In Silico Multivariate Regressio n Analysis and Validation Studies on Selective MMP-13 Inhibitors

G. Nirmala
Senior Asst professor
Dept of C.S.E
Sir C R Reddy college of Engineering
ELURU – 534007, A.P, India

A. Yesubabu, PhD
Professor
HOD C.S.E
Sir C R Reddy college of Engineering ELURU – 534007, A.P, India

P. Seetharamaiah, PhD
Emeritus professor
,Dept of C.S.E
Andhra University
Vishakapatnam – 530003
A.P, India

ABSTRACT
QSAR(Quantitative Structure Activity Relationship) studies were carried out on a set of 72 α-sulfone hydroxamatesas Matrix Metalloproteinase-13 (MMP-13) inhibitors using multiple regression procedure. Outliers were removed based on Relative Error calculation and Extent of Extrapolation. The activity contributions of these compounds were determined from regression equation and the validation procedures such as external set cross-validation r^2, ($R^2_{cv, ext}$) and the regression of observed activities versus predicted activities and vice versa for validation set was described to analyze the predictive ability of the QSAR model. Parameters concerning predictive ability of QSAR model and Y-randomization tests were found to be within the limits. From a set of 5 models, an accurate and reliable QSAR model involving six descriptors was chosen based on the FIT Kubinyi function, which defines the statistical quality of the model. The generated model could be useful in designing more potent inhibitors of MMP-13.

Keywords
α-sulfone hydroxamates, QSAR,

Multiple regression, Cross validation, Outliers,

FIT Kubinyi, descriptors, MMP-13.

1. INTRODUCTION
Matrix metalloproteinases (MMPs) belong to the metzinc in super family (i.e.) they bind zinc at the catalytic site and have a conserved 'Met-turn' motif. Matrix metalloproteinase (MMPs) play an important role in the tissue modeling and remodeling of the extra cellular matrix in both physiologic and pathologic states and thus plays an important role in tumor progression[1,2]. Matrix metalloproteinases are structurally similar, but differ in substrate specificity, in that each MMP has the ability to degrade particular subset of matrix proteins[3]. Abnormal activity of these enzymes has been related to a variety of pathologic processes, involving metastasis, angiogenesis, cardiovascular disease, osteoarthritis, and rheumatoid arthritis[4]. The development of potent subclass selective inhibitors of these enzymes has been challenging, and they rely on a small number of zinc binding motifs [5].

MMP inhibitors may bind to members of the structurally-related ADAMs family (A Disintegrin and Metalloprotease) [6] , leading to undesired joint effects. Therefore, an alternative approach to design selective MMP inhibitors with reduced side effects is to optimize the inhibition of the single MMP isozymes that should confer the most therapeutic benefit, reducing the probability of off-target protease inhibition [7].MMP-13 is an attractive isozyme to pursue; MMP-13 rapidly degrades type II collagen and is associated with pathology. The isozyme is upregulated in osteoarthritis joints and in cancer[8].

Therefore in search of potent MMP-13 inhibitors, a novel series of various α-sulfone hydroxamates reported in three papers[9] were considered to perform structure-activity relationship (SAR) studies as they delineate the structural requirements for potency of inhibitors. QSAR studies have been investigated on the basis of the fact that the biological activity of the compound is a function of its physicochemical properties. From literature it was observed that several attempts were made to build QSAR models of various nonzinc chelating compounds[10**]**, piperazine analogs[11], carboxylic acid based compounds[12],N-hydroxy-a-phenylsulfony-l acetamide[13] and docking based QSAR [14] studies were reported. Moreover, none of the QSAR studies reported on α-sulfone hydroxamate analogs that covered two or more different kinds of ligands. Hence, a QSAR study on ligands with observable structure diversity, if possible, will definitely lead to more universal and robust QSAR models for designing novel compounds against MMP-13.

To address such powerful models covering different types of ligands, here, we report QSAR studies on 57 α-sulfone hydroxamate analogs, respectively, to investigate the influence of molecular structure on biological activity.

2. METHODS
2.1 Data set
A set of 72 compounds biological data[15,16]reported in literature were utilized to obtain a reliable and robust QSAR model. The inhibitory activities of these derivatives reported in terms of IC_{50} concentration values and their structures along with bioactivities are given in Table 1. The structures were sketched using ISIS Draw 2.3 (www\\mdli.com) software and the descriptors were calculated using Tsar Software.

2.2 Multivariate Regression Analysis
QSAR models were constructed on complete and training sets, respectively Validation was done internally using leave-one-out (LOO) technique and dependent variable ($log1/IC_{50}$) and independent variables was established by linear multiple regression analysis. Significant descriptors were chosen based on the statistical data of analysis. The Generated QSAR equation on the parameters like correlation coefficient (r),

Standard error of estimate externally by predicting the activities of validation set. The relationship between (s), F-value,Cross-validation r^2 (q^2) and predictive residual sum of squares (PRESS). Cross-validation was calculated using leave-one-out (LOO) technique over 2 random trials with F to leave and F to enter being 2 in F stepping to include the most

significant variables in generating the QSAR model.

