ISSN: 2754-6705

Journal of Mathematical & Computer Applications

SCIENTIFIC Research and Community

Research Article

Open Access

QSAR Study of 4-Benzylpyridinone Derivatives: Non-Nucleoside Human Immunodeficiency Virus Type 1 Reverse Transcriptase Inhibitors

Alejandro Morales-Bayuelo

Basic and Clinical Research Group of the Universidad del Sinú, Universidad del Sinú, School of Medicine, Cartagena de Indias, Colombia

ABSTRACT

To study a series of derivatives of 4-Benzylpyridinone with inhibition activity on inverse transcriptase of HIV-1 using computational tools and bioinformatics. A Linear Regression Model (LRM) associated that anti-HIV-1 activity was obtained. This model was correlated with charge on nitrogen atom at the position 3 of 4-BP structure and the total energy of molecule. The LRM was statistically significant with p-value < 0.001, R2 = 0.9006, and a cross validation coefficient of R2cross = 0.8482. A discriminating analysis classified the molecules as active and inactive with high predictive power. In addition, six new molecules based on theoretical analysis were proposed as anti- HIV-1 activity candidate to be synthesized.

An equation model that relates the anti-HIV-1 activities of molecules studied was obtained. This model depends basically of two molecular descriptors, the charge on the nitrogen atom in the position 3 and total energy. The charge on the nitrogen atom descriptor gives us an idea of more active site of the molecule while the energy is related with on overall stability of the molecule.

***Corresponding author**

Alejandro Morales-Bayuelo, Basic and Clinical Research Group of the Universidad del Sinú, Universidad del Sinú, School of Medicine, Cartagena de Indias, Colombia

Received: September 07, 2023; **Accepted:** September 13, 2023, **Published:** October 01, 2023

Keywords: Derivatives of 4-Benzylpyridinone, HIV-1, QSAR, Theoretical Calculations

Introduction

In order to explore physicochemical properties of a set of derivatives of 4-benzylpyridinone (BP) (Figure 1, Table 1) a Quantitative Structure Activity Relationships (QSAR) was used to correlate molecular properties through of a series of theoretical molecular descriptors that account for hydrophobicity, electronic properties, and steric effects of molecules, among others; these descriptors were determined by well-established computational methods under the assumptions that "*the biological activity can be correlated with physical-chemical properties that are adequately described by the set of descriptors*". The use of these descriptors has allowed obtaining a robust model that related molecular descriptors and biological activity providing insights of the mechanism of action of the group of molecules [1-4].

The BP derivatives have been proven selective and potent non-nucleoside inhibitors of the Human Immunodeficiency Virus type 1 Reverse Transcriptase (HIV-1 RT) [5,6]. The human immunodeficiency virus (HIV) is a retrovirus, which contain ribonucleic acid (RNA) as genetic material (retrovirials family), characterized by integrating into the genome of the infected cell in the form of acid deoxyribonucleic (DNA). The most important property of HIV is its replication ability on a reverse transcription that involves the formation of DNA from RNA, using biological machinery of the infected cell [5,6].

Non-nucleoside HIV-1 RT inhibitors have attracted special attention due to these compounds may be drug candidates providing a unique mode of action which do not form non-covalent and non-competitive binding at the allosteric site present only in the HIV-1 RT enzyme [7].

In addition, the interest for HIV-1RT inhibitors has been increased recently due to the emergence resistance shown by some HIV variants to current drugs used to control the AIDs disease.

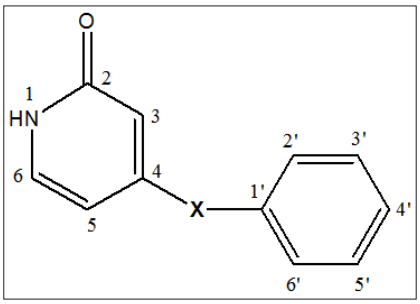


Figure 1: Structure of 4-Benzylpyridinone

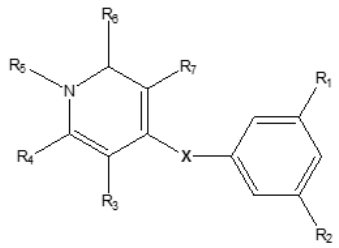
Correlating the anti-human immunodeficiency virus (HIV) activity with specific structural characteristics of non-nucleoside inhibitors is a theoretical approach to the understanding of the biological inhibition mechanism of action of the 4-benzylpyridinone derivatives. The aimed of this paper is estimate inhibitory power of 4-benzylpyridinone and establishing the differences among its

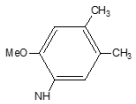
derivates. In this way, it is possible to get some valuable information that can be used as the basis for seeking possible new candidate compounds for proposing to make new syntheses with higher biological activities for this disease which has been affected thousands of people around the world.

