Moroccan Journal of Chemistry
ISSN : 2351-812X
Copyright © 2023,
University of Mohammed Premier Oujda Morocco

Mor. J. Chem., 2023, Volume 11, Issue 02, Page 300-317
https://revues.imist.ma/index.php/morjchem

In silico design of new pyrimidine-2,4-dione derivatives as promising inhibitors for HIV Reverse Transcriptase-associated RNase H using 2D-QSAR modeling and (ADME/Tox) properties

Y. El Masaoudy[1], K. Tabti[1], Y. Koubi[1], H. Maghat[1,*], T. Lakhlifi[1], M. Bouachrine[1,2]

[1]*Molecular Chemistry and Natural Substances Laboratory (MCNSL), Department of Chemistry, Faculty of Sciences, University Moulay Ismail, Meknes, Morocco*
[2]*EST Khenifra, Sultan Moulay Slimane University, Beni Mellal, Morocco.*
**Corresponding author, Email address: h.maghat@umi.ac.ma*

Received *17 Dec 2022,*
Revised *18 Feb 2023,*
Accepted *20 Feb 2023*

Citation*: El Masaoudy Y., Tabti K., Koubi Y., Maghat H., Lakhlifi T., Bouachrine M. (2023) In silico design of new pyrimidine-2,4-dione derivatives as promising inhibitors for HIV Reverse Transcriptase-associated RNase H using 2D-QSAR modeling and (ADME/Tox) properties, Mor. J. Chem., 14(2), 300-317. Doi: https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i2.35455*

Abstract: The main target of present QSAR modeling is to pave the way for the development of new pyrimidine-2,4-dione derivatives and predict their HIV reverse transcriptase-associated RNase H inhibitory activity. To accelerate this process, linear and non-linear models of thirty-nine pyrimidine-2,4-dione derivatives have been constructed by exploiting PCA, MLR, and MNLR statistical techniques available in the XLSTAT software, as well as the (DFT/ Beck3LYP/6 − 31G (d, p)) approach. Among the 16 quantum and physicochemical descriptors measured, only four optimal molecular descriptors have been employed to perform QSAR models, i.e., density, number of H-bond acceptors, octanol/water partition coefficient, and LUMO energy. The Loo/cross-validation procedure, the Y-scrambling test, Golbraikh-Tropsha's criteria and the applicability area have all been utilized to evaluate the linear model's performance accuracy. Likewise, the nonlinear model's predictive power has been measured internally through the Loo/cross-validation procedure with coefficient $R^2_{CV(LOO)}$ and externally through test set compounds with external prediction coefficient R^2_{pred}. Herein, both MLR and MNLR models which exhibited excellent performance and met OECD criteria were exploited to predict inhibitory activities. By analyzing the structural characteristics of the studied compounds encoded in the afore-mentioned descriptors along with their effects on pIC_{50} inhibitory activity, we have been able to design eleven new chemical inhibitors. All of these inhibitors with new substituents displayed significantly higher HIV RT-associated RNase H inhibitory activities than the existing ones, as well as satisfactory results in silico ADME/Toxicity assessments.

Keywords: Pyrimidine-2,4-dione; HIV RT RNase; 2D-QSAR; Lipinski's RO5; ADME/Tox properties

1. Introduction

The human immunodeficiency virus (HIV) infection remains a serious and life-threatening condition, despite the major diagnostic and therapeutic advances and it is a global pandemic and major public health issue (De Broucker, 2013).

HIV is a retrovirus belonging to the lentivirus family that attacks immune system cells with CD4 receptors on their surface (Hammache *et al.,* 2000), these are in the first place CD4+ T lymphocytes.

HIV is wrapped with a capsid containing an RNA genome and 3 viral enzymes: reverse transcriptase, integrase, and protease. HIV reverse transcriptase (RT) is a key enzyme in the viral cycle that ensures a complex step at the cytoplasmic level. It is the first viral enzyme to be targeted by anti-HIV drugs and is responsible for the high variability of HIV within each individual (Gerbouin and Grellet, 2017, Wang *et al.,* 2018).

HIV Reverse Transcriptase is an enzyme that has several associated activities, including DNA polymerase that can copy an RNA or a DNA model and ribonuclease H (RNase H) that degrades the viral RNA genome during reverse transcription (Ilina and Parniak, 2008). These 2 activities work in synergy to achieve transcription in the opposite direction of the standard direction. That is, allows the synthesis of complementary DNA from viral RNA, before its integration into the genome of the infected cell and produces persistent infection and viral reservoirs (Julias *et al.,* 2002, Nikolenko *et al.,* 2005).

The RNase H (RH) activity of HIV reverse transcriptase (RT) has therefore been considered an attractive target for drug discovery and development for antiretroviral therapy (Gerbouin and Grellet, 2017, Klumpp and Mirzadegan, 2006). In contrast, there is a class of drugs called reverse transcriptase inhibitors that slow down or block the action of the enzyme. This category consists of two subtypes: nucleoside reverse transcriptase inhibitors (NRTIs), commonly known as nucleoside analogs, their action on reverse transcriptase is done by competition with natural nucleosides, it blocks the elongation of the viral DNA chain, and non-nucleoside reverse transcriptase inhibitors (NNRTIs), commonly known as non-nucleoside analogs which act selectively on reverse transcriptase by blocking DNA polymerase-dependent RNA and DNA activities (Spence *et al.,* 1995, Hang *et al.,* 2007, Cihlar, *et al.,* 2010). Nevertheless, resistance is observed particularly because of the occurrence of mutations in reverse transcriptase.

The purpose behind this theoretical research was to identify a new class of pyrimidine-2,4-dione derivatives with enhanced anti-HIV reverse transcriptase-associated RNase H activity. Toward this end, we attempted to propose predictive 2-dimensional-QSAR models (linear and nonlinear) based on certain quantum and physicochemical parameters along with three statistical techniques (PCA, MLR, and MNLR). The afore-mentioned 2D-QSAR models were submitted to an internal evaluation using the simplest cross-validation procedure, i.e., the Leave-one-out method, and external evaluation by external prediction coefficient R^2_{pred}. The linear model was subjected to more stringent and critical testing in addition to previous validations, including Y-randomization, parameters suggested by Golbraikh and Tropsha (Golbraikh and Tropsha, 2002; Tropsha, 2010) as well as applicability area. Finally, Lipinski' physicochemical parameters (HBAs, Rot B, HBDs, $logP_{o/w}$, and MWT) (Lipinski *et al.,* 1997) and the pharmacokinetic and toxicity (ADME/Tox) properties were computed to determine the fate of the candidate compounds.

2. Materials and Methods

2.1 Dataset

To perform QSAR models, we have chosen a dataset of thirty-nine pyrimidine-2,4-dione derivatives that were reported as HIV RT RNase H inhibitors by Lei Wang et al (Wang *et al.,* 2018). The activity of these inhibitory derivatives was expressed as IC_{50} ranging from $(0.0076 - 7\ \mu M)$. The IC_{50} value of each compound was transformed into pIC_{50} (M) using the formula $pIC_{50} = -\log_{10}(IC_{50})$ and then invested as a dependent variable for the 2-dimensional-QSAR analysis. The general molecular structure of 39 chemical inhibitors is shown in **Figure 1**. The substituents (R_1, R_2, R_3 and R_4) and

pIC_{50} values of all chemical compounds are represented in **Table 1**. The above dataset was randomly divided into two independent groups. A group of 31 samples (training set) was selected to formulate the model equation. The remaining 8 samples (test set) were utilized to measure the performance of the constructed model.

