ISSN 2348-6201

QSAR investigation on benzimidazole derivatives in Trichomonosis' disease

Alireza Nemati Rashtehroodi[1], *Ghasem Ghasemi*[2],

[1] *Young Researchers club and elite, Islamic Azad University, Rasht Branch, Rasht, Iran*
Mr.Alireza.Nemati@gmail.com

[2] *Department of Chemistry, Islamic Azad University, Rasht Branch, Rasht, Iran*
Ghasemi@iaurasht.ac.ir

ABSTRACT

Globally trichomoniasis affects approximately 152 million people as of 2010 (2.2% of the population). It is more common in women (2.7%) than males (1.4%). The American Social Health Association estimates trichomoniasis affects 7.4 million previously unaffected Americans each year and is the most frequently presenting new infection of the common sexually transmitted diseases. On the pattern, QSAR study has been done on benzimidazole derivatives as potent inhibitors with trichomonicidal activity. Genetic algorithm (GA), artificial neural network (ANN), stepwise multiple linear regression (stepwise-MLR) were used to create then on non-linear and linear QSAR models. Geometry optimization of compounds was carried out by B3LYP method employing 6–31G (2d) basis set. HyperChem, Gaussian 03W, and Dragon (version 5.5) software programs were used for geometry optimization of the molecules and calculation of the quantum chemical descriptors. Finally, Unscrambler program was used for the analysis of data. The root-mean square errors of the training set and the test set for GA–ANN model using jack-knife method, were 0.1840, 0.5051 and R2 was 0.70. Also, the R and R2 values in the gas phase were obtained 0.78, 0.61 from GA-stepwise MLR model. According to the obtained results, we find out GA-ANN model is the most favorable method toward the other statistical methods. Also, we would suggest that compounds No. 20, 33, 58, 48 and 47 as the most appropriate structure for the design of drugs to pharmacists.

Indexing terms/Keywords

Benzimidazole derivatives; QSAR model; Genetic Algorithm; Artificial Neural Network.

Academic Discipline And Sub-Disciplines

Biotechnology; Organic Chemistry; Inorganic Chemistry and etc.

SUBJECT CLASSIFICATION

Medicinal Chemistry Subject Classification; 21st Iranian Seminar of Organic Chemistry- Ilam University

TYPE (METHOD/APPROACH)

Genetic algorithm (GA), artificial neural network (ANN), stepwise multiple linear regression (stepwise-MLR) were used to create then on non-linear and linear QSAR models. QSAR study has been done on benzimidazole derivatives as potent inhibitors with trichomonicidal activity.

Council for Innovative Research

Peer Review Research Publishing System

Journal: JOURNAL OF ADVANCES IN BIOTECHNOLOGY

Vol .4 , No. 3

www.cirjbt.org , jbteditor@gmail.com

ISSN 2348-6201

INTRODUCTION

Trichomoniasis (or “trich”) is a very common sexually transmitted disease (STD) that is caused by infection with a protozoan parasite called Trichomonas vaginalis. Although symptoms of the disease vary, most women and men who have the parasite cannot tell they are infected. Trichomoniasis is considered the most common curable STD. In the United States, an estimated 3.7 million people have the infection, but only about 30% develop any symptoms of trichomoniasis. Infection is more common in women than in men, and older women are more likely than younger women to have been infected [1]. Trichomoniasis can increase the risk of getting or spreading other sexually transmitted infections. For example, trichomoniasis can cause genital inflammation that makes it easier to get infected with the HIV virus, or to pass the HIV virus on to a sex partner [2]. Benzimidazole is a heterocyclic aromatic organic compound. It is an important pharmacophore and a privileged structure in medicinal chemistry. This compound is bicyclic in nature which consists of the fusion of benzene and imidazole. Nowadays is a moiety of choice which possesses many pharmacological properties [3].Structure-activity modeling plays an important role in government programs in support of protecting human populations from exposure to environmental contaminants [4]. Specifically, computational methods to identify chemicals that may pose endocrine disruption hazard for additional in vitro or in vivo testing are important prioritization approaches [5].QSAR is an effective way of optimizing or correlating the biological activity within congeneric series with certain structural features or with atomic, group or molecular descriptors, such as lipophilicity, polarizability, and electronic and steric properties. Previously, we reported several examples of adopting a QSAR approach to probe the nature of interactions of various classes of ligands towards the cyclooxygenase enzyme [6-12]. 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. [13], [14].