2.3 Cross-validation

Cross-validation is a popular technique used to test the reliability of QSAR models. In this study, leave-one-out (LOO) technique was utilized to create a number of modified data sets by deleting the first row and its value predicted using the rest of the data. Likewise, each row is left in turn, so that the value of each row is predicted from all others.

2.4 Molecular Descriptors

Forty molecular descriptors were selected in the study: topological shape and connectivity indices, total dipole and lipole, molecular weight, h-bond donors,h-bond acceptors, logP and rotatable bondcounts. A semi-empirical molecular orbital package was used to calculate thermodynamic property like heat of formation and electrostatic properties like HOMO (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied

Molecular Orbital) components.

2.5 Predictive Ability of QSAR model

Predictive ability of the generated model was estimated externally by predicting the activities of validation set. This criterion may not be sufficient for a QSAR model to be truly predictive[17]. An additional condition for high predictive ability of QSAR model is based on external set cross-validation r^2, ($R^2_{cv,ext}$) and the regression of observed activities against predicted activities and vice versa for validation set, if the following conditions are satisfied[18].Calculations relating to $R^2_{cv,ext}$, R_0^2 and the slopes, *k* and *k'* are based on regression of observed values against predicted values and vice versa.

$$R^2_{cv,ext} > 0.5 \quad (1)$$

$$R^2 > 0.6 \quad (2)$$

$$(R^2 - R_0^2) / R^2 < 0.1 \text{ or } (R^2 - R_0'^2) / R^2 < 0.1 \quad (3)$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \quad (4)$$

2.6 Y-randomization

This test ensures the robustness of a QSAR model[19] and to assess the multiple linear regression models obtained by descriptor selection[20]. In y-randomization test, the dependent variable or y-data is randomly shuffled and a new QSAR model is developed keeping X-data intact. The new models are expected to have low $R2$ and $Q2$ values, which determine the statistical significance of the original model.

Table 2. Logarithmic molar concentration values of training and validation sets and descriptor values of the proposed QSAR model (Eq. 8)

ID	Activity log(1/IC_{50})	Training Set [b] log (1/ IC_{50})	Validation Set [c] Log (1/ IC_{50})	Total Lipole	Lipole Z Component	KAlpha2 index	6-membered aromatic rings	Rotatable Bonds	LUMO
1 [a]	-0.23		-0.394	19.906	-4.881	9.154	2	5	-0.910
2 [a]	-0.431		-0.472	20.285	-5.385	9.340	2	5	-0.905
3 [a]	-0.886		-0.981	21.600	3.085	9.383	2	5	-0.885
4	-0.954	-0.752		22.490	-3.570	9.562	2	5	-0.889
5	-0.903	-0.922		16.415	3.701	9.979	2	6	-0.897
6	-1.301	-1.166		27.389	5.609	9.955	2	5	-1.000
7	-0.279	-0.580		23.407	-5.599	9.383	2	5	-0.900
8	-1.456	-1.187		24.581	3.118	9.615	2	5	-0.875
9	-0.778	-1.007		23.598	2.777	9.178	2	5	-0.861
10	-1.243	-1.234		20.466	2.165	10.004	2	6	-0.80

									5
11	-1.628	-1.220		26.038	2.613	9.587	2	5	-0.865
12	-1.515	-1.070		24.846	1.037	9.408	2	5	-0.855
13a	-1.459		-1.411	28.326	3.205	9.979	2	5	-0.886
14	-1.544	-1.355		21.553	1.782	10.637	2	7	-0.803
15	-1.053	-1.175		21.035	3.062	9.383	2	5	-0.811
16	-1.528	-1.322		32.211	3.530	11.037	3	6	-0.865
17	-1.057	-0.854		30.110	1.962	9.794	3	5	-0.859
18	-0.845	-1.246		20.586	2.928	10.231	2	6	-0.847
19	-1.041	-1.322		19.134	-2.081	11.472	2	8	-0.794
20	-1.057	-1.351		18.971	-1.883	11.472	2	8	-0.785
21	-1.845	-1.497		26.511	-2.201	11.083	2	7	-0.799
22	-0.519	-0.638		18.081	-2.539	9.485	2	6	-0.839
23	-0.38	-0.900		27.671	-2.401	10.323	2	6	-0.962
24	-0.204	-0.323		16.206	-2.971	11.952	2	10	-1.063
25	0.18	0.045		38.120	-21.711	12.216	2	10	-1.131
26	-0.903	-0.820		27.448	5.135	11.304	2	9	-1.092
27	-0.914	-1.162		18.328	1.880	11.442	1	9	-1.042
28	-0.914	-0.891		23.086	2.945	11.442	1	11	-1.093
29	0.523	0.237		15.462	-9.420	11.757	2	10	-1.118
30	0.699	0.429		38.125	-27.091	12.216	2	10	-1.177

