Journal of the Korean Chemical Society
2016, *Vol. 60, No. 4*
Printed in the Republic of Korea
http://dx.doi.org/10.5012/jkcs.2016.60.4.225

A DFT and QSAR Study of Several Sulfonamide Derivatives in Gas and Solvent

Robabeh Sayyadi kord Abadi[†,*], Asghar Alizadehdakhel[‡], and Soghra Tajadodi Paskiabei[†]

[†]*Department of Chemistry, Rasht Branch, Islamic Azad University, P.O. Box 41335-3516, Rasht, Iran.*
[*]*E-mail: sayyadi@iaurasht.ac.ir*
[‡]*Department of Chemical Engineering, Islamic Azad University, Rasht Branch, P.O. Box 41335-3516, Rasht, Iran.*
(Received January 15, 2016; Accepted June 2, 2016)

ABSTRACT. The activity of 34 sulfonamide derivatives has been estimated by means of multiple linear regression (MLR), artificial neural network (ANN), simulated annealing (SA) and genetic algorithm (GA) techniques. These models were also utilized to select the most efficient subsets of descriptors in a cross-validation procedure for non-linear -log (IC50) prediction. The results obtained using GA-ANN were compared with MLR-MLR, MLR-ANN, SA-ANN and GA-ANN approaches. A high predictive ability was observed for the MLR-MLR, MLR-ANN, SA-ANN and MLR-GA models, with root mean sum square errors (RMSE) of 0.3958, 0.1006, 0.0359, 0.0326 and 0.0282 in gas phase and 0.2871, 0.0475, 0.0268, 0.0376 and 0.0097 in solvent, respectively (N=34). The results obtained using the GA-ANN method indicated that the activity of derivatives of sulfonamides depends on different parameters including DP03, BID, AAC, RDF035v, JGI9, TIE, R7e+, BELM6 descriptors in gas phase and Mor 32u, ESpm03d, RDF070v, ATS8m, MATS2e and R4p, L1u and R3m in solvent. In conclusion, the comparison of the quality of the ANN with different MLR models showed that ANN has a better predictive ability.

Key words: QSAR, Sulfonamides, Molecular descriptors

INTRODUCTION

The sulfonamide group is considered as a pharmacopoeia which is present in a number of biologically active molecules, particularly in antimicrobial agents.[1-5] It is also present in inhibitors of carbonic anhydrase,[6-10] anticancer[11] and anti-inflammatory agents,[12] which are derivatives of sulfonamides.

Most diseases that involve G-protein receptors in the central nervous system cause abnormal behavior, due to drug addiction to sulfonamides. Recent studies have shown that in regulating other receptors that interact with drug and other substance abuse, the opioid receptors play an important role.[13-16]

One of the most important aspects in chemometrics that provide important information useful for molecular design and medicinal chemistry is the Quantitative structure activity relationship (QSAR).[17-19] QSAR models are mathematical equations that create a relationship between chemical structures and biological activities. The first step in the QSAR study is to find a set of descriptors with higher impact on biological activity.[20-23] In QSAR models, a wide range of descriptors are used, which can be constitutional, geometrical etc.

Several QSAR studies[26,27] have been carried out involving the use of an effective computational method to examine the inhibition mechanism.

In the present study, the multiple linear regressions (MLR) as linear models, and artificial neural networks (ANN), simulated annealing (SA) and genetic algorithm (GA)[21-25] as non-linear models were applied to investigate the QSAR in sulfonamide derivatives. Various QSAR models have been used to select the best descriptors for the important prediction of inhibitory activity of sulfonamide compounds, and then these models were compared.

THEORY AND COMPUTATIONAL METHODS

General methods

The geometric optimizations of sulfonamide compounds were carried out using Gaussian 03W at B3lyp/6-31g.[28] Polarized continuum model (PCM) was applied to consider the non-specific solvent effect, and all molecules were optimized in H_2O solvent.

3226 molecular descriptors in topological, geometrical, MoRSE,[30,31] RDF,[31,32] GETAWAY,[33,23] auto-correlations[34] and WHIM [35,36] groups were calculated using the Dragon program.[29] In three steps, the number of descriptors was reduced through an objective feature selection.

At first, in the dataset of sulfonamide compounds, the descriptors that had the same value of at least 70% were removed, and thereafter, the descriptors with correlation coefficient less than 0.25 with the dependent variable (-log IC50) were considered redundant and removed.[37] After

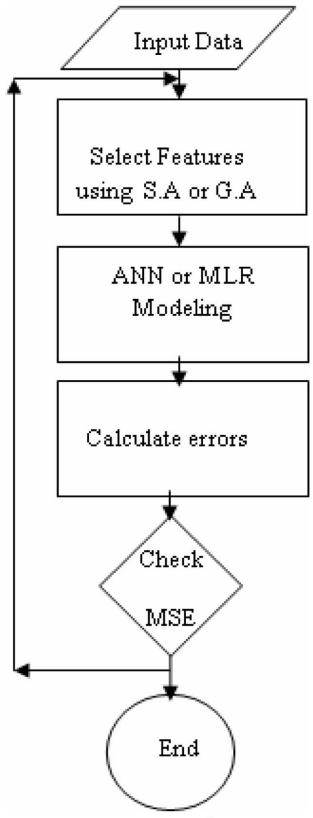

***Figure* 1.** The employed procedure for finding optimum descriptors of the ANN models.

these two steps, the number of descriptors was reduced to 1047 in the gas phase and 1110 in the solvent phase. Stepwise multiple linear regression procedure was used for rejection of descriptors. The QSAR method with high correlation coefficient (R), low standard deviation, least numbers of independent variables, high ability to predict and high F statistic value is an ideal method.[38]

The best subset of descriptors selected in (MLR) was fed into neural networks in the MLR-ANN method. The neural networks used in this study were all three-layer feed-forward network. The networks were trained using the TSET members with Levenberg-Marquart algorithm.[39] In SA-ANN and GA-ANN methods, 1047 and 1110 descriptors in the gas and solvent phase were considered as possible input of the ANN and fed into the input layer of the ANNs in GA-ANN and SA-ANN models (*Fig.* 1). All calculations in the present study were done in Matlab environment (V 7.12, The Mathworks,Inc), SA, GA and Neural Fitting toolbox.

