting is not known beforehand and hence, initially, the weights are given a random value. The process of updating the weights to a correct set of values is called "Training or Learning", which is mostly achieved by means of Backpropagation (BP) algorithm [16]. The BP is a generalization of the least mean squared algorithm that modifies network weight to minimize the mean squared error between the desired and actual outputs of the network. The BP uses supervised learning in which the network is trained using data for which inputs as well as desired outputs are known [17].

The application of ANN's in solving different problems in pharmacy is receiving growing attention when it comes to data analysis problems [18]. It is mainly because they are applicable in every situation in which a relationship between predictor variables (inputs) and predicted variables (output) exists, even when that relationship is very complex and not easy to express in the usual terms of correlation or differences between groups. Therefore, anywhere that there are problems of prediction, classification or control, neural networks prove to be helpful [19]. Accordingly, in this study, neural computing is used for building an efficient model in order to evaluate the relationship between physicochemical properties and bioactivity of antifungal coumarins.

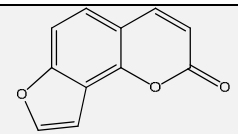
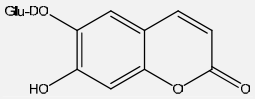
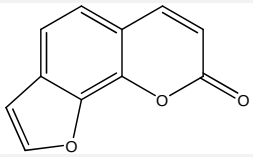
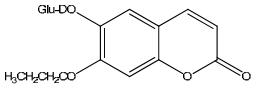
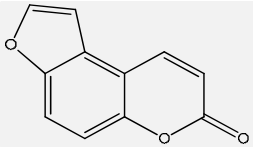
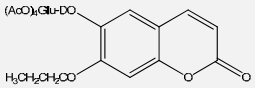
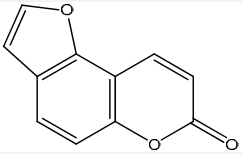
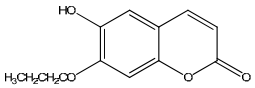
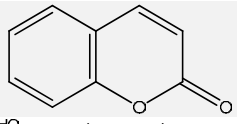
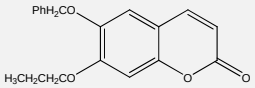
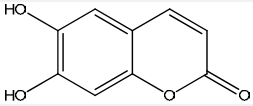
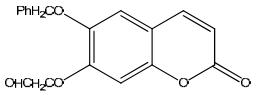
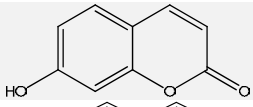
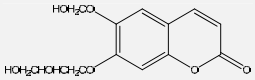
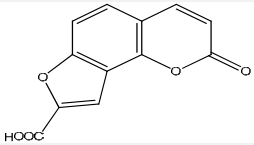
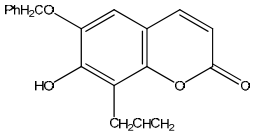
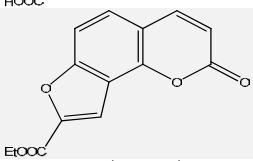
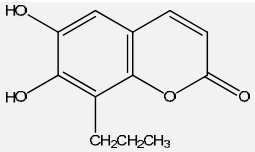
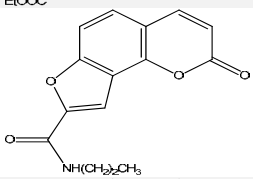
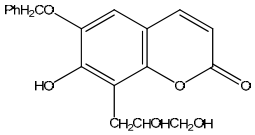
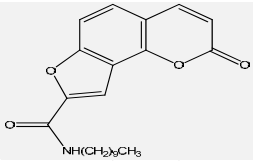
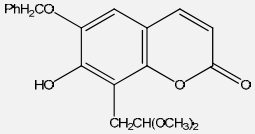
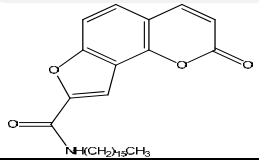
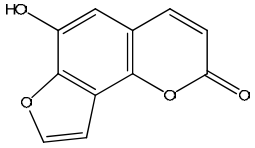
Materials and Methods

