Digest Journal of Nanomaterials and Biostructures Vol. 4, No.1, March 2009, p. 19 - 43

PREDICTION OF ANTICANCER / NON-ANTICANCER DRUGS BASED ON COMPARATIVE MOLECULAR MOMENT DESCRIPTORS USING ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

PRADEEP KUMAR NAIK*, AMIYA PATEL[a]
Department of Bioinformatics and Biotechnology, Jaypee University of Information Technology, Waknaghat, Distt.-Solan, Himachal Pradesh, India (Pin-173215).
[a]School of Life Sciences, Sambalpur University, Jyoti Vihar, Burla, Sambalpur-768019, Orissa, India

The structure-activity relationship (QSAR) model developed discriminate anticancer / non-anticancer drugs using machine learning techniques: artificial neural network (ANN) and support vector machine (SVM). The ANN used here is a feed-forward neural network with a standard back-propagation training algorithm. The performance was compared using 13 shape and electrostatic (Molecular Moments) descriptors. For the complete set of 13 molecular moment descriptors, ANN reveal a superior model (accuracy = 86.7%, Q_{pred} = 76.7%, sensitivity = 0.958, specificity = 0.805 Matthews correlation coefficient (MCC) = 0.74) in comparison to the SVM model (accuracy = 84.28%, Q_{pred} = 74.28%, sensitivity = 0.9285, specificity = 0.7857, MCC = 0.6998). These methods were trained and tested on a non redundant data set of 180 drugs (90 anticancer and 90 non-anticancer). The proposed model can be used for the prediction of the anti-cancer activity of novel classes of compounds enabling a virtual screening of large databases.

(Receieved January 11, 2009; accepted January 20, 2009)

Keywords: Artificial neural network, comparative molecular moment descriptors, support vector machine, drug design, structure activity relationship.

1. Introduction

A number of natural and synthetic products have been found to exhibit anticancer activity against tumor cell lines [1-5]. 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 [6-8].

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 [9,10]. Currently two classifier systems are most often used in these applications: PLS based classifiers [11,12] and various types of artificial neural networks (ANN) [13]. Typically, these systems yield an average overall accuracy of 80% correct predictions for binary decision tasks following the "likeness concept" in virtual screening [10]. Xue et al. [14]

*Corresponding author: pknaik73@rediffmail.com

have successfully used the probabilistic neural networks for the classification of 102 active compounds from diverse medicinal plants with anticancer activity using molecular descriptors. The support vector machine (SVM) approach was first introduced by Vapnik as a potential alternative to conventional artificial neural networks [15,16]. Its popularity has grown ever since in various areas of research and first applications in molecular informatics and pharmaceutical research have been described [17]. Although SVM can be applied to multiclass separation problems, its original implementation solves binary class/nonclass separation problems. Here we described application of ANN and SVM to the anticancer/non-anticancer drugs classification problem which employs a class/nonclass implementation of ANN and SVM. Both SVM and ANN algorithms can be formulated in terms of learning machines. The standard scenario for classifier development consists of two stages: training and testing. During first stage the learning machine is presented with labeled samples which are basically *n*-dimensional vectors with a class membership label attached. The learning machine generates a classifier for prediction of the class label of the input coordinates. During the second stage, the generalization ability of the model is tested.

Currently various sets of molecular descriptors are available [18-20]. For application to class/nonclass classification of compounds, the molecules are typically represented by *n*-dimensional vectors. In this work, we focused on CoMMA i.e. Comparative Molecular Moment Analysis [21] for calculating the molecular descriptors. It utilizes information from moment expansions of molecular mass and charge, up through and inclusive of second order to perform molecular comparison. It also allows deriving shape and electrostatic descriptors. CoMMA uses the lower order moments of the molecular mass and charge distributions in addition to one higher-order multipole moment of the charge density distribution, namely, the quadrupole moment as well as a description of the relationship between the two distributions by projections of the electrostatic moments upon the principal component inertial axes. This, together with the ability to perform similarity assignments between different molecules without the requirement of molecular superposition makes the CoMMA descriptors a powerful three-dimensional representation of molecular structure.

In the present paper an attempt has been made to develop a structure-activity relationship model that could help to predict novel classes of compounds having anticancer activity. We have used two machine-learning techniques: a support vector machine (SVM) and an Artificial Neural Network (ANN) as binary classifiers.

2. Materials and methods

Data Set

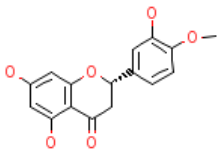
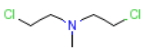
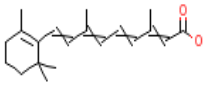
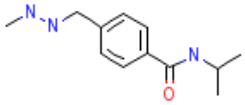
To discriminate between the anticancer and non-anticancer drugs, a data set of 180 drug molecules consisted of 90 non redundant anticancer and the same number of non redundant non-anticancer drugs were used for training, validation and testing. The 3D structure of all the drug molecules in Mol_2 format is obtained from the DRUG BANK database [22]. The complete list of the drug molecules used along with their properties is given in the supplementary material.

Prediction of molecular moment's descriptors

A set of 13 molecular moment descriptors were computed for each chemical structure following Comparative Molecular Moment Analysis (CoMMA) that uniquely encode all the 180 structures, some of which are shown in Table 1. It includes two electrostatic descriptors p and q. The descriptor 'p' calculates the multiple moment descriptors; depending upon the definition of center-of-mass of molecule. Whereas, the descriptor 'q' define the quadrupole moment which depends upon the definition of the center-of-dipole. Besides, p and q it includes 3 moments of inertia, Ix, Iy and Iz with respect to three axes X, Y and Z. Eight additional descriptors that relate the charge to the distribution of mass, dx, dy, dz, the magnitudes of projections of the dipole upon the principal inertial axes, Px, Py, Pz, and two components of the quadruple tensor written in the

frame of the principal inertial axes, q_{xx} and q_{yy} are also included [23]. The moment descriptors provide a succinct representation of the three-dimensional distribution of molecular mass, shape and charge.

Table 1. A set of 13 shape and electrostatistic descriptors predicted using CoMMA.

	1	2	3	4
STRUCTURE	 CC C16H14O6 Hesperetin	 C5H11Cl2N Mechlorethamine	 C20H28O2 Tretinoin	 C12H19N3O Procarbazine
DESCRIPTORS	Ix = 387.815 Iy = 2115.55 Iz = 2503.37 Px = 0.331978 Py = 0.428902 Pz = 0 P = 0.542371 Q = 3.70709 qxx = 2.31823 qyy = 1.38886 dx = 2.45122 dy = 3.39761 dz = 0	Ix = 37.8164 Iy = 585.882 Iz = 623.698 Px = 0.000451848 Py = 0.153977 Pz = 0 P = 0.153978 Q = 1.04965 qxx = -1.04965 qyy = -9.03893e-06 dx = 0.122569 dy = 2.38712 dz = 0	Ix = 243.311 Iy = 3596.59 Iz = 3839.9 Px = 0.162838 Py = 0.0761897 Pz = 0 P = 0.179781 Q = 1.30762 qxx = 0.234843 qyy = 1.07277 dx = 1.21492 dy = 1.08092 dz = 0	Ix = 115.937 Iy = 1550.96 Iz = 1666.9 Px = 0.106196 Py = 0.352981 Pz = 0 P = 0.368609 Q = 2.24136 qxx = 2.05533 qyy = 0.186034 dx = 1.47717 dy = 0.462454 dz = 0

Implementation of the Neural Network Predictor

Neural network predictor was implemented using the Stuttgart neural network simulator (SNNS) (http://www-ra.informatik.uni-tuebingen.de/SNNS/) in Microsoft Windows environment with cygwin. A feed-forward neural network with back-propagation algorithm was used to discriminate between anticancer and non-anticancer drugs. The neural network consisting of 13 input nodes, 4 hidden nodes and 1 output node. For each drug molecule in the training and testing sets, we have transformed 13 network input parameters into the normalized values varying from 0 to 1. Similarly, the output parameters from the ANN were in the range of [0:1]. During the learning phase, a value of 1 was assigned for the anticancer drugs and 0 for non-anticancer drugs. 100 independent training runs were performed to evaluate the average predictive power of the network. The corresponding counts of the false/true positive and negative predictions were estimated using 0.1 and 0.9 cut-off values for non-anticancer and anticancer drugs respectively. Thus, an anticancer drug from the testing set was considered correctly predicted by the ANN only when its output value ranged from 0.9 to 1.0. For each non-anticancer drug of the testing set the correct prediction was assumed if the corresponding ANN output lies between 0 and 0.1. Thus, all network output values ranging from 0.2 to 0.9 have been ultimately considered as incorrect predictions (rather than undetermined or non-defined).