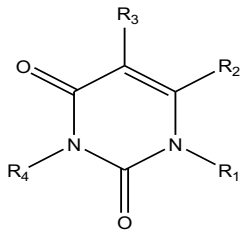
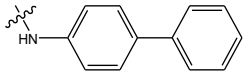
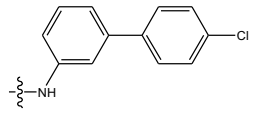
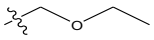
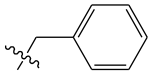
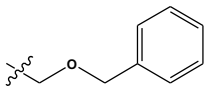
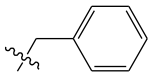
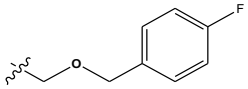
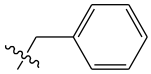
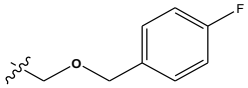
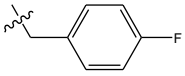
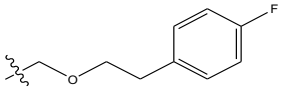
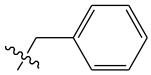
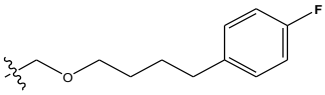
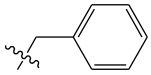
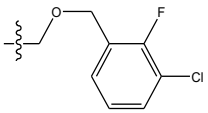
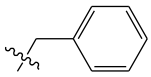
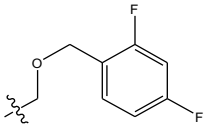
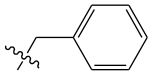
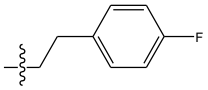
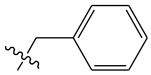
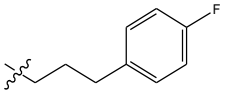
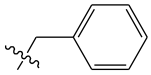
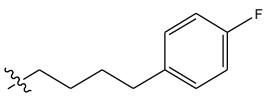
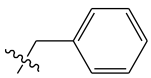
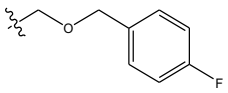
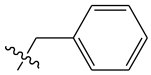
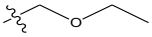
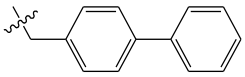
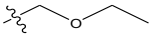
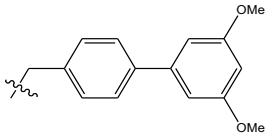
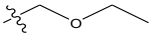
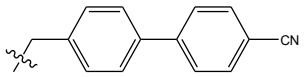
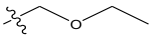
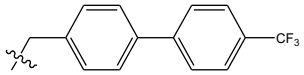
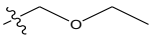
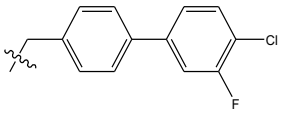
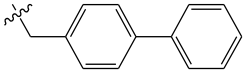
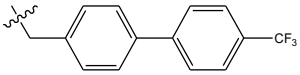
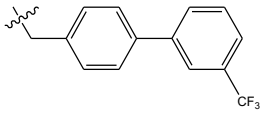
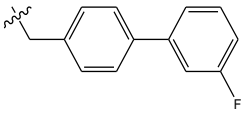
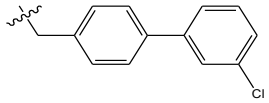
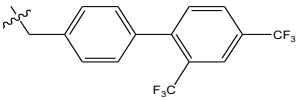
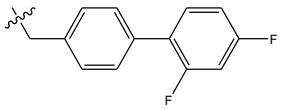
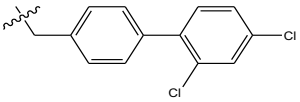
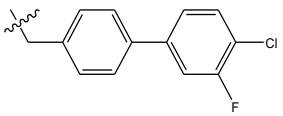
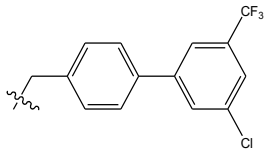
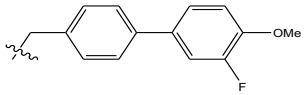
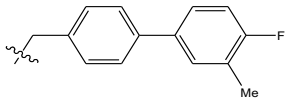
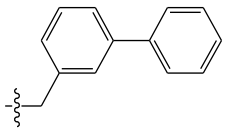
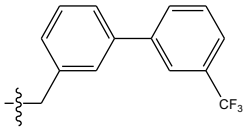
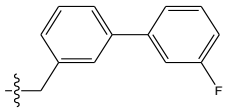
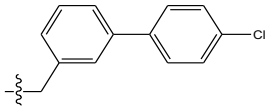
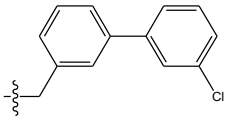
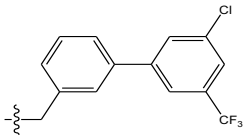
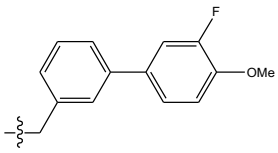
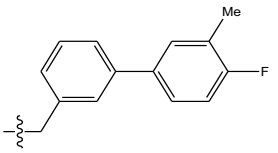


Figure 1. Two-dimensional pyrimidine-2,4-dione core representation

Table 1. Substituents of thirty-nine pyrimidine-2,4-dione and their pIC_{50} values

$N°$	pIC_{50}	R_1	R_2	R_3	R_4
1	6.602	H		H	OH
2	6.658	H		H	OH
3	5.167			$-HC(CH_3)_2$	OH
4	5.602			$-HC(CH_3)_2$	OH
5	5.678			$-HC(CH_3)_2$	OH
6	5.208			$-HC(CH_3)_2$	OH
7	5.553			$-HC(CH_3)_2$	OH
8	5.745			$-HC(CH_3)_2$	OH
9	6.046			$-HC(CH_3)_2$	OH
10	5.745			$-HC(CH_3)_2$	OH
11	5.180			$-HC(CH_3)_2$	OH
12	5.420			$-HC(CH_3)_2$	OH
13	6.046			$-HC(CH_3)_2$	OH

14	5.155			$-HC(CH_3)_2$	Me
15	5.208			Me	OH
16	5.174			Me	OH
17	5.284			Me	OH
18	5.678			H	OH
19	6.367			H	OH
20	7.796	H		H	OH
21	7.328	H		H	OH
22	7.602	H		H	OH
23	7.553	H		H	OH
24	7.796	H		H	OH
25	7.071	H		H	OH
26	7.824	H		H	OH
27	8.036	H		H	OH
28	8.027	H		H	OH
29	7.824	H		H	OH
30	8.119	H		H	OH
31	7.959	H		H	OH

32	7.796	H		H	OH
33	7.432	H		H	OH
34	7.745	H		H	OH
35	7.921	H		H	OH
36	7.699	H		H	OH
37	7.420	H		H	OH
38	8.114	H		H	OH
39	8.056	H		H	OH

2.2 Molecular descriptors: drawing and QSAR calculation

The theoretical approach requires the use of molecular descriptors to acquire structural data about chemical compounds (Consonni and Todeschini, 2010). Two-dimensional 2D and three-dimensional 3D structures were constructed by ChemDraw Ultra 8.0 and GaussView 6.0. To establish a consistent correlation between inhibitory activity and structural properties of molecules and to develop 2D-QSAR models, numerous molecular descriptors reported in **Table 2** are computed for each chemical compound in the dataset using three software (Tabti *et al.,* 2022b).

Table 2. Descriptors included in the two-dimensional QSAR modeling

Dipole moment μ, Total Energy E_T, LUMO Energy, HOMO Energy, HOMO-LUMO Energy gap ΔE, Electrophilicity index ω, Hardness η, Electronegativity χ, Number of H-Bond Acceptors HBAs, number of H-bond Donors HBDs, Partition coefficient $\log P_{o/w}$, Density d, Parachor Pc, Refractive Index n, Polarizability P, Surface tension S.
$\Delta E = E_{LUMO} - E_{HOMO}, \eta = \Delta E/2, \chi = -(E_{LUMO} + E_{HOMO})/2, \omega = \chi^2/2\eta$

Both ACD/ChemSketch (https://chemaxon.com) and MarvinSketch (https://www.acdlabs.com) software were applied to measure physicochemical descriptors, and the Gaussian 09W (Frisch *et al.,* 2016) software package was also applied to measure quantum chemical parameters through the density functional theory (DFT) technique (Cramer, 2004) by using functional Becke-3-Lee-Yang-Parr (Becke, 1993, Lee *et al.,* 1988) together with (6-31G (d, p)) basis set (Hariharan and Pople, 1973).

2.3. Statistical Methods

To better understand and interpret the influence of the physicochemical properties of 39 pyrimidine-2,4-dione derivatives on their inhibitory activity expressed as pIC_{50}, there is a need for QSAR statistical methods that can be used to analyze molecular descriptors and perform robust models. For this purpose, we used three techniques implemented in the XLSTAT software (XLSTAT version 2021.1), both multiple linear regression (MLR) and multiple nonlinear regression (MNLR), as well as the principal component analysis (PCA). Molecular descriptors were first subjected to PCA by exploiting the correlation matrix, which serves to decrease the number of the descriptors computed and to select the relevant and appropriate ones to quantitatively describe and explain the response activity (Koubi *et al.,* 2019). Subsequently, those selected molecular descriptors are used in the MLR to propose a linear model that could predict the inhibitory activities of novel molecules. Normally, the MNLR method has often been used in the field of QSAR modeling to improve the structure-activity relationship. In this regard, we developed an MNLR model based on the data of the descriptors and training set compounds proposed by the MLR (Tabti *et al,* 2020, Elidrissi *et al,* 2018, Lafridi *et al,* 2020).

2.4. Model Selection and validation procedures

In the present research, statistical characteristics associated with optimal model, internal and external assessment were made to verify the quality of the model developed, its stability and its ability to predict biological activities according to OECD criteria (OECD, 2007). In the first stage, we used different statistical characteristics to evaluate the quality and the reliability of the suggested model such as, squared correlation coefficient ($R^2 > 0.7$), adjusted R^2 ($R^2_{adj} > 0.7$), mean squared error MSE (low value), and F-value (high value). In the second stage, the stability and predictive capacity of the proposed QSAR model are evaluated utilizing several internal and external validation criteria, precisely Loo cross-validation procedure by Loo-cv coefficient $Q^2_{cv(loo)} > 0.5$, Y-randomization parameters (R^2_{Y-rand}, Q^2_{Y-rand}) and Golbraikh-Tropsha's parameters (R^2_{pred}, k, k′, $(R^2 - R^2_0)/R^2$, $(R^2 - R'^2_0)/R^2$ and $|R^2_0 - R'^2_0|$) (Golbraikh and Tropsha, 2002, Tropsha, 2010). Then, once the model is validated internally and externally, it is needed to elucidate the Domain of Applicability (AD) by the leverage method for chemical compounds involved in the development of the QSAR model in order to check its reliability for predicting the activities with high confidence.

