FABAD J. Pharm. Sci., 48, 3, 411-430, 2023
Doi: 10.55262/fabadeczacilik.1309814
RESEARCH ARTICLE

Interpretable QSAR Modelling for QSAR-Based Virtual Screening of 3H-Thiazolo[4,5-b] pyridin-2-one Derivatives as Potential Antioxidant Drug Candidates

Olena KLENINA[*°]

Interpretable QSAR Modelling for QSAR-Based Virtual Screening of 3H-Thiazolo[4,5-b] pyridin-2-one Derivatives as Potential Antioxidant Drug Candidates

SUMMARY

A quantitative structure-activity relationship (QSAR) study has been carried out for 32 N3 substituted 3H-thiazolo[4,5-b] pyridin-2-one derivatives as potential antioxidant drug candidates. The genetic algorithm (GA) and multiple linear regression analysis (MLRA) were used as appropriate techniques for descriptor selection and correlation model generation. The four best regressions for predicting the ability to scavenge the DPPH radical were generated as three-parameter QSAR models with the highest statistical characteristics and predictive power. It was shown that a set of 2D, 3D, and Molecular properties descriptors play a crucial role in antioxidant activity enhancement. Small hydrophilic molecules with the minimal distance of specific atoms and fragments from the center of mass, neglectable electronic density redistribution between the distant atoms, and molecules keeping strong symmetry of electronegative atoms along the 1st principal component axe exhibit higher activity. Validation parameters of the generated models allow us to state that they satisfy the statistical requirements for their goodness-of-fitting with no current overfitting. The predictive ability of the constructed models was assessed with both internal and external validation approaches and estimated with the leave-one-out and leave-group-out cross-validation coefficients (Q^2_{LOO} and Q^2_{LGO}). The values of Q^2_{LOO} (0.7060 ÷ 0.7480) and Q^2_{LGO} (0.6647 ÷ 0.7711) are reasonable, showing that the models are significant and robust to predict the free radical scavenging activity of the compounds from both training and validation sets. Applicability domain-defining technique was employed in the obtained models, and indicated that most structures were adequately represented by the chemical space of the models.

Key Words: *3H-thiazolo[4,5-b]pyridin-2-one, QSAR, multiple linear regression (MLR), antioxidant activity, applicability domain*

Potansiyel Antioksidan İlaç Adayları Olarak 3H-Tiazolo[4,5-b] piridin-2-on Türevlerinin QSAR Tabanlı Sanal Taraması için Yorumlanabilir QSAR Modellemesi

ÖZ

Potansiyel antioksidan ilaç adayları olarak 32 N3 sübstitüe 3H-tiazolo[4,5-b]piridin-2-on türevleri için kantitatif yapı-aktivite ilişkisi (QSAR) çalışması yapılmıştır. Tanımlayıcı seçimi ve korelasyon modelleri oluşturmak için uygun teknikler olarak genetik algoritma (GA) ve çoklu doğrusal regresyon analizi (MLRA) kullanıldı. DPPH radikalini temizleme yeteneğinin tahmini için en iyi dört regresyon, en yüksek istatistiksel özelliklere ve öngörü yeteneğine sahip üç parametreli QSAR modelleri olarak üretildi. Bir dizi 2D, 3D ve Moleküler özellik tanımlayıcılarının, antioksidan aktiviteyi artırmada çok önemli bir rol oynadığı gösterilmiştir. Belirli atomların ve fragmanların kütle merkezinden minimum uzaklığa sahip olduğu küçük hidrofilik moleküller, uzak atomlar arasındaki ihmal edilebilir elektron yoğunluğu dağılımı ve 1. ana bileşen ekseni boyunca elektronegatif atomların güçlü simetrisini koruyan moleküller daha yüksek aktivite sergiler. Oluşturulan modellerin doğrulama parametreleri, bunların mevcut aşırı uyum olmadan uyumun iyiliği için istatistiksel gereksinimleri karşıladığını belirtmemize olanak tanır. Oluşturulan modellerin tahmin yeteneği, hem iç hem de dış doğrulama yaklaşımıyla değerlendirildi ve birini dışarıda bırak ve grubu dışarıda bırak çapraz doğrulama katsayıları (Q^2_{LOO} ve Q^2_{LGO}) ile tahmin edildi. Q^2_{LOO} (0.7060 ÷ 0.7480) ve Q^2_{LGO} (0.6647 ÷ 0.7711) değerleri makul olup, modellerin hem eğitim hem de doğrulama setlerinden bileşiklerin serbest radikal yakalama aktivitesini tahmin etmek için anlamlı ve sağlam olduğunu göstermektedir. Elde edilen modellerde uygulanabilirlik alanı tanımlama tekniği kullanılmış ve çoğu yapının modellerin kimyasal uzayı tarafından yeterince temsil edildiği belirtilmiştir.

Anahtar Kelimeler: *3H-tiazolo[4,5-b]piridin-2-on, QSAR, çoklu doğrusal regresyon (MLR), antioksidan aktivite, uygulanabilirlik alanı*

Received: 08.06.2023
Revised: 27.08.2023
Accepted: 29.08.2023

[*] Orcid ID: 0000-0002-8946-3698, Department of General, Bioinorganic, Physical and Colloidal Chemistry, Faculty of Pharmacy, Danylo Halytsky Lviv National Medical University, Lviv, Ukraine; Departament of Chemistry and Biochemistry, Faculty of Pharmacy, University San Pablo-CEU, Urbanización Montepríncipe, 28668 Boadilla del Monte, España

[°] Corresponding Author; Olena Klenina
Phone: +380 980 519 298, E-mail: olena_klenina@yahoo.com

INTRODUCTION

Nowadays, the discovery of effective antioxidant agents among small molecules is a recent problem that requires new methodological approaches development for the synthesis of novel compounds and their pharmacological activity screening. At the same time, it is also a society-relevant task of life sciences (Vieira & Santos, 2017). Environmental stress factors like pollution, drought, temperature, excessive light intensities, and nutritional limitations can increase the production of reactive oxygen and nitrogen species (Pizzino et al., 2017). Their highly reactive potential is discussed to be responsible for some human diseases, e.g. cancer and cardiovascular diseases. It can cause oxidative damage to proteins, DNA, and lipids in both humans and microorganisms (Pisoschi & Pop, 2015). Despite the tremendous progress in natural antioxidants, elaboration with the new purification and composition methodologies development and novel synthetic antioxidants and drug-like molecules designing, possessing strong antioxidant effects is still a challenging task (Kontogiorgis et al., 2005).

Both thiazole and pyridine scaffolds are of the highest priority in modern medicinal chemistry (De et al., 2022). Different approaches have been developed and introduced for synthesis of new thiazole-based and pyridine-based heterocyclic compounds as biologically active substances (Chaban et al., 2018; Nazir et al., 2023). Also, numerous reports concerning the variety of biological effects possessed by thiazolopyridine annulated system derivatives have been currently published, including their derivatives evaluations as potent antioxidant (Shi et al., 2009), antifungal (Othman et al., 2021), anticancer (Chaban et al., 2012), antimicrobial (El-Mawgoud, 2019), anti-inflammatory (Kamat et al., 2020) and tuberculostatic (Chaban et al., 2014) agents. Diverse kinds of activities, good drug-like properties, and structural functionalization possibilities have led to increasing interest in the design, pharmacological evaluation, and virtual screening proceedings of thiazolopyridine analogs.

In the last decades, virtual screening approaches and tools, including quantitative structure-activity and structure-property relationship analysis, computer molecular modeling methodologies, and combinatorial synthesis in combination with high-throughput total screening became the current pipeline in hit biologically active compounds discovering in early stages (Neves et al., 2018). Modern ligand-based strategies including structure similarity search, quantitative structure-activity relationship (QSAR) analysis and modeling of pharmacophores for libraries of low molecular weight compounds, are in the process of development (Muegge & Oloffa, 2006; Ekins et al., 2007). Thus, methods of designing new drug-like compounds based on such strategies increase the efficiency of creating potential drugs significantly at the stage of lead identification and their structural optimization. QSAR analysis was developed and introduced into the drug design process as the statistically significant tool to correlate structural descriptors of compounds with their biological activities or toxicity. During the recent decades, QSAR became an incredible part of achieving more potent biologically active compounds *via* hit identification and hit-to-lead optimization. Nowadays, drug design with QSAR analysis workflow is commonly used as one of the preliminary stages of high-throughput screening technologies as the labor-, time-, and cost-efficient technique to obtain compounds with desired biological properties. Knowledge of structure-activity regularities makes it possible to predict ways for the synthesis of novel and more effective drug candidates within particular chemical series, contributes to a deeper understanding of their action mechanisms and opens the possibility of virtual screening for compounds before their direct synthesis using the obtained QSAR models.

This research aimed to develop interpretable QSAR models for a series of 3*H*-thiazolo(4,5-*b*)pyridin-2-one derivatives as potential antioxidant drug candidates, which can be used further for QSAR-based *in silico* screening of virtual libraries in the same chemical domain.

