International Journal of
Mathematical Modelling & Computations
Vol. 01, No. 04, 2011, 263 - 269

Understanding Behavior of Antineoplastic Molecules Based on MLR Models

S. Hajiali[a], M. Doroudian[b*] , and H. Borna [c]

[a] *Department of Genetics, Tehran Medical Branch, Islamic Azad University, Tehran, Iran.*
[b] *Young Researchers Club, Central Tehran Branch, Islamic Azad University, Tehran, Iran.*
[c] *Young Researchers Club, Medical University Branch, Islamic Azad University, Tehran, Iran.*

Received: 13 July 2011; Accepted: 10 September 2011.

Abstract. New statistic based models provide a wide area of prediction equipments for different science areas. Among these fields biology have just entered the contest of interdisciplinary sciences. Drug discovery is a long and expensive process which could be decreased with theoretical approaches. In this study, 500 reported assayed anti cancer molecules were extracted from Science Direct articles, sketched, optimized with AM1 basis set with Gaussian, and analyzed for their bioactivities with PASS. Molecular and physicochemical descriptors such as MW, H Donor/Acceptor, $\log P$, Parachor, and etc of 15 screened compounds were calculated with Dragon and applied for Stepwise Multi Linear Regression process within SPSS. Calculations present that H Donor, H Acceptor, Polarizability, and Refractivity descriptors are the most impacting factors on emergence of antineoplastic activity with higher than 91% accuracy. A valid model is achieved which will be utilized for future molecule discovery and prediction.

Keywords: Antineoplastic, Multi Linear Regression, Model, PASS, Gaussian.

Index to information contained in this paper

1. Introduction

Nowadays growing technology has different side effects on industrializing areas. Life styles have changed and concepts of health care been ignored. Fresh nutrients are changed into canned pasteurized stuff and vitamins or minerals are shifted into tablets. In such conditions various diseases or disorders are widely emerged such as cancers, viral illnesses, sexual disorders and etc. cancers are the most abundant diseases among other illnesses which cost too much to treat [1]. Every day new anti cancer drugs are discovered or commercialized to help ill people but even newer

*Corresponding author. Email: mohamad.doroudian@gmail.com.

and more powerful drugs are needed to be designed in order to overcome new malignancies and drug resistances [6]. Discovering new drugs and pushing them through screening processes from in-vitro to clinical trials need more than 500 to 800 million dollars and 8 to 10 years [8].

Along with development of theoretical basics of science and invention of advanced approaches in statistical and numerical methods, computers have become more powerful than before. Theoretical methods are getting too improved that some times by pass all expensive laboratory experiments. Multi linear Regressions (MLR), Quantitative Structure and Activity Relationships (QSAR) and etc are of the most famous mathematical models with biological and chemical usages [8]. This study focuses on applications of MLR models on design and development of new antineoplastic molecules which are compatible with natural molecules.

2. Materials and Methods

In order to extract naturally active and assayed molecules, approximately all articles of Science Direct publishing group between 2006 and 2010 were studied for reports of anti cancer compounds and their sources or targets. 3D structures of molecules were sketched with Marvin Sketch, ChemAxon series. Structures were optimized for proper data extraction with Gaussian, Gaussian Inc., AM1 semi empirical basis set. In this research, molecules derived from microorganisms such as bacteria, alga, protozoa and etc are studied for their biological activities with PASS software [9]. Some well studied and important Physicochemical and Molecular descriptors such as MW, Parachore, LogP, Molar Refractivity, Surface Tension, Polarizability, Refractivity, H Donor/Acceptor, Polar Surface Area, and Molar Volume have been extracted with DRAGON 5.4. Regressions were calculated and analyzed for all descriptors and biological activities with SPSS 16. The best MLR model was achieved after comprehensive analysis [7].

3. Results

More than 500 molecules were gathered after mining of articles. Compounds derived from microorganisms were screened and sketched. 16 molecules among all 50 molecules of microorganisms had very high or absolute antineoplastic activity. Descriptors were derived and analyzed with SPSS. Table 1 describes the molecules and activities along with their descriptor.

Table 1. Names, Activities, and descriptors of screened molecules with Antineoplastic activities