EXPERIMENTAL

The structures of the benzimidazole derivatives used in this study [15] (Table 1). 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 5.5) programs were employed to calculate the molecular descriptors. All calculations were performed using Gaussian 03W program series. Geometry optimization of compounds was carried out by B3LYP method employing 6–31G (2d) 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 3224 theoretical descriptors were selected and calculated. Finally, Unscrambler (version 9.7) program was used for analysis of data and statistical methods. 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, the method of stepwise multiple linear regression (stepwise, MLR) to select the most appropriate descriptor of all descriptor was used. Totally 3224 descriptors were generated that were too many to be fitted in 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 0.4 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.

RESULTS AND DISCUSSIONS

Based on the types of variable selection method and also the types of the feature mapping technique, these models can be shown as GA-ANN Jack–Knife and cross validation. It revealed that the GA–ANN model was much better than other models is given in Table 2. Considering the experimental error, the overall prediction of the log (1/IC50) values was quite satisfactory.

The efficiency of the QSAR model to predict log (IC50) value was also estimated using the internal cross-validation method, this resulted predictions of the log (1/IC50) using Jack–Knife model is given in Table 3.

In the present study, linear variable selection methods were used to select the most significant descriptors (stepwise MLR) (Table 4).

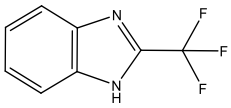
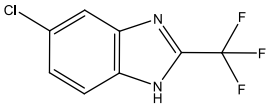
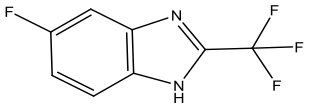
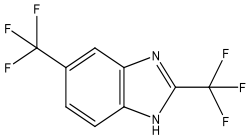
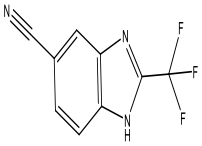
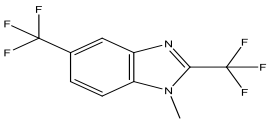
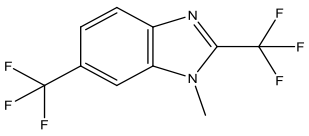
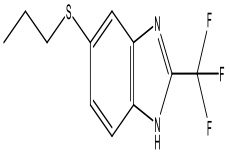
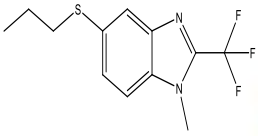
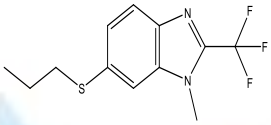
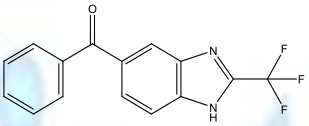
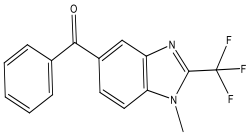
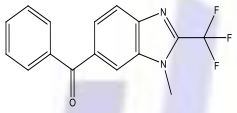
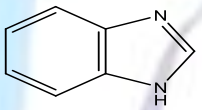
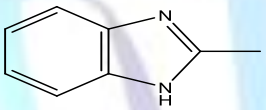
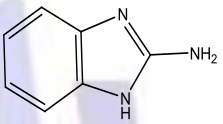
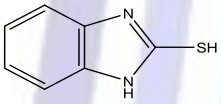
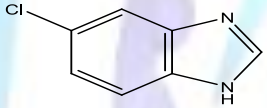
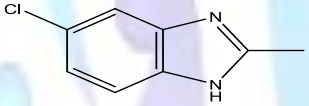
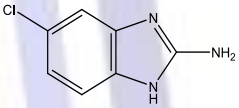
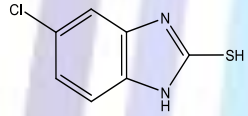
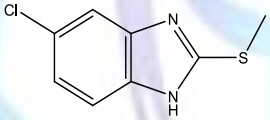
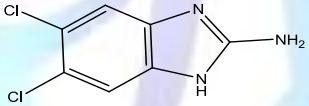
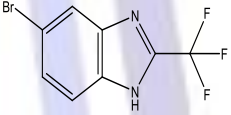
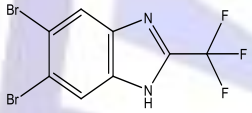
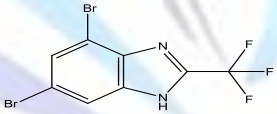
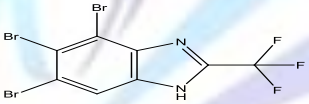
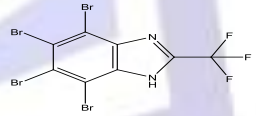
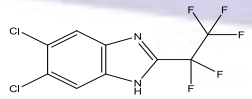
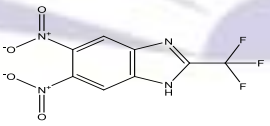
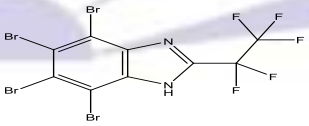
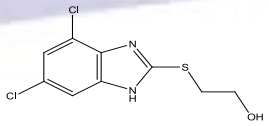
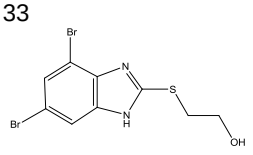
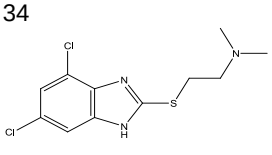
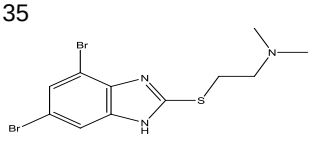
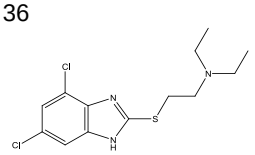
The six most significant descriptors which were selected by stepwise MLR are as follows:

BLI, RDF090m, G(N..N), MATS5v, RDF035u, HATS1e.

Results from the stepwise MLR method is shown in Figure 2. According to this figure, R2 between the expected values and the observed values were near to a stright line. Also, the R and R2 values were obtained as 0.78, 0.61 from stepwise MLR model. Electronegativities, atomic masses and atomic van der Waals volumes were important descriptors in this study.

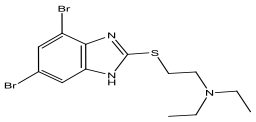
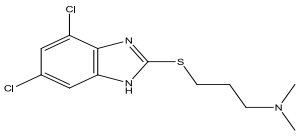
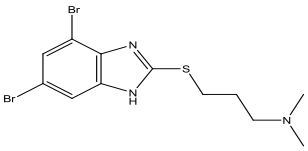
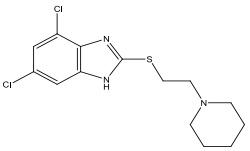
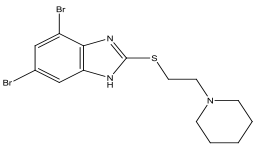
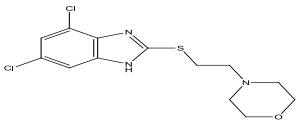
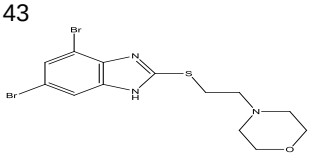
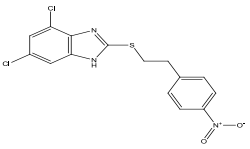
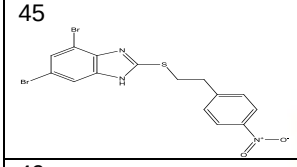
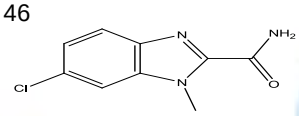
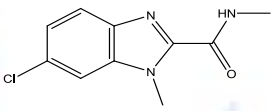
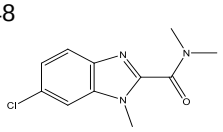
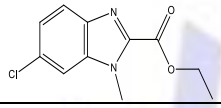
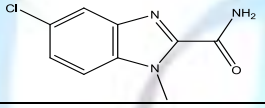
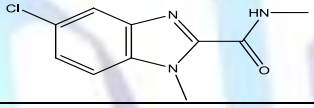
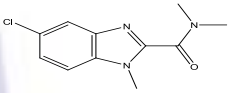
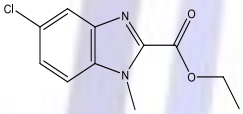
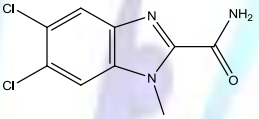
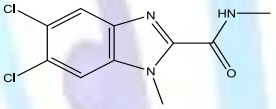
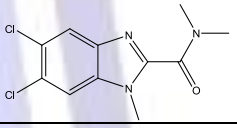
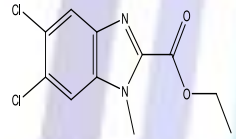