In the present work we have taken a dataset containing a series of 4-benzylpyridinone compounds with HIV-1 RT inhibitory activities from literature [6]. The 4-benzylpyridinone derivative compounds used in the work inhibit a wide range of the HIV-1 RT activities with values of IC_{50} values from 0.2 to 41400 through a high diversity of structural dataset.

Also, were calculated a variety of electronic descriptors from molecular structures but only one of them were included in the adjusted model. The descriptors used in the model of this study were chosen for their non-colinearity and its ability to describe the molecular structure of the compounds under study. The model obtained in this paper predicts that molecules of the dataset with the higher positive charge on the nitrogen atom on BP and the higher total energy value experiment a better biological activity $\log(1/IC_{50})$.

Table 1: Substituted 4-Benzylpyrodinone Molecules with HIV-1 RT Inhibitory Activities



Compound	Substituent								IC50 Observed (nM)
	R1	R2	R3	R4	R5	R6	R7	X	
BP-1	CH_3	CH_3	CH_3CH2	CH_3	H	O=	NH_2	CH_2	17
BP-2	CH_3	CH_3	CH_3CH2	CH_3	H	O=	$NHCOCH_3$	CH_2	100
BP-3	CH_3	CH_3	CH_3CH_2	CH_3	H	O=	NHCHO	CH_2	3
BP-4	CH_3	CH_3	CH_3CH_2	CH_3	H	O=	NHCHO	S	6600
BP-5	CH_3	CH_3	CH_3CH_2	CH_3	H	O=	$NHCH_3$	CH_2	30
BP-6	CH_3	CH_3	CH_3CH_2	CH_3	H	O=	$NHCO_2CH_2CH_3$	CH_2	460
BP-7	CH_3	CH_3	CH_3CH_2	CH_3	H	O=	$N(CH_3)_2$	CH_2	0.2
BP-8	CH_3	CH_3	CH_3	H	H	O=	$NHCO_2CH_2CH_3$	S	7500
BP-9	CH_3	CH_3	CH_3	H	H	O=	$NHCOCH_2CH_3$	S	41400
BP-10	CH_3	CH_3	CH_3CH_2	CH_3	-	CH3O-	$NHCOC(CH_3)_3$	CH_2	1000
BP-11	CH_3	CH_3	CH_3CH_2	CH_3	-	CH3O-	$NHCOC(CH_3)_3$	CHOH	1000
BP-12	CH_3	CH_3	CH_3CH_2	CH_3	-	CH3O-	$NHCOC(CH_3)_3$	CO	10000
BP-13	CH_3	CH_3	CH_3CH_2	CH_3	H	O=		CH_2	1000
BP-14	H	H	CH_3CH_2	CH_3	H	O=	NH_2	CH_2	730

Computational and Theoretical Details

Modeling

Molecular structures of BPs (Tables 1), were optimized with semi-empirical MO AM1 Methods, in Gaussian program version 03 [8-11]. Stable conformations of molecules obtained at the minimum energy through molecular optimization were used as input to perform the calculation of molecular descriptors from a higher computational calculation level using DFT/ B3LYP method, with 6-31G basis sets [12-15].

The most stable structure of each compound was used for computing several physico-chemical descriptors such as thermodynamic and electronic properties including atomic charges. All descriptors considered in the model were the independent variable while the biological activity was the dependent variable. All possible combinations of parameters were used in order to generate a robust QSAR model in this study.

Statistical Analysis

Relationships between the biological activity log(1/IC50) and the molecular descriptors generated were established by the backward stepwise multiple regression technique. The best model was selected on the basis of the highest multiple regression coefficient of determination (R2). The statistical predictive power of model was performed by assessing the cross-validation (CV) test. The

CV correlation coefficient (Rcross) was computing between one leaving-out log(1/IC_{50}) generated and the experimental log(1/IC_{50}) values [16]. Higher Rcross values are an indicative of the robustness of linear regression model. An additional predictive power of the model was carried out by performing a linear discriminant analysis (LDA). The LDA function was obtained to classify calculated and experimental log(1/IC_{50}) values [17,18]. All statistical calculations were performed using the R statistical program [19].

