The 8th International Electronic Conference on Medicinal Chemistry (ECMC 2022)
01–30 NOVEMBER 2022 | ONLINE

QSAR and Drug-likeness Studies of Thiadiazole Derivatives Against Lung Cancer

Chaired by DR. ALFREDO BERZAL-HERRANZ;
Co-Chaired by PROF. DR. MARIA EMÍLIA SOUSA

pharmaceuticals MDPI

Mouad Mouhsin[1*] , Mustapha Oubenali[1] , Samir Chtita[2], Mohamed Mbarki[1] and Ahmed Gamouh[1]

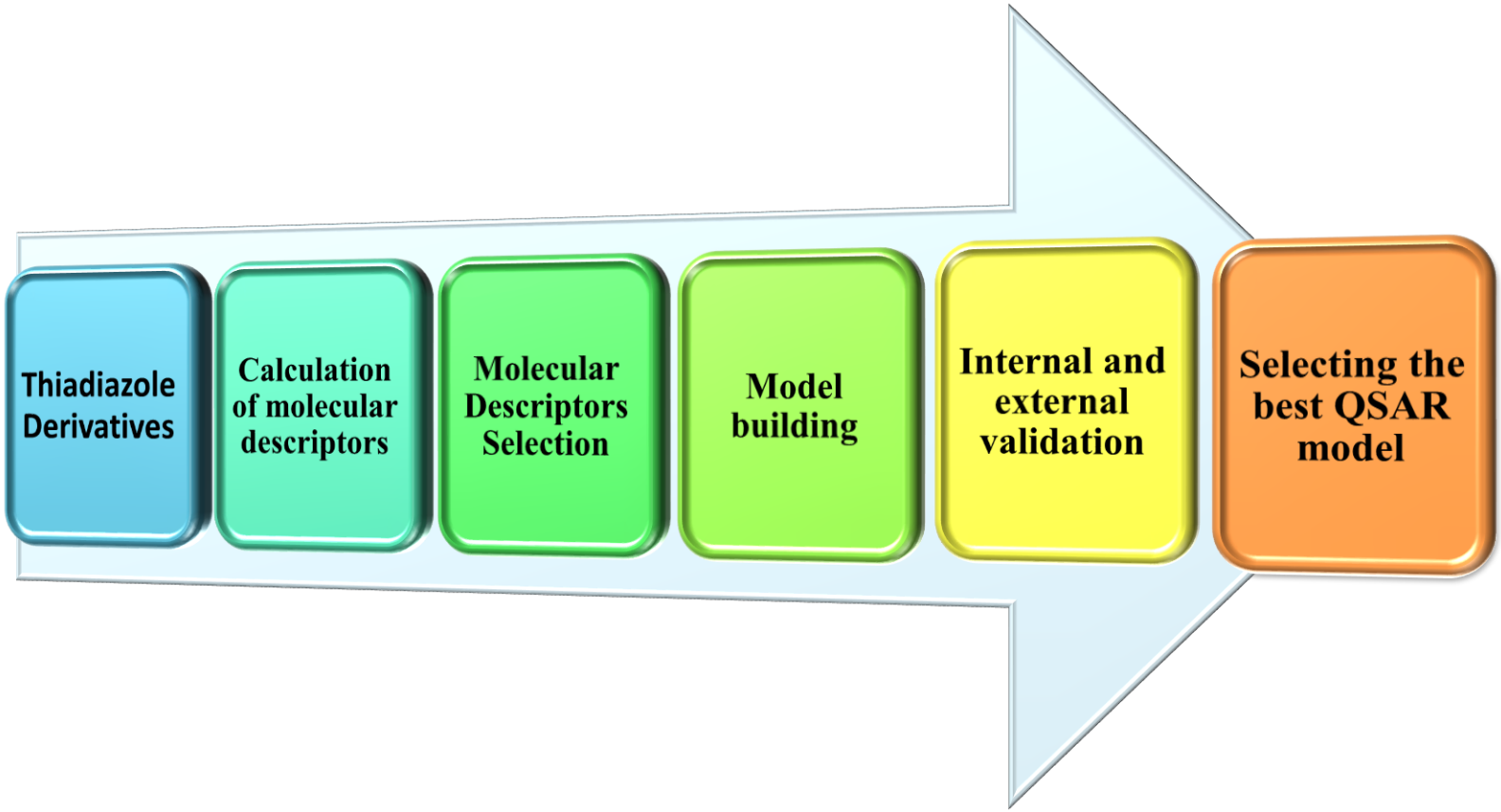
[1] Team of ACCNE - Laboratory of Engineering in Chemistry and Physics of Matter, Faculty of Science and Technologies, Beni Mellal, Morocco.

