Original Research Article

Prediction of antitumor activity of some 1, 5-N, N'–disubstituted-2- (substituted benzenesulphonyl)-glutamamides as antitumor agents using QSAR approach

Gourav Chouhan[1], Yash Shukla[1], Suparna Ghosh[1], K. Anita[1], Ruchi Dubey Sharma[1], Shweta Sharma[1*]

[1]Department of Chemistry, Career College, BHEL, Bhopal-4602023, Madhya Pradesh, INDIA.
Email: shwetasharma2703@yahoo.com

Abstract

A quantitative structure–activity relationship (QSAR) analysis was performed on a data set of 5 - N-substituted-2-(substituted Benzenesulphonyl)-Glutamaminesderivatives for their antitumor activity. Several types of descriptors, including 2D autocorrelation, 3DMorse and WHIM descriptors were used to derive a quantitative relationship between antitumor activity and structural properties. A multiple linear regression analysis shows that five-parametric model found to be best for modeling log (BA) activity of present set of compounds. For the best QSAR model the statistics were R2=0.8262, Q=20.910, N=28 for the present set of compounds. This model was further validated using the leave-one-out (LOO) cross-validation approach and Ridge regression analysis.
Key Words: QSAR, 2D autocorrelation, Ridge regression analysis, Model.

[*]**Address for Correspondence:**
Dr. Shweta Sharma, Department of Chemistry, Career College, BHEL, Bhopal-4602023, Madhya Pradesh, INDIA.
Email: shwetasharma2703@yahoo.com

Access this article online

Quick Response Code:

Website: www.statperson.com

Accessed Date: 26 March 2018

INTRODUCTION

Around the world, tremendous resources are being invested in prevention, diagnosis, and treatment of cancer. Cancer is the second leading cause of death in Europe and North America. Discovery and development of anticancer agents are the key focus of several pharmaceutical companies as well as non- profit government andnon-government organizations,like theNational Cancer Institute (NCI) in the United States, the European Organization for Research and Treatment of Cancer(EORTC),and the British Cancer Research Campaign (CRC)[1]. Accordingly we need new approach for cancer therapy. Among the wide spectrum of cancer treatment strategies, chemotherapy has significant role in management of cancer. Therefore discovery of novel anti cancer agents is one of the necessities in cancer research. The main objectives of anti cancer drug development are decreasing the toxicity and maximizing the efficacy. The discovery of new compounds with antitumoral activity has become one of the most important goals in medicinal chemistry. QSAR is among the most extensively used computational methodology for analogue-based drug design. Cancer has been described as a nitrogen trap. Tumor cells are consumers of avid glutamine. Glutamine (GLN) which is easily converted into glutamic acid in cell, supplies its amide nitrogen to tumor cells in the biosynthesis of purine and pyrimidine bases. Isoglutamines, which have 1-N-amide instead of 5-N-amide and 5- COOH instead of 1-COOH in glutamine, shows antitumor activity. Thus, compounds containing both 1- and 5- amido group may have potential antitumor activity.

Although it is worthy to mention that we have successfully modelled the various activity of many compounds using qsar methodology.

MATERIALS AND METHOD

In the present study for the development of new potential antitumor agents through computer aided drug designing we have used 1, 5- N, N' – Disubstituted-2- (substituted Benzenesulphonyl) –Glutamamides, this was synthesised and tested by Jha and co-workers. The structures and biological activity data were also taken from their studies.[2].

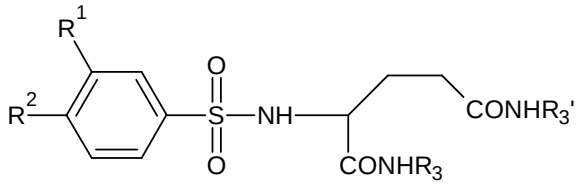

Figure 1:

Presentation of data: Table 1 records structural details of 28 1,5- N,N' – Disubstituted-2- (substituted Benzenesulphonyl) – Glutamamides derivatives. From the large pool of RDF, WHIM and GETAWAY descriptors we have selected a few to carryout multiple regression analysis. This selection of descriptors was done by using variable selection for multiple regression analysis available with the NCSS software[3]. The calculated values of such descriptors are presented in Table 2. The intercorrelatedness among the descriptors and their correlation with the activity values Log BA is presented in Table 3. The regression parameters as well as the quality of different models containing one to several correlating parameters are summarized in Table 4. Using the best five -parametric model, we have estimated and compared the values of activity. Such a comparison is demonstrated in Table 5. Finally, all the proposed models are validated by cross-validation method (Table 6). The presence/absence of co- linearity, if any, was examined by Ridge regression parameters[4] (Table 7, Figs. 3 and 4).