Results and Discussion

In the present study, a QSAR model was performed over a set of 14 BP derivatives [6]. A wide range of molecular descriptors were calculated. The descriptors included in the adjusted model were chosen for pair wise colinearity below of 80% and its ability to describe the molecular structure of the compounds under study.

The best linear regression model (Equation 1) describe the independent variable log(1/IC_{50}) on function of charge on the nitrogen atom at the position 3 (Q_N) and the total energy (E) as parameters of the dependent variable. This model is characterized by a coefficient of determination of R^2 = 0.9006 and a cross validation coefficient of Rcross = 0.921. This finding suggests that the model explains the 90.06% of the variability with a model robustness of 92.1%.

$$\text{Log}\,(1/IC_{50}) = 11.7643 + 1{,}05015\text{x}10^{-5}\,E + 9{,}72234Q_N \quad \text{(Equation1)}$$

R^2 = 0.9006, R^2_{cross} = 0.8482, p-value = 0.0000, F = 45.3, n = 13.

A correlation between log(1/IC_{50}) calculated obtained from multiple linear regression model (Equation 1) and experimental values was made (Figure 2 and Table 2). The linear regression model was obtained from the dataset excluding the BP-14 compound (Table 1). This BP-14 was considered as an outlier from dataset in which whose molecular behavior is dissimilar from the rest of structures. The calculations excluding the BP-14 compound were of R^2 = 0.9006 for the coefficient of determination, standard error of estimate (SEE) values $\left(SEE = \sqrt{\frac{PRESS}{n}}\right)$ SEE = 0.466 where prediction error sum of the squared (PRESS),

$PRESS = \sum (Y_{predict} - Y_{actual})^2$, and q^2=0.8896 for predictive correlation coefficient q^2 $\left(q^2 = 1 - \frac{\sum (Y_{predict} - Y_{actual})^2}{(Y_{actual} - Y_{mean})^2}\right)$

while the same calculation including the structure BP-14 were lower for R^2 = 0.8052, q^2=0.7819, and higher for SEE = 0.653. From the chemical point of view the BP-14 revealed to be a compound structurally different from the rest of other compounds. The unusual behavior of BP-14 can be explained due to the two methyl groups at the position 3' and 5' on benzyl ring are replaced by hydrogen atoms on the BP-14 contrary to the rest of compounds. For instance, when two 3',5'-methyl groups are placed on the structure BP-14 replacing the hydrogen atoms and resulting the structure BP-1, the activity measured as log(1/IC_{50}) increased the biological activity from -2.863 for BP-14 to -1.23 for BP-1. Therefore, the substitution of methyl groups at position 3' y 5' on the 4-BP play an important role on the activity of these compounds.

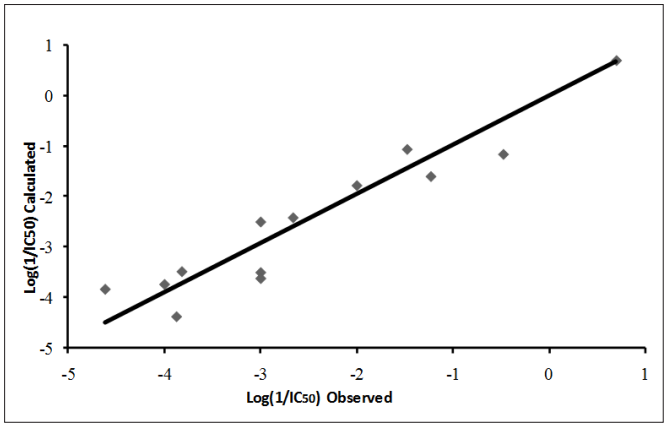


Figure 2: Plot of observed and calculated log (1/IC50) using Equation 1

The multiple linear regression model (Equation 1) is able to predict that derivate of 4-benzylpiridinone compounds with higher both positive charge on the nitrogen atom at position 3 and total energy values shown an increasing trend on biological activity log(1/IC_{50}). Therefore, it is necessary to consider both descriptors explain satisfactorily the biological activity of BP derivatives. In addition, the most active molecules are obtained when the substituted groups on nitrogen atom at position 3 are strong electron attractive.