MATERIAL AND METHODS

Dataset curation workflow

Molecular 2*D* structures of thiazolo(4,5-*b*)pyridin-2-ones were drawn with ACD/ChemSketch chemical formulas redactor (https://www.acdlabs.com/, assessed March 30, 2023). Later, they were converted to 3*D* structures using Hyper-Chem 7.5 software. Molecular mechanics energy minimization of all compounds was carried out using the MM+ force field, and repeated minimization was performed using the semi-empirical AM1 quantum-chemical method for closed-shell systems until the root-mean-square (RMS) deviation of 0.01 kcal/mol was achieved. Conformations of compounds were optimized through the AM1 method with the global minima selection among all energy-minimal conformers. 3*D* globally minimized structures as *hin* HyperChem output were converted into SMILE (.smi) format, structural data file (SDF) was prepared with E-BABEL on-line version. 20 Subsets of molecular descriptors were calculated using E-DRAGON software (Tetko et al., 2005). The structural parameters calculated after discarding the constant and the near constant values were saved and further analyzed. Before starting the construction of the models, the descriptor normalization procedure was carried out, and the values of all generated descriptors were scaled as (**1**):

$$X_{ij}^{n} = \frac{X_{ij} - X_{j,\min}}{X_{i,\max} - X_{j,\min}} \qquad (\mathbf{1}),$$

where X_{ij} and X_{ij}^{n} are the original and normalized values of the descriptor j (j = 1, 2, … , **K**) for i^{th} compound (i = 1, 2,…, 32), respectively; $X_{j,\min}$ and $X_{j,\max}$ are the minimal and the maximal values for the j^{th} descriptor. Thus, for all normalized descriptors, the following criteria are fair: min (X_{ij}^{n})= 0 and max (X_{ij}^{n}) = 1. The overall data set was then split into training and test (validation) sets manually using the activity sampling approach in the ratio 24:8, according to which the training set consisted of 24 compounds (75% of all), and the validation set contained 8 compounds (25%). The selection of the optimal set of molecular descriptors was carried out using the genetic algorithm. Variables selection was carried out within each of the descriptor modules (0*D*-,1*D*-, 2*D*, 3*D*, and module "Other") firstly using a previously reported approach (Suleiman et al., 2014). The Multiple Linear Regression (MLR) method was applied to generate structure-antioxidant activity QSAR models with the training set compounds using the BuiltQSAR program (de Oliveira & Gaudio, 2000).

Statistical data analysis, internal prediction of models` stability, and external validation

The constructed models were evaluated by the values of the statistical indicators such as the coefficient of determination R^2, the standard deviation s and the value of the Fisher test F. Adjusted regression coefficient R^2_{adj} was used to ensure that all independent variables (predictors) contribute with the equal significance to explain dependent (target) variable. R^2_{adj} was defined as (**2**) (Ouattara, 2017):

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \qquad (\mathbf{2}),$$

where n – total set size, k – the number of independent variables (the number of descriptors in the model), n-k-1 – degrees of freedom. The Sum of squares for the regression ($SS_{regression}$ *or SSR*) was estimated as the sum of the differences between *the* predicted ***value*** of the activity and the mean of the dependent variable for the training set compounds (**3**). In contrast, the Mean square for the regression ($MS_{regression}$ or *MSR*) was an estimate of the variance of the regression (**4**):

$$SSR = \sum\nolimits_{i=1}^{n} \left(Y_{i\ pred} - \overline{\overline{Y}_{training}}\right)^2 \qquad (\mathbf{3}).$$

$$MSR = SSR / DF_{regression} \qquad (\mathbf{4}),$$

where $DF_{regression}$ is the degree of freedom for the regression: $DF_{regression} = k$. The internal validation of derived models was carried out with the leave-one-out cross-validation coefficient (Q^2_{LOO}), predicted residual sums of squares standard deviation (S_{PRESS}), the standard deviation error in prediction (S_{DEP}), and leave-one-out *C.V.* (cross-validation) estimator. The leave-one-out (LOO) cross-validation technique involves

removing an observation from the training set, then constructing a new model, and finally predicting, by this model, the activity of the observation removed. The cross-validation coefficient is calculated at the end of the cycle (He &Jurs, 2005) (formula **5**):

$$SSR = \sum_{i=1}^{n} (Y_{i\ pred} - \overline{Y}_{training})^2 \quad (5),$$

where and are the experimental and predicted biological activity values for molecule *i* of the training set, respectively. The summation in this and all the following equations are done over all *n* compounds of the training set. *SSE* is the Sum of squares for errors (**6**), that is, the sum of the differences between the observed values of the activity and the predicted ones (formula **6**). The Mean square error (*MSE*) (formula **7**) is the average squared distance between the observed and predicted values.

$$SSR = \sum_{i=1}^{n} (Y_{i\exp} - Y_{ipred})^2 \quad (\mathbf{6});$$

$$MSE = SSE / DF_{res} \quad (\mathbf{7}),$$

where DF_{res} is the degree of freedom for the residuals: $DF_{res} = n - k - 1$. For the training set compounds $DF_{res\ (training)} = 24 - 3 - 1 = 20$. *MSE* is a risk function corresponding to the expected value of the squared error loss. The predictive power of models was validated with the external set compounds. The value of leave-group-out cross-validation coefficient Q^2_{LGO} was used as the quantitative characteristics of the external validation (Golbraikh &Tropsha, 2002), which was calculated as (**8**):

$$Q^2_{LGO} = 1 - \frac{\sum (Y_{i\exp(test)} - Y_{ipred(test)})^2}{\sum (Y_{i\exp(test)} - \overline{Y}_{test})^2} \quad (\mathbf{8}),$$

where and are the activity values for the validation set compounds, observed and predicted with the corresponding model, respectively; is the mean value of the experimental activity of the validation (test) set compounds.

Applicability domain (AD) determining

The applicability domain (AD) was assessed using the Williams plot (Weaver & Gleeson, 2008). The standardized residuals d_i for predicted free radical scavenging activity for the validation set compounds were calculated as the residuals in activity divided by their standard deviations (**9**):

$$d_i = \frac{e_i}{\sqrt{MS_{res}}} = \frac{Y_{i\exp(test)} - Y_{i\ pred(test)}}{\sqrt{MS_{res}}} \quad (\mathbf{9}),$$

where (**10**) is the Mean square for residuals (**11**) (the mean square in an estimate of the activity variance).

$$MS_{res} = SS_{res} / DF_{res\ (test)} \quad (\mathbf{10});$$

$$SSres = \sum_{i=1}^{n} (Y_{i\exp(test)} - Y_{ipred(test)})^2 \quad (\mathbf{11}).$$

The applicability domain is the mechanical structural requirements derived from interactive hypothesis generation and testing in the design of the training set. AD was employed to confirm that the obtained model can be considered reliable. Williams plot or leverage approach was used to measure the influence of descriptors on the model (Tropsha et al., 2003). The leverage value ($\boldsymbol{h_i}$) shows the distance of a compound from the centroid of ***X***, which is defined as: , where $\boldsymbol{x_i}$ is the descriptor row vector of the query molecule; *X* is ($m \times p$) characteristic matrix of the data set (*m* is the number of the training set samples and *p* is the number of descriptors). The diagonal elements in this matrix represent the leverage values (h_i) for the molecules in the dataset. The critical leverage value (h^*) is defined as: $h^* = 3(k + 1) / n$, where *k* is the number of descriptors in the model and *n* is the total number of compounds in the training set. For our data set: $h^* = 3\ (3+1)/24 = 0.50$. The leverage values for activities of the training and validation sets compounds were calculated with a H_i-Calculator by DTC Lab (Roy et al., 2015).

RESULTS AND DISCUSSION

Data set preparation

Synthesis and molecular modeling of novel drug candidates based on fused *3H*-thiazolo(4,5-*b*)pyridin-2-one scaffold had extensive development during the recent decade (Chaban et al., 2013; Chaban et al., 2016; Klenina et al., 2017). The synthesis of a series of N^3-substituted *3H*-thiazolo(4,5-*b*)pyridin-2-one derivatives as potential antioxidant drug candi-

dates was previously reported (Chaban et al., 2016; Chaban et al., 2019). The set of compounds used in the present study comprises a series of N'-(2-(5,7-dimethyl-2-oxo-thiazolo(4,5-*b*)pyridine-3-yl)-acetyl) carboxylic acids hydrazides **1**-**10 (**Figure 1), hetarylsulfanyl derivatives of N'-(5,7-dimethyl-2-oxo-thiazolo(4,5-*b*)pyridine-3-yl)-acetyl hydrazide acetic acid (compounds **11**-**14**), 3-(5-mercapto-(1,3,4) oxodiazole-2-yl-methyl)-5,7-dimethyl-3*H*-thiazolo(4,5-*b*)pyridine-2-one (compound **15**), 2-(5,7-dimethyl-2-oxo-thiazolo(4,5-*b*)pyridine-3-yl)-N-(4-oxo-2-thioxo-thiazolydine-3-yl)-acetamide (compound **16**), N-(5-(4-arylidene)-4-oxo-2-thioxo-thiazolydine-3-yl)-2-(5,7-dimethyl-2-oxo-thiazolo(4,5-*b*)pyridine-3-yl)-acetamides (compounds **17**-**22**) and S-substituted 3-(5-mercapto-(1,3,4)oxodiazole-2-yl-methyl)-5,7-dimethyl-3*H*-thiazolo(4,5-*b*) pyridine-2-ones (compounds **23**-**32**). Thus, a data set of thirty-two compounds was generated. The antioxidant activity evaluation of all tested compounds was reported as a spectrophotometric DPPH assay based on the ability of antioxidant drug candidates to possess free radical scavenging potency. The percentage of free-radical-scavenging activity expressed *via* percent inhibition is listed in Table 1S of the Supplementary materials.

In the present research, the overall set of compounds was split into training and validation (test) sets manually based on the criterion of their antioxidant activity (Golbraikh & Tropsha, 2000; Golbraikh et al., 2003). All compounds were divided into four groups according to their activity values: the most active compounds with radical scavenging activity values of 34.27 and 36.14% (compounds **1** and **3**) composed the first group, followed by compounds possessing high antioxidant activity at the level of 10.05 - 20.55% (namely compounds **4**, **10**, **13**, **16**, **19**, **26**, **27** and **28**). The third group contained compounds exhibiting moderate antioxidant activity in the range of 8.11 - 9.74% (compounds **2**, **8**, **9**, **11**, **12**, **14**, **15**, **17**, **18**, **22**, **23**, **24**, **29**, and **31**), and the low-active compounds composed the fourth group with their antioxidant activity evaluated within 4.95 – 7.43% and contained compounds **5**, **6**, **7**, **20**, **21**, **25**, **30** and **32**.