Data set

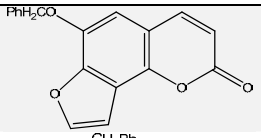
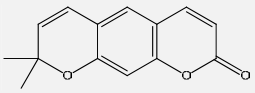
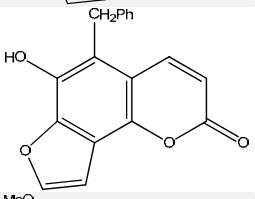
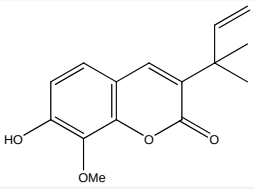
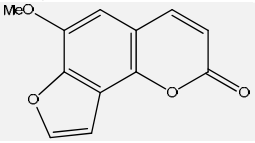
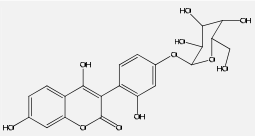
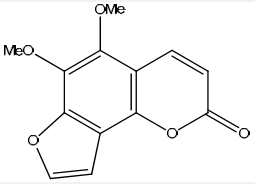
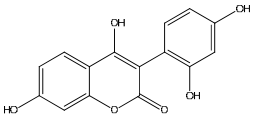
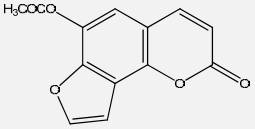
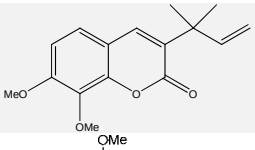
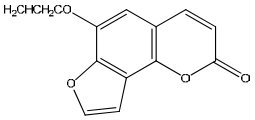
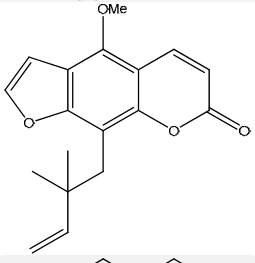
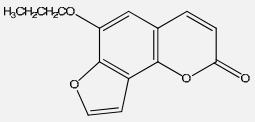
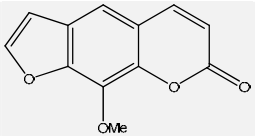
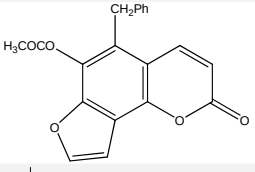
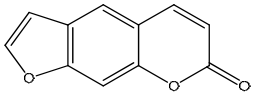
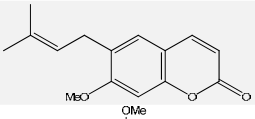
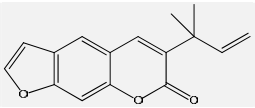
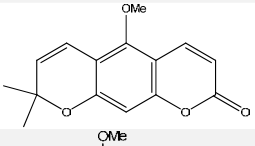
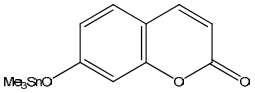
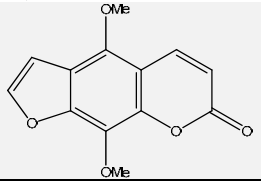
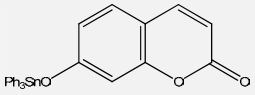
The data set was composed of 68 coumarins and coumarin derivatives selected on the basis of antifungal activity. Antifungal activity of compounds from Table 1 that were screened by the well dilution method has been taken from the literature [20-27].

Authors encountered problems related to reporting of antifungal activity according to the two different forms of minimal inhibitory

