Journal of Computational Science & Engineering 12 (2014) 438-444

Journal of Computational Science & Engineering

Available at http://www.asocse.org ISSN 1710-4068

QSAR Study of the Anti-cancer Activity of 38 Compounds in Different Cancer Cell Lines Based on Neural Networks

Lei Liu[1], Xiurui Han[1], Hongzong Si[2,*], Lianhua Cui[1]

[1]*Department of PublicHealth, Qingdao University, Qingdao, Shandong 266071, China*
[2]*Institute for Computational Science and Engineering, Laboratory of New Fibrous Materials and Modern Textile, the Growing Base State Key Laboratory, Department of Pharmacy, Qingdao University, Qingdao, Shandong 266071, China*

Article Information

Article history:
Received 5 April 2014
Revised 25 May 2014
Accepted 5 June 2014
Available online 9 June 2014

Keywords:
QSAR
Anticancer activity
RENs
MLP
RFP

Abstract

Using the MLP and the RBF, we established QSAR models for anti-cancer activity of 38 compounds in 5 cancer cell lines we established. These compounds are a novel class of anticarcinogen named tricyclic 5:7:5-fused diimidazo [4, 5-d:4', 5-f][1,3] diazepines. The carcinoma cell lines involved in this research are A549 lung cancer, MCF-7 breast cancer, MDA-MB-231 breast cancer, OVCAR-3 ovarian cancer and PC-3 prostate cancer. Descriptors of each molecular were calculated and some optimal were selected for building models. Accuracies of these models based on both methods in training groups and test groups are over 90%, showing perfect predictive ability. These QSAR models indicates that MLP and the RBF are great valuable in providing guidance for designing and synthesizing of anticancer drugs.

1. Introduction

The human RNA helicase DDX3 belonging to the DEAD box protein family of ATP-dependent RNA helicases is identified as a significant target in anticancer and antiviral chemotherapy [1]. The name "DEAD box" was given because of the conserved amino acid sequence D-E-A-D (Asp-Glu-Ala-Asp) existed in all member of the family [2]. DDX3 is ubiquitous in almost all species. It may promote progress of some kinds of cancers by means of impedance of death receptor-mediated apoptosis. Another possible way to aid cancer progression Richard S. Jope found is that DDX3 increases the level of transcription factor Snail in cells, which can repress the expression of cellular adhesion proteins [3]. DDX3 regulates the activity of caspase-7 induced by camptothecin in cells expressing wild-type p53, and caspase-7 is predominant executioner caspase in these cells [4].

* Corresponding author: sihz@126.com

2014 © The American Computational Science Society. All rights reserved.

In MCF-10A breast cancer cells, overexpression of DDX3 leads to an epithelial-mesenchymal-like transformation and facilitate the risk of neoplastic transformation of breast epithelial [5]. Though the known of DDX3 is still limited, it has attracted extensive attention for its commendable feasibility in anticancer and antiviral therapy.

Ring-expanded nucleosides (RENs), a certain type of nucleoside analogs, have been reported as promising drugs with biological activities against a wide spectrum of cancers and viruses with little toxicity to the host cells [6]. RENs inhibit activity of DDX3 and replication of HIV-1 toxicity in ex vivo cell culture or in vivo in mice [7]. Some tribenzoyl derivatives exhibit distinct suppression of human prostate cancer cells PC-3 and DU-145 [8].

Prediction of activities of medicines from their molecular structure has been needed for scientists for a long period. Quantitative structure-activity relationship (QSAR) is now a well-established and highly credible technique that correlates properties and activities of a compound with its structure, through a series of molecular features of this compound, called descriptors. Accordingly, the approach has the advantage of saving time and costing in drug design and investigation. Additionally, once the correlation has been acquired, it would predict the same activities or properties of other molecules that have not yet been measured or prepared [9]. We have established QSAR models for these compounds based on the Gene Expression Programming (GEP) [10]. In this study, using the Artificial neural networks (ANNs), we made other models for these compounds. These models showed promising predictive ability.

ANNs are computational algorithms with remarkable performance in classification and prediction. ANNs have been applied in various domains such as medicine, psychology, environmental science [11, 12, 13] for 20 or so years and they are still developing rapidly. An advance of ANNs is their remarkable ability to obtain a general result from complicated or imprecise data. ANNs consist of classes of architectures. Multi Layer Perceptron (MLP) and Radial Basis Function (RBF) were used in this study for their acknowledged well performance in classification.