Implementation of the SVM

SVM learning was implemented using SVMlight [24] available at http://svmlight.joachims.org. In this study the regression mode of SVM was used to model the discrimination between anticancer and non-anticancer drugs. Window size of 13 nodes was used as input to SVM, where each node corresponds to a molecular moment descriptor. The SVM model was also trained and tested on the data set of 90 anticancer and 90 non-anticancer drugs. Assuming that we have number of drugs $x_{i \in R}{}^{d}$ (i = 1, 2…..N) with corresponding target values $y_i \in$ ‹target value›, the x_i corresponds to the molecular moment descriptors representing a drug molecule presented to SVM for learning. Here, the target value is either +1, representing an anticancer drug or -1, representing a non-anticancer drug. The kernel chosen was Radial Basis Function (RBF) with regression mode. The QSAR model develops in this study using ANN and SVM is depicted in Figure 1.

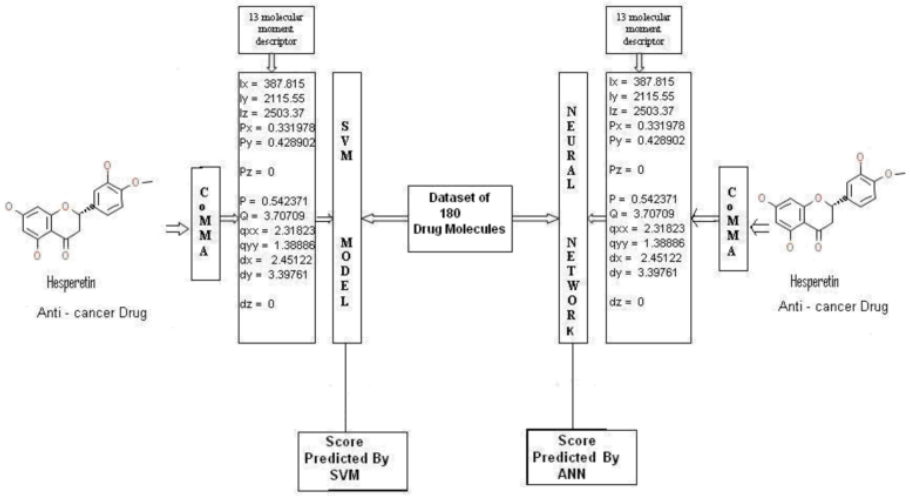


Fig. 1. Pictorial representation of the prediction methods of anticancer and non-anticancer drugs using ANN and SVM.

Fourfold cross-validation

The prediction accuracy of both the trained ANN and SVM model was tested using a fourfold cross validation technique. The models were trained and tested on a datasets of 90 anticancer and 90 non-anticancer drugs. Both the data sets were divided into a total of 3 subsets: training set, validation set and a testing set, each of which consisted of 60 drug molecules (30 anticancer and 30 non anticancer molecules). These three sets were randomly reordered and selected, i.e. the procedure of division of 180 molecules into 3 subsets was done four times. The reason for this division is to avoid over-fitting of the neural network model since it has the capability to learn even the experimental noise of the data. 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. The performance measures given have been averaged over the four testing datasets.

Performance measures

The prediction results from both SVM and ANN were evaluated for test dataset using the following statistical measures.

1. Accuracy of the methods: The accuracy of prediction for neural network and SVM model were calculated as follows:

$$Q_{ACC} = \frac{P+N}{T}, \text{ T = (P+N+O+U)}$$

Where *P* and *N* refer to correctly predicted anticancer and non-anticancer drugs, *O* and *U* refer to incorrectly predicted anticancer and non-anticancer drugs, respectively.

2. The Matthews correlation coefficient (MCC) is defined as:

$$MCC = \frac{(P \times N) - (O \times U)}{\sqrt{(P+U) \times (P+O) \times (N+U) \times (N+O)}}$$

3. Sensitivity (Q_{sens}) and specificity (Q_{spec}) of the prediction methods are defined as:

$$Q_{sens} = \frac{P}{P+U}$$

$$Q_{spec} = \frac{N}{N+O}$$

4. Q_{Pred} (Probability of correct prediction) is defined as:

$$Q_{pred} = \frac{P}{P+O} \times 100$$

3. Results and discussion

Prediction of anticancer / non-anticancer drugs from the comparative molecular moment descriptors has not been undertaken so far. However, the works on the classification of drugs and non-drugs were reported by applying the methods of support vector machines [25], probability-based classification [26], the ANN [27-29] and the Bayesian Neural Networks [13]. Here, we have explored the learning potentials of machine learning techniques, namely, ANN and SVM for the differentiation of anticancer drugs from the non-anticancer drugs. The present work is practically more important as it uses only a limited number of structural descriptors which are easy for statistical analysis due to their limited number. This QSAR model can be useful for virtual screening, combinatorial library design and data mining of anticancer drugs. The average values of the 13 CoMMA descriptors independently calculated for anticancer and non-anticancer drugs are shown in Table 2. It revealed that the two classes of compounds were clearly separated, hence, these descriptors appeared appropriate for building QSAR model of ‘anticancer-likeness’. Table 3 contains the resulting values of specificity, sensitivity, accuracy and other performance measures

of separation of anticancer and non-anticancer drugs in the testing sets using ANN and SVM. The ANN model revealed a superior model (accuracy = 86.7%, Q_{pred} = 76.7%, sensitivity = 0.958, specificity = 0.805 and MCC = 0.74) in comparison to the SVM model (accuracy = 84.28%, Q_{pred} = 74.28%, sensitivity = 0.9285, specificity = 0.7857 and MCC = 0.6998). Figure 2a & b represents the average predicted range of the output values for the four testing sets consisting of equal number of anticancer and non-anticancer drugs using ANN and SVM. As it can readily be seen from the graph, the vast majority of the predictions has been contained within (0.0 – 0.1) for non-anticancer and (0.9 – 1.0) for anticancer drugs which also illustrates that 0.1 and 0.9 cut-offs values provide very adequate separation of two bioactive classes using.

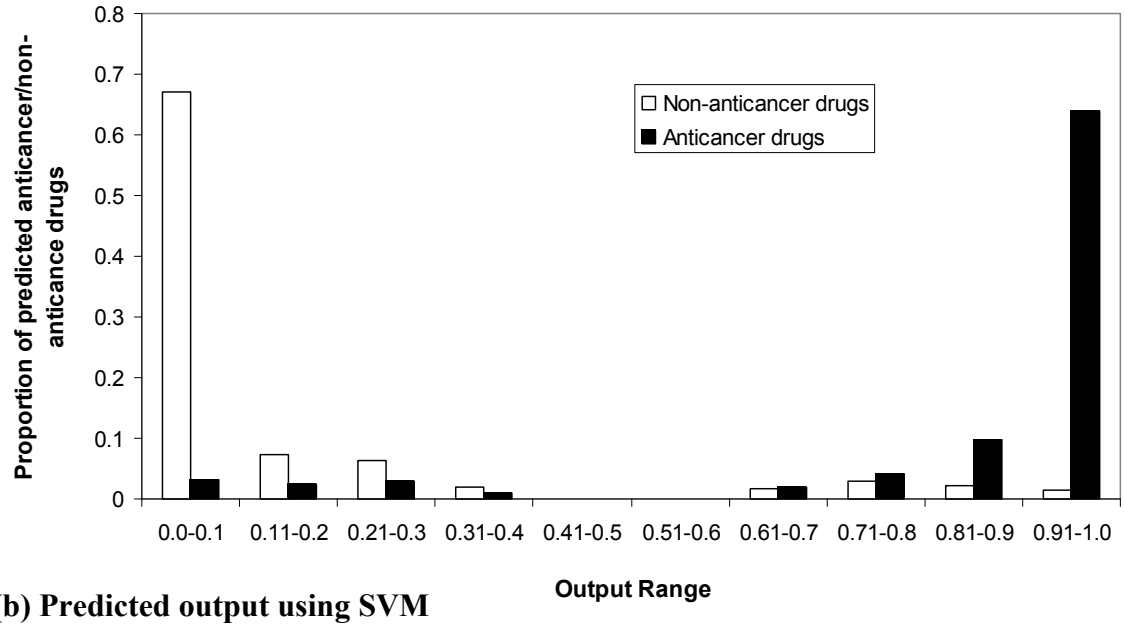
Table 2. The average values of 13 descriptors independently calculated for anticancer and non- anticancer drugs using CoMMA.

