Int. J. Bio-Inorg. Hybr. Nanomater., 11(4): 163-171, Winter 2022

Linear QSAR study of Sulfonamide drugs using by imperialist competitive algorithm

M. Kia[1*]

[1]*Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran*

Received: 21 May 2022; Accepted: 24 July 2022

ABSTRACT: Multiple linear regression (MLR) as modeling tools and Imperialist Competitive Algorithm (ICA) as optimization techniques were employed to choose the best set of descriptors for linear -log(IC_{50}) (the empirical negative logarithm half maximal inhibitory concentration) prediction of the Sulfonamide derivatives. A high predictive ability was observed for the MLR-ICA model with the best number of empires/ imperialists (nEmp=20) and (nEmp=60) with root mean sum square errors (RMSE) of 0.03375 and 0.036665 in gas phase and in solvent, respectively. The results obtained using the MLR-ICA method indicated that the activity of the derivatives of Sulfonamide depends on different parameters such as HVcpx, RDF090u, E1v, Wap, R5e, Mor15v, MPC08, RDF115p descriptors in the gas phase and including RDCHI,MATS1v, RDF115v, RDF080v, D/Dr06, piPC05,BEHp6 and G3m descriptors in the solvent phase. It was concluded that simultaneous utilization of MLR-ICA method can lead to a more comprehensive understanding of the relation between physico-chemical, structural or theoretical molecular descriptors of drugs to their biological activities and facilitate designing of new drugs.

Keywords: *ICA Algorithm, Monte Carlo Method, QSAR, Sulfonamide drugs*

INTRODUCTION

Sulfonamide drugs were the first antibiotics to be used systemically, and paved the way for the antibiotic revolution in medicine and Sulfonamides of derivatives present in inhibitors of carbonic anhydrase [1-5], anticancer [6] and anti-inflammatory agents [7]. Quantitative structure activity relationship (QSAR) [8-11] is one of the most efficacious approaches for designing new chemical identities and understanding the action mechanisms of drugs. QSAR have several variable selection models including multiple linear regression (MLR), genetic algorithm (GA), simulated annealing algorithm (SA) etc. [12]. Imperialist Competitive Algorithm (ICA) is a new population-based optimization algorithm that has recently been introduced for dealing with different kinds of optimization problem [11-14]. The total power of an empire depends on both the power of its colonies and power of the imperialist country because the ICA starts with an initial population called countries and most powerful countries are selected as imperialists and the rest form the colonies of these imperialists and most powerful empires tend to increase their power while weak empires collapse. All empires try to take possession of colonies of other empires and

(*) Corresponding Author - e-mail: kia@iaurasht.ac.ir

control them [13, 15 and 16]. Imperialist competitive algorithm (ICA) is a new population based optimization algorithm [13-19] that starts with an initial population called countries that they are divided into two types: colonies and imperialists that all together form some empires. The total power of an empire depends on both the power of its colonies and power of the imperialist country because the ICA starts with an initial population called countries and most powerful countries are selected as imperialists and the rest form the colonies of these imperialists and most powerful empires tend to increase their power while weak empires collapse. All empires try to take possession of colonies of other empires and control them [15]. In the present study, multiple linear regressions as linear modeling tool and Imperialist Competitive Algorithm as optimization method was applied to investigate the QSAR in some Sulfonamide anticancer drugs.

COMPUTAIONAL METHODS

Details of geometry optimizations of compounds were given in our previous work. As it was described, B3lyp/6-31g by Gaussian 09W [20] was utilized to optimize the geometries of 34 Sulfonamide anticancer drugs and Dragon program [21] was used for calculation of 3226 molecular descriptors for each of the 34 compounds [21]. Modeling and optimizing calculations were carried out using Matlab. 2014a [14]. SPSS [17] program was used to reduce the number of descriptors to 1047 and 1110 descriptors in the gas and solvent phases, respectively [18]. A stepwise multiple linear regression procedure based on the stepwise method was used for the inclusion or rejection of descriptors in the screened models and the final selected 8 descriptors in the gas and solvent phase. The numbers of the most effective descriptors (8 descriptors) chosen by a stepwise multiple linear regression procedure in our previous work was used as a basis for the number of descriptors in this work. The employed ICA algorithm generated random points (matrix indices of descriptors) called the initial Countries that are the counterpart of Chromosomes in GA and it is an array of values of a candidate solution of optimization problem. Empires are sub-populations of

