



QSAR STUDIES OF COMARIN DERIVATIVES AS POTENT HIV – 1 INTEGRASE INHIBITORS

Enas Basher Ahmed^{1*} Ahmed E. M. Saeed¹

Department of Chemistry, Collage of Science, Sudan University of Science and Technology, Khartoum, Sudan.

***Corresponding Author: Enas Basher Ahmed**

Department of Chemistry, Collage of Science, Sudan University of Science and Technology, Khartoum, Sudan.

Article Received on 19/01/2018

Article Revised on 09/02/2018

Article Accepted on 01/03/2018

ABSTRACT

A series of coumarins derivatives (containing 25 compounds) as HIV-1 integrase inhibitors were subjected to quantitative structure-activity relationship (QSAR) analysis. For building the regression models two different variable selection approaches namely, genetic function approximation multiple linear regression (MLR) were used and compared to predict the HIV-1IN inhibition activity. Based on prediction, the best validation model for inhibition activity with squared correlation coefficient of determination (R^2) = 0.9127, 0.9128 and 0.907. The results of the present study are used for designing more potent HIV-1IN inhibitors.

KEYWORDS: Coumarins, HIV-1 integrase inhibition, QSAR, Biological activity.

1. INTRODUCTION

Acquired immunodeficiency syndrome (AIDS), reported in 1981,^[1] is a fatal disorder resulting from a chronic persistent infection by the human retrovirus, human immunodeficiency virus (HIV).^[2] Today, AIDS is considered as one of the most devastating diseases faced by mankind, with an estimation of 34 million people living with HIV worldwide at the end of 2010 according to Joint United Nations Programme on HIV/AIDS.^[3] Up to now, successful chemotherapy has not been developed. Currently, Reverse Transcriptase (RT) and Protease (PT) inhibitors are the main targets for the majority of available drugs for HIV treatment. However, toxicity and rapid development of resistance to these inhibitors are the main issues related to the current therapy.^[4] Therefore, the development of new anti-HIV agents with varied structure and mechanisms of action is of great importance. HIV-1 integrase (HIV-1IN) is a very attractive and unexplored target for developing of new anti-HIV drugs as it plays a vital role in replication cycle and it has no cellular counterpart.^[5-7]

Various compounds exhibit HIV-1IN inhibitory activity, including lignanoides,^[8] aurointricarboxylic acids,^[9] dicaffeoyl quick acids and analogues diary sulfones. Unfortunately, all of stated inhibitors have the 1,2-dihydroxy (catechol) moiety, separated by an appropriate linker, so all of them have significant cytotoxicity because of catechol moiety autoxidation to reactive quinone species.^[10] To overcome this problem, a series of coumarin derivatives which do not contain catechol

functionality but possess good HIV-1IN inhibition activity was synthesized.^[12]

The importance of quantitative structure-reactivity relationship (QSAR) studies in modern drug design is well established since QSAR can make the early prediction of activity-related characteristics of drug candidates and can eliminate molecules with undesired properties.^[13] The main goal of QSAR approach is to correlate the biological activity of a series of compounds with the calculated molecular properties in terms of descriptors.^[14] Thousands of molecular descriptors are used in QSAR studies for the purpose of encoding molecules chemical and structural features,^[15,16] with great importance of topological descriptors calculated on the basis of molecular graphs. As a part of ongoing effort the present work is aimed to derive some statistically significant QSAR models for coumarin derivatives to correlate anti-HIV-1 IN activity to its physicochemical properties.^[17]

Coumarins are an important group of naturally occurring compounds widely distributed in the plant kingdom. However, they have been produced synthetically for many years for commercial use. Coumarin and its derivatives have been attracting great interest because of their importance in synthetic organic chemistry. Many products which contain the subunit of coumarin exhibit useful and diverse biological activity. These compounds find their application in pharmaceuticals, fragrances, agrochemicals, and insecticides.^[18,19]

2. MATERIALS AND METHODS

A dataset of coumarin derivatives containing 25 compounds with well-defined activity,^[19] was selected for QSAR study. The compounds which do not have well defined activity were excluded from dataset. The biological activity data in the form of IC₅₀ (molar concentration of the drug leading to 50% inhibition of enzyme Integrase) value in μM (micromoles) were converted into negative logarithmic dose in moles (pic₅₀) for QSAR Analysis.

