Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Pharmacy
2(1), 2013, 45-53

ISSN 2321 – 0958

Original Research Article

DEVELOPMENT OF QSAR MODEL FOR STUDYING SULFONAMIDE DERIVATIVES AGAINST CARBONIC ANHYDRASE USING MULTIPLE LINEAR REGRESSION WITH McGOWAN VOLUME AND ALOGP AS MOLECULAR DESCRIPTORS

Ajeet*

Department of Pharmaceutical Chemistry, S. D. College of Pharmacy and Vocational Studies, Bhopa Road, Muzaffarnagar, PIN-251001 U.P., India,

Abstract:
Here sulfonamide analogues have been used to correlate the inhibition constant with the McGowan volume and ALogP descriptors for studying the Quantitative Structure Activity Relationship (QSAR). Correlation may be an adequate predictive model which can help to provide guidance in designing and subsequently yielding greatly specific compounds that may have reduced side effects and improved pharmacological activities. We have used Multiple Linear Regression (MLR) for developing QSAR model. For the validation of the developed QSAR model, statistical analysis such as cross validation test (as internal validation), quality factor, fischers test, root mean square deviation (RMSD), standard deviation, variance, Y-randomization test etc.; have been performed and all the tests validated this QSAR model with fraction of variance r^2 = 0.8138 and LOO-CV q^2 = 0.7887.

Key words- QSAR; Multiple Linear Regression; sulfonamide; carbonic anhydrase.

Introduction:

Epilepsy is a disorder of the brain characterized by recurrent unprovoked seizures. These seizures are symptoms of abnormal neuronal activity in the central nervous system. In spite of the large therapeutic arsenal of old and new antiepileptic drugs (AEDs), approx. 30% of epileptic patients are affected by seizures. In many cases the clinical use of AEDs is prohibited by their side effects. Therefore, a substantial need remains to discover novel molecules for the development of new effective and safer AEDs.[1,2]

While talking about the sulfonamides as anticonvulsants, we found that it should contains bifunctional 5-membered heterocycles composed of a sulfonamide, an amide and 1,3,4-thiazole nucleus for developing a highly potent carbonic anhydrase (CA) such as Acetazolamide, topiramate etc. These contain sulphamate moity which is essential for anhydrase inhibition. The aromatic and heteroaromatic sulfonamides having valproyl

For Correspondence:
ajeet_pharma111@rediffmail.com

Received on: March 2013
Accepted after revision: March 2013

and some other lipophilic moieties have been found to possess anticonvulsant and as well as carbonic anhydrase inhibiting properties.[3]

In the present study, we developed a QSAR model on a series of sulfonamide analogues with respect to their logKi. The QSAR studies are perfect tool for understanding the drug design process in terms of their chemical-pharmacological activity interaction, along with it is also used in toxicology and pesticide research. QSAR studies can focus on mechanism of action of ligands with Human, bacteria, virus, membranes, enzymes etc. It can also be used for the evaluation of the metabolism, absorption, distribution and excretion phenomena. The QSAR methodology comprises of computationally derived descriptors to correlate with pharmacological activities. These descriptors are principally of four types such as electronic, stearic, hydrophobic and topological indices[4]. The descriptors used for developing the QSAR model are McGowan[5] volume and AlogP (Atomic partition coefficient). Rational Drug Design helps to facilitate and fasten the drug designing process, which involves various methods to identify novel compound.[6]

All the bioactivity values and information about 2D structure of sulfonamide analogues were taken from literature.[7] LogKi is a variable that comprises the bioactivity parameter for the QSAR model. In order to calculate the 2D molecular descriptors, PaDEL descriptor software[8], which incorporate CDK library for descriptor calculation has been used after optimitizing the sulfonamide analogues. For the development of QSAR model, Multiple Linear Regression[4] has been employed and all were validated through statistics.

