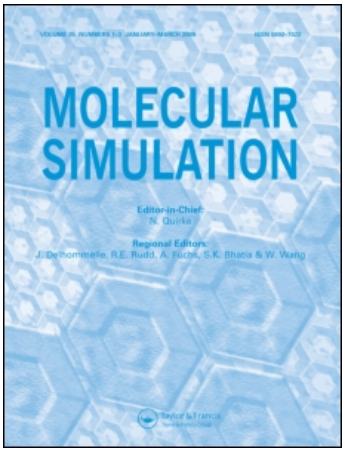


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

QSAR-based prediction of anti-HCV activity of thiourea derivatives

Ravichandran Veerasamy[a]
[a] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, AIMST University, Semeling, Kedah, Malaysia

Online publication date: 15 October 2010

To cite this Article Veerasamy, Ravichandran(2010) 'QSAR-based prediction of anti-HCV activity of thiourea derivatives', Molecular Simulation, 36: 11, 878 — 886
To link to this Article: DOI: 10.1080/08927022.2010.490399
URL: http://dx.doi.org/10.1080/08927022.2010.490399

Molecular Simulation
Vol. 36, No. 11, September 2010, 878–886

Taylor & Francis
Taylor & Francis Group

QSAR-based prediction of anti-HCV activity of thiourea derivatives

Ravichandran Veerasamy*

Department of Pharmaceutical Chemistry, Faculty of Pharmacy, AIMST University, Semeling 08100, Kedah, Malaysia

(*Received 18 March 2010; final version received 27 April 2010*)

Quantitative structure–activity relationship (QSAR) studies were performed on a series of thioureas to explore the physico-chemical parameters responsible for their activity against the hepatitis C virus (HCV)-infected AVa5 cell. The physico-chemical parameters were calculated using WIN CAChe 6.1. Multiple linear regression analysis, after the variables selection by factor analysis, was performed to derive QSAR models which were further evaluated for their statistical significance and predictive power by internal and external validation. The developed QSAR model had the correlation coefficient (R) = 0.928 and cross-validated squared correlation coefficient (Q^2) = 0.751. The selected significant QSAR model indicates that hydrophobicity, dielectric energy, valence connectivity index (order 1), conformational minimum energy and highest occupied molecular orbital of the whole molecule play an important role in the anti-HCV activity of thioureas.

Keywords: QSAR; anti-HCV; FA-MLR; hydrophobicity; thioureas

1. Introduction

Hepatitis C virus (HCV) is a blood-borne virus which was identified in 1989 [1]. Six genotypes with several subtypes within each genotype have been described for this member of the Flaviviridae family [2]. It is estimated that more than 170 million people worldwide are infected by HCV [3]. Most of those with acute hepatitis C develop a chronic infection. Chronic hepatitis C is often asymptomatic but may lead to liver cirrhosis and hepatocellular carcinoma within the next decade [4]. Liver failure related to HCV infection is, in many countries, the leading cause of liver transplantation [5]. There is currently no vaccine or a direct antiviral agent for HCV [6]. The standard of care for treating HCV is a combination of pegylated interferon and ribavirin. However, this therapy in patients with genotype 1 is not very successful and is associated with serious side effects [7]. Therefore, there is an urgent need to discover and develop new anti-HCV agents that are more effective, have less cytotoxicity and are better tolerated by patients.

Gonzalez-Diaz et al. discussed in detail topological indices [8], connectivity indices [9] and MARCH-INSIDE approach [10] in their reviews. Prado-Prado et al. [11,12] reported quantitative structure–activity relationship (QSAR) studies on antiviral drugs vs. different viral species. Our research group developed a few QSAR models to predict anti-HIV activity [13–17], HIV-1 integrase inhibitory activity [18] and HIV-1 protease inhibitory activity [19]. In continuation of such efforts, in the present work, I have performed QSAR analysis for anti-HCV activity of thiourea derivatives [20], using the molecular modelling software WIN CAChe 6.1 and statistical software SPSS 11.5.

The purpose of the present study is to investigate the physico-chemical parameters responsible for the anti-HCV activity of thioureas, explore the correlation between them and obtain more information for designing novel thiourea derivatives with lesser cytotoxicity and potent anti-HCV activity.

There is high structural diversity in the selected series of thiourea derivatives; therefore, this selected series of compounds is used in the present QSAR study. QSAR analysis has been performed and a QSAR model established to guide further structural optimisation and to predict the anti-HCV activity and physico-chemical properties of clinical drug candidates.

2. Materials and methods

All the molecular modelling studies, reported herein, were performed using WIN CAChe 6.1 (Product of Fujitsu Private Limited, Tokyo, Japan, http://www.cachesoftware.com/contacts/japan.shtml) and the QSAR models were executed with SPSS 11.5 statistical software.