Tricyclic 5:7:5-fused diimidazo [4,5-d:4',5'-f][1,3] diazepines are a novel category of RENs synthetized recently. Recent researches have proved that some of them had high activity against a number of cancer cell lines [14, 15]. This study aims to explore an efficient, stable method to predict anticancer activities of the compounds by their structures. Based on these experimental data, we established the QSAR models based on ANNs (MLP and BRF), separately, and selected out the optimal one for each kind of cancer. These models show excellent predict ability for anticancer activities of the series of compounds.

2. The ANNs theory

Artificial Neural Networks (ANNs) come from the simulation of human neural system. ANNs have a powerful capability of learning and reasoning from observations. It is the capability that allows the ANNs to seek out even highly non-linear relationships among complicated data.

As in nature, the function of ANNs can be owed to their elementary units usually called neurons as well as the extensive contacts among them called weights. In ANNs, these neurons are logically arranged into a series of layers: the input layer where each neuron represents an independent variable, the hidden layer the number of which may be various and the output layer. A typical ANN structure is showed in Figure 1. Each neuron located in a layer is connected with all neurons in next layer. In learning or training process, ANNs regulate the weights between neurons to form a predicted model for inputs and outputs. ANNs contain a series of different architectures. Multilayer Perceptron (MLP) and Radial Basis Function (RBF) are widely applied in classification as reliable approaches [16].

MLP consist of one input layer, one or more hidden layers and one output layer. MLP uses the back-propagation learning algorithm. Information

2014 © The American Computational Science Society. All rights reserved.

consisted in input layer is transmitted to the output layer by function of hidden layers. Each neuron in hidden layers sums weighted data consisted in inputs with errors. Then it transmits the sum to the output layer or the next hidden layer by a non-linear function, usually the sigmoid transfer function. The result of MLP is a combination of each neuron's result in output layer. MLP uses the back-propagation algorithm and keep regulating weights in training to get the global error minimized.

RBF network has an input layer, a hidden layer containing a large number of neurons and an output layer. The activation function of the neuron in hidden layer is radial basis function. It is a powerful non-linear function and gives the RBF network respectable modeling ability. Each radial basis function has two parameters: the center location of the function and its bias [17]. In RBF network, input data is transformed into hidden layer in an unsupervised form. The weight between the hidden layer and the output layer is training based on the output in a supervised way.

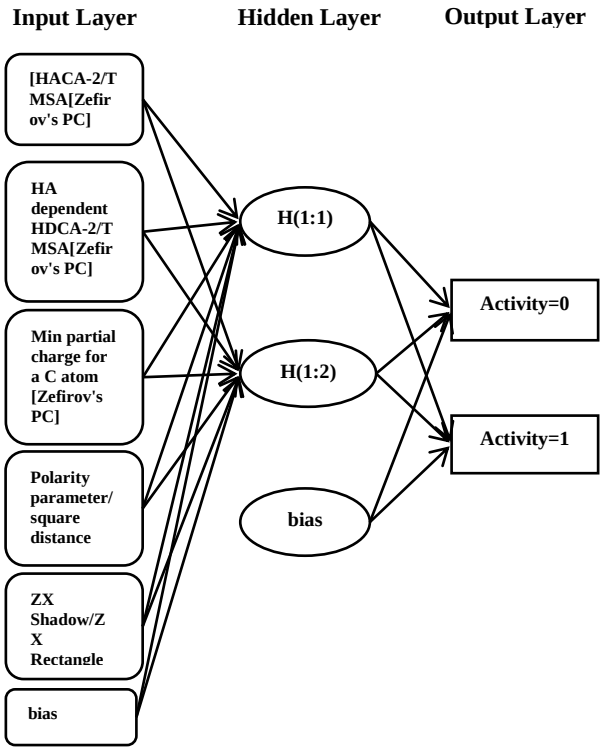


Figure 1. The structure of the ANNs implemented

3. Materials and Methods

3.1 Data set

The structures and experimental anticancer activity data of 38 tricyclic 5:7:5-fused diimidazo [4, 5-d:4', 5-f][1, 3] diazepines are taken from [14, 15]. Having considered a part of non-effective compounds to be contained in modeling, we decide to search a classification model, instead of a more commonly method-regression model. Original experimental data are divided into two classes: the active and the non-active. Then the active compounds are separated into a training set and a test set, and the non-active ones are dealt with all the same, randomly, as listed in the table 1. But that one compound is whether in the training set or in the test one is unknowable. We can only stipulate proportion of each group. The training set is employed for building the model and the test set is for evaluating its prediction ability.