	Ix	**Iy**	**Iz**	**Px**	**Py**	**Pz**	**P**	**Q**	**qxx**	**qyy**	**dx**	**dy**	**dz**
Anti-cancer	927.5	4269.3	5123.1	0.393	0.449	0.040	0.688	2.762	0.133	0.325	291.58	221.3	0.157
Non-anti-cancer	1468.7	5873.5	6907.2	0.262	0.277	0.192	0.526	7.019	2.656	-1.045	2.843	3.27 e+14	1.36 e+14

Table 3. Performance Measures of SNNS and SVM_Light classifiers using 13 CoMMA descriptors.

Method	**Accuracy (%)**	**QPred (%)**	**Mathew correlation co-efficient (MCC)**	**Sensitivity (Sn)**	**Specificity (Sp)**
Artificial Neural Network (ANN)	86.67	76.67	0.74	0.95	0.80
Support Vector Machine (SVM)	84.28	74.28	0.67	0.93	0.79

(a) Predicted output using ANN



(b) Predicted output using SVM

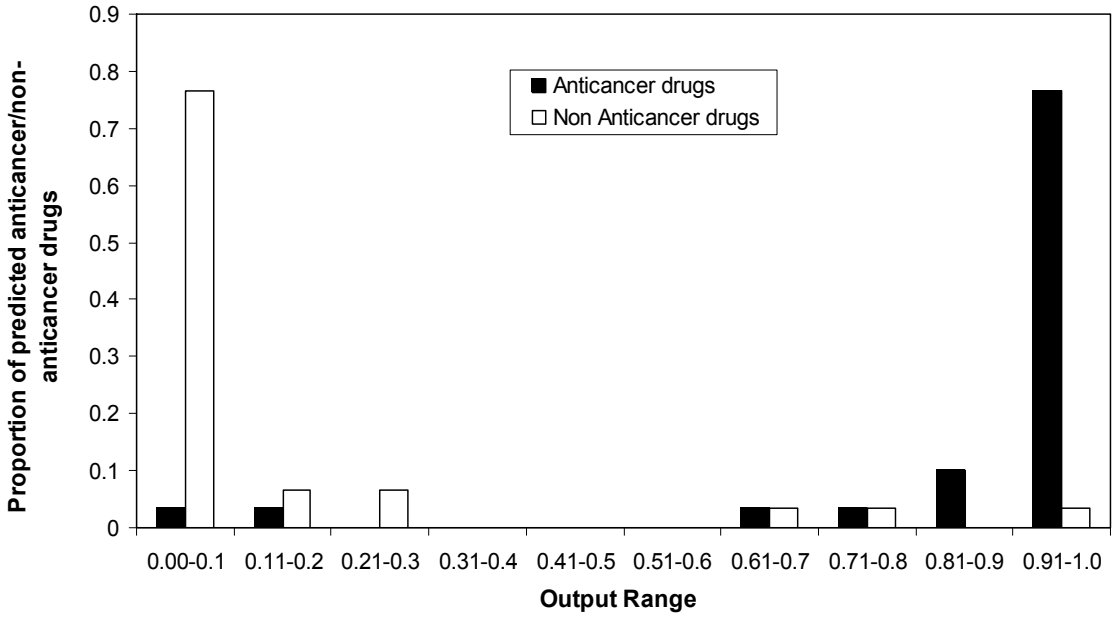


Fig. 2. Distribution of the output values of anticancer and non-anticancer drugs of test dataset from the (a) ANN and (b) SVM prediction model.

The results on discrimination of anticancer compounds by the ANN and SVM based QSAR solutions built upon the 'CoMMA' descriptors clearly demonstrate an adequacy and good predictive power of the developed model. Thus, there is strong evidence, that the 'CoMMA' descriptors do adequately reflect the structural properties of organic chemicals which are relevant for their anticancer activity (though the mechanisms of action of different classes of anticancer drugs are completely different). These observations could be explained by the fact that the parameters calculated by CoMMA cover a broad range of properties of bound atoms and molecules related to their shape, electrostatics, molecular mass, quadruple moment, moment of inertia, etc.

4. Conclusions

The results have shown that ANN and SVM can be used to predict the activity of drugs from calculated information derived from structure and available physicochemical descriptors. These results are particularly interesting from a clinical perspective. This is the first time, an instance of ANN and SVM has been applied with success from the limited set of molecular moment descriptors of a drug molecule as input. Intelligent systems such as this could markedly reduce costs of experimental approaches of prediction of anticancer drugs.

Acknowledgements

The calculation of all molecular descriptors with the program CoMMA obtained from IBM Thomas J. Watson Research Center is gratefully acknowledged.

References

[1] M.E. Bracke, R.M.L. Van Cauwenberge, M.M. Mareel, Clin. Metastasis **2**, 161 (1984).
[2] R. Giavazzi, A. Garofalo, G. Damia, S. Garattini, M. D'Incalci, British J. of Cancer **57(3)**, 277 (1988).
[3] S. Mukherjee, V. Kumar, A.K. Prasad, H.G. Raj, M.E. Bracke, C.E. Olsen, S.C. Jain, V.S. Parmar, Bioorganic Med. Chem. **9(2)**, 337 (2001).
[4] B.W. Vanhoecke, H.T. Depypere, A. De Beyter, S.K. Sharma, V.S. Parmar, D. De Keukeleire, M.E. Bracke, Pure App. Chem. **71(1)**, 65 (2005).
[5] L. Sartor, E. Pezzato, I. Dell'Aica, R. Caniato, S. Biggin, S. Garbisa, Biochem. Pharm. **64(2)**, 229 (2002).
[6] P. Buchwald, N. Bodor, N. Drugs Future **27,** 577 (2002).
[7] E.A. Lunney, Med. Chem. Research **8**, 352 (1998).
[8] W.L. Jorgensen, Science **303**, 1813 (2004).
[9] D.E. Clark, S.D. Pickett, Drug Disco. Today **5**, 49 (2000).
[10] G. Schneider, H.J. Bohm, Drug Discov. Today **7**, 64 (2002).
[11] S. Wold, Chemometrics Intel. Lab. Sys. **23, 149 (**1994).
[12] M. Forina, M.C. Casolino, C. de la Pezuela Martinez, J. Pharma. Biomed. Analysis **18**, 21 (1998).
[13] A. Ajay, W.P. Walters, M.A. Murcko, J. Med. Chem. **41(18)**, 3314 (1998).
[14] C.X. Xue, X.Y. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, J. Pharm. Biomed. Anal. **38(3)**, 497 (2005).
[15] V.N. Vapnik, Statistical Learning Theory (New York, Wiley) (1998).
[16] J. Weston, F. Perez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, B. Schoelkopf, Bioinformatics **19**(6), 764 (2003).
[17] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Comp. Chem. **26**, 5 (2001).
[18] D.E. Clark, Combi. Chem. High Throughput Screen. **4(6)**, 477 (2001).
[19] F. Atkinson, S. Cole, C. Green, H. van de Waterbeemd, Curr. Med. Chem. **2(3)**, 229 (2002).
[20] D.E. Clark, Drug Discov. Today **8**, 927 (2003).
[21] B.D. Silverman, D.E. Platt, M. Pitman, I. Rigoutsos, Perspectives in Drug Discov. and Design **399**(11), 183 (1998).
[22] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, Nucl. Acids Res. **34(1)**, D668 (2006).
[23] B.D. Silverman, Rev. in Med. Chem. **1(1)**, 1 (2001).
[24] T. Joachims, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD) ACM (2002).
[25] E. Byvatov, G. Schneider, App. Bioinf. **2(2)**, 67 (2003).