3. Results and Discussion

3.1 Principal Component Analysis (PCA)

Among the 16 descriptors calculated (**Table 3**) which quantitatively describe the structures of the 39 chemical compounds, only the descriptors that display a high correlation with the activity and a low correlation between them have been exploited to establish a strong and efficient MLR model to predict the inhibitory activity of pyrimidine-2,4-dione derivatives. To this end, these molecular descriptors were submitted to PCA analysis by using the correlation matrix (**Table 4**). We started by excluding all descriptors that have no effect on activity, i.e., those descriptors having correlation coefficients between them and inhibitory activity of less than or equal to 0.1, along with the descriptors that are highly correlated (with a correlation coefficient |r| approximately equal to 1.00). After that, we retained the descriptors that show the lowest correlation with the remaining of them. Because of the results of this analysis, we can select the following descriptors to be utilized in the generation of a multiple linear regression model: E_{HOMO}, $logP_{o/w}$, n, Pc, E_{LUMO}, HBAs, HBDs and d.

Table 3. Values of descriptors calculated for the 39 pyrimidine-2,4-dione derivatives

N°	pIC_{50}	E_{HOMO}	Pc	$logP$	HBA	HBD	E_T	μ	E_{LUMO}	ΔE	η	?	ω	n	S	d	P
1	6.602	-0.220	600.5	2.36	4	3	-1007.465	8.256	-0.053	0.167	0.084	0.137	0.112	1.715	73.5	1.44	31.96
2	6.658	-0.225	637.6	2.96	4	3	-1467.059	6.703	-0.059	0.166	0.083	0.142	0.121	1.718	74.5	1.519	33.9
3	5.167	-0.233	693.3	2.77	4	1	-1071.458	5.964	-0.043	0.190	0.095	0.138	0.100	1.577	51.6	1.23	33.99
4	5.602	-0.238	827.1	4.14	4	1	-1263.199	5.686	-0.045	0.193	0.097	0.142	0.104	1.614	55.6	1.256	41.87
5	5.678	-0.238	834.4	4.28	4	1	-1362.431	6.451	-0.047	0.191	0.096	0.143	0.106	1.603	54.5	1.297	41.86
6	5.208	-0.240	841.8	4.42	4	1	-1461.662	5.606	-0.050	0.190	0.095	0.145	0.111	1.593	53.4	1.337	41.86
7	5.553	-0.232	874.5	4.57	4	1	-1401.746	6.595	-0.046	0.186	0.093	0.139	0.104	1.596	53.3	1.274	43.7
8	5.745	-0.229	954.6	5.46	4	1	-1480.381	4.978	-0.047	0.182	0.091	0.138	0.105	1.584	51.3	1.235	47.37
9	6.046	-0.244	871.6	4.89	4	1	-1822.022	4.394	-0.050	0.194	0.097	0.147	0.111	1.609	55.7	1.356	43.8
10	5.745	-0.242	841.8	4.42	4	1	-1461.663	5.490	-0.049	0.193	0.097	0.146	0.110	1.593	53.4	1.337	41.86
11	5.180	-0.234	814.1	4.51	3	1	-1287.226	5.411	-0.045	0.189	0.095	0.140	0.103	1.607	53.7	1.271	41.17
12	5.420	-0.232	854.2	4.95	3	1	-1326.541	6.502	-0.044	0.188	0.094	0.138	0.101	1.599	52.5	1.249	43.01
13	6.046	-0.229	894.2	5.4	3	1	-1365.858	7.031	-0.044	0.185	0.093	0.137	0.101	1.593	51.5	1.229	44.84
14	5.155	-0.233	857.5	4.51	3	0	-1326.587	4.597	-0.037	0.196	0.098	0.135	0.093	1.572	45.5	1.2	43.1
15	5.208	-0.226	787.2	3.69	4	1	-1223.892	6.083	-0.044	0.182	0.091	0.135	0.100	1.61	52.9	1.255	40.11
16	5.174	-0.213	904.5	3.37	6	1	-1452.942	6.508	-0.043	0.170	0.085	0.128	0.096	1.589	50.1	1.254	45.41
17	5.284	-0.239	834.9	3.54	5	1	-1316.134	6.516	-0.067	0.172	0.086	0.153	0.136	1.653	68.2	1.34	42.17
18	5.678	-0.244	812.9	4.17	4	1	-1521.611	6.063	-0.054	0.190	0.095	0.149	0.117	1.574	48.9	1.367	40.22
19	6.367	-0.238	795.4	4.04	4	1	-1743.397	5.418	-0.052	0.186	0.093	0.145	0.113	1.616	56.5	1.395	40.19
20	7.796	-0.235	612.3	2.65	3	2	-991.426	6.660	-0.043	0.192	0.096	0.139	0.101	1.659	61.7	1.347	31.94
21	7.328	-0.247	674.3	3.53	3	2	-1328.461	5.207	-0.058	0.189	0.095	0.153	0.123	1.594	51.3	1.438	33.91
22	7.602	-0.246	674.3	3.53	3	2	-1328.461	4.804	-0.054	0.192	0.096	0.150	0.117	1.594	51.3	1.438	33.91
23	7.553	-0.240	619.6	2.79	3	2	-1090.657	5.528	-0.048	0.192	0.096	0.144	0.108	1.643	60	1.402	31.93
24	7.796	-0.243	649.4	3.25	3	2	-1451.020	5.091	-0.051	0.192	0.096	0.147	0.113	1.664	63.1	1.427	33.88
25	7.071	-0.248	736.3	4.4	3	2	-1665.490	4.095	-0.065	0.183	0.092	0.157	0.134	1.547	44.2	1.507	35.88
26	7.824	-0.239	627	2.93	3	2	-1189.887	5.538	-0.047	0.192	0.096	0.143	0.107	1.628	58.3	1.456	31.93
27	8.036	-0.246	686.5	3.86	3	2	-1910.608	4.678	-0.051	0.195	0.098	0.149	0.113	1.668	64.4	1.498	35.82
28	8.027	-0.242	656.8	3.39	3	2	-1550.247	4.639	-0.056	0.186	0.093	0.149	0.119	1.649	61.4	1.478	33.87
29	7.824	-0.248	711.4	4.13	3	2	-1788.053	5.547	-0.063	0.185	0.093	0.156	0.131	1.601	52.8	1.503	35.85
30	8.119	-0.219	678.2	2.63	4	2	-1205.177	6.124	-0.043	0.176	0.088	0.131	0.098	1.625	57.2	1.388	34.58
31	7.959	-0.231	657.9	3.3	3	2	-1129.979	6.245	-0.044	0.187	0.094	0.138	0.101	1.633	57.5	1.365	33.85
32	7.796	-0.236	612.3	2.65	3	2	-991.425	6.534	-0.045	0.191	0.096	0.141	0.103	1.659	61.7	1.347	31.94
33	7.432	-0.245	674,3	3.53	3	2	-1328.461	6.291	-0.054	0.191	0.096	0.150	0.117	1.594	51.3	1.438	33.91
34	7.745	-0.241	619.6	2.79	3	2	-1090.657	6.179	-0.049	0.192	0.096	0.145	0.110	1.643	60	1.402	31.93
35	7.921	-0.239	649.4	3.25	3	2	-1451.020	5.998	-0.051	0.188	0.094	0.145	0.112	1.664	63.1	1.427	33.88
36	7.699	-0.243	649.4	3.25	3	2	-1451.019	6.118	-0.051	0.192	0.096	0.147	0.113	1.664	63.1	1.427	33.88
37	7.420	-0.247	711.4	4.13	3	2	-1788.053	5.864	-0.062	0.185	0.093	0.155	0.129	1.601	52.8	1.503	35.85
38	8.114	-0.221	678.2	2.63	4	2	-1205.176	6.441	-0.045	0.176	0.088	0.133	0.101	1.625	57.2	1.388	34.58
39	8.056	-0.232	657.9	3.3	3	2	-1129.979	6.199	-0.045	0.187	0.094	0.139	0.103	1.633	57.5	1.365	33.85

