Aouidate *et al. Chemistry Central Journal (2018) 12:32*
https://doi.org/10.1186/s13065-018-0401-x

Chemistry Central Journal

RESEARCH ARTICLE

Open Access

QSAR study and rustic ligand-based virtual screening in a search for aminooxadiazole derivatives as PIM1 inhibitors

Adnane Aouidate[1*], Adib Ghaleb[1], Mounir Ghamali[1], Samir Chtita[1], Abdellah Ousaa[1], M'barek Choukrad[1], Abdelouahid Sbai[1], Mohammed Bouachrine[2] and Tahar Lakhlifi[1]

Abstract

Background: Quantitative structure–activity relationship (QSAR) was carried out to study a series of aminooxadiazoles as PIM1 inhibitors having pk_i ranging from 5.59 to 9.62 (k_i in nM). The present study was performed using Genetic Algorithm method of variable selection (GFA), multiple linear regression analysis (MLR) and non-linear multiple regression analysis (MNLR) to build unambiguous QSAR models of 34 substituted aminooxadiazoles toward PIM1 inhibitory activity based on topological descriptors.

Results: Results showed that the MLR and MNLR predict activity in a satisfactory manner. We concluded that both models provide a high agreement between the predicted and observed values of PIM1 inhibitory activity. Also, they exhibit good stability towards data variations for the validation methods. Furthermore, based on the similarity principle we performed a database screening to identify putative PIM1 candidates inhibitors, and predict their inhibitory activities using the proposed MLR model.

Conclusions: This approach can be easily handled by chemists, to distinguish, which ones among the future designed aminooxadiazoles structures could be lead-like and those that couldn't be, thus, they can be eliminated in the early stages of drug discovery process.

Keywords: PIM1, Aminooxadiazoles, QSAR model, Applicability domain, MLR, Virtual screening

Introduction

Proviral integration site for Moloney murine leukemia virus (PIM) is a family of serine/threonine protein kinases that are widely expressed and are involved in cell survival and proliferation as well as a number of other signal transduction [1, 2]. This family is composed of three isoforms: PIM1, PIM2, and PIM3 that share a high level of sequence homology and exhibit some functional redundancy. Over-expression of PIM1 and PIM2 kinases has been reported in hematologic malignancies also in solid tumors such as diffuse large B cell lymphomas (DLBCL) and prostate cancer [3], thus, these findings make it an attractive target for cancer therapy [1].

Several heterocycles have been studied with different approaches so far, as 5-(1H-indol-5-yl)-1,3,4-thiadiazol-2-amines [4] and pyrrolo carbazole [5], thiazolidine [6] including many clinical compounds as SGI-1776 [7] and AZD-1208 [8] that have been found to be able to inhibit PIM1 kinase and exhibit an anti-cancer activity. However, no PIM1 inhibitor has crossed all stages of drug discovery process and approved as a drug yet, therefore there is always a need to discover and identify new PIM1 inhibitors. Consequently, in order to reduce time and cost, in addition to design and identify more potent PIM inhibitors, theoretical research can circumvent

*Correspondence: a.aouidate@hotmail.fr
[1] MCNSL, School of Sciences, Moulay Ismail University, Meknes, Morocco
Full list of author information is available at the end of the article

Springer Open

© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

these difficulties and allow obtaining precise data while taking advantage of the rapid progress of computing chemical descriptors, which can be obtained easily from publicly available software and servers. Descriptors can be exploited to build a quantitative structure–activity relationship (QSAR) model to enable calculation of the activity and prediction of the efficacy of new potent aminooxadiazoles. In the recent years, many QSAR studies have been developed on different PIM1 heterocycle inhibitors [9, 10], despite, it would be worthwhile to extend these data and develop QSAR studies on new PIM1 inhibitors. Recently, a series of some potent PIM1 inhibitors: have been designed and reported by Wurz et al. [11]. We believe that this is the first QSAR study performed on the reported activities of this series. That prompted us to aim an in silico study based on it to design new molecules with enhanced inhibitory activity.

Quantitative structure activity relationship is one of the most common approach in computer aided drug design [12] as well as in many other applications, including predictive toxicology, and risk assessment [13, 14]. QSAR studies are based on the fact that the biological activities of organic molecules depend on their chemical structures, and can be quantitatively described by chemometrics models. This approach has a wide application for evaluating the potential impact of chemicals on human health, and technological processes as in the pharmaceutical industry and drug discovery [15]. Thus, it is necessary to develop a QSAR model for the prediction of activity before synthesis of new PIM1 inhibitors. A successful QSAR model not only, helps to understand relationships between the structural properties and biological activity of any class of molecules, but also provides researchers a deep analysis about the lead molecules to be used in further studies [16].

The present study aims to derive QSAR models, which explain the relationship between the anti-cancer activity and the structure of 34 compounds based on physicochemical descriptors using several chemometric methods such as genetic functional algorithm for variable selection GFA, multiple linear regression MLR and non-linear regression MNLR for modeling and William's plot for applicability domain. Finally, PubChem database was virtually screened using the most active compound in the series as a reference molecule.

Materials and methods

For QSAR studies a series of 34 aminooxadiazoles with reported activity values were compiled from the literature [11]. The activity was expressed as k_i and is defined as the binding affinity constants of aminooxadiazoles to PIM1 kinase. Because the inhibitory activity values cover a wide range, they are converted into logarithm units ($pk_i = -\log k_i$) (k_i in nM) for modelling purposes. Figure 1 and Table 1 show the substituted structures of the studied compounds. For modeling, the data set was split into two sets. Twenty-seven molecules were chosen based on the activity variation to represent the quantitative model (training set) and the rest were used to test the performance of proposed model (Test set). Additionally leave-one-out protocol and Y-Randomization were performed on the training set for internal validation of the obtained models.