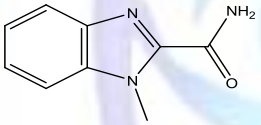
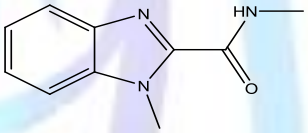
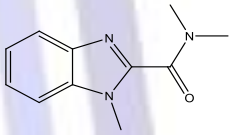
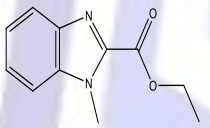
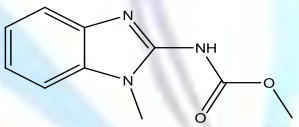
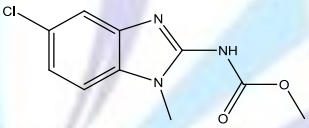
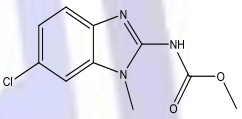
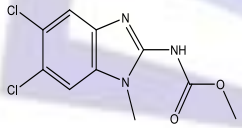
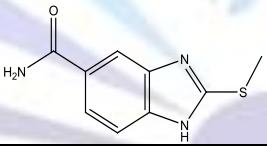
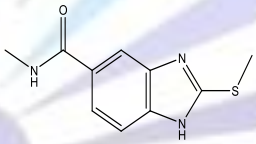
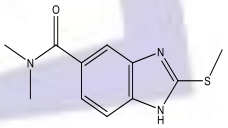
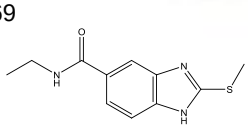
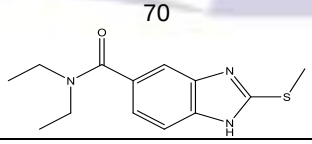
ISSN 2348-6201

Table1. Structures of benzimidazole derivatives used for QSAR model building

1 	2 	3 	4 
5 	6 	7 	8 
9 	10 	11 	12 
13 	14 	15 	16 
17 	18 	19 	20 
21 	22 	23 	24 
25 	26 	27 	28 
29 	30 	31 	32 
33 	34 	35 	36 

ISSN 2348-6201

Table 1: Continued

37 	38 	39 	40 
41 	42 	43 	44 
45 	46 	47 	48 
49 	50 	51 	52 
53 	54 	55 	56 
57 	58 	59 	60 
61 	62 	63 	64 
65 	66 	67 	68 
69 	70 		

ISSN 2348-6201

Table 2. The statistical parameters of different constructed QSAR models

Parameters / Method	RMSE $_{test}$	RMSE $_{train}$	R^2
GA–ANN Jack–Knife (gas)	0.5051	0.1840	0.70
GA–ANN (gas) cross validation	0.7316	0.1439	-

Table 3. Experimental and predicted values of log (1/IC50) using Jack–Knife method.