The mean square error of all the models was calculated using the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_o)^2}{n}} \quad (1)$$

where y_i is the desired output, y_o is the predicted value by model, and n is the number of molecules in this study's data set.

RESULTS AND DISCUSSION

Thirty four different sulfonamide derivatives were selected as a sample set, and the geometry of the compounds was optimized using Gaussian 09W at B3LYP/6-31 g. All the optimized Sulfonamide compounds are shown in *Fig.* 2.

Linear and non-linear feature selection methods, such as MLR-MLR (stepwise-MLR), MLR-ANN, SA-ANN, MLR-GA and GA-ANN, were used to select the most significant descriptor.

SPSS[40] software was used for stepwise MLR models as shown in *Table* 10. The RMSE in MLR-MLR for predicted activity was found to be 0.39576 in gas phase and 0.2871 in solvent phase. Also, the correlation coefficient (R^2) calculated for the PSET was 0.8226 in gas phase and 0.90671 in solvent phase.

Table 10 shows that MLR-MLR method is better than other linear methods (MLR-PLS1 and MLR-PCR). The definition of the descriptors in the MLR-MLR method is shown in *Table* 1.

The descriptors, which were selected using the MLR-MLR model were fed into the neural networks to establish the MLR-ANN model. In this model, the RMSE for predicted activity and TSET compounds were found to be 0.1006, 0.0475 and 0.1162, 0.0458 in gas and solvent phase, respectively (*Table* 9).

To establish the SA-ANN, MLR-GA and GA-ANN models, the 1047 and 1110 descriptors in gas and solvent phase were fed into the neural network to select the best descriptors, also 3 neurons in the hidden layer of the GA-ANN model were used in this study (*Fig.* 1).

The descriptors, which were selected using the QSAR models are shown in *Tables* 1-8. These parameters relate the structure to the activity of the optimized compounds.

MATS5e and GATS2p (*Tables* 1 and 2), GATS3e and ATS4v (*Table* 4), ATS8m, and MATS2e (*Table* 8) are 2D autocorrelation descriptors. The 2D-autocorrelation descriptors explain how the values of certain functions, at intervals equal to the lag, are correlated.[41]

EEig0 (*Table* 1), EEig13d (*Tables* 2 and 5) and ESPm03d (*Table* 8) are Edge adjacency indices. The Edge adjacency relationships in molecular graphs have been used to define