Molecular modeling

All modeling studies were performed using the SYBYL-X 2.0 molecular modeling package (Tripos Inc., St. Louis, USA) running on a windows 7, 32 bit workstation. Three-dimensional structures were built using the SKETCH option in SYBYL. All compounds were minimized under the Tripos standard force field [17] with Gasteiger–Hückel atomic partial charges [18] by the Powell method with a gradient convergence criterion of 0.01 kcal/mol Å. To describe the compound structural diversity and in order to obtain validated QSAR models. The optimized structures were saved in sdf format, and transferred to PaDEL-Descriptor version 2.18 tool kits, topological descriptors encode the chemical properties have been calculated for each aminooxadiazole, using PaDEL server [19]. Only three suitable ones have been chosen as relevant descriptors for the studied inhibitory activity: Mannhold LogP (MLogP) and two Burden modified eigenvalues (SpMax1_Bhi and SpMin6_Bhm).

Methodology

After the calculation of descriptors, a Genetic Function Algorithm (GFA) analysis was performed to select the relevant molecular descriptors [20, 21]. The selected descriptors were then used to perform an MLR study until a valid model including: the critical probability P value < 0.05 for all descriptors and for the complete model, The Fisher static, the coefficient of determination, the mean squared error and the multi-collinearity test, internal and external validations, in addition to the Y-Randomization. Those selected descriptors were exploited to generate the applicability domain, then to evaluate a non-linear model. Later, the proposed model was used to identify aminooxadiazoles analogues in PubChem database and predict their PIM1 inhibitory activities.

Statistical analysis

In the present study, XLSTAT version 2013 [22] was used to perform multiple linear regression (MLR) and non-linear regression (MNLR), which are two statistical methods used to derive a mathematical relationship between

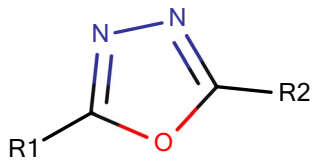


Fig. 1 The chemical structure of the studied compounds

a property of a given system and a set of descriptors that encode chemical information. A Genetic Algorithm tool was used to carried out the Genetic algorithm analysis (GFA) to reduce the number of the variables of the data set and choose the pertinent ones, in which, the mutation probability and smoothing parameter were set to 0.1 and 0.5, respectively. GFA in this study serves to select descriptors that were applied as input in multiple linear regression (MLR), multiple non-linear regression (MNLR) and applicability domain (AD).

Validation

The main objective of a QSAR study is to obtain a model with the highest predictive and generalization abilities. In order to evaluate the predictive ability of the developed QSAR models, two principals (internal validation and external validation) were performed. For the internal validation the leave-one-out cross-validation (Q^2) was used to evaluate the internal stability and of the present models. A high Q^2 value means a high internal predictive power of a QSAR model and a good robustness. Nevertheless, the study of Globarikh [23] indicated that there is no correlation between the value of Q^2 for the training set and predictive ability of the test set, revealing that the Q^2 is still insufficient for a reliable estimation of the model's predictive power for all new compounds. Thus, the external validation remains the only way to determine both the generalizability and the true predictive ability of QSAR models for new chemicals. For this reason, the statistical external validation was applied to the models as described by Globarikh and Tropsha. Roy and Roy [23–25] using a test set.

Y-Randomization test

The obtained models were further validated by the Y-Randomization method [21]. The dependent vector (pk_i) is randomly shuffled many times and after every iteration, a new QSAR model is developed. The new QSAR models are expected to have lower Q^2 and R^2 values than those the original models. This technique is carried out to eliminate the possibility of the chance correlation. If higher values of the Q^2 and R^2 are obtained, it means that an acceptable QSAR can't be generated for this data set because of the structural redundancy and chance correlation.

Results and discussion

Data set for analysis

A QSAR study was carried out on 34 aminooxadiazoles for the first time in order to establish a quantitative relationship between the PIM1 inhibitory activity and their chemical structures. The three selected descriptors by GFA method among 1543 other ones firstly calculated by PaDEL server are shown in Table 2.

Multiple linear regression (MLR)

Based on the selected descriptors a mathematical linear model was proposed to predict quantitatively the physicochemical effects of substituents on the PIM1 inhibitory activity of the 34 molecules using multiple linear regression. The linear model using this method includes three molecular descriptors: the total energy SpMin6_Bhm, the energy MLogP and the surface tension SpMax1_Bhi.

$$Y = a_0 + \sum_{i=1}^{n} a_i x_i$$

The following equation represents the best obtained linear QSAR model using the regression linear multiple (MLR) method:

$$pK_i = 43.24 + 8.396 \times (\boldsymbol{SpMin6_Bhm}) - 1.93 \times (\boldsymbol{MLogP}) - 9.65 \times (\boldsymbol{SpMax1_Bhi}) \quad (1)$$

$N = 27$, $R = 0.838$, $R^2 = 0.714$, $Q^2 = 0.60$, $MSE = 0.29$, $F = 19.12$, $P < 0.0001$.