Molecule		Observed log (1/IC50)		Predicted Jack–Knife		ΔΘ Jack–Knife	
1	36	5.500	7.390	5.6596	7.5133	0.1596	0.1233
2	37	6.350	7.530	6.1778	7.8486	0.1722	0.3186
3	38	5.500	7.070	5.7521	8.0758	0.2521	1.0058
4	39	6.630	7.990	4.9851	6.7325	1.6449	1.2575
5	40	5.640	7.140	5.0489	6.6407	0.5911	0.4993
6	41	5.390	7.690	4.6203	7.8830	0.7697	0.1930
7	42	5.270	7.290	5.4971	7.2359	0.2271	0.0541
8	43	6.460	7.230	7.0341	7.2593	0.5741	0.0293
9	44	6.700	7.960	5.2236	8.7301	1.4764	0.7701
10	45	5.590	8.480	5.8807	8.0535	0.2907	0.4265
11	46	4.550	6.960	5.0208	6.6984	0.4708	0.2616
12	47	4.530	6.980	4.8003	6.9558	0.2703	0.0242
13	48	4.970	6.630	4.9427	6.6072	0.0273	0.0228
14	49	6.440	7.720	6.8496	7.0862	0.4096	0.6338
15	50	6.520	6.730	6.9146	6.6625	0.3946	0.0675
16	51	6.540	6.450	4.3453	6.6933	2.1947	0.2433
17	52	6.710	6.680	5.7913	6.4232	0.9187	0.2568
18	53	6.690	7.570	7.4746	7.0281	0.7845	0.5419
19	54	6.790	6.870	6.9099	6.6415	0.1199	0.2285
20	55	6.710	6.650	6.7043	6.6977	0.0057	0.0477
21	56	6.870	7.120	6.9996	6.6515	0.1296	0.4685
22	57	7.030	7.530	7.5474	7.2970	0.5174	0.2330
23	58	6.980	6.780	7.0607	6.7669	0.0807	0.0131
24	59	5.800	6.980	7.0071	6.8356	1.2071	0.1444
25	60	6.660	6.370	5.4246	6.5479	1.2354	0.1779
26	61	6.720	7.070	6.6052	7.1918	0.1148	0.1218
27	62	6.570	5.880	7.4969	6.0867	0.9269	0.2067
28	63	8.700	6.360	6.4421	4.6311	2.2579	1.7289
29	64	6.520	6.460	5.9910	6.3281	0.5290	0.1319
30	65	6.240	6.270	7.8581	6.3517	1.6181	0.0817
31	66	5.000	6.670	8.2423	6.0290	3.2423	0.6410
32	67	6.160	6.790	6.9014	7.1973	0.7414	0.4073
33	68	6.830	7.540	6.8411	6.6210	0.0111	0.9190
34	69	7.140	7.260	6.4149	7.3575	0.7251	0.0975
35	70	6.510	6.050	5.7899	7.4083	0.7201	1.3583

ISSN 2348-6201

Table 4. Descriptors values by stepwise MLR model for correlation 0.4 to up.