Table 4. Matrix of molecular descriptor correlations

Var	pIC_{50}	E_{HOMO}	logP	E_T	S	μ	P	E_{LUMO}	n	ΔE	HBA	η	Pc	?	HBD	ω	d
pIC_{50}	**1**																
E_{HOMO}	-0.246	**1**															
logP	-0.566	-0.221	**1**														
E_T	**0.054**	0.452	-0.557	**1**													
S	0.354	0.168	-0.574	0.195	**1**												
μ	**-0.074**	0.577	-0.355	0.540	0.421	**1**											
P	-0.812	0.196	0.829	-0.354	-0.460	-0.103	**1**										
E_{LUMO}	-0.204	0.547	-0.013	0.501	-0.195	0.176	0.201	**1**									
n	0.450	0.186	-0.616	0.229	**0.976**	0.412	-0.532	-0.104	**1**								
ΔE	**0.100**	-0.663	0.243	-0.072	-0.368	-0.508	-0.046	0.264	-0.308	**1**							
HBA	-0.582	0.531	0.028	-0.047	0.027	0.274	0.523	0.039	-0.089	-0.577	**1**						
η	**0.100**	-0.663	0.243	-0.072	-0.368	-0.508	-0.046	0.264	-0.308	**1.000**	-0.577	**1**					
Pc	-0.817	0.153	0.843	-0358	-0.537	-0.142	**0.994**	0.179	-0.616	-0.017	0.512	-0.017	**1**				
?	0.258	**-0.910**	0.148	-0.537	-0.010	-0.455	-0.225	-0.845	-0.067	0.293	-0.358	0.293	-0.186	**1**			
HBD	0.788	-0.044	-0.666	0.144	0.616	0.232	-0.833	-0.398	0.668	-0.305	-0.350	-0.305	-0.852	0.225	**1**		
ω	0.209	-0.637	0.046	-0.515	0.154	-0.241	-0.204	**-0.993**	0.066	-0.154	-0.107	-0.154	-0.178	0.899	0.358	**1**	
d	0.737	-0.457	-0.405	-0.343	0.405	-0.174	-0.686	-0.709	0.422	-0.108	-0.365	-0.108	-0.682	0.643	0.809	0.701	**1**

3.2 Multiple linear regression (MLR)

The optimal linear model for the pIC_{50} of pyrimidine-2,4-dione derivatives is obtained by the equation below, which includes four molecular descriptors (E_{LUMO}, d, $logP_{o/w}$ and HBAs) as well as their data of the statistical characteristics mentioned in **Table 5**:

$$pIC_{50} = 1{,}482 + 49{,}364 \times E_{LUMO} + 8{,}297 \times d - 0{,}425 \times logP_{o/w} - 0{,}587 \times HBAs$$

According to the R^2 squared correlation coefficient value, it is evident that the efficiency of the linear model developed with the above molecular descriptors to explain the variability of inhibitory activity is approximately 78%. Also, the values of other statistical characteristics R^2adj, MSE and F obtained by the MLR method reflect the reliability of this linear model. Furthermore, the model has a high Loo-cv coefficient that meets the requirement $Q^2_{cv(Loo)} > 0.5$. This allows us to conclude that the

linear model for inhibitory activity predictions has high stability. **Tables 7** and **8** summarize the values of actual activity $(pIC_{50})_{act}$ versus those of predicted activity $(pIC_{50})_{pred}$ computed through the MLR model of both the training set and test set, respectively. The scatter graph of actual $(pIC_{50})_{act}$ and predicted $(pIC_{50})_{pred}$ values of the linear model obtained by MLR is illustrated in **Figure 2**.

Table 5. Statistical characteristics of the developed linear model

Parameter	Equation	Value
R^2	$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \overline{Y}_{obs})^2}$	0.781
R^2adj	$R^2_{adj} = \frac{(N-1)R^2 - p}{N - p - 1}$	0.747
$Q^2_{cv(Loo)}$	$Q^2_{cv(Loo)} = 1 - \frac{\sum(Y_{\mathbf{pred}(train)} - Y_{obs(train)})^2}{\sum(Y_{obs(train)} - \overline{Y}_{obs(train)})^2}$	0.620
MSE	$MSE = \frac{\sum(Y_{obs} - Y_{pred})^2}{N}$	0.315
F	$F = \left(\frac{N-p-1}{p}\right) \cdot \frac{\sum(Y_{pred} - \overline{Y}_{pred})^2}{\sum(Y_{obs} - Y_{pred})^2}$	23.158

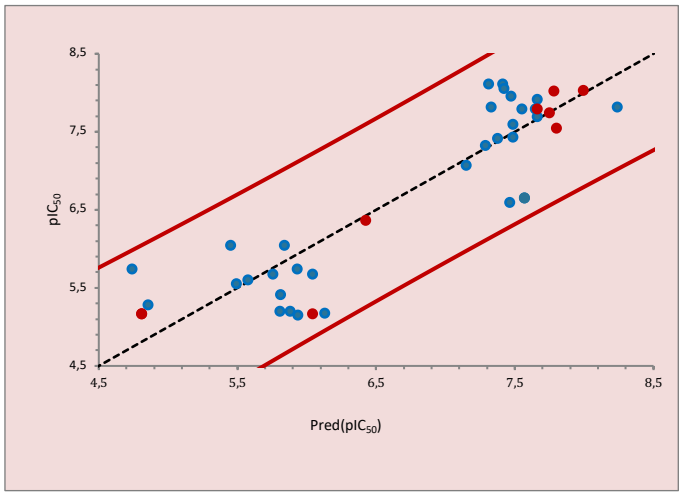


Figure 2. Graphical representation of actual and predicted activities obtained by MLR model

3.3 Multiple nonlinear regression (MNLR)

The main focus of this step is to realize a nonlinear model that may be capable of improving descriptors-activity relationship (Idrissi-Taourati et al, 2015). The MNLR equation for the pIC_{50} of pyrimidine-2,4-dione derivatives, which is based on the descriptors selected by the MLR model, is as follows:

$$pIC_{50} = -63{,}098 + 425{,}945 \times E_{LUMO} + 122{,}987 \times d - 2{,}436 \times logP_{o/w} - 1{,}908 \times HBAs + 3457{,}114 \times E^2_{LUMO} - 40{,}864 \times d^2 + 0{,}302 \times logP^2_{o/w} + 0{,}174 \times HBAs^2$$

The statistical parameters characterizing our optimal MNLR model are shown in **Table 6**. Based on the values of the determination coefficient (a higher R^2) and mean squared error (a lower MSE) for the MNLR model, it is evident that the generated non-linear model is more reliable. The MNLR model's predictive capacity was evaluated internally using the (Leave-one-out) cross-validation technique by Loo-cv coefficient $Q^2_{cv(Loo)} > 0.5$, and externally by external prediction coefficient

$R^2_{pred} > 0.6$. Indeed, the coefficients of Loo-cv $Q^2_{cv(Loo)}$ and external test R^2_{pred} are both greater than 0.6. These good statistical characteristics are considered as an indicator or as evidence that the non-linear QSAR model has a high internal and external predictive capacity. Based on the results obtained by this analysis, we can conclude that the MNLR model is sufficiently robust and thus may be exploited to predict accurately the studied activities pIC_{50} of new effective chemical inhibitors. **Tables 7** and **8** summarize the values of actual activity $(pIC_{50})_{act}$ versus those of predicted activity $(pIC_{50})_{pred}$ computed through the MNLR model of both the training set and test set, respectively. The scatter graph of actual $(pIC_{50})_{act}$ and predicted $(pIC_{50})_{pred}$ values of the nonlinear model obtained by MNLR is illustrated in **Figure 3**.

Table 6. Statistical data of the nonlinear model

Statistical parameter	Value
N_{train}	31
R	0.922
R^2	0.850
R^2_{pred}	0.943
$Q^2_{cv(Loo)}$	0.689
MSE	0.254

Table 7. Values of actual and predicted pIC_{50} of the training set compounds

N°	pIC_{50}	MLR		MNLR	
		Pred(pIC_{50})	Residual	Pred(pIC_{50})	Residual
1	6.602	7.463	-0.861	7.490	-0.888
2	6.658	7.567	-0.910	6.924	-0.266
4	5.602	5.575	0.027	4.990	0.612
5	5.678	5.757	-0.079	5.554	0.123
6	5.208	5.881	-0.674	5.924	-0.716
7	5.553	5.492	0.061	5.316	0.237
8	5.745	4.741	1.004	4.944	0.801
9	6.046	5.839	0.207	6.347	-0.302
10	5.745	5.931	-0.186	6.008	-0.263
11	5.180	6.129	-0.949	6.042	-0.861
12	5.420	5.809	-0.389	5.906	-0.486
13	6.046	5.452	0.594	5.784	0.262
14	5.155	5.935	-0.780	5.619	-0.464
15	5.208	5.807	-0.599	5.119	0.089
17	5.284	4.854	0.430	5.284	0.000
18	5.678	6.039	-0.361	5.993	-0.315
20	7.796	7.649	0.147	8.007	-0.211
21	7.328	7.290	0.038	7.191	0.137
22	7.602	7.487	0.115	7.346	0.256
25	7.071	7.147	-0.076	7.334	-0.264
26	7.824	8.237	-0.413	8.258	-0.435
29	7.824	7.327	0.497	7.263	0.561
30	8.119	7.410	0.709	7.794	0.325
31	7.959	7.473	0.486	7.686	0.272
32	7.796	7.550	0.246	7.764	0.032
33	7.432	7.487	-0.055	7.346	0.086
35	7.921	7.663	0.258	7.578	0.343
36	7.699	7.663	0.036	7.578	0.121
37	7.420	7.377	0.044	7.257	0.164
38	8.114	7.312	0.802	7.551	0.563
39	8.056	7.423	0.632	7.568	0.487