The influence of size of substitute groups on nitrogen atom at position 3 was also observed that bigger size groups shown better and significant biological activity which can related with the interaction of the allosteric site of the enzyme which is crucial to inhibit the transcription of viral RNA to a double chain of viral DNA. Further studies are necessary on molecular docking to know the role of substituent group size.

A linear discriminating function (LDF) (Equation 2) with two predictive variables and statistical significantly P-value less than 0.05 ranks the dataset of 14 molecules into two groups. The first subset of molecules are considered inactive compound to an inhibition activity of the enzyme reverse transcriptase HIV-1 with values log (1/IC_{50}) between -4.67 to -3.00 while the second part of dataset with values of log(1/IC_{50}) between -2.863 and 0.69 were classified as active compound (Figure 3).

$$F = \text{CONSTANT} + c_1E + c_2\,QN \quad \text{(Equation 2)}$$

	Inactive	Active
CONSTANT	- 206.862	-134.812
E (Kcal/mol)	-0.000266283	-0.00021018
QN	-228.411	-188.806

n=14 P-value= 0.002
F = linear discriminating function
c_1 and c_2 are coefficient for E and QN

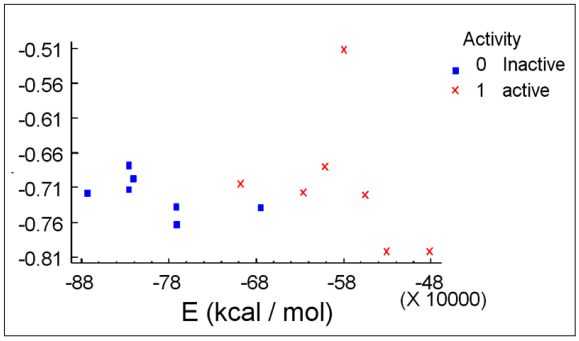


Figure 3: Low and high activity classification of 4-BP derivate compounds by Linear discriminating function

The LDF obtained was enabled to predict 85.7% of all cases correctly classified. The higher percentage of molecules correctly classified as low and high activity is an indicative of power of predictive of the biological activity for the model presented [16].

A qualitative structure activity relationship analysis on 4-BP dataset shows that methylene fragments of X on the 4-benzylpiridinone led the best anti-HIV-1 activities (IC_{50}) ranging from 0.2 to 1000 nM while compound where sulphur atom is replaced by the methylene group the activities are lower ranging from 6600 to 41400 nM. In addition, the molecules 10, 11 and 12 did not show significant biological activity. It can be explained by the reason that in the molecules have been replaced by carbonyl groups at the N-2 position instead of methoxy group (OCH_3), giving as a result a different chemical environment. The substituted OCH_3 group increases the intermolecular forces.

The effect of hydrogen atoms in the methyl groups is compensated with the presence of a carbonyl tert-butyl group on the nitrogen atom at position N-3 which affects the charges of this atom. This means that the charges decrease due to presence of attracting group instead of weakly electrons of nitrogen atom which it is bonded. These features allow us to suggest that the carbonyl group in position N-2 is important to activity of this molecule. The methoxy group also makes that the pyridinones are protected to interact with the active site of reverse transcriptase.

However, the best activities are controlated by replacing high electro-attractive groups on the nitrogen to confer higher positive charge on this atom. However, these groups do not help to decrease the total energy of the molecule, thereby achieving a way to control of inhibiting activity of the reverse transcriptase of HIV-1 in the CD4 T-lymphocytes.

Table 2: Substituted 4-benzylpyrodinone molecules and log($1/IC_{50}$) observed and calculated HIV-1 RT inhibitors

Compound	Energy (Kcal/mol)	QN	Log($1/IC_{50}$) observed	Log($1/IC_{50}$) calculated	LDF classification
BP-1	-530631	-0.802	-1.230	-1.6054	active
BP-2	-626394	-0.717	-2.000	-1.7847	active
BP-3	-601723	-0.68	-0.477	-1.1659	active
BP-4	-826901	-0.676	-3.820	-3.4917	inactive
BP-5	-555284	-0.72	-1.477	-1.0671	active
BP-6	-698244	-0.705	-2.663	-2.4226	*active
BP-7	-579938	-0.512	0.699	0.6962	active
BP-8	-874093	-0.717	-3.875	-4.3859	inactive
BP-9	-826906	-0.712	-4.617	-3.8418	*inactive
BP-10	-675701	-0.738	-3.000	-2.5067	inactive
BP-11	-772206	-0.737	-3.000	-3.5104	inactive
BP-12	-771459	-0.762	-4.000	-3.7456	inactive
BP-13	-821412	-0.696	-3.000	-3.6285	inactive
BP-14	-481298	-0.802	-2.863	-1.0874	active

*Molecule incorrectly classified by linear discriminating function of Equation 2.