[26] Y. Takaoka, Y. Endo, S. Yamanobe, H. Kakinuma, T. Okubo, Y. Shimazaki, T. Ota, S. Sumiya, K. Yoshikawa, J. Chem. Inf. and Comp. Sc. **43(4)**, 1269 (2003).
[27] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design: An Introduction, (Germany, John Wiley) (1999).
[28] C.Y. Zhao, R.S. Zhang, H.X. Liu, C.X. Xue, S.G. Zhao, X.F. Zhou, M.C. Liu, T. Fan, J. Chem. Inf. and Sc. **44(6)**, 2040 (2004).
[29] J. Galvez, J.V. de Julian-Ortiz, R. Garcia-Demenech, J. Mol. Graphics and Model. **20(1)**, 84 (2001).

Dataset used for training, validation and testing of the model.

Anticancer dataset (90 molecules):

Serial Number	Accession I.D. of DrugBank	Name of the molecule	Molecular Formula	Description / Property / Usage (Taken from DrugBank)
1.	APRD00007	Ifosfamide	$C_7H_{15}Cl_2N_2O_2P$	For third line chemotherapy of germ cell testicular *cancer*. It should ordinarily be used in combination with a prophylactic agent for hemorrhagic cystitis, such as mesna.
2.	APRD00016	Anastrozole	$C_{17}H_{19}N_5$	For treatment of breast *cancer* in post-menopausal women.
3.	APRD00021	Amifostine	$C_5H_{15}N_2O_3PS$	For reduction in the cumulative renal toxicity in patients with ovarian *cancer* (using cisplatin) and moderate to severe xerostomia in patients undergoing post-operative radiation treatment for head and neck cancer.
4.	APRD00028	Tramadol	$C_{16}H_{25}NO_2$	Indicated in the treatment of moderate to severe pain. Tramadol is used to treat postoperative, dental, *cancer*, and acute musculosketetal pain and as an adjuvant to NSAID therapy in patients with osteoarthritis.
5.	APRD00042	Bicalutamide	$C_{18}H_{14}F_4N_2O_4S$	For treatment (together with surgery or LHRH analogue) of advanced *prostatic cancer.*
6.	APRD00064	Amsacrine	$C_{21}H_{19}N_3O_3S$	For treatment of acute myeloid *leukaemia*.
7.	APRD00078	Porfimer	$C_{68}H_{74}N_8O_{11}$	Indicated in the treatment of *esophageal cancer.*
8.	APRD00090	Dexrazoxane	$C_{11}H_{16}N_4O_4$	For reducing the incidence and severity of cardiomyopathy associated with doxorubicin administration in women with metastatic *breast cancer.*
9.	APRD00100	Aprepitant	$C_{23}H_{21}F_7N_4O_3$	For the prevention of nausea and vomiting associated with highly emetogenic *cancer* chemotherapy, including high-dose cisplatin (in combination with other antiemetic agents). Animal and human Positron Emission Tomography (PET) studies with Aprepitant have shown that it crosses the blood brain barrier and occupies brain NK1 receptors. Animal and human studies show that Aprepitant augments the

				antiemetic activity of the 5-HT_3-receptor antagonist ondansetron and the corticosteroid ethasone and inhibits both the acute and delayed phases of cisplatin induced emesis.
10.	APRD00101	Vinorelbine	$C_{45}H_{54}N_4O_8$	For the treatment of non-*small-cell lung carcinoma*
11.	APRD00115	Chlorambucil	$C_{14}H_{19}Cl_2NO_2$	For treatment of chronic lymphatic (lymphocytic) *leukemia*, malignant lymphomas including lymphosarcoma, giant follicular lymphoma, and Hodgkin's disease. Chlorambucil is an antineoplastic in the class of alkylating agents and is used to treat various forms of *cancer*.
12.	APRD00117	Hesperetin	$C_{16}H_{14}O_6$	For lowering cholesterol and, possibly, otherwise favorably affecting lipids. *In vitro* research also suggests the possibility that hesperetin might have some *anticancer* effects and that it might have some anti-aromatase activity, as well as activity again.
13.	APRD00118	Melphalan	$C_{13}H_{18}Cl_2N_2O_2$	For the palliative treatment of multiple *myeloma* and for the palliation of non-resectable epithelial *carcinoma* of the ovary.
14.	APRD00123	Tamoxifen	$C_{26}H_{29}NO$	For the treatment of *breast cancer*.
15.	APRD00124	Dactinomycin	$C_{62}H_{86}N_{12}O_{16}$	For the treatment of Wilms' *tumor*, childhood rhabdomyosarcoma, Ewing's sarcoma and metastatic, nonseminomatous *testicular cancer* as part of a combination chemotherapy and/or multi-modality treatment regimen
16.	APRD00125	Venlafaxine	$C_{17}H_{27}NO_2$	For the treatment of severe depression. It is used to treat melancholia, generalized anxiety disorder (GAD), panic disorder, post-traumatic stress disorder, and hot flashes in *breast cancer* survivors.
17.	APRD00144	Exemestane	$C_{20}H_{24}O_2$	For the treatment of advanced *breast cancer* in postmenopausal women whose disease has progressed following tamoxifen therapy.
18.	APRD00150	Nilutamide	$C_{12}H_{10}F_3N_3O_4$	For use in combination with surgical castration for the treatment of *metastatic prostate cancer* (Stage D2)
19.	APRD00174	Clonidine	$C_9H_9Cl_2N_3$	Clonidine is an antihypertensive

				agent and an epidural agent for *refractory cancer pain*.
20.	APRD00201	Gemcitabine	$C_9H_{11}F_2N_3O_4$	For the first-line treatment of patients with *metastatic breast cancer*, locally advanced (Stage IIIA or IIIB), or metastatic (Stage IV) *non-small cell lung cancer* and as first-line treatment for patients with *adenocarcinoma* of the pancreas.
21.	APRD00202	Pentostatin	$C_{11}H_{16}N_4O_4$	For the treatment of hairy cell *leukaemia* refractory to alpha interferon.
22.	APRD00203	Capecitabine	$C_{15}H_{22}FN_3O_6$	For the treatment of patients with *metastatic breast cancer* resistant to both paclitaxel and an anthracycline-containing chemotherapy regimen.
23.	APRD00209	Streptozocin	$C_8H_{15}N_3O_7$	For the treatment of malignant neoplasms of pancreas *(metastatic islet cell carcinoma)*.
24.	APRD00239	Etoposide	$C_{29}H_{32}O_{13}$	For use in combination with other chemotherapeutic agents in the treatment of refractory *testicular tumors* and as first line treatment in patients with *small cell lung cancer*. Also used to treat other malignancies such as lymphoma, non-lymphocytic leukemia, and glioblastoma multiforme.
25.	APRD00246	Calcitriol	$C_{27}H_{44}O_3$	Calcitriol has been found to induce differentiation and/or inhibit cell proliferation in a number of malignant cell lines including *human prostate cancer cells*.
26.	APRD00249	Mechlorethamine	$C_5H_{11}Cl_2N$	For the palliative treatment of Hodgkin's disease (Stages III and IV), lymphosarcoma, chronic myelocytic or chronic lymphocytic *leukemia*, polycythemia vera, mycosis fungoides, and bronchogenic *carcinoma*. Also for the palliative treatment of metastatic *carcinoma* resulting in effusion.
27.	APRD00255	Orlistat	$C_{29}H_{53}NO_5$	In the March 15, 2004 issue of Cancer Research, [1] Steven J. Kridel *et al.* state that orlistat may also inhibit growth of *prostate cancer*, and in theory may be useful in treating other *cancers*, by interfering with the metabolism of fats.
28.	APRD00259	Paclitaxel	$C_{47}H_{51}NO_{14}$	Used in the treatment of Kaposi's sarcoma and *cancer of*