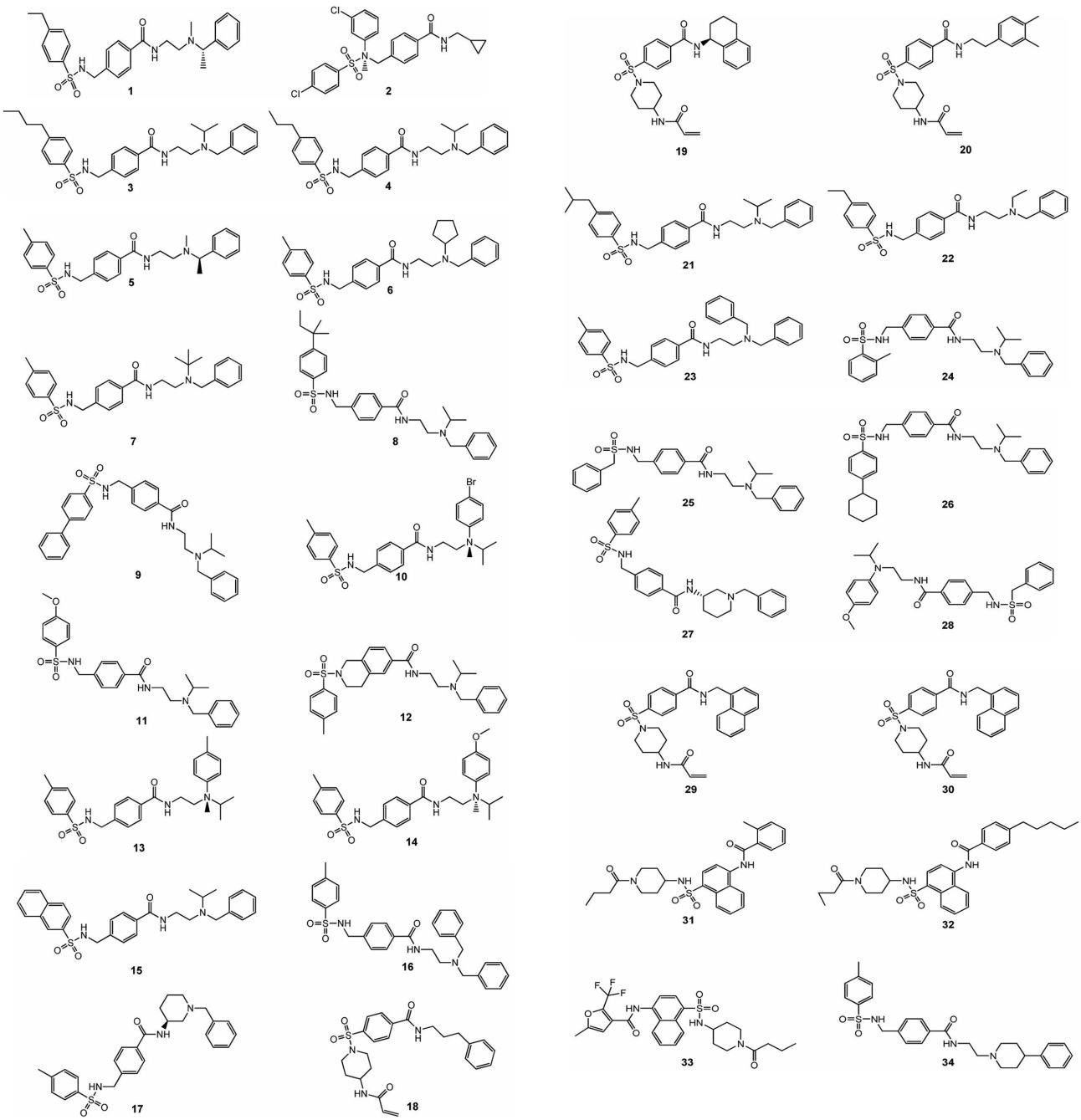

***Figure* 2.** Optimized structure of the compounds used to build QSAR models with B3lyp/6-31g in gas phase.

a new topographic index.[41]

RDF 130p, RDF 115v, RDF095v, RDF035v, RDF070v, and RDF 115p (*Tables* 1, 2, 5, 7, 8, and 6) are RDF descriptors. The radial distribution function (RDF) descriptors are based on the distance distribution in the molecule.[42]

IC5 (*Table* 2) and IC0 (*Table* 5), and AAC (*Table* 7) are information indices. The total information content (I) is obtained by multiplying the mean information content by the number of elements:[43]

G1s and G1v (*Table* 1), L1m (*Table* 3), KM (*Table* 6), L1u (*Table* 6), and TP (*Table* 5) are WHIM descriptors. WHIM descriptors are built in such a way to capture the relevant molecular 3D information regarding molecular size, shape, symmetry and atom distribution with respect to invariant reference frames.[31]

R4e− and R5p− (*Table* 1), R7u+ (*Table* 2) and H6v, RTe,

***Table* 1.** The best selected descriptors using MLR-MLR method in gas phase

Descriptor	Definition	Type
MAT S5e	Moran autocorrelation – Lag 5/weighted by atomic Sanderson electronegativities	2D autocorrelations
GAT S2p	Geary autocorrelation – Lag 2 / weighted by atomic polarizabilities	2D autocorrelations
EEig03r	Eigenvalue 03 from edge adj. matrix weighted by resonance integrals	Edge adjacency indices
RDF130p	Radial distribution function -13.5 / weighted by atomic polarizabilities	RDF descriptors
G1s	1st component symmetry directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors
G1v	1st component symmetry directional WHIM index / weighted by atomic van der Waals	WHIM descriptors
R4e+	R maximal autocorrelation of lag 4 / weighted by atomic van der Waals volumes	GETAWAY descriptors
R5P+	R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities	GETAWAY descriptors

***Table* 2.** The best selected descriptors using MLR-MLR method in solvent phase

Descriptor	Definition	Type
IC5	Information content index (neighborhood symmetry of 5-order)	Information indices
GATS2p	Geary autocorrelation – Lag 2 / weighted by atomic polarizabilities	2D autocorrelations
EEig13d	Eigenvalue 13 from edge adj. matrix weighted by dipole moments	Edge adjacency indices
RDF115v	Radial distribution function -11.5 / weighted by atomic van der Waals	RDF descriptors
Mor08u	3D-MoRSE-Signal 08 / weighted	3D-MoRSE descriptors
Mor17v	3D-MoRSE-Signal 17 / weighted by atomic van der Waals volumes	3D-MoRSE descriptors
Mor23e	3D-MoRSE-Signal 23 / weighted by atomic Sanderson electronegativities	3D-MoRSE descriptors
R7u+	R maximal autocorrelation of lag 7 / unweighted	GETAWAY descriptors

***Table* 3.** The best selected descriptors using SA-ANN method in gas phase

Descriptor	Definition	Type
H6v	H autocorrelation of lag 6 / unweighted	GETAWAY descriptors
GGI6	Topological charge index of order 6	Topological charge indices
RTe	R total index / weighted by atomic Sanderson electronegativities	GETAWAY descriptors
R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors
MPC05	Molecular path count of order 05	Walk and path counts
L1m	1st component size directional WHIM / weighted by atomic masses	WHIM descriptors
F06[C-C]	Frequency of X-X at topological distance 06	2D frequency fingerprints
BEHm6	Highest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	Burden eigenvalues

***Table* 4.** The best selected descriptors using SA-ANN method in solvent phase

Descriptor	Definition	Type
H5m	H autocorrelation of lag 5 / weighted by atomic masses	GETAWAY descriptors
F09[c-o]	Frequency of C-O at topological distance 09	2D frequency fingerprints
VED2	Average eigenvector coefficient sum from distance matrix	Eigenvalue – based indices
GATS3e	Geary autocorrelation – lag 3 / weighted by atomic Sanderson electronegativities	2D autocorrelations
Mor02v	3D-MoRSE – signal 02 / weighted by atomic vander Waals volumes	3D-MoRSE descriptors
BELm	Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses	Burden eigenvalues
Mor17u	3D-MoRSE – signal 17 / unweighted	3D-MoRSE descriptors
ATS4V	Broto-Moreau autocorrelation of a topological structure- lag 4 / weighted by atomic vander Waals volumes	2D autocorrelations