Molecule	BLI	RDF090m	G(N..N)	MATS5v	RDF035u	HATS1e
1	0.770	0.000	0.000	-0.777	8.632	0.309
2	0.814	0.000	0.000	-0.618	9.205	0.307
3	0.738	0.000	0.000	-0.644	7.425	0.316
4	0.720	0.007	0.000	-0.449	6.741	0.299
5	0.745	2.750	12.084	-0.556	7.876	0.306
6	0.744	3.562	0.000	-0.378	9.644	0.274
7	0.744	0.000	0.000	-0.316	9.661	0.267
8	1.031	3.195	0.000	-0.311	8.252	0.239
9	1.039	2.008	0.000	-0.288	11.449	0.228
10	1.039	1.478	0.000	-0.243	10.103	0.182
11	0.798	2.394	0.000	-0.172	14.556	0.236
12	0.814	2.844	0.000	-0.159	15.661	0.200
13	0.814	1.160	0.000	-0.140	16.225	0.209
14	0.854	0.000	0.000	0.000	6.172	0.170
15	0.893	0.000	0.000	0.250	6.745	0.242
16	0.835	0.000	0.000	-1.000	5.578	0.281
17	0.940	0.000	0.000	0.440	6.607	0.175
18	0.907	0.000	0.000	-0.375	3.915	0.380
19	0.938	0.000	0.000	-0.083	6.149	0.241
20	0.885	0.000	0.000	-0.656	5.514	0.274
21	0.981	0.000	0.000	0.047	5.985	0.173
22	1.174	0.009	0.000	0.067	5.510	0.262
23	0.929	0.000	0.000	-0.556	5.455	0.271
24	0.897	0.000	0.000	-0.452	5.511	0.305
25	1.009	0.000	0.000	-0.348	5.529	0.298
26	1.009	0.000	0.000	-0.282	5.448	0.304
27	1.110	0.000	0.000	-0.257	5.385	0.297
28	1.199	0.000	0.000	-0.066	5.322	0.295
29	0.818	1.188	0.000	-0.477	6.351	0.257
30	0.689	0.001	22.089	-0.283	12.888	0.320
31	1.117	10.969	0.000	-0.091	9.635	0.324
32	1.131	0.032	0.000	0.892	6.680	0.200
33	1.286	0.035	0.000	0.396	6.737	0.195
34	1.154	2.445	11.081	0.824	8.702	0.223
35	1.292	8.763	11.079	0.231	8.799	0.219

Table 4. Continued.

Molecule	BLI	RDF090m	G(N..N)	MATS5v	RDF035u	HATS1e
36	1.212	4.858	11.114	0.845	11.788	0.213
37	1.336	10.554	11.112	0.225	11.840	0.210
38	1.172	3.712	12.559	-0.032	11.161	0.188
39	1.303	7.388	12.958	-0.181	13.479	0.202
40	1.181	2.789	11.059	0.834	13.516	0.191
41	1.294	8.541	11.058	0.222	13.571	0.187
42	1.124	3.391	11.090	0.433	8.631	0.212
43	1.237	9.326	11.085	0.161	8.667	0.208
44	1.017	3.645	19.861	-0.041	11.425	0.147
45	1.117	4.472	19.861	-0.086	11.394	0.145
46	0.878	0.000	6.546	-0.331	7.230	0.228
47	0.910	0.001	6.546	-0.189	10.063	0.214
48	0.921	0.023	6.546	-0.109	14.653	0.187
49	0.935	0.000	0.000	-0.289	9.782	0.203
50	0.878	0.248	6.546	-0.409	7.814	0.230
51	0.910	0.248	6.546	-0.233	10.055	0.217
52	0.921	0.047	6.546	-0.137	12.731	0.188
53	0.935	0.091	0.000	-0.332	9.791	0.206
54	0.914	0.248	6.546	-0.325	7.751	0.228
55	0.941	0.248	6.546	-0.209	9.992	0.216
56	0.950	0.265	6.681	-0.138	10.763	0.226
57	0.964	0.006	0.000	-0.288	9.729	0.206
58	0.838	0.000	6.546	-0.455	7.878	0.229
59	0.875	0.086	6.760	-0.218	7.763	0.199
60	0.889	0.094	6.678	-0.103	12.055	0.223
61	0.904	0.085	0.000	-0.346	9.839	0.203
62	0.841	0.134	0.000	-0.706	9.794	0.226
63	0.876	0.039	0.000	-0.635	9.465	0.223
64	0.876	0.261	0.000	-0.569	9.531	0.221
65	0.908	0.007	0.000	-0.531	11.92	0.269
66	1.053	0.191	10.968	0.390	9.777	0.241
67	1.074	0.217	10.968	0.397	12.720	0.229
68	1.076	0.077	10.968	0.398	15.368	0.191
69	1.109	0.411	10.968	0.398	14.421	0.212
70	1.144	0.106	10.968	0.392	14.087	0.159

CONCLUSION

In our study, the linear methods were used to select the most significant descriptors. The GA-stepwise MLR and GA-ANN were used to construct a quantitative relation between the activities of benzimidazole derivatives and their calculated descriptors. GA-ANN has been successfully used for finding a QSAR model for benzimidazole derivatives. It provides the best results in comparison with other studied methods. The results of the Jack-Knife, compounds No. 20, 33, 58, 48 and 47 have the smallest difference between the observed and predicted values exist and are proposed for drug design.