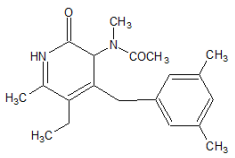
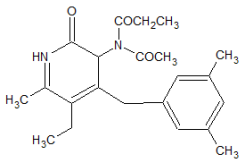
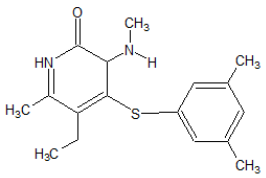
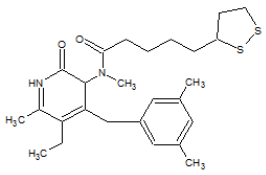
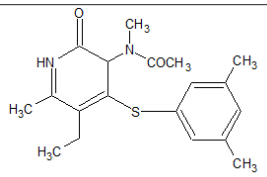
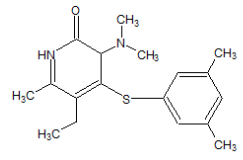
An analysis on the sulfur atoms placed at the position N-4 replace the methylene group of the molecules 2 and 3 maintaining the activity. It is assumed that the sulfur has a higher electronegativity, for these molecules are necessary to take into account the substituents that are in the position of the nitrogen N3. The values of total energy (Table 1) and charge on the nitrogen atom are depicted of the molecules 2 and 3 the values of total these values have the same trend that provides for the model, being the molecule 3 more active that the molecule 2 this is due to the molecule 3 has a high energy and high charge on the atom nitrogen that the 2.

The compound 13 has low activity, the reason for this behaviour could be the fact that the molecule has two bulky groups at position N-3 and N-4 which possibly to prevent such a structure does not fit into the active site of reverse transcriptase. In the derivatives of BP-8 and -9 which have IC_{50} values of 7500 and 41,400 nM respectively, are observed that the values of total energy are -874092.96 kcal/mol for the BP-8 and -826905.74 kcal /mol for the BP-9. These two molecules have the lower levels of energy and biological activity of the full range of study molecules, this is possibly due to structural differences, and these two compounds do not have the replacements of ethyl group in position 5 and N-ethyl group in position N-6 in the main ring of BP, having only a methyl group at position N-5.

On the other hand, a series of new substituent have been proposed (Table 3) in specific sites of the molecules considered in order to get new compounds as potential RT inhibitors from theoretical parameters values of total energy and charge on the nitrogen atom and then were tested in the linear regression model (Equation 1) in order to find the log ($1/IC_{50}$) calculated. In addition, the proposal structures were classified as active and inactive by linear discriminating function of Equation 2.

The molecules proposal as possible candidate for synthesis (Table 3) were calculated physicochemical parameters such as the charge values on the nitrogen atom and total energy and derivate corresponding log($1/IC_{50}$) predicted. It is significant that LDA function predicted well the inactivity of molecule with values of IC_{50} below of 1000 nM but fail to predict as active the BP(b) structure. Under theoretical consideration was proposal very high activities structures BP(c) y BP(f) with outstanding IC50 estimated values of 0,080and 0,070 nM respectively. It should be noted that these compounds present possible potent inhibitors of the RT but obviously it is needed to experimental synthesis to corroborate this prediction.

Table 3: Probable New Transcriptase Inverse Inhibitors

Compound	Structure	E (Kcal/mol)	QN	Log($1/IC_{50}$) calculated	LDF classification
BP (a)		-651047	-0.606	-0.9644	activity
BP (b)		-746809	-0.677	-2.6603	inactivity*
BP (c)		-780468	-0.254	1.0987	activity
BP (d)		-1297970	-0.621	-7.9039	inactivity
BP (e)		-865806	-0.231	0.4262	activity
BP (f)		-795639	-0.232	1.1533	activity

*Structure incorrectly classified by linear discriminating function of Equation 2.