31	-0.519	0.032		36.924	-21.541	12.424	2	10	-1.148
32	-0.255	0.075		15.425	-13.068	12.638	2	11	-1.072
33a	-0.954		-1.063	21.740	-11.868	12.563	2	9	-0.850
34a	-1.389		-1.081	5.881	-0.332	10.208	1	7	-0.845
35	-1.301	-1.143		12.108	-1.999	11.562	2	8	-0.803
36	-0.954	-1.343		19.879	0.714	10.720	2	5	-0.894
37	-0.792	-1.306		13.839	-3.765	12.359	2	7	-0.901
38	-1.699	-1.957		17.552	-9.589	12.844	1	8	-0.820
39	-1.258	-1.190		8.401	-5.347	10.714	1	5	-0.928
40	-1.699	-1.591		3.026	2.421	11.133	1	5	-0.924
41	-1.447	-1.262		9.714	-7.971	11.342	1	6	-0.908
42	-1.653	-1.548		3.120	-1.597	12.391	1	8	-0.892
43	-1.4	-0.937		25.927	-22.148	13.202	2	8	-0.910
44	-1.602	-1.504		3.282	2.177	11.737	1	7	-0.932
45	-1.954	-1.621		4.854	-2.314	12.603	1	8	-0.902
46	-0.826	-0.852		20.113	-17.776	12.112	2	6	-0.908
47	-1.255	-1.125		16.936	-13.444	12.758	2	7	-0.892
48	-0.778	-0.887		20.803	-16.085	12.112	2	6	-0.930
49	-0.826	-0.973		10.788	-8.581	12.801	2	7	-0.972
50	-1.086	-0.873		27.407	-24.006	12.384	2	6	-0.907
51	-1.029	-1.259		19.834	-10.226	12.136	2	6	-0.89

								1
52	-0.806	-1.207	10.721	-6.161	12.345	2	6	-0.918
53	-1.477	-0.997	10.673	-6.020	12.136	2	6	-0.964
54	-1.428	-1.186	10.453	-5.633	12.345	2	6	-0.932
55	-0.602	-1.198	27.193	-16.613	12.384	2	6	-0.913
56	-1.442	-1.368	23.283	-15.301	12.708	2	7	-0.829
57	-1.845	-1.866	31.953	-17.031	14.475	2	9	-0.855

[a] Validation set molecules.

[b] Calculated values from Equation 8

[c] Predicted values from Equation 8

3. RESULTS AND DISCUSSION

3.1 Complete Data set

Multivariate regression analysis with F stepping (F to enter and F to leave being 2) and cross-validation by leaving-out-one row, to test the predictive power, resulted in inertia moments, lipole components, shape flexibility and 6-membered rings as the most significant descriptors. Equation 5 represents the linear QSAR model from a complete set of 72 inhibitors.

log ($1/IC_{50}$)= 0.78127909*Inertia Moment 1 Size

+ 1.0273268*Inertia Moment 1 Length

- 0.19030687*Total Lipole

- 0.12560831*Lipole X component

- 0.22934534*Lipole Z component

- 0.81264067*Shape Flexibility

- 0.67134702*Randic Topological ndex

- 0.1585972*6-membered aliphaticrings

- 0.83904165

r =0.839, r^2 = 0.705, q^2 =0.60, F = 18.798, n = 72, s = 0.398 (5)

3.2 Outlier Detection

The criterion for removing outliers is based on Relative Error calculation and Extent of Extrapolation.

3.3 Relative Error calculation

This method was employed to calculate the relative error (Eq. 6) of all compounds in the data set. From Table 3, it cannot be stated that the model predicted wrongly for the highlighted compounds, instead it can be emphasized that the model prediction led to a high relative error for compounds 8, 12, 15, 29-30, 36-38, 42, 44-45 and 52 (Table 3) and hence these compounds should be excluded from the study as they influence the outcome in a significant manner.

Relative Error= Residual Value / Actual Value(6)

Table 3. Relative error calculation on complete data set. Compounds 5 and 40 are regarded as outliers based on extent of extrapolation graph. Remaining compounds 8, 12, 15, 29-30, 36-38, 42, 44-45 and 52 are disregarded from analysis.

S. No.	Compound No.	Actual Value	Predicted Value	Residual Value	Relative Error
1	2_2a.mol	-0.230	-0.391	0.161	-0.701
2	2_2b.mol	-0.431	-0.662	0.231	-0.536
3	2_2c.mol	-0.886	-0.916	0.030	-0.033
4	2_2d.mol	-0.954	-0.796	-0.158	0.166
5	2_2e.mol	-2.114	-1.202	-0.912	0.431
6	2_2f.mol	-0.903	-0.913	0.010	-0.011
7	2_2g.mol	-1.301	-1.051	-0.250	0.192
8	2_2h.mol	0.201	-0.478	0.679	**3.376**