2.1 Biological data

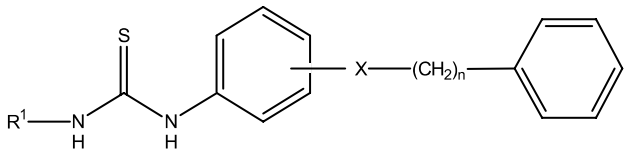
In the present work, the anti-HCV activity of 45 thiourea compounds (Tables 1 and 2) reported by Kang et al. [20]

*Email: phravi75@rediffmail.com

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.490399
http://www.informaworld.com

Table 1. Structure of thioureas.



Compounds 1–17

S. no	Compound	R^1	X	EC_{50}
1	1	H	O	0.494
2	5	H	O	0.561
3	6	H	O	0.335
4	7	H	O	0.231
5	8	H	O	0.159
6a	9	H	O	0.048
7	10	H	O	0.047
8	11	H	O	0.097
9	12	H	O	1.012
10	13	H	O	0.197
11	14	H	NH	0.143
12	15	Me	O	0.822
13	16	Et	O	0.685
14a	17	*n*-Pr	O	0.708
15	18	Ph	O	31.497
16	19	CH_2Ph	O	7.238
17	20	CH_2CH_2Ph	O	13.388

Note: a, test-set compounds.

has been used. All the anti-HCV activities used in the present study were expressed as $pEC_{50} = -\log EC_{50}$, where EC_{50} is the micromolar concentration of the compounds producing 50% reduction in the HCV-infected AVa5 cell growth (stated as the means of at least two experiments). The compounds which did not show confirmed anti-HCV activity in the above-cited literature have not been taken for our study.

2.2 Optimisation of the molecule structure

All the 45 compounds (39 compounds in the training set and 6 in the test set; the test- and training-set compounds were chosen manually so that low, moderate and high activity compounds were present in approximately equal proportions in both sets) were built on the workspace of the molecular modelling software WIN CAChe 6.1. The energy minimisation was done by the geometry optimisation of molecules using MM2 (molecular mechanics) followed by the semi-empirical PM3 method available in the MOPAC module until the root mean square gradient value became smaller than 0.001 kcal/mol Å. The most stable structure for each compound was generated and used for calculating various physico-chemical descriptors such as thermodynamic properties, steric energy and electronic properties.

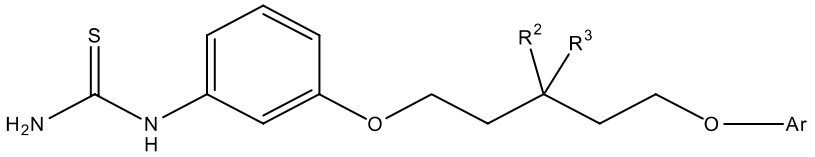
2.3 Calculation of descriptors

The physico-chemical properties were calculated on a project leader file of the modelling software WIN CAChe 6.1. In the present study, the calculated descriptors were conformational minimum energy (CME), zero-order connectivity index ($^0\chi$), first-order connectivity index ($^1\chi$), second-order connectivity index ($^2\chi$), dipole moment (DM), size of the smallest ring (SSR), heat of formation (HF) at its current geometry after the optimisation of the structure, dipole vector X, Y or Z (DVX, DVY or DVZ), molecular weight (MW), dielectric energy (DE), octanol–water partition coefficient ($\log P$), squared partition coefficient ($\log P^2$), molar refractivity (MR), shape index order one (SI1), shape index order two (SI2), shape index order three (SI3), zero-order valence connectivity index ($^0\chi_v$), first-order valence connectivity index ($^1\chi_v$), second-order valence connectivity index ($^2\chi_v$), highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) and solvent accessible surface area (SASA). The physico-chemical parameters involved in the selected QSAR model are given in Table 3 and full physico-chemical parameters data will be provided on request.

2.4 Development and validation of QSAR models

The SPSS 11.5 software was used to generate QSAR models. Factor analysis (FA) was used as the data-preprocessing step to identify the important predictor variables contributing to the response variable and to avoid collinearities among them, although classical approach of the multiple linear regression technique was used as the final statistical tool for developing QSAR relations. In a typical FA procedure, the data matrix is first standardised, correlation matrix and subsequently reduced correlation matrix are constructed, eigenvalue problem is then solved and the factor pattern is obtained from the corresponding eigenvectors. The principal objectives of the FA are to display multidimensional data in a space of lower dimensionality with minimum loss of information (explaining $>95\%$ of the variance of the data matrix) and to extract the basic features behind the data with the ultimate goal of interpretation and/or prediction. FA was performed on the data-set(s) containing biological activity and all descriptor variables, which were to be considered. The factors were extracted by the principal component method, and then rotated by VARIMAX rotation to obtain Thurston's simple structure [21]. The simple structure is characterised by the property that as many variables as possible fall on the coordinate axes when presented in a common factor space, so that the largest possible number of factor loadings becomes zero. This is done to obtain a numerically comprehensive picture of the relatedness of the variables. Only the variables with non-zero loadings in such factors where biological activity also has non-zero

Table 2. Structure of thioureas.