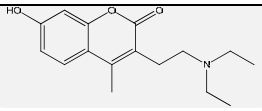
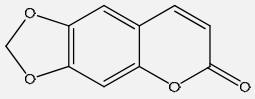
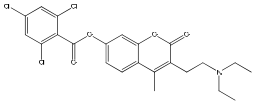
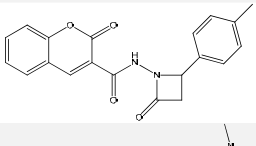
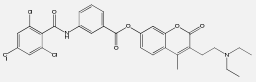
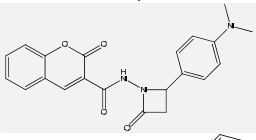
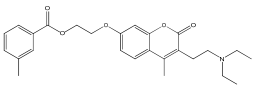
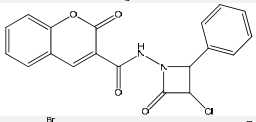
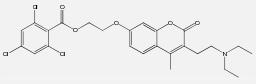
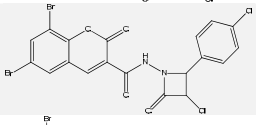
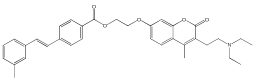
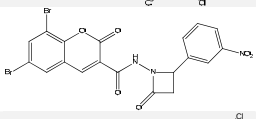
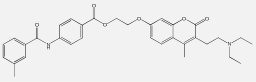
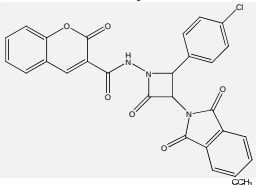
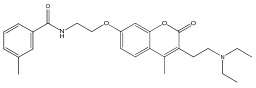
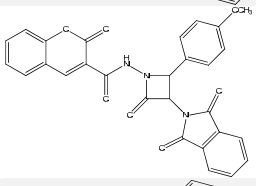
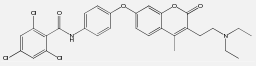
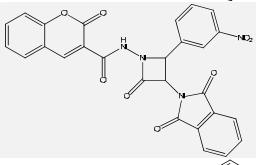
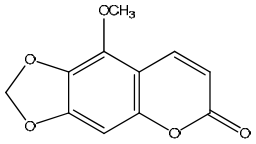
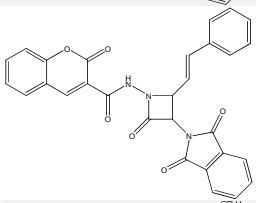
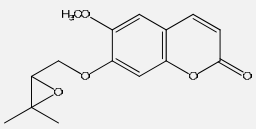
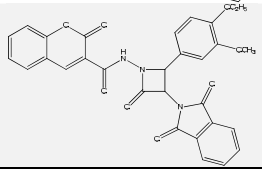
Table 1. Structure and bioactivity of studied coumarins

Number	Compound	MIC(*μg/ml*) observed	MIC(*μg/ml*) predicted*	Ref	Number	Compound	MIC(*μg/ml*) observed	MIC(*μg/ml*) predicted*	Ref
1		62.5	291	20	13		1000	285	20
2		250	290	20	14		1000	293	20
3		250	290	20	15		1000	262	20
4		250	290	20	16		1000	159	20
5		1000	264	20	17		1000	272	20
6		1000	282	20	18		1000	230	20
7		1000	302	20	19		500	166	20
8		2000	341	20	20		1000	225	20
9		1000	282	20	21		250	279	20
10		250	274	20	22		1000	269	20
11		250	137	20	23		1000	281	20
12		250	125	20	24		250	309	20

Contd.

Number	Compound	MIC(μg/ml) observed	MIC(μg/ml) predicted*	Ref	Number	Compound	MIC(μg/ml) observed	MIC(μg/ml) predicted*	Ref
25		500	280	20	36		25	252	21
26		500	286	20	37		93.75	232	21
27		500	301	20	38		512	267	22
28		500	315	20	39		64	322	22
29		250	290	20	40		78.75	181	23
30		500	284	20	41		22.6	230	23
31		500	279	20	42		42.65	284	23
32		62.5	290	20	43		31.4	290	23
33		64	205	21	44		16.65	264	23
34		70	237	21	45		5	189	24
35		80	279	21	46		25	215	24

Contd.

Number	Compound	MIC(*μg/ml*) observed	MIC(*μg/ml*) predicted*	Ref	Number	Compound	MIC(*μg/ml*) observed	MIC(*μg/ml*) predicted*	Ref
47		500	270	25	58		250	287	26
48		15.6	137	25	59		3752	3557	27
49		15.6	138	25	60		3321	3332	27
50		31.3	131	25	61		4310	3774	27
51		15.6	136	25	62		1979	1682	27
52		15.6	143	25	63		3478	3041	27
53		7.8	141	25	64		2705	2547	27
54		125	129	25	65		2150	2343	27
55		7.8	136	25	66		2035	2490	27
56		250	282	26	67		3256	2486	27
57		250	205	26	68		1870	1714	27

(*) The observed MICs and structures of coumarin compounds are derived from mentioned references in the table, but predicted MICs have been calculated by our ANN model.

concentration (MIC) and 50% inhibitory concentration (IC_{50}) which disabled the analysis of data set with adequate care. To make the dataset uniform, we multiplied the IC_{50} values

by two to obtain a close equivalent of MIC level. Thus, the number generated is approximately equal to MIC for complete inhibition. Preliminary results have shown that coumarins possess considerable antifungal activity [5]. Therefore, antifungal screening results of isolates of *C. albicans* were used for the modeling of activity against this microorganism.