R6u+ (*Table* 3), H5m (*Table* 4), R7e+ (*Table* 7) and R4p, R3m (*Table* 8) are GETAWAY descriptors. GETAWAY (Geometry, Topology, and Atom-Weights Assembly) descriptors encode the geometrical information obtained from the molecular matrix, the topological information obtained from the molecular graph and the information obtained from atomic weights, which are specially designed with the aim of matching the 3D-molecular geometry.[31]

Mor08u, Mor17v, Mor23e (*Table* 2) and Mor02v, Mor17u (*Table* 4) and Mor 17e (*Table* 5), and Mor 32u (*Table* 8)

***Table* 5.** The best selected descriptors using MLR-GA method in gas phase

Descriptor	Definition	Type
EEig13d	Eigenvalue 13 from edge adj. matrix weighted by dipole moments	Edge adjacency indices
Mor17e	3D-MoRSE – signal 17 / unweighted	3D-MoRSE descriptors
TP	T total size index / weighted by atomic polarizabilities	WHIM descriptors
J	Balaban distance connectivity index	Topological descriptors
RDF095v	Radial distribution function – 9.5 / unweighted by atomic van der Waals volumes	RDF descriptors
IC0	Information content (neighborhood symmetry of 0-order)	Information indices
Eig1m	Leading eigenvalue from mass weighted distance matrix	Eigenvalue-based indices
QYYM	QQY COMMA2 value / weighted by atomic masses	Geometrical descriptors

***Table* 6.** The best selected descriptors using MLR-GA method in solvent phase

Descriptor	Definition	Type
KM	K global shape index / weighted by atomic masses	WHIM descriptors
VAR	Variation	Topological descriptors
MWCO3	Molecular walk count of order 03	Walk and path count
AEige	Absolute eigenvalue sum from electro negativity weighted distance matrix	Eigenvalue-based indices
SOK	Kier symmetry index	Topological descriptors
TI2	Second mohar index TI2	Topological descriptors
BEHp6	Highest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities	Burden eigenvalues
RDF115p	Radial distribution function – 11.5 / weighted by atomic polarizabilities	RDF descriptors

***Table* 7.** The best selected descriptors using GA-ANN method in gas phase

Descriptor	Definition	Type
DP03	Molecular profile no. 03	Randic molecular profiles
BID	Balaban ID number	Walk and path counts
AAC	Mean information index on atomic composition	Information indices
RDF035v	Radial distribution function – 3.5 / weighted by atomic polarizabilities	RDF descriptors
JGI9	Mean topological charge index of order 9	Topological charge indices
TIE	E –state topological parameter	Topological descriptors
R7e$^-$	R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors
BELm6	Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	Burden eigenvalues

***Table* 8.** The best selected descriptors using GA-ANN method in solvent phase

Descriptor	Definition	Type
Mor32u	3D-MoRSE – signal 32 / unweighted	3D-MoRSE descriptors
ESpm03d	Spectral momen 03 from edge adj. matrix weighted by dipole moments	Edge adjacency indices
RDF070v	Radial distribution function – 7.0 / weighted by atomic van der Waals volumes	RDF descriptors
ATS8m	Broto-Moreau autocorrelation of a topological structure-lag 8 / weighted atomic masses	2D autocorrelations
MATS2e	Moran autocorrelation- lag 2 / weighted by atomic Sanderson electronegativities	2D autocorrelations
R4p	R autocorrelation of lag 4 / weighted by atomic polarizabilities	GETAWAY descriptors
L1u	1st component size directional index / unweighted	WHIM descriptors
R3m	R autocorrelation of lag 3 / weighted atomic masses	GETAWAY descriptors

are 3D-MoRSE descriptors. The 3D-MoRSE descriptors were obtained through the molecular transformation employed in electron diffraction studies.[43]

GGI6 (*Table* 3), JGI9 (*Table* 7) are topology charge indices. The Topological Charge Indices were proposed to evaluate the charge transfer between pairs of atoms and therefore, the global charge transfer in the molecule.[31]

MPC05 (*Table* 3), MWC03 (*Table* 6), BID (*Table* 7) are walk and path count. The molecular walk count of kth order (MWCk) is the total number of walks of the kth length in

***Table* 9.** Statistical parameters of different nonlinear QSAR models

QSAR Model	Predicted		Train	
	R^2	RSME	R^2	RSME
MLR-ANN (Gas)	0.8925	0.1006	0.8894	0.1162
MLR-ANN (Solvent)	0.9472	0.0475	0.9518	0.0458
SA-ANN (Gas phase)	0.9603	0.0359	0.9565	0.0422
SA-ANN (Solvent)	0.9700	0.0268	0.9824	0.0179
MLR-GA (Gas phase)	0.9633	0.0326	0.9471	0.0343
MLR-GA (Solvent)	0.9587	0.0376	0.9628	0.041
GA-ANN (Gas phase)	0.9716	0.0282	0.9591	0.0331
GA-ANN (Solvent)	0.9894	0.0097	0.9877	0.0117

***Table* 10.** Statistical parameters of different linear QSAR models in gas and solvent phase

QSAR Model	R^2	RMSE
MLR-PLS1 (Gas phase)	0.8224	0.3960
MLR-PCR (Gas phase)	0.8023	0.4178
MLR-MLR (Gas phase)	0.8226	0.3958
MLR-PLS1 (Solvent phase)	0.7953	0.4252
MLR-PCR (Solvent phase)	0.7942	0.4263
MLR-MLR (Solvent phase)	0.9067	0.2871

the hydrogen suppressed molecular graph.[31]