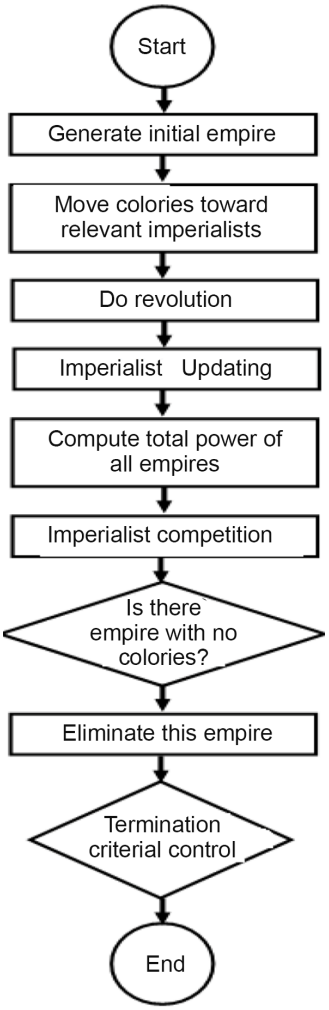
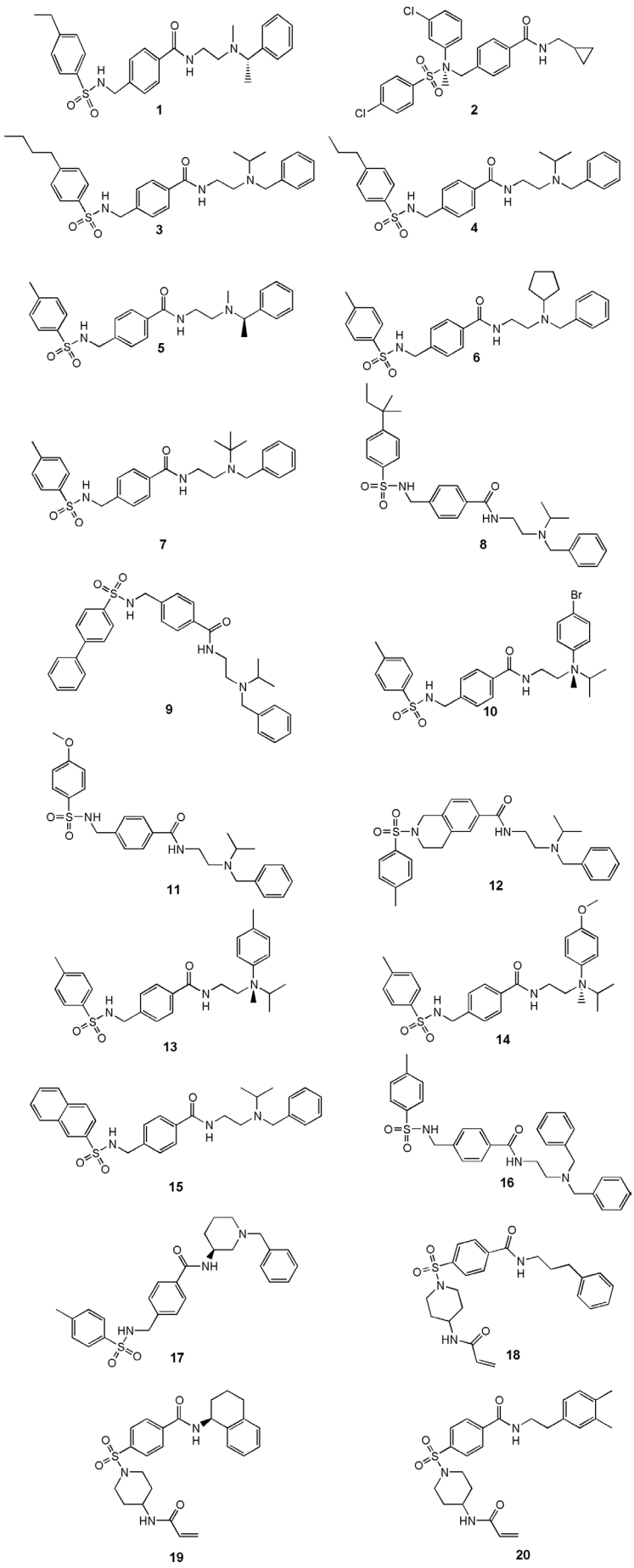


Fig. 1. Flowchart of the Imperialist Competitive Algorithm.

countries. Assimilation, which can be regarded as a primitive form of Particle Swarm Optimization [19] moves all non-best countries (called colonies) in an empire toward the best country (called imperialist) in the same empire to find the colonies with lowest error (RMSE of predicted $-\log(IC_{50})$ using MLR versus empirical values). Fig. 1 shows the imperialist competitive algorithm.