Molecular parameters used: DRAGON software[5] has been used for calculation of all RDF, WHIM and GETAWAY descriptors. They include: RDF 080u, E1e, G2p, G3p and H7v. In fact before this study, topological parameters have been very successfully used in our research for modeling different activities of drug molecules[6-13].

RESULTS AND DISCUSSION

Table 1: Structural details and LogBA values for the compounds used in present study

Comp. no.	R^1	R^2	R^3	$R^{3'}$	Log BA
1	NO_2	H	H	H	1.606
2	NO_2	H	CH_3	CH_3	1.808
3	NO_2	H	C_2H_5	C_2H_5	1.64
4	NO_2	H	n- C_3H_7	n- C_3H_7	1.275
5	NO_2	H	n-C_4H_9	n-C_4H_9	1.431
6	NO_2	H	i-C_3H_7	i-C_3H_7	1.693
7	NO_2	H	n-C_6H_{13}	n-C_6H_{13}	1.534
8	NO_2	H	C_6H_5	C_6H_5	1.728
9	Cl	Cl	H	H	1.624
10	Cl	Cl	CH_3	CH_3	1.874
11	Cl	Cl	C_2H_5	C_2H_5	1.774
12	Cl	Cl	n- C_3H_7	n- C_3H_7	1.458
13	Cl	Cl	n-C_4H_9	n-C_4H_9	1.57
14	Cl	Cl	i-C_3H_7	i-C_3H_7	1.744
15	Cl	Cl	n-C_6H_{13}	n-C_6H_{13}	1.462
16	Cl	Cl	C_6H_5	C_6H_5	1.741
17	H	Br	H	H	1.582
18	H	Br	CH_3	CH_3	1.962
19	H	Br	C_2H_5	C_2H_5	1.782
20	H	Br	n- C_3H_7	n- C_3H_7	1.843
21	H	Br	n-C_4H_9	n-C_4H_9	1.696
22	H	Br	i-C_3H_7	i-C_3H_7	1.839
23	H	Br	n-C_6H_{13}	n-C_6H_{13}	1.696
24	H	C_2H_5	H	H	1.58
25	H	C_2H_5	CH_3	CH_3	1.542
26	H	C_2H_5	C_2H_5	C_2H_5	1.642
27	H	C_2H_5	n- C_3H_7	n- C_3H_7	1.46
28	H	C_2H_5	n-C_4H_9	n-C_4H_9	1.769

Table 2: Calculated values of 2D autocorrelation, RDF, WHIM, and GETAWAY descriptors for the compounds used in present study

Sr. No.	RDF080u	E1e	G2p	G3p	H7v
1	8.503	0.560	0.170	0.170	0.160
2	8.771	0.506	0.156	0.204	0.178
3	14.536	0.474	0.162	0.177	0.141
4	19.518	0.492	0.148	0.157	0.131
5	19.434	0.505	0.153	0.171	0.210
6	15.902	0.471	0.148	0.191	0.189
7	21.107	0.510	0.172	0.206	0.320
8	16.352	0.477	0.156	0.188	0.201
9	4.369	0.492	0.164	0.164	0.042
10	8.523	0.480	0.171	0.177	0.105
11	10.037	0.476	0.179	0.190	0.155
12	18.249	0.477	0.176	0.172	0.187
13	18.144	0.480	0.169	0.173	0.256
14	17.244	0.447	0.149	0.167	0.239
15	27.437	0.492	0.159	0.191	0.350
16	17.240	0.462	0.165	0.200	0.233
17	6.213	0.469	0.246	0.211	0.042
18	6.621	0.455	0.158	0.189	0.100
19	11.602	0.463	0.174	0.218	0.139
20	15.671	0.464	0.149	0.190	0.169
21	17.711	0.470	0.162	0.154	0.236
22	15.247	0.432	0.149	0.154	0.218
23	23.308	0.474	0.165	0.214	0.382
24	11.318	0.536	0.158	0.183	0.067
25	12.640	0.503	0.153	0.169	0.144
26	15.187	0.498	0.159	0.215	0.189
27	18.771	0.488	0.173	0.185	0.194
28	19.308	0.489	0.143	0.200	0.246

