Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Distinguishing compounds with anticancer activity by ANN using inductive QSAR descriptors

Kunal Jaiswal[1] and Pradeep Kumar Naik[1,*]

[1]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Distt.-Solan, Himachal Pradesh, India-173215; Pradeep Kumar Naik* - E-mail: pradeep.naik@juit.ac.in; Phone: 91 1792 239227; * Corresponding author

received May 19, 2008; revised June 17, 2008; accepted July 06, 2008; published July 30, 2008

Abstract:
This article describes a method developed for predicting anticancer/non-anticancer drugs using artificial neural network (ANN). The ANN used in this study is a feed-forward neural network with a standard back-propagation training algorithm. Using 30 'inductive' QSAR descriptors alone, we have been able to achieve 84.28% accuracy for correct separation of compounds with- and without anticancer activity. For the complete set of 30 inductive QSAR descriptors, ANN based method reveals a superior model (accuracy = 84.28%, Q_{pred} = 74.28%, sensitivity = 0.9285, specificity = 0.7857, Matthews correlation coefficient (MCC) = 0.6998). The method was trained and tested on a non redundant data set of 380 drugs (122 anticancer and 258 non-anticancer). The elaborated QSAR model based on the Artificial Neural Networks approach has been extensively validated and has confidently assigned anticancer character to a number of trial anticancer drugs from the literature.

Keywords: artificial neural network; inductive QSAR descriptors; anticancer drugs; non-anticancer drugs

Background:
A number of natural and synthetic products have been found to exhibit anticancer activity against tumor cell lines **[1, 2]**. Eventually, the number of anticancer drugs is increasing exponentially day by day. Hence, discrimination between anticancer and non-anticancer drugs is a major challenge in current cancer research. The worldwide pharmaceutical industry is investing in technologies for high-throughput screening (HTS) of such compounds. Therefore, development of *in silico* techniques for anticancer drug screening is the demand of today's anticancer drug discovery. The use of computational tools for discrimination of anticancer drugs from lead molecules prior to their chemical synthesis will accelerate the drug discovery processes in the pharmaceutical industry **[3]**.

Early-phase virtual screening and compound library design often employs filtering routines, which are based on binary classifiers and are meant to eliminate potentially unwanted molecules from a compound library **[4, 5]**. Currently two classifier systems are most often used in these applications: PLS-based classifiers **[6, 7]** and various types of artificial neural networks **[8, 9]**. Quantitative structure activity relationship (QSAR) science uses a broad range of atomic and molecular properties ranging from merely empirical to the '*ab initio*' computed. The most commonly used QSAR based methods can include up to thousands of descriptors readily computable for extensive molecular datasets. Such varieties of available descriptors in combination with numerous powerful statistical and machine learning techniques such as Artificial Neural Networks (ANN) allow distinguishing biologically active from non-active substances **[10, 11]**.

Currently various sets of molecular descriptors are available **[12]** and thus for application to anticancer drug/non-drug classification of compounds, the molecules can be typically represented by n-dimensional vectors **[10, 11]**. In the current work, we focused on the 'inductive' QSAR descriptors **[13]** for anticancer/non-anticancer drug classification. These include various local parameters calculated for certain kinds of bound atoms (for instance; for most positively/negatively charges etc), groups of atoms (for substituent with the largest/smallest inductive or steric effect within a molecule) or computed for the entire molecule. All these descriptors (except the total formal charge) depend on the actual spatial structure of molecules. These inductive descriptors found broad application for quantification of antibacterial activity of synthetic cationic polypeptides **[13]**. The demand for computational screening methodology is clear in all areas of human therapeutics. However, the field of anti-cancer drugs has a particular need for computational solutions enabling rapid identification of novel therapeutic leads. QSAR approaches for classification of anticancer compounds against non-anticancer agents represents an important and valuable task for the modern QSAR research.

The main objective of this study was to develop a scheme for encoding relevant information from molecular structure into a format which is suitable for use in ANN and to develop a

QSAR model of the binary classification of anticancer/non-anticancer drugs with predictive capabilities, which so far has been unattainable.

Methodology:

Dataset

To investigate the possibility of using the inductive QSAR descriptors for creation of an effective model of discrimination between anticancer/non-anticancer drugs, we have considered a dataset of 380 structurally heterogeneous compounds including 122 non-redundant anticancer **[14, 15]** and 258 non-redundant non-anticancer drugs. All the 122 anticancer drugs were taken from the NCI anti-cancer agent mechanism database **[16]** and have been proved to have well known mechanism of action (Table 1 under supplementary material) whereas; all the 258 non-redundant non-anticancer drugs were taken from DrugBank **[17]**.