The established models are judged by the statistical keys, such as, R^2 is the coefficient of determination, F is the Fisher statistic and MSE is the mean squared error. Higher coefficient of determination and lower mean squared error indicate that the model is more reliable. A P smaller than 0.05 means that the obtained equation is statistically significant at the 95% level. The leave one out cross-validated correlation coefficient LOO ($Q^2 = 0.60$) illustrates the reliability of the model by focusing on the sensitivity of the model towards the elimination of any single data point. A value of Q^2 greater than 0.5 is the basic criteria to qualify a model as valid [23].

The multi-collinearity between the three chosen descriptors was evaluated by calculating their variation inflation factors VIF as shown in Table 3. The VIF [26] was defined as $1/(1 - R^2)$, where R is the coefficient of correlation between one descriptor and all the other descriptors in the proposed model. A VIF value greater than 5.0 indicates that the model is unstable; a value between 1.0 and 4.0 indicates that the model is acceptable. Accordingly, it has been found that the descriptors used in the proposed model have very low-inter-correlation.

Negative values in the regression coefficients show that the indicated variables (MLogP and SpMax1_Bhi)

Table 1 Observed activities of studied aminooxadiazoles

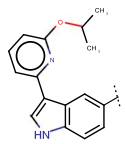
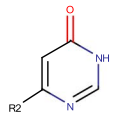
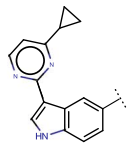
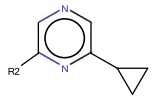
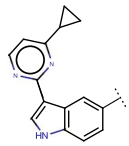
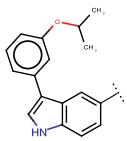
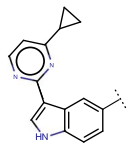
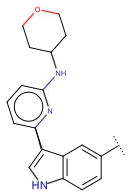
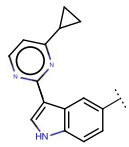
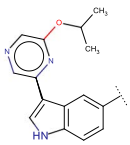
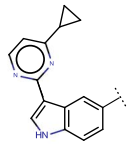
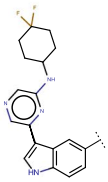
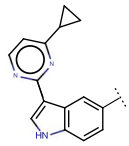
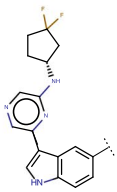
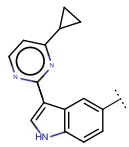
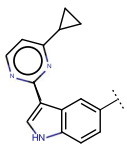
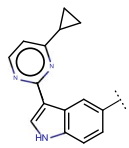
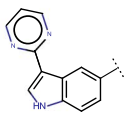
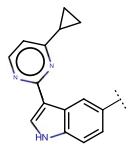
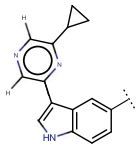
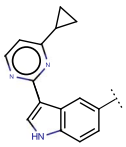
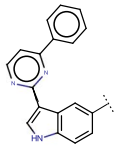
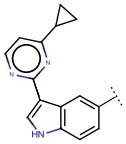
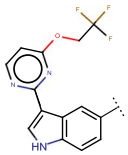
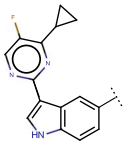
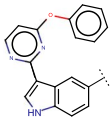
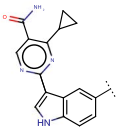
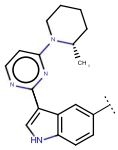
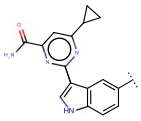
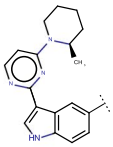
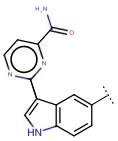
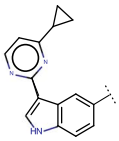
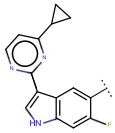
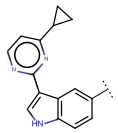
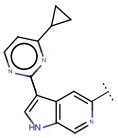
N	R_1	R_2	pk_i	N	R_1	R_2	pk_i
1	NH_2		8.769	**18**			6.943
2	NH_2		5.591	**19**			7.494
3*	NH_2		7.259	**20***	c-C_3H_5		8.080
4	NH_2		8.677	**21***	i-Bu		6.600
5	NH_2		8.522	**22**	Ph		6.939
6	NH_2		8.795	**23***	$NH\text{-}CH_2CF_3$		9.619
7	NH_2		9.284	**24**	NH-i-Pr		9.075
8	NH_2		8.853	**25**	NH-c-C_3H_5		8.920
9	NH_2		6.823	**26**	NH-oxetane		8.657

Table 1 continued

10	NH_2		8.699	**27**	NH-*t*-Bu		9.259
11	NH_2		7.508	**28***	NH-3-F-Ph		7.161
12	NH_2		8.677	**29***	NH-*i*-Pr		8.886
13	NH_2		7.832	**30**	NH-*i*-Pr		9.346
14	NH_2		8.568	**31**	NH-*i*-Pr		8.795
15	NH_2		8.958	**32**	NH-*i*-Pr		7.267
16			7.284	**33**	NH-*i*-Pr		8.920
17			8.522	**34***	NH-*i*-Pr		8.443

*Test set

contribute negatively to the value of pk_i, whereas positive value in the regression coefficient of variable (SpMin6_Bhm) indicates that the greater the value of the variable, the greater the value of the pk_i.

The predicted values computed using this MLR model with the experimental values for the training and test sets are shown in Table 4, and plotted in Fig. 2. The selected descriptors (Eq. 1) in the MLR model are then used as the input variables to perform the multiple nonlinear regression (MNLR).