Table 3: Correlation matrix

	LogBA	RDF080u	E1e	G2p	G3p	H7v
Log BA	1.000					
RDF080u	-0.430	1.000				
E1e	-0.461	-0.090	1.000			
G2p	-0.118	-0.347	-0.013	1.000		
G3p	0.240	0.023	0.014	0.328	1.000	
H7v	-0.101	0.852	-0.101	-0.303	0.178	1.000

Table 4: Regression parameters and quality of correlations

Model No.	Parameters Used	A_i=(1.....6)	B	Se	R^2	R^2_A	F	Q=R/Se
1	RDF080u	-0.0123(±0.0050)	1.8391	0.0870	0.1850	0.1537	5.904	4.944
2	E1e	-2.7865(±1.0526)	3.0032	0.0855	0.2123	0.1820	7.009	5.389
3	G2p	-0.9842(±1.6185)	1.8167	0.0957	0.0140	0.0000	0.370	1.236
4	G3p	2.0006(±1.5890)	1.2854	0.0935	0.0575	0.0212	1.585	2.565
5	H7v	-0.1922(±0.3722)	1.6914	0.0959	0.0101	0.0000	0.267	1.048
6	E1e RDF080u	-3.0454(±0.9115) -0.0136(±0.0043)	3.3313	0.0738	0.4366	0.3915	9.687	8.953
7	E1e G2p	-2.7966(±1.0629) -1.0357(±1.4608)	3.1777	0.0863	0.2279	0.1661	3.689	5.532
8	E1e G3p	-2.8067(±1.0314) 2.0532(±1.4235)	2.6331	0.0838	0.2728	0.2147	4.690	6.233
9	E1e H7v	-2.8775(±1.0638) -0.2840(±0.3356)	3.1002	0.0860	0.2343	0.1730	3.824	5.628
10	E1e G2p	-3.1355(±0.8483) -2.7491(±1.2382)	3.8747	0.0686	0.5326	0.4742	9.116	10.638

	RDF080u	-0.0169(±0.0043)						
	E1e	-3.0700(±0.8739)						
11	G3p	2.1506(±1.2015)	2.9480	0.0707	0.5030	0.4408	8.095	10.031
	RDF080u	-0.0137(±0.0041)						
	E1e	-2.9074(±0.7099)						
12	H7v	1.7749(±0.4266)	3.2710	0.0574	0.6727	0.6317	16.440	14.289
	RDF080u	-0.0361(±0.0064)						
	E1e	-2.9965(±0.6167)						
	G2p	-2.6796(±0.8992)						
13	H7v	1.7570(±0.3702)	3.8012	0.0498	0.7638	0.7228	18.598	17.549
	RDF080u	-0.0391(±0.0056)						
	E1e	-2.9294(±0.7106)						
	G3p	1.0101(±1.0241)						
14	H7v	1.6405(±0.4481)	3.0955	0.0574	0.6859	0.6313	12.559	14.428
	RDF080u	-0.0345(±0.0066)						
15	E1e	-3.0786(±0.5418)						
	G2p	-3.5957(±0.8536)						
	G3p	2.3675(±0.8431)	3.5712	0.0437	0.8262	0.7866	20.910	20.800
	H7v	1.4357(±0.3443)						
	RDF080u	-0.0363(±0.0050)						

Table 5: Observed and Estimated LogBA values using model no 15 (Table 4).

Comp. No.	Obs. Log BA	Est. LogBA	Residual
1	1.606	1.559	0.047
2	1.808	1.872	-0.064
3	1.640	1.623	0.017
4	1.275	1.375	-0.100
5	1.431	1.467	-0.036
6	1.693	1.735	-0.042
7	1.534	1.563	-0.029
8	1.728	1.681	0.047
9	1.624	1.757	-0.133
10	1.874	1.739	0.135
11	1.774	1.770	0.004
12	1.458	1.483	-0.025
13	1.570	1.604	-0.034
14	1.744	1.771	-0.027
15	1.462	1.443	0.019
16	1.741	1.737	0.004
17	1.582	1.577	0.005
18	1.962	1.953	0.009
19	1.782	1.814	-0.032
20	1.843	1.730	0.113
21	1.696	1.602	0.094
22	1.839	1.829	0.010
23	1.696	1.727	-0.031
24	1.580	1.471	0.109
25	1.542	1.620	-0.078
26	1.642	1.695	-0.053
27	1.460	1.481	-0.021
28	1.769	1.677	0.092