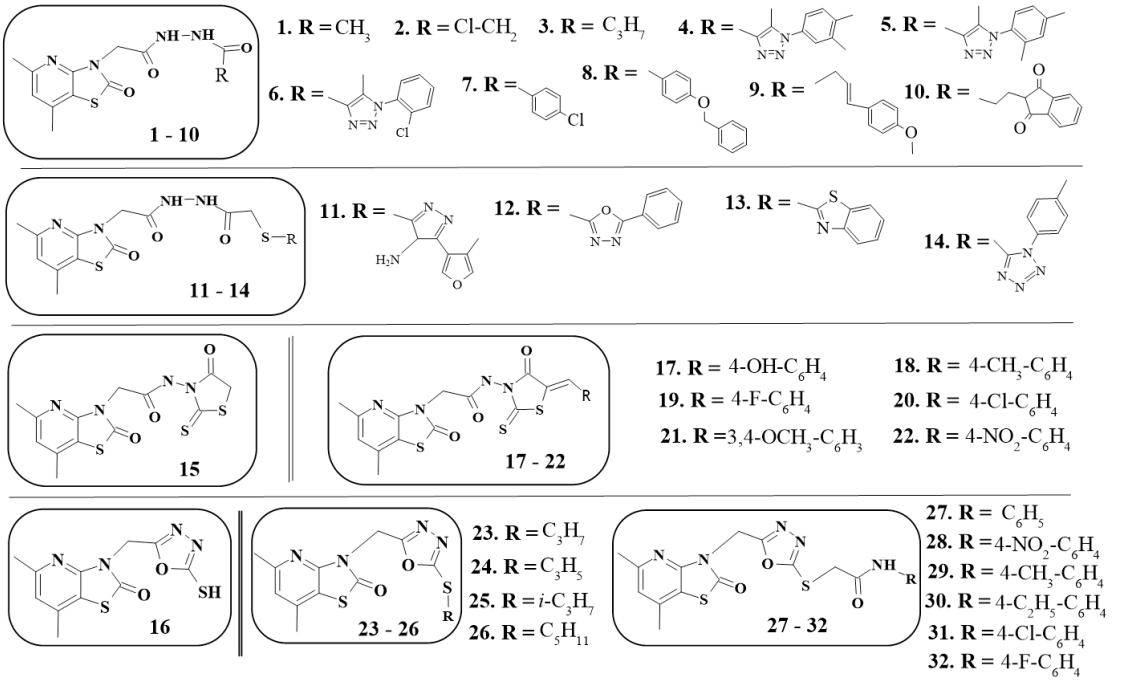


Figure 1. General structures of N[3] substituted 3*H*-thiazolo(4,5-*b*)pyridin-2-one derivatives and chemical structures of *R* substituents

The database was split into training and validation sets in proportion of 75% and 25%, respectively. The test set was composed of all four groups of compounds in the appropriate ratio: one compound (**3**) from the first group, three compounds (**4**, **10**, **16**) from the second one, and two compounds each from the third (compounds **14**, **31**) and the fourth (compounds **5**, **25**) groups.

Chemical structures optimization and molecular descriptors calculation

We have used 2*D* structures of thiazolo(4,5-*b*) pyridin-2-ones, which were then converted into 3*D* ones, optimized, and utilized for molecular descriptors calculation. The structural parameters calculated after discarding descriptors with the constant and near-constant values were saved. Before starting the construction of QSAR models, the descriptor normalization procedure was carried out, and the values of all generated descriptors were scaled in the range of 0÷1. Descriptors with high pairwise correlation, determined based on correlation matrix analysis, were excluded from the multidimensional descriptor space. As a result, a set of 571 descriptors was obtained for both training and test sets compounds.

Descriptors Selection Strategy, QSAR Models Generation

In the present work, the selection of the optimal set of molecular descriptors was carried out using a genetic algorithm within each of the descriptor modules, namely 0*D*-,1*D*-, 2*D*, 3*D*, and module "Other" firstly. The most significant descriptors from each dimensionality module were introduced to the final data set. MLR was used to generate structure - antioxidant activity models as multivariate linear regressions within the training set compounds.

Minimally recommended values of validation parameters for a generally acceptable QSAR model if the following conditions are satisfied: $R^2 > 0.6$; $Q^2 > 0.6$, and $R^2_{pred} > 0.5$ (Golbraikh et al., 2002). The most reliable models were evaluated using the following statistical parameters: n (the number of compounds in regression), determination coefficient R^2, standard deviation s, and F-test, which reflected the ratio of the variance explained by the model and the variance due to the error in the regression. The following validation parameters were also calculated: Q^2_{LOO} as leave-one-out cross-validation coefficient, standard deviation of the sum of squares of prediction error S_{PRESS}, and standard deviation of error of prediction S_{DEP}. Among the generated models, four three-parameter QSAR models were selected with the highest statistical characteristics and predictive ability:

Inhibition, % = - 26.5686 MATS4m + 12.7944 Mor13u – 33.0080 ALOGPS_logP + 34.3967 (**1**)

($n = 24$; $R = 0.901$; $s = 2.972$; $F = 28.867$; $p < 0.0001$; $Q^2_{LOO} = 0.7480$; $S_{Press} = 4.458$; $S_{DEP} = 4.157$);

Inhibition, % = + 18.4827 GATS2m + 16.7743 GATS5m – 30.3645 SPAN + 9.6505 (**2**)

($n = 24$; $R = 0.894$; $s = 3.077$; $F = 26.485$; $p < 0.0001$; $Q^2_{LOO} = 0.7106$; $S_{Press} = 4.307$; $S_{DEP} = 4.017$);

Inhibition, % = - 27.2730 MATS4m -10.5559 RDF130m -22.8937 ALOGPS_logP + 39.1485 (**3**)

($n = 24$; $R = 0.865$; $s = 3.449$; $F = 19.728$; $p < 0.0001$; $Q^2_{LOO} = 0.7060$; $S_{Press} = 5.234$; $S_{DEP} = 4.881$);

Inhibition, % = + 18.7576 GATS5m - 26.5965 BEHm8 + 9.6814 G1e + 15.2389 (**4**)

($n = 24$; $R = 0.865$; $s = 3.443$; $F = 19.816$; $p < 0.0001$; $Q^2_{LOO} = 0.7230$; $S_{Press} = 5.172$; $S_{DEP} = 4.823$).

Models **1**-**4** were constructed using molecular descriptors of 2*D* and 3*D* groups and descriptor ALOGPS_logP from the subset "Others" with low pairwise correlation. The normalized values of molecular descriptors used to build models **1**-**4** are listed in Tables 2S and 3S of the Supplementary materials. To assess the prediction accuracy of the generated models, the prediction errors and the prediction error standard deviations were calculated for each model (Tables 4S and 5S of the Supplementary materials).

The coefficients of the molecular descriptors suggest that the "Molecular properties" **ALOGPS_logP**

for Model **1,** 3*D* Geometrical descriptor **SPAN** for Model **2**, 2*D* autocorrelation descriptor **MATS4m** for Models **1** and **3**, and 2*D* Eigenvalue of Burden matrix **BEHm8** descriptor for Model **4** are the most impactful descriptors for enhancing antioxidant activity of *3H*-thiazolo(4,5-*b*)pyridine-2-one derivatives.

QSAR models interpretation

According to the 5th principle for the validation for regulatory purposes of QSAR models adopted by OECD, a mechanistic interpretation for a generated QSAR model should be made, if possible. Thus, when a mathematical model is established, the researcher should be in a position of the activity mechanism interpretation.

All constructed models contain 2*D* autocorrelation descriptors. In particular, models **1** and **2** contain Moran autocorrelation coefficient with lag 4, weighted by atomic masses (**MATS4m**), while models **2** and **4** contain Geary autocorrelation coefficients with lags 2 and 5, weighted by atomic masses (**GATS2m** and **GATS5m**). In general, *2D* autocorrelation descriptors represent the topological structure of compounds and describe the mutual correlation of certain properties of atoms in intervals equal to the sums of topological distances in the corresponding structural fragments (Helguera et al., 2008). The presence of lags 4 2*D* Autocorrelations in QSAR regressions **1** and **3** may be reviewed as the association of activity information content with structural fragments of such size. It should be noted that mass-weighted Moran auto-correlation coefficient **MATS4m** negatively contributed to the free radical scavenging activity. Based on these models' interpretation, it can be stated that the presence of structural fragments with the sum of topological distances (lag) equal to 4 in the molecules of the training set substances, whose terminal atoms have high atomic masses, is undesirable. Multiply regressions also utilize 2*D* Geary autocorrelations; namely, model **2** contains a **GATS2m** descriptor while **GATS5m** is introduced into models **2** and **4**. For both mentioned descriptors, the regression coefficients have positive signs. Based on the interpretation and analysis of Geary autocorrelations contribution in activity according to models **2** and **4** it can be asserted that the presence of structural fragments with sums of topological distances equal to 2 and 5, whose terminal atoms have high atomic masses in the molecules corresponds to the antioxidant activity enhancing. Model **4** also incorporates the 2*D* descriptor **BEHm8** (Highest eigenvalue n. 8 of Burden matrix / weighted by atomic masses). It belongs to the BCUT group, where the descriptor is based on a weighted version of the Burden matrix, which considers both the connectivity and atomic properties (Consonni & Todeschini, 2009). The weights are various atom properties placed along the diagonal of the Burden matrix. The atomic mass weighting scheme is employed in the **BEHm8** descriptor. Derived model **4** comprises the **BEHm8** descriptor, which makes a negative contribution to the activity. This fact could be related to the importance of the Burden matrix eigenvalues on the antioxidant activity and may be interpreted as follows: the increase in radical scavenging activity is observed when the molecules contain only lightweight atoms characterized by low electronegativities. At the same time, the electron density redistribution between distant atoms and groups of atoms is undesirable. Model **2** contains the value of the 3*D* descriptor **SPAN**, one of the numerous geometrical descriptors subsets. Geometrical descriptors set consolidates different conformationally dependent descriptors based on the molecular geometry. The simplest geometrical descriptor is the SPAN size descriptor, which is defined as the radius of the smallest sphere, centered at the molecule`s center of mass, that completely comprises all atoms of the molecule (Consonni & Todeschini, 2009): $R = \max_i r_i$. Normalized **SPAN** descriptor values contribute negatively to the antioxidant activity, as evidenced by the negative sign of the regression coefficient in model **2**. So, it may be suggested that the molecules` sizes increasing due to the increasing the distances between certain atoms and fragments and the centers of masses leads to a decrease in the radical scaveng-