3.2 Calculation of descriptors

All molecular structures are drawn into the ChemOffice 2004. They are pre-optimized by means of MM+ molecular force field and semi-empirical method AM1 in turn by Hyperchem Release [18]. Then MOPAC [19] is used for a more precise optimization and the output files are employed by CODESSA [20] to calculate descriptors of these compounds. A total of 163 calculators are calculated including geometrical, structural, topological, geometrical and quantum chemical descriptors. Finally, four to five most appropriate are selected by heuristic method of CODESSA to build the model and are showed in table 2.

3.3 Development of model by Artificial Neural Networks

MLP and RBF included in ANNs of IBM SPSS Statistics 18 are applied to build prediction models. To make results with less volatility, we operate the random number generators and set the fixed number as 2000000. The anticancer activity value (0 or 1) is input as a dependent variable and

2014 © The American Computational Science Society. All rights reserved.

descriptors as covariant. In ANNs, the order of those covariant may affect the result. We try different arrangements and select the best.

4. Results and discussion

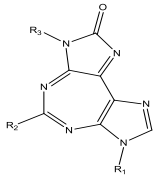
4.1 The ANNs model

The classification results of MLP and RBF models are listed in table 1.

Table 1 shows that the predicted value of anticancer activity correspond finely with the experimental results for most molecular, with the exception of 7a and 11c, of which models make unsatisfying predictions about anticancer activities in all five kinds of cancers. The false prediction of 7a may be attributed to the molecular weight. Observing compounds from 1 to 9b, we find that while substituents on R1 and R3 site of these compounds are the same, substituents on R2 site are getting more complex and molecular weights getting heavier. With the increasing molecular weights are decreasing activities. Compound 1, whose molecular weight is the smallest, is active. When the molecular weights increase, even the small change from hydrogen atom to a methyl on R2 site, the activities. Compound 1, whose molecular weight activities go away. The decreasing trends of activities raise the false prediction of 7a which is just near the boundary. Difference between compound 1 and 11c is the replacement of methoxyl group by fluorine atom on the site R1. We presume that anticancer activities may be let down by halogen atoms and the decrease gets sharper when it ranges from fluorine to iodine. Thus, similarly to 7a, the decreasing trend makes activities of the compound 11c which is around the boundary unpredictable. We hope getting more experimental data of activities of compounds with halogen atoms to justify the assumption above.

After getting rid of the compound 7a and 11c, these models achieve a perfect prediction on anticancer activities.

Table 1 Experimental and predicted values of anti-cancer activities for five kinds of cancers.



comp	R_1	R_2	R_3	A549lung			MCF-7breast			MDA Breast			OVCAR-3 ovarian			PC-3 prostate		
				exp	MLP	RBF	exp.	MLP	RBF	exp.	MLP	RBF	exp.	MLP	RBF	exp.	MLP	RBF
1	4-MeO-$C_6H_4CH_2$	H	4-MeO-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7a	4-MeO-$C_6H_4CH_2$	CH_3	4-MeO-$C_6H_4CH_2$	0	1*	1*	0	0	0	0	1*	0	0	0	0	0	1*	0
7b	4-MeO-$C_6H_4CH_2$	CH_2CH_3	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8a	4-MeO-$C_6H_4CH_2$	Ph	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8b	4-MeO-$C_6H_4CH_2$	4-CH_3Ph	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9a	4-MeO-$C_6H_4CH_2$	$(CH_2)_2COOH$	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9b	4-MeO-$C_6H_4CH_2$	$(CH_2)_3COOH$	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13a	$PhCH_2$	H	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13b	CH_2CH_3	H	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

2014 © The American Computational Science Society. All rights reserved.