Compounds 18–45

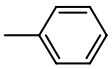
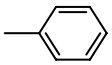
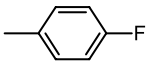
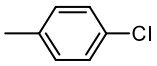
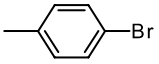
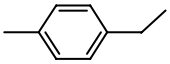
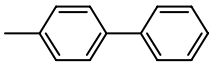
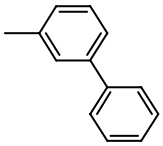
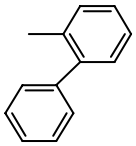
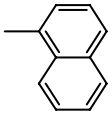
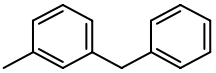
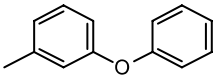
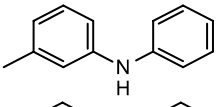
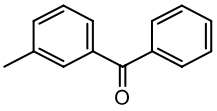
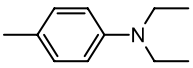
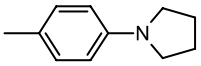
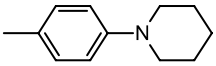
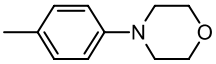
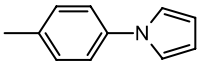
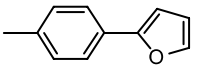
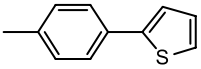
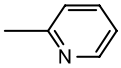
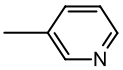
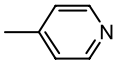
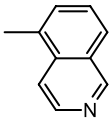
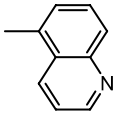
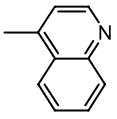
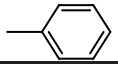
S. no	Compound	R^2	R^3	Ar	EC_{50}
18a	25	H	H		0.059
19	26	Me	H		0.104
20	27	H	H		0.072
21	28	H	H		0.072
22	29	H	H		0.072
23	30	H	H		0.218
24	32	H	H		0.191
25	33	H	H		0.214
26	34	H	H		0.097
27	35	H	H		0.058
28a	36	H	H		0.058
29	37	H	H		0.107
30	38	H	H		0.177
31	39	H	H		0.218
32	40	H	H		0.594

Table 2 – *continued*

S. no	Compound	R^2	R^3	Ar	EC_{50}
33	41	H	H		0.913
34	42	H	H		0.407
35	43	H	H		0.485
36	44	H	H		0.255
37a	45	H	H		0.136
38	46	H	H		0.121
39	47	H	H		0.279
40	48	H	H		1.134
41	49	H	H		2.992
42	50	H	H		0.481
43	51	H	H		0.250
44	52	H	H		0.411
45a	60	Me	Me		0.105

Note: a, test-set compounds.

loading were considered important in explaining the variance of the activity. Furthermore, variables with non-zero loadings in different factors were combined in a multivariate equation.

Statistical measures used were as follows: n, number of compounds in regression; R, correlation coefficient; R^2, squared correlation coefficient; F-test (Fischer's value) for statistical significance; SEE, standard error of estimation; Q^2, cross-validated correlation coefficient.

Internal validation was carried out by the leave-one-out (LOO) method using the statistical software STATISTICA. The cross-validated correlation coefficient, Q^2, was calculated using the following equation:

$$Q^2 = 1 - \frac{\text{PRESS}}{\sum_{i=1}^{N}(y_i - y_\text{m})^2},$$

$$\text{PRESS} = \sum_{i=1}^{N}(y_{\text{pred},i} - y_i)^2,$$

where y_i is the activity for training-set compounds, y_m is the mean observed value, corresponding to the mean of the values for each cross-validation (CV) group, and $y_{\text{pred},i}$ is the predicted activity for y_i. The predictive ability of the

Table 3. Selected physico-chemical parameters of thioureas.