2.1 Molecular Modeling and Generation of Molecular Descriptors

The molecular modeling study was performed using Chemo- Office version 10.00 software. Structure of all compounds was drawn using chemo draw Ultra module of the program and transferred to Chem3D Ultra module to create the three-dimensional (3D) structure. All the drawn structures compound was selected and used for calculating various physicochemical parameters like log

p, density, polarizability, molar refractivity, surface tension.

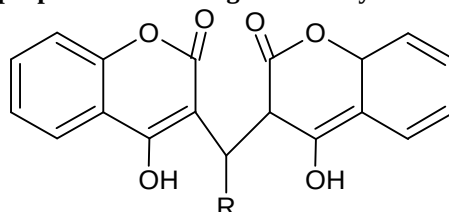
2.2 Variable Selection and Model Generation

Although many molecular descriptors are available, only a subset of them is statistically significant in terms of correlation with biological activity. Thus, it is very important to address the variable selection method for deriving the optimal QSAR model. The model was generated by using multiple linear regression (MLR) to select the best possible variables as well as for the generation of QSAR models.

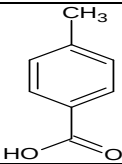
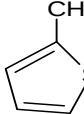
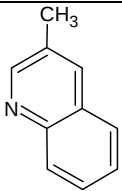
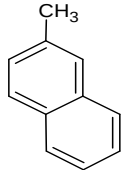
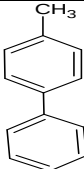
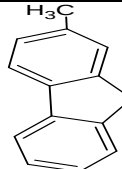
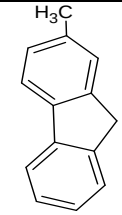
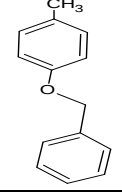
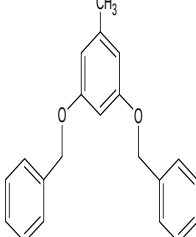
2.3 MLR method

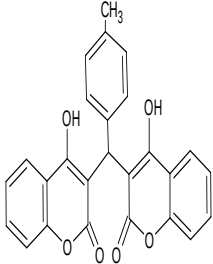
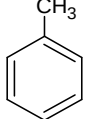
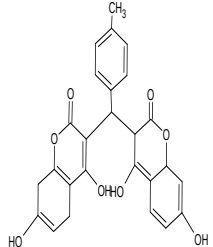
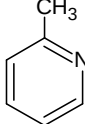
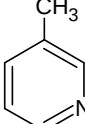
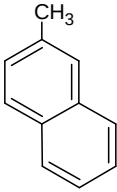
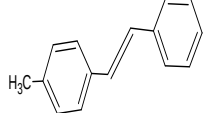
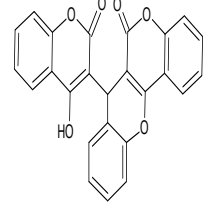
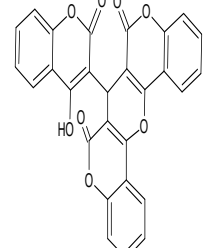
In order to establish a correlation between physicochemical parameters as independent variables and pic₅₀ as dependent variable employing sequential multiple linear regression analysis method, dataset was transferred to the statistical program SPSS Inc. version 16.

Table 1: Structure, physicochemical properties and biological activity of coumarin derivatives.



Comp No	R	pic ₅₀	Log p	Pola Cm3	Surface T (dyne cm ⁻¹)	Density (g/cm ³)	MR (cm ³)
1		4.3665	5.38	-4.36E-23	79.9	1.517	109.86
2		3.8927	4.65	-4.43E-23	88.1	1.585	111.74
3		3.752	4.35	-4.69E-23	82.4	1.558	118.42
4		4.055	5.49	-4.92E-23	76.8	1.47	124.17
5		4.301	5.11	-4.61E-23	88.4	1.612	116.41

6		4.3187	5.06	-4.63E-23	88.9	1.605	116.79
7		4	5.06	-4.29E-23	86.8	1.601	108.24
8		4.4685	4.54	-4.05E-23	82.7	1.58	102.16
9		4.5376	5.25	-4.99E-23	86.8	1.549	125.79
10		4.7212	6.61	-5.06E-23	80.9	1.511	127.7
11		4.8538	7.14	-5.33E-23	72.8	1.449	134.45
12		5	7.33	-5.45E-23	80.8	1.513	137.4
13		5.2596	8	-5.84E-23	77.2	1.449	147.25
14		5.0705	6.95	-5.59E-23	73.8	1.454	141.03
15		5.0969	8.9	-6.83E-23	70.3	1.415	172.19