Descriptors calculation and selection

A set of 50 inductive descriptors have been calculated initially for all the 380 drugs. During calculation the hydrogen atoms were suppressed and only the heavy atoms have been taken into account. The inductive QSAR descriptors were calculated from values of atomic electro-negativities and radii by using the custom SVL-scripts downloaded from the SVL exchanger **[18]** and implemented within the MOE package (Chemical Computing Group Inc 2005). To avoid cross correlation among the independent variables, we have computed pair-wise correlation among all the 50 QSAR parameters and removed those inductive descriptors which formed any linear dependence with $R \geq 0.9$. As a result of this procedure, only 30 inductive QSAR descriptors have been selected (Table 2 see supplementary material). The normalized values (in the scale of 0-1) of these 30 parameters have been used to generate QSAR models.

Composition of the training and testing sets

For effective training of the network (primarily to avoid over-fitting), we have used the training sets of 342 compounds (including 100 anticancer drugs) randomly derived out of the 380 molecules. Such random sampling has been performed 20 times and 20 independent QSAR models have been created. In each training run the remaining 10 percent of the compounds were used as the testing set in order to evaluate the average predictive ability of the method. The given performance measures have been averaged over five QSAR models.

ANN model for classification of anticancer/non-anticancer drugs

In order to relate the inductive descriptors to anticancer activity of the studied molecules we have employed the standard back-propagation ANN using Stuttgart Neural Network Simulator package **[19]**. The ANN used in this study consists of 30 input nodes, depicting 30 inductive QSAR descriptors and 1 output node. The number of nodes in the hidden layer varied from 2 to 40 in order to find the optimal network that allows most accurate separation of anticancer/non-anticancer drugs in the training sets. During the learning phase, a value of 1 was assigned for the anticancer drugs and 0 to the others. For each configuration of the ANN (with 2, 4, 6, 8, 10, 12, 14, 20 and 40 hidden nodes respectively) 20 independent training runs were performed to evaluate the average predictive power of the network. The corresponding counts of the true positive, true negative, false positive and false negative predictions have been estimated using 0.4 and 0.6 cut-off values for non-anticancer and anticancer respectively. Thus, an anticancer drug from the testing set has been considered classified correctly by the ANN only when its output value ranged from 0.6 to 1.0. Similarly, for each non-anticancer drug of the testing set, the correct classification has been obtained if the ANN output lay between 0 and 0.4. Thus, all network output values ranging from 0.4 to 0.6 have been ultimately considered as incorrect predictions (rather than undetermined or non-defined).

Performance measures

The prediction results from neural network model were evaluated using the following statistical measures like accuracy, Matthews correlation coefficient (MCC), sensitivity (Q_{sens}), specificity (Q_{spec}), probability of correct prediction (Q_{pred}) by using the equations given under supplementary material.

Results and discussion:

The accuracy of distinguishing of anticancer compounds by the artificial neural networks built upon the 'inductive' descriptors clearly demonstrates the adequacy and good predictive power of the developed QSAR model. There is strong evidence that the introduced inductive descriptors do adequately reflect the structural properties of chemicals, which are relevant to their anticancer activity. This observation is not surprising, considering the inductive QSAR descriptors calculated should cover a very broad range of proprieties of bound atoms and molecules related to their size, polarizability, electro-negativity, compactness, mutual inductive, steric influence and distribution of electronic density, etc. The average value for both the classes were separated to quite an extent on the graph and the selected 30 inductive descriptors should allow building of an effective QSAR model for binary classification.

Considering the most important implication of the "anticancer-likeness" model is its potential use for identification of novel anticancer drug candidates from electronic databases, we have calculated the parameters of the positive predictive values (PPV) for the networks while varying the number of hidden nodes. Taking into account the PPV values for the networks with the varying number of the hidden nodes along with the corresponding values of sensitivity, specificity and general accuracy, we have selected neural network with six hidden nodes as the most efficient among the studied ANNs (Table 2 in supplementary material). The ANN with 30 input, 6 hidden and 1 output nodes has allowed the recognition of 84% of anticancer and 84% of non-anticancer compounds on average. The output from this 30-6-1 network has also demonstrated very good separation on positive (anticancer) and negative (non-