F06[C-C] (*Table* 6) and F09[C-O] are 2D frequency fingerprints descriptors. Fragment descriptors are representations of local atomic environments.[31]

BEHm6 (*Table* 3), Belem (*Table* 4), BEHp6 (*Table* 6), and BELm6 are Burden eigenvalue descriptors. The B matrix has been defined as the number of atoms, bond order between two atoms or the electronegativity of the atoms.[31]

VED2 (*Table* 4), Eig1m (*Table* 5), and Adige (*Table* 6) are eigenvalue based indices descriptors. The Eigenvalue Sum Descriptors are computed from Weighted Distance Matrices of a Hydrogen-depleted Molecular Graph.

QYYM (*Table* 5) and DP03 (*Table* 7) are geometrical and Rancid molecular profiles. The Rancid molecular profile DP_k is derived from the distance distribution moments of the geometric matrix G as the average row sum of its entries raised to the k^{th} power and normalized by the factor k!.[31]

The geometrical variables incorporate information about the magnitude of the displacement between the molecular centroid (center of mass) and the polarizability-field (center of charge).[4]

SOK, TI2 (*Table* 6), J (*Table* 5) and TIE (*Table* 7) are topological descriptors. Topological index mathematically encode information regarding the structure of molecules, which have been depicted as graphs and are often sensitive to size, shape, branching, cyclicity and, to a certain extent, the electronic characteristics of molecules.[31]

***Table* 11.** Observed and predicted values of -logIC50 by using GA- ANN in gas phase

Compounds	Predicted	-Log IC50	Residues
1	0.024777	0.126	0.1012
2	0.314696	0.27	-0.0447
3	0.761097	0.775	0.0139
4	-0.35922	-0.352	0.007
5	-0.52847	-0.38	0.1485
6	0.900297	0.863	-0.0373
7	1.541817	1.542	0.0002
8	0.385464	0.244	-0.1415
9	0.341333	0.365	0.0237
10	0.232471	0.244	0.0125
11	0.137561	0.094	-0.0436
12	2.505598	2.509	0.0035
13	-0.63773	-0.609	0.0287
14	0.270955	0.274	0.0031
15	0.860356	0.796	-0.0643
16	-0.85898	-0.873	-0.014
17	-0.81244	-0.829	-0.0166
18	0.222192	0.114	-0.1082
19	0.488902	0.495	0.006
20	0.693254	0.745	0.0517
21	0.24559	0.226	-0.0196
22	-0.2322	-0.299	-0.0668
23	-0.85898	-0.873	-0.014
24	-0.39159	-0.369	0.0226
25	-0.31709	-0.544	-0.2269
26	-0.06302	-0.086	-0.0229
27	-0.81244	-0.829	-0.0166
28	0.274863	0.345	0.0701
29	0.59767	0.769	0.1713
30	0.434135	0.438	0.004
31	3.517797	3.523	0.005
32	1.425166	1.442	0.0168
33	2.215925	1.329	-0.0887
34	-0.61029	-0.589	0.0213

***Table* 12.** Observed and predicted values of logIC50 by using GA-ANN in solvate phase

Compounds	Predicted	-Log IC50	Residues
1	0.14092	0.126	-0.0149
2	0.269564	0.27	0.0004
3	0.770259	0.775	0.005
4	-0.34609	-0.352	-0.006
5	-0.3915	-0.38	0.0115
6	0.871084	0.863	-0.008
7	1.531976	1.542	0.01
8	0.297004	0.244	-0.053
9	0.383159	0.365	-0.0182
10	0.243234	0.244	0.0008
11	0.095581	0.094	-0.002
12	2.60947	2.509	-0.1005
13	-0.6053	-0.609	-0.004
14	0.235868	0.274	0.0381
15	0.235868	0.796	0.5601
16	-0.88056	-0.873	0.007
17	-0.82805	-0.829	-0.0009
18	0.118651	0.114	-0.005
19	0.492411	0.495	0.003
20	0.744692	0.745	0.0003
21	0.203927	0.226	0.0221
22	-0.30701	-0.299	0.008
23	-0.88056	-0.873	0.008
24	-0.36735	-0.369	-0.002
25	-0.54838	-0.544	0.004
26	0.378371	-0.086	-0.4644
27	-0.82805	-0.829	-0.0009
28	0.357496	0.345	-0.0125
29	0.761015	0.769	0.008
30	0.437997	0.438	0.000003
31	3.309107	3.523	0.2139
32	1.668023	1.442	-0.226
33	1.329267	1.329	-0.0003
34	-0.58374	-0.589	-0.005

The statistical parameters of all QSAR models are shown in *Tables* 9 and 10. In train, a computation of 80% sulfonamid compounds is used. In the GA-ANN model, the RMSE and R-square were calculated as 0.0282 and 0.9716 in gas phase and 0.0097 and 0.9894 in the solvent phase, respectively, therefore, GA-ANN model was better than the other models and as such, only the descriptors used in this model were evaluated in this study. These descriptors are shown in *Tables* 7 and 8. The observed and predicted values of -logIC50 using Matlab program are shown in *Tables* 11 and 12. The plot showing the variation of observed versus predicted -logIC50 values are shown in *Figs.* 3 and 4.