16		5.6989	8.55	-7.67E-23	95.5	1.644	193.47
17	H	4.3344	3.43	-3.53E-23	103.1	1.748	89.15
18		4.7644	5.26	-4.50E-23	97	1.654	113.62
19		6.4317	5.15	-7.95E-23	121.6	1.81	200.46
20		4.0757	3.77	-4.43E-23	105.1	1.701	111.71
21		4.2076	3.77	-4.43E-23	105.1	1.701	111.71
22		5.3767	6.49	-5.21E-23	96.1	1.633	131.47
23		5.1549	7.88	-5.99E-23	89.7	1.553	151.02
24		3.9172	8	-7.72E-23	118.2	1.8	194.78
25		4.4473	6.75	-8.24E-23	123.1	1.85	207.95

Selection of training and test sets is one of the most important steps in QSAR analysis. There must be sufficient structural diversity which could cover the complete range of variation in biological activity. The dataset was divided into a training set of 18 compounds (18, 19, 4, 17, 20, 6, 12, 22, 1, 13, 25, 16, 7, 11, 21, 26, 9, and 5) and test set of 8 compounds (2, 3, 8, 10, 14, 15, 23 and 24). After selection of training set and test set compounds, the dataset was further subjected to MLR analysis to predict the HIV-1IN inhibition activity of test set compounds using SPSS Inc. version 16.0.

2.4 Validation of QSAR Models

The QSAR models were developed by MLR methods and evaluated using the following statistical parameters: n (the number of compounds in regression), correlation coefficient of determination (r), r^2 , F , Kubinyi function (FIT), variance, SEE, P , statistical significance. Kubinyi function (FIT) is a closely related statistical parameter to F test. The greater the FIT value the better the linear equation. standard deviation of sum of square of difference between predicted and observed values (SPRESS) and standard deviation of error of prediction (SDEP). For reliability of the model, probable error of correlation (PE) was also calculated. If the value of correlation- correlation coefficient (r) is more than six times of PE, then the expression is good and reliable (23). Finally, the derived QSAR models were used for the prediction of the biological activity of the compounds in the test set and pred_r^2 was calculated for evaluating the prediction ability of the models.

3. RESULTS AND DISCUSSIONS

3.1. QSAR study

The inhibitory activities of studied compounds against HIV-1 IN are reported with the pic50 ($\log 1/C$) values. C is the molar concentration of the observed HIV-1 IN inhibitory activity.

Generally, the number of compounds should be greater than four times the number of descriptors in order to get a reliable model using multiple linear regression. In this study, we reduced down our pool of descriptors using principal components analysis for building multilinear QSAR models. The treatment started with the reduction in the number of molecular descriptors. If two descriptors inter correlated highly with each other, then only one of them was selected; descriptors with insignificant variance for the data set treated were also rejected. This helps to speed up the descriptor selection and reduce the probability of including unrelated descriptors by chance.

The correlation between the different physio-chemical descriptors as independent variable and the pic50 as dependent variable was determined. Some of the statistically significant models are given below and their statistical measures are listed:

3.1.1 Model 1

This model relates the biological activity to $\log p$ and molar refractivity with $R^2 = 0.9127$ for the training set is a very good indication of the fitness of the model to the data

$$\text{Pic50} = 4.258(\pm 0.394) + 0.569(\pm 0.119) \log p - 0.029(\pm 0.0030) \text{Mr cm}^3$$

$$N=8 \quad F = 37.078 \quad R^2 = 0.9127 \quad P \text{ value} = 0.006 \quad \text{mse} = 0.042$$

Model = significant coefficient's = significant.

3.1.2 Model 2

This model relates the biological activity to $\log p$ and Polarizability with $R^2 = 0.9128$ for the training set is a good indication of the fitness of the model to the data.

$$\text{Pic50} = 4.256(\pm 0.394) + 0.660(\pm 0.119) \log p + 7.264E22(\pm 15.592) \text{Pola}$$

$$N=8 \quad F = 13.681 \quad R^2 = 0.9128 \quad P \text{ value} = 0.002 \quad \text{mse} = 0.042$$

Model = significant coefficient's = significant.