Descriptors generation

Eleven attributes have been generated for the description of selected coumarin derivatives that included eight quantum chemical descriptors; molar refractivity (cm^3), molar volume (cm^3), parachor (cm^3), index of refraction, surface tension (*dyne/cm*), density (*g/cm*), polarizability (10-24cm^3), molecular mass (*Da*) and three regular calculated descriptors (% carbon, % hydrogen, % oxygen). Calculation of quantum chemical descriptors was preceded by molecular geometry optimization based on the PM3 semiempirical approach. Both semiempirical and regular calculations were carried out by ACDLAB 11.02 release 21, May 2008 for *in vacuo* systems. Besides, quantum chemical descriptors, the regular calculated descriptors, % carbon, % hydrogen, and % oxygen) were included in the pool that make better understanding of structure–function activity of coumarin antifungal.

Learning tools

In this study the artificial neural network application of Easy-NNplus 8.0 release 2007, was utilized for SAR model development. Since this technique has been thoroughly described in the reference [28], a detailed description of the method has been omitted. However, a specific implementation of the method for this study is given below.

A standard feed-forward network, with back propagation rule and with one, two or three hidden layer architecture was chosen. The physico-chemical descriptors were used as the inputs, while MIC was the output of the network architecture. In order to avert an over-fitting problem, which is usually produced by more weights due to higher numbers of neurons in input and hidden layers [29], the number of neurons was kept to minimum. However, to produce the optimum architecture, powerful enough to model the functions and keep the errors below 0.05%, number of nodes in the hidden layer(s) were varied.

Model validation

Model validation process provides a reasonable mean for understanding and approach to molecular design and action mechanism analysis. Applied primary validation methods involved the use of random number generators as a part of the learning process. In order to analyze the influence of inherent randomness on the prediction stability, ten 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, while a correlation coefficient (CO) of accuracies obtained across ten repetitions was established as a measure of learning stability. Also cross-validation was applied by leave-n-out method.

Results

The results of this paper are based on investigation and analysis of collected or calculated data of several coumarin structural descriptors. The artificial neural network system was performed to build a powerful model for prediction of lead and template antifungal coumarins. Table 2 shows results of the various architectures of the neural network system. The numbers of hidden layer nodes were varied according to different node numbers and layers. One of the best architectures, considering the correlation behavior and output cycles of calculation was 11-8-4-1. The importance of an input descriptor is determined by the sum of the absolute values of the weights of all the outgoing architecture connections from the input node to the next layer. Some factors, such as surface tension, percent of oxygen, index of refraction, and percentage H have appeared among the most important factors. The least important de-

Table 2. Various architecture of neural network and their criteria used in this study

Architecture	Layer number	Number of training cycles	Average error for training set	Average error for validation set
11-4-1	1	363	0.009987	0.008889
11-7-1	1	258	0.009941	0.009839
11-14-1	1	320	0.009998	0.009787
11-16-1	1	327	0.009981	0.008973
11-5-4-1	2	615	0.009987	0.009876
11-8-4-1	2	333	0.009924	0.00459
11-8-7-1	2	435	0.009932	0.009567
11-8-12-1	2	350	0.00996	0.00657
11-4-5-4-1	3	1259	0.09999	0.07789
11-8-4-4-1	3	1554	0.09999	0.09054
11-12-5-4-1	3	1198	0.08812	0.08639
11-12-7-3-1	3	947	0.06812	0.07687

scriptor was determined as the density. A range of predicted activity varied from 125.6796 to 3774.3753. The correlation coefficients between the experimental and the predicted MIC value pertaining to all the coumarins was 0.984 (Figure 1).

Compounds 67, 15, and 5 corresponded to the highest error that was generated during the training cycles. Y-Randomization result showed that the classification accuracy for randomized data sets was significantly lower than for the original data sets (data not shown) and hence we concluded that there is no evidence of over-fitting in our models. Cross validation is done by leave-some-out (some= 4) validating method. Validation showed that average of absolute errors was 0.379.