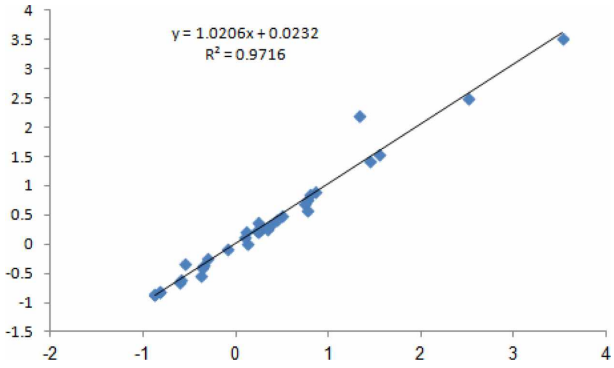

***Figure* 3.** Plot between observed vs predicted -log (/IC50) by using GA-ANN descriptors in gas phase.

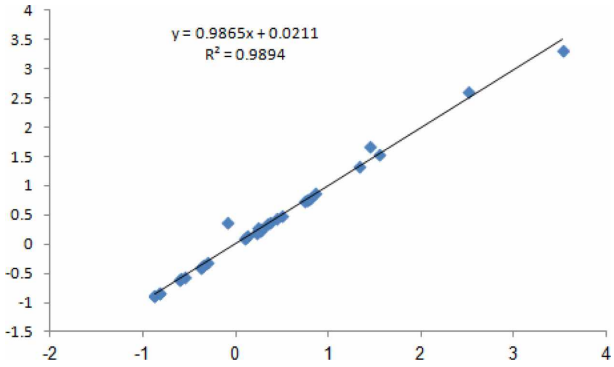

***Figure* 4.** Plot between observed vs predicted -log (IC50) by using GA-ANN descriptors in solvate phas.

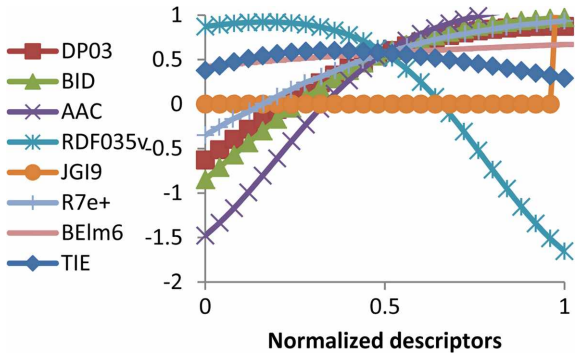

***Figure* 5.** Plot between -log IC50 experimental versus the DP03, BID, AAC, RDF035v, JGI9, TIE, R7e+, and BELm6 normalized descriptors in the gas phase.

The plots of the DP03, BID, AAC, RDF035v, JGI9, TIE, R7e+, and BELm6 descriptors (*Fig.* 5) in the gas phase and Mor 32u, ESPm03d, RDF070v, ATS8m and MATS2e, R4p, L1u, and R3m descriptors in solvent phase (*Fig.* 6) versus the experimental negative logarithm half maximal inhibitory concentration (-logIC50) values were plotted using Excel program. The descriptors values in GA-ANN method in gas and solvent phase were normalized using the equation (2) in Excel program.

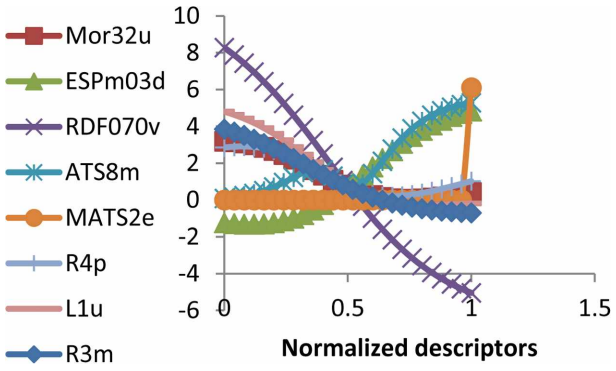

***Figure* 6.** Plot between experimental -log IC50 value versus the Mor32u, ESpm03d, RDF070v, ATS8m, MATS2e, R4p, L1u, and R3m normalized descriptors in the solvent phase.

$$Xnormalized = (Xi-Xmin)/(Xmax-Xmin) \quad (2)$$

The charts in gas phase show that the experimental negative logarithm half maximal inhibitory concentration (-logIC50) value increases with increasing DP03 (Molecular profile no.3), BID (Balaban ID number), R7e+ (weighted by atomic Sanderson electronegativities), and BElm6 (Weighted by atomic masses) descriptors. Thus the half maximal inhibitory concentration (IC50) value is reduced. Therefore, the aforementioned descriptors are the best among the eight descriptors in the gas phase. As the RDF035v (weighted by atomic polarizabilities) descriptor increased, the experimental negative logarithm half maximal inhibitory concentration (-logIC50) value decreased. In JGI9 (topological charge index) descriptor of about 0.8, response do not change. But between 0.8-1 values, an increased experimental negative logarithm half maximal inhibitory concentration rate is shown in the bar chart. As the TIE (E-state topological parameter) descriptor increased up to 0.4, the experimental negative logarithm half maximal inhibitory concentration (-log IC50) value increased and then, the increased TIE (that are sensitive to size, shape, and the electronic characteristics of molecules) descriptor decreased the experimental negative logarithm half maximal inhibitory concentration value. Charts in solvent show that as Mor 32u (indicates that the size of the inhibitor molecule has certain effect on the extent of the interaction between the drug and molecule), RDF070v (weighted by atomic van der waals volumes), R4p (weighted by atomic polarizabilities), L1u (size direction index), and R3m (weighted by atomic masses) descriptors are increased, the experimental negative logarithm half maximal inhibitory concentration (-logIC50) value is reduced. However, with an increase in ATS8m (Broto-Moreau autocorrelation of a topological structure), the amount of experimental negative logarithm half