9	2_2i.mol	-0.279	-0.475	0.196	-0.702
10	2_2j.mol	-1.456	-1.060	-0.396	0.272
11	2_2l.mol	0.222	-0.471	0.693	**3.120**
12	2_2m.mol	0.000	-0.325	0.325	-
13	2_3a.mol	-0.778	-0.818	0.040	-0.052
14	2_3b.mol	-1.243	-1.275	0.032	-0.025
15	2_3c.mol	0.155	-0.615	0.770	**4.966**
16	2_3d.mol	-1.628	-1.037	-0.591	0.363
17	2_3e.mol	-1.515	-0.964	-0.551	0.364
18	2_3f.mol	-1.459	-1.532	0.073	-0.050
19	2_3g.mol	-1.544	-1.224	-0.320	0.207
20	2_3h.mol	-1.053	-0.913	-0.140	0.133
21	2_3i.mol	-1.528	-1.414	-0.114	0.074
22	2_3j.mol	-1.057	-0.765	-0.292	0.276
23	2_3k.mol	-0.845	-0.900	0.055	-0.065
24	2_3l.mol	-1.041	-1.442	0.401	-0.385
25	2_3m.mol	-1.057	-1.319	0.262	-0.248
26	2_3n.mol	-1.845	-1.390	-0.455	0.247
27	2_8a.mol	-0.519	-0.275	-0.244	0.470
28	2_8b.mol	-0.380	-0.117	-0.263	0.691
29	2_8c.mol	0.377	-0.162	0.539	**1.429**
30	3_12a.mol	-0.204	-0.417	0.213	**-1.045**
31	3_12b.mol	-0.204	-0.210	0.006	-0.031
32	3_12c.mol	0.180	0.360	-0.180	-0.999
33	3_12d.mol	-0.903	-0.869	-0.034	0.038
34	3_12e.mol	-0.914	-0.912	-0.002	0.003
35	3_12f.mol	-0.914	-1.089	0.175	-0.191
36	3_13.mol	0.046	-0.281	0.327	**7.107**
37	3_14.mol	0.699	-0.132	0.831	**1.188**
38	3_15a.mol	-0.230	0.462	-0.692	**3.009**
39	3_15b.mol	0.523	0.506	0.017	0.033
40	3_15c.mol	0.699	1.156	-0.457	-0.654
41	3_15d.mol	-0.519	-0.367	-0.152	0.292
42	3_15e.mol	-0.079	-0.569	0.490	**-6.197**
43	3_16.mol	-0.255	-0.208	-0.047	0.185
44	3_17.mol	0.046	0.409	-0.363	**-7.895**
45	3_3a.mol	0.398	-0.172	0.570	**1.432**
46	4a_1.mol	-1.544	-1.133	-0.411	0.266
47	4aa_1.mol	-0.954	-1.128	0.174	-0.183
48	4b_1.mol	-1.389	-0.974	-0.415	0.299

49	4c_1.mol	-1.301	-1.227	-0.074	0.057
50	4d_1.mol	-0.954	-1.072	0.118	-0.124
51	4e_1.mol	-0.792	-0.921	0.129	-0.162
52	4f_1.mol	-0.643	-1.443	0.800	**-1.245**
53	4g_1.mol	-1.699	-1.869	0.170	-0.100
54	4h_1.mol	-1.258	-1.333	0.075	-0.060
55	4i_1.mol	-1.699	-1.319	-0.380	0.224
56	4j_1.mol	-1.447	-1.063	-0.384	0.265
57	4k_1.mol	-1.653	-1.495	-0.158	0.096
58	4l_1.mol	-1.400	-1.136	-0.264	0.189
59	4m_1.mol	-1.602	-1.527	-0.075	0.047
60	4n_1.mol	-1.954	-1.260	-0.694	0.355
61	4o_1.mol	-0.826	-1.042	0.216	-0.262
62	4p_1.mol	-1.255	-1.482	0.227	-0.181
63	4q_1.mol	-0.778	-0.849	0.071	-0.091
64	4r_1.mol	-0.826	-1.000	0.174	-0.211
65	4s_1.mol	-1.086	-0.971	-0.115	0.106
66	4t_1.mol	-1.029	-1.051	0.022	-0.021
67	4u_1.mol	-0.806	-1.286	0.480	-0.595
68	4v_1.mol	-1.477	-0.918	-0.559	0.379
69	4w_1.mol	-1.428	-1.345	-0.083	0.058
70	4x_1.mol	-0.602	-1.151	0.549	-0.912
71	4y_1.mol	-1.442	-1.114	-0.328	0.228
72	4z_1.mol	-1.845	-2.068	0.223	-0.121

3.4 Extent of Extrapolation

Outliers should be removed in order to obtain the best statistical result[21]. After employing relative error calculation, the data set was selected to plot extent of extrapolation graph plotted using MedCalc software

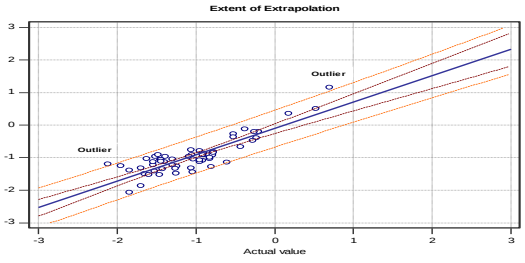


Fig 1: Extent of extrapolation graph displaying compounds (5)and (40) as outliers

Middle Dark line: regression line

Dotted lines: 95% confidence level

Solid lines: 95% prediction level

From the above graph, it is evidenced that all values lie within 95% prediction levels, whereas compounds (5) and (40) fall outside the region. After extrapolation of regression line further, it was observed that the predicted activities for some of the target compounds would fall within 95% prediction level.

3.5 QSAR Model

A new QSAR model was attempted by dividing the set as a 57 molecule training set and a 6 molecule validation set (Table 2) based on hierarchical clustering after rejecting outliers from the data set. More specifically, the selection of molecules in the training set was made according to the biological action and molecular structure; so that representatives of a wide range of structures with different substituents and activity were included. The distribution of

activity values for the validation set follows the similar distribution of the activity values for the training set. The results obtained from the multiple linear regression procedure with varied number of descriptors are shown in Table 4 with their statistics. Table 5 represents the predictive ability of all newly generated models. Given below are a set of 5 different models obtained and are statistically significant (Table 4).