13c	4-F-$C_6H_4CH_2$	H	4-MeO-$C_6H_4CH_2$	1	0*	0*	1	0*	0*	1	1	0*	0	0	0	1	1	0*
15a	4-MeO-$C_6H_4CH_2$	H	Ph	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15b	4-MeO-$C_6H_4CH_2$	H	4-MeO-Ph	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15c	4-MeO-$C_6H_4CH_2$	H	4-MeO-Ph-$C_6H_4CH_2CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	1*	0
15d	4-MeO-$C_6H_4CH_2$	H	CH_2CH_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15e	4-MeO-$C_6H_4CH_2$	H	CH_2COOEt	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0
15f	4-MeO-$C_6H_4CH_2$	H	CH_2CH_2Cl	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
15g	4-MeO-$C_6H_4CH_2$	H	$PhCH_2$	1	1	1	1	1	0*	1	1	1	1	1	1	1	1	0*
15h	4-MeO-$C_6H_4CH_2$	H	4-F-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15i	4-MeO-$C_6H_4CH_2$	H	4-Cl-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15j	4-MeO-$C_6H_4CH_2$	H	4-Br-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15k	4-MeO-$C_6H_4CH_2$	H	3-MeO-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15l	4-MeO-$C_6H_4CH_2$	H	2-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15m	4-MeO-$C_6H_4CH_2$	H	3-F-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15n	4-MeO-$C_6H_4CH_2$	H	3,4-$(F)_2$-$C_6H_3CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15o	4-MeO-$C_6H_4CH_2$	H	3-Cl-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15p	4-MeO-$C_6H_4CH_2$	H	3-Br-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15q	4-MeO-$C_6H_4CH_2$	H	3-F_3C-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15r	4-MeO-$C_6H_4CH_2$	H	4-OH-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15s	4-MeO-$C_6H_4CH_2$	H	4-$(CH_3OCH_2CH_2)$O-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15t	4-MeO-$C_6H_4CH_2$	H	4-$(HOCH_2CH_2)$O-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15u	4-MeO-$C_6H_4CH_2$	H	3-Pyridylmethyl	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15v	4-MeO-$C_6H_4CH_2$	H	4-$(CH_3)_2N$-$C_6H_4CH_2$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15w	4-MeO-$C_6H_4CH_2$	H	CH_2CH_2Piperidinyl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15x	4-MeO-$C_6H_4CH_2$	H	CH_2CH_2Morpholinyl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4-MeO-$C_6H_4CH_2$	2,4,5-$(CO_2H)_3$-C_6H_2	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	4-MeO-$C_6H_4CH_2$	H	2-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	4-MeO-$C_6H_4CH_2$	H	2,4-$(F)_2$-C_6H_3	0	0	0	0	1*	0	0	0	0	0	0	0	0	0	0
14	2-Cl-$C_6H_4CH_2$	H	4-MeO-$C_6H_4CH_2$	0	0	0	0	0	0	0	1*	0	0	0	0	0	1*	0

The star(*) means the false prediction.

2014 © The American Computational Science Society. All rights reserved.

Table 2 The involved molecular descriptors.

A549 Lung cancer	MCF-7 Breast cancer	MDA-MB-231 Breast cancer	OVCAR-3 Ovarian cancer	PC-3 Prostate cancer
HACA-2/TMSA[Zefirovv's PC]	ZX Shadow/ZX Rectangle	ZX Shadow/ZX Rectangle	ZX Shadow/ZX Rectangle	ZX Shadow/ZX Rectangle
HA dependent HDCA-2/TMSA[Zefirovv's PC]	Polarity parameter/square distance	Polarity parameter/square distance	Polarity parameter/square distance	Polarity parameter/square distance
Min partial charge for a C atom [Zefirov's PC]	Min partial charge for a C atom [Zefirov's PC]	Min partial charge for a C atom [Zefirov's PC]	Min partial charge for a C atom [Zefirov's PC]	Min partial charge for a C atom [Zefirov's PC]
Polarity parameter/square distance	MAX partial charge for a O atom[Zefirov's PC]	YZ Shadow/YZ Rectangle	Max partial charge for a O atom[Zefirov's PC]	MAX partial charge for a O atom[Zefirov's PC]
ZX Shadow/ZX Rectangle	Relative number of single bonds		Relative number of single bonds	Average structural Information content (order 1)

From table 2 we observe that there are three descriptors involved in all 5 models: Min partial charge for a C atom [Zefirov's PC], Polarity parameter/square distance and ZX Shadow/ZX Rectangle. It indicates that the three descriptors have an important effect on anticancer activities for these five kinds of cancers. Limited by finite experimental data and the classification approach, whether these descriptors make a positive or a negative effect on anticancer activities remains inconclusive.