***Table* 13.** Physico-chemical descriptors in GA-ANN method in gas and solvent phase

Descriptors	Parameters
RDF035v	Molecular polarizability
JGI9	Charge over the atoms in a molecule
ATS8m, BElm6	Weighted by atomic mass
ESpm03d	Dipole moment of the molecule
RDF070v	Van der Waals volumes
MATS2e, R7e+	Weighted by atomic Sanderson electronegativities

***Table* 14.** The common selected descriptors using QSAR methods

Descriptors	QSAR method
R4e+, R5p+	MLR-ANN in gas phase
R7u+	MLR-ANN in solvent
RTe, R6u+	SA-ANN in gas phase
H5m	SA-ANN in solvent
R7e+	GA-ANN IN gas phase
R4p, R3m	GA-ANN in solvent phase

maximal inhibitory concentration is first increased and then reduced, and finally a sharp increase is achieved. In increased ESPm03d (Spectral momen 03 edge adj. matrix weighted by dipole moments), a constant process experimental negative logarithm half maximal inhibitory concentration (-logIC50) is seen, and then subsequently increased. MATS2e (weighted by atomic Sanderson electronegativities) descriptor, which increased the amount of 0.8 changes in the experimental negative logarithm half maximal inhibitory concentration (-logIC50), cannot be seen. But from 0.8 to 1, an increase probe was seen in the experimental negative logarithm half maximal inhibitory concentration (-log IC50) value.

Selected descriptors that are common between all the QSAR methods are shown in *Table* 14. The GETAWAY

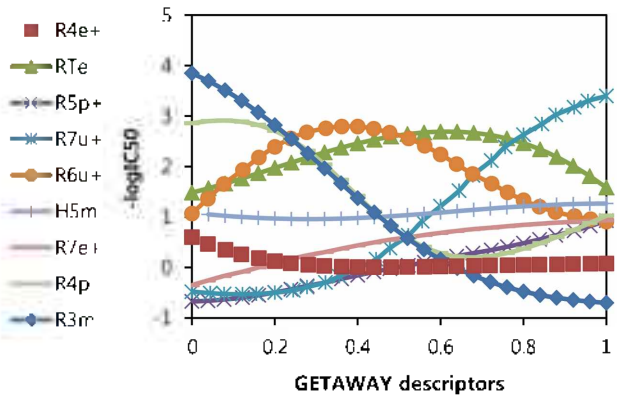

***Figure* 7.** Plot between GETAWAY descriptors versus -log (IC50).

***Table* 15.** Statistical parameter and QSAR model from the previous literatures

Compounds	QSAR model	Statistical parameters
4-benzylidene amino benzene sulphonamide derivatives	PLSR	R^2pred=0.8482
Para-substitued aromatic sulfonamides(pentaparamatric)	ES-SWR algorithm	R^2pred=0.7296
Sulfonamide derivatives	Leave-one-out (LOO) method	R=0.881
Sulfonamide compounds which includes five acetazolamide derivatives	MLR method	R^2=0.94

descriptors played an important role in predicting the -log-IC50 of Sulfonamide compounds. The plots of the GET-AWAY descriptors versus the experimental negative logarithm half maximal inhibitory concentration (-logIC50) values were plotted using Excel program.

Fig. 7 shows that the -logIC50 value increase with increasing R5p, R7u+, H5m, R7e+ descriptors. As the R3m descriptor increased the –log IC50 value decreased.

However, with an increased in RTe and R6u+ descriptors the amount of -logIC50 is first increased and then reduced (Fig. 7). In R4p descriptor the amount of -logIC50 is first decreased and then increased. In increased R4e+, a constant process -logIC50 is seen. R5p, H5m, R7e+, RTe, R4p, R3m are physico-chemical descriptors and These are polarizability, weighted by atomic masses and sanderson electronegativities.

Table 13 shows physico-chemical descriptors in GA-ANN method in gas and solvent phase.the physico-chemical descriptors were found to have an important role in change in activity (*Fig.* 5, 6). These descriptors reduce the half maximal inhibitory concentration (IC50).

Statistical parameter and QSAR model of the sulfonamide compounds from the previous literatures are presented on the *Table* 15.[45–48] It shows that the results of GA-ANN method in this work (*Table* 9) is better than the other QSAR models in previous studies.

CONCLUSION

Among the QSAR models used in this study, the non-linear feature selection models were demonstrated to be better than their linear methods, and the results of GA-ANN method were better than the other non-linear models used. These results also proved that DP03, BID, AAC, RDF035v, JGI9, TIE, R7e$^+$, BELm6 descriptors in the gas phase and Mor32u, ESpm03d, RDF070v, ATS8m, MATS2e, R4p, L1u, R3m descriptors in the solvent phase were more significant than other descriptors in building this QSAR model and predicting the biological activity of Sulfonamides substitution patterns.

Acknowledgment. Publication cost of this paper was supported by the Korean Chemical Society.