Discussion

The artificial neural networks (ANNs) have become an important modeling technique in numerous areas of chemistry and pharmacy

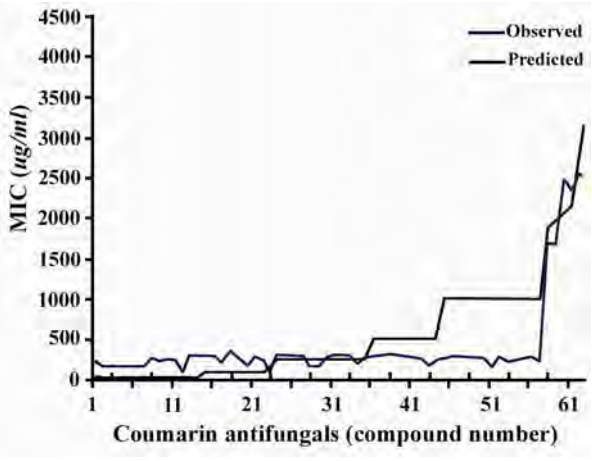


Figure 1. Plot of predicted activity versus the observed one

[30]. The mathematical adaptability of ANN commends them as a powerful tool for pattern classification and building predictive models. A particular advantage of ANNs is their inherent ability to incorporate nonlinear dependencies between the dependent and independent variables without using an explicit mathematical function.

This study presents an approach to correlate the antifungal activity score data for a data set of drug-like molecules with the structural descriptors. In this study a nonlinear modeling technique of artificial neural network (ANN) with back propagation learning algorithm and sigmoid activation function was used. In this work, a MLP network [29] was developed and used to obtain a nonlinear SAR model. Topologically, it consisted of input, hidden, and output layers of neurons or units connected by weights. Each input layer node corresponded to a single independent variable (physicochemical descriptor) with the exception of the bias node. Similarly, each output layer node corresponded to a different dependent variable (property under investigation). In this study, all descriptors were derived solely from molecular structures which did not require experimental data or expensive theoretical calculations (to be obtained).

The ANN model was trained only on the training set since the validation set was used to monitor the external prediction error and thus to avoid overtraining. Among the 11 architectures constructed, the best ANN architecture we found was 11–8–4–1. That is, in the first layer eleven inputs comprised of

eleven input descriptors, hidden layer comprised of seven neurons, and the last output layer comprised of one neuron for the property modeled. The statistical criteria obtained for the ANN model are shown in Table 2.

As it can be seen from this table the error for the training set is quite low. In addition, the errors for the validation set are also low showing the good prediction ability. The range of observed and predicted data criterion is very close to each other, that is, the overall prediction is close to experimental. Also, from these result we can conclude that the ANN model satisfactorily predicts the classification nature of the experimental data. Here, we should take into account that a large number of molecular descriptors are usually used in SAR methods. The specific biological action of drugs is frequently described by hydrophobic, electronic, steric and physicochemical properties. Physicochemical properties characterize the pharmacodynamic properties in the ligand– receptor interaction. They define the ability of the drug to join to the receptor.

The results of this ANN-based study indicate that surface tension is one of the most important factors in coumarin bioactivity. Surface tension of the molecule causes it to creep around the membrane, leading to formation of a layer of loaded molecules at the cell membrane quickly [31]. This finding could describe how the LogP is the main sensitivity descriptor of the trained network. Sensitivity analysis is a measure of how the outputs change when the inputs are changed. Result of this paper could help to predict bioactivity of new coumarins.