anticancer) predictions, which revealed a superior model (accuracy = 84.28%, Q_{pred} = 74.28%, sensitivity = 0.9285, specificity = 0.7857, Matthews correlation coefficient (MCC) = 0.6998) (Table 2 in supplementary material). The vast majority of the predictions for the testing sets consisting of ⅓ of anticancer and ⅔ of non-anticancer compounds, has been contained within 0.0 - 0.4 for non-anticancer and 0.6 - 1.0 for anticancer drugs which also illustrates that 0.4 and 0.6 cut-off values provide very adequate separation of two bioactive classes (Table 3 and 4 (see supplementary material) feature the output values from the 30-6-1 ANN for the training and testing sets respectively). Presumably, accuracy of the approach operating by the inductive descriptors can be improved even further by expanding the QSAR descriptors or by applying more powerful classification technique such as Support Vector Machine. Use of merely statistical techniques in conjunction with the inductive QSAR descriptors would also be beneficial, as they allow interpreting individual descriptor contributions into molecular "anticancer-likeness". Nonetheless, despite certain drawbacks, it is obvious that the developed ANN-based QSAR model operating by the inductive descriptors has demonstrated very high accuracy and can be used for mining electronic collections of chemical structures for novel anticancer candidates.

An application of the model

The developed QSAR model of distinguishing anticancer drugs was validated further based on the anticancer compounds published in the journal 'Nature Review Drug Discovery', July 2004, spplement HOT DRUGS 2004; and 'Current Pharmaceutical Design', 2000. The "experimental" anticancer drugs cited by the Nature Review includes Gefitinib (an inhibitor of Tyrosine Kinase) and Abarelix (inhibit production of androgens involved in prostrate cancer). The drugs Etoposide and Teniposide and their involvement in cancer treatments are published in Current Pharmaceutical Design **[20]**. The corresponding structural formulas and their prediction results as anticancer drugs were presented in Table 6 under supplementary material. The predicted output of all the 12 drugs was above 0.60, the threshold value for predicting as anticancer drugs by the model. These results demonstrate that the ANN-based binary classifier of anticancer/non-anticancer drugs is adequate and can be considered an effective tool for '*in silico*' anticancer drugs screening. The results also demonstrate that the inductive parameters readily accessible from atomic electronegativities, covalent radii and interatomic distances can produce a variety of useful QSAR descriptors to be used in '*in silico*' chemical research.

Conclusion:

The results of the present work demonstrate that a variety of atomic, substituent and molecular properties which can be computed within the framework of inductive and steric effects, inductive electro-negativity and molecular capacitance represent a powerful arsenal of 3D QSAR descriptors for modern '*in silico*' drug research. Using only 30 inductive descriptors with no additional independent parameters, we have achieved 84.28% accuracy for distinguishing compounds with and without anticancer activity. The selected set of inductive descriptors possesses a number of important merits. They are 3D and stereo-sensitive which can be easily computed from fundamental properties of bound atoms and molecules and possess much defined physical meaning. This ANN-based model for anticancer drug prediction can be used as a powerful QSAR tool for filtering out lead molecules to discover novel anticancer drugs.

References:

[01] M. E. Bracke, *et al., Anticancer Agents Med Chem*., 8: 171 (2008) [PMID: 18288920]

[02] L. Sartor, *et al., Biochemical Pharmacology*, 64: 229 (2002) [PMID: 12123743]

[03] P. Buchwald and N. Bodor, *J Med Chem.,* 49: 883 (2006) [PMID: 16451054]

[04] D. E. Clark and S. D. Pickett, *Drug Discovery Today*, 5: 49 (2000) [PMID: 10652455]

[05] G. Schneider and H. J. Böhm, *Drug Discovery Today*, 7: 64 (2002) [PMID: 11790605]

[06] S. Wold, *J Pharm Biomed Anal.*, 9: 589 (1994) [PMID: 1790178]

[07] M. Forina, *et al., J Pharm. Biomed. Anal.,* 18: 21 (1998) [PMID: 9863940]

[08] J. Sadowski, *Current Opinion in Chemical Biology*, 4: 280 (2000) [PMID: 10826977]

[09] G. Schneider, *Neural Network*, 13: 15 (2000) [PMID: 10935454]

[10] A. Ajay, *et al., Journal of Medicinal Chemistry*, 41: 3314 (1998) [PMID: 9719583]

[11] J. Sadowski and H. A. Kubinyi, *Journal of Medicinal Chemistry*, 41: 3325 (1998) [PMID: 9719584]