Molecule Name	Antineoplastic Activity	MW	MR	MV	Parachore	ST	Pol	logP	PSA	Hd	Ha	Refractivity
caffeic acid phenethyl ester	1	284.3	81.43	224.4	616.7	56.9	32.28	3.92	66.76	3	2	81.16
Calicaemicin	0.938	1368.34	322.66	871.5	2582.2	77	127.91	4.8	308.44	8	27	332.63
Chromomycin A3	0.593	1191.31	294.58	845.8	2440.8	69.3	116.78	3.74	340.88	8	22	290.43
Cordycepin	1	251.24	59.1	130.8	407.1	93.6	24.43	-1.4	119.31	3	7	61.9
Cytochalasin B	1	249.6	135.03	396.8	1072.3	53.3	52.25	4.08	95.86	3	4	136.42
Depsipeptide	1	540.69	172.02	460.02	1155.7	39.7	56.3	1.08	142.7	4	7	142.1
Discodermolide	1	593.8	165.6	543.5	1393.1	43.1	65.64	4.17	154.55	5	6	166.77
Doxorubicin	1	543.52	131.52	336.6	1054	96.4	52.13	1.5	206.07	6	12	134.59
Epothiolone	1	507.68	138.37	446.5	1134.2	41.6	54.85	4.12	109.25	2	7	134.76
Mithramycin A	1	1085.15	262.26	728.3	2174.2	79.4	103.96	2.07	358.2	11	24	257
Mitomycin	1	334.32	80.79	213.6	633.5	77.2	32.02	-1.05	146.89	3	7	83.26
Okadiac acid	1	805	210.35	627.7	58.3	58.3	83.39	5.13	182.83	5	13	210.78
Pinocembrin	1	256.25	68.37	184.8	520.2	62.7	27.1	3.14	66.76	2	4	69.3
Rhizoxin	1	625.7	168.22	536.6	1364.8	41.8	66.68	3.72	133.15	1	7	168.66
Sirolimus	1	914.17	246.69	773.4	2071	51.4	97.79	7.45	195.43	3	12	250.66
Staurosporine	1	466.53	128.82	298.7	841.1	62.8	51.06	3.97	69.45	2	4	132.37

Figure 1 presents a sample molecule in PASS platform which is predicted for its different activities. This software uses Multilevel Neighborhoods of Atoms (MNA) model in order to predict the activity of compounds with mean accuracy of higher than 95%.

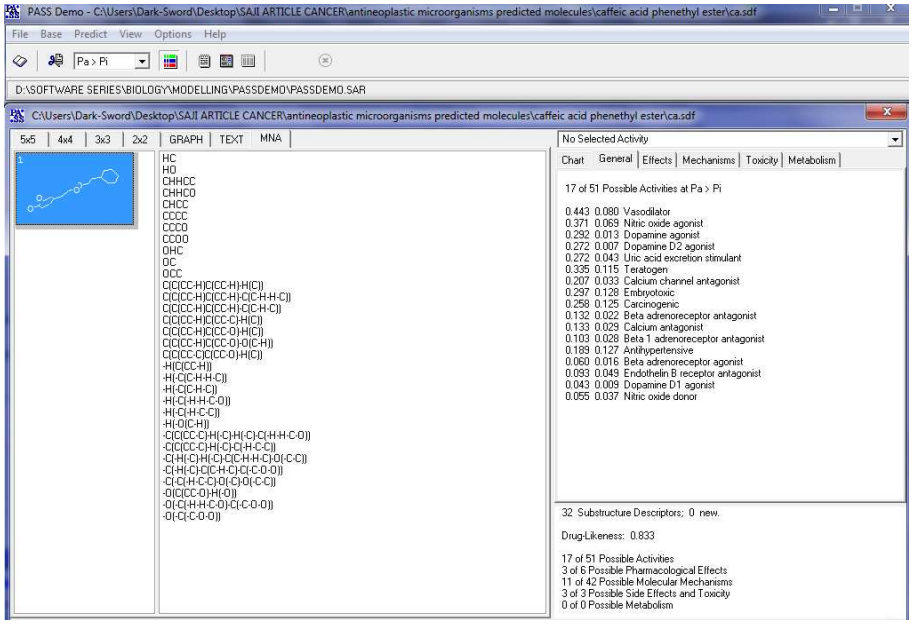

Figure 1. PASS software platform while predicting activities of Caffeic Acid phenethyl ester.

Step wise reduction of descriptors and best multi linear regression used to achieve the latest model in which both number of descriptors and accuracy of prediction are high and meaningful [2]. Table 2 demonstrates possible error and R2 of predicted model. Table 3 specifies precision of model with ANOVA analysis of it.

Table 2. Statistical parameters of built BMLR model, a. Predictors: (Constant), Refractivity, Hd, Ha, Polarizability, Dependent Variable: Antineoplastic Activity

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.958[a]	0.917	0.835	0.007

Table 3. ANOVA test for confirmation of achieved BMLR model, a. Predictors: (Constant), Refractivity, Hd, Ha, Polarizability, Dependent Variable: Antineoplastic Activity

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.003	4	0.000	11.107	0.003[a]
Residual	0.000	10	0.000	-	-
Total	0.004	14	-	-	-

Coefficients of descriptors are presented in Table 4 which are ingredients of equation for prediction of bioactivity.

So the equation for prediction of antineoplastic activity of unknown molecule is presented in Equation 1 according to information of Table 4.