Table 4. Descriptor data and statistical values of model equations.

Descriptor	**Coefficient**				
	Model-1	**Model-2**	**Model-3**	**Model-4**	**Model-5**
Total Lipole	-0.014	-0.027	-0.028	+0.065	
Lipole Z Component	-0.041	-0.047	-0.046		-0.042
Kier ChiV2 (path) index	-0.701	-	-		
Kier ChiV3 (cluster) index	+1.845	-	-		
Balaban Topological index	-3.984	-	-		
Number of Cl Atoms	-0.341	-	-		
6-membered aliphatic rings	-0.390	-	-		
H-bond Donors	+0.376	-	-		+0.223
KAlpha2 index	-	-0.408	-0.404		
6-membered aromatic rings	-	+0.548	+0.594	+0.208	+0.234
Rotatable Bonds	-	+ 0.153	+0.147	+0.249	+0.221
LUMO	-	-2.84	-3.017	-3.735	-3.560
Dipole Moment Z Component				-0.042	+0.061
Shape Flexibility				-0.763	-0.730
H-bond Acceptors				+0.212	+0.193
HOMO				+0.262	+0.264
Constant	+7.916	-0.795	-1.012	+0.216	-0.188
Statistics					
r	0.858	0.868	0.864	0.895	0.899
r^2	0.737	0.755	0.747	0.802	0.809
q^2	0.602	0.798	0.934	0.936	0.879
F	14.75	22.61	21.76	21.36	19.31
n	51	51	51	51	51
PRESS	4.307	4.035	3.915	3.338	3.188
s	0.663	0.302	0.298	0.281	0.278
No. of Descriptors	8	6	6	8	9
Equation No.	7	8	9	10	11

Table 5. Predictive ability of validation sets for all 5 equations obtained as models.

Var[a]	$R^2_{cv,ext}$ (q^2)	R^2	*k*	*k'*	Eq[b]	Eq[c]
8	0.602	0.737	1.040	0.960	0.003	0.0003
6	0.798	0.755	0.989	1.010	0.092	0.027
6	0.934	0.747	1.014	0.985	0.007	0.003
8	0.936	0.802	1.074	0.988	0.064	0.019
9	0.879	0.809	0.906	1.103	0.038	0.008

[a] number of significant variables

[b] $(R^2 – R_0^2) / R^2$

[c] $(R^2 – R_0'^2) / R^2$

3.6 FIT Kubinyi function

To define the statistical quality of activity prediction, the number of variables that enter in a QSAR model are compared by FIT Kubinyi function (Eq. 9), a criteria closely related to F value was proven to be useful[22].

The main feature of the F value is its sensitivity to changes in k, if k is small and its lower sensitivity if k is large. The FIT criterion has a low sensitivity towards changes in k values, as long as they are small numbers, and a substantially increasing sensitivity for large k values [23].

FIT = $R^2 (n – k – 1) / (n + k^2) (1 – R^2)$ (9)

Where *n* is the number of compounds in training set and *k* is the number of variables in the QSAR equation. The best model will be the one that posses a high value of this function.

Table 6. Statistical parameters of the regression models obtained for all QSAR models.

Eq No.	r^2	*k*	n	FIT
7	0.737	8	51	1.026
8	**0.751**	**6**	**51**	**1.558**
9	0.747	6	51	1.500
10	0.802	8	51	1.485
11	0.809	9	51	1.316

According to the statistical values of the models reported in Table 6, we choose the model with six variables since this showed high FIT than others. The observed, calculated and predicted values of the statistically significant six parameter QSAR model (Eq.8)is presented inTable 4.

Equation 8 accounts for the significant correlation of descriptors with biological activity and displayed good internal predictivity as shown by q^2 value of 0.798 and was able to explain 75.5 % variance of inhibitory activities of derivatives. Observed verses predicted values of molecules in training and validation set are shown graphically in Figure 2a.

The proposed QSAR model Eq. 8 illustrated the predictive ability of Eqs. 1-4 and depicted graphically in Figures 2b and 2c.

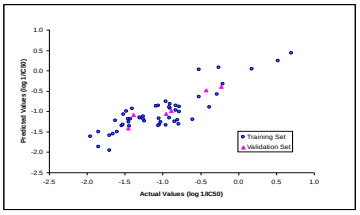


Fig 2a : Observed and predicted values of molecules in training and validation set

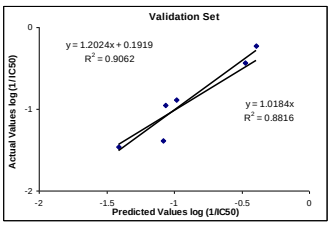


Fig 2b: Regression plot between observed vs. predicted values of compounds from validation set justifying the predictive ability of QSAR model Eq.8

The model was further validated by applying the y-randomization test. As the low R^2 and Q^2 values indicate that the results obtained in our original model (Eq. 8) are not due to chance correlation.

