Indian Journal of Innovations & Developments, *Vol 3(1), 1–5, February 2014*
ISSN (Print): 2277-5382
ISSN (Online) : 2277-5390

Quantitative Structure Activity Relationship (QSAR) Studies of some Phthalimide Analogues as HIV-1 Integrase Inhibitors by Heuristic Method

Saeid Khodadoust*

Young Researchers and Elite Club, Branch Dehdasht, Islamic Azad University, Dehdasht, Iran;
saeid.kh64@gmail.com

Abstract

A Quantitative Structure Activity Relationship (QSAR) was developed by the Heuristic Method (HM) to study a series of phthalimide analogue compounds. Five classe's descriptors were calculated by the software CODESSA in this study. HM was used both for pre-selecting molecular descriptors and for developing the linear model. The inhibition activity of phthalimide analogues toward pIC_{50}, inhibition of HIV-1 integrase enzyme, was correlated with six descriptors, with a Squared Correlation Coefficient (R^2) of 0.8828 and Squared Cross-validated Correlation Coefficient (R^2_{cv}) of 0.8564, respectively. The stability and validity of the model were evaluated by validation of RMSE = 0.1857. This paper provided a way to predict the inhibition activity of phthalimide analogue toward pIC_{50} from their structures alone and gave some insight into structural features related to phthalimide analogue.

Keywords: AIDS, HIV-1 Integrase, QSAR, Phthalimide, MOPAC

1. Introduction

The Quantitative Structure Activity Relationship (QSAR) process is used to quantitatively correlate chemical structure with a well-defined process, such as biological activity, chemical reactivity, physicochemical properties, or environmental behavior. Molecular descriptors are mathematical values that describe the structures or shapes of the molecules, and they are used to predict the activity of molecules as well as their physical and chemical properties. These descriptors may be empirical or theoretical values. A QSAR model can be constructed by using many types of descriptors, such as topological, quantum chemical, geometrical, and spectral descriptors.

The Human Immunodeficiency Virus (HIV)[1] is the causative agent of the Acquired Immunodeficiency Syndrome (AIDS). An estimated 36 million people worldwide are currently living with HIV; however, there is still no known cure or vaccination against it. There are three viral enzymes encoded by the pool gene of HIV-1, namely Reverse Transcriptase (RT), Protease (PR), and Integrase (IN). Currently, the combination of RT and PR inhibitors, although highly effective, produces unwanted side effects like toxicity and patient adherence, among others[2]. HIV-1 IN functions in a two-step manner by initially removing a dinucleotide unit from the 3′-ends of the viral DNA (termed 3′-processing). The 3′-processed strands are then transferred from the cytoplasm to the nucleus where they are introduced into the host DNA following 5 base-pair offset cleavages of opposing host strands (termed 3′-strand transfer or end joining). There is obviously a requirement for a functional integrase in HIV-1 replication and has no cellular homologue[3]. This is one of the reasons why IN represents an attractive and validated target for chemotherapeutic intervention and has become a focus of Anti-AIDS drug design efforts.

Through the years, QSAR studies about activity of the inhibition of the HIV-1 integrase or the synthesis of new compounds with this activity have been carried out. Different descriptors and variable selection techniques

*Author for correspondence

were used to develop models capable of relating the activity to the structure[4]. In spite of the development of the Highly Active Anti-Retroviral Therapy (HAART)[5], there is an emergent need to search for new Anti-HIV agents. The main reasons are serious adverse side effects of the available drugs and the emergence of drug-resistance (including cross-resistance)[6]. Attention has been given to the development of drugs that act on new targets, such as the Host Protein Cell[7–10] and other viral structures, such as the enzyme HIV-1 Integrase (HIV-1 IN)[11,12]. The HIV-1 IN is, actually, a major breakthrough in AIDS research[13–16]. The present study improves the QSAR model with geometrical descriptors, electrostatic descriptors and quantum-chemical descriptors generated from the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) software[17].