References

1. Denning DW, Evans GV, Kibbler CC, Richardson MD, Roberts MM, Rogers TR, et al. Guidelines for the investigation of invasive fungal infections in haematological malignancy and solid organ transplantation. Eur J Clin Microbiol Infec Dis 1997;16(6):424-436.
2. Coleman DC, Rinaldi MG, Haynes KA, Rex JH, Summerbell RC, Anaissie EJ, et al. Importance of Candida species other than Candida albicans as opportunistic pathogens. Med Mycol 1998;36 (Suppl1):156-165.
3. Warnock DW. Fungal infections in neutropenia: current problems and chemotherapeutic control. J Antimicro Chemother 1998;41(Suppl D):95-105.
4. Pfaller MA. Epidemiology of fungal infections: the promise of molecular typing. Clin Infect Dis 1995;20:1535-1539.
5. Tiew P, Ioset JR, Kokpal U, Chavasiri, W, Hostettmam K. Anti-fungal, anti-oxidant and larvicidal activities of compounds isolated from the heart wood of Mansonia gagei. Phytother Res 2003;17 (2):190-193.
6. Zaha AA, Hazem A. Antimicrobial activity of two novel coumarin derivatives: 3-cyanonaptho [1,2-(e)] pyran-2-one and 3-cyano coumarin. New Microbiol 2002; 25(2):213-222.
7. Lewis RJ, Singh OMP, Smith CV, Skarzynski T, Maxwell A, Wonacott AJ, et al. The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by X-ray crystallogarphy. EMBO J 1996;15(6):1412-1420.
8. Kawase M, Motohasi N, Sakagami H, Kanamato T, Nakashima H, Fereczy L, et al. Antimicrobial activity of trifluoromethyl ketones and their synergism with promethazine. Int J Antimicrob Agents 2001;18(2):161-165.
9. Kulkarni MV, Pujar BJ, Patil VD. Studies on coumarins II. Arch Pharm 1983;316(1):15-21.
10. Sardari S, Dezfulian M. Cheminformatics in anti-infective agents discovery. Mini Rev Med Chem 2007;79(2):181-189.
11. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural networks (ANN) modelling and its application in pharmaceutical research. J Pharm Biomed Anal 2000;22(5):717-727.
12. Zitko V. Chemometrics in environmental analysis. Chemomet Intel Lab Sys 1998;40:119-120
13. Sun Y, Peng Y, Chen Y, Shukla AJ. Application of artificial neural networks in the design of controlled release drug delivery systems. Adv Drug Del Rev 2003;55(9):1201-1215.
14. Wesołowski M, Konieczynski P. Thermoanalytical, chemical and principal component analysis of plant drugs. Int J Pharm 2003;262 (1-2):29-37.
15. Debeljak Z, Marohnic V, Srecnik G, Medic-Sÿaric M. Novel approach to evolutionary neural network based descriptor selection and QSAR model development. J Comput Aided Mol Des 2005;19 (12):835-855.

16. Baumann K. Cross-validation as the objective function for variable selection techniques. Trends Analyt Chem 2003;22(6):395-406.

17. Yamamura S. Clinical application of artificial neural networks (ANN) modeling to predict pharmacokinetic parameters of severely ill patients. Adv Drug Del Rev 2003;55(9):1233-1251.

18. Zupan J, Novic M, Ruisanchez I. Kohonen and counterpropagation artificial neural networks in analytical chemistry. Chemomet Intellig Lab Syst 1997;38(1);1-23.

19. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J Chem Inf Comput Sci 2003;43(6):1947-1958.

20. Sardari S, Mori Y, Horita K, Micetich RG, Nishibe S, Daneshtalab M. Synthesis and antifungal activity of coumarins and angular furanocoumarins. Bioorg Med Chem 1999;7(9): 1933-1940.

21. Godoya MFP, Victor SR, Bellini AM, Guerreiro G, Rochab WC,Bueno OC, et al. Inhibition of the symbiotic fungus of leaf-cutting ants by coumarins. J Braz Chem Soc 2005;16(3):669-672.

22. El-Seedi HR. Antimicrobial arylcoumarins from Asphodelus microcarpus. J Nat Prod 2007;70(1): 118-120.

23. Daoubi M, Duran-Patron R, Hmamouchi M, Hernandez-Gala´R, Ahmed Benharref, 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.

24. 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(1):134-144.

25. Mouri T, Yano T, Kochi S, Ando T, Hori M. Synthesis and antifungal activity of new 3,4,7-Trisubstituted coumarins. J Pestic Sci 2005;30(3): 209-213.

26. Stein AC, Alvarez S, Avancini C, Zacchino S, Poser GV. Antifungal activity of some coumarins obtained from species of Pterocaulon (Asteraceae). J Ethnopharmacol 2006;107(1):95-98.

27. 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.

28. Soltani S, Keymanesh K, Sardari S. Evaluation of structural features of membrane acting antifungal peptides by artificial neural networks. J Biol Sci 2008;8:834-845.

29. Bhatia MS, Ingale KB, Choudhari PB, Bhatia NM, Sawant RL. Application of quantum and physico-chemical molecular descriptors utilizing principal components to study mode of anticoagulant activity of pyridyl chromen-2-one derivatives. Bioorg Med Chem 2009;17(4):1654-1662.

30. Burns JA, Whitesides GM. Feed-forward neural networks in chemistry: mathematical systems for classification and pattern recognition. Chem Rev 1993;93(8):2583-2601.

31. Fiszelew A, Britos P, Ochoa A, Merlino H, Fernández E, García-Martínez R. Finding optimal neural network architecture using genetic algorithms. Res Comp Sci 2007;27:15-24.