ing activity of compounds. Model **3** incorporates a 3*D* **RDF130m** descriptor (Radial Distribution Function - 13.0 / weighted by atomic masses) with a negative regression coefficient. RDF descriptors are molecular descriptors obtained by radial basis functions centered on different interatomic distances (from 0.5Å to 15.5 Å) (González et al., 2008). The radial distribution function (RDF) is a 3*D* conformational molecular descriptor defined based on the distribution of interatomic distances in a molecule. Formally, the RDF of a group of *n* atoms can be interpreted as the probability of finding an atom in a spherical volume with the radius *R*. Based on the analysis of model **5**, it can be argued that free radical scavenging activity enhancing for compounds under study is ensured by the negative contribution of **RDF130m** descriptor, which corresponds to the atomic radius of 13.0 Å. The interpretation of the obtained QSAR model in the sense of specific contributions of substituents and other features of the molecular steric structure indicates the presence of the linear relationship between the activity of the compounds and 3*D* molecular distribution of atomic masses in a spherical volume with the radius of 13.0 Å. Thus, the decreasing antioxidant activity may be caused by heavy atoms availability within this volume. Model **1** was derived with 3D-MoRSE descriptor **Mor13u** (signal 13 / unweighted). 3D-MoRSE (3D Molecule Representation of Structures based on Electron diffraction) descriptors are calculated from the model of the IR spectrum using the general scattering function (Consonni & Todeschini, 2009). Thus, 3*D* Molecule Representation of Structure based on Electron diffraction (MoRSE) descriptors provide 3*D* information from the 3*D* coordinates by using the same transform as electron diffraction (which uses it to prepare theoretical scattering curves). 3D-MoRSE – signal 13/unweighted descriptor **Mor13u** is incorporated into regression **1** for antioxidant activity of N^3 substituted of *3H*-thiazolo(4,5-*b*)pyridin-2-ones making the positive contribution in the activity enhancing. Thus, the increase in activity occurs when the electron beam scattering would be possibly more intensive, mainly on account of groups of any atoms located at a distance of 13 Å. Model **4** contains 3*D* descriptor **G1e** from the WHIM group (1^{st} component directional WHIM index / weighted by Sanderson electronegativity). WHIM descriptors (Weighted Holistic Invariant Molecular descriptors) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes (Todeschini & Gramatica, 2002). WHIM descriptors capture relevant molecular 3*D* information regarding molecular size, shape, symmetry, and atom distribution concerning invariant reference frames. Derived model **4** for antioxidant activity of the compounds under study comprises 1^{st} component directional WHIM index / weighted by Sanderson electronegativity **G1e**, which makes a positive contribution to the activity. This fact could be related to the importance of strong symmetry keeping with the atoms possessing high electronegativity, like Oxygen or Chlorine, along the 1^{st} principal component axe. QSAR models **1** and **3** also incorporate **ALOGPS_logP** descriptor (Ghose–Crippen octanol-water partition coefficient), which refers to a group of molecular properties calculated from models and some empirical descriptors (Ghose et al., 1999). Ghose–Crippen octanol-water partition coefficient ALOGPS_logP is calculated based on a model of the specific contribution of each functional group to the overall value of the partition coefficient of the compound. The analysis of models **1** and **3**, in which the regression coefficients for the normalized ALOGPS_logP descriptor have negative signs, indicates that the antioxidant activity of *3H*-thiazolo(4,5-*b*)pyridin-2-ones increases with a decrease in the values of ALOGPS_logP, which corresponds to their hydrophilic properties increasing.

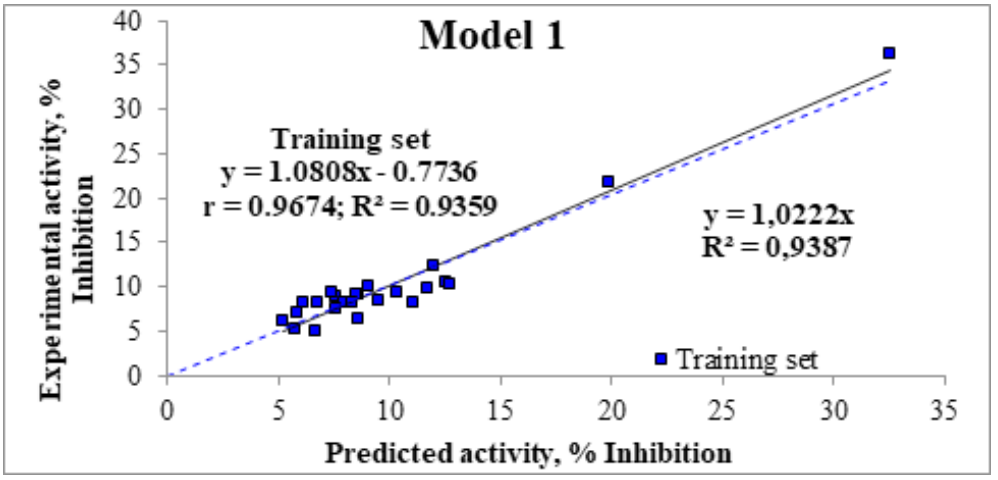
Statistical evaluation of the constructed models

Based on the validation parameters of the generated models (Table 1), it may be stated that they all satisfy the statistical requirements for their goodness of fit with no current overfitting.

Table 1. Statistical parameters for QSAR models 1 – 4

No	Statistical parameter	Model 1	Model 2	Model 3	Model 4
1.	n	24	24	24	24
2.	k	3	3	3	3
3.	$r_{(observed\ vs.\ predicted)}$	0.9674	0.7989	0.7474	0.7483
4.	$R^2_{(observed\ vs.\ predicted)}$	0.9359	0.9064	0.8915	0.8318
5.	R^2_{adj}	0.7842	0.7687	0.7095	0.7105
6.	$R^2 - R^2_{adj}$	0.1168	0.1253	0.1555	1.1545
7.	s	2.9721	3.077	3.4485	3.4428
8.	$DF_{regression}$	3	3	3	3
9.	SSR	764.9861	752.2926	703.8127	704.6043
10.	MSR	254.9954	250.7642	234.6042	234.8681
11.	F	28.8667	26.4848	19.7275	19.8157

The goodness of fitting for QSAR models generated with k parameters for the training set consisting of n compounds was assured by maximizing Pearson correlation coefficient r, determination coefficient R^2, adjusted regression coefficient R^2_{adj} and F-test criterion while minimizing $R^2 - R^2_{adj}$ and standard deviation s. It should be noted that the degree of freedom for the regression equals the number of independent parameters: $DF_{regression} = k$. $SS_{regression}$ (SSR) means the Sum of squares for the regression, while $MS_{regression}$ (MSR) means the Mean square for the regression. The regression sum of squares is interpreted as the amount of total variation explained by the model. SSR is a variation in Y_i associated with the regression line, . It may be referred to as the measure that describes how well our line fits the data. $MS_{regression}$ is an estimate of the regression variance, $MS_{regression} = SS_{regression} / DF_{regression}$. If this value of $SS_{regression}$ is equal to the sum of squares total, it means our regression model captures all the observed variability and is perfect. Correlation coefficients and coefficients of determination between observed and predicted activities for the training set compounds were estimated as internal validation predictivity. Figure 2 shows the regression between experimental and predicted values of free radical scavenging activity for Models **1**-**4**. The determination coefficients for observer endpoints versus experimental oner for the derived models are between 0.8318 and 0.9359, so the linear models explain 83.18% - 93.59% of variation in experimental activity. R^2_{adj} values for the models calculated as 0.7095 - 0.7842 ensure high enough significance of all introduced variables.