2. Methodology

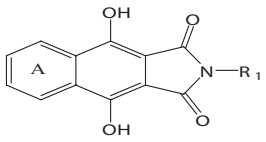
2.1 Data Sets

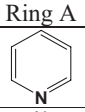
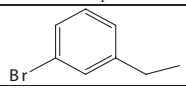
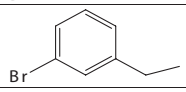
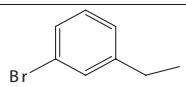
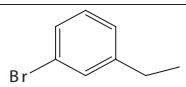
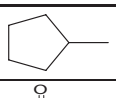
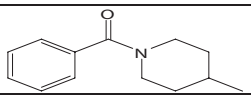
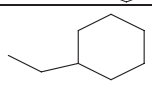
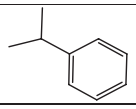
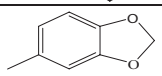
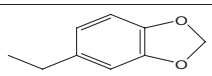
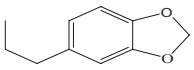
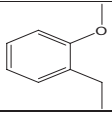
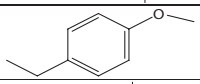
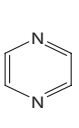
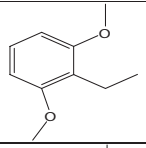
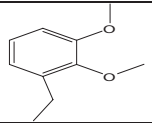
In the present study, we used a data set of 39 HIV-1 Integrase inhibitors for which the pIC_{50} for 3′-processing were reported. The data set for this investigation was obtained from the previous studies[18]. A complete list of the compounds' names and their corresponding pIC_{50} are summarized in Table 1. Determination of pIC_{50} values was achieved by plotting drug concentration versus percentage of inhibition and by measuring the concentration at which 50% inhibition occurred. In order to guarantee the linear distribution of the dependent variable, we calculated the natural logarithm of the IC_{50} values and obtained the model using these values.

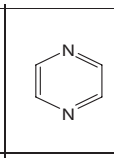
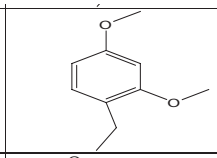
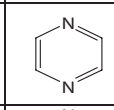
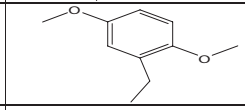
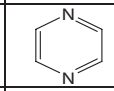
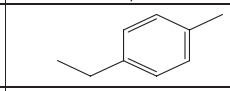
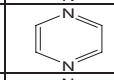
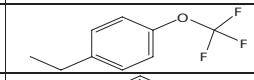
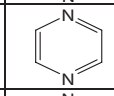
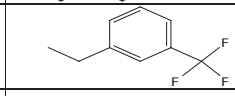
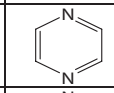
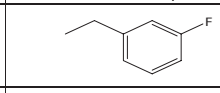
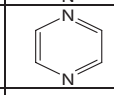
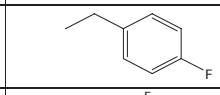
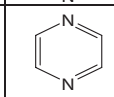
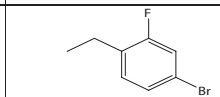
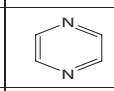
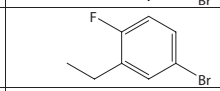
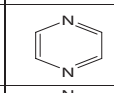
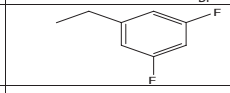
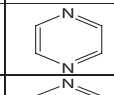
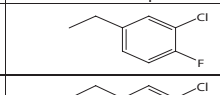
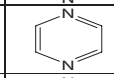
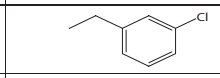
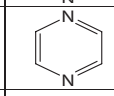
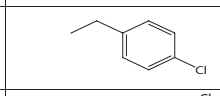
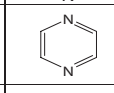
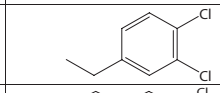
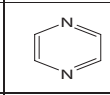
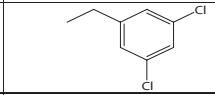
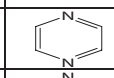
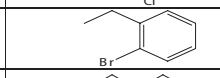
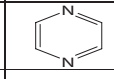
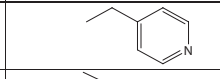
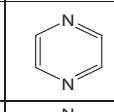
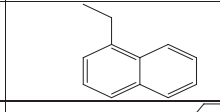
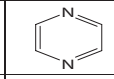
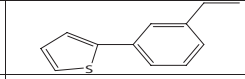
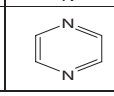
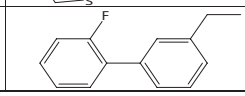
2.2 Calculation of Descriptor