In ICA-MLR method, respectively 1047 and 1110 descriptors in the gas and solvent phases were considered as the population matrix. In ICA-MLR method the number of decision variables (nVar) and number of empires/ imperialists (nEmp) were considered 1 to 8 in the gas and solvent phase and 10, 20, 30 respectively. The maximum numbers of iterations were considered 500 and Population Size and Upper Bound of Variables were considered 1047 in gas phase and 1110 in solvent phase


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

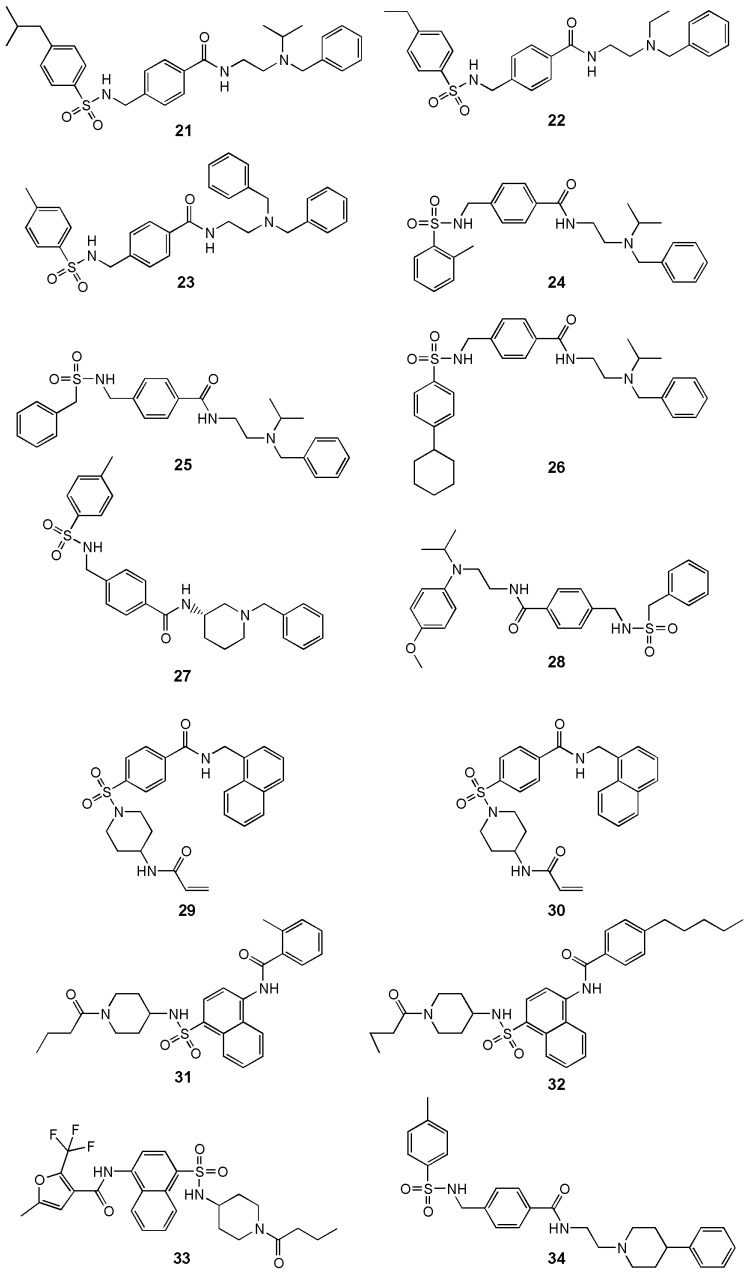


Fig. 2. Optimized structure of the compounds used to build QSAR models with B3lyp/6-31g in gas phase [18].

RESULTS AND DISCUSSION

All studied Sulfonamide derivatives [18] are presented in Fig. 2.