Modeling parameters and structure optimization

The 2D structure construction, energy minimization and geometry optimization of the selected sulfonamide derivatives were carried out by using ChemDraw Ultra 7.0 and Chem3D Pro 7.0 (Cambridge Soft Corporation, 100 Cambridge Park Drive, Cambridge MA, 02140 USA) on an Intel(R) Core(TM)2 Duo Central Processing Unit T6670 @ 2.20 GHz and 4.00 GB of RAM, running the Windows 7 Home Basic, 64-bit compatible operating system. The energy minimization was carried out to minimum RMS Gradient of 0.100, with step interval of 2.0 Fs and frame interval of 10 Fs. Descriptors calculated for the training set are given in Table 1.

Materials and Method:

Table 1. Descriptors calculated for the training set.

1. ALOGP	2. Molecular linear free energy relation
3. APol	4. Petitjean number
5. Aromatic atoms count	6. Rotatable bonds count
7. Aromatic bonds count	8. Rule of five
9. Atom count	10. TPSA
11. Autocorrelation (arge)	12. VadjMa
13. Autocorrelation (Mass)	14. Weight
15. Autocorrelation (Polarizability)	16. Weighted path
17. BCUT	18. Wiener numbers
19. Bond count	20. XlogP
21. BPol	22. Zagreb index
23. Carbon types	24. CPSA

25. Chi chain	26. Gravitational index
27. Chi cluster	28. Length over breadth
29. Chi path cluster	30. Moment of inertia
31. Chi path	32. Petitjean shape index
33. Eccentric connectivity index	34. WHIM
35. Atom type electrotopological state	36. Largest Pi system
37. Fragment complexity	38. Longest aliphatic chain
39. Hbond acceptor count	40. Mannhold LogP
41. Hbond donor count	42. McGowan volume
43. Kappa shape indices	44. MDE
45. Largest chain	

Descriptor selection

The selection of descriptors among the calculated descriptors for the multiple linear regression analysis is based on the correlation matrix. This matrix is prepared and analyzed for the least correlated descriptors. The correlation matrix is given in Table 2

Table 2. Correlation matrix

	Eccentricity	AlogP	McGowanVol.	Mol. Wt.	Frag. complxity
Eccentricity	1				
AlogP	0.1732	1			
McGowan Vol.	0.7225	-0.2973	1		
Mol. Wt.	0.54	-0.2766	0.92	1	
Frag. complxity	0.7166	0.149	0.7098	0.5633	1

McGowan volume

It has been widely used for the analysis of physicochemical and biochemical properties in chemistry and drug discovery field, because McGowan volumes are unavailable for ions, its application is limited to only neutral compounds.[5]

Partition co-efficient (logP)

The partition coefficient is a ratio of concentrations of un-ionized compound between the two solutions. To measure the partition coefficient of ionizable solutes, the pH of the aqueous phase is adjusted such that the predominant form of the compound is un-ionized. The logarithm of the ratio of the concentrations of the un-ionized solute in the solvents is called log *P*: The log P value is also known as a measure of lipophilicity.

AlogP is termed as atomic logP.

Model validation

The QSAR model validation was carried with statistical analysis.

Statistical parameters

Fraction of variance (r^2): The value of fraction of variance may vary between 0 (means model without explanatory power) and 1 (means perfect model). QSAR model having $r^2 > 0.6$ will only be considered for validation[4].

LOO-CV (leave-one-out cross validation) Test (q^2): Cross validation is the most commonly used technique, in which compounds with different proportions are removed from original data set and

developed a new QSAR model in order to verify the internal predictive ability of the original QSAR model. A QSAR model must have $q^2 > 0.5$ for the predictive ability.[4]

Standard deviation (s): The smaller s value is always required for the predictive QSAR model.

r^2-q^2 < 0.3: The difference between r^2 and q^2 should never be exceeding by 0.3. A large difference suggests the following: presence of outliers, over-fitted model, and presence of irrelevant variables in data.[4]

Quality Factor (Q): Overfitting and chance correlation, due to excess number of descriptors, can be detected by Q value. Positive value for this QSAR model suggests its high predictive power and lack of overfitting.