Table 4. Coefficients of selected descriptors of model, a. Dependent Variable: Antineoplastic Activity

Model	Unstandardized Coefficients		Standardized Coefficients	t	sig.
	B	Std. Error	Beta		
(Constant)	0.991	0.007	-	138.862	0.000
Polarizability	0.006	0.002	12.211	3.230	0.014
H bond donor	−0.003	0.002	−0.468	−1.514	0.174
H bond acceptor	−0.006	0.001	−2.636	−4.120	0.004
Refractivity	−0.002	0.001	−10.294	−3.304	0.013

So the equation for prediction of antineoplastic activity of unknown molecule is presented in Equation 1 according to information of Table 4.

$$\text{Antineoplastic Activity} = 0.991 + 0.006(\text{Polarizability}) - 0.003(\text{H Bond Donor}) - 0.006(\text{H Bond Acceptor}) - 0.002(\text{Refractivity}) \quad (1)$$

Figure 2 presents expected data against observed data which have been entered according to Table 1. This plot specifies that there were no unfit data in entered data.

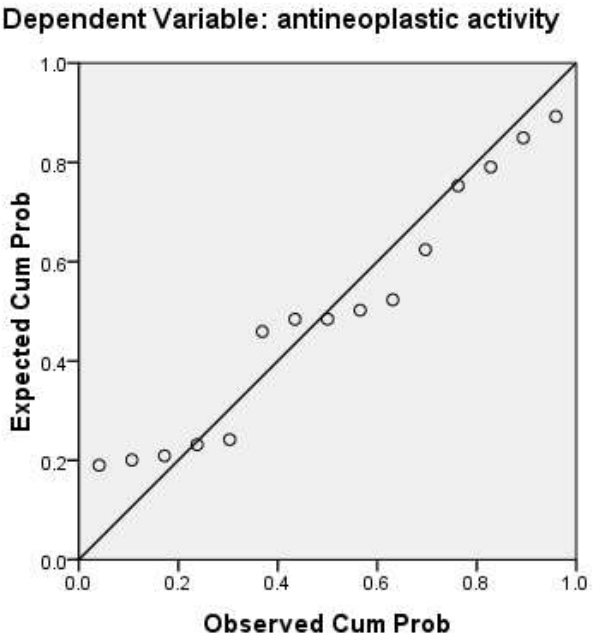

Figure 2. Expected and observed data plot of model prediction activity.

4. Discussion

This article focuses on establishment of a relationship between some molecular and physicochemical characteristics of few compounds with their bioactivities. As it is visible in Table 5 drug likenesses of screened molecules are higher than 90%. It determines that nearly all natural molecules are like drugs and compatible with natural systems (8). It indirectly helps design of new drug candidates, because not only achieved formula specifies exact descriptors, but also nature of molecules has Lipinski rules inside [7].

Table 5. Drug likeness of screened compounds

Molecule Name	Drug Likeness	Molecule Name	Drug Likeness
Caffeic Acid Phenethyl Ester	0.833	Epothiolone	0.991
Calicaemicin	0.991	Mithramycin A	0.994
Chromomycin A3	0.994	Mitomycin	0.994
Cordycepin	0.992	Okadiac acid	0.991
Cytochalasin B	0.994	Pinocembrin	0.981
Depsipeptide	0.995	Rhizoxin	0.994
Discodermolide	0.995	Sirolimus	0.995
Doxorubicin	0.994	Staurosporine	0.992

This equation now is being used by researching team to design new molecules with antineoplastic activity in order to be synthesized in chemistry labs and assayed against different cancerous cell lines.

References

[1] American Cancer Society,. Cancer Facts and Figures, Atlanta, Ga: American Cancer Society, (2011).
[2] Ayati M.,Ghasemi J. B. , A QSPR Study of GC/MS Retention Data 85 volatile organic compounds as air pollutant materials by multivariate method, CHA4CH, **3** (2010).
[3] Behrangi N., Hashemi M., Doustar Y., Borna H., Camptothecins And Their Novel Anticancer Properties Evaluated By Using Pass Method, Advances in Environmental Biology, **5(9)**, (2011) 2551-2556.
[5] Hanahan D., Weinberg R. A., The Hallmarks of Cancer, Cell, **100**, (2000), 57-70.
[5] Hashemi M., Borna H., Dadras O., Shafaroodi H., A. Akbarzade, S. Hajiali, Mechanistical Approach Through Discovery of New Generations of Anti Inflammatory Drugs, J Proteomics Bioinform, **4(12)**, (2011),284-288.
[6] Jemal A., Thun M. J., Ries L. A., Howe HL, Weir HK, et al. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use and tobacco control. J Natl Canc Inst, **100**, (2008) 1672-94.
[7] Kasam V. K., In silico drug discovery on computational Grids for finding novel drugs against neglected diseases, Ch **1**, (2009), 1-20.
[8] Vahdani S., Bayat Z., A Quantitative Structure-Activity Relationship (QSAR) Study Of Anti Cancer Drugs, Der Chemica Sinica, **2 (4)**, (2011), 235-243.
[9] Varnek A., Tropsha A., Chemoinformatics Approaches to Virtual Screening, **6**, (2008), 182 - 216.