

A QSAR Study on Biological Activities of Bisphosphonates Compounds as Anticancer Drugs

M. Talebi, Gh. Ghasemi, H. Kefayati

Abstract—Bisphosphonates (BPs) are drugs used to slow down or prevent bone damage. BPs treatment can stop some types of cancer from spreading into the bone. Studies have also shown that bisphosphonates, especially nitrogen-containing bisphosphonates can help with myeloma, breast cancer and prostate cancer. In this research, we wish to report, Quantitative structure activity relationship (QSAR) study on biological activities of some compounds of BPs. Genetic algorithm (GA), artificial neural network (ANN) and multiple linear regression (MLR) were used to create the nonlinear and linear QSAR models. In the gas phase the root-mean-square errors of the training set and the test set for GA-ANN model using jack-knife method, were 0.1363, 0.3335 and R^2 was 0.951. Also, the R and R^2 values in the gas phase were obtained 0.934, 0.872 from GA-MLR model. We find out GA-ANN model is the most favorable method toward the other statistical methods.

Keywords—anticancer drugs, Bisphosphonates, Genetic algorithm (GA), QSAR model

I. INTRODUCTION

CANCER is the uncontrolled growth of abnormal cells anywhere in a body. Cancer is a leading cause of death worldwide. Deaths from cancer worldwide are projected to continue rising, with an estimated 12 million deaths in 2030. Many studies have demonstrated various effects of bisphosphonates on several cancer cells and it is accepted that their anti-tumor activity is related to interference with the mevalonate pathway[1]. Bisphosphonates are a chemical class of compounds in widespread use since the 1970s for the management of disorders of bone metabolism, such as Paget's disease and osteoporosis[2]. Bisphosphonates, especially nitrogen-containing bisphosphonates, are widely used to block bone destruction in cancer patients with bone metastasis because they are effective inhibitors of osteoclast-mediated bone resorption. In addition to their antiresorptive effects, preclinical evidence strongly suggests that nitrogen-containing bisphosphonates exert direct and indirect anticancer activities through inhibition of tumor cell functions, enhancement of the cytotoxic activity of chemotherapy agents, inhibition of tumor angiogenesis, and stimulation of antitumor immune reactions.

nitrogen-containing bisphosphonates also inhibit the growth of MDA-MB-231 breast cancer cells by inducing G1/S cell cycle arrest[3]. *Drug design* frequently relies on computer modeling techniques. Computer-assisted drug design uses computational chemistry to discover, enhance, or study drugs and related biologically active molecules. The most fundamental goal is to predict whether a given molecule will bind to a target and if so how strongly. *QSAR* attempts to find consistent relationship between biological activity and molecular properties, so that these "rules" can be used to evaluate the activity of new compounds. Today, QSARs are being applied in many disciplines with much emphasis in drug design. Over the years of development, many methods, algorithms and techniques have been discovered and applied in QSAR studies[4],[5]. Genetic Algorithms (GAs) are often used to improve the performance of other AI methods. A genetic algorithm maintains a population of candidate solutions for the problem at hand, and makes it evolve by iteratively applying a set of stochastic operators. The aim of this study is to assess QSAR models reliability, using GA-ANN methods for prediction of new anticancer compounds.