				the lung, ovarian, and breast.
29.	APRD00260	Cladribine	$C_{10}H_{12}ClN_5O_3$	For the treatment of active hairy cell *leukemia* (leukemic reticuloendotheliosis) as defined by clinically significant anemia, neutropenia, thrombocytopenia, or disease-related symptoms.
30.	APRD00268	Trimetrexate	$C_{19}H_{23}N_5O_3$	For use, with concurrent leucovorin administration (leucovorin protection), as an alternative therapy for the treatment of moderate-to-severe *Pneumocystis carinii* pneumonia (PCP) in immunocompromised patients, including patients with the acquired immunodeficiency syndrome (AIDS). Also used to treat several types of *cancer* including *colon cancer.*
31.	APRD00292	Lomustine	$C_9H_{16}ClN_3O_2$	Indicated primarily for the treatment of brain *tumours*. Also for the treatment of breast cancer, Hodgkin's disease, *lung cancer*, malignant melanoma, multiple myeloma, Non-Hodgkin's lymphomas, *ovarian cancer*, *pancreatic cancer*, and *renal cell cancer*.
32.	APRD00347	Fentanyl	$C_{22}H_{28}N_2O$	For the treatment of *cancer* patients with severe pain that breaks through their regular narcotic therapy.
33.	APRD00351	Palonosetron	$C_{19}H_{24}N_2O$	For the treatment of nausea and vomiting associated with *cancer* chemotherapy.
34.	APRD00359	Cisplatin	$Cl_2H_4N_2Pt$	For the treatment of metastatic *testicular tumors*, metastatic *ovarian tumors* and advanced *bladder cancer*.
35.	APRD00361	Epirubicin	$C_{27}H_{29}NO_{11}$	For use as a component of adjuvant therapy in patients with evidence of axillary node tumor involvement following resection of primary *breast cancer*
36.	APRD00362	Tretinoin	$C_{20}H_{28}O_2$	For the induction of remission in patients with acute promyelocytic leukemia (APL), French-American-British (FAB) classification M3 (including the M3 variant); For the topical treatment of acne vulgaris, flat warts and other skin conditions (psoriasis, ichthyosis congenita, icthyosis vulgaris, lamellar icthyosis, keratosis palmaris et

				plantaris, epidermolytic hyperkeratosis, senile comedones, senile keratosis, keratosis follicularis (Darier's disease), and *basal cell carcinomas*.); For palliative therapy to improve fine wrinkling, mottled hyperpigmentation, roughness associated with photodamage.
37.	APRD00364	Paroxetine	$C_{19}H_{20}FNO_3$	It is used to treat depression resistant to other antidepressants, depression complicated by anxiety, panic disorder, social and general anxiety disorder, obsessive-compulsive disorder (OCD), premenstrual dysphoric disorder, premature ejaculation, and hot flashes of menopause in women with *breast cancer*.
38.	APRD00391	Toremifene	$C_{26}H_{28}ClNO$	For the treatment of *metastatic breast cancer* in postmenopausal women with estrogen receptor-positive or receptor-unknown tumors
39.	APRD00392	Vindesine	$C_{43}H_{55}N_5O_7$	For the treatment of acute *leukaemia*, malignant lymphoma, Hodgkin's disease, acute erythraemia and acute panmyelosis
40.	APRD00396	Conjugated Estrogens	$C_{18}H_{21}NaO_5S$	For the treatment of moderate to severe vasomotor symptoms associated with the menopause, atrophic vaginitis, osteoporosis, hypoestrogenism due to hypogonadism, castration, primary ovarian failure, *breast cancer* (for palliation only), and Advanced *androgen-dependent carcinoma* of the prostate (for palliation only)
41.	APRD00400	Raloxifene	$C_{28}H_{27}NO_4S$	Raloxifene is used in the prevention of postmenopausal osteoporosis and *breast cancer*.
42.	APRD00408	Cyclophosphamide	$C_7H_{15}Cl_2N_2O_2P$	For management of malignant lymphomas, multiple myeloma,leukemias, mycosis fungoides (advanced disease), neuroblastoma (disseminated disease), adenocarcinoma of the ovary, retinoblastoma and *carcinoma* of the breast.
43.	APRD00430	Raltitrexed	$C_{21}H_{22}N_4O_6S$	For the treatment of *malignant neoplasm* of colon and rectum
44.	APRD00433	Testosterone	$C_{19}H_{28}O_2$	Testosterone is an antineoplastic hormonal agent primarily used in the treatment of *prostate cancer*.
45.	APRD00466	Carboplatin	$C_6H_{12}N_2O_4Pt_2$	For the initial treatment of

				advanced *ovarian carcinoma* in established combination with other approved chemotherapeutic agents. One established combination regimen consists of PARAPLATIN and cyclophosphamide.
46.	APRD00481	Ondansetron	$C_{18}H_{19}N_3O$	Ondansetron is an antinauseant and antiemetic agent indicated for the prevention of nausea and vomiting associated with moderately-*emetogenic cancer chemotherapy* and for the prevention of postoperative nausea and vomiting.
47.	APRD00495	Vincristine	$C_{46}H_{56}N_4O_{10}$	For treatment of acute *leukaemia*, malignant lymphoma, Hodgkin's disease, acute erythraemia, acute panmyelosis
48.	APRD00516	Fluorouracil	$C_4H_3FN_2O_2$	For the topical treatment of multiple actinic or solar keratoses. In the 5% strength it is also useful in the treatment of superficial *basal cell carcinomas* when conventional methods are impractical, such as with multiple lesions or difficult treatment sites. Fluorouracil injection is indicated in the palliative management of some types of cancer, including colon, rectum, breast, stomach and pancrease.
49.	APRD00518	Dolasetron	$C_{19}H_{20}N_2O_3$	For the prevention of nausea and vomiting associated with moderately-*emetogenic cancer chemotherapy*, including initial and repeat courses and prevention of postoperative nausea and vomiting
50.	APRD00559	Marimastat	$C_{15}H_{29}N_3O_5$	For the treatment of various *cancers*
51.	APRD00571	Marinol	$C_{21}H_{30}O_2$	For the treatment of anorexia associated with weight loss in patients with AIDS, and nausea and vomiting associated with *cancer chemotherapy* in patients who have failed to respond adequately to conventional antiemetic treatments
52.	APRD00573	Pemetrexed	$C_{20}H_{21}N_5O_6$	For the treatment of malignant pleural mesothelioma and locally advanced or metastatic non-*small cell lung cancer (NSCLC)* after prior chemotherapy
53.	APRD00579	Irinotecan	$C_{33}H_{38}N_4O_6$	For the treatment of metastatic