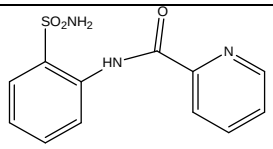
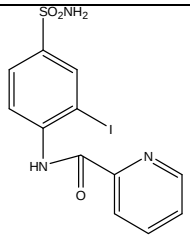
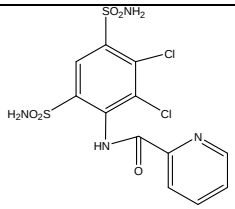
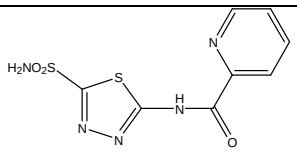
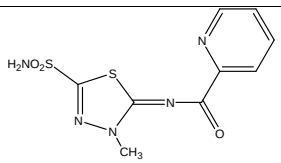
Fischer Statistics (F): The F value of QSAR model was compared with their literature value at 95% level.

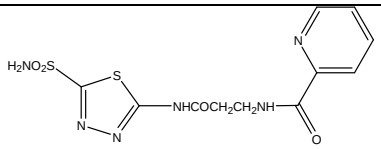
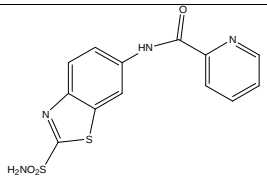
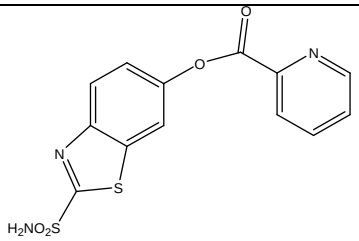
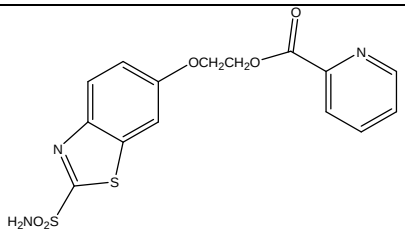
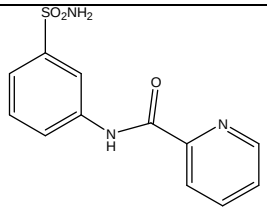
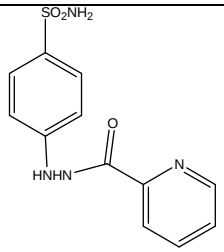
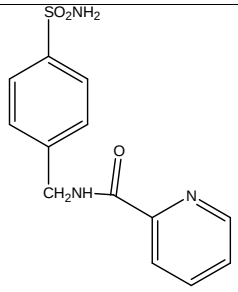
Results and Discussion

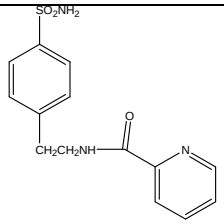
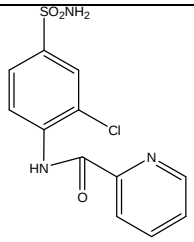
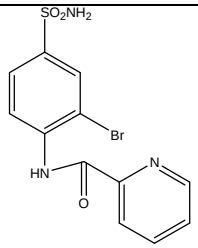
From the data in Table 3, QSAR equation have been developed, here 95% confidence intervals are given in parantheses.

logKi = 7.418976 (3.163464) -2.895065 (1.677445) (McGowan volume) -1.821205 (0.6095455) (AlogP)

Table 3. Descriptors used to derive QSAR equation along with bioactivities

	Training set	logKi			Descriptors used	
S.no.		Obs.	Pred.	Diff.	McGowan	AlogP
1		4.33	3.844167	0.485833	1.6837	-0.7136
2		2.78	2.519927	0.260073	1.9634	-0.4311
3		2.71	3.147376	-0.43738	2.3307	-1.3595
4		1.6	1.889393	-0.28939	1.6122	0.4734
5		1.49	2.59652	-1.10652	1.7101	-0.0705