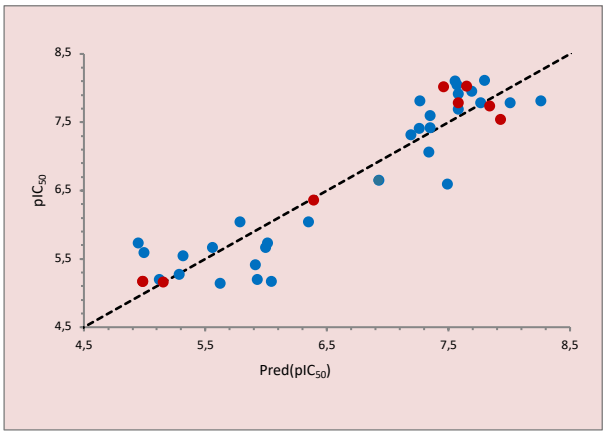


Figure 3. Graphical representation of actual and predicted activities obtained by MNLR model

Table 8. Values of actual and predicted pIC_{50} of the test set compounds

N°	pIC_{50}	MLR		MNLR	
		Pred(pIC_{50})	Residual	Pred(pIC_{50})	Residual
3	5.167	6.040	-0.872	5.153	0.015
16	5.174	4.810	0.364	4.984	0.190
19	6.367	6.425	-0.059	6.389	-0.023
23	7.553	7.799	-0.246	7.925	-0.372
24	7.796	7.663	0.133	7.578	0.218
27	8.036	7.993	0.043	7.649	0.388
28	8.027	7.780	0.247	7.456	0.571
34	7.745	7.749	-0.005	7.835	-0.090

3.4 Y-randomization test

The realistic stability of our linear model may be proved using the Y-scrambling approach. Currently, it is regarded as a mandatory procedure for ensuring that the QSAR model's development did not occur by chance (Shameera Ahamed et al., 2019). It consists of randomly shuffling the pIC_{50} values for the training set (33 samples) to generate random models for several iterations with the same descriptors of the optimal model (El Masaoudy et al., 2020, Khaldan et al., 2020). In the present case, this test was repeated one hundred times. Overall, the obtained models after every random trial are expected to show lower statistical characteristics than those of the optimal model. Indeed, analyzing the statistical characteristics for the random models listed in **Table 9** revealed that the R^2_{Y-rand} and Q^2_{Y-rand} values for each randomized model are lower than those (R^2 and $Q^2_{cv(Loo)}$) of the model suggested by the MLR. Furthermore, the average value of R^2_{Y-rand} and Q^2_{Y-rand} are 0.140 and -0.278, respectively. The obtained results demonstrated that the MLR model has excellent internal predictive ability and high stability.

3.5 Golbraikh and Tropsha's criteria

Although the linear model exhibited stability and good internal predictive power, as previously assessed by Loo-cv and Y-scrambling procedures, this prediction is inaccurate and insufficient. So, to check whether this linear model is real or not to accurately predict the studied activities, we calculated all characteristics or criteria proposed by researchers Golbraikh and Trobsha. Taking into account the overall results mentioned in **Table 10**, it is apparent that all of the external validation criteria were fulfilled, indicating that we have achieved a perfect linear model with excellent predictive ability.

Table 9. Statistical characteristics of 100 random models

Rand	R^2_{Y-rand}	Q^2_{Y-rand}	Rand	R^2_{Y-rand}	Q^2_{Y-rand}	Rand	R^2_{Y-rand}	Q^2_{Y-rand}	Rand	R^2_{Y-rand}	Q^2_{Y-rand}
Rand 1	0.047	-0.335	Rand 26	0.093	-0.439	Rand 51	0.066	-0.310	Rand 76	0.117	-0.657
Rand 2	0.273	-0.034	Rand 27	0.136	-0.236	Rand 52	0.274	-0.014	Rand 77	0.131	-0.378
Rand 3	0.361	0.092	Rand 28	0.126	-0.426	Rand 53	0.080	-0.525	Rand 78	0.079	-0.265
Rand4	0.149	-0.607	Rand 29	0.333	-0.258	Rand 54	0.147	-0.145	Rand 79	0.309	0.028
Rand 5	0.129	-0.294	Rand 30	0.135	-0.175	Rand55	0.078	-0.302	Rand 80	0.040	-0.348
Rand 6	0.081	-0.372	Rand 31	0.145	-0.193	Rand 56	0.164	-0.166	Rand 81	0.296	0.045
Rand 7	0.232	-0.243	Rand 32	0.129	-0.212	Rand 57	0.045	-0.289	Rand 82	0.256	0.025
Rand 8	0.101	-0.248	Rand 33	0.184	-0.136	Rand 58	0.103	-0.239	Rand 83	0.225	-0.153
Rand 9	0.075	-0.290	Rand 34	0.255	-0.131	Rand 59	0.074	-0.252	Rand 84	0.108	-0.254
Rand 10	0.115	-0.350	Rand 35	0.087	-0.339	Rand 60	0.114	-0.226	Rand 85	0.191	-0.464
Rand 11	0.088	-0.288	Rand 36	0.234	-0.249	Rand 61	0.133	-0.205	Rand 86	0.226	-0.359
Rand 12	0.241	-0.117	Rand 37	0.062	-0.377	Rand 62	0.190	-0.165	Rand 87	0.165	-0.122
Rand 13	0.093	-0.303	Rand 38	0.234	-0.197	Rand 63	0.166	-0.108	Rand 88	0.046	-1.016
Rand14	0.107	-0.418	Rand 39	0.279	0.040	Rand 64	0.156	-0.177	Rand 89	0.086	-0.331
Rand15	0.189	-0.115	Rand 40	0.108	-0.557	Rand 65	0.144	-0.384	Rand 90	0.026	-0.436
Rand16	0.060	-0.399	Rand 41	0.111	-0.225	Rand 66	0.264	-0.061	Rand 91	0.045	-0.603
Rand17	0.096	-0.278	Rand 42	0.117	-0.208	Rand 67	0.179	-0.302	Rand 92	0.044	-0.464
Rand18	0.348	0.067	Rand 43	0.340	-0.099	Rand 68	0.117	-0.174	Rand 93	0.058	-0.361
Rand19	0.116	-0.227	Rand 44	0.029	-0.495	Rand 69	0.019	-0.415	Rand 94	0.363	0.102
Rand20	0.068	-0.261	Rand 45	0.088	-0.296	Rand 70	0.075	-0.364	Rand 95	0.108	-0.211
Rand21	0.049	-0.452	Rand46	0.100	-0.294	Rand 71	0.073	-0.339	Rand 96	0.026	-0.689
Rand22	0.247	-0.157	Rand 47	0.155	-0.221	Rand 72	0.139	-0.187	Rand 97	0.187	-0.211
Rand23	0.112	-0.223	Rand 48	0.090	-0.343	Rand 73	0.195	-0.131	Rand 98	0.092	-0.368
Rand24	0.052	-0.629	Rand 49	0.202	-0.436	Rand 74	0.116	-0.182	Rand 99	0.205	-0.125
Rand25	0.050	-0.370	Rand 50	0.037	-0.441	Rand 75	0.098	-0.248	Rand100	0.065	-0.484

Table 10. External validation results of MLR model

Parameter	Equation	Value	Threshold
R^2_{pred}	$R^2_{pred} = 1 - \frac{\sum(Y_{pred(test)} - Y_{obs(test)})^2}{\sum(Y_{obs(test)} - \overline{Y}_{obs(train)})^2}$	0.905	> 0.6
R^2_0	$R^2_0 = 1 - \frac{\sum(Y_{pred} - k \times Y_{pred})^2}{\sum(Y_{pred} - \overline{Y}_{pred})^2}$	0.998	> 0.6
R'^2_0	$R'^2_0 = 1 - \frac{\sum(Y_{obs} - k' \times Y_{obs})^2}{\sum(Y_{obs} - \overline{Y}_{obs})^2}$	0.999	> 0.6
$\lvert R^2_0 - R'^2_0 \rvert$	$\lvert R^2_0 - R'^2_0 \rvert$	0.001	< 0.3
$\frac{(R^2 - R^2_0)}{R^2}$	$\frac{(R^2 - R^2_0)}{R^2}$	-0.278	< 0.1
$\frac{(R^2 - R'^2_0)}{R^2}$	$\frac{(R^2 - R^2_0)}{R^2}$	-0.279	< 0.1
k	$k = \frac{\sum(Y_{obs} \times Y_{pred})}{\sum Y^2_{pred}}$	0.994	$0.85 \leq k \leq 1.15$
k'	$k' = \frac{\sum(Y_{obs} \times Y_{pred})}{\sum Y^2_{obs}}$	1.004	$0.85 \leq k' \leq 1.15$

3.6 Applicability domain

In line with principle 3 of the Organization for Economic Co-operation and Development (OECD) (OECD, 2007), after the QSAR model is validated internally and externally, the applicability area of the training set must be defined. It is an area of chemical space that allows us to detect each compound that lies within or outside this region using Williams' plot (Gramatica, 2007). The leverage threshold is expressed by the relation: $h^* = 3\,p/k$ (Netzeva et al., 2005), where p refers to the number of model descriptors plus one ($p = 5$) and k refers to the number of samples in the training set ($k = 31$). Only the compounds found inside the DA may be considered able to accurately predict the anti-HIV activity of new pyrimidine-2,4-dione derivatives (i.e., when the leverage chemical h is lower than 0.484). Conversely, the response is regarded as an outlier for each chemical compound involved in the development of the linear model when the leverage chemical exceeds the threshold ($h > 0.484$). From **Figure 4**, it has been observed that the leverage value of each compound does not reach the threshold, except compound n°17, which has a leverage value that well surpasses the critical value.