				colorectal cancer (first-line therapy when administered with 5-fluorouracil and leucovorin).
54.	APRD00594	Fludarabine	$C_{10}H_{13}FN_5O_7P$	For the treatment of adult patients with B-cell chronic lymphocytic *leukemia* (CLL) who have not responded to or whose disease has progressed during treatment with at least one standard alkylating-agent containing regimen
55.	APRD00627	Medroxyprogesterone	$C_{22}H_{32}O_3$	Used as a contraceptive and to treat amenorrhea, abnormal uterine bleeding, endometriosis, endometrial and *renal cell carcinomas*, and pulmonary disorders such as chronic obstructive pulmonary disease (COPD), Pickwickian syndrome, and other hypercapnic pulmonary conditions.
56.	APRD00640	Testolactone	$C_{19}H_{24}O_3$	For palliative treatment of advanced *breast cancer* in postmenopausal women.
57.	APRD00652	Altretamine	$C_9H_{18}N_6$	For use as a single agent in the palliative treatment of patients with persistent or recurrent *ovarian cancer* following first-line therapy with a cisplatin and/or alkylating agent-based combination.
58.	APRD00654	Fulvestrant	$C_{32}H_{47}F_5O_3S$	For the treatment of hormone receptor positive *metastatic breast cancer* in postmenopausal women with disease progression following antiestrogen therapy.
59.	APRD00662	Valrubicin	$C_{34}H_{36}F_3NO_{13}$	*Bladder cancer*
60.	APRD00664	Busulfan	$C_6H_{14}O_6S_2$	For use in combination with cyclophosphamide as a conditioning regimen prior to allogeneic hematopoietic progenitor cell transplantation for *chronic myelogenous leukemia*.
61.	APRD00687	Topotecan	$C_{23}H_{23}N_3O_5$	For the treatment of *metastatic carcinoma* of the *ovary* and *small cell lung cancer* following the failure of first-line chemotherapy.
62.	APRD00691	Ethinyl Estradiol	$C_{20}H_{24}O_2$	For treatment of moderate to severe vasomotor symptoms associated with the menopause, female hypogonadism, prostatic carcinoma-palliative therapy of advanced disease, *breast cancer*, as an oral contraceptive, and as emergency contraceptive.

63.	APRD00695	Procarbazine	$C_{12}H_{19}N_3O$	For use with other *anticancer drugs* for the treatment of stage III and stage IV Hodgkin's disease.
64.	APRD00698	Leucovorin	$C_{20}H_{23}N_7O_7$	For the treatment of osteosarcoma (after high dose methotrexate therapy). Used to diminish the toxicity and counteract the effects of impaired methotrexate elimination and of inadvertent overdosages of folic acid antagonists, and to treat megaloblastic anemias due to folic acid deficiency. Also used in combination with 5-fluorouracil to prolong survival in the palliative treatment of patients with advanced *colorectal cancer*.
65.	APRD00703	Cerulenin	$C_{12}H_{17}NO_3$	Inhibition of FAS by cerulenin leads to cytotoxicity and apoptosis in human *cancer cell lines*, an effect believed to be mediated by the accumulation of malonyl-coenzyme A in cells with an upregulated FAS pathway.
66.	APRD00708	Vinblastine	$C_{46}H_{58}N_4O_9$	For treatment of *breast cancer*, *testicular cancer*, lymphomas, neuroblastoma, Hodgkin's and non-Hodgkin's lymphomas, mycosis fungoides, histiocytosis, and Kaposi's sarcoma.
67.	APRD00715	Chlorotrianisene	$C_{23}H_{21}ClO_3$	Used to treat symptoms of menopause, deficiencies in ovary function (including underdevelopment of female sexual characteristics and some types of infertility), and in rare cases, *prostate cancer*.
68.	APRD00809	Azacitidine	$C_8H_{12}N_4O_5$	The cytotoxic effects of azacitidine cause the death of rapidly dividing cells, including *cancer cells* that are no longer responsive to normal growth control mechanisms. Non-proliferating cells are relatively insensitive to azacitidine.
69.	APRD00828	Bortezomib	$C_{19}H_{25}BN_4O_4$	For treatment of multiple *myeloma* in patients who have not been successfully treated with at least two previous therapies.
70.	APRD00878	Clofarabine	$C_{10}H_{11}ClFN_5O_3$	For the treatment of pediatric patients 1 to 21 years old with relapsed or refractory acute lymphoblastic *leukemia* after at least two prior regimens.

71.	APRD00920	Diethylstilbestrol	$C_{18}H_{20}O_2$	Used in the treatment of *prostate cancer*. Previously used in the prevention of miscarriage or premature delivery in pregnant women prone to miscarriage or premature delivery.
72.	APRD00932	Docetaxel	$C_{43}H_{53}NO_{14}$	For the treatment of patients with locally advanced or *metastatic breast cancer* after failure of prior chemotherapy.
73.	APRD00938	Dromostanolone	$C_{23}H_{36}O_3$	For use in females, for palliation of androgenresponsive recurrent *mammary cancer* in women who are more than one year but less than five years postmenopausal.
74.	APRD00951	Erlotinib	$C_{22}H_{23}N_3O_4$	For the treatment of patients with locally advanced or *metastatic non-small cell lung cancer* after failure of at least one prior chemotherapy regimen.
75.	APRD00981	Fluoxymesterone	$C_{20}H_{29}FO_3$	In males, used as replacement therapy in conditions associated with symptoms of deficiency or absence of endogenous testosterone. In females, for palliation of androgenresponsive recurrent *mammary cancer* in women who are more than one year but less than five years postmenopausal.
76.	APRD00984	Flutamide	$C_{11}H_{11}F_3N_2O_3$	For the management of locally confined Stage B2-C and Stage D2 *metastatic carcinoma of the prostate*
77.	APRD00997	Gefitinib	$C_{22}H_{24}ClFN_4O_3$	For the continued treatment of patients with locally advanced or metastatic non-small cell lung cancer after failure of either platinum-based or docetaxel chemotherapies.
78.	APRD01002	Granisetron	$C_{18}H_{24}N_4O$	For the prevention of: nausea and vomiting associated with initial and repeat courses of *emetogenic cancer* therapy, including high-dose cisplatin.
79.	APRD01021	Hydromorphone	$C_{17}H_{19}NO_3$	For the relief of moderate to severe pain such as that due to surgery, *cancer*, trauma/injury, burns, myocardial infarction and colic.
80.	APRD01030	Imiquimod	$C_{14}H_{16}N_4$	Imiquimod can be used to treat certain types of skin cancer called *superficial basal cell carcinoma*.
81.	APRD01066	Letrozole	$C_{17}H_{11}N_5$	For the extended adjuvant

				treatment of early *breast cancer* in postmenopausal women who have received 5 years of adjuvant tamoxifen therapy. Also for first-line treatment of postmenopausal women with hormone receptor positive or hormone receptor unknown locally *advanced or metastatic breast cancer*. Also indicated for the treatment of *advanced breast cancer* in postmenopausal women with disease progression following antiestrogen therapy.
82.	APRD01067	Levamisole	$C_{11}H_{12}N_2S$	For adjuvant treatment in combination with fluorouracil after surgical resection in patients with Dukes' stage C *colon cancer*. Also used to treat malignant melanoma and *head/neck cancer*.
83.	APRD01084	Masoprocol	$C_{18}H_{22}O_4$	Used for the treatment of actinic keratoses (precancerous skin growths that can become *malignant* if left untreated).
84.	APRD01092	Megestrol	$C_{22}H_{30}O_3$	For the treatment of anorexia, cachexia, or an unexplained, significant weight loss in patients with a diagnosis of acquired immunodeficiency syndrome (AIDS). Also used to treat *breast cancer, endometrial cancer, and prostate cancer* in Canada and some other countries.
85.	APRD01127	Nabilone	$C_{24}H_{36}O_3$	Used for the control of nausea and vomiting, caused by chemotherapeutic agents used in the treatment of *cancer*, in patients who have failed to respond adequately to conventional antiemetic treatments.
86.	APRD01161	Pamidronate	$C_3H_{11}NO_7P_2$	For the treatment of moderate or severe hypercalcemia associated with *malignancy.*
87.	APRD01246	Tazarotene	$C_{21}H_{21}NO_2S$	Tazarotene is associated with a significant reduction in atypical melanocytes and keratocytes - cells considered to be precursors of *skin cancer*.
88.	APRD01294	Zoledronate	$C_5H_{10}N_2O_7P_2$	For the treatment of *hypercalcemia* of *malignancy*. Also for the treatment of patients with *multiple myeloma* and patients with documented bone metastases from solid *tumors*, in conjunction with standard antineoplastic therapy.