[12] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, *Wiley-VCH*, *New York* (2000)

[13] A. Cherkasov and B. Jankovic, *Molecules,* 9: 1034 (2004) [PMID: 18007503]

[14] J. N. Weinstein, *et al.*, *Science*, 258: 447 (1992) [PMID: 1411538]

[15] W. W. Van Osdol, *et al., Journal of the National Cancer Institute*, 86: 1853 (1994) [PMID: 7990160]

[16] http://dtp.nci.nih.gov/docs/cancer/searches/standard_mechanism.html

[17] D. S. Wishart, *et al., Nucleic Acids Research*, 34: D668 (2006) [PMID: 16381955]

[18] http://svl.chemcomp.com/index.php

[19] http://www.ra.cs.uni-tuebingen.de/SNNS/

[20] M. Gordaliza, *et al., Current Pharmaceutical Design*, 6: 1811 (2000) [PMID: 11102564]

Edited by P. Kangueane

Citation: Jaiswal and Naik, Bioinformation 2(10): 441-451 (2008)

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Supplementary material

Equations

Accuracy (Q_{ACC}) $Q_{ACC} = (TP + TN) / (TP + TN + FP + FN)$ → **(1)**

where TP, FP, TN and FN refer to true positives, false positives, true negatives and false negatives, respectively.

Matthews correlation coefficient (MCC)

$$MCC = \frac{(TP.TN - FP.FN)}{\sqrt{(TP+FN).(TP+FP).(TN+FN).(TN+FP)}}$$ → **(2)**

Sensitivity (Q_{sens}) $Q_{sens} = TP / (TP + FN)$ → **(3)**

Specificity (Q_{spec}) $Q_{spec} = TN / (TN + FP)$ → **(4)**

Probability of correct prediction (Q_{pred}) $Q_{pred} = (TP / (TP + FP)) \times 100$ → **(5)**

Tables

Anti-cancer drugs by mechanism	Number of drug molecules
Alkylating agents	36
Antimitotic agents	13
Topoisomerase I inhibitors	24
Topoisomerase II	15
inhibitorsRNA/DNA antimetabolites	18
DNA antimetabolites	16
Total	122

Table 1: A dataset of 122 anti-cancer drugs used in the study with their mechanism of action.

Descriptor	Characterization	Descriptor	Characterization
Average_EO_Pos	arithmetic mean of electronegativities of atoms with positive partial charge	*Most_Pos_Rs_i_mol*	Steric influence *Rs(atom→molecule)* OF the most positively charged atom to the rest of a molecule
Average_Hardness	arithmetic mean of hardnesses of all atoms of a molecule	*Most_Pos_Sigma_i_mol*	Largest positive atomic inductive parameter *σ*(atom→molecule)* for atoms in a molecule
*Average_Neg_Charge**	Arithmetic mean of negative partial charges on atoms of a molecule	*Smallest_Neg_Softness*	Smallest atomic softness among values for negatively charged atoms
Average_Neg_Hardness	arithmetic mean of hardnesses of atoms with negative partial charge	*Smallest_Pos_Hardness*	Smallest atomic hardness among values for positively charged atoms
Average_Neg_Softness	Arithmetic mean of softnesses of atoms with negative partial charge	*Smallest_Pos_Softness*	Smallest atomic softness among values for positively charged atoms
*Average_Pos_Charge**	Arithmetic mean of positive partial charges on atoms of a molecule	*Smallest_Rs_i_mol*	Smallest value of atomic steric influence *Rs(atom→molecule)* in a molecule
Average_Pos_Softness	Arithmetic mean of softnesses of atoms with positive partial charge	*Smallest_Rs_mol_i**	Smallest value of group steric influence *Rs(molecule→atom)* in a molecule
*Largest_Neg_Hardness**	Largest atomic hardness among values	*Softness_of_Most_Neg*	Atomic softness of an atom with