As a first trial, 1000 number of iterations were done to find the most powerful empires and, subsequently, the best descriptors. A plot of the best cost values versus the number of iterations is represented in Fig. 3. It implies that there is no variation in the best cost (MSE) after about 300 iterations. However, in order to ensure that the best descriptors are captured, the number of iterations for the rest of computations was set to 500.

In order to check the robustness of the selected descriptors by varying the model parameters, the effects of number of empires on them were studied by changing the number of empires from 10 to 20 and 30. For choosing the most suitable number of empires, the model was run using different number of empires and 1 to 8 number of descriptors (according to our previous work [18]). The effects of number of selected descriptors on the chosen descriptors and the prediction quality (according to R2 and RMSE) were inves-

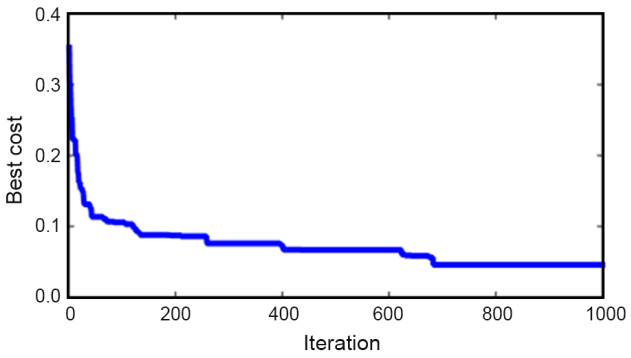


Fig. 3. Plot between Best Cost values compared to the variation of iteration.

tigated for each of the gas and solution phases. The obtained results for the gas phase are given en Table 1. In the ICA-MLR model with Max.It=500, nEmp=20, nVar=8 in gas phase and Max.It=500, nEmp=30, nVar=8 in solvent phase the RMSE and R-square were calculated as 0.0337 and 0.9618 in the gas phase and 0.0454 and 0.9486 in the solvent phase for the predicted activity, respectively (Table 1). The statistical parameters of ICA-MLR approaches with increasing the nEmp 40 to 100 in gas phase and solvent phase with MaxIt=500 are shown in Tables 2,3.

Tables 2, 3 show that in the MLR-ICA model with Max.It=1000, nEmp=60, nVar=8, the RMSE and R-square were calculated as 0.0366 and 0.9585 in the solvent phase for the predicted activity, respectively, therefore, MLR-ICA model with nEmp=20 in gas phase and nEmp=60 in solvent phase were better than the other models and as such, the descriptors used in this model were important (Tables 2, 3). The plot showing the variation of observed versus predicted – $\log IC_{50}$ values are shown in Figs. 4 and 5 in gas and solvent phase, respectively.

Table 1. Statistical parameters of ICA-MLR models in gas and solvent phase (Max.It=500)

nVar- nEmp	Predicted (Gas phase)		Predicted (Solvent phase)	
	R^2	RMSE	R^2	RSME
1-10	0.4582	0.478379	0.4582	0.478379
2-10	0.5894	0.362595	0.5999	0.353327
3-10	0.6722	0.289485	0.7184	0.248667
4-10	0.72732	0.213919	0.8103	0.167473
5-10	0.842	0.139493	0.859	0.124484
6-10	0.8694	0.11535	0.9011	0.087334
7-10	0.8928	0.094652	0.9168	0.073453
8-10	0.9221	0.068807	0.9521	0.042252
1-20	0.4582	0.478379	0.4582	0.478379
2-20	0.5894	0.362595	0.5999	0.353327
3-20	0.6667	0.29432	0.7184	0.248667
4-20	0.7509	0.219933	0.8103	0.167473
5-20	0.8087	0.168963	0.859	0.124484
6-20	0.8996	0.088661	0.9011	0.087334
7-20	0.9114	0.078228	0.9243	0.066832
8-20	0.9618	0.03375	0.9578	0.037251
1-30	0.4582	0.478379	0.4582	0.478379
2-30	0.5894	0.362595	0.5999	0.353327
3-30	0.7116	0.254672	0.7184	0.248667
4-30	0.7629	0.209349	0.8103	0.167473
5-30	0.8335	0.147023	0.8476	0.134598
6-30	0.8471	0.135002	0.8888	0.098205
7-30	0.907	0.082093	0.9359	0.056634
8-30	0.9315	0.060523	0.9486	0.045403