Table 6: Cross validated parameters for the best obtained models

Model No.	Parameters used	PRESS/SSY	R^2_{cv}	S_{PRESS}	PSE
15	E1e G2p G3p H7v RDF080u	0.2104	0.7896	0.0723	0.0641

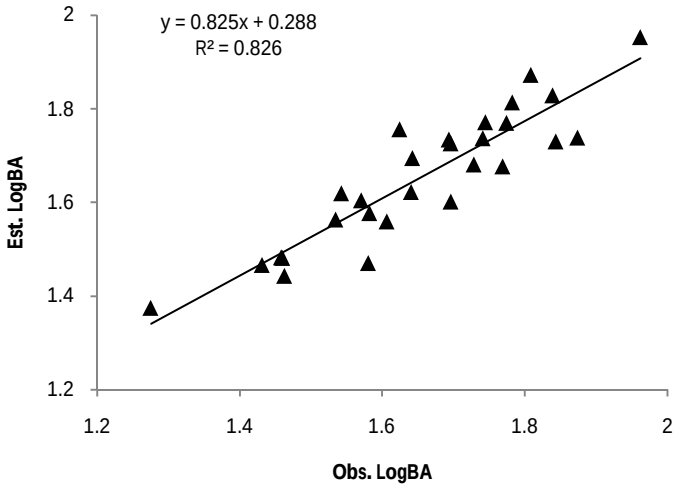

Figure 2: Correlation between Observed and estimated LogBA using model (Table 4).

Table 7: Ridge analysis for the best five-parametric model

Model No.	Parameters used	VIF	Tolerence	Eigenvalue	Condition no.
15	RDF080u	3.9337	0.2542	2.0663	1.00
	E1e	1.0157	0.9845	1.2832	1.61
	G2p	1.3349	0.7491	0.9917	2.08
	G3p	1.2915	0.7743	0.5257	3.93
	H7v	4.1235	0.2425	0.1328	15.55