	for negatively charged atoms		the most negative charge
Largest_Rs_i_mol	Largest value of atomic steric influence *Rs(atom→molecule)* in a molecule	*Softness_of_Most_Pos*	Atomic softness of an atom with the most positive charge
Most_Neg_Charge	Largest partial charge among values for negatively charged atoms	*Sum_Hardness**	Sum of hardnesses of atoms of a molecule
Most_Neg_Rs_i_mol	Steric influence R*s(atom→molecule)* OF the most negatively charged atom to the rest of a molecule	*Sum_Neg_Hardness*	Sum of hardnesses of atoms with negative partial charge
*Most_Neg_Rs_mol_i**	Steric influence R*s(molecule→atom)* ON the most negatively charged atom in a molecule	*Sum_Pos_Hardness*	Sum of hardnesses of atoms with positive partial charge
*Most_Neg_Sigma_i_mol**	Largest negative atomic inductive parameter *σ*(atom→molecule)* for atoms in a molecule	*Sum_Neg_Sigma_mol_i**	Sum of all negative group inductive parameters *σ*(molecule →atom)*within a molecule
*Sum_Neg_Sigma_mol_i**	Sum of all negative group inductive parameters *σ*(molecule →atom)* within a molecule	*Total_Charge_Formal**	Sum of charges on all atoms of a molecule (formal charge of a molecule)
*Most_Pos_Charge**	Largest partial charge among values for positively charged atoms	*Total_Neg_Softness**	Sum of softnesses of atoms with negative partial charge

Table 2: The thirty 'Inductive QSAR Descriptors' used in the study.

Hidden nodes	Specificity	Sensitivity	Accuracy Q(Total)	Q (Pred in %)	MCC
2	0.7674	0.9259	0.8285	71.42	0.6750
4	0.7674	0.9259	0.8285	71.42	0.6750
6	0.7857	0.9285	0.8428	74.28	0.6998
8	0.7674	0.9259	0.8285	71.42	0.6750
10	0.7674	0.9259	0.8285	71.42	0.6750
12	0.7500	0.9230	0.8142	68.57	0.6504
14	0.7500	0.9230	0.8142	68.57	0.6504
20	0.7500	0.9230	0.8142	68.57	0.6504
40	0.7500	0.9230	0.8142	68.57	0.6504

Table 3: Parameters of specificity, sensitivity, accuracy and positive predictive values for prediction of anticancer and non-anticancer compounds by the artificial neural networks with the varying number of hidden nodes. The cut-off values 0.4 and 0.6 have been used for negative and positive predictions respectively.

Name	Output	Name	Output
Anticancer		Maytansine	0.850
asaley	0.973	Rhizoxin	0.981
busulfan	0.702	carboxyphthalatoplatinum	0.982
Thiopurine	0.881	Taxol derivative	0.733
CBDCA	0.938	chlorozotocin	0.606
CCNU	0.825	cis-platinum	0.983
CHIP	0.977	clomesone	0.770
Taxol	0.920	Vincristine sulfate	0.984
cyclodisone	0.926	Camptothecin	0.974
dianhydrogalactitol	0.604	Camptothecin Na salt	0.970
fluorodopan	0.985	Aminocamptothecin	0.938
hepsulfam	0.974	Hydroxycamptothecin	0.860