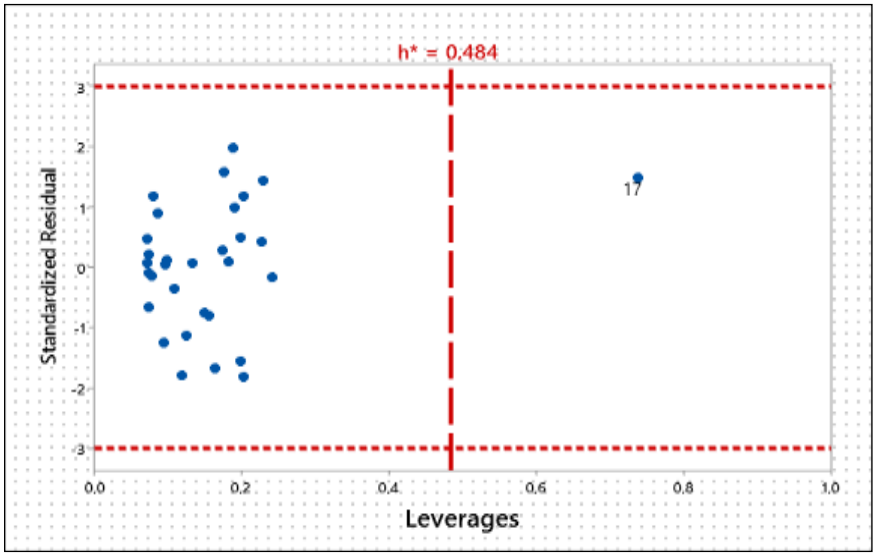


Figure 4. Model training set's applicability domain

3.7 Proposed new candidate compounds

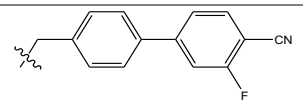
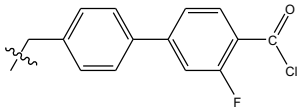
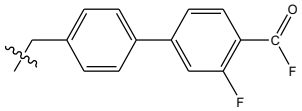
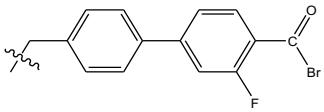
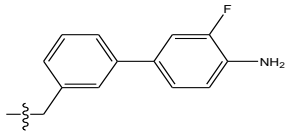
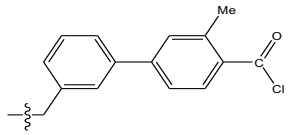
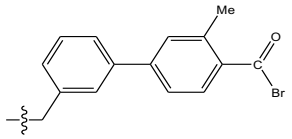
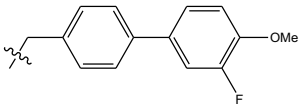
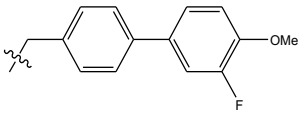
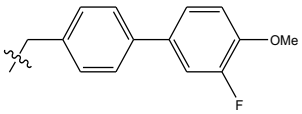
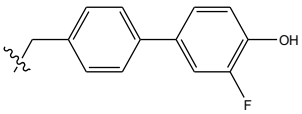
The MLR equation shows that lowest unoccupied molecular orbital energy (E_{LUMO}) and density (d) have a positive influence on pIC_{50} activity, unlike the number of H-bond acceptors (HBAs) and octanol-water partition coefficient $logP_{o/w}$ have a negative influence. Accordingly, the inhibitory activity may be increased by increasing the E_{LUMO} and density values, then decreasing the $logP_{o/w}$ and HBAs values.

The t-test value allows for determining the significance of each descriptor participating in this linear model on the pIC_{50} inhibitory activity values. By comparing the absolute t-test values for descriptors E_{LUMO}, d, $logP_{o/w}$ and HBAs, which are $|1.768|$, $|3.073|$, $|-2.314|$, and $|-2.671|$, respectively, it is obvious that the descending order of influence of these descriptors towards activity is as follows: $d > HBAs > logP_{o/w} > E_{LUMO}$.

Decreasing the number of heteroatoms in pyrimidine-2,4-dione derivatives, such as oxygen O and nitrogen N atoms and halogens, leads to decreasing the HBAs value. In addition, the presence of a hydrophilic group (e.g., carboxyl COOH, hydroxyl OH and amino NH_2) may result in a decrease in $logP_{o/w}$ value. Density is a steric descriptor related to the mass and size of the molecule. Decreasing the size of substituents may lead to an increase in the value of density (d). The lowest energy level in a chemical species that has an empty orbital is the LUMO. The E_{LUMO} is a quantum descriptor that measures the ability of a chemical species to accept a pair of electrons and hence its susceptibility toward nucleophilic attack. When a molecule has a large LUMO energy, it behaves as a hard electrophile that will receive electrons more easily. Consequently, to increase the value of E_{LUMO} having a positive sign in the equation, we'll have to replace the pyrimidine-2,4-dione derivatives with positively charged or neutral species which are more susceptible to nucleophilic attack (e.g., haloformyl COX and carboxyl COOH).

In short, the different information captured about the structural properties of the 39 pyrimidine-2,4-dione derivatives by the above descriptors led to the design of a novel set of chemical inhibitors with improved activity, as shown in **Table 11**, by adding appropriate substituents based on the proposed model equations.

Table 11. New candidate compounds with inhibitory activity

N°	R_1	R_2	R_3	R_4	E_{LUMO}	d	$logP_{o/w}$	HBAs	pIC_{50}	
									MLR	MNLR
1	H		H	OH	-0.079	1.49	2.65	4	6.471	**8.174**
2	H		H	OH	-0.092	1.504	2.98	4	5.805	**10.088**
3	H		H	OH	-0.085	1.473	2.69	4	6.016	**8.955**
4	H		H	OH	-0.092	1.671	3.13	4	7.127	**8.871**
5	H		H	OH	-0.043	1.455	1.96	4	**8.250**	**8.952**
6	H		H	OH	-0.087	1.420	3.35	4	5.197	**8.636**
7	H		H	OH	-0.088	1.579	3.50	4	6.403	**8.829**
8	H		COCl	OH	-0.107	1.495	2.75	5	4.500	**13.835**
9	H		COOH	OH	-0.079	1.504	2.22	6	5.595	**8.262**
10	H		Br	OH	-0.050	1.626	3.46	4	**8.686**	6.524
11	OH		OH	OH	-0.053	1.731	2.24	6	**8.754**	5.36

3.8 Lipinski's RO5 and ADME/Tox prediction

The role of five RO5 (or the set of rules), introduced by Christopher Lipinski, is a tool that plays a pivotal role in the drug discovery path. It is employed to estimate the drug likeness of a molecule with favorable physicochemical properties to be an orally bioavailable drug (Chen et al., 2020). According to Lipinski's RO5 (Lipinski et al., 1997), if a molecule meets the following criteria, then it could be a more promising orally active drug: octanol/water partition coefficient $logP_{o/w} \leq 5$, number of H-bond acceptors HBAs ≤ 10, number of H-bond donors HBDs ≤ 5, molecular weight MWT $\leq$

$500\,g/mol$ and number of rotatable bonds $\text{Rot B} \leq 10$. The number of HBAs, Rot B and HBDs, $\log P_{o/w}$ and MWT of the newly designed pyrimidine-2,4-dione derivatives were calculated using the SwissADME online tool (Daina et al., 2017) (**Table 12**). All the results obtained for the drug-likeness properties (HBAs, Rot B and HBDs, $\log P_{o/w}$ and MWT) of the newly proposed inhibitors are in accord with the role of five. As a result, these molecules showed promising bioavailability for oral drugs.

In silico ADME/Tox prediction, characterized by five pharmacokinetic properties (PK) (Absorption, Distribution, Metabolism, Excretion, Toxicity), is a crucial stage of the discovery process that guides the selection of qualified candidates for drug design as well as the removal of candidate drugs that could fail during clinical trials (Tsaioun et al., 2009, Santos et al., 2015, Tabti et al., 2022a). For this purpose, in silico PK parameters of eleven candidates reported in **Table 11** were estimated by using the pkCSM online tool (Pires, et al., 2015). The ADME/Toxicity results are summarized in **Table 13**.