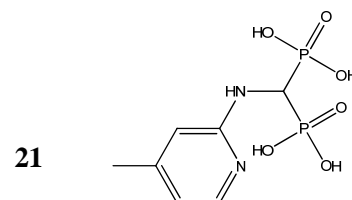
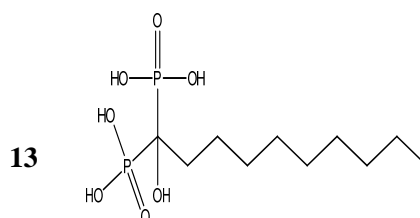
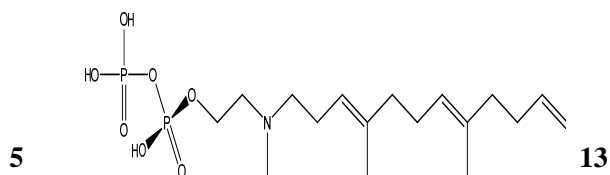
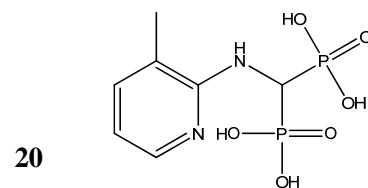
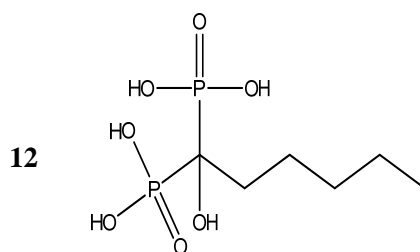
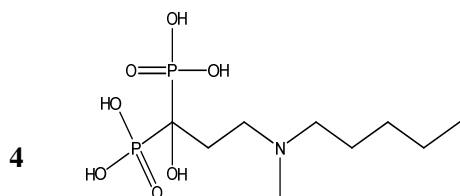
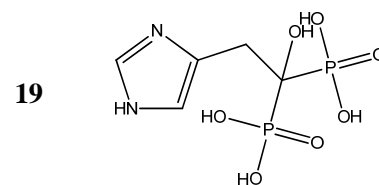
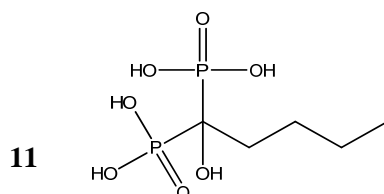
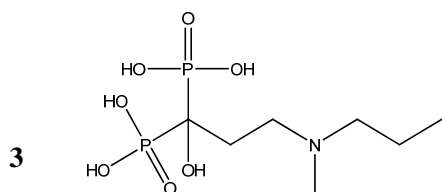
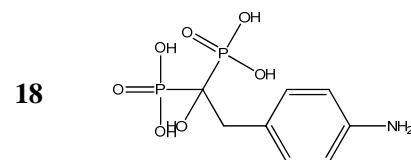
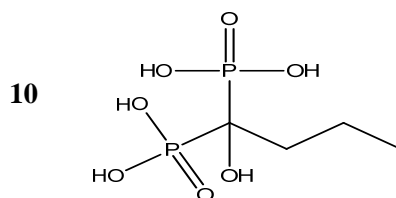
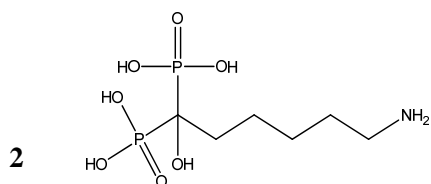
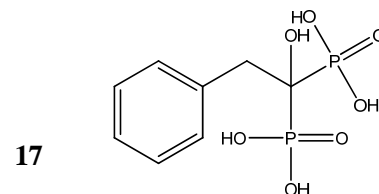
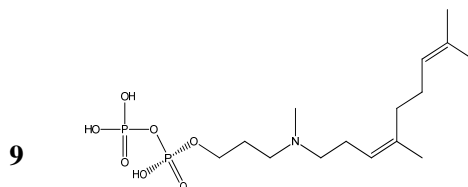
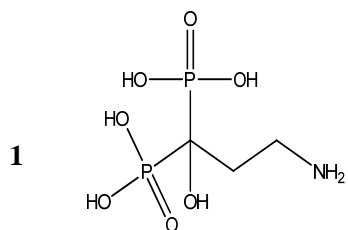
II. COMPUTATIONAL DETAILS

The 3D structures of the molecules were generated using the built optimum option of Hyperchem software (version 8.0). Then, the structures were fully optimized based on the ab initio method, using DFT level of theory. Hyperchem, ChemOffice and Dragon (version 3.0) programs were employed to calculate the molecular descriptors. All calculations were performed using Gaussian 98W program series. Geometry optimization of compounds was carried out by B3LYP method employing 6-31G (d) basis set. In this study, the independent variables were molecular descriptors and the dependent variables were the actual half maximal inhibitory concentration (IC₅₀) values. Overall, more than 1498 theoretical descriptors were selected and calculated. These descriptors can be classified into several groups including: (i) geometrical, (ii) MoRSE, (iii) galvestopol, (iv) autocorrelations descriptors. For each compound in the training sets, the correlation equation was derived with the same descriptors. Then, the obtained equation was used to predict log (1/IC₅₀) values for the compounds from the corresponding test sets. In the present work, stepwise multiple linear regression (stepwise-MLR) and GA variable subset selection methods were used for the selection of the most relevant descriptors from all of the descriptors. These descriptors would be used as inputs of the ANN. Totally 1498 descriptors were generated that were too many to be fitted in

F. M. Talebi, S. Gh. Ghasemi and T. H. Kefayati are with the Department of Chemistry, Islamic Azad University-Rasht branch, Iran (e-mail: ramo.6468@yahoo.com, Ghasemi@iaurasht.ac.ir, Haskefayati@gmail.com)

our models. So, it was necessary to reduce the number of descriptors through an objective feature selection which was performed in three steps. First, descriptors that had the same value for at least 70% of compounds within the dataset were removed. In next step, descriptors with correlation coefficients less than 0.3 with the dependent variable were regarded

redundant and removed. Finally, since highly correlated descriptors provide approximately identical information, a pair wise correlation was performed. When their correlation coefficient exceeded 0.90, one of two descriptors was randomly removed. GA was utilized as the mean for non-linear feature selection.