ACKNOWLEDGMENTS

The authors thank the Research vice Presidency of Islamic Azad University, Rasht Branch, for their encouragement, permission, and financial support.

REFERENCES

[1] Workowski KA., Berman S. MMWR Recomm Rep., 59: 1-110 (2010).

[2] Sutton M., Sternberg M., Koumans EH., McQuillan G., Berman S., Markowitz L., Clin Infect Dis., 45: 1319-1326 (2007).

[3] Barker HA., Smyth RD., Weissbach H., Toohey JI., Ladd JN., Volcani BE., Journal of Biological Chemistry., 235: 480-488 (1960).

[4] Demchuk E., Ruiz P., Chou S., Fowler B.A., SAR/QSAR methods in public health practice., Toxicol. Appl. Pharmacol., 254: 192–197 (2011).

[5] Tsakovska I., Pajeva I., Alov P., Worth A., Recent advances in the molecular modeling of estrogen receptor-mediated toxicity., Adv. Protein Chem. Struct. Biol., 85: 217–251 (2011).

[6] Prasanna S., Manivannan E., Chaturvedi S.C., Bioorg. Med. Chem. Lett., 14: 4005 (2004).

[7] Prasanna S., Manivannan E., Chaturvedi S.C., QSAR Comb. Sci., 23: 621 (2004).

[8] Prasanna S., Manivannan E., Chaturvedi S.C., Arch. Pharm. Pharm. Med. Chem., 337: 440 (2004).

[9] Prasanna S., Manivannan E., Chaturvedi S.C.. Bioorg. Med. Chem. Lett., 15: 313 (2005).

[10] Prasanna S., Manivannan E., Chaturvedi S.C., Bioorg. Med. Chem. Lett., 15: 2097 (2005).

[11] Prasanna S., Manivannan E., Chaturvedi S.C., Indian J. Chem. B., 43: 2249 (2004).

[12] Manivannan E., Prasanna S., Chaturvedi S.C., Indian J. Biochem. Biophys., 41: 179 (2014).

[13] He L., Jurs P.C., J. Molecular Graphics and Modelling., 23: 503–523 (2005).

[14] Eldred D. V., Weikel C. L., Jurs P.C., Kaiser K.L.E., Chem. Res. Toxicol., 12: 670–678 (1999).

[15] Pérez-Villanueva J.L., Medina-Franco J.R., Caulfield Th., Hernández-Campos A., Hernández-Luis F., Yépez-Mulia L., Castillo R., European Journal of Medicinal Chemistry., 46: 3499-3508 (2011).

Author' biography with Photo

Alireza Nemati Rashtehroodi received his BSc degree in Chemistry in 2009, MSc in Organic Chemistry in 2012 from Payame Noor University, Sari, Iran and Master's thesis did in Chemometrics field from Payame Noor University, Sari, Iran in 2012 under supervision of Dr. Ghasem Ghasemi and Dr. Sattar Arshadi. His research fields include: chemometrics (Quantitative Structure - Activity Relationships) and nanothechnology in Computational chemistry, study of drug delivery.

Education

Payame Noor University, Sari Branch

Master of Science (M.Sc), organic Chemistry, A

2010 – 2012

Activities and Societies: Member of the University Research Committee(2005-2009)/ Member of journal of "Synthesis" as an editor and writer (2005-2009)/ Chairman of the Executive Committee Student Student Regional Conference of Chemistry (2008)/ President of the Scientific Society of Chemistry (2008)/ Member of Iranian Chemical Society (2013-now)/ Member ofYoung Researchers Club and Elite (2011-now)/ Member of Quality Control Officials Association of Guilan (2013- now)/ Member of Society Food, Cosmetics province of Guilan (2013-now).

Yours sincerely,
Alireza Nemati Rashtehroodi