[2] Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, Sidi Othman, Box 7955, Casablanca, Morocco

* Corresponding author:mouhsinmouad@gmail.com

Faculté des Sciences et Techniques
Béni Mellal

QSAR and Drug-likeness Studies of Thiadiazole Derivatives Against Lung Cancer



ECMC 2022
The 8th International Electronic Conference on Medicinal Chemistry
01–30 NOVEMBER 2022 | ONLINE

Abstract:
This study was aimed at building a robust quantitative structure–activity relationship (QSAR) to predict the anti-proliferate activity of 1,3,4-thiadiazole derivatives against the A549 lung cancer cell lines. The semi-empirical PM7 parametrization approach was used to optimize the complete set of 1,3,4-thiadiazole derivatives and various classes of molecular descriptors have been calculated. We built models using Fisher score and the best subset selection for feature selection, and the final model was developed using the multiple linear regression technique, all in accordance with the rigorous Organization for Economic Co-operation and Development (OECD) requirements. Furthermore, various internationally agreed severe validation parameters were used to validate the model. Overall, our established model for quick prediction should be relevant to new, untested, or not yet produced compounds that fall within the applicability domain (AD) of the model.

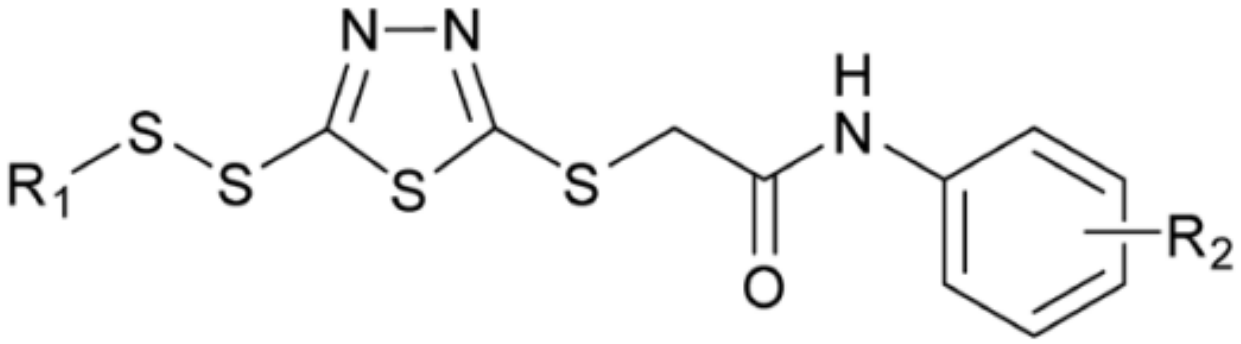
Keywords: QSAR; Thiadiazole derivatives; A549; PM7; OECD

Introduction

- Despite significant progress in understanding cancer pathophysiology, it continues to be one of the leading causes of death.
- Novel more effective anticancer treatments are urgently needed.
- Computer-aided drug design techniques have a significant impact in shortening the drug discovery development process
- ***Objective***: Construct a robust quantitative structure–activity relationship (QSAR) to predict the anti-proliferate activity of 1,3,4-thiadiazole derivatives against the lung cancer cell lines.

Dataset

The data set of 1,3,4-thiadiazole derivatives with anti-proliferative activity in A549 cancer cell lines was assembled from synthesis and experimental determination studies . The data set contains a total of 33 molecules . The activities of these compounds expressed by IC_{50} values used as the dependent variables for QSAR modelling analyses.



Structure optimization and descriptor generation

- The PaDEL , Mordred , and ChemoPy programs were used to calculate the descriptors based on the 3D geometry representation of molecules.
- As a result, we have over 4000 descriptors, which include 1D, 2D, 3D, and quantum molecular descriptors.
- Intercorrelated descriptors and descriptors with constant and near-constant value were removed during the initial data pre-treatment