89.	APRD01304	Sorafenib	$C_{21}H_{16}ClF_3N_4O_3$	*Anticancer Agent*. For the treatment of patients with advanced *renal cell carcinoma*.
90.	EXPT00120	1,6-Fructose Diphosphate (Linear Form)	$C_6H_{14}O_{12}P_2$	The drug target of this drug is *Lung cancer antigen* NY-LU-1.

Non Anticancer dataset (90 molecules):

Serial Number	**Accession I.D. of DrugBank**	**Name of the molecule**	**Molecular Formula**	**Description / Property / Usage (Taken from DrugBank)**
1.	EXPT00370	Adenosine-3'-5'-Diphosphate	$C_{10}H_{15}N_5O_{10}P_2$	Mediates the metabolic *activation* of *carcinogenic* N-hydroxyarylamines to DNA binding products and could so participate as modulating factor of cancer risk
2.	APRD00880	Clomifene	$C_{26}H_{28}ClNO$	Clomifene can lead to multiple ovulation, and hence increasing the risk of twins. In comparison to purified FSH, the rate of ovarian hyperstimulation syndrome is low. There may be an *increased risk of ovarian cancer*, and weight gain.
3.	APRD00263	Ganciclovir	$C_9H_{13}N_5O_4$	*Suspected cancer agent.*
4.	APRD00438	Sulfanilamide	$C_6H_8N_2O_2S$	Side effects include itching, burning, skin rash, redness, swelling, or other sign of irritation not present before use of this medicine and long-term use of sulfonamides may cause *cancer of the thyroid gland.*
5.	EXPT01291	1, 4-Dithiothreitol	$C_4H_{10}O_2S_2$	Its target induces apoptosis in *cancer* cells.
6.	APRD00924	Dimenhydrinate	$C_{24}H_{28}ClN_5O_3$	Dimenhydrinate is an antiemetics drug combination that contains diphenhydramine and theophylline. It is *not effective* in the

				treatment of nausea associated with *cancer chemotherapy*.
7.	APRD00738	Ethanol	C_2H_6O	For therapeutic neurolysis of nerves or ganglia for the relief of intractable chronic pain in such conditions as inoperable cancer and trigeminal neuralgia (tic douloureux), in patients for whom neurosurgical procedures are contraindicated.
8.	EXPT01171	Diisopropylphosphono Group	$C_6H_{14}O_3P_1$	Complement Factor B inhibitor
9.	EXPT01172	5-Deoxyflavanone	$C_{15}H_{12}O_4$	Chalcone--Flavonone Isomerase 1 inhibitor
10.	EXPT01174	2-Deoxy-Glucitol-6-Phosphate	$C_6H_{15}O_8P_1$	Myo-Inositol-1-Phosphate Synthase inhibitor
11.	EXPT01175	D-Glucuronic Acid	C_6H_8O	Chondroitinase Ac & B inhibitor
12.	EXPT01176	Digalactosyl Diacyl Glycerol (Dgdg)	$C_{51}H_{96}O_{15}$	Chlorophyll A-B Binding Protein, Chloroplast inhibitor
13.	EXPT01177	1-[Glycerolylphosphonyl]-2-[8-(2-Hexyl-Cyclopropyl)-Octanal-1-Yl]- 3-[Hexadecanal-1-Yl]-Glycerol	$C_{39}H_{75}O_{10}P_1$	Flavohemoprotein inhibitor
14.	EXPT01178	(2r)-Amino(4-Hydroxyphenyl)Acetic Acid	$C_8H_9N_1O_3$	Feglymycin inhibitor
15.	EXPT01179	2'-Deoxyguanosine-5'-Diphosphate	$C_{10}H_{15}N_5O_{10}P_2$	Nucleoside Diphosphate Kinase II inhibitor
16.	EXPT01180	D-Glutamic Acid	$C_5H_9N_1O_4$	Thermolysin inhibitor
17.	EXPT01181	D-Glutamine	$C_5H_{10}N_2O_3$	Glutamate Racemase inhibitor
18.	EXPT01182	2'-Deoxyguanosine-5'-Monophosphate	$C_{10}H_{14}N_5O_7P_1$	Rc-Rnase6 Ribonuclease inhibitor
19.	EXPT01183	3,6-Anhydro-D-Galactose-2-Sulfate	$C_6H_{10}O_8S_1$	Iota-Carrageenase inhibitor
20.	EXPT01184	2'-Deoxyguanosine-5'-Triphosphate	$C_{10}H_{16}N_5O_{13}P_3$	Nucleoside Diphosphate Kinase II inhibitor
21.	EXPT01185	(2s,3s)-Trans-Dihydroquercetin	$C_{15}H_{12}O_7$	Leucoanthocyanidin Dioxygenase inhibitor

22.	EXPT01186	2,3-Didehydroalanine	$C_3H_5N_1O_2$	Lantibiotic Mersacidin inhibitor
23.	EXPT01187	3,4-Dihydroxybenzoic Acid	$C_7H_6O_4$	Lipoxygenase-3 inhibitor
24.	EXPT01188	3,4-Dihydroxycinnamic Acid	$C_9H_8O_4$	Photoactive Yellow Protein inhibitor
25.	EXPT01189	Heme D	$C_{34}H_{32}N_4O_{10}Fe_1$	Nitrite Reductase inhibitor
26.	EXPT01190	Dihydrofolic Acid	$C_{19}H_{21}N_7O_6$	Dihydrofolate Reductase inhibitor
27.	EXPT01191	3-Dehydroshikimate	$C_7H_{10}O_5$	3-Dehydroquinate Dehydratase inhibitor
28.	EXPT01192	2,6-Dimethyl-7-Octen-2-Ol	$C_{10}H_{20}O_1$	Odorant-Binding Protein inhibitor
29.	EXPT01193	5-Hydroxy Norvaline	$C_5H_{11}N_1O_3$	Sst1-Selective Somatosatin Analog inhibitor
30.	EXPT01194	Deoxycholic Acid	$C_{24}H_{40}O_4$	Major Pollen Allergen Bet V 1-L inhibitor
31.	EXPT01195	3-Decyl-2,5-Dioxo-4-Hydroxy-3-Pyrroline	$C_{14}H_{23}N_1O_3$	Glycolate Oxidase inhibitor
32.	EXPT01196	3,4-Dihydro-5-Methyl-Isoquinolinone	$C_{10}H_{11}N_1O_1$	Poly inhibitor
33.	EXPT01197	(2s)-Hydroxy(4-Hydroxyphenyl)Ethanenitrile	$C_8H_7N_1O_2$	Beta-Glucosidase inhibitor
34.	EXPT01198	3-Amino-4,5-Dihydroxy-Cyclohex-1-Enecarboxylate	$C_7H_{10}N_1O_{41}$	3-Dehydroquinate Dehydratase Arod inhibitor
35.	EXPT01199	Dihydrotestosterone	$C_{19}H_{30}O_2$	Sex Hormone-Binding Globulin inhibitor
36.	EXPT01200	2-(3,4-Dihydroxyphenyl)Acetic Acid	$C_8H_8O_4$	Homoprotocatechuate 2,3-Dioxygenase inhibitor
37.	EXPT01201	Octamethylenediamine	$C_8H_{20}N_2$	Polyamine Oxidase inhibitor
38.	EXPT01202	3,4-Dichloroisocoumarin	$C_9H_4O_2Cl_2$	Factor D inhibitor
39.	EXPT01203	4,4'[1,6-Hexanediylbis(Oxy)] Bisbenzenecarboximidamide	$C_{20}H_{26}N_4O_2$	Transcriptional Regulator Qacr inhibitor
40.	EXPT01205	2,5-Dideoxy-2,5-Imino-D-Glucitol	$C_6H_{13}N_1O_4$	D-Xylose Isomerase inhibitor
41.	EXPT01206	4'-Deaza-1'-Aza-2'-Deoxy-1'-(9-Methylene)-Immucillin-H, (3r,4r)-N-[9-Deazahypoxanthin-9-Yl)Methyl]-4-Hydroxymethyl-	$C_{12}H_{19}N_4O_{31}+$	Purine Nucleoside Phosphorylase inhibitor