hycanthone	0.982	Camptothecin acetate	0.915
melphalan	0.985	14-Chloro-20(S)-camptothecin hydrate	0.973
Methyl CCNU	0.985	9-Amino-20-(R,S)-camptothecin	0.984
Mitomycin C	0.984	Camptothecin analog	0.630
Piperazine	0.978	7-Chlorocamptothecin	0.963
Piperazinedione	0.979	Camptothecin analog-monohydrochloride	0.967
Pipobroman	0.984	Camptothecin,20-O-((4-(2-hydroxyethyl)-1-piperazino)OAC	0.980
Porfiromycin	0.980	Camptothecin, 9-methoxy-	0.955
GLYCINATE	0.847	Camptothecin, 4-ethyl-4-hydroxy-11-methoxy	0.983
Teroxirone	0.960	11-Formyl-20(RS)-camptothecin	0.901
Tetraplatin	0.976	11-Hydroxymethyl-20(RS)-camptothecin	0.700
Thio-tepa	0.966	Camptothecin phosphate	0.850
PALA	0.786	Camptothecin-20-O-(N,N-dimethyl)glycinate HCl	0.577
m-AMSA	0.658	Camptothecin lysinate HCl	0.962
Yoshi-864	0.919	Camptothecin glutamate HCl	0.971
Colchicines	0.919	Camptothecin butylglycinate ester hydrochloride	0.907
Mitoxantrone	0.938	Camptothecin hemisuccinate sodium salt	0.946
Dolastatin 10	0.963	Spirohydantoin mustard	0.984
menogaril	0.600	Camptothecin ethylglycinate ester hydrochloride	0.965
Oxanthrazole	0.694	Morpholino-ADR	0.984
Rubidazone	0.791	Halichondrin b	0.951
VM-26	0.723	Amonafide	0.977
VP-16	0.981	Uracil nitrogen mustard	0.979
L-alanosine	0.736	Anthrapyrazole derivative	0.665
5-azacytidine	0.636	Pyrazoloacridine	0.975
Acivicin	0.705	Bisantrene HCL	0.970
An antifol	0.901	Daunorubicin	0.980
3-HP	0.978	Deoxydoxorubicin	0.983
Pyrazofurin	0.912	Colchicines derivative	0.975
Name	**Output**	**Name**	**Output**
Trimetrexate	0.822	N,N-dibenzyl daunomycin	0.884
Ara-C	0.909	L-Aspartic acid, aminopterine,	0.895
Beta-TGDR	0.963	L-Aspartic acid, aminopterine- sesquihydrate	0.915
cyclocytidine	0.983	Aspartic acid, N-[2-chloro-4-[[(2, 4-diamino-6-pteridinyl)methyl]amino]benzoyl]-, monohydrate, L-	0.955
Guanazole	0.911	Baker's soluble antifol	0.971
hydroxyurea	0.983	Dichlorallyl lawsone	0.937
Macbecin II	0.837	Aphidicolin glycinate	0.929
pyrazoloimidazole	0.942	5-aza-2'-deoxycytidine	0.873
Thioguanine	0.739	5,6-dihydro-5-azacytidine	0.938
5-HP	0.850	Methotrexate derivative	0.857
Alpha-TGDR	0.882	cyanomorpholinodoxorubicin	0.908
Alpha-TGDR	0.954	2'-deoxy-5-fluorouridine	0.915
Brequinar	0.924	Inosine glycodialdehyde	
Ftorafur	0.943	Triethylenemelamine	
Non-anticancer			
2-amino-4-picoline	0.258	5-bromosalicylic acid acetate	0.258
bezafibrate	0.256	5-nitro-2propoxyacetanilide	0.280
binifibrate	0.319	5-nitro-2propoxyacetanilide	0.258
bisoprolol	0.184	acecarbromal	0.431
bitolterol	0.004	aceclofenac	0.258
bucloxic acid	0.258	acefylline(c,d,e,g)	0.541
Bromfenac	0.258	emorfazone	0.348
bufexamac	0.327	bromisovalum	0.258
Alphaprodine	0.108	bromodiphenhydramine	0.057
Alprenolol	0.249	acetylsalicylic acid	0.158
Amosulalol	0.328	alminoprofen	0.248
Anileridine	0.218	Bufuralo	0.008

Antipyrine	0.168	Bunitrolol	0.238
Antrafenine	0.258	Bumadizon	0.418
Apazone	0.290	Butallylonal	0.032
Apronalide	0.259	Butanilicaine	0.293
Bamifylline	0.257	Butibufen	0.258
Capuride	0.066	butidrine hydrochloride	0.015
carbiphene	0.258	Butoctamide	0.255
carbocloral	0.313	diethylbromoacetamide	0.031
carbromal	0.257	difenpiramie	0.004
carbuterol	0.258	diflunisal	0.258
carfimate	0.263	dilevalol	0.162
carprofen	0.258	dioxadrol	0.279
carteolol	0.000	dipyrocetyl	0.000
carvedilol	0.259	carsalam	0.004
doxofylline	0.255	celiprolol	0.001
droperidol	0.000	cetirizine	0.258
dyphylline	0.315	chlorobutanol	0.311
Name	**Output**	**Name**	**Output**
chlorothen	0.001	dipyrone	0.001
chlorprothixene	0.041	doxefazepam	0.002
chlorcyclizine	0.270	ephedrine	0.258
cinmetacin	0.410	eprozinol	0.259
embramine	0.244	Etafedrine	0.335
enfenamic acid	0.010	etaqualone	0.258
epanolol	0.256	etersalate	0.259
epirizole	0.246	ethinamate	0.258
estazolam	0.258	ethoxazene	0.004
etamiphyllin	0.229	ciprofibrate	0.261
eterobarb	0.237	clenbuterol	0.261
ethenzamide	0.002	clinofibrate	0.179
ethoheptazine	0.050	clometacin	0.323
cinromida	0.070	clonixin	0.259
clemastine	0.095	clordesmetildiazepam	0.783
clidanac	0.017	cropropamide	0.259
clofibric acid	0.258	fentanyl	0.213
clometiazol	0.078	floctafenine	0.000
cloranolol	0.251	fluoresone	0.289
clozapine	0.248	lornoxicam	0.258
fluphenazine	0.039	loxoprofen	0.256
loxapina	0.234	medibazine	0.000
mecloqualone	0.258	formoterol	0.095
flupirtine	0.282	flurazepam	0.481
flutropium bromide	0.255	fluspirilene	0.005
fluproquazone	0.256	medrylamine	0.214
flurbiprofen	0.292	mepindolol	0.256
methafurylene	0.179	mequitazine	0.260
methyltyrosine	0.260	methaphenilene	0.512
metiapine	0.003	methyldopa	0.611
metofoline	0.051	gentisic acid	0.337
metron	0.002	glucametacin	0.254
pyrilamine	0.348	haloperidide	0.180
hydroxyzine	0.035	hexapropymate	0.126
ibuprofen	0.005	methyldopa	0.003
indenolol	0.262	methyprylon	0.250
isoetharine	0.002	metipranolol	0.259
morazone	0.258	metoprolol	0.252
moxastine	0.030	mexiletine	0.068
naproxen	0.058	pyrrobutamine	0.218
nefopam	0.259	ibufenac	0.268