as a learning set

Data splitting

33 molecules

22 molecules (67%) as a training set

11 molecules (33%) as a Test set

Descriptor selection

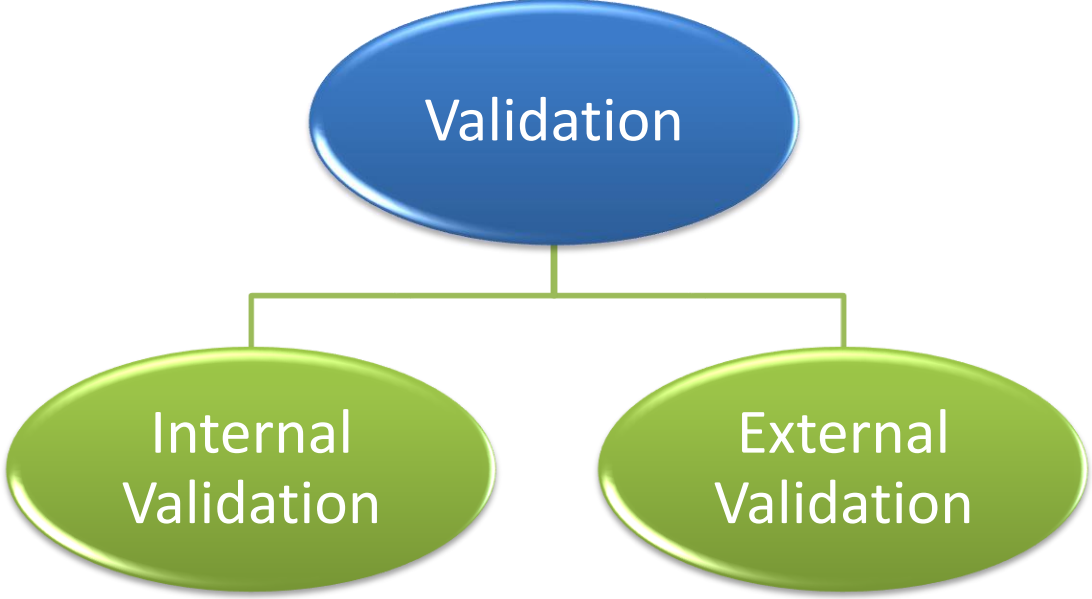
- F-regression from sklearn python package has been used as a feature selection method to search for those descriptors with significant explained variance.

1) First step: the correlation coefficient between each descriptor and biological activity is calculated.
2) Second step: the value of F score is computed and converted to p value
3) Third step: 'selectKBest' (from sklearn) has been used which allows the selection of several variables based on p-value

Model developement

- ✓ Using these 100 descriptors, we ran the best subset selection to generate models.

- ✓ Best subset regression is a linear regression analysis technique for developing exploratory models. This method examines all potential models with a given set of variables and provides the best-fitting models that have one, two, or more variables.



- R^2
- Q^2_{LOO}
- $r^2_{m(LOO)}$

- R^2_{pred}
- Q^2_{F2}
- $r^2_{m(test)}$

- Y-Randomization test

Results and discussion

- The molecular descriptors obtained for the Thiadiazole derivatives were pre-treated, as previously stated.
- Finally, using the ‘best subset selection' technique, 100 descriptors were examined.
- Using a combination of 4, 5, and 6 descriptors, many models were created using the MLR approach.
- Based on the two parameters Q^2_{LOO} and Q^2_{f1}, the best model for each combination is selected.

Models

Model	Equation
1	IC_{50}= -3.313 + 1.029TDB8e - 11.081RDF130v + 7.071Mor06 - 1.758RDFU5
2	IC_{50}= 9.625 + 1.285TDB8e - 2.715RDF130s + 8.271Mor06 - 21.561GATSm4 - 2.006RDFU5
3	IC_{50}= 79.852 - 0.042TDB9v + 0.061PNSA-1 + 11.889Mor06 - 3.533Mor04m - 1.603RDFU5 + 17.994MoRSEU15

Internal validation results for the three models

Parameters	Model 1	Model 2	Model 3
R^2	0.802	0.803	0.869
R^2_{adj}	0.756	0.741	0.816
Q^2_{LOO}	0.676	0.685	0.751
Standard Error of Estimate	3.927	4.043	3.406
F-value (95%)	17.258	13.035	16.560
$\overline{r^2_{m(LOO)}}$	0.594	0.604	0.681
$\Delta r^2_{m(LOO)}$	0.065	0.051	0.025