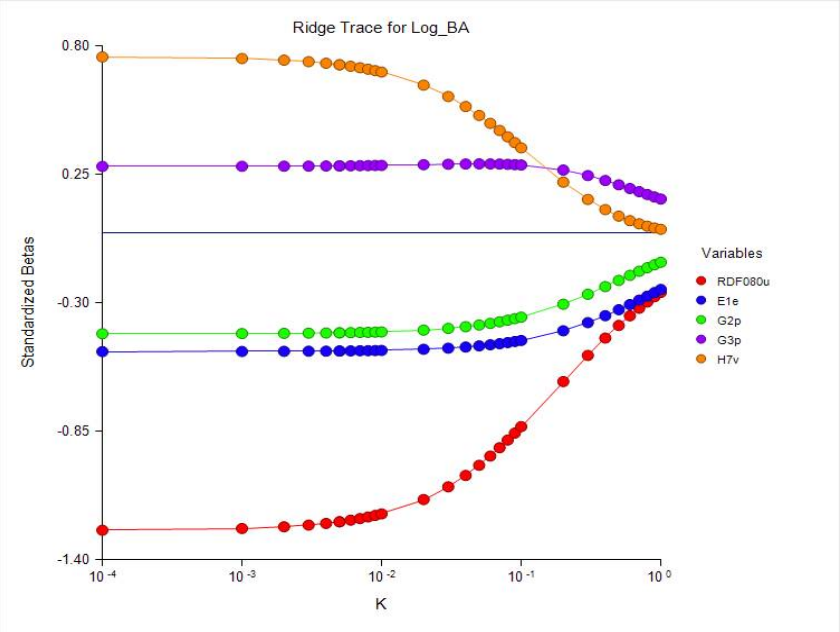

Figure 3: Ridge trace for best five-parametric model

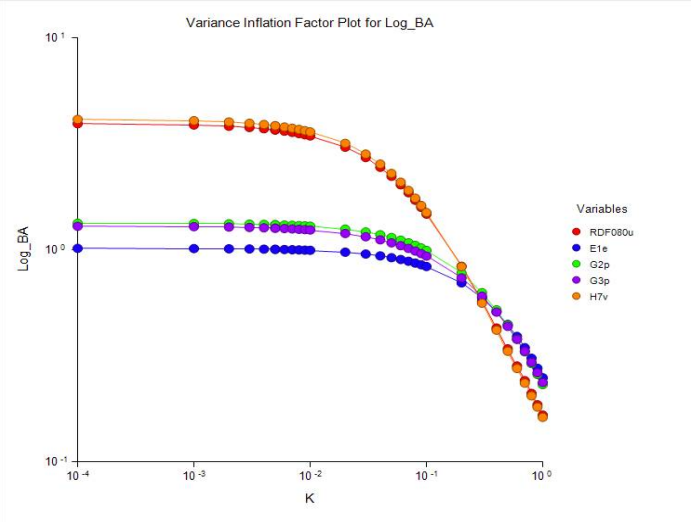

Figure 4: VIF plot for five-parametric model.

The data presented in Table 4 indicates that statistically allowed model start pouring using two or more parameters as correlating descriptors. We observed that in all these higher parametric models E1e is invariably present as one of the correlating descriptors. By examination of Table 4 we also observed that both R^2 and R^2_A go on increasing with each addition of descriptor in the regression analysis. This indicates that addition of descriptor in each case is favorable for the exhibition of the activity.

One-variable model: Among all the models the best one-parametric model contains E1e having R^2 value equal to 0.2123. The model is as below.

LogBA = -2.7865(±1.0526) E1e+ 3.0032

N=28, R^2=0.2123, R^2A=0.1820, Se=0.0855, F=7.00, Q=5.389 (1)

Here, and here after N is the number of compound, Se is the standard error of estimation, R^2 is the square of correlation coefficient, R^2_{Adj} is the adjusted R^2, F is the Fisher's ratio, and Q is the Pogliani's quality factor which is the ratio of R/Se.(Pogliani,1994,1996)[14-15].

Two -variable model

When we add RDF080u to the one- parametric model, a two - parametric model with improved R^2 value is obtained. The value of R^2 is equal to 0.4366, which is better than the previous value. The model is given below

LogBA = -3.0454(±0.09115) E1e - 0.0136(±0.0043)RDF080u +3.3313

N=28, R^2=0.4366, R^2_A=0.3915, Se=0.0738, F=9.687,Q=8.953 (2)

Three-variable model

Further addition of H7v yielded a three -parametric model with R^2=0.6727.

LogBA = -2.9074(±0.7099) E1e +1.7749(±0.4266)H7v - 0.0361(±0.0064) RDF080u + 3.2710

N=28, R^2=0.6727, R^2_A=0.6317, Se=0.0574, F=16.440, Q=14.289 (3)

Four -variable model

In the above model, addition of G2p yielded a four - parametric model with R^2=0.7638.

LogBA = -2.9965(±0.6167)E1e-2.6796(±0.8992)G2p +1.7570(±0.3702)H7v-0.0391(±0.0056) RDF080u + 3.8012

N=28, R2=0.7638, R^2_A=0.7228, Se=0.0498, F=18.598, Q= 17.549 (4)

Five -variable model

However, on the basis of R^2 and Q values another model with E1e, G2p, G3p, H7v, and RDF080u, as correlating parameters was obtained which is the best model for modelling the LogBA value of 28 compounds acting as antitumoral agents. The model is given below.

LogBA = -3.0786(±0.5418)E1e -3.5957(±0.8536)G2p +2.3675(±0.8431)G3p+ 1.4357(±0.3443)H7v - 0.0363(±0.0050) RDF080u + 3.5712

N=28, R2=0.8262, R^2_A=0.7866, Se=0.0437, F=20.910, Q=20.800 (5)