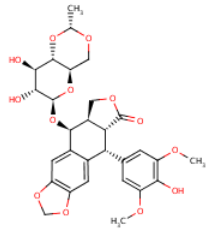
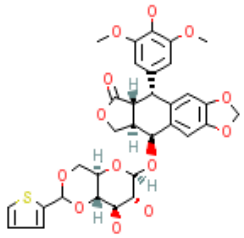
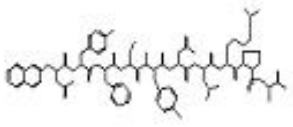
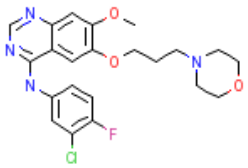
isonixin	0.254	ibuproxam	0.255
isoxicam	0.258	fosazepam	0.254
ketorolac	0.229	meparfynol	0.257
indomethacin	0.258	meprobamate	0.000
Name	**Output**	**Name**	**Output**
octopamine	0.260	ipratropium bromide	0.000
oxaceprol	0.260	morphine	0.259
oxanamide	0.258	nadoxolol	0.284
oxitropium bromide	0.260	narcobarbital	0.254
propyphenazone	0.101	orphenadrine	0.331
reprotero	0.000	oxametacine	0.000
proxibarbital	0.031	oxaprozin	0.296
phenacetin	0.013	oxprenolol	0.258
pindolol	0.258	protokylol	0.259
piperidione	0.257	salicylamide O-acetic acid	0.435
tertatolol	0.438	proxyphylline	0.258
thenyldiamine	0.008	phenylbutazone	0.127
tiaprofenic acid	0.237	phenyltoloxamine(a,c,g)	0.254
toliprolol	0.245	pipebuzone	0.001
tolmetin	0.399	thenaldine	0.003
tolpropamine	0.207	theobromine	0.276
trifluperidol	0.008	procaterol	0.298
trimethadione	0.258	prolintane	0.248
zolamine	0.251	pronethalol	0.002
thioridazine	0.003	tripelennamine	0.099
triazolam	0.148	tulobuterol	0.034
triclofos	0.034	vinylbital	0.284
trifluoperazine	0.363	xibenolol	0.343
zomepirac	0.247		

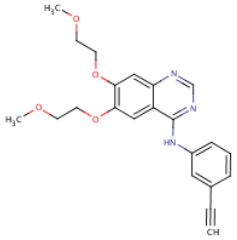
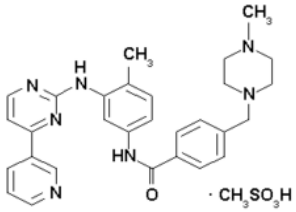
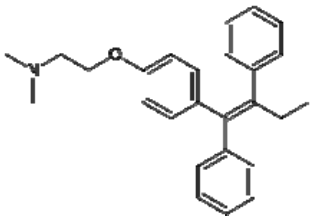
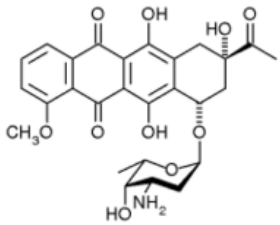
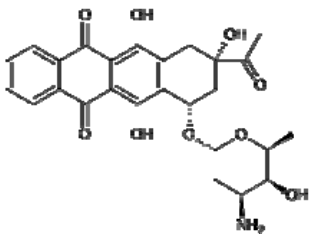
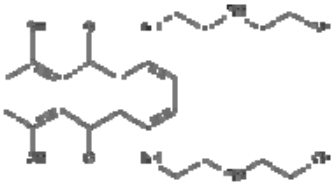
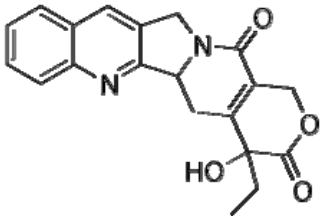
Table 4: Compounds of the training set and output values from the trained neural network with six hidden nodes.