S. no	CME	DE	HOMO	$^{1}\chi_{v}$	log *P*	log P^2
1	−8.770	−1.187	−8.668	5.541	3.626	13.148
2	−8.861	−1.080	−8.640	5.979	3.721	13.846
3	−5.407	−1.209	−8.683	6.479	3.973	15.785
4	−6.719	−1.192	−8.624	6.979	4.369	19.088
5	−6.958	−1.178	−8.616	7.479	4.765	22.705
6a	−3.831	−1.135	−8.610	7.979	5.162	26.646
7	−1.998	−1.190	−8.615	8.479	5.558	30.891
8	−4.152	−1.277	−8.646	8.979	5.954	35.450
9	−4.607	−0.985	−8.700	7.985	5.162	26.646
10	−6.317	−1.184	−8.551	7.979	5.162	26.646
11	−1.167	−1.266	−8.564	8.090	4.988	24.880
12	−1.294	−1.071	−8.521	8.441	5.519	30.459
13	−1.733	−1.026	−8.516	9.001	5.861	34.351
14a	−0.934	−0.982	−8.461	9.384	6.274	39.363
15	1.057	−0.987	−8.505	10.101	7.196	51.782
16	−12.852	−0.949	−8.508	10.559	7.295	53.217
17	−6.845	−1.027	−8.526	11.059	7.547	56.957
18a	0.588	−1.274	−8.641	8.119	4.460	19.892
19	5.160	−1.297	−8.647	8.513	4.791	22.954
20	0.186	−1.320	−8.657	8.218	4.600	21.160
21	−0.206	−1.294	−8.654	8.596	4.978	24.780
22	0.333	−1.272	−8.660	9.011	5.252	27.584
23	15.156	−1.364	−8.639	8.642	4.207	17.699
24	−6.531	−1.350	−8.652	10.190	6.145	37.761
25	−5.451	−1.349	−8.643	10.190	6.145	37.761
26	−11.859	−1.307	−8.636	10.196	6.145	37.761
27	−12.266	−1.316	−8.585	9.529	5.462	29.833
28a	−6.967	−1.373	−8.640	10.647	6.541	42.785
29	−4.909	−1.363	−8.635	10.348	5.889	34.680
30	5.618	−1.403	−8.548	10.440	5.712	32.627
31	8.100	−1.523	−8.653	10.644	5.682	32.285
32	17.461	−1.404	−8.177	10.300	5.410	29.268
33	19.583	−1.426	−8.423	10.385	5.050	25.503
34	25.750	−1.408	−8.249	10.885	5.446	29.659
35	29.523	−1.476	−8.408	10.463	4.382	19.202
36	8.838	−1.395	−8.548	9.769	4.698	22.071
37a	8.284	−1.375	−8.516	9.675	4.628	21.418
38	12.466	−1.364	−8.654	10.554	4.971	24.711
39	21.451	−1.525	−8.629	7.978	3.847	14.799
40	23.360	−1.455	−8.663	7.968	3.148	9.910
41	23.805	−1.463	−8.678	7.968	3.148	9.910
42	12.196	−1.433	−8.673	9.379	4.151	17.231
43	13.709	−1.429	−8.678	9.389	4.550	20.703
44	17.165	−1.478	−8.682	9.389	4.550	20.703
45a	6.596	−1.318	−8.678	8.826	5.224	27.290

Note: a, test-set compounds.

selected model was also confirmed by external R^2, R^2_{CVext}, $r^2 - r_0^2/r^2$, $r^2 - r_0'^2/r^2$, k, k', r^2_{m} and R^2_p [22,23]. The formula to calculate the R^2_{CVext} is:

$$R^2_{\mathrm{CVext}} = 1 - \frac{\sum_{i=1}^{\mathrm{test}} (y_{\mathrm{exp,i}} - y_{\mathrm{pred,i}})^2}{\sum_{i=1}^{\mathrm{test}} (y_{\mathrm{exp,i}} - \bar{y}_{\mathrm{tr}})^2},$$

where $\bar{y}_{\mathrm{tr}}$ is the averaged value for the dependent variable for the training set.

The robustness of a QSAR model was ensured by *Y*-randomisation test. In this technique, the dependent variable vector is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The random number simulation has shown that the probability of chance correlation as well as the degree of inflation of internal figures of merit is considerable in small data-sets and in the case of low object-to-variable ratios. Therefore, it became a common practice in QSAR studies to assess the probability of chance correlation with the help of a permutation test (scrambling) when variable selection is applied [24–27]. The new QSAR models (after several repetitions) are expected to have low R^2 and Q^2 values. An acceptable QSAR model cannot be obtained for the specific

modelling method and data, in the event of the opposite happening [23].

3. Results and discussion

It was observed that six factors explained the data matrix to the extent of 95% (which were obtained by using anti-HCV activity data and the associated physico-chemical parameters). Table 4 shows that the biological activity is highly loaded with factors 4 (highly loaded in DM), 5 (highly loaded in SSR, HF and DVZ) and 2 (highly loaded in DE, LUMO, $\log P$, $\log P^2$ and CME), moderately loaded with factor 3 (highly loaded in HOMO and CME), and poorly loaded with factors 1 and 6 (highly loaded in MR, MW, $\log P$, $\log P^2$, all in the order of connectivity, valence connectivity and shape index). Based on the FA, the following equation was derived with six variables:

$$\begin{aligned} \mathrm{pEC}_{50} = &-14.094(\pm4.474) - 0.117(\pm0.063)^1\chi_v \\ &- 2.953(\pm0.606)\mathrm{DE} + 2.742(\pm0.341) \\ &\log P - 0.254(\pm0.032)\log P^2 - 0.601(\pm0.489) \\ &\mathrm{HOMO} - 0.024(\pm0.007)\mathrm{CME} \quad n = 39, \\ &R = 0.928, R^2 = 0.862, \; R^2_{\mathrm{adj}} = 0.836, \\ &\mathrm{SEE} = 0.253, \; F = 33.17, \; P < 0.001, \\ &Q^2 = 0.751, \; S_{\mathrm{PRESS}} = 0.342, \; \mathrm{SDEP} = 0.293. \end{aligned} \tag{1}$$