Table 2. Statistical parameters of ICA-MLR models in gas phase with different nEmp (Max.It=500)

nVar- nEmp	Predicted (Gas phase)	
	R^2	RMSE
8-40	0.9367	0.055914
8-50	0.9271	0.064397
8-60	0.9157	0.074458
8-70	0.9123	0.077426
8-80	0.9417	0.051521
8-90	0.9397	0.053252
8-100	0.9397	0.053252

Table 3. Statistical parameters of ICA-MLR models in solvent phase with different nEmp (Max.It=500).

nVar- nEmp	Predicted (Solvent phase)	
	R^2	RMSE
8-40	0.9545	0.040217
8-50	0.9327	0.059386
8-60	0.9585	0.036665
8-70	0.9537	0.040923
8-80	0.957	0.037977
8-90	0.9405	0.052518
8-100	0.9544	0.040242

The plot of the variation of empirical in comparison with predicted $-\log IC_{50}$ values has been shown in Figs. 4 and 5 for gas and solvent phase respectively. As it is understood by these Figs. the developed model has acceptable accuracy for both phases.

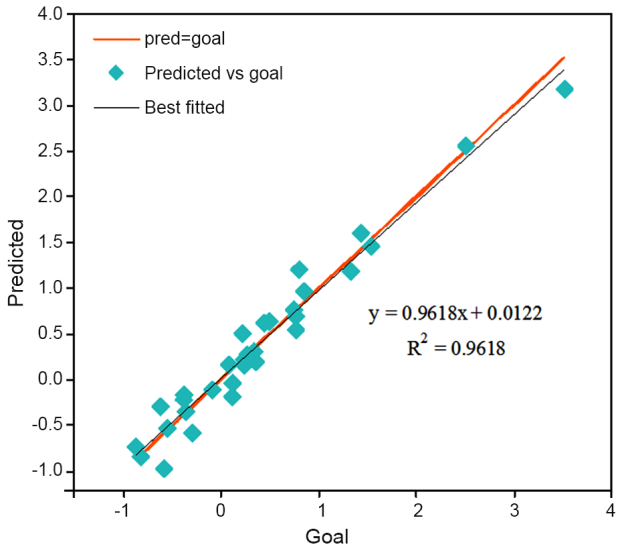


Fig. 4. Plot between predicted values versus Goal with nVar=8 and nEmp=20, in gas phase.

The last chosen descriptors with nEmp=20 in gas phase and with nEmp=60 in solvent phase and 8 descriptors have been presented in Tables 4, 5.

The presented information in these two tables show that RDF115p (weighted by atomic polarizabilities, E1v (weighted by atomic van der Waals volumes), R5e(weighted by atomic Sanderson electronegativities), Reciprocal distance Randic-type index and BEHp6(weighted by atomic masses) in gas phase and RDCHI descriptor (connectivity indices), weighted by atomic van der Waals volumes, atomic Sanderson electronegativities [29] in solvent phase are the most

Table 4. The best selected descriptors using ICA-MLR Method in gas phase (Max.It=500, nEmp=20).

Descriptor	Definition	Type
HVcpx	Graph vertex complexity index	Information indices
RDF090u	Radial Distribution Function-9.0/ unweighted	RDF descriptors
E1v	1st component accessibility directional WHIM index /weighted by atomic van der Waals volumes	WHIM descriptors
Wap	All-path winer index	Topological descriptors
R5e	R autocorrelation of lag 5/ weighted by atomic Sanderson electronegativities	GETAWAY descriptors
Mor15v	3D-MoRSE-signal 15.0/weighted by atomic van der Waals volumes	3D-MoRSE descriptors
MPC08	Molecular path count of order 8	Walk and path counts
RDF115p	Radial Distribution Function-11.5/ weighted by atomic polarizabilities	RDF descriptors

Table 5. The best selected descriptors using MLR-ICA Method with nVar=8 and nEmp=60 in solvent phase.