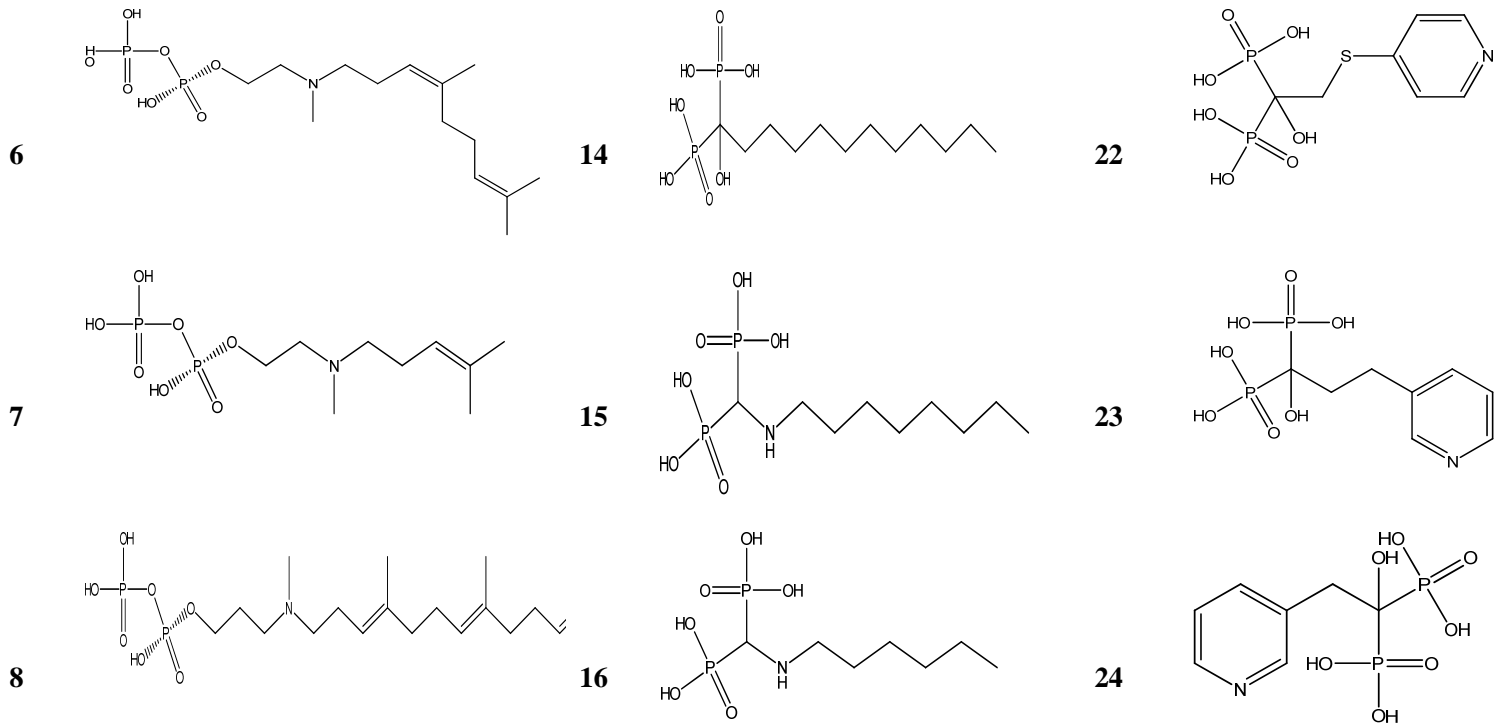


Fig. 1 The molecular structure of bisphosphonates analogues

TABLE I
EXPERIMENTAL AND PREDICTED VALUES OF LOG (1/IC50) USING JACK-KNIFE MODEL

Calculated (Jack-Knife) gas	Observed log (1/IC50)
2.2500	2.1625
2.8400	2.3624
2.5200	2.3057
1.9200	2.4632
-0.8500	-0.5183
-0.1300	-0.1412
2.3800	1.5468
-0.4300	-0.4748
-0.5100	-0.3381
2.7900	2.5210
2.3000	2.6558
1.7200	1.3038
-0.1400	0.9336
-0.0400	0.1543
0.3400	1.0206
1.2800	1.9576
2.3400	2.3718
2.2600	2.5285
2.3400	2.4222
2.3400	2.5751
2.4200	2.7906
2.7400	2.1798
2.6100	2.3946
2.5400	1.9134