Multiples non-linear regression (MNLR)

The nonlinear regression model was also used to evaluate the effect of the substituents in the studied aminooxadiazoles on the PIM1 inhibitory activity, improve the structure–activity relationship in quantitative manner.

Training set used in MLR and descriptors selected by GFA were used in this method to build the non-linear model. The best regression performance was selected according to the coefficient of determination R^2 and the mean squared error MSE, a pre-programmed function in

Table 2 The values of three relevant molecular descriptors used in the best QSAR model

No	pk_i	SpMin6_Bhm	MLogP	SpMax1_Bhi	No	pk_i	SpMin6_Bhm	MLogP	SpMax1_Bhi
1	8.769	1.276	2.67	4.187	**18**	8.522	1.352	2.890	4.225
2	5.591	0.707	2.01	4.135	**19**	6.943	1.228	2.889	4.172
3*	7.259	1.276	2.889	4.188	**20***	7.494	1.372	3.330	4.190
4	8.677	1.361	2.78	4.188	**21***	8.080	1.372	3	4.191
5	8.522	1.266	2.45	4.187	**22**	6.600	1.364	3.11	4.190
6	8.795	1.361	2.56	4.188	**23***	6.939	1.350	3.329	4.194
7	9.284	1.349	2.449	4.188	**24**	9.619	1.334	2.449	4.190
8	8.853	1.286	2.56	4.189	**25**	9.075	1.340	2.89	4.190
9	6.823	1.015	2.23	4.186	**26**	8.920	1.372	2.89	4.190
10	8.699	1.295	2.56	4.190	**27**	8.657	1.341	2.78	4.190
11	7.508	1.015	2.01	4.188	**28**	9.259	1.341	3	4.190
12	8.677	1.278	2.89	4.194	**29***	7.161	1.350	3.11	4.190
13	7.832	1.276	2.78	4.187	**30***	8.886	1.338	2.78	4.191
14	8.568	1.361	2.78	4.188	**31**	9.346	1.349	2.78	4.192
15	8.958	1.361	2.78	4.188	**32**	8.795	1.341	2.78	4.192
16	7.284	1.360	3.11	4.226	**33**	7.267	1.249	2.45	4.188
17	8.522	1.276	2.67	4.187	**34***	8.920	1.335	2.78	4.195

* Test set

the XLSTAT was used to evaluate the nonlinear regression model as follows:

$$y == a + (bX_1 + cX_2 + dX_3 + eX_4 \ldots) + \left(fX_1^2 + gX_2^2 + hX_3^2 + iX_4^2 \ldots\right).$$

where X_1, X_2, X_3, X_4...: represent the variables, and *a, b, c, d*...: represent the parameters.

The resulting equation is as follows:

$$pK_i = -19641.39 - 47.63 \times (\boldsymbol{SpMin6_Bhm}) + 15.79 \times (\boldsymbol{MLogP}) + 9356.96 \times (\boldsymbol{SpMax1_Bhi}) + 21.6 \times (\boldsymbol{SpMin6_Bhm})^2 - 3.10 \times (\boldsymbol{MLogP})^2 - 1113.59 \times (\boldsymbol{SpMax1_Bhi})^2 \quad (2)$$

N = 27, R = 0.910, R^2 = 0.812, Q^2 = 0.56, MSE = 0.22.

The leave one out cross-validated correlation coefficient LOO (Q^2 = 0.56) illustrates the reliability of the model by focusing on the sensitivity of the model towards the elimination of any single data point. A value of Q^2 greater than 0.5 is the basic criteria to qualify a model as valid [23]. It can be seen clearly from the key statistical indicators, coefficient of determination R^2, mean squared error MSE and, value of Q^2, that the predicting ability of this model is better than that of the linear model (MLR). The enhancement in the predictive ability was due to the involvement of the squared terms in the nonlinear model.

The predicted values computed using this MNLR model for the training and test sets are shown in Table 4, and plotted in Fig. 3.

Applicability domain

The utility of a QSAR model is its accurate prediction ability for new chemical, so, once the QSAR model is built, its domain of applicability (AD) must be defined. A model is considered valid only if it is able to make predictions within its training domain and only the prediction for new compounds falling within its applicability domain can be regarded credible and not model extrapolations. The most common method to define the AD, it is based on the determination of the leverage value of each compound [25]. The Williams plot [The plot of standardized residuals versus leverage values (*h*)] is used in the present study to visualize the AD of the QSAR model.

$$h_i = x_i^T \left(X^T X\right)^{-1} x_i$$

where the x_i is the descriptor vector of the considered compound, X is the descriptor matrix derived from the training set descriptor values, the threshold is defined as:

$$h^* = \frac{3(k+1)}{n}$$

where n is the number of compound in the training set, k is the number of the descriptors in the proposed model, a leverage (*h*) greater than the threshold (*h**) indicates that the predicted response is an extrapolation of the model and, consequently, it can be unreliable.