The values given in parentheses are 95% confidence intervals of the regression coefficients. The selected descriptors were CME, DE, $^1\chi_v$, $\log P$, $\log P^2$ and HOMO. These properties were capable of elucidating 86.2% and predicting 83.6% of the total variance of the anti-HCV activity data. There was no inter-correlation between the selected descriptors. The calculated anti-HCV activity data by Equation (1) are given in Table 5. The residual values of observed and calculated, and observed and predicted anti-HCV activities are graphically shown in Figures 1 and 2.

The selected model exhibited good internal and external predictivity. The proposed QSAR model is predictive as it satisfies the conditions $R^2_{\mathrm{CVext}} = 0.870 > 0.5$ and $R^2_{\mathrm{pred}} = 0.757 > 0.6$; $r^2_{\mathrm{m}} = 0.753 > 0.5$; $R^2_P = 0.663 > 0.5$; $k = 1.023$; $k' = 0.935$ $(0.85 < k$ or $k' < 1.15)$; $r^2 - r^2_0/r^2 = -0.243$; $r^2 - r'^2_0/r^2 = -0.158$ $(r^2 - r^2_0/r^2$ or $r^2 - r'^2_0/r^2 < 0.1)$. We obtained low R^2 and Q^2 values (maximum R^2 value is 0.271 and maximum Q^2 is 0.116) for the *Y*-randomisation test, when only the finally selected model is subjected to the randomisation test.

Two hundred runs of the permutation test were carried out for LOO-CV. Approximately three Q^2 values were equal to or even larger than the original Q^2 values of 0.751. Hence, the probability of chance correlation is about 1.5% (3/201) of the results. Consequently, the LOO-CV model should not be trusted, because Baumann [28] reported that the conservative maximum probability of chance correlation should be less than 1%. The

Table 4. Factor loadings of the variables after VARIMAX rotation.

Variable	1	2	3	4	5	6	Communality
pEC_{50}	−0.130	−0.428	0.210	0.600	0.430	−0.121	0.904
CME	0.245	−0.592	−0.568	−0.133	−0.248	0.340	0.928
$^0\chi$	0.985	−0.093	0.079	0.013	−0.041	−0.022	0.988
$^1\chi$	0.980	−0.068	0.094	−0.013	−0.077	−0.052	0.982
$^2\chi$	0.963	−0.133	0.125	0.026	−0.088	−0.086	0.976
DM	−0.177	0.152	−0.454	0.607	0.291	−0.283	0.815
DVX	−0.072	0.736	−0.161	0.422	−0.172	0.263	0.849
DVY	0.294	−0.533	−0.025	−0.184	0.255	0.029	0.870
DVZ	−0.132	0.610	0.071	0.450	−0.416	0.382	0.915
DE	−0.291	0.873	0.113	−0.179	0.026	−0.082	0.899
HF	−0.297	0.544	0.328	−0.193	−0.316	−0.396	0.786
HOMO	0.438	0.259	−0.793	−0.092	0.084	−0.192	0.940
LUMO	0.043	0.827	−0.176	−0.153	0.172	0.132	0.887
SI1	0.991	−0.066	0.058	0.017	0.003	0.038	0.991
SI2	0.969	0.058	0.027	−0.023	0.064	0.158	0.972
SI3	0.818	0.136	0.022	0.053	0.218	0.377	0.881
SASA	0.978	−0.145	0.029	0.069	−0.024	0.018	0.985
$^0\chi_v$	0.996	−0.015	0.027	0.017	−0.021	−0.018	0.993
$^1\chi_v$	0.992	0.034	0.005	0.021	−0.043	−0.044	0.990
$^2\chi_v$	0.967	0.035	0.037	0.080	−0.055	−0.068	0.952
$\log P$	0.643	0.662	0.259	0.068	0.150	−0.137	0.964
MR	0.989	0.055	0.065	0.018	−0.044	−0.067	0.992
MW	0.972	−0.140	0.090	0.008	−0.035	−0.033	0.974
SSR	−0.216	0.172	0.215	−0.245	0.651	0.302	0.817
$\log P^2$	0.629	0.678	0.271	−0.009	0.116	−0.110	0.954
Variance	0.480	0.200	0.084	0.052	0.050	0.038	0.904

Table 5. Calculated and predicted (LOO) anti-HCV activity of training and test sets.