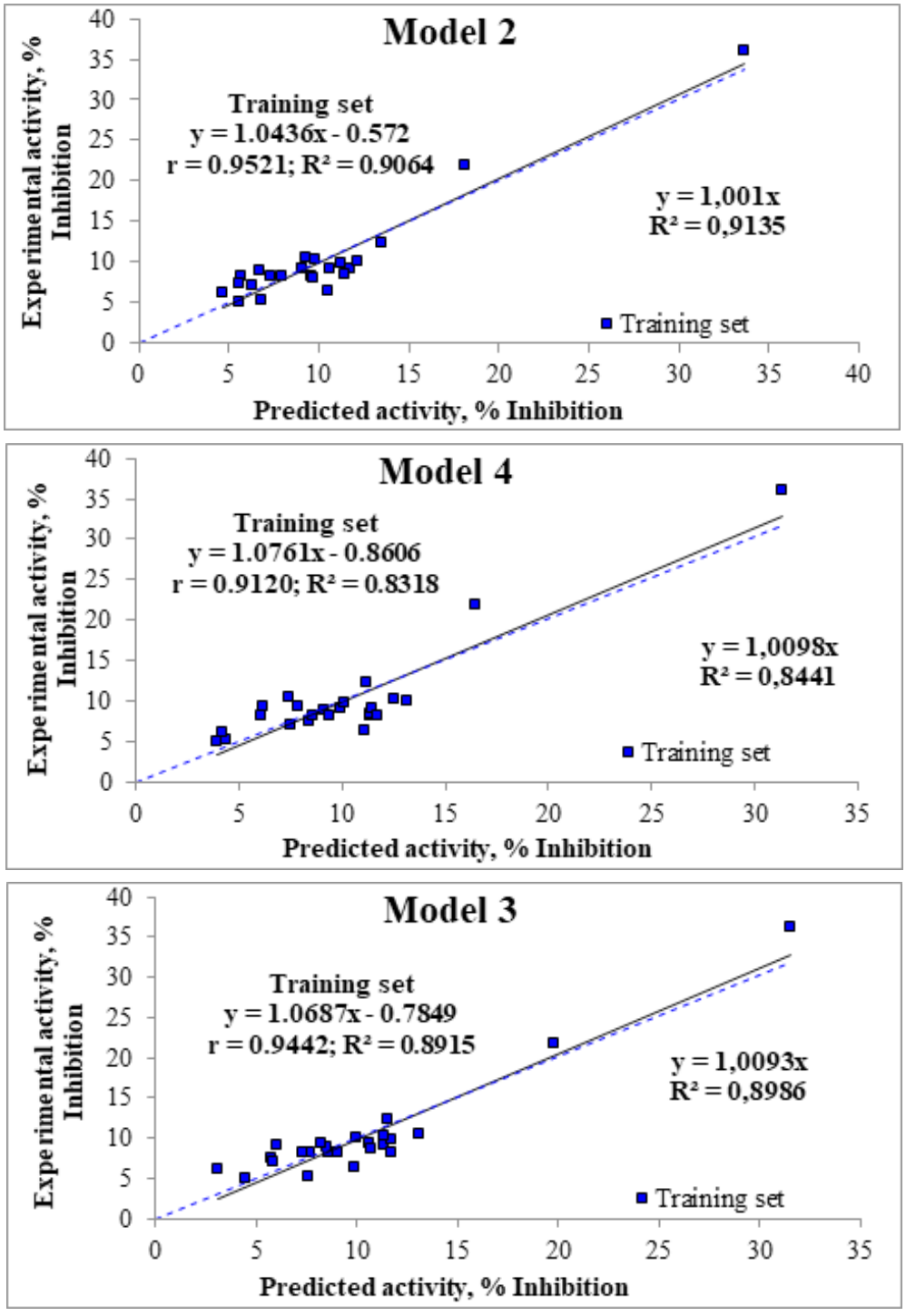


Figure 2. Correlation between observed and predicted activity values for constructed Models 1 - 4 for the training set compounds together with linear fit statistical parameters

The statistical significance of the constructed models given by the *F*-test in the range of 19.7275 - 28.8667 is also satisfactory. Thus, generated QSAR models could approximate the experimental values properly according to their statistical analysis performance parameters.

Internal and external validation of the derived models

The predictive ability (or predictive power) of a model is ensured by using test set data on the model, which in turn is developed by the training set. To validate the constructed QSAR models, we applied leave-one-out (LOO) and leave-group-out (LGO) cross-validation procedures. In the case of an internal validation procedure, each object is removed from the original training dataset in turn, and the remaining reduced data set is converted into a new training set used for model generation and the response prediction for the excluded compound. The outcome of this procedure is a leave-one-out cross-validation coefficient Q^2_{LOO}. For the developed models **1**-**4**, we evaluated the accuracies of QSAR models also using the difference $|R^2 - Q^2_{LOO}|$ absolute value, supposing that it tends to minimize for a truly predictive model. The values of Q^2_{LOO} (0.7060 ÷ 0.7480) are reasonable, showing that the models are significant and robust to predict the free radical scavenging activity of the compounds under the study. The absolute values of the difference between R^2 and Q^2_{LOO} are in the range of 0.0252 - 0.0886, within the suggested limit (Kiralj & Ferreira 2009) of $|R^2 - Q^2_{LOO}| < 0.3$, which is an indication that the model does not have data overfitting. Internal validation parameters are summarized in Table 2.

Table 2. Internal validation parameters for QSAR models 1 – 4

No	Validation parameter	**Model 1**	**Model 2**	**Model 3**	**Model 4**
1.	$\boldsymbol{Q^2_{LOO}}$	0.7480	0.7106	0.7060	0.7230
2.	$\boldsymbol{\lvert R^2 - Q^2_{LOO} \rvert}$	0.0638	0.0886	0.0422	0.0252
3.	$\boldsymbol{S_{Press}}$	4.4575	4.3073	5.2342	5.1716
4.	$\boldsymbol{S_{DEP}}$	4.1567	4.0165	4.8809	4.8225
5.	$\boldsymbol{DF_{res\ (training)}}$	20	20	20	20
6.	***SSE***	176.6712	189.3646	237.8446	237.053
7.	***MSE***	8.8336	9.4682	11.8922	11.8526
8.	***C.V.***	29.3314	30.3669	34.0328	33.9761

For internal validation, we also applied S_{DEP} (standard deviation of error of predictions) and the standard deviation of the predicted residual error sum of squares S_{PRESS}. Both S_{PRESS} and S_{DEP} values display a tendency to depreciation to ensure that generated models possess enough predictive power. Mean squared error (*MSE*) measures error in statistical models using the average squared difference between observed and predicted values. Highly predictive linear regression should display minimized *SSE*. The degree of freedom for residuals (errors) for the training set compounds $DF_{res(\text{training})} = n - k - 1$, where n is the number of observations, and k is the number of parameters in the model. The cross-validation *C.V.* estimator for leave-one-out validation is used to indicate the lowest *MSE*. Many authors have suggested that the only way to estimate the true predictive power of a QSAR model is to compare the predicted and observed activities for the validation set compounds (Golbraikh et al., 2003; Gramatica, 2020). Following this guideline, we fulfilled predictive ability external validation of the constructed models. The experimental and predicted values and the residues obtained for the compounds of the two sets, are given in Table 3.

Table 3. Experimental and predicted values of free radical scavenging activity for the test set compounds, Q^2_{LGO}, and standardized residues for developed Models 1 - 4

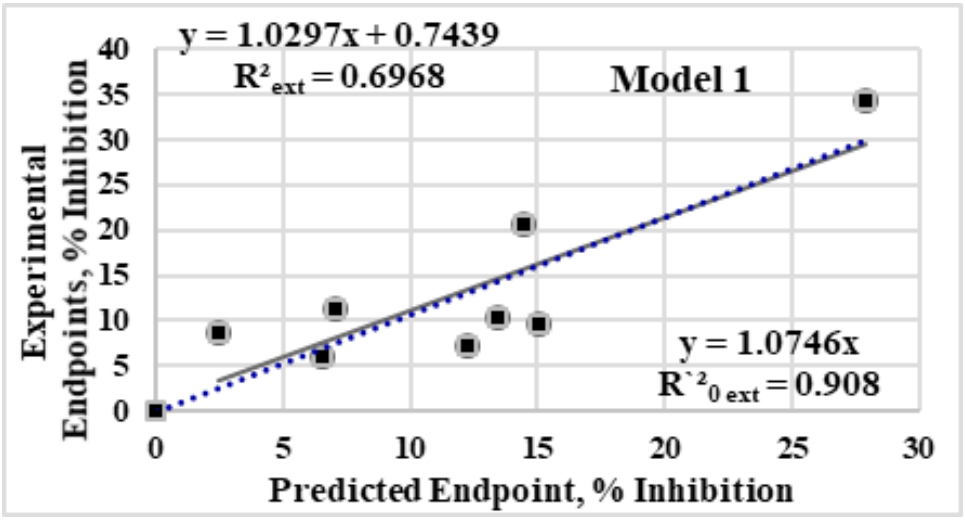
Compound	Observed activity, % Inhibition	Predicted activity, % Inhibition				Standardized residues			
		Models				Models			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
3	34.27	27.872	25.676	26.342	26.749	0.925	0.830	1.041	1.223
4	11.25	7.026	8.393	9.036	4.557	0.627	0.925	0.238	0.711
5	6.12	6.534	7.195	4.420	2.241	-0.259	0.081	0.075	0.023
10	10.45	13.477	13.610	14.150	7.724	-0.799	-0.968	-0.634	-0.076
14	9.50	15.054	12.587	12.608	7.124	-1.292	-0.863	-1.177	-0.170
16	20.55	14.430	16.619	14.038	23.898	0.948	0.481	0.948	-0.959
25	7.15	12.234	12.960	10.480	12.271	-1.186	-1.210	-0.791	-1.181
31	8.70	2.472	6.082	4.538	2.060	1.036	1.064	0.893	0.628
Q^2_{LGO}		**0.7208**	**0.7477**	**0.7711**	**0.6647**				

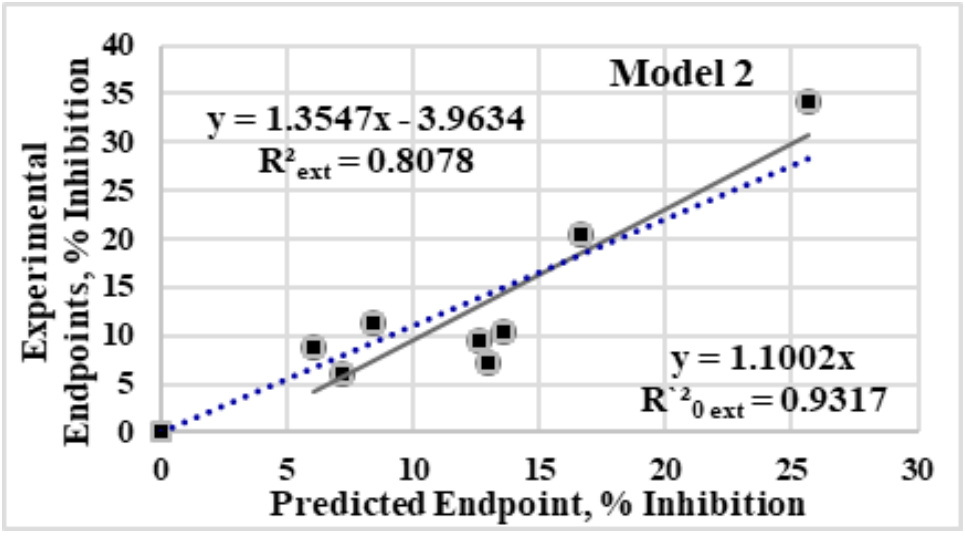
Additionally, the standardized residuals in free radical scavenging activity prediction using Models **1**-**4** were calculated (Table 3). All the residuals of predicted activity values were between -1.292 and 1.223, which indicated that developed models have good accuracy and reliability for predicting the antioxidant activity of N^3-substituted 3*H*-thiazolo(4,5-*b*) pyridin-2-ones.