The Williams plot of the presented MLR model is shown in the Fig. 4, the applicability domain is established inside

Table 3 Multi-colinearity test

Variables	SpMin6_Bhm	MLogP	SpMax1_Bhi
VIF	3.035	2.201	1.869

Table 4 Observed values and calculated values of pk_i according to different methods

No	pk_i (obs)	pk_i (pred)	
		MLR	MNLR
1	8.769	8.345	8.328
2	5.591	5.358	5.535
3*	7.259	7.916	8.045
4	8.677	8.838	9.037
5	8.522	8.695	8.261
6	8.795	9.272	9.197
7	9.284	9.383	9.036
8	8.853	8.629	8.514
9	6.823	7.021	7.510
10	8.699	8.694	8.595
11	7.508	7.429	6.996
12	8.677	7.880	8.171
13	7.832	8.143	8.186
14	8.568	8.838	9.024
15	8.958	8.838	9.024
16	7.284	7.824	7.648
17	8.522	8.192	8.241
18	6.943	7.664	7.010
19	7.494	7.852	7.486
20*	8.080	8.481	8.772
21*	6.600	8.211	8.299
22	6.939	7.623	7.309
23*	9.619	9.235	8.919
24	9.075	8.439	8.655
25	8.920	8.706	9.016
26	8.657	8.653	8.856
27	9.259	8.228	8.390
28*	7.161	8.092	8.144
29*	8.886	8.621	8.856
30	9.346	8.708	8.984
31	8.795	8.641	8.909
32	7.267	8.544	8.174
33	8.920	8.556	8.902
34*	8.522	8.785	8.804

* Test set

a squared area within ±2 standard deviation and a leverage threshold h^* of 0.44. As shown in the Williams plot the majority of the compounds in the data set are in this area, except one (Compound **2**) in training set exceeds the threshold and it is considered as an outlier compound. This erroneous prediction could probably be attributed to the R_2 position, whereas, the majority of compounds are substituted by an indole linked to another moiety at this position this compound has just an indole moiety at the R_2 position. Also, compound 22 in the test set is wrongly predicted (>3 s), but with lower leverage values ($h < h^*$) and that could probably be attributed to a different mechanism of action rather than to molecular structures [25].

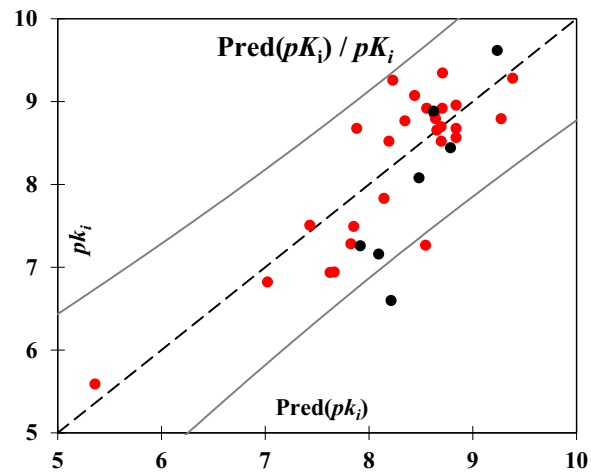


Fig. 2 Graphical representation of predicted and observed activity (pk_i) values calculated by MLR

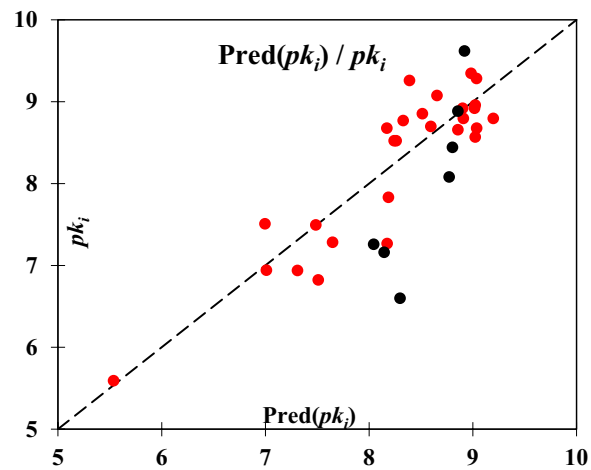


Fig. 3 Graphical representation of predicted and observed activity (pk_i) values calculated by MNLR

Y-Randomization

The Y-Randomization method was carried out to validate the MLR and MNLR models. Several random shuffles

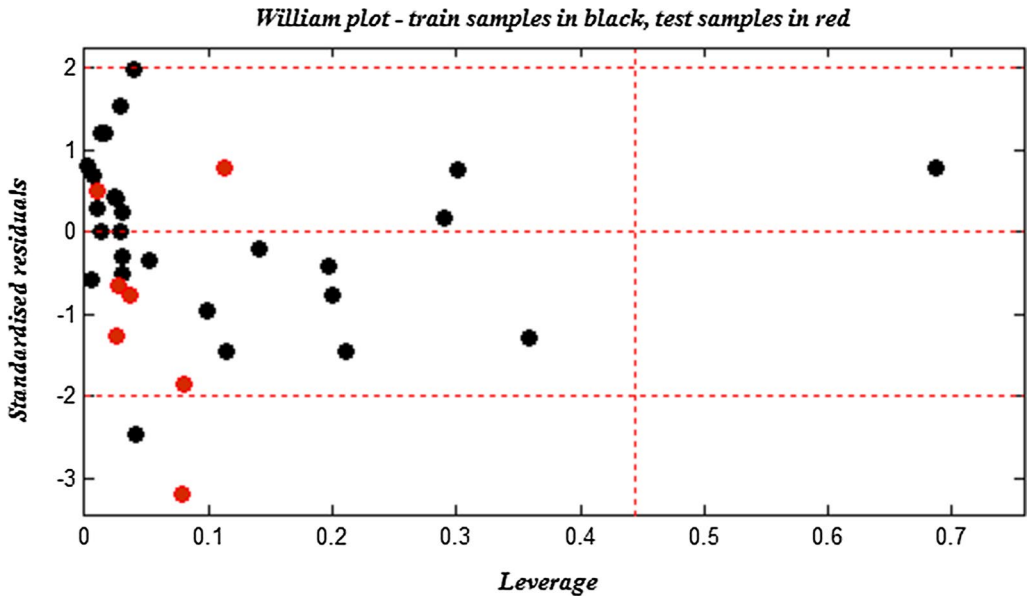


Fig. 4 Williams plot for the training set and external validation for the PIM1 inhibitory activity of aminooxadiazole compounds, listed in Table 1 ($h^*=0.44$ and residual limits ± 2)