Name	Output	Name	Output
		Anticancer	
AZQ	0.984	methotrimeprazine	0.117
BCNU	0.881	gemfibrozil	0.001
Thiocolchicine	0.924	glafenine	0.260
Trityl cysteine	0.730	glutethimide	0.000
Vinblastine sulfate	0.685	haloperidol	0.248
chlorambucil	0.984	hexobarbital	0.259
Mitozolamide	0.874	isofezolac	0.259
Nitrogen mustard	0.985	isopromethazine	0.265
PCNU	0.946	ketoprofen	0.221
Aminopterin	0.848	labetalol	0.000
methotrexate	0.912	niceritrol	0.260
Allocolchicine	0.880	nifenalol	0.262
Doxorubicin	0.984	probucol	0.133
Non-anticancer		proglumetacin	0.000
acetanilide	0.023	promazine	0.260
acetazolamide	0.263	propanolol	0.260
bucetin	0.227	thiothixene	0.292
bufetolol	0.148	thonzylamine	0.269
Butofilolol	0.252	timolol	0.358
Carbidopa	0.257	tretoquinol	0.356
Arotinolol	0.159	triprolidine	0.129
cetamolol	0.257	viminol	0.257
chlorhexadol	0.006	xenbucin	0.214
chloropyramine	0.009	salsalate	0.003
disulfiram	0.258	salverine	0.003

doxylamine(b,f,g,i)	0.000	secobarbital	0.001
droxicam	0.256	Phenopyrazone	0.258
ectylurea	0.232	pirprofen	0.418
chlorpheniramine	0.002	lefetamine	0.006
chlorthenoxacin	0.258	nicoclonate	0.266
cinchophen	0.094	nipradilol	0.259
cinnarizine	0.130	nordiazepam	0.459
moprolol	0.259	nitrazepam	0.210
enprofylline	0.248	novonal	0.282
fenoterol	0.388	salacetamide	0.684
fentiazac	0.197	salicylamide	0.254
flufenamic acid	0.251	phenoperidine	0.258
meclofenamic acid(f)	0.265	piroxicam	0.000
ronifibrate	0.259		

Table 5: Compounds of the testing set and the corresponding output values from the trained neural network with six hidden nodes.

Compound name	Structure	Compound ID (Drug bank ID)	Prediction
Etoposide		APRD00239	0.999
Teniposide		APRD00649	1.000
Abarelix		BTD00051	1.000
Gefitinib		APRD00997	1.000

Erlotinib		APRD00951	0.982
Imatinib (tyrosine kinase inhibitor)		APRD01028	0.985
Tamoxifen (estrogen receptor inhibitor)		APRD00123	0.985
Daunorubicin (inhibit DNA synthesis by intercalating with base pair)		APRD00521	0.984
Idarubicin (antitumor antibiotic)		APRD00126	0.915
Mitoxantrone (type II topoisomerase inhibitor)		APRD00371	0.985
Camptothecin (Topoisomerase I inhibitor)		CAS No. 7689-03-4	0.988

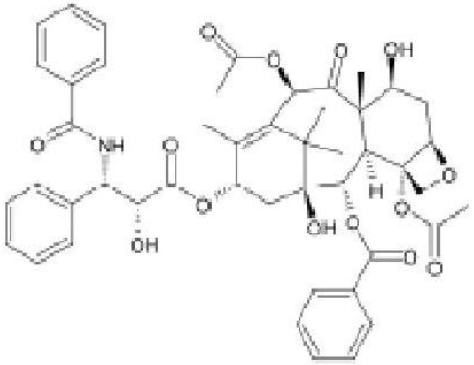
aclitaxel (Antimitotic agent)		APRD00259	0.994

Table 6: Structural formulas and prediction results from the neural network for some anticancer drugs (validation set).