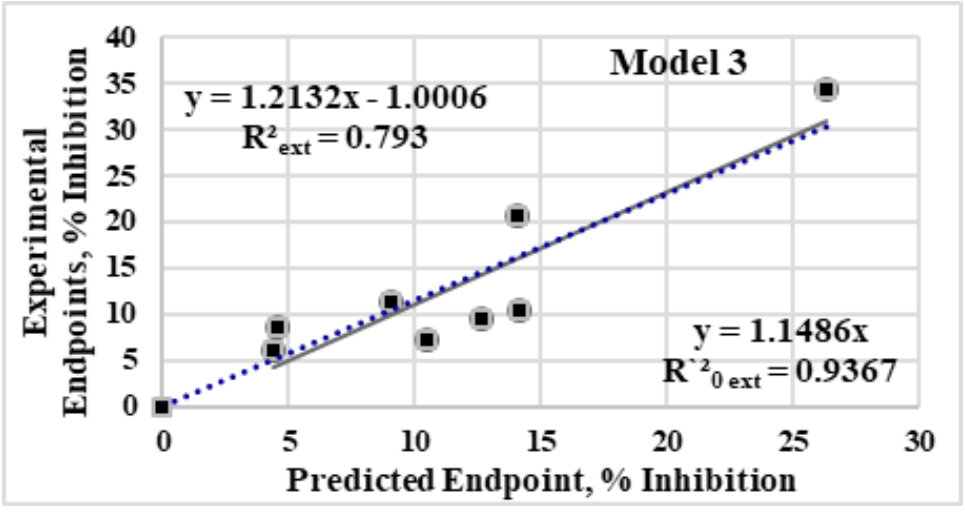
Acceptable QSAR predictive models should satisfy the following conditions (Golbraikh & Tropsha, 2002; Gramatica, 2020): (**i**) $Q^2_{LGO} > 0.5$; (**ii**) $R^2_{ext} > 0.6$; (**iii**) $(R^2_{ext} - R^2_{0\,ext})/R^2_{ext} < 0.1$ and $0.85 \le k \le 1.15$ or $(R^2_{ext} - R`^2_{0\,ext})/R^2_{ext} < 0.1$ and $0.85 \le k` \le 1.15$; (**iv**) $|R^2 - R`^2_{0\,ext}| < 0.1$, where R^2_{ext} is the coefficient of determination between the predicted and observed activities; , and are the coefficients of determination for predicted versus observed activities and observed versus predicted activities, respectively, for regressions through the origin; slopes *k* and *k*′ of the regression lines through the origin. The values of leave-group-out cross-validation coefficients Q^2_{LGO} for all developed models ranged from 0.6647 to 0.7711, as listed in Table 3. The values of the external validation criteria are well illustrated in Figure 3 and are summarized in Table 4.

Table 4. External validation criteria values for Models 1 – 4

No.	Validation criteria	Models			
		Model 1	**Model 2**	**Model 3**	**Model 4**
(ii)	r_{ext}	0.8347	0.8988	0.8905	0.8787
	R^2_{ext}	0.6968	0.8078	0.7930	0.7721
(iii)	$R^2_{0\,ext}$	0.9121	0.9317	0.9367	0.9057
	$(R^2_{ext} - R^2_{0\,ext})/R^2_{ext}$	-0.3090	-0.1534	-0.1812	-0.1730
	k	0.8450	0.8468	0.8155	0.8274
	k`	1.0746	1.1002	1.1486	1.0946
(iv)	$\lvert R^2 - R`^2_{0\,ext} \rvert$	0.0238	0.0253	0.0452	0.0739


y = 1.0297x + 0.7439
R²ext = 0.6968
Model 1
y = 1.0746x
R`²0 ext = 0.908
Experimental Endpoints, % Inhibition
Predicted Endpoint, % Inhibition
0
5
10
15
20
25
30
35
40


Model 2
y = 1.3547x - 3.9634
R²ext = 0.8078
y = 1.1002x
R`²0 ext = 0.9317
Experimental Endpoints, % Inhibition
Predicted Endpoint, % Inhibition
0
5
10
15
20
25
30
35
40


Model 3
y = 1.2132x - 1.0006
R²ext = 0.793
y = 1.1486x
R`²0 ext = 0.9367
Experimental Endpoints, % Inhibition
Predicted Endpoint, % Inhibition
0
5
10
15
20
25
30
35
40

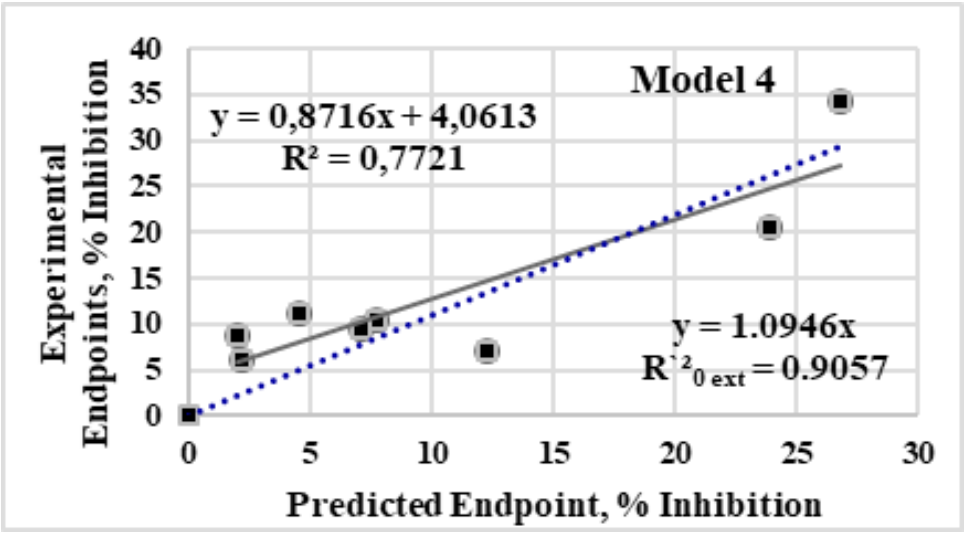
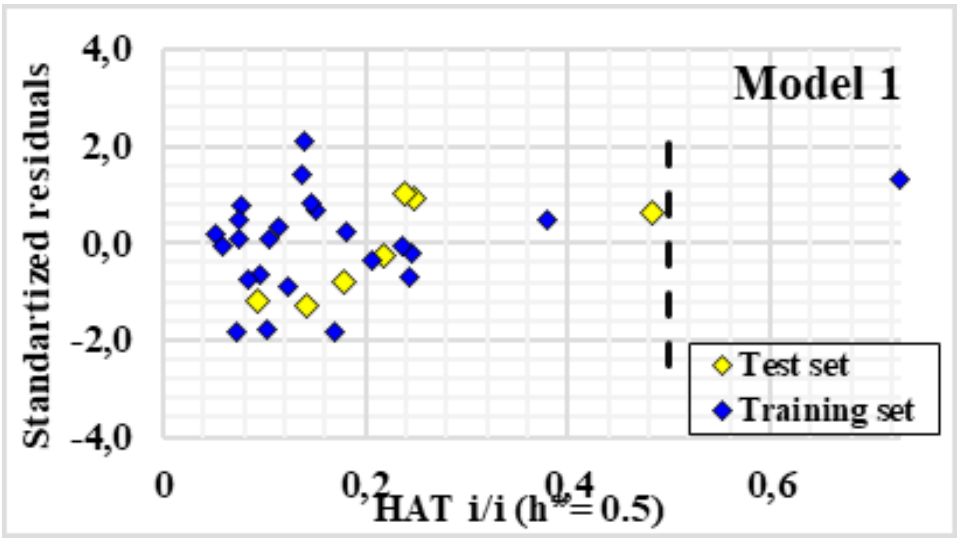


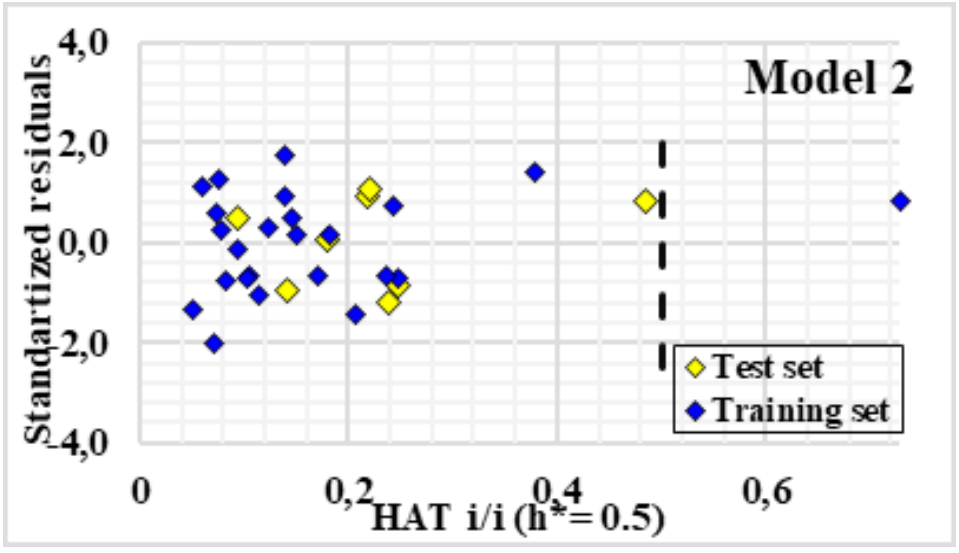
Figure 3. Experimental endpoints vs. Predicted endpoints for Models 1 – 4