Descriptor	Definition	Type
RDCHI	Reciprocal distance Randic-type index	Connectivity indices
MATs1v	Morau autocorrelation lag 1/weighted by atomic van der Waals volumes	2D-autocorrelation
RDF115v	Radial Distribution Function-11.5/ weighted by atomic van der Waals volumes	RDF descriptors
RDF080v	Radial Distribution Function-8.0/ weighted by atomic van der Waals volumes	RDF descriptors
D/Dr06	Distance/ detour ring index of order	Topological descriptors
PiPC05	Molecular multiple path count of order 5	Walk and path count
BEHp6	Highest eigenvalue n.6 of Burden matrix/weighted by atomic masses	Burden eigenvalues
G3m	3st component symmetry directional WHIM index/ weighted by atomic Sanderson electronegativities	WHIM descriptors

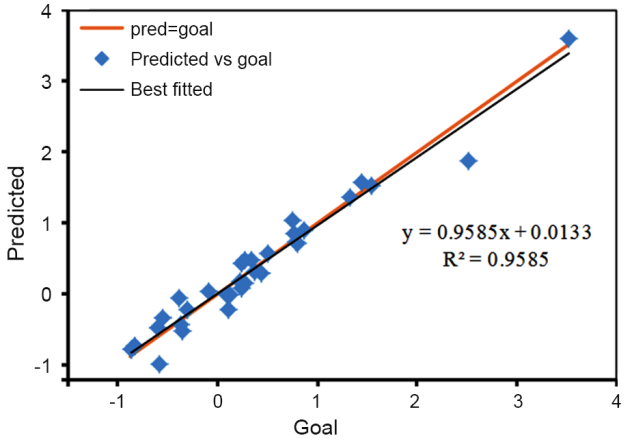


Fig. 5. Plot between predicted values versus Goal with nVar=8 and nEmp=60, in solvent phase.

important descriptors for designing this class of drugs.

CONCLUSION

In this study, MLR-ICA approach was employed to study the structure-activity relationships of 34 Sulfonamide Anticancer Drugs. The best descriptors with nEmp= 20 in gas phase and with nEmp= 60 in solvent phase were more significant than other descriptors. These results proved that HVcpx, RDF090u, E1v, Wap, R5e, Mor15v, MPC08 and RDF115p descriptors in the gas phase and RDCHI, MATS1v, RDF115v, RDF080v, D/Dr06, piPC05,BEHp6 and G3m descriptors in the solvent phase were more significant than other descriptors in building QSAR model and predicting biological activity of Sulfonamide substitution patterns. These descriptors are shown that for designing this class of drugs atomic van der Waals volumes, atomic Sanderson electronegativities, and atomic polarizabilities in gas phase and atomic van der Waals volumes, atomic Sanderson electronegativities, in the solvent phase are important. Also the obtained results can be employed for designing new anti-cancer drugs.

ACKNOWLEDGEMENTS

The support of the Islamic Azad University, Rasht Branch is gratefully acknowledged.

REFERENCES

[1] Zimmerman, S., Innocenti, A., Casini, A., Ferry, J.G., Scozzafava. A., Supuran, C.T. (2004), Carbonic anhydrase inhibitors. Inhibition of the prokariotic beta and gamma-class enzymes from Archaea with sulfonamide. Bioorg. Med. Chem. Lett., 14, 6001-6006.

[2] Garaj, V., Puccetti, L., Fasolis, G., Winum, J-Y., Montero, J-L., Scozzafava, A., Vullo, D., Inno-

centia. A., Supurana, C.T. (2004) Carbonic anhydrase inhibitors: synthesis and inhibition of cytosolic/tumor-associated carbonic anhydrase isozymes I, II, and IX with sulfonamides incorporating 1,2,4-triazine moieties. Bioorg. Med. Chem. Lett,, 14, 5427-5433.

[3] Puccetti, L., Fasolis, G., Vullo, D., Chohan, Z.H., Scozzafavab. A., Supuranb. C.T. (2005). Carbonic anhydrase inhibitors. Inhibition of cytosolic/ tumor-associated carbonic anhydrase isozymes I, II, IX, and XII with Schiff's bases incorporating chromone and aromatic sulfonamide moieties and their zinc complexes. Bioorg. Med. Chem. Lett., 15, 3096-3101.

[4] Lehtonen, J.M., Parkkila, S., Vullo, D., Casini, A., Scozzafavac, A. and Supuranc, C.T. (2004). Carbonic anhydrase inhibitors, Inhibition of cytosolic isozyme XIII with aromatic and heterocyclic sulfonamides: a novel target for the drug design. Bioorg. Med. Chem. Lett., 14, 3757-3762.