6		1.3	1.714086	-0.41409	2.0429	-0.115
7		1	1.781432	-0.78143	1.9578	-0.0167
8		1.04	0.847986	0.192014	1.9382	0.527
9		1	0.349587	0.650413	2.1927	0.3961
10		4.29	3.844167	0.445833	1.6837	-0.7136
11		4.35	4.383118	-0.03312	1.762	-1.134
12		3.04	3.082127	-0.04213	1.7816	-0.4508

13		3.04	2.3332	0.7068	1.8795	-0.1952
14		2.78	2.756089	0.023911	1.8276	-0.3449
15		2.79	2.450827	0.339173	1.8802	-0.2609

Validation of QSAR model

A quantitative assessment of model robustness has been performed through model validation. All the statistical results of model validation have been given in Table 4.

Table 4. Results of statistical validation

n/p (>=4)	r^2	q^2	$r^2 - q^2 < 0.3$	Q	RMSD	variance	F
7.5	0.8138	0.7887	0.025	0.7978	0.131	0.321	26.08

n= no. of molecules taken for modeling, p= no. of descriptors used

Internal validation

Y-Randomization Test*:* To establish the QSAR model robustness, this technique is being used widely. For this test, the dependent variable vector is randomly shuffled, and a new QSAR model is developed using the unchanged independent variable. This process was repeated for five times. The statistical data of r^2 for five runs are given in Table 5. The values $r^2<0.6$ in Y-randomization test confirm the robustness of this QSAR model.[4]

Table 4: Results of internal validation: Y-randomization test (5 runs)

	Shuffled observed logKi				
	Run 1	**Run 2**	**Run 3**	**Run 4**	**Run 5**
1	1	4.29	1	4.35	1.3
2	1	4.35	1.04	3.04	1
3	1.04	3.04	1	3.04	1.04
4	4.29	3.04	4.29	2.78	1

5	4.35	2.78	4.35	2.79	4.29
6	3.04	2.79	3.04	4.33	4.35
7	3.04	4.33	3.04	2.78	3.04
8	2.78	2.78	2.78	2.71	3.04
9	2.79	2.71	2.79	1.6	2.78
10	4.33	1.6	4.33	1.49	2.79
11	2.78	1.49	2.78	1.3	4.33
12	2.71	1.3	2.71	1	2.78
13	1.6	1	1.6	1.04	2.71
14	1.49	1.04	1.49	1	1.6
15	1.3	1	1.3	4.29	1.49
r^2	0.3	0.052	0.31	0.019	0.014

According to the developed QSAR model, the sulfonamide analogues must have negative McGowan volume for enhanced activity. Moving towards the effects of the AlogP on the bioactivity of derivatives of sulfonamide as carbonic anhydrase inhibitor, the developed QSAR model suggest that a negative AlogP will definitely be favourable to the activity, as discussed by Verma and Hansch (2010)[4], Ajeet (2012)[6]. A comparison (multiple linear regression plots) of observed values and predicted values of logKi for sulfonamide analogues used for development of QSAR equation is shown in Figure 1 and multiple linear graph for pattern analysis is shown in Figure 2.