Table 5 Q^2 and R^2 values after several Y-Randomization tests

Iteration	MLR		MNLR	
	Q^2	R^2	Q^2	R^2
1	0.390	0.235	0.031	0.350
2	0.120	0.094	0.095	0.008
3	0.290	0.124	0.079	0.190
4	0.340	0.129	−0.264	0.290
5	0.180	0.263	−0.160	0.335
6	0.160	0.194	−0.522	0.140
7	0.20	0.075	−0.105	0.006
8	0.130	0.043	0.230	0.026
9	0.140	0.116	0.120	0.196
10	0.230	0.031	0.060	0.131

Table 6 The statistical results of MLR and MNLR models with validation techniques

Method/parameter	R	R^2	Q^2	R^2_{test}	MSE
MLR	0.838	0.712	0.60	0.81	0.29
MNLR	0.910	0.812	0.56	0.75	0.22

of the dependent variable (pk_i) were performed then after every shuffle, a QSAR was developed and obtained results are shown in Table 5. The low Q^2 and R^2 values obtained after every shuffle indicate that the good result in our original MLR and MLR models are not due to a chance correlation of the training set.

External validation

To test the prediction ability of the obtained models, it is required the use of a test set for external validation. Thus, the models generated on the training set using 26 aminooxadiazoles were used to predict the PIM1 inhibitory activity of the remaining molecules. The parameters of the performance of the generated models are shown in Table 6. It can be seen clearly that the MNLR is statically better than the MLR model in terms of coefficient of determination, but the MLR has a better predictive ability and good internal stability.

Among the obtained models for this series, the MLR model has the highest prediction ability for the test set ($R^2_{test}=0.81$), also the highest cross-validation coefficient ($Q^2=0.60$), all that support the applicability of the proposed MLR prediction model. However, both the results obtained by the MLR and MNLR should be regarded as satisfactory for predicting the PIM1 activity using the proposed descriptors.

Virtual screening for aminooxadiazole analogues and prediction of their PIM1 inhibitory activities

Overall, this study can be used to screen chemical databases to identify new PIM1 inhibitors as well as to predict their inhibitory activities. Therefore, the built MLR model was used to screen the PubChem database, by searching compounds had 95% similarity with the most

Table 7 Predicted values and calculated *h* of pk_i (k_i in nM) of the sixteen identified hits

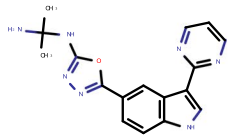
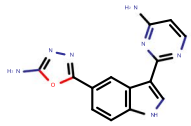
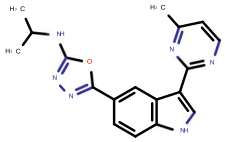
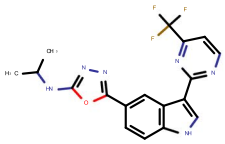
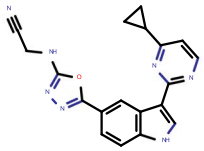
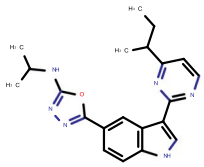
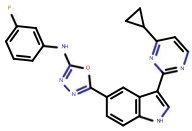
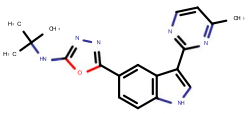
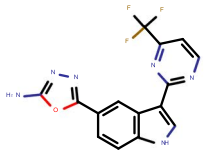
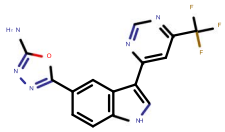
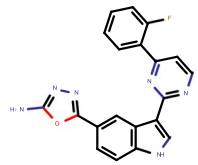
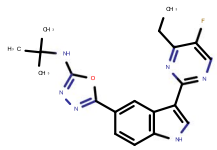
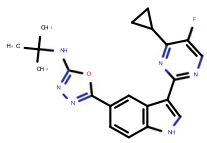
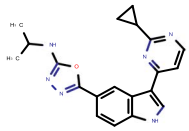
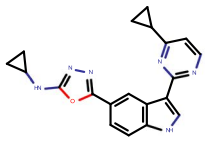
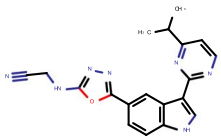
N	Molecular structure	Pubchem CID	Pred(pk_i) for PIM1	h
1		123719668	9.029	0.0892
2		123533604	8.048	0.1831
3		123527918	8.746	0.0408
4		90256832	9.210	0.1813
5		88938076	9.283	0.0890
6		68356835	9.291	0.1007
7		68356805	8.561	0.2014
8		68356785	8.533	0.0739
9		68356773	7.888	0.2468
10		68356701	7.887	0.2459

Table 7 continued

11		68328341	8.547	0.0631
12		68328223	8.806	0.0477
13		68328001	8.871	0.0649
14		66554899	8.970	0.0532
15		66554898	8.698	0.1189
16		66554728	9.100	0.0760