[5] Güzel, O., Innocenti, Scozzafava, A., Salman. A., Supuran, C.T. (2009) Carbonic anhydrase inhibitors. Phenacetyl-, pyridylacetyl- and thienylacetyl-substituted aromatic sulfonamides act as potent and selective isoform VII inhibitors. Bioorg. Med. Chem. Lett., 19, 3170-3173.

[6] Scozzafava, A., Owa, T., Mastrolorenzo. A., Supuran, C.T. (2003). Anticancer and antiviral sulfonamide. Curr. Med. Chem., 10, 925-953.

[7] Weber, A., Casini, A., Heine, A., Kuhn, D., Supuran, C.T. (2004). Unexpected Nanomolar Inhibition of Carbonic Anhydrase by COX-2-Selective Celecoxib: New Pharmacological Opportunities Due to Related Binding Site Recognition. J. Med. Chem. 47, 550-557.

[8] Liao, S.Y., Chen, J.C., Qian, L., Shen, Y. (2008). QSAR studies and molecular design of phenanthrene-based tylophorine derivatives with anticancer activity. QSAR Comb. Sci., 27, 280–288.

[9] Wang, X., Yang, W., Xu, X., Zhang, H., Wang, Y.S. (2010). Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. Curr. Med. Chem., 17, 2788-803.

[10] Yang, G-F., Huang, X. (2006). Development of Quantitative Structure-Activity Relationships and Its Application in Rational Drug Design. Current Pharm. Des., 12, 4601-4611.

[11] Srivastava, H.K., Pasha, F.A., Singh, P.P. (2005). Atomic Softness-Based QSAR Study of Testosterone. Int. J. Quantum. Chem., 103, 237.

[12] Consonni, V., Todeschini, R., Pavan, M., Gramatica, P. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. J. Chem. Inf. Compute Sci., 42, 693-705.

[13] Atashpaz-Gargari, E. and Lucas, C. (2007). Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: IEEE Congress on Evolutionary Computation, 4661–4667.

[14] Hosseini, S., Al Khaled, A. (2014). A survey on the Imperialist Competitive Algorithm metaheuristic: Implementation in engineering domain and directions for future research. Appl. Soft Comput, 24, 1078-1094

[15] Mousavi Rad, S., Akhlaghian Tab, F., Mollazade, K. (2012). Application of imperialist competitive algorithm for feature selection: A case study on bulk rice classification. Int. J. Compute. Appl., 40(16), 41-48.

[16] Sayyadi kord Abadi, R., Alizadehdakhel, A., Dorani Shiraz, S. (2017). Ab Initio and QSAR Study of Several Etoposides as Anticancer Drugs: Solvent Effect. Russ. J. Phys. Chem. B, 11, 307.

[17] Todeschini, R., Consonni, V. (2000). Handbook of Molecular Descriptors, Wiley-VCH.

[18] Sayyadi kord Abadi, R., Alizadehdakhel, A., Tajadodi Paskiabei, S. (2016). A DFT and QSAR Study of Several Sulfonamide Derivatives In Gas and Solvent. J. Korean Chem. Soc., 60, 225-234.

[19] Lin, J.-L., Tsai, Yu-H., Yu, C-Y., Li, M-S. (2012). Interaction Enhanced Imperialist Competitive Algorithms. Algorithms, 5, 433-448.

[20] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G. E., Robb, M.A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J. L. , Hada,M., Ehara, M., Toyota,K., Fukuda, R., Hasegawa, J., Ishida, M.,

Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Peralta, Jr., J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers,E., Kudin,K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M.,. Knox, J. E., Cross, J.B.,Bakken, V.,. Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli,C., Ochterski, J.W., Martin,R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J., Dapprich, S., Daniels,A.D., Farkas,Ö., Foresman, J. B., Ortiz, J.V., Cioslowski, J., Fox,D. J. (2009). Gaussian 09, Gaussian, Inc., Wallingford CT.

[21] Dragon 3.0 Evaluation Version. Available online: http://www.disat.unimib.it/chm.

[22] SayyadikordAbadi, R., Alizadehdakhel, A. (2018). Feature Extraction using Linear and Non-Linear QSAR Study on Several Taxol Derivatives as Anticancer Drugs. Rev. Roum. Chim., 63, 171-180.

AUTHOR (S) BIOSKETCHES

Majid Kia, Assistant Professor, Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran, *E-mail: kia@iaurasht.ac.ir*