S. no	pEC_{50} (μM) Observed anti-HCV activity (μM)[a]	Calculated anti-HCV activity	Predicted anti-HCV activity (LOO)
Training set			
1	0.306	0.281	0.275
2	0.251	−0.015	−0.083
3	0.475	0.532	0.539
4	0.636	0.670	0.673
5	0.799	0.799	0.799
7	1.328	0.912	0.870
8	1.013	1.067	1.083
9	−0.005	0.420	0.580
10	0.706	0.771	0.777
11	0.845	0.945	0.957
12	0.085	0.428	0.474
13	0.164	0.208	0.216
15	−1.498	−0.876	−0.682
16	−0.860	−1.139	−1.261
17	−1.127	−1.238	−1.309
19	0.983	0.885	0.880
20	1.143	0.858	0.844
21	1.143	0.990	0.981
22	1.143	1.054	1.047
23	0.662	0.522	0.511
24	0.719	0.752	0.756
25	0.670	0.741	0.749
26	1.013	0.637	0.601
27	1.237	0.752	0.729
29	0.971	0.696	0.667
30	0.752	0.714	0.711
31	0.662	1.039	1.110
32	0.226	0.168	0.125
33	0.040	0.510	0.571
34	0.390	0.199	0.123
35	0.314	0.132	0.084
36	0.593	0.505	0.494
38	0.917	0.743	0.720
39	0.554	0.719	0.763
40	−0.055	−0.302	−0.400
41	−0.476	−0.272	−0.191
42	0.318	0.479	0.506
43	0.602	0.809	0.829
44	0.386	0.931	0.982
Test set			
6	1.319	0.985	
14	0.150	0.030	
18	1.229	1.109	
28	1.237	1.549	
37	0.866	1.013	
45	0.979	1.224	

[a] The anti-HCV activity EC_{50} values (in μM) were converted into $-\log EC_{50}$ (pEC_{50}, in μM).

distributions of Q^2 values for leave-33%-out (L33%O) CV are less than their respective original values, indicating a low probability of chance correlation. Therefore, LOO-CV without constraints is most susceptible to chance correlation and overfitting. Constraints on the LOO-CV, on the other hand, effectively decrease the risk of chance correlation in this case.

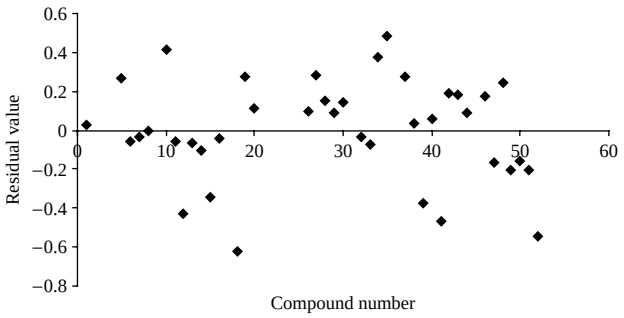


Figure 1. Residual plot between experimental and calculated anti-HCV activities of thiourea derivatives (training set).

Finally, the applicability domain was established for Equation (1), determining the leverage values for each compound. Figure 3 shows the Williams plot, i.e. the plot of standardised residuals (*y*-axis) vs. leverages (*x*-axis) for each compound of the training set. From this plot, the applicability domain is established inside a squared area within ±2.5 standard deviations and a leverage threshold $h^* = 0.538$ ($h^* = 3p'/n$, p' being the number of model parameters +1 and n, the number of compounds). As can be seen from Figure 3, all compounds of the training and test sets are inside the square area. For future predictions, predicted anti-HCV activity data must be considered reliable only for those chemicals that fall within the applicability domain on which the model was constructed [29].

Moreover, it is not possible to use the reported QSAR models to predict the activity of any type of molecules vs. HCV. The applicability domain of the derived QSAR models is only the different substituted thiourea derivatives. However, it is very important to point out an eventual QSAR model disappointment: activity cliffs [30]. It is possible because similar molecules can show

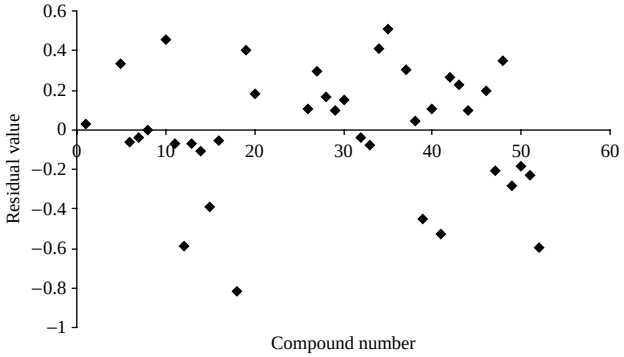


Figure 2. Residual plot between experimental and predicted (LOO) anti-HCV activities of thiourea derivatives (training set).

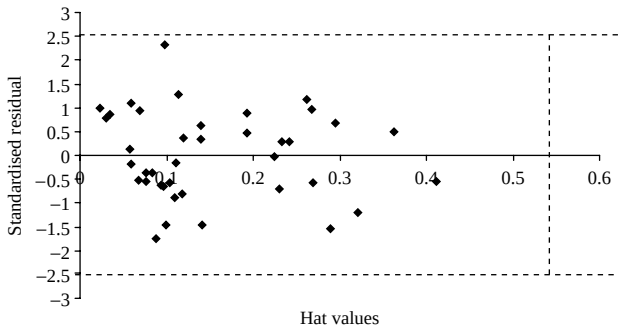


Figure 3. Williams plot: plot of standardised residuals (*y*-axis) vs. leverages (hat values; *x*-axis) for each compound.

significantly different biological activities. For these molecules, activities are often mispredicted, even when the overall predictivity of the models is high.