Table 12. Physicochemical property parameters of new pyrimidine-2,4-dione derivatives

Proposed compound	Molecular Weight	Octanol-water partition coefficient	Number of rotatable bonds	Number of H-bond acceptors	Number of H-bond donors
1	337.30	2.56	3	5	2
2	374.75	3.04	4	5	2
3	358.30	2.89	4	6	2
4	419.20	3.07	4	5	2
5	327.31	2.26	3	4	3
6	370.79	3.05	4	4	2
7	415.24	3.17	4	4	2
8	404.78	3.05	5	6	2
9	386.33	2.40	5	7	3
10	421.22	3.41	4	5	2
11	360.29	1.67	3	7	4
Rule	$\text{MWT} \leq 500\,g/mol$	$\log P_{o/w} \leq 5$	$\text{Rot B} \leq 10$	$\text{HBAs} \leq 10$	$\text{HBDs} \leq 5$

According to intestinal absorption values ($\text{HIA} > 30\%$), all newly designed pyrimidine-2,4-dione derivatives may be absorbed through the human intestine. In addition, most of these derivatives are soluble ($\log S > -4$), while other derivatives (8 and 10) are moderately soluble ($-6 < \log S \leq -4$).

A new derivative (2, 3, 4, 6, 7, 8, 9 and 10) with a logBB value of < -1 does not readily cross the BBB, whereas a new derivative (3, 8, 9 and 11) with a logPS value of < -3 is poorly at penetrating the CNS. Accordingly, derivatives 1 and 5 possess the ability to cross the barriers.

It is important to evaluate which derivatives in the new dataset (**Table 11**) may be substrates or inhibitors of the CYP450 3A4 isoenzyme. The metabolic properties of all eleven pyrimidine-2,4-dione derivatives were evaluated. As a result, these derivatives, except for 8 and 10, turned out to be non-inhibitors of CYP450 3A4.

The obtained results for excretion property (**Table 13**) indicated that all newly designed pyrimidine-2,4-dione derivatives, except for 1 and 9, had a low value of total tolerance log (CLtot).

The Ames toxicity test is a mutagenicity test utilized to assess the novel derivative for its potential to be a mutagenic (Mortelmans et el., 2000). A negative Ames test suggests that this derivative is non-mutagenic and therefore does not pose a cancer risk. Our Ames test results revealed that all newly proposed pyrimidine-2,4-dione derivatives are non-mutagenic except for derivative 5 that is capable of inducing cancer.

Table 13. ADME/Tox predictions of newly proposed pyrimidine-2,4-dione derivatives

N°	Absorption		Distribution		Metabolism							Excretion	Toxicity
	WS	HIA	BBT	CNS	Substrate CYP		Inhibitor CYP					TC	AT
					CYP 2D6	CYP 3A4	CYP 1A2	CYP 2C19	CYP 2C9	CYP 2D6	CYP 3A4		
1	-3.533	75.54	-0.896	-2.6	No	Yes	Yes	No	Yes	No	No	0.707	No
2	-3.386	73.358	-1.212	-2.594	No	Yes	Yes	Yes	Yes	No	No	-0.368	No
3	-3.194	69.218	-1.284	-3.245	No	Yes	Yes	No	No	No	No	0.181	No
4	-3.45	73.304	-1.239	-2.562	No	Yes	Yes	Yes	Yes	No	No	-0.432	No
5	-3.45	70.236	-0.979	-2.647	No	Yes	Yes	No	Yes	No	No	0.297	Yes
6	-3.647	75.686	-1.025	-2.472	No	Yes	Yes	Yes	Yes	No	No	-0.462	No
7	-3.72	75.631	-1.052	-2.439	No	Yes	Yes	Yes	Yes	No	No	-0.527	No
8	-4.876	71.719	-1.319	-3.442	No	Yes	No	Yes	Yes	No	Yes	-0.317	No
9	-3.072	36.716	-1.397	-3.649	No	No	No	No	No	No	No	0.658	No
10	-4.551	91.559	-1.191	-2.523	No	Yes	Yes	Yes	Yes	No	Yes	-0.338	No
11	-3.923	59.619	-1.308	-3.614	No	No	Yes	No	No	No	No	0.25	No

Abbreviations: Water Solubility (WS, log mol/L), Human Intestinal Absorption (HIA, % Absorbed), Blood-Brain Barrier (BBB, log BB), Central Nervous System (CNS, log PS), Cytochrome P450 (CYP), Total Clearance (TC, log ml/min/kg), Ames Toxicity (AT).

Conclusion

In this research paper, we have developed linear and nonlinear models through two 2D-QSAR techniques (MLR and RNLM) to formulate quantitative relationships between HIV RT RNase H inhibitory activity for 39 pyrimidine-2,4-dione derivatives and their physicochemical characteristics.

The two regression models used to describe the HIV RT-associated RNase H inhibitory activity of pyrimidine-2,4-dione derivatives revealed that density (d), octanol/water partition coefficient ($\log P_{o/w}$), LUMO energy (E_{LUMO}), and the number of hydrogen bond acceptors (HBAs) are the pertinent descriptors.

According to the validation principles recommended by the OECD, the ability of the above models to predict the inhibitory activity of pyrimidine-2,4-dione derivatives was confirmed.

The obtained results in this study demonstrated that the presence of a hydrophilic group and a hard electrophile, as well as reducing the size of substituents and the number of heteroatoms in pyrimidine-2,4-dione derivative, can increase the inhibitory activity. Consequently, the eleven pyrimidine-2,4-dione derivatives that were proposed as chemical inhibitors exhibited good inhibitory activities. In particular, inhibitors 2 and 8 showed the highest activity with $pIC_{50} = 10.088M$ and $pIC_{50} = 13.835M$, respectively.

Lipinski's parameters (HBAs, Rot B and HBDs, $\log P_{o/w}$, MWT) and in silico ADME/Toxicity predictions suggest that most novel inhibitors seem to be orally bioavailable, show favorable pharmacokinetic properties, and do not pose a cancer risk. So, these chemical inhibitors are suitable candidates for drug discovery and development. Finally, it can be said that in silico approaches employed in this research paper yielded promising results that could help medicinal chemists to design more potent anti-HIV RT RNase H drugs.

Acknowledgement: We are grateful to the ''Association Marocaine des Chimistes Théoriciens'' (AMCT) for its pertinent help concerning the programs.

Disclosure statement: *Conflict of Interest:* The authors declare that there are no conflicts of interest.
Compliance with Ethical Standards: This article does not contain any studies involving human or animal subjects.

References

Addinsoft (2021). XLSTAT statistical and data analysis solution. New York, USA. https://www.xlstat.com.

Beck A. D., (1993) Density-functional thermochemistry. III. The role of exact exchange. The Journal of Chemical Physics, 98(7), 5648–5652.

Cihlar T., and Ray A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. Antiviral research, 85(1), 39-58. https://doi.org/10.1016/j.antiviral.2009.09.014

Consonni V., and Todeschini R. (2010) Molecular Descriptors. In: Puzyn T., Leszczynski J., and Cronin M. T., (Eds.), Recent Advances in QSAR Studies, C*hallenges and Advances in Computational Chemistry and Physics.* Springer, Dordrecht, 8, 29–102.

ChemAxon - Software Solutions and Services for Chemistry & Biology. https://chemaxon.com/ (accessed Jul. 12, 2021).

Chemistry Software for Analytical and Chemical Knowledge Management. https://www.acdlabs.com/ (accessed Jul. 12, 2021).

Chen X., Li H., Tian L., Li Q., Luo J., and Zhang Y. (2020) Analysis of the Physicochemical Properties of Acaricides Based on Lipinski's Rule of Five. *Journal of computational biology,* 27(9), 1397-1406. https://doi.org/10.1089/cmb.2019.0323

Cramer C. J. (2004) Essentials of Computational Chemistry: Theories and Models, 2nd ed, John Wiley & Sons Ltd.

Daina A., Michielin O., and Zoete V. (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports,* 7(1), 1-13. https://doi.org/10.1038/srep42717

De Broucker T. (2013) Neurological complications associated with human immunodeficiency virus (HIV). *Pratique neurologique-FMC*, 4(4), 213-228. https://doi.org/10.1016/j.praneu.2013.10.002

Elidrissi B., Ousaa A., Ghamali M., Chtita S., Ajana M. A., Bouachrine M., Lakhlifi T. (2018) Quantitative Structure–Activity Relationship (QSAR) Studies of Some Glutamine Analogues for Possible Anticancer Activity, *Mor. J. Chem.* 6 (4) 752-766.

Idrissi-Taourati A., Adad A., Guenoun F., Lakhlifi T., Hamidi M., Bouachrine M. (2015) QSAR Models predict Olfactive threshold relationships for Pyrazine derivatives, *Mor. J. Chem.* 3 (3) 379-393.

El Masaoudy Y., Aanouz I., Moukhliss Y., Koubi Y., Maghat H., Lakhlifi T., and Bouachrine M. (2020) 2D-QSAR study of the antimicrobial activity of a series of 5-(substituted benzaldehyde) thiazolidine-2,4-dione derivatives against Staphylococcus aureus by Multiple Linear Regression method. *Journal of Materials and Environmental Science,* 11(11), 1914-1927.

Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Petersson G. A., Nakatsuji H., Li, X., Caricato M., Marenich A. V., Bloino J., Janesko B. G., Gomperts R., Mennucci B., Hratchian H. P., Ortiz J. V., Izmaylov A. F., Sonnenberg J. L., Williams-Young D., Ding F., Lipparini F., Egidi F., Goings J., Peng B., Petrone A., Henderson T., Ranasinghe D., Zakrzewski V. G., Gao J., Rega N., Zheng G., Liang W., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Throssell K., Montgomery J. A. Jr., Peralta J. E., Ogliaro F., Bearpark M. J., Heyd J. J., Brothers E. N., Kudin K. N., Staroverov V. N., Keith T. A., Kobayashi R., Normand J., Raghavachari K., Rendell A. P., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Millam J. M., Klene M., Adamo C., Cammi R., Ochterski J. W., Martin R. L., Morokuma K., Farkas O., Foresman J. B., Fox D. J. Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009.

Gerbouin O., and Grellet J. (2017) Human immunodeficiency virus: viral cycle and epidemiology. Actualites Pharmaceutiques, 56(564), 1-2. https://doi.org/10.1016/j.actpha.2016.12.001

Golbraikh A., and Tropsha A. (2002) Beware of q2!. *Journal of molecular graphics and modelling,* 20(4), 269-276. https://doi.org/10.1016/S1093-3263(01)00123-1

Gramatica P. (2007) Principles of QSAR models validation: internal and external. *QSAR & combinatorial science*, 26(5), 694-701. https://doi.org/10.1002/qsar.200610151

Hammache D., Yahi N., and Fantini J. (2000) Glycosphingolipids and virus-cell fusion: current data highlighting the role of membrane microdomains in the HIV-1 infection cycle. *Oléagineux, Corps gras, Lipides,* 7(5), 449–455. https://doi.org/10.1051/ocl.2000.0449

Hang J. Q., Li Y., Yang Y., Cammack N., Mirzadegan T., and Klumpp K. (2007) Substrate-dependent inhibition or stimulation of HIV RNase H activity by non-nucleoside reverse transcriptase inhibitors (NNRTIs). *Biochemical and Biophysical Research Communications,* 352(2), 341–350.

https://doi.org/10.1016/j.bbrc.2006.11.018
Hariharan P. C., and Pople J. A., (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta,* 28(3), 213-222. https://doi.org/10.1007/BF00533485
Ilina T., and Parnia M. A., (2008) Inhibitors of HIV-1 Reverse Transcriptase. *Advances in Pharmacology,* 56, 121–167. https://doi.org/10.1016/S1054-3589(07)56005-9
Julias J. G., McWilliams M. J., Sarafianos S. G., Arnold E., and Hughes S. H. (2002) Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage in vivo. *Proceedings of the National Academy of Sciences,* 99(14), 9515–9520. https://doi.org/10.1073/pnas.142123199
Nikolenko G. N., Palmer S., Maldarelli F., Mellors J. W., Coffin J. M., and Pathak V. K. (2005) Mechanism for nucleoside analog-mediated abrogation of HIV-1 replication: Balance between RNase H activity and nucleotide excision. *Proceedings of the National Academy of Sciences,* 102(6), 2093–2098. https://doi.org/10.1073/pnas.0409823102
Khaldan A., El khatabi K., El-mernissi R., Sbai A., Bouachrine M., Lakhlifi T. (2020) Combined 3D-QSAR Modeling and Molecular Docking Study on metronidazole-triazole-styryl hybrids as antiamoebic activity, *Mor. J. Chem.* 8 (2) 527-539.
Klumpp K. & Mirzadegan T. (2006) Recent Progress in the Design of Small Molecule Inhibitors of HIV RNase H. Current Pharmaceutical Design, 12(15), 1909–1922. doi.org/10.2174/138161206776873653
Koubi Y., Moukhliss Y., Sbai A., El Masaoudy Y., Maghat H., EL-Mernissi R., Aziz Ajana M., Bouachrine M., Lakhlifi T., (2019) Antimicrobial evaluation against Escherichia coli (MTCC 1652) by using 1, 4-disubstituted 1, 2, 3-triazole and derivatives: QSAR study. *Arabian J. Chem. Environ. Res,* 06(1), 57–69.
Lee C., Yang W., and Parr R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B,* 37(2), 785–789. https://doi.org/10.1103/PhysRevB.37.785
Lafridi H., Ousaa A., Zgou H., Bouachrine M., (2020) QSAR modeling of antiradical properties of phenolic compounds using DFT calculations, *Mor. J. Chem.* 8 (4) 841-852.
Lipinski C. A., Lombardo F., Dominy B. W., and Feeney P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 23(1-3), 3-25. https://doi.org/10.1016/j.addr.2012.09.019
Mortelmans K., and Zeiger E. (2000) The Ames Salmonella/microsome mutagenicity assay. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 455(1-2), 29–60. https://doi.org/10.1016/S0027-5107(00)00064-6
Netzeva T. I., Worth A. P., Aldenberg T., Benigni R., Cronin M. T. D, Gramatica P., Jaworska J. S., Kahn S., Klopman G., Marchant C. A., Myatt G., Nikolova-Jeliazkova N., Patlewicz G. Y., Perkins R., Roberts D. W., Schultz T. W., Stanton D. T., van de Sandt J. J. M., Tong W., Veith G., Yang C. (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Alternatives to Laboratory Animals,* 33(2), 155-173. https://doi.org/10.1177/026119290503300209
OECD Guidance Document on the Validation of QSAR Models, Organization for Economic Co-operation & Development, Paris, 2007.
Pires D. E. V., Blundell T. L., and Ascher D. B. (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry,* 58(9), 4066-4072. https://doi.org/10.1021/acs.jmedchem.5b00104
Tabti K., Elmchichi L., Sbai A., Maghat H., Bouachrine M., Lakhlifi T., Ghosh A. (2022) In silico design of novel PIN1 inhibitors by combined of 3D-QSAR, molecular docking, molecular dynamic simulation and ADMET studies. *Journal of Molecular Structure,* 1253, 132291, doi.org/10.1016/j.molstruc.2021.132291

Tabti K., Sbai A., Maghat Hamid., Bouachrine M., Lakhlifi T., (2020) QSAR studies of new compounds based on thiazole derivatives as PIN1 inhibitors via statistical methods. *RHAZES: Green and Applied Chemistry*, 9, 70-91.

Tabti K., El Mchichi L., Moukhliss Y., Singh G., Sbai A., maghat H., bouachrine M., Lakhlifi T. (2022). CoMFA Topomer, CoMFA, CoMSIA, HQSAR, docking molecular, dynamique study and ADMET study on phenyloxylpropyl isoxazole derivatives for coxsackie virus B3 virus inhibitors activity, *Mor. J. Chem*., 10, 703-725, https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i4.34319

Tropsha A. (2010) Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular informatics,* 29(6-7), 476-488. https://doi.org/10.1002/minf.201000061

Tsaioun K., Bottlaender M., and Mabondzo A. (2009) ADDME – Avoiding Drug Development Mistakes Early: central nervous system drug discovery perspective. *BMC neurology,* 9(1), 1-11. https://doi.org/10.1186/1471-2377-9-S1-S1

Santos L. H., Ferreira R. S., and Caffarena E. R. (2015) Computational drug design strategies applied to the modelling of human immunodeficiency virus-1 reverse transcriptase inhibitors. *Memórias do Instituto Oswaldo Cruz,* 110(7), 847-864. https://doi.org/10.1590/0074-02760150239

Shameera Ahamed T. K., Rajan V. K., and Muraleedharan K. (2019) QSAR modeling of benzoquinone derivatives as 5-lipoxygenase inhibitors. *Food Science and Human Wellness*, 8(1), 53-62. https://doi.org/10.1016/j.fshw.2019.02.001

Spence R. A., Kati W. M., Anderson K.S., and Johnson K.A. (1995) Mechanism of Inhibition of HIV-1 Reverse Transcriptase by Nonnucleoside Inhibitors, *Science,* 267 (5200), 988–993. https://doi.org/10.1126/science.7532321

Wang L., Tang J., Huber A. D., Casey M. C., Kirby K. A., Wilson D. J., Kankanala J., Xie J., Parniak M. A., Sarafianos S. G., and Wang Z. (2018) 6-Arylthio-3-hydroxypyrimidine-2,4-diones potently inhibited HIV reverse transcriptase-associated RNase H with antiviral activity. *European journal of medicinal chemistry*, 156, 652–665. https://doi.org/10.1016/j.ejmech.2018.07.039

Wang L., Tang J., Huber A. D., Casey M. C., Kirby K. A., Wilson D. J., Kankanala J., Parniak M. A., Sarafianos S. G., and Wang Z. (2018) 6-Biphenylmethyl-3-hydroxypyrimidine-2,4-diones potently and selectively inhibited HIV reverse transcriptase-associated RNase H. European journal of medicinal chemistry, 156, 680–691. https://doi.org/10.1016/j.ejmech.2018.07.035