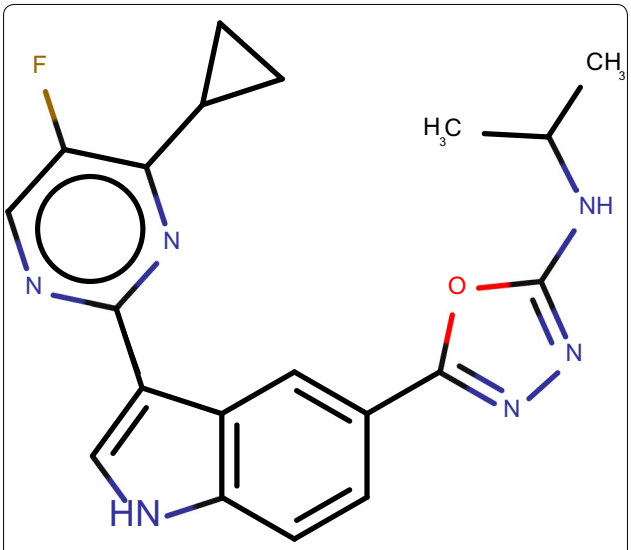


Fig. 5 Reference structure of aminooxadiazole model with lowest binding constant k_i

active compound of the studied series (Compound **29**) and fulfilling the Lipinski's rule of bioavailability [27]. Sixteen compounds were identified as shown in Table 7 and their pk_i values were predicted in addition to their leverages (h) to check if they fall in the AD of the proposed model. (Table 7, Figs. 5 and 6).

It can be seen from the Fig. 6 that all identified compounds have $h < h^*$, ($h^* = 0.44$) so their predicted values are regarded reliable.

Conclusion

To predict the PIM1 inhibitory activity of a series substituted aminooxadiazoles, two unambiguous models were developed in this study with topological descriptors. A good stability and prediction ability were exhibited by MLR and MNLR models, on the same set of descriptor. Furthermore, the obtained results from each model on this series of compounds are quite similar, no one of the established models is considered better than the other.

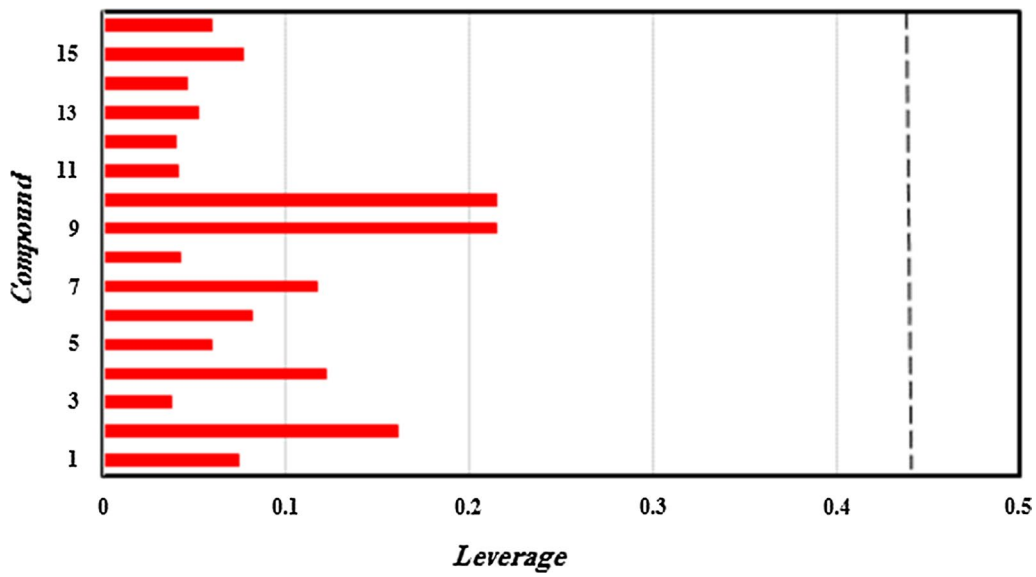


Fig. 6 Leverage values of the screened compounds from the PubChem database for the PIM1 inhibitory activity, listed in Table 7 ($h^* = 0.44$)

So, the MLR and MNLR models are regarded as effective tools to predict PIM1 inhibitory activity of substituted aminooxadiazoles based on the proposed descriptors. The predictive ability of the transparent model MLR was excellent enough to be used to virtually screen novel PIM1 inhibitors from PubChem database.

Finally, we combined a machine learning approach using unambiguous MLR-QSAR model with PubChem database filtering concept to provide a rustic ligand-based virtual screening protocol. As a result, 16 potentially aminooxadiazole analogues as PIM1 inhibitors were identified. This study provides the theoretical basis and specific chemicals for PIM1 inhibitors, which can help the experimental research groups to search for potential anticancer.

Abbreviations

QSAR: quantitative structure activity relationship; PIM: proviral integration site for Moloney murine leukaemia virus kinases; MLR: multiple linear regression; MNLR: multiple nonlinear regression; AD: applicability domain; GFA: Genetic Function Algorithm; Q^2: cross-validated determination coefficient; R^2: non-cross-validated correlation coefficient; MSE: standard error of the estimate; F: F test value; R^2_{test}: external validation determination coefficient.

Authors' contributions

AA proposed the work, AA carried out the QSAR studies, arranged the results and drafted the manuscript under the guidance of MC, AS, MB, and TL. AA and AG, MG, AO and SC did manuscript revision and final shape. All authors read and approved the final manuscript.

Author details

[1] MCNSL, School of Sciences, Moulay Ismail University, Meknes, Morocco.
[2] High School of Technology, Moulay Ismail University, Meknes, Morocco.