REFERENCES

1. Katzung, B. G. In *Basic and Clinical Pharmacology*; 6th ed.: University of California: San Francisco, 1995.
2. Joshi, S.; Khosla, N. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3747.
3. Joshi, S.; Khosla, N.; Tiwari, P. In VitroStudy of Some Medicinally Important Mannich Bases Derived from an Antitubercular Agent. *Bioorg. Med. Chem.* **2004**, *12*, 571.
4. Anand, N. Sulfonamides and Sulfones, In *Burger's Medicinal Chemistry and Drug Discovery*; M. E. Wolff, Ed.: John Wiley & Sons Inc.: New York, **1996**; pp 527.
5. Kamal, A.; Khan, M. N. A.; Reddy, K. S.; Rohini, K.; Sastry, G. N.; Sateesh, B.; Sridhar, B. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5400.
6. Zimmerman, S.; Innocenti, A.; Casini, A.; Ferry, J. G.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 6001.
7. Garaj, V.; Puccetti, L.; Fasolis, G; Winum, J.-Y.; Montero, J.-L.; Scozzafava, A.; Vullo, D.; Innocentia, A.; Supurana, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5427.
8. Puccetti, L.; Fasolis, G.; Vullo, D.; Chohan, Z. H.; Scozzafavab, A.; Supuranb, C. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3096.
9. Lehtonen, J. M.; Parkkila, S.; Vullo, D.; Casini, A.; Scozzafavac, A.; Supuranc, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3757.
10. Güzel, O.; Innocenti, A.; Scozzafava, A.; Salman, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3170.
11. Scozzafava, A.; Owa, T.; Mastrolorenzo, A.; Supuran, C. T. *Curr. Med. Chem.* **2003**, *10*, 925.
12. Weber, A.; Casini, A.; Heine, A.; Kuhn, D.; Supuran, C. T.; Scozzafava, A.; Kiebe, G. *J. Med. Chem.* **2004**, *47*, 550.
13. Pubchem Home Page. https://pubchem.ncbi/nlm.nih.gov (accessed March 2, 2004).
14. Dhawan, B. N.; Cesselin, F.; Raghubir, R.; Reisine, T.; Bradley, P. B.; Portoghese, P. S.; Hamon, M. *Pharmacol. Rev.* **1996**, *48*, 567.
15. Janecka, A.; Fichna, J.; Janecki, T. *Curr. Top. Med. Chem.* **2004**, *4*, 1.
16. Waldhoer, M.; Bartlett, S. E.; Whistler, J. L. *Annu. Rev. Biochem.* **2004**, *73*, 953.
17. Schmidi, H. *Chemom. Intell. Lab. Sys.* **1997**, *37*, 125.
18. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619.
19. Wold, S.; Trygg, J.; Berglund, A.; Antii, H. *Chemom.*

Intell. Lab. Syst. **2001**, *58*, 131.
20. Horvath, D.; Mao, B. *QSAR. Comb. Sci.* **2003**, *22*, 498.
21. Putta, S.; Eksterowicz, J.; Lemmen, C.; Stanton, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1623.
22. Gupta, S.; Singh, M.; Madan, A. K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 272.
23. Consonni, V.; Todechine, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693.
24. Kirkpatrick, S.; Gelatt, Jr. C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
25. Cerný, V. O. *J. Optim. Theory Appl.* **1985**, *45*, 41.
26. Winkler, D. A. *Brief. Bio. Inform.* **2002**, *3*, 73.
27. Guha, R.; Serra, J. R.; Jurs, P. C. *J. Mol. Graph. Model.* **2004**, *23*.
28. DeMelo, E. B.; Ferreira, M. M. *Eur. J. Med. Chem.* **2009**, *44*, 3577.
29. Todeschini, R. Milano Chemometrics and QSAR Research Group. http://www.disat.unimib.it/chem (accessed 2000).
30. Schuur, J. H.; Selzer, P.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334.
31. Todeschini, R.; Consonni, V. *Hand Book of Molecular Descriptors*; *Wiley-VCH.*: 2000.
32. Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vibr. Spectrosc.* **1999**, *19*, 151.
33. Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682.
34. Gramatica, P.; Consonni, V.; Todeschini, R. *Chemosphere* **1999**, *38*, 1371.
35. Gramatica, P.; Consonni, V.; Todeschini, R. *Chemosphere* **2000**, *41*, 763.
36. Fatemi, M. H.; Gharaghani, S. *Bioorg. Med. Chem.* **2007**, *15*, 7746.
37. Jalali-Heravi, M.; Parastar, F. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 147.
38. Levenberg, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics* **1944**, *2*, 164.
39. Horvath, D.; Mao, B. *QSAR. Comb. Sci.* **2003**, *22*, 498.
40. SPSS (Version19). http://www.sps ssc ien ce.com (accessed 2010).
41. Asadollahi, T.; Dadfarnia, S.; Mohammad, A.; Shabani, H.; Ghasemi, J. B. *MATCH Commun. Math. Comput. Chem.* **2014**, *71*, 287.
42. Strand website. www.strandls.com/sarchitect/.../desctheory (accessed Oct 24, 2014).
43. Schuur, J. H.; Selzer, P.; Gasteiger, J. *J. Chem. Inform. Comput. Sci.* **1996**, *36*, 334.
44. Silverman, B. D. *J. Chem. Inform. Comput. Sci.* **2000**, *40*, 1470.
45. Sisodiya, D.; Dashora, K. *Int. J. of Phyto. Pharm.* **2014**, *4*, 153.
46. Melagraki, G.; Afantitis, A.; Sarimveis, H.; Igglessi-Markopoulou, O.; Supura, C. T. *Bioorg. Med. Chem.* **2006**, *14*, 1108.
47. Jaiswal, D.; Karthikeyan, C.; Shirastava, S. K.; Trivedi, P. *Internet Electron. J. Mol. Des.* **2006**, *5*, 345.
48. Eroglu, E.; Turkmen, H.; Guler, S.; Palaz, S.; Oltulu, O. *Int. J. Mol. Sci.* **2007**, *8*, 145.