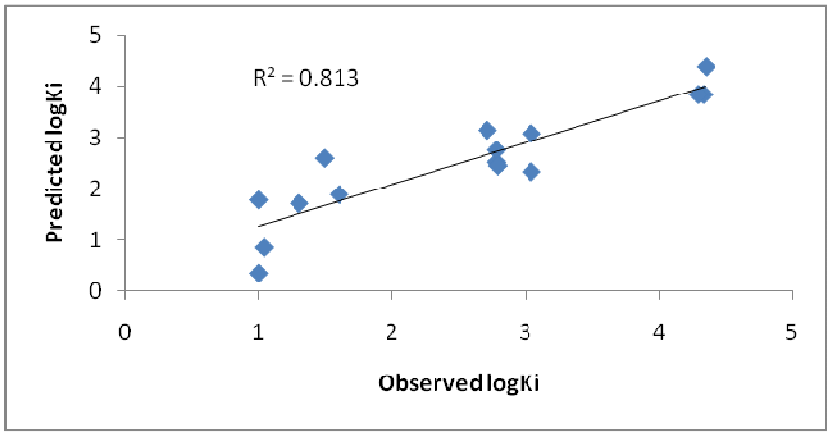


Fig. 1. Multiple linear regression plot for QSAR study

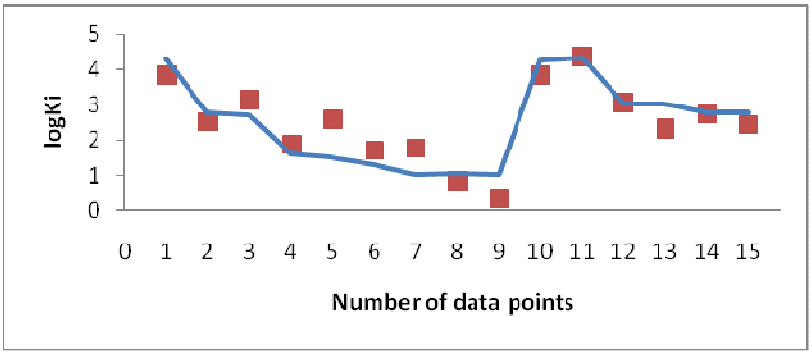


Fig. 2. Multiple linear graph between Number of data points and bioactivities

Conclusion

With deluged data of QSAR studies of sulfonamide analogues, we could draw a number of conclusions. On the basis of discussion given earlier we could conclude that for developing the novel sulfonamide analogues on the basis of McGowan volume and AlogP, we have to select such groups or substituent which decrease the McGowan volume as well as AlogP of the novel molecules and we can evaluate the novel molecules for their logKi values on the basis of the derived QSAR equation. While playing with McGowan volume of molecule during its synthesis, we can alter its physiological and biochemical properties for gaining positive attitude toward its bioactivity values and reducing the toxicity. For developing a drug like anticonvulsant, it is necessary to alter the partition coefficient of novel molecule for crossing the blood brain barrier. So, with the help of the developed model we can identify the logKi values of novel designed molecules and alter their structural properties accordingly before synthesizing them.

References

1. Naama. H., Meir. B., Bogdan. W., Richard. H. F., Boris. Y. 2010. Syntheses and Evaluation of Anticonvulsant Profile and Teratogenicity of Novel Amide Derivatives of Branched Aliphatic Carboxylic Acids with 4-Aminobenzensulfonamide. Journal of Medicinal Chemistry. 53: 4177-4186.
2. Tripathi. L., Kumar. P., Singh. R., and Stable. J. P. 2012. Design, synthesis and anticonvulsant evaluation of novel N-(4-substituted phenyl)-2-[4-(substituted) benzylidene]-hydrazinecarbothioamides. European Journal of Medicinal Chemistry. 47: 153-166.
3. Pooja. S. B. and Sharma. P. K. 2012. New antiepileptic agents: structure-activity relationship. Medicinal Chemistry Research. 21: 1491-1508.
4. Verma. R. P. and Hansch. C. 2010. QSAR modeling of taxane analogues against colon cancer. European Journal of Medicinal Chemistry. 45: 1470-1477.
5. Yuan. H. Z., Michael. H. A., and Andreas. M. Z. 2003. Determination of McGowan Volumes for Ions and Correlation with van der Waals Volumes. Journal of Chemical Information and Computer Sciences. 43(6): 1848–1854.
6. Ajeet. 2012. QSAR modeling of recombinant human stromelysin inhibitors– MLR approach. International Journal of Advanced Pharmaceutical and Biological Sciences. 2(2): 171-175.

7. Singh. S., Singh. J., Ingle. M., Mishra. R., and Padmakar. V. K. 2006. A QSAR study on carbonic anhydrase inhibition: predicting logKi(hCAI) by using (SO_2NH_2) NMR chemical shift as a molecular descriptor. Arkivoc. xvi: 1-15.

8. Yap. C. W. 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry. 32(7): 1466-1474.