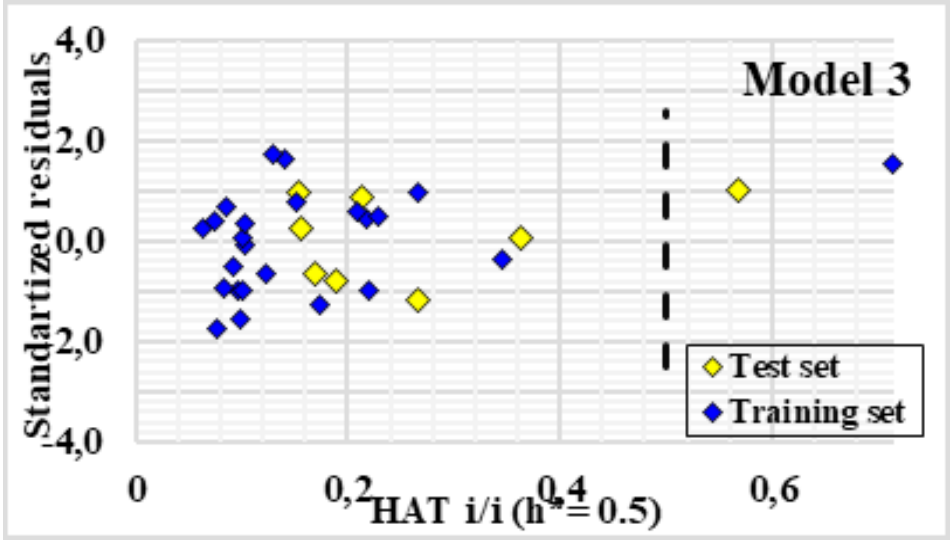
According to Table 4, we demonstrated that developed QSAR models displayed sufficient predictive results under the external validation as they met all the above-listed criteria demands, which ensured the adequate predictive ability of Models **1**-**4**.

Applicability domain defining

The applicability domain (AD) defining for the developed models is one of the most critical aspects of the validation workflow. The AD expresses that QSARs are inescapably associated with restrictions in the categories of chemical structures, physicochemical properties, and mechanisms of action for which the models can generate reliable predictions. The model applicability domain is the theoretical chemical space of the compounds defined by the descriptors and the modeled activity in which the acceptable QSAR model can make reliable predictions (He & Jurs, 2005; Weaver & Gleeson, 2008). The structural features of drug-like molecules should be in close proximity, and a model should be able to outlier those compounds that are far away in their structures from the majority of the set used to construct and validate the model. One of the standard tools used to visualize the AD of a QSAR model is the Williams plot created as standardized residuals in prediction versus leverage values (h_i) for each i^{th} sample. Leverage is a measure of how far away the independent variable values of an observation are from those of the other observations. It is a standardized measure of the distance of x_i from the center of x space. Thus, after generating and evaluating the model, the AD defining technique was employed to confirm that the obtained models can be considered reliable. Williams plots for both training and test sets compounds were generated for Models **1**-**4** (Figure 4). The threshold leverage was defined at the level of $h^* = 0.50$.


Model 1
Standartized residuals
4,0
2,0
0,0
-2,0
-4,0
0
0,2
0,4
0,6
HAT i/i (h*= 0.5)
Test set
Training set


Model 2
Standartized residuals
4,0
2,0
0,0
-2,0
-4,0
0
0,2
0,4
0,6
HAT i/i (h*= 0.5)
Test set
Training set


Model 3
Standartized residuals
4,0
2,0
0,0
-2,0
-4,0
0
0,2
0,4
0,6
HAT i/i (h*= 0.5)
Test set
Training set

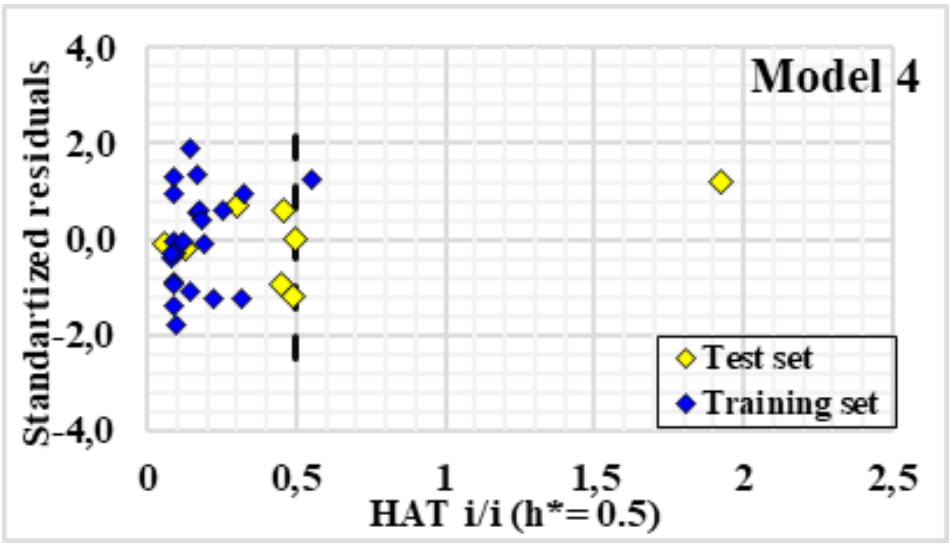


Figure 4. Williams plots for applicability domains of Models 1-4

As shown in Figure 4, for Models **1** and **2**, just compound **1** from the training set lies out of the domain of applicability, while for Models **3** and **4**, in addition to compound **1**, compound **31** from the validation set, is also out of AD. It may be pointed out that these compounds have the highest observed free radical scavenging activity: 36.14% and 34.27%, respectively. Thus, the developed models are applicable for drug-like compounds possessing moderate antioxidant activity, as the most active compounds from both sets fell outside of the AD as outliers.

CONCLUSION

The present study allows us to conclude that GA-MLR QSAR analysis can explain the free radical scavenging activity of N^3 substituted *3H*-thiazolo(4,5-*b*) pyridin-2-one derivatives. It was shown that a set of highlighted 2*D,* 3*D*, and Molecular properties descriptors play the defining role in antioxidant activity estimation of the compounds. Interpretation of generated QSAR models allows to conclude that the presence of structural fragments with the sum of topological distances equal to 2 and 5 with heavy terminal atoms enhances the antioxidant activity of the compounds under study (Geary autocorrelations GATS2m and GATS5m with lags 2 and 5/weighted by atomic masses make positive contributions to the activity). In contrast, similar fragments presence with the sum of topological distances equal to 4 (Moran autocorrelation MATS4m of lag 4 weighted by mass) is undesirable. The presence of SPAN geometrical 3*D* descriptor and BEHm8 Burden eigenvalue matrix 2*D* descriptor in derived models, both contributing negatively, ensures that free radical scavenging activity increasing corresponds to small molecules with the minimal distance of specific atoms and fragments from the center of mass and to the molecules without heavy atoms, the electronic density redistribution between the distant atoms and groups of atoms should be neglectable. The linear relationship between the activity and 3*D* molecular distribution of atomic masses in spherical volume with a radius of 13.0 Å indicates that the antioxidant activity decreasing may be caused by heavy atoms availability within this volume (negative contribution of RDF130m descriptor). The positive contribution of 3D-MoRSE – signal 13/unweighted Mor13u descriptor in the activity ensures that the increase in activity occurs when the electron beam scattering is more intensive, mainly on account of groups of any atoms located at a distance of 13 Å. Some symmetry rules should be complied in the molecules of potent antioxidants, namely strong symmetry of atoms possessing high electronegativity, like Oxygen or Chlo-

rine, along the 1st principal component axe (positive contribution of 1st component directional WHIM index/weighted by Sanderson electronegativity G1e in the activity). The antioxidant activity of the compounds also increases with the hydrophilic properties of the substances increasing (negative contribution of Ghose-Crippen octanol-water partition coefficient ALOGPS_logP in activity). The developed models can be considered reliable, robust, and efficient, capable of predicting the antioxidant effect of novel thiazolopyridines, which is proved by considering their statistical accuracy and the appropriate values of the statistical estimators. The applicability domains were defined for the models. The AD indicated that most structures were adequately represented by the chemical space of the models, so the developed models are applicable for drug-like compounds possessing moderately high antioxidant activity. Models resulting from the predictive QSAR modeling workflow may be used for the virtual screening of antioxidant activity to prioritize the selection of 3*H*-thiazolo(4,5-*b*)pyridin-2-one derivatives for the experimental activity validation.

ACKNOWLEDGMENTS

This research was conducted as part of the scientific research project «Directed search for «drug-like» molecules among new azoles using the strategy of *in silico* design», carried out at Danylo Halytsky Lviv National Medical University, Lviv, Ukraine, state registration number 0121U109330, supported by The Ministry of Health of Ukraine. O. K. thanks Universidad San Pablo CEU, Madrid, Spain, for a Postdoctoral Contract for Ukrainian Researchers; she is exceedingly grateful to Prof. Beatriz de Pascual-Teresa Fernández, all the Group members for their solidarity with Ukraine and their support aimed to assist the Ukrainian University community.