As the data set contains only 28 compounds no higher parametric correlation is permitted. Therefore, the five-parametric model obtained above is the best model for estimating LogBA activity of proposed set of compounds. The predictive power of this model comes out to be 0.8262, indicating that about 82% of the data is explained by this model. The estimated activity values using the best five -parametric model has been reported in Table 5, and are in good agreement with the observed ones confirming that the proposed five-parametric model is best suitable for modeling, estimating LogBA activity of present set of compounds. All the above models have been tested using cross validated parameters. These parameters are reported in Table 6. It is worth mentioning

that PRESS is a good estimate of the real predictive power of the model. If PRESS is smaller than SSY, the model predicts better than chance and can be considered statistically significant. Table 6 shows that in this regard, all the models proposed by us are better than chance and are statistically significant. The ratio PRESS / SSY can be used to calculate the approximate confidence interval of the prediction of new compounds. To be a reasonable QSAR model, this ratio should be smaller than 0.4. The models proposed by us are found to have this ratio smaller than 0.4 and the model expressed by equation 1, 2, 3, 4 and 5 has the excellent predictive power. The developed models are cross-validated by leave-one-out method. The high values observed in case of eqn. 5 (R^2_{CV} = 0.78) are indicative of their reliability in prediction of biological activity. Another cross-validated parameter related to uncertainty of prediction, the PSE, is calculated. The lowest value of PSE for model 15 (eq.5) supports its highest predictive potential (power). The highest R^2_{CV} and lowest PSE for the model 15 shows that this is the most appropriate model for modeling LogBA value of 28 compounds used in the present study. We have further carried out analysis to test model-15 whether it suffers from the defect due to co- linearity. For this we have, subjected this model to Ridge analysis and calculated Ridge traces [Fig. 3 and 4]. All these results have finally demonstrated that the proposed model-15 is the most appropriate model for modeling the activity and that it is devoid of any co- linearity defect.

CONCLUSION

On the basis of above discussion and a close look of equation no. 5 we may conclude that the negative coefficients of E1e -the first component accessibility directional WHIM index weighted by sanderson electronegativity, G2p- second component symmetry directional WHIM index weighted by polarizability, RDF080u- Radial distribution function-080 unweighted and the positive coefficients of G3p- third component symmetry directional WHIM index weighted by polarizability and H7v- H autocorrelation of lag7 weighted by Vander waals Volume are favorable for modeling the biological activity of present set of compounds.

REFERENCES

1. S. Narang and D. S. Desai, “Anticancer drug development”. In : Lu Y, Mahato RI, editors. Pharmaceutical Perspectives of Cancer Therapeutics New York: Springer Verlag; pp 49-92, 2009.
2. B. Debnath, K. Srikanth, S. Banerjee, T. Jha, “Internet Electronic Journal of Molecular Design”, Vol. 1, pp. 488-502, 2002.
3. NCSS statistical analysis software, www.ncss.com
4. S. Chaterjee, A.S. Hadi, B. Price, Regression Analysis by Examples, 3rd Ed., Wiley, New York, 2000, 51.
5. DRAGON software for calculation of topological indices, www.disat.unimib.it
6. S. Sharma, B. Shaik, V.K. Agrawal “Journal of current pharma and research” Vol. 6 (2), pp. 1777-1785, 2016.
7. S. Sharma, B. Shaik, S.P. Gupta, V.K. Agrawal, “Journal of Applied biopharmaceutical and pharmacokinetics”, vol. 4, pp. 1-12, 2016.
8. S. Sharma, K. Anita, V.K. Agrawal, S. Shrivastav, “Moroccan Journal of Chemistry”, Vol.3 (2) pp. 212-221, 2015.
9. K. Anita, V.K. Agrawal, B. Shaik, S. Sharma, “International Journal of Pharmaceutical Sciences and Research” 5 (8), pp. 3333-3341, 2014.
10. S. Sharma, K. Anita, B. Shaik, I. Ahmed, V.K. Agrawal, “J. Engg. Sc. Mgmt”.Ed. 6 (II),pp. 151–155, 2013.
11. S. Sharma, V.K. Agrawal, K. Anita, B. Shaik, S. Shrivastava, “IOSR -Journal of Applied Chemistry (IOSR-JAC)” 6 (3), pp. 41-48, 2013.
12. K. Anita, S. Sharma, B. Shaik, “IOSR- Journal of Applied Chemistry (IOSR-JAC)” 6 (3),pp. 9-17, 2013.
13. K. Anita, S. Sharma, B. Shaik, S. Shrivastava, “IOSR Journal of Applied Chemistry (IOSR-JAC)” Vol. 4, (1) pp. 26-33, 2013.
14. L. Pogliani, “J. Phys. Chem.”, 100. pp. 18065-18077, 1996.
15. L. Pogliani L, “Amino Acids”., 6, pp.141-153, 1994.

Source of Support: None Declared
Conflict of Interest: None Declared