In Equation (1), the contribution of HOMO is very small (the difference between the coefficient and the confidence interval is very small), even though this parameter was included in the final model. The reason is that the statistical significance and predictability of the equation was reduced badly when this descriptor was removed. The negative contribution of HOMO to anti-HCV activity indicates that the compounds having high HOMO energy lead to less potent anti-HCV compounds. HOMO is the highest energy level in the molecule that contains electrons. Molecules with low HOMOs are more capable of accepting electrons. Therefore, electronegative (electron-accepting) groups are conducive to anti-HCV activity of thioureas. The negative coefficient of DE showed that the increase in the dielectric constant of the compounds is not favourable for the anti-HCV activity of thiourea derivatives.

The negative contribution of $^{1}\chi_{v}$ to anti-HCV activity indicates that highly branched groups and long-chain groups are not favourable for anti-HCV activity. The molecules may not be fitted into the binding site when the molecules having highly branched groups increase the valence connectivity index value of the molecules. The presence of higher alkyl groups or aromatic groups may lead to potent anti-HCV compounds. This is supported by compounds **6**–**10** with a chain length of two, three, four, five and six carbons that increased the potency to about 1.5- to 10-fold with respect to compound **1**. However, going to longer seven-carbon homologue (**11**) was not favourable for anti-HCV activity.

The negative contribution of CME to anti-HCV activity indicates that the thiourea compounds having high CME values are not conducive to anti-HCV activity. The QSAR showed a parabolic relationship of anti-HCV activity with the log *P*. The negative sign of $\log P^{2}$ indicates that highly hydrophobic groups are not good for improving the anti-HCV activity of thiourea derivatives. This is supported by replacement of the phenyl group (**32**) in the *para*-position of the biphenyl moiety with diethyl amino (**40**), pyrrolidinyl (**41**), piperidinyl (**42**), morpholinyl (**43**) and pyrrolyl (**44**) groups causing a loss of activity.

4. Conclusion

In this study, it was possible to obtain a QSAR model for a set of 45 thiourea derivatives that have the capability of inhibiting the *in vitro* anti-HCV activity. The LOO-CV and L33%O-CV methods, the *Y*-randomisation technique, permutation test and the external validation indicated that the model is significant, robust and has good internal and external predictability. The resulting findings can be helpful in the development and optimisation of new thiourea derivatives as anti-HCV inhibitors. The reported QSAR models in this study might be used for predicting only the anti-HCV activity of thiourea derivatives.

Acknowledgements

The author is thankful to the AIMST University for providing the necessary facilities to carry out this work.

References

[1] Q.L. Choo, G. Kuo, A.J. Weiner, L.R. Overby, D.W. Bradley, and M. Houghton, *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*, Science 244 (1989), pp. 359–362.
[2] P. Simmonds, E.C. Holmes, T.A. Cha, S.W. Chan, F. McOmish, B. Irvine, E. Beall, P.L. Yap, J. Kolberg, and M.S. Urdea, *Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region*, J. Gen. Virol. 74 (1993), pp. 2391–2399.
[3] D. Lavanchy, R.H. Purcell, F.B. Hollinger, C. Howard, A. Alberti, M.C. Kew, and G.M. Dusheiko, *Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium*, J. Viral. Hepat. 6 (1999), pp. 35–47.
[4] B.W. Dymock, P.S. Jones, and F.X. Wilson, *Novel approaches to the treatment of hepatitis C virus infection*, Antiviral Chem. Chemother. 11 (2000), pp. 79–96.
[5] N. Boyer and P. Marcellin, *Pathogenesis, diagnosis and management of hepatitis C*, J. Hepatol. 32 (2000), pp. 98–112.
[6] S. Abrignani, M. Houghton, and H.H. Hsu, *Perspectives for a vaccine against hepatitis C virus*, J. Hepatol. 31 (1999), pp. 259–263.
[7] J.H. Hoofnagle and L.B. Seeff, *Peginterferon and ribavirin for chronic hepatitis C*, New Engl. J. Med. 355 (2006), pp. 2444–2451.
[8] H. Gonzalez-Diaz, S. Vilar, L. Santana, and E. Uriarte, *Medicinal chemistry and bioinformatics – Current trends in drugs discovery with networks topological indices*, Curr. Top. Med. Chem. 10 (2007), pp. 1015–1029.
[9] H. Gonzalez-Diaz, Y. Gonzalez-Diaz, L. Santana, F.M. Ubeira, and E. Uriarte, *Proteomics, networks and connectivity indices*, Proteomics 8 (2008), pp. 750–778.
[10] H. Gonzalez-Diaz, F. Prado-Prado, and F.M. Ubeira, *Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach*, Curr. Top. Med. Chem. 8 (2008), pp. 1676–1690.
[11] F.J. Prado-Prado, F. Borges, E. Uriarte, L.G. Peraecopyz-Montoto, and H. Gonzalez-Diaz, *Multi-target spectral moment: QSAR for antiviral drugs vs. different viral species*, Anal. Chim. Acta. 651 (2009), pp. 159–164.