		Pyrrolidin-3-Ol		
42.	EXPT01207	Methylphosphonic Acid Diisopropyl Ester	$C_7H_{17}O_3P_1$	Parathion Hydrolase inhibitor
43.	EXPT01208	1,4-Diethylene Dioxide	$C_4H_8O_2$	Epsin inhibitor
44.	EXPT01209	Dinor-N(Omega)-Hydroxy-L-Arginine	$C_4H_{10}N_4O_3$	Arginase 1 inhibitor
45.	EXPT01210	Disordered Solvent	H_2O_1	Alpha-1-Purothionin inhibitor
46.	EXPT01211	D-Isovaline	$C_5H_{11}N_1O_2$	Antiamoebin I inhibitor
47.	EXPT01212	Dcka, 5,7-Dichlorokynurenic Acid	$C_{10}H_5N_1O_3Cl_2$	N-Methyl-D-Aspartate Receptor Subunit 1 inhibitor
48.	EXPT01213	Decanoic Acid	$C_{10}H_{20}O_2$	Daptomycin inhibitor
49.	EXPT01214	4-[2-(3-Benzyloxycarbonylamino-4-Cyclohexyl-1-Hydroxy-2-Oxo-Butylamino)-5-Guanidino-Pentanoylamino]-4-(1-Carboxy-2-Cyclohexyl-Ethylcarbamoyl)-Butyric Acid	$C_{38}H_{57}N_7O_{10}$	Tricorn Protease inhibitor
50.	EXPT01215	D-Lactic Acid	$C_3H_6O_3$	2-Haloacid Dehalogenase inhibitor
51.	EXPT01216	D-Leucine	$C_6H_{13}N_1O_2$	Gramicidin A inhibitor
52.	EXPT01217	2-Hexyloxy-6-Hydroxymethyl-Tetrahydro-Pyran-3,5-Diol	$C_{12}H_{24}O_5$	Histo-Blood Group Abo System Transferase inhibitor
53.	EXPT01218	Di-Linoleoyl-3-Sn-Phosphatidylcholine	$C_{44}H_{80}N_1O_8P_1$	Phosphatidylcholine Transfer Protein inhibitor
54.	EXPT01219	D-Lysine	$C_6H_{14}N_2O_2$	Catabolic Alanine Racemase Dadx inhibitor
55.	EXPT01221	Dimethylallyl Diphosphate	$C_5H_{12}O_7P_2$	Farnesyl Diphosphate Synthase inhibitor
56.	EXPT01222	5,6-Dimethylbenzimidazole	$C_9H_{10}N_2$	Nicotinate Mononucleotide:5,6-Dimethylbenzi inhibitor
57.	EXPT01224	Dimethylformamide	$C_3H_7N_1O_1$	Elastase inhibitor
58.	EXPT01225	Dimethylglycine	$C_4H_9N_1O_2$	Sarcosine Oxidase inhibitor
59.	EXPT01226	2,3-Dimethylimidazolium Ion	$C_5H_9N_{21}+$	Cytochrome C Peroxidase inhibitor
60.	EXPT01227	1-Deoxymannojirimycin	$C_6H_{13}N_1O_4$	Mannosyl-Oligosaccharide Alpha-1,2-Mannosida

				inhibitor
61.	EXPT01228	Terminal Dimethyl	C_2H_6	Lysozyme inhibitor
62.	EXPT01229	Alpha-Difluoromethylornithine	$C_6H_{12}N_2O_2F_2$	Ornithine Decarboxylase inhibitor
63.	EXPT01230	Dmp450(Inhibitor of Dupont Merck)	$C_{33}H_{38}N_4O_3$	Hiv-1 Protease inhibitor
64.	EXPT01231	Dimethyl Sulfoxide	$C_2H_6O_1S_1$	Four-Helix Bundle Model inhibitor
65.	EXPT01232	2,3-Dihydroxy-Valerianic Acid	$C_6H_{12}O_4$	Acetohydroxy-Acid Isomeroreductase inhibitor
66.	EXPT01233	3,5-Dinitrocatechol	$C_6H_4N_2O_6$	Catechol O-Methyltransferase inhibitor
67.	EXPT01234	Deamido-Nad+	$C_{21}H_{27}N_6O_{15}P_{21}+$	Nh inhibitor
68.	EXPT01235	2,4-Dinitrophenol	$C_6H_4N_2O_5$	Pentaerythritol Tetranitrate Reductase inhibitor
69.	EXPT01236	1-Deoxy-Nojirimycin	$C_6H_{13}N_1O_4$	D-Xylose Isomerase inhibitor
70.	EXPT01237	7,8-Diamino-Nonanoic Acid	$C_9H_{20}N_2O_2$	Dethiobiotin Synthetase inhibitor
71.	EXPT01238	3-Amino-Alanine	$C_3H_9N_2O_{21}+$	Edap : Ace-Ile-Trp-Glu-Ser-Gly-Lys-Leu-Il inhibitor
72.	EXPT01239	Dnqx	$C_8H_2N_4O_6$	Glutamate Receptor 2 inhibitor
73.	EXPT01240	2-Amino-6-Oxo-Hexanoic Acid	$C_6H_{11}N_1O_3$	Glutaminase-Asparaginase inhibitor
74.	EXPT01241	2,4-Dihydroxybenzoic Acid	$C_7H_6O_4$	P-Hydroxybenzoate Hydroxylase inhibitor
75.	EXPT01242	2',3'-Dideoxycytidine-5'-Monophosphate	$C_9H_{14}N_3O_6P_1$	Cytidylate Kinase inhibitor
76.	EXPT01244	4-(3,12,14-Trihydroxy-10,13-Dimethyl-Hexadecahydro-Cyclopenta[a]Phenanthren-17-Yl)-5h-Furan-2-One	$C_{23}H_{34}O_5$	Diga16 inhibitor
77.	EXPT01245	Beta-Hydroxy Aspartic Acid	$C_4H_7N_1O_5$	Coagulation Factor X inhibitor
78.	EXPT01246	Dalfopristin	$C_{34}H_{50}N_4O_9S_1$	Streptogramin A Acetyltransferase inhibitor
79.	EXPT01247	2'-Deoxymaltose	$C_{12}H_{22}O_{10}$	Beta-Amylase inhibitor
80.	EXPT01248	Domoic Acid	$C_{15}H_{21}N_1O_6$	Glutamate Receptor, Ionotropic Kainate 2 inhibitor

81.	EXPT01249	Dihydroorotic Acid	$C_5H_6N_2O_4$	Dihydroorotase inhibitor
82.	EXPT01250	Delta-Bis(2,2'-Bipyridine)Imidazole Osmium (Ii)	$C_{23}H_{20}N_6Os_{12}+$	Azurin inhibitor
83.	EXPT01252	L-N(Omega)-Nitroarginine-2,4-L-Diaminobutyric Amide	$C_{10}H_{22}N_8O_4$	Nitric-Oxide Synthase, Endothelial inhibitor
84.	EXPT01253	N-{(4s)-4-Amino-5-[(2-Aminoethyl)Amino]Pentyl}-N'-Nitroguanidine	$C_8H_{21}N_7O_2$	Nitric-Oxide Synthase, Endothelial inhibitor
85.	EXPT01254	L-N(Omega)-Nitroarginine-(4r)-Amino-L-Proline Amide	$C_{11}H_{22}N_8O_4$	Nitric-Oxide Synthase, Endothelial inhibitor
86.	EXPT01255	Dpb-T	$C_{17}H_{19}N_2O_8P_1$	5 inhibitor
87.	EXPT01256	Dipyrromethane Cofactor	$C_{20}H_{24}N_2O_8$	Porphobilinogen Deaminase inhibitor
88.	EXPT01257	D-Phenylalanine	$C_9H_{11}N_1O_2$	Sandostatin inhibitor
89.	EXPT01258	Diphosphate	O_7P_{24}-	Nh inhibitor
90.	EXPT01259	D-Proline	$C_5H_9N_1O_2$	Tetrameric Beta-Beta-Alpha Mini-Protein inhibitor