The 2-D structures of the 39 molecules were drawn with Hyperchem software. Then obtained structures were pre-optimized by using MM+ molecular mechanics force field, and then a further precise optimization was done with the AM1 semi-empirical method. The molecular structures were optimized using the Polak–Ribiere algorithm till the root mean square gradient was 0.01, and then input the optimized files into MOPAC. The MOPAC output files were used by the CODESSA2.63 program to calculate five classes of descriptors: Constitutional (number of various types of atoms and bonds, number of rings and molecular weight, etc.); Topological (Wiener Index, Randic Indices

Table 1. Structural and HIV-1 integrase inhibition data of tricyclic phthalimide analogues



No	Ring A	R_1	pIC50
1			6.42
2			6.59
3			5.44
4			6.68
5		-CH_3	4.31
6			4.98
7			5.62
8			5.66
9			5.98
10			5.00
11			6.25
12			5.85
13			6.09
14			6.25
15			5.70
16			6.37

17			6.37
18			6.11
19			5.80
20			5.69
21			6.37
22			6.25
23			6.39
24			6.36
25			6.45
26			6.43
27			6.75
28			6.66
29			6.38
30			6.95
31			6.73
32			6.20
33			5.17
34			5.85
35			6.17
36			6.30

37			6.00
38			5.84
39			5.69

and Kier–Hall Shape Indices, etc.); Geometrical (Moments of Inertia, Molecular Volume and Molecular Surface Area, etc.); Electrostatic (Minimum and Maximum Partial Charges, Polarity Parameter and Charged Partial Surface Area Descriptors, etc.); and Quantum Chemical (Reactivity Indices, Dipole Moment, HOMO and LUMO Energies, etc.) and about 640 descriptors were calculated.

2.3 Statistical Analysis for Model Validation

A large number of molecular descriptors can be calculated by the CODESSA program on the basis of the geometrical and electronic features of the molecules. Structural attributes include topological connectivity indices and properties depending on the charge distribution in the molecule. CODESSA uses statistical structure–property correlation techniques for the analysis of experimental data in combination with the calculated molecular descriptors. Several strategies are available for the effective search of the best multi-parameter correlation in the large space of the natural descriptors. Finally, values of the molecular property of interest can be predicted by means of one or more multi-parameter correlation equation(s), obtained in the previous steps. The Heuristic Method (HM)[19] implemented in CODESSA PRO was employed for selecting the 'best' regression model.

Once molecular descriptors were generated, the HM in CODESSA was used to pre-select descriptors and built the linear model[20,21]. Its advantages included high speed and no software restrictions on the size of the dataset. The HM could either quickly give a good estimation about what quality of correlation to expect from the data or derive several best regression models. The HM proceeded with pre-selecting descriptors by eliminating:

1. Descriptors that was not available for each structure.
2. Descriptors having a small variation in magnitude for all structures.
3. Descriptors that gave an F-test's value below 1.0 in the one-parameter correlation.

4. Descriptors whose t-test values were less than the user-specified value.

This procedure ordered the descriptors by the decreasing correlation coefficient when used in one-parameter correlations.

As a next step, the program calculated the pair correlation matrix of descriptors and further reduced the descriptor pool by eliminating highly correlated descriptors. After the pre-selection of descriptors, multi-linear regression models were developed in a stepwise procedure. Thus, descriptors and correlations were ranked according to the values of the F-test and the correlation coefficient. Starting with the top descriptor from the list, two parameter correlations were calculated. In the following steps new descriptors were added one-by-one until the pre-selected number of descriptors in the model was achieved. The goodness of the correlation was tested by the Correlation Coefficient (R^2), F-test value (F), and the Squared Standard Error (S^2). The stability of correlations was tested against the cross-validated coefficient (R^2_{cv}), which described the stability of a regression model obtained by focusing on the sensitivity of the model to the elimination of any single data point. Briefly, for each data point, the regression was recalculated with the same descriptors but for the dataset without this point. The obtained regression was used to predict the value of this point, and the set of estimated values calculated in this way was correlated with the experimental values.