[12] F.J. Prado-Prado, O. Martinez de la Vega, E. Uriarte, F.M. Ubeira, K.C. Chou, and H. Gonzalez-Diaz, *Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug–drug complex networks*, Bioorg. Med. Chem. 17 (2009), pp. 569–575.
[13] V. Ravichandran and R.K. Agrawal, *Predicting anti-HIV activity of PETT derivatives: CoMFA approach*, Bioorg. Med. Chem. Lett. 17 (2007), pp. 2197–2202.
[14] V. Ravichandran, P.K. Jain, V.K. Mourya, and R.K. Agrawal, *QSAR study on some arylsulfonamides as anti-HIV agents*, Med. Chem. Res. 16 (2007), pp. 342–351.
[15] V. Ravichandran, V.K. Mourya, and R.K. Agrawal, *QSAR study of novel 1,1,3-trioxo [1,2,4]-thiadiazine (TTDs) analogues as potent anti-HIV agents*, Arkivoc XIV (2007), pp. 204–212.
[16] V. Ravichandran, B.R. Prashanthakumar, S. Sankar, and R.K. Agrawal, *Predicting anti-HIV activity of 1,3,4-thiazolidinone derivatives: 3D QSAR approach*, Eur. J. Med. Chem. 44 (2009), pp. 1180–1187.
[17] V. Ravichandran, B.R. Prashanthakumar, S. Sankar, and R.K. Agrawal, *Comparative molecular similarity indices analysis for predicting anti-HIV activity of phenyl ethyl thiourea (PET) derivatives*, Med. Chem. Res. 17 (2008), pp. 1–11.
[18] V. Ravichandran, S. Shalini, K. Sundram, and S.A. Dhanaraj, *QSAR study of substituted 1,3,4-oxadiazole naphthyridines as HIV-1 integrase inhibitors*, Eur. J. Med. Chem. (2010), (inpress), DOI: 10.1016/j.ejmech.2010.02.062.
[19] V. Ravichandran, V.K. Mourya, and R.K. Agrawal, *Prediction of HIV-1 protease inhibitory activity of 4-hydroxy-5,6-dihydropyran-2-ones: QSAR study*, J. Enzym. Inhib. Med. Chem. (2008), (in press).
[20] I.J. Kang, L.W. Wang, C.C. Lee, Y.C. Lee, Y.S. Chao, T.A. Hsu, and J.H. Chern, *Design, synthesis, and anti-HCV activity of thiourea compounds*, Bioorg. Med. Chem. Lett. 19 (2009), pp. 1950–1955.
[21] R. Franke and A. Gruska, *Principal component and factor analysis,* in *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, ed., VCH, Weinheim, 1995, pp. 113–163.
[22] A. Tropsha, P. Gramatica, and V.K. Gombar, *The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models*, QSAR Comb. Sci. 22 (2003), pp. 1–9.
[23] P.P. Roy, S. Paul, I. Modra, and K. Roy, *On two novel parameters for validation of predictive QSAR models*, Molecules 14 (2009), pp. 1660–1701.
[24] M. Hahn and D. Rogers, *Receptor surface models. 2. Application to quantitative structure–activity relationships studies*, J. Med. Chem. 38 (1995), pp. 2091–2102.
[25] S.S. So and M. Karplus, *Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation*, J. Med. Chem. 40 (1997), pp. 4347–4359.
[26] H. Kubinyi, F.A. Hamprecht, and T. Mietzner, *Three-dimensional quantitative similarity–activity relationships (3D QSiAR) from SEAL similarity matrices*, J. Med. Chem. 41 (1998), pp. 2553–2559.
[27] K. Baumann, *Distance profile: A transnationally and rotationally invariant 3D structure descriptor capturing steric properties of molecules*, Quant. Struct. Act. Relat. 21 (2002), pp. 507–519.
[28] K. Baumann, *Cross-validation as the objective function for variable selection techniques*, Trends Anal. Chem. 22 (2003), pp. 395–406.
[29] M. Vighi, P. Gramatica, F. Consolaro, and R. Todeschini, *QSAR and chemometric approaches for setting water quality objectives for dangerous chemicals*, Ecotoxicol. Environ. Safety 49 (2001), pp. 206–220.
[30] G. Maggiora, *On outliers and activity cliffs – Why QSAR often disappoints*, J. Chem. Inf. Model. 46 (2006), pp. 1535–1535.