Acknowledgements

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

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 2 September 2017 Accepted: 9 March 2018
Published online: 22 March 2018

References

1. Nawijn MC, Alendar A, Berns A (2011) For better or for worse: the role of Pim oncogenes in tumorigenesis. Nat Rev Cancer 11:23–34. https://doi.org/10.1038/nrc2986
2. Santio NM, Vahakoski RL, Rainio E-M et al (2010) Pim-selective inhibitor DHPCC-9 reveals Pim kinases as potent stimulators of cancer cell migration and invasion. Mol Cancer 9:279. https://doi.org/10.1186/1476-4598-9-279
3. Brault L, Gasser C, Bracher F et al (2010) PIM serine/threonine kinases in the pathogenesis and therapy of hematologic malignancies and solid cancers. Haematologica 95:1004–1015. https://doi.org/10.3324/haematol.2009.017079
4. Wu B, Wang HL, Cee VJ et al (2015) Discovery of 5-(1H-indol-5-yl)-1,3,4-thiadiazol-2-amines as potent PIM inhibitors. Bioorg Med Chem Lett 25:775–780. https://doi.org/10.1016/j.bmcl.2014.12.091
5. Gadewal N, Varma A (2012) Targeting Pim-1 kinase for potential drug-development. Int J Comput Biol Drug Des 5:137–151. https://doi.org/10.1504/IJCBDD.2012.048303
6. Bataille CJR, Brennan MB, Byrne S et al (2017) Thiazolidine derivatives as potent and selective inhibitors of the PIM kinase family. Bioorg Med Chem 25:2657–2665. https://doi.org/10.1016/j.bmc.2017.02.056

7. Chen LS, Redkar S, Taverna P et al (2011) Mechanisms of cytotoxicity to Pim kinase inhibitor, SGI-1776, in acute myeloid leukemia. Blood 118:693–702. https://doi.org/10.1182/blood-2010-12-323022
8. Keeton EK, McEachern K, Dillman KS et al (2014) AZD1208, a potent and selective pan-Pim kinase inhibitor, demonstrates efficacy in preclinical models of acute myeloid leukemia. Blood 123:905–913. https://doi.org/10.1182/blood-2013-04-495366
9. Amin SA, Adhikari N, Shukla V et al (2017) Structural findings of pyrazolo[1,5-a]pyrimidine compounds for their Pim-1/2 kinase inhibition as potential anticancer agents. Indian J Biochem Biophys 54:32–46
10. Kumar PT, Piyush G, Kumar JS (2017) 3D QSAR studies on some 5-(1H-Indol-5-yl)-1, 3, 4-thiadiazol-2 amines as potential PIM-1 inhibitors. J Young Pharm 9:162–167. https://doi.org/10.5530/jyp.2017.9.32
11. Wurz RP, Pettus LH, Jackson C et al (2015) The discovery and optimization of aminooxadiazoles as potent Pim kinase inhibitors. Bioorg Med Chem Lett 25:847–855. https://doi.org/10.1016/j.bmcl.2014.12.067
12. Dearden JC (2016) The history and development of quantitative structure-activity relationships (QSARs). Int J Quant Struct Relationsh 1:1–44. https://doi.org/10.4018/IJQSPR.2016010101
13. Kar RK, Ansari MY, Suryadevara P et al (2013) Computational elucidation of structural basis for ligand binding with *Leishmania donovani* adenosine kinase. Biomed Res Int 2013:1–14. https://doi.org/10.1155/2013/609289
14. Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? where are you going to? J Med Chem 57:4977–5010. https://doi.org/10.1021/jm4004285
15. Aouidate A, Ghaleb A, Ghamali M, et al. (2018) Furanone derivatives as new inhibitors of CDC7 kinase: development of structure activity relationship model using 3D QSAR, molecular docking, and in silico ADMET. Struct Chem. https://doi.org/10.1007/s11224-018-1086-4
16. Gupta SP, Mathur AN, Nagappa AN et al (2003) A quantitative structure-activity relationship study on a novel class of calcium-entry blockers: 1-[(4-(aminoalkoxy)phenyl)sulphonyl]indolizines. Eur J Med Chem 38:867–873
17. Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose tripos 5.2 force field. J Comput Chem 10:982–1012. https://doi.org/10.1002/jcc.540100804
18. Purcell WP, Singer JA (1967) A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. J Chem Eng Data 12:235–246. https://doi.org/10.1021/je60033a020
19. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474. https://doi.org/10.1002/jcc.21707
20. Klekota J, Roth FP (2008) Chemical substructures that enrich for biological activity. Bioinformatics 24:2518–2525. https://doi.org/10.1093/bioinformatics/btn479
21. Veerasamy R, Rajak H, Jain A et al (2011) Validation of QSAR models-strategies and importance. Int J Drug Des Discov 2:511–519. https://doi.org/10.1016/j.febslet.2005.06.031
22. XLSTAT 2013 (2013) Data analysis and statistical solution for Microsoft Excel. Addinsoft, Paris
23. Golbraikh A, Tropsha A (2002) Beware of q^2! J Mol Graph Model 20:269–276. https://doi.org/10.1016/S1093-3263(01)00123-1
24. Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 27:302–313. https://doi.org/10.1002/qsar.200710043
25. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26:694–701. https://doi.org/10.1002/qsar.200610151
26. O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690. https://doi.org/10.1007/s11135-006-9018-6
27. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1:337–341. https://doi.org/10.1016/j.ddtec.2004.11.007

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com