3. Result and Discussion

In order to select the structural descriptors responsible for the pIC_{50} of phthalimide analogues, the stepwise regression was performed by the HM. The heuristic correlations provided the optimal equations for different numbers of descriptors in the range of 1–8. Figure 1 shows the relationship of R^2 and R^2_{cv} with the number of descriptors. As could be seen from Figure 1, R^2 and R^2_{cv} increase as the number of descriptors increase from 1 to 8. To avoid 'over-parameterization' of the model, an increase in the R^2 values of less than 0.02 was chosen as the breakpoint criterion[22]. The best model contained six molecular descriptors. The regression coefficients of the descriptors and their physical - chemical meaning are listed in Table 2. The obtained model has a correlation coefficient R^2=0.8828, F=40.1786, s^2=0.0420 and R^2_{cv}=0.8564 with RMSE=0.1857. The plot of observed versus predicted pIC_{50} is illustrated in Figure 2.

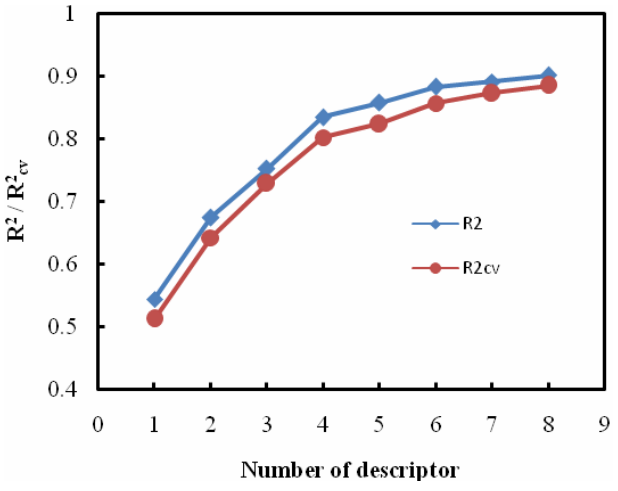


Figure 1. Number of descriptors vs. R^2 and R^2_{cv} values.

4. Conclusions

Our present effort to correlate the pIC_{50} of 39 drugs with theoretically calculated molecular descriptors has led to a relatively successful QSAR model of the drug molecules. All the descriptors involved were calculated solely from the chemical structures and had a definite physicochemical meaning corresponding to different intermolecular interactions. The present work provided an effective and simple method for prediction of the pIC_{50} for the 39 phthalimide analogue compounds.

5. Acknowledgement

The author expresses appreciation to the Young Researchers and Elite Club, Branch Dehdasht, Islamic Azad University, Dehdasht, Iran for financial support of this work. Also author makes a point of his sincere thanks to Prof. Natarajan Gajendran.

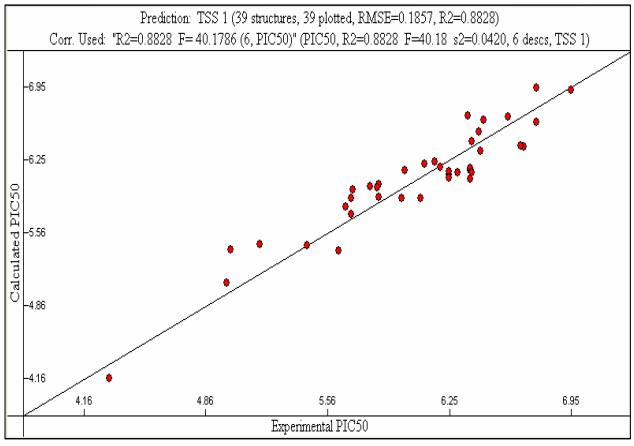


Figure 2. Experimental pIC_{50} vs. the predicted pIC_{50} by HM.

Table 2. Descriptors, coefficients and t-test values for the HM model

Descriptor	Chemical meaning	Coefficient	t-Test
Constant	Intercept	9.9982e+0	8.2695
PMIA	Principal Moment of Inertia A / # of atoms	-3.2186e+3	-8.8557
RNCG	RNCG Relative negative charge (QMNEG/QTMINUS) [Quantum-Chemical PC]	3.9132e+1	7.5376
MERIN	Max Electroph. React. Index for a N atom	-9.1157e+0	-3.5664
TMEI	Tot Molecular Electrostatic Interaction / # of atoms	-1.1495e+0	-5.0878
DPSA-3	DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	4.9199e-2	4.0412
HASA-1	HASA-1 [Zefirov's PC]	-8.8366e-3	-3.3022