3.1.3 Model 3

This model relates the reciprocal of inhibitory concentration ($1/\text{pic50}$) to $\log p$ and Polarizability with $R^2 = 0.907$ for the training set is a very good indication of the fitness of the model to the data.

$$1/\text{Pic50} = 0.237 (\pm 0.021) - 0.034(\pm 0.006) \log p - 3.789E21(\pm 8.180E20) \text{Pola}$$

$$N=8 \quad F = 23.499 \quad R^2 = 0.907 \quad P \text{ value} = 0.007 \quad \text{mse} = 0.000$$

Model = significant coefficient's = significant.

3.1.4 Model 4

This model relates the reciprocal of inhibitory concentration ($1/\text{pic50}$) to $\log p$ and molar refractivity with $R^2 = 0.907$ for the training set is a very good indication of the fitness of the model to the data.

$$1/\text{Pic50} = 0.236(\pm 0.103) - 0.034(\pm 0.005) \log p + 0.001(\pm 0.063) \text{Mr cm}^3$$

$$N=8 \quad F = 13.681 \quad R^2 = 0.907 \quad P \text{ value} = 0.006 \quad \text{mse} = 0.001$$

Model = significant coefficient's = significant.

3.1.5 Model 5

This model relates the reciprocal of inhibitory concentration ($1/\text{pic50}$) to $\log p$ and surface tension with $R^2 = 0.727$ for the training set is a very good indication of the fitness of the model to the data.

$$1/\text{Pic50} = 0.193 (\pm 0.047) - 0.012(\pm 0.004) \log p + 0.001(\pm 0.000) \text{ST}$$

$$N=8 \quad F = 19.215 \quad R^2 = 0.727 \quad P \text{ value} = 0.008 \quad \text{mse} = 0.000$$

Model = significant coefficient's = significant.

3.1.6 Model 6

This model relates the biological activity to $\log p$ and surface tension with $R^2 = 0.753$ for the training set is a very good indication of the fitness of the model to the data.

$$\text{Pic50} = 5.134(\pm 0.919) + 0.228(\pm 0.086) \log p - 0.025(\pm 0.009) \text{ST}$$

$$N=8 \quad F = 17.560 \quad R^2 = 0.753 \quad P \text{ value} = 0.006 \quad \text{mse} = 0.120$$

Model = significant coefficient's = significant.

3.1.7 Model 7

This model relates the biological activity to log p and density with $R^2 = 0.753$ for the training set is a very good indication of the fitness of the model to the data.

$$\text{Pic50} = 0.163(\pm 1.976) + 0.187(\pm 0.089) \log p - 3.127(\pm 1.208) D$$

N=8 F = 19.215 $R^2 = 0.753$ P value = 0.0075 mse = 0.1178

Model = significant coefficient's = significant.

3.1.8 Model 8

This model relates the reciprocal of inhibitory concentration (1/pic50) to log p and density with $R^2 = 0.725$ for the training set is a very good indication of the fitness of the model to the data.

$$1/\text{Pic50} = 0.041 (\pm 0.103) - 0.009(\pm 0.005) \log p + 0.158(\pm 0.063) D$$

N=8 F = 19.215 $R^2 = 0.725$ P value = 0.008 mse = 0.000

Model = significant coefficient's = significant.

Model 1, 2, 3 and 4 are the best models since it shows the best correlation coefficient of determination $R^2 = (0.9127, 0.9128, 0.907 \text{ and } 0.907)$. Further, smaller standard error of estimate, higher *F* value demonstrates satisfactory predictive ability of these models. These models were generated by regression analysis of HIV-1 IN inhibitory activity and have good correlation between biological activity and descriptor as indicated by R^2 , low standard deviation of the model demonstrates accuracy of the model. The model showed overall significance level better than 99%. The p-value are 0.002, 0.006, 0.007 and 0.006 (<0.01), which shows there are significant relationship between descriptor and biological activity.

Model 5, 6, 7 and 8 show correlation coefficient $R = (0.753, 0.727 \text{ and } 0.725)$ respectively and it were considered to be a poor predictor of activity compared to the models 1,2,3 and 4 since they possess lower correlation coefficient.

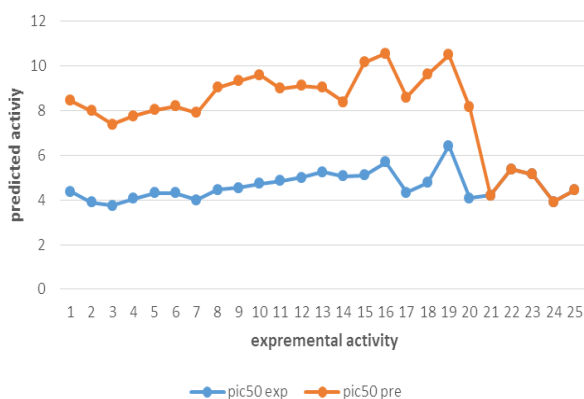


Figure 1: Scatter plot between the observed and predicted activity of training and test set (Integration inhibition activity).