Table 3 Accuracies of two models.

	MLP		RFP	
	training	test	training	test
A549	95.50%	93.80%	92%	100%
MCF	91.70%	100%	96%	92.30%
MDA	91.70%	100%	100%	90.90%
OVA	100%	100%	100%	100%
PC	92.60%	90.90%	92%	100%

Accuracies of two models are listed in Table 3. Accuracies of these models in training groups and test groups are over 90%, showing perfect predictive ability.

5. Conclusion

In this study, using the MLP and the RBF we build QSAR models to predict the anticancer activities of tricyclic 5:7:5-fused diimidazo [4, 5-d:4', 5-f][1,3] diazepines against five kinds of cancer cell lines. We get the optimal model for each kind of cancer. These models show powerful and stable predict ability, which indicates that the MLP and the RBF are excellent methods for predicting anticancer activities of this kind of compound.

Acknowledgements

This work was supported by Qingdao University Innovative Teaching Laboratory Research Project and the Shandong Province Department of Project: No. 2012YD18038 and No. 2012YD18042.

References

[1] E. Jankowsky, Trends in Biochemical Sciences, 6(2010) 19-29

[2] S. Rocak, P. Linder, N.R. Mol. Cell Biol, 5 (2004) 232-241

2014 © The American Computational Science Society. All rights reserved.

[3] M. Sun, L. Song, T. Zhou, G.Y. Gillespie, R.S. Jope, Biochimica et Biophysica Acta, 1813 (2011) 438-447
[4] M. Sun, T. Zhou, E. Jonasch, R.S. Jope, Biochimica et Biophysica Acta, 1833(2013) 1489-1497
[5] M. Botlagunta, F. Vesuna, Y. Mironchik, A. Raman, A. Lisok, P.W. Jr, S. Mukadam, P.V. Diest, J.H. Chen, P. Farabaugh, A.H. Patel, V. Raman, Oncogene, 27 (2008) 3912-3922
[6] R.S. Hosmane, Current Topics in Medicinal Chemistry, 2(2002) 1093-1109
[7] V. Yedavalli, N. Zhang, H.R. Cai, K.T. Jeang, R. Hosmane, Antiviral Research, 2(2008) A68
[8] J.K. Gill, L.J Wang, M. Bretner, R. Newman, N. Kyprianou, R.S. Hosmane, Nucleosides, Nucleotides and Nucleic Acids, 20(2011) 1043-1045
[9] A.R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Chem.Rev, 10(2010) 5714–5789
[10] X.R. Han, X.C. Li, H.Z. Si, C.Z. Ge, H. Gao, H.L. Zhai, Advanced Materials Research, 850-851 (2014) 1291-1294
[11] C. Smyser, T. Inder, J. Shimony, J. Hill, A. Degnan, A. Snyder, J. Neil, Cereb. Cortex, 12 (2010) 2852-2862
[12] Bowers, S. Jeffrey, Psychol. Rev, 1(2009)220
[13] S. Bayar, I. Demir, G.O. Engin, Ecotoxicology and Environmental Safety, 3 (2009) 843-850
[14] A. Kondaskar, S. Kondaskar, J.C. Fishbein, B.A. Carter-Cooper, R.G. Lapidus, M. Sadowska, M.J. Edelman, R.S. Hosmane, Bioorganic & Medicinal Chemistry, 21(2013) 618-631
[15] A. Kondaskar, S. Kondaskar, R. Kumar, J.C. Fishbein, N. Muvarak, R.G. Lapidus, M. Sadowska, M.J. Edelman, G.M. Bol, F. Vesuna, V. Raman, R.S. Hosmane, ACS Med. Chem, 2(2011)252 -256
[16] S. Cohen, N. Intrator, Multiple Classifier Systems, 2709(2003) 227-235
[17] I. Yilmaz, O. Kaynar, Expert Systems with Applications, 38 (2011) 5958-966
[18] HyperChem. 4.0, Hypercube, 1994
[19] JJP Stewart, Comput. Chem, 1989.
[20] A.R. Katritzky, V.S. Lobanov, M. Karelson, Reference Manual, Version 2.0, 1994

2014 © The American Computational Science Society. All rights reserved.