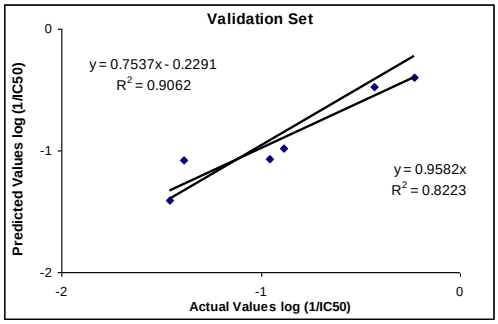


Fig 2c: Regression plot between predicted vs.observed values of compounds from validation set justifying the predictive ability of QSAR model *Eq.8*

The model was further validated by applying the y-randomization test. As a model selection included F-stepping ,random shuffles of the dependent as well as independent variables were performed and the result are shown in Table 7.The low R2 and Q2 values indicate that the results obtained in our original model(Eq.8) are not due to chance correlation

Table 7. R^2 and Q^2 values after several y-randomization tests

Iteration	R^2	Q^2		Iteration	R^2	Q^2
1	0.28	0.12		11	0.15	0.41
2	0.25	0.22		12	0.24	0.22
3	0.15	0.33		13	0.26	0.10
4	0.38	0.25		14	0.12	0.38
5	0.11	0.12		15	0.35	0.24
6	0.14	0.11		16	0.28	0.41
7	0.29	0.33		17	0.18	0.34
8	0.21	0.38		18	0.44	0.14
9	0.28	0.31		19	0.17	0.11
10	0.10	0.45		20	0.09	0.15

Table 8. Inter-correlation between descriptors utilized in generating QSAR model eq. 8

	Total Lipole	Lipole Z Component	KAlpha2 index	6-membered aromatic rings	Rotatable Bonds	LUMO
Total Lipole	1	-0.297	-0.095	0.605	0.104	-0.218
Lipole Z Component	-0.297	1	-0.660	-0.087	-0.389	0.337
KAlpha2 index	-0.095	-0.660	1	-0.158	0.629	-0.267
6-membered aromatic rings	0.605	-0.087	-0.158	1	-0.188	0.097
Rotatable Bonds	0.104	-0.389	0.629	-0.188	1	-0.572
LUMO	-0.218	0.337	-0.267	0.097	-0.572	1

Inter-correlation between descriptors utilized in developing QSAR model (Eq. 8) is given in Table 8. From the table it is well known that the descriptors appeared in the final model are not highly inter-correlated.

The generated QSAR model (Eq. 8) indicates that a high value of LUMO energy contributes negatively to the activity.Electron-withdrawing substituent, such as halogens, lower the energy of LUMO. Molecules with low-lying LUMOs have greater tendencies to accept electrons than those with high-energy LUMOs. As LUMO increases (relative to other molecules) the molecule becomes less reactive[20]. Thus, designing analogs with electron-withdrawing substituents should improve activity. From equation 8 it can be observed that an increase in 6-membered aromatic rings would enhance MMP-13 inhibition. Most of the studied inhibitors contain

linear aliphatic groups and have the tendency to rotate the compound within the active site region. Positive correlation of rotatable bonds term with activity indicates more the number of such groups in the molecule, more active it would be. On the other hand, reduction of lipophilic character on the compounds would increase bioactivity.

4. CONCLUSION

Out of five QSAR models generated on the data set with reasonable chemical diversity demonstrated eq.8 to be a promising method and the six descriptors [Total Lipole, Lipole Z Component, KAlpha2 index,6-membered aromatic rings, Rotatable Bonds and LUMO] were found to be important in enhancing MMP-13 inhibition. The predictive ability of the model and the internal and external validation procedures illustrated the accuracy on one hand and offered a useful alternative to the time consuming experiments for MMP-13 inhibition, on the other. Considering the advantages of QSAR techniques, this work indicates that accurate predictions can be achieved with computational analysis in a reliable manner and we plan to extend the procedure presented here to perform docking of MMP-13 inhibitors, and moreover, experimental evaluation of their biological activities would help in designing more potent compounds.

5. REFERENCES

[1] PD Brown, R Giavazzi, 1995 *Matrixmetalloproteinase inhibition: A review of anti-tumour activity. Ann. Oncol. 6:967-974.*

[2] J Gross, CM Lapiere. 1962 *Collagenolytic activity in amohibian tissues - a tissue culture assay. Proc Natl Acad Sci USA 48:1014-1022*

[3] P Vihinen, R Ala-aho, VM Kahari 2005 *Matrix metalloproteinases as therapeutic targets in cancer. Curr. Cancer Drug Targets 5:203-220.*

[4] Christian K. Engel 2005 *Structural Basis for the Highly Selective Inhibition of MMP-13. Chemistry & Biology 12:181–189*

[5] H Matter,M Schudok, 2004 *Recent advances in the design of matrix metalloproteinase inhibitors. Curr. Opin. Drug Disc. Devel. 7:513–535.*

[6] ZR Wasserman 2005 *Chem. Biol. 12:143.*

[7] Yvette M Fobian, John N Freskos, Thomas E Barta, Louis J Bedell, Robert Heintz, Darren J Kassab, James R Kiefer, Brent V Mischke, John M Molyneaux, Patrick Mullins, Grace E Munie, Daniel P Becker 2011 *MMP-13 selective alpha-sulfone hydroxamates: Identification of selective P10 amides. Bioorganic & Medicinal Chemistry Letters 21:2823–2825.*