From the scatter plot the trends lines of experimental activity and predictive activity are almost similar which indicate that the predictive activity is reliable and could be used to predict the activities of coumarin derivatives.

4. CONCLUSION

QSAR analysis was performed on a series of coumarins derivatives compounds using molecular modeling program Chemo- Office version 10.00. QSAR models were proposed for inhibitory activity of the coumarins derivatives using multiple regression analysis method. The selected models were checked for multi collinearity and correlation with F-test value. The predictive power of each model was validated with determination coefficient R^2 method and cross validation Q^2 method. The generated QSAR models were very informative as they showed statistical significance more than 99% and good prediction ability. From the prediction ability of dataset and training set it is confirmed that the present QSAR study holds true for different sets of compounds.

5. REFERENCES

- Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A. Pneumocystis carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men - Evidence of a New Acquired Cellular Immunodeficiency. N Engl J Med, 1981; 305(24): 1425-31.
- Koup RA, Merluzzi VJ, Hargrave KD, Adams J, Grozinger KJ. Inhibition of human immunodeficiency virus type 1 replication by the dipyridociazepinone. J Infect Dis, 1991; 163(5): 966-70.
- UNAIDS, 2011. World AIDS Day Report 2011.
- Erickson JW, Burt SK. Structural mechanisms of HIV drug resistance. Annu Rev Pharmacol Toxicol, 1996; 36: 545-71.
- Goldgur Y, Craigie R, Cohen GH, et al. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. Proc Natl Acad Sci, 1999; 96(23): 13040-3.
- Bhatt H, Patel P, Pannecouque C. Discovery of HIV-1 integrase inhibitors: pharmacophore mapping, virtual screening, molecular docking, synthesis, and biological evaluation. Chem Biol Drug Des, 2014; 83(2): 154-66.
- De Carvalho LL, Maltarollo VG, de Lima EF, Weber KC, Honorio KM, da Silva AB. Molecular features related to HIV integrase inhibition obtained from structure- and ligand-based approaches. PLoS One., 2014; 9(1): e81301.
- Eich E, Pertz H, Kaloga M, et al. (-)-Arctigenin as a Lead Structure for Inhibitors of Human Immunodeficiency Virus Type-1 Integrase. J Med Chem, 1996; 39(1): 86-95.
- Mazumder A, Raghavan K, Weinstein J, Kohn KW, Pommier Y. Inhibition of human immunodeficiency virus type-1 integrase by curcumin. Biochem Pharmacol, 1995; 49(8): 1165-70.

10. Galati G, Sabzevari O, Wilson JX, O'Brien PJ. Prooxidant activity and cellular effects of the phenoxyl radicals of dietary flavonoids and other polyphenolics. *Toxicology*, 2002; 177(1): 91-104.
11. Cavalieri EL, Li KM, Balu N, et al. Catechol orthoquinones: The electrophilic compounds that form depurinating DNA adducts and could initiate cancer and other diseases. *Carcinogenesis*, 2002; 23(6): 1071-7.
12. Zhao H, Neamati N, Hong H, et al. Coumarin-based inhibitors of HIV integrase. *J Med Chem*, 1997; 40(2): 242-9.
13. Hansch C, Hoekman D, Gao H. Comparative QSAR: toward a deeper understanding of chemicobiological interactions. *Chem Rev*, 1996; 96(3): 1045-76.
14. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci*, 2003; 22(1): 69-77.
15. Karelson M. *Molecular Descriptors in QSAR/QSPR*. Wiley- Interscience, New York, 2000.
16. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, 2000.
17. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA. New QSPR study for the prediction of aqueous solubility of drug-like compounds. *Bioorg Med Chem*, 2008; 16(17): 7944-55.
18. E.A. Gunnewegh, A.J. Hoefnagel, R.S. Downing, H. van Bekkum, *Recl. Trav. Chim. Pays-Bas* 115 (1996) 226.
19. W.C. Sun, K.R. Gee, R.P. Haugland, *Bioorg. Med. Chem. Lett.*, 1998; 8: 3107.