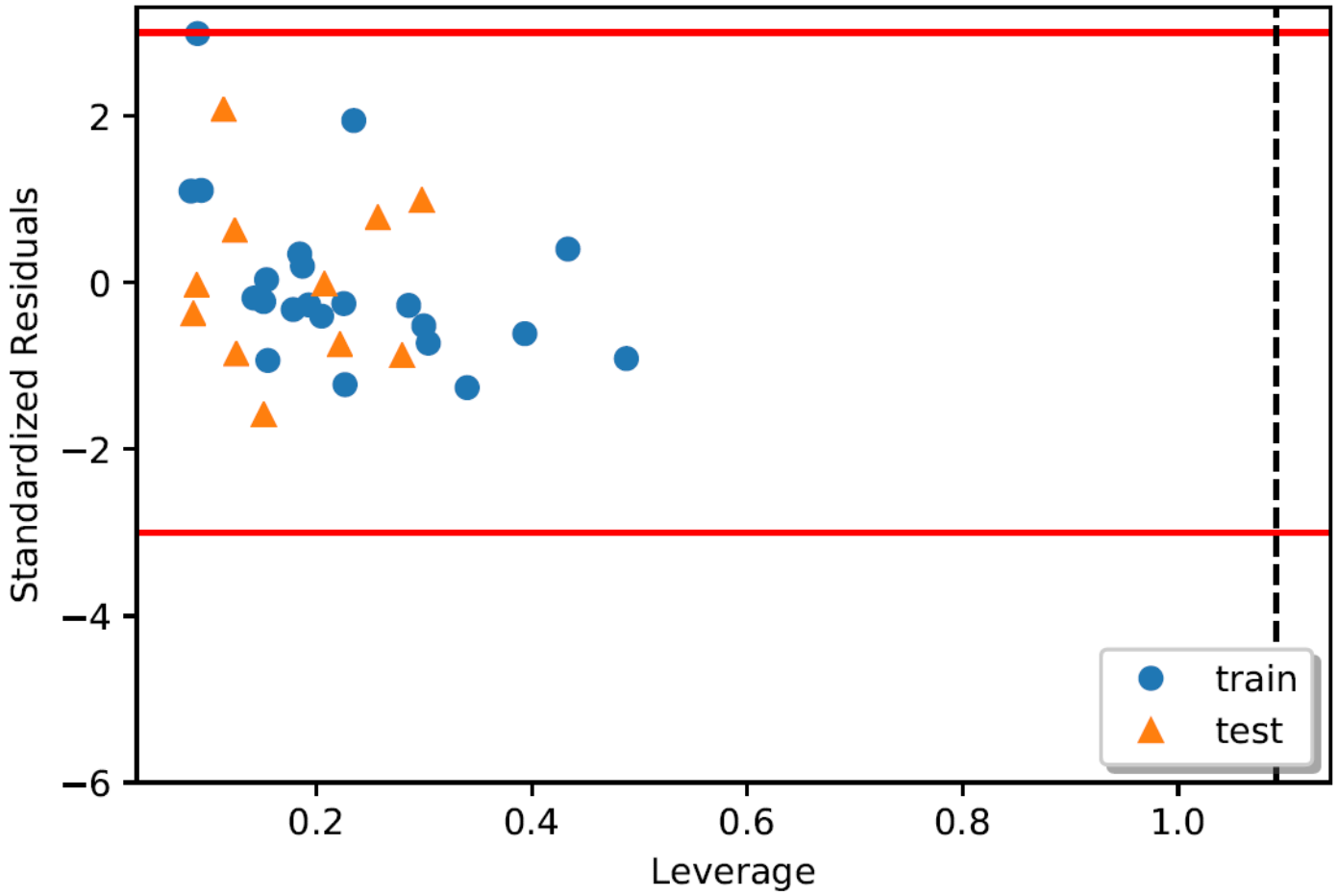
External validation results for the three models

Parameters	Model 1	Model 2	Model 3
R^2	0.602	0.675	0.819
Q_{f1}^2	0.601	0.668	0.812
Q_{f2}^2	0.601	0.668	0.812
RMSEP	4.941	4.506	3.388

The correlation matrix

	TDB9v	PNSA-1	Mor06	Mor04m	RDFU5	MoRSEU15	ACTIVITY
TDB9v	1	0.385	0.538	-0.571	-0.501	0.501	0.560
PNSA-1		1	0.217	-0.046	-0.368	0.192	0.441
Mor06			1	0.007	-0.132	0.340	0.579
Mor04m				1	0.368	-0.252	-0.465
RDFU5					1	-0.039	-0.461
MoRSEU15						1	0.593

the applicability domain



Conclusions

- In this study, we have developed an MLR-regression-based QSAR model from 33 compounds having defined anti-proliferate activity against the lung cancer cell lines

- The most relevant descriptors were chosen using a variety of variable selection strategies, and the final model was built using the multiple linear regression technique. The model was thoroughly validated using both internal and external validation metrics, with the results demonstrating the generated model's reliability and utility