[8] Johnson AR, Pavlovsky AG, Ortwine, DF, Prior F, Man CF, Bornemeier DA, Banotai CA, Mueller WT, McConnell P, Yan C, Baragi V, Lesch C, Roark WH, Wilson M, Datta K, Guzman R, Han HK, Dyer RD (2007) J. *Biol Chem. 282:27781.*

[9] Stephen A Kolodziej, Susan L Hockerman, Gary A DeCrescenzo, Joseph J McDonald, Debbie A Mischke, Grace E Munie, Theresa R Fletcher, Nathan Stehle, Craig Swearingen, Daniel P Becker 2010 *MMP-13 selective isonipecotamide a-sulfone hydroxamates. Bioorganic & Medicinal Chemistry Letters 20:3561–3564.*

[10] Stephen A Kolodziej, Susan L Hockerman, Terri L Boehm, Jeffery N Carroll, Gary A DeCrescenzo, Joseph J McDonald, Debbie A Mischke, Grace E Munie, Theresa R Fletcher, Joseph G Rico, Nathan W Stehle, Craig Swearingen, Daniel P Becker 2010 *Orally bioavailable dual MMP-1/MMP-14 sparing, MMP-13 selective a-sulfone hydroxamates. Bioorganic & Medicinal Chemistry Letters 20:3557–3560*

[11] Fobian YM, Freskos JN, Barta TE, Bedell LJ, Heintz R, Kassab DJ, Kiefer JR, Mischke BV, Molyneaux JM, Mullins P, Munie GE, Becker DP (2011) *MMP-13 selective alpha-sulfone hydroxamates: identification of selective P1' amides. Bioorganic & Medicinal Chemistry Letters 21:2823–2825.*

[12] Sharma BK, Singh P 2013 *Chemometric Descriptor Based QSAR Rationales for the MMP-13 Inhibition Activity of Non-Zinc-Chelating Compounds. Med chem 3:168-178.*

[13] G.B. Kulkarni, A. Siva Reddy, Konda Ramadevi, Devarapalli Kezia and M. Narasimha Rao 2009 *3D-QSAR Studies of Matrix Metalloproteinase-13 Inhibitors. Rasayan J Chem. 2:407-414.*

[14] BK Sharma, P Singh and YS Prabhakar 2013 *QSAR rationale of Matrix metalloproteinase inhibition activity in a class of carboxylic acid based compounds. British Journal of Pharmaceutical Research 3:697-721.*

[15] Michael Ferna´ndez and Julio Caballero 2007 *QSAR modeling of matrix metalloproteinase inhibition by N-hydroxy-a-phenylsulfonylacetamide derivatives. Bioorganic & Medicinal Chemistry 15:6298–6310.*

[16] Xi L, Li S, Yao X, Wei Y, Li J, Liu H, Wu 2014 *In Silico Study Combining Docking and QSAR Methods on a Series of Matrix Metalloproteinase 13 Inhibitors. Arch. Pharm. doi: 10.1002/ardp.201400200.*

[17] Golbraikh A, Tropsha A 2002 *Beware of q2!. J. Mol. Graph. Model. 20:269-276.*

[18] Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O 2006 A novel *QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromenes. Bioorg. Med. Chem. 14:6686-6694.*

[19] Kim D, Hong S, Lee D 2006 *The Quantitative Structure-Mutagenicity Relationship of Polycylic Aromatic Hydrocarbon Metabolites. Int. J. Mol. Sci. 7:556-570.*

[20] Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O 2006 *Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques. J. Comput. Aided Mol. Des. 20:83-95.*

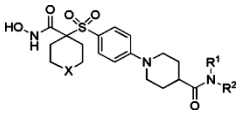
[21] Kubinyi H 1994 *Variable selection in QSAR studies. II. A Highly Efficient Combination of Systematic Search and Evolution. Quant.* Struct. Act. Relat. 13:393-401.

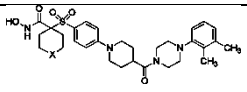
[22] Kubinyi H 1994 Variable selection in QSAR studies. I. An evolutionary algorithm. Quant. Struct. Act. Relat. 13:285-294.

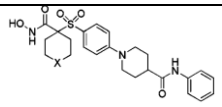
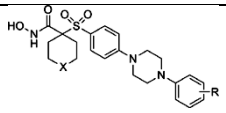
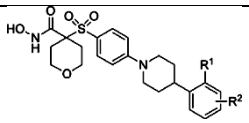
[23] Hall LH, Mohney B, Kier LB 1991 *The* Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. J. Chem. Inf. Comput. Sci. 31:76-*82*

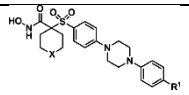
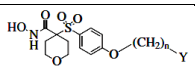
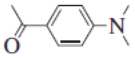
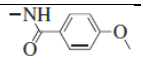
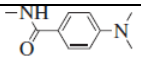
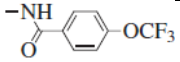
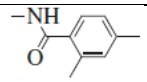
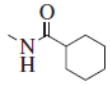
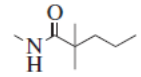
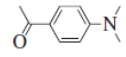
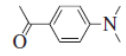
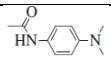
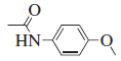
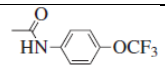
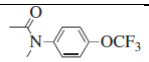
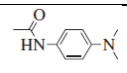
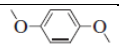
6. APPENDIX

Table 1. Structures and biological activities of α-sulfone hydroxamate derivatives as MMP-13 inhibitors.