TABLE II
DESCRIPTORS VALUES FOR GA-MLR MODEL

Molecule	JGI1	MATS5p	FDI	Mor30u
1	0.500	0.017	0.984	0.276
2	0.400	0.048	0.992	0.029
3	0.438	0.173	0.958	0.224
4	0.389	0.184	0.974	0.295
5	0.339	0.089	0.956	1.301
6	0.370	0.100	0.904	0.330
7	0.417	0.120	0.953	0.269
8	0.328	-0.027	0.977	0.470
9	0.354	-0.040	0.912	0.588
10	0.500	-0.069	0.996	0.117
11	0.462	0.058	0.978	0.276
12	0.429	0.052	0.987	0.223
13	0.333	0.040	1.000	0.395
14	0.300	0.029	1.000	0.601
15	0.324	0.016	0.996	0.557
16	0.367	0.02	0.975	0.000
17	0.406	0.054	0.995	-0.125
18	0.441	0.064	0.986	0.230
19	0.433	0.085	1.000	0.045
20	0.406	0.214	0.972	0.398
21	0.438	0.147	0.977	-0.116
22	0.382	0.151	0.999	0.438
23	0.382	0.239	1.000	-0.140
24	0.406	-0.031	0.996	-0.026

III. RESULTS AND DISCUSSIONS

The structures of the bisphosphonates compounds used in this study are shown in Fig. 1. The efficiency of the QSAR model to predict log (IC₅₀) value was also estimated using the internal cross-validation method. The resulted predictions of the log (1/IC₅₀) in gas phase is given in Table 1. Two linear and non-linear variable selection methods were used to select the most significant descriptors (stepwise-MLR and GA) (Table 2). The selected descriptors through these methods were used to construct some linear and non-linear models by using MLR and ANN methods. Based on the types of variable selection method and also the types of the feature mapping technique, these models can be shown as MLR- ANN, GA-MLR and GA- ANN[6]-[8]. It revealed that the GA-ANN model was much better than other models. The four most significant descriptors which were selected by GA are as follows:

JGI1, Mor30u, MATS5p, FDI.

These GA selected descriptors were used as inputs for the construction of ANN model.

As can be seen from this table, topological charge and atomic polarizability were important descriptors in our study.

In the present study, two linear and non-linear variable selection methods were used to select the most significant descriptors. The MLR, ANN and GA were used to construct a quantitative relation between activities of bisphosphonates compounds and their calculated descriptors.

IV. CONCLUSION

In the present study, two linear and non-linear variable selection methods were used to select the most significant descriptors, and the MLR, ANN and GA were used to construct a quantitative relation between the activities of bisphosphonates compounds and their calculated descriptors. ANN has been successfully used for finding a QSAR model for bisphosphonates compounds. It provides the best results among those we have tested. Our present attempt to correlate

the log (1/IC₅₀) with theoretically calculated molecular descriptors has led to a relatively successful QSAR model that relates this complex

The results obtained from this work indicate that the linear regression and ANN models exhibit reasonable prediction capabilities. Though the linear model was developed mainly for the purpose of structure-activity interpretation, the ANN model was primarily developed for predictive ability and classification.

ACKNOWLEDGMENT

We thank the Research Vice Presidency of Islamic Azad University, Rasht Branch for their encouragement, permission and financial support.

REFERENCES

- [1] M. T. Valenti, L. D. Carbonare, F. Bertoldo, L. Donatelli, V. Lo Cascio, European Journal of Pharmacology. 2008, 1-2, 36-42
- [2] R. Graham, G. Russell, Bone, 2007, 40, 5, S21-S25
- [3] M. A. Merrell, S. Wakchoure, P. P. Lehenkari, K. W. Harris, K. S. Selander, European Journal of Pharmacology, 2007, 570, 1-3, 27-37
- [4] L. He, P. C. Jurs, Journal of Molecular Graphics and Modelling. 2005, 23, 503-523
- [5] D.V. Eldred, C.L. Weikel, P.C. Jurs, K.L.E. Kaiser, Chem. Res. Toxicol, 1999, 12, 670-678
- [6] de Weijer, A.P. Buydens, L. Kateman, G. Chemom. Intell Lab. Syst, 1992, 16, 77-86.
- [7] Tominaga, Y. Chemom. Intell. Lab. Syst. 1999, 49, 105-115.
- [8] Manallack, D.T. Livingstone, D.J. Eur. J. Med. Chem, 1999, 34, 195-208.
- [9] S. E. Papapoulos, Best Practice & Research Clinical Endocrinology & Metabolism. 2008, 22, 831-847.
- [10] M. L. Lolli, B. Rollando, P. Tosco, S. Chaurasia, A. Di stillo, L. Lazzarato, E. Gorrassini, R. Ferracini, S. Bosso, R. Fruttero, A. Gasco, Bio organic & Medicinal Chemistry. 2010, 18, 2428-2438.
- [11] M. C. Winter, I. Holen, R.E. Coleman, Cancer Treatment Reviews. 2008, 34, 453-475.