6. References

1. Hu R, Doucet JP, Delamar M, Zhang R. QSAR models for 2-amino-6-arylsulfonylbenzonitriles and congeners HIV-1 reverse transcriptase inhibitors based on linear and nonlinear regression methods. Eur J Med Chem. 2009; 44(5):2158–71.
2. Richman DD. HIV chemotherapy. Nature. 2001; 410(1): 995–1001.
3. Makhija MT, Kulkarni VM. QSAR of HIV-1 integrase inhibitors by genetic function approximation method. Bioorg Med Chem. 2002 May; 10(5):1483–97.
4. Saíz-Urra L, González MP, Fall Y, Gómez G. Quantitative structure–activity relationship studies of HIV-1 integrase inhibition. 1. GETAWAY descriptors. Eur J Med Chem. 2007; 42(1):64–70.
5. Hagmann M. Study confirms effectiveness of antiretroviral drugs for HIV patients. Bull World Health Organ. 2003; 81(1):918–19.
6. Castro HC, Loureiro NI, Pujol-Luz M, Souza AM, Albuquerque MG, Santos DO et al. HIV-1 reverse transcriptase: a therapeutical target in the spotlight. Curr Med Chem. 2006; 13(3):313–24.
7. Melo EB, Silveira AG. α- and β-Glucosidase inhibitors: chemical structure and biological activity. Tetrahedron. 2006; 62(44):10277–302.
8. Melo EB, Carvalho I. α- and β-Glucosidases como alvos moleculares para desenvolvimento de fármacos. Quimica Nova. 2006; 29(1):840–43.
9. Rosenkilde MM, Gerlach LO, Jakobsen JS, Skerlj RT, Bridger GJ, Schwartz TW. Molecular Mechanism of AMD3100 Antagonism in the CXCR4 Receptor: transfer of binding site to the cxcr3 receptor. J Biol Chem. 2004; 279(4):3033–41.
10. Craigie R. HIV integrase, a brief overview from chemistry to therapeutics. J Biol Chem. 2001; 276(2):23213–16.
11. Souza MVN, Almeida MV. Drogas anti-VIH: passado, presente e futuras perspectives. Quimica Nova. 2003; 26(1):366–72.
12. Yuan H, Parrill AL. QSAR studies of HIV-1 integrase inhibition. Bioorg Med Chem. 2002 Dec; 10(12):4169–83.
13. Opar A. New HIV drug classes on the horizon. Nat Rev Drug Discov. 2007; 6:258–59.
14. Summa V, Petrocchi A. 4,5-Dihydroxypyrimidine carboxamides and N-Alkyl-5-hydroxypyrimidinone carboxamides are potent, selective hiv integrase inhibitors with good pharmacokinetic profiles in preclinical species. J Med Chem. 2006; 49(23):6646–49.
15. Roy K, Leonard JT. QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners using molecular connectivity and E-state parameters. Bioorg Med Chem. 2004; 12(4):745–54.
16. Borges de Melo E, Miguel Castro Ferreira M. Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors. Eur J Med Chem. 2009; 44(9):3577–83.
17. University of Florida and Semichem Inc. CODESSA PRO software version 2.7.2. 1995–2004.
18. Verschueren WG, Dierynck I, Amssoms KI, Hu L, Boonants PM, Pille GM et al. Design and optimization of tricyclic phtalimide analogues as novel inhibitors of HIV-1 integrase. J Med Chem. 2005; 48(6):1930–40.
19. Semichem and the University of Florida. CODESSA References Manual V. 2.13. 1995–1997.
20. Luan F, Xue C, Zhang R, Zhao C, Liu M, Hu Z et al. Prediction of retention time of a variety of volatile organic compounds based on the heuristic method and support vector machine. Anal Chim Acta. 2005; 537(1–2): 101–10.
21. Oblak M, Randic M, Solmajer T. Quantitative Structure Activity Relationship (QSAR) of flavonoid analogues. 3. inhibition of $p56^{lck}$ protein tyrosine kinase. J Chem Inform Comput Sci. 2000 Jul–Aug; 40(4):994–1001.
22. Tamm K, Fara DC, Katritzky AR, Burk P, Karelson M. A quantitative structure–property relationship study of lithium cation basicities. J Phys Chem A. 2004; 108(21): 4812–18.