ID	X	NR^1R^2	IC_{50} (nM)	Log $1/IC_{50}$
1	O	Allyl(methyl)amino	35.0	-1.544
2	O	Methyl(prop-2-ynyl)amino	24.5	-1.389
3	N-cyclopropyl	Benzyl(methyl)amino	20.0	-1.301
4	O	3,4-Dihydroisoquinolin-2(1H)-yl	9.0	-0.954
5	O	6,7-Dimethoxy-3,4-dihydroisoquinolin-2(1H)-yl	6.2	-0.792
6	O	3,5-Dimethylpiperidin-1-yl	4.4	-0.643
7	N-CH_2CH_2OMe	3,5-dimethylpiperidin-1-yl	50.0	-1.699
8	O	cis-2,6-Dimethylmorpholin-4-yl	18.1	-1.258
9	O	4-Acetylpiperazin-1-yl	50.0	-1.699
10	O	4-Isopropylpiperazin-1-yl	28.0	-1.447
11	O	4-(2-Methoxyethyl)piperazin-1-yl	45.0	-1.653
12	O	4-Phenethylpiperazin-1-yl	25.4	-1.405
13	O	4-(2-Hydroxyethyl)piperazin-1-yl	40.0	-1.602
14	O	4-(2-(Dimethylamino)ethyl)piperazin-1-yl	90.0	-1.954
15	O	4-(2-Fluorophenyl)piperazin-1-yl	6.7	-0.826
16	O	4-(2-Methoxyphenyl)piperazin-1-yl	18.0	-1.255
17	O	4-(4-Fluorophenyl)piperazin-1-yl	6.0	-0.778
18	O	4-(4-Acetylphenyl)piperazin-1-yl	6.7	-0.826
19	O	4-(2,4-Dimethylphenyl)piperazin-1-yl	12.2	-1.086
20	O	4-(Pyridin-2-yl)piperazin-1-yl	10.7	-1.029
21	O	4-(Pyrimidin-2-yl)piperazin-1-yl)	6.4	-0.806
22	O	4-(Pyridin-4-yl)piperazin-1-yl	30.0	-1.477
23	O	4-(Pyrazin-2-yl)piperazin-1-yl	26.8	-1.428
				
24	O	-	4.0	-0.602

25	N-Cyclopropyl	-	27.7	-1.442
26	NCH2CH2OMe	-	70	-1.845
				
27	NCH2CH2OMe	-	9.0	-0.954
				
	X	**R**	**IC_{50} (nM)**	**Log $1/IC_{50}$**
28	O	H	1.7	-0.230
29	O	2-F	2.7	-0.431
30	O	2-Me	7.7	-0.886
31	O	2-Cl	9.0	-0.954
32	O	2-MeO	130	-2.114
33	O	3-MeO	8.0	-0.903
34	O	3-CF_3	20	-1.301
35	O	4-MeO	0.63	0.201
36	O	4-Me	1.9	-0.279
37	O	2,4-diMe	28.6	-1.456
38	N-cPr	H	3.3	-0.519
39	N-cPr	4-CF_3	2.4	-0.380
				
	R^1	**R^2**	**IC_{50} (nM)**	**Log $1/IC_{50}$**
40	H	H	6.0	-0.778
41	MeO	H	17.5	-1.243
42	H	4-Cl	0.7	0.155
43	Cl	H	42.5	-1.628
44	Me	H	32.7	-1.515
45	CF_3	H	28.8	-1.459
46	EtO	H	35.0	-1.544
47	OH	H	11.3	-1.053
48	4-F-C_6H_4	H	33.7	-1.528
49	2,3-(CH=CH)	naphthyl	11.4	-1.057
50	Me	4-MeO	7.0	-0.845
51	MeO	4-diMeO	11.0	-1.041
52	MeO	5-diMeO	11.4	-1.057
53	MeO	5-iPr	70	-1.845

				
	X	**R^1**	**IC_{50} (nM)**	**Log $1/IC_{50}$**
54	O	CF_3	0.6	0.222
55	O	CL	1.0	0.000
56	cPr-N	OCH_3	0.42	0.377
				
	n	**Y**	**IC_{50} (nM)**	**Log $1/IC_{50}$**
57	4		0.4	0.398
58	3		1.6	-0.204
59	3		1.6	-0.204
60	3		0.66	0.180
61	3		8.0	-0.903
62	3		8.2	-0.914
63	3		8.2	-0.914
64	2		0.9	0.046
65	3		0.2	0.699
66	3		1.7	-0.230
67	3		0.3	0.523
68	3		0.2	0.699
69	3		3.3	-0.519
70	3		1.2	-0.079
71	4		1.8	-0.255
72	4		0.9	0.046