Conclusions

An equation model that relates the anti-HIV-1 activities of molecules studied was obtained. This model depends basically of two molecular descriptors, the charge on the nitrogen atom in the position 3 and total energy. The charge on the nitrogen atom descriptor gives us an idea of more active site of the molecule while the energy is a descriptor is related with on overall stability of the molecule. To get a better biological activity, the substituent groups on the nitrogen atom in the 3 position must possess the characteristic of being an electron attractive group because these kinds of substituent attract electrons on nitrogen atom to be greater positive charge and thus the results of biological activity in the model would be more satisfactory.

References

1. Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-Chemical Descriptors in QSAR/QSPR Studies. Chem Rev 96: 1027-1044.
2. Katritzky AR, Fara DC, Petrukhin RO, Tatham DB, Maran U, et al. (2002) The present utility and future potential for medicinal chemistry of QSAR/QSPR with whole molecule descriptors. Curr Top Med Chem 2: 1333-1356.
3. Payares P, Díaz D, Olivero J, Vivas-Reyes R, Gómez I (1997) Prediction of the gas chromatographic relative retention times of flavonoids from molecular structure. J Chromatogr A 771: 213-219.
4. Selassie CD, Mekapati SB, Verma RP (2002) QSAR: then and now. Curr Top Med Chem 2: 1357-1379.
5. Paolucci S, Baldanti F, Maga G, Cancio R, Zazzi M, et al. (2004) Antimicrob Agents Chemother 48: 4611-4617.
6. Dollé V, Nguyen CH, Legraverend M, Aubertin AM, Kirn A, at al. (2000) Eur J Med Chem 43: 3949.
7. Schafer W, Friebe WG, Leinert H, Mertens A, Poll T, et al. (1993) Non-nucleoside inhibitors of HIV-1 reverse transcriptase: molecular modeling and x-ray structure investigations. J Med Chem 36: 726-732.
8. Dewar MJS, Hwang CH, Kuhn DRJ (1991) Am Chem Soc 113: 735.
9. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107: 3902-3909.
10. Dewar MJS, C Jic (1993) Tetrahedron. 49: W5003.
11. Gaussian 98, Revision A.7. Gaussian, Inc., Pittsburgh, PA, 1998. Gaussian 98
12. (a) Parr RG, Yang W (1989) Density Functional Theory of Atoms and Molecules, Oxford University Press, New York https://global.oup.com/academic/product/density-functional-theory-of-atoms-and-molecules-9780195092769?cc=us&lang=en&.
    (b) Parr RG, Yang W (1995) Density-Functional Theory of the Electronic Structure of Molecules. Annu Rev Phys Chem 46: 701-728.
    (c) Geerlings P, De Proft F, Langenaeker W (2003) Conceptual Density Functional Theory. Chem Rev 103: 1793-1874.
    (d) Morales-Bayuelo A, Baldiris R, Torres T, Torres JE, Vivas-Reyes R (2012) In J Quant Chem 112: 2681.
13. (a) Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98: 5648-5652.
    (b) Lee C, Yang W, Parr RG (1998) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37: 785.
    (c) Stephens PJ, Delvin FJ, Chablowski CF, Firsch MJ (1994) Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. J Phys Chem 98: 11623-11627.
14. Hehre WJ, Ditchfield R, Pople JA (1972) Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. J Chem Phys 56: 2257-2261.
    (b) Francl MM, Petro WJ, Hehre WJ, Binkley JS, Gordon MS, et al. (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. J Chem Phys 77: 3654-3665.
15. Hehre WJ, Radom L, Schleyer PVR, Pople JA (1986) Ab Initio Molecular Orbital Theory. Wiley New York https://www.wiley.com/en-us/AB+INITIO+Molecular+Orbital+Theory-p-9780471812418#:~:text=a%20dummy%20description.-,Description,compares%20them%20to%20experimental%20results..
16. Katritzky AR, Gordeeva EV (1993) Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. J Chem Inf Comput Sci 33: 835-857.
17. Wold S, Eriksson L, Clementi S (1995) Statistical Validation of QSAR Results, in H. Van Waterbeemd, (Ed.). Chemometric Methods in Molecular Design, VCH Verlagsgesellschaft mbH, Weinhiem, https://download.e-bookshelf.de/download/0000/6032/39/L-G-0000603239-0002288628.pdf.
18. Miller JN, Miller JC (2005) Statistics and Chemometrics for Analytical Chemistry. 5th Edition. Pearson Prentice Hall https://bisakimia.com/wp-content/uploads/2017/09/21202653-miller-m-james-statistics-and-chemometrics-for-analytical-chemistry-5th-ed.pdf.
19. R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL http://www.R-project.org.