AUTHOR CONTRIBUTION STATEMENT

Study conception and design, data collection and preparation, computational part, obtaining the results and their analysis, conclusions formulation and preparation of the manuscript (OK)

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest.

REFERENCES

Chaban, T. I., Panchuk, R. R., Klenina, O. V., Skorokhyd, N. R., Ogurtsov, V. V., & Chaban, I. G. (2012). Synthesis and evaluation of antitumor activity of some thiazolo(4,5-*b*) pyridines. *Biopolymers and Cell*.

Chaban, T. I., Ogurtsov, V. V., Chaban, I. G., Klenina, O. V., & Komarytsia, J. D. (2013). Synthesis and antioxidant activity evaluation of novel 5, 7-dimethyl-3 H-thiazolo(4,5-*b*)pyridines. Phosphorus, Sulfur, and Silicon and the Related Elements, 188(11), 1611-1620.

Chaban, T., Klenina, O., Drapak, I., Ogurtsov, V., Chaban, I., & Novikov, V. (2014). Synthesis of some novel thiazolo(4,5-*b*)pyridines and their tuberculostatic activity evaluation. *Chemistry and Chemical Technology*, 89: 287-292.

Chaban, T. I., Klenina, O. V., Zimenkovsky, B. S., Chaban, I. G., Ogurtsov, V. V., & Shelepeten, L. S. (2016). Synthesis of novel thiazolo(4,5-*b*)pyridines as potential biologically active substances. Der Pharma Chemica, 8(19), 534-542.

Chaban, T., Klenina, O., Ogurtsov, V., Chaban, I., Lelyukh, M., & Harkov, S. (2018). Thiazolo(5,4-*d*) pyrimidines and thiazolo(4,5-*d*)pyrimidines: A review on synthesis and Pharmacological importance of their derivatives. Pharmacia, 65(2), 54-70.

Chaban, T., Ogurtsov, V., Mahlovanyy, A., Sukhodolska, N., Chaban, I., Harkov, S., & Matiychuk, V. (2019). Antioxidant properties of some novel derivatives thiazolo(4,5-*b*)pyridine. Pharmacia, 66(4), 171-180. https://doi.org/10.3897/pharmacia.66.e36764

Consonni, V., & Todeschini, R. (2009). *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing/Volume II: Appendices, References*. John Wiley & Sons.

De, S., SK, A. K., Shah, S. K., Kazi, S., Sarkar, N., Banerjee, S., & Dey, S. (2022). Pyridine: the scaffolds with significant clinical diversity. *RSC Advances*, *12*(24), 15385-15406. https://doi.org/10.1039/d2ra01571d

Ekins, S., Mestres, J., & Testa, B. (2007). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British journal of pharmacology*, *152*(1), 9-20. https://doi.org/10.1038/sj.bjp.0707306

El-Mawgoud H.K.A. (2019). Synthesis, in-vitro cytotoxicity and antimicrobial evaluations of some novel thiazole based heterocycles. *Chemical & Pharmaceutical Bulletin,* 67(12):1314-1323. https://doi.org/10.1248/cpb.c19-00681

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of combinatorial chemistry*, *1*(1), 55-68. https://doi.org/10.1021/cc9800071

Golbraikh, A., & Tropsha, A. (2000). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. Molecular diversity, 5, 231-243. https://doi.org/10.1023/A:1021372108686

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

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. D., Lee, K. H., & Tropsha, A. (2003). Rational selection of training and test sets for the development of validated QSAR models. Journal of computer-aided molecular design, 17, 241-253. https://doi.org/10.1023/a:1025386326946

González, M. P., Gándara, Z., Fall, Y., & Gómez, G. (2008). Radial Distribution Function descriptors for predicting affinity for vitamin D receptor. *European journal of medicinal chemistry*, *43*(7), 1360-1365. https://doi.org/10.1016/j.ejmech.2007.10.020

Gramatica, P. (2020). Principles of QSAR modeling: comments and suggestions from personal experience. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, *5*(3), 61-97. https://doi.org/10.4018/IJQSPR.20200701.oa1

He, L., & Jurs, P. C. (2005). Assessing the reliability of a QSAR model's predictions. *Journal of Molecular Graphics and Modelling*, *23*(6), 503-523. https://doi.org/10.1016/j.jmgm.2005.03.003

Helguera, A. M., Combes, R. D., Gonzalez, M. P., & Cordeiro, M. N. D. S. (2008). Applications of 2D descriptors in drug design: a DRAGON tale. *Current topics in medicinal chemistry*, *8*(18), 1628-1655. https://doi.org/10.2174/156802608786786598

Kamat, V., Santosh, R., Poojary, B., Nayak, S. P., Kumar, B. K., Sankaranarayanan, M., ... & Vootla, S. K. (2020). Pyridine-and thiazole-based hydrazides with promising anti-inflammatory and antimicrobial activities along with their in silico studies. *ACS omega*, *5*(39), 25228-25239. https://doi.org/10.1021/acsomega.0c03386

Kiralj, R., & Ferreira, M. (2009). Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society*, *20*, 770-787. DOI: http://dx.doi.org/10.1590/S0103-50532009000400021

Klenina, O., Chaban, T., Zimenkovsky, B., Harkov, S., Ogurtsov, V., Chaban, I., & Myrko, I. (2017). QSAR modeling for antioxidant activity of novel N3substituted 5,7-dimethyl-3H-thiazolo(4, 5-b) pyridin-2-ones. Pharmacia, 64(4), 49-71.

Kontogiorgis, A. C., Pontiki, A. E., & Hadjipavlou-Litina, D. (2005). A review on quantitative structure-activity relationships (QSARs) of natural and synthetic antioxidants compounds. *Mini Reviews in Medicinal Chemistry*, *5*(6), 563-574. https://doi.org/10.2174/1389557054023233

Muegge I., Oloffa S. (2006). Advances in virtual screening. *Drug Discov. Today: Technologies*, 3 (4), 405–411. DOI: https://doi.org/10.1016/j.ddtec.2006.12.002

Nazir, M. S., Nawaz, A., Aslam, S., Ahmad, M., Zahoor, A. F., & Mohsin, N. U. A. (2023). Synthetic strategies for thiazolopyridine derivatives. *Synthetic Communications*, *53*(7-8), 519-546. https://doi.org/10.1080/00397911.2023.2183363

Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in pharmacology*, *9*, 1275. https://doi.org/10.3389/fphar.2018.01275

de Oliveira, D. B., & Gaudio, A. C. (2000). BuildQSAR: a new computer program for QSAR analysis. *Quantitative Structure-Activity Relationships: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis*, *19*(6), 599-601. https://doi.org/10.1002/1521-3838(200012)19:6<599::AID-QSAR599>3.0.CO;2-B

Othman, I. M., Gad-Elkareem, M. A., Radwan, H. A., Badraoui, R., Aouadi, K., Snoussi, M., & Kadri, A. (2021). Synthesis, structure-activity relationship and in silico studies of novel pyrazolothiazole and thiazolopyridine derivatives as prospective antimicrobial and anticancer agents. *ChemistrySelect*, *6*(31), 7860-7872. DOI: https://doi.org/10.1002/slct.202101622

Ouattara, O., Affi, T. S., Koné, M. G. R., Bamba, K., & Ziao, N. (2017). Can Empirical Descriptors Reliably Predict Molecular Lipophilicity? A QSPR Study Investigation. *Int. J. of Engineering Research and Application*, *7*(5), 50-56.

Pisoschi, A. M., & Pop, A. (2015). The role of antioxidants in the chemistry of oxidative stress: A review. *European journal of medicinal chemistry*, *97*, 55-74. https://doi.org/10.1016/j.ejmech.2015.04.040

Pizzino, G., Irrera, N., Cucinotta, M., Pallio, G., Mannino, F., Arcoraci, V., ... & Bitto, A. (2017). Oxidative stress: harms and benefits for human health. *Oxidative medicine and cellular longevity*, *2017*. https://doi.org/10.1155/2017/8416763

Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, *145*, 22-29. http://dx.doi.org/10.1016/j.chemolab.2015.04.013

Shi, F., Li, C., Xia, M., Miao, K., Zhao, Y., Tu, S., ... & Ma, N. (2009). Green chemoselective synthesis of thiazolo(3, 2-a)pyridine derivatives and evaluation of their antioxidant and cytotoxic activities. *Bioorganic & medicinal chemistry letters*, *19*(19), 5565-5568. https://doi.org/10.1016/j.bmcl.2009.08.046

Suleiman, M. M., Isaev, S. G., Klenina, O. V., & Ogurtsov, V. V. (2014). Synthesis, biological activity evaluation and QSAR studies of novel 3-(aminooxalyl-amino)-and 3-(carbamoylpropionylamino)-2-phenylamino-benzoic acid derivatives. *J. Chem. Pharm. Res*, *6*(5), 1219-1235

Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., ... & Prokopenko, V. V. (2005). Virtual computational chemistry laboratory–design and description. Journal of computer-aided molecular design, 19, 453-463. https://doi.org/10.1007/s10822-005-8694-y

Todeschini, R., & Gramatica, P. (2002). New 3D Molecular Descriptors: The WHIM theory and QSAR Applications. *Perspectives in Drug Discovery and Design, 2*, 355-380. 10.1007/0-306-46857-3_19

Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69-77. https://doi.org/10.1002/QSAR.200390007

Vieira, A. J. S. C., & Santos, P. M. P. (2017). A Tentative Classification of Antioxidants: Which Role They Play when Protecting Biological Targets from Oxidative Stress Induced Damage. *J Med Chem Drug Des*, *1*(1). http://dx.doi.org/10.16966/2578-9589.101

Weaver, S., & Gleeson, M. P. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics and Modelling*, *26*(8), 1315-1326. https://doi.org/10.1016/j.jmgm.2008.01.002