Molecular Docking and QSAR Studies for Modeling the Inhibitory Activity of Pyrazole-benzimidazolone Hybrids as Novel Inhibitors of Human 4-hydroxyphenylpyruvate dioxygenase Against Type I Tyrosinemia Disease

Nidal Naceiri Mrabti [1], **Hanae Naceiri Mrabti** [2] , **Er-Rajy Mohammed** [1], **Khalid Dguigui** [1], **Latifa Doudach** [3], **Zineb Khalil** [4], **Abdelhakim Bouyahya** [5,*] , **Gokhan Zengin** [6,*], **Menana Elhallaoui**

[1] Engineering Materials, Modeling and Environmental Laboratory, Faculty of Science, University Sidi Mohammed Ben Abdellah, Dhar Mehraz, B.P. 1796, Atlas, Fes, Morocco
[2] Laboratory of Pharmacology and Toxicology, Bio-Pharmaceutical and Toxicological Analysis Research Team, Faculty of Medicine and Pharmacy, Mohammed V University in Rabat, BP 6203, Rabat, Morocco
[3] Biomedical Engineering Department, National School of Arts and Crafts Rabat (ENSAM) Mohammed V University in Rabat, BP 6203 Rabat, Morocco
[4] Laboratory of Medicinal Chemistry, Drug Sciences Research Center, Faculty of Medicine and Pharmacy, Mohammed V University in Rabat, BP 6203, Rabat, Morocco
[5] Laboratory of Human Pathologies Biology, Department of Biology, Faculty of Sciences, Mohammed V University, Rabat, Morocco
[6] Department of Biology, Science Faculty, Selcuk University, 42130 Konya, Turkey
[7] Department of Chemistry, Faculty of Science, University Sidi Mohammed Ben Abdellah, Dhar Mehraz, B.P. 1796, Atlas, Fes, Morocco
* Correspondence: boyahyaa-90@hotmail.com (A.B.); biyologzengin@gmail.com (G.Z.);

Scopus Author ID 57190813643

Abstract: Pyrazole-benzimidazolone is a novel inhibitor of human 4-Hydroxyphenylpyruvate dioxygenase of the HPPD receptor. To analyze this inhibitory activity, a QSAR study was performed using 33 compounds. The MLR multiple linear regression method and the NN neural network were used to establish the predicted model. This model was validated by the external validation method. The molecular electronic descriptors were calculated by the DFT method. This study shows that the activity of the compounds is strongly correlated with the five descriptors selected by the MLR method. The correlation coefficients calculated by MLR and then by NN are $R = 0.878$ and $R = 0.978$; they allow us to evaluate the proposed quantitative model. The latter allowed us to predict the activity of new HPPD receptor inhibitors and to show the good predictive competence of the established QSAR model. Based on these results, we performed docking in the active site of HPPD protein with the most active compound (12), and we deduced that this compound specifically interacts with are GLN 307, ASN 423, and PHE 392 at the HPPD protein binding pocket; thus, it showed additional interactions with VAL 228, PRO 280, SER 267, LEU 265, LYS 421, GLN 379, GLY 420, and LEU 368, This interaction is similar to that of NTBC with the active site.

Keywords: QSAR; MLR; neural network; cross validation; ADMET; molecular docking.

1. Introduction

Type I tyrosinemia is an autosomal recessive disease [1]; it manifests itself in the liver and kidneys with a series of symptoms that may vary in clinical severity [2]. The symptoms

include hepatomegaly, severe neurological attacks, and elevated tyrosine and methionine concentrations in plasma and urine. Type I tyrosinemia is linked to a deficiency of fumarylacetoacetate hydrolase (FAH) [3], the enzyme involved in the final phase of tyrosine catabolism [4]. The absence of FHA activity results in the accumulation of tyrosine metabolites, particularly fumaryl acetoacetate and maleyl acetoacetate, which, after reduction and decarboxylation, are converted to succinylacetone [5]. The presence of succinylacetone in the urine is one way to diagnose the disease. Since the availability of NTBC [6], liver transplantation is not the only treatment for the disease [7]. NTBC blocks tyrosine catabolism by inhibiting 4-hydroxyphenylpyruvate dioxygenase (HPD) [8,9], thereby delaying the progression of liver degeneration [10] (Figure 1). Therefore, the only two sources of tyrosine are protein-rich foods (milk, eggs, meat, nuts, and avocado, etc.,) and the conversion of phenylalanine through the action of phenylalanine hydroxylase [11], NTBC binds closely to 4-hydroxyphenylpyruvate dioxygenase, The half-life of NTBC in humans would be 54 h, in healthy adult men [12], for this reason, the researchers thought to synthesize drugs, based on the (NTCB) as a powerful reference against Tyrosimia 1 [13,14].

Benzimidazole compounds are well-known pharmacophores with diverse biological activities and are promising candidates for the future development of drugs for clinical use such as Astemizole, Omeprazole, Lansoprazole, Pantoprazole, etc., [15]. This is mainly due to the ability of this class of molecules to target and interact with different sites effectively. In addition, other members of this class have also been reported to have antiviral, antiparasitic, antifungal, antioxidant, antituberculosis, antifertility, antimicrobial, anticancer, and antihistamine activities [16].

The major problem for people with type 1 tyrosinemia is the shortage of effective medicines for this disease, as it is infrequent that about 1 in 2 million people are affected by the disease each year [17]. Therefore, investing in this type of disease would be very costly, and researchers are focusing in particular on diseases that are heavily ravaged.

QSAR studies have become a pillar of modern chemistry. In QSAR analysis, the main objective is to develop statistical models capable of predicting the biological activity of new compounds not yet tested. For this reason, we used the QSAR method to gain on two sides, such as optimizing the time and cost of searching for new drugs.

Biological activity is closely correlated with regions of space, and to identify these, and we calculated multidimensional molecular descriptors (Table 1). QSAR models can be generated using statistical methods. This study required the use of 33 benzimidazole derivatives, which Yu-Ling Xu et al. [18] evaluated as HPPD inhibitors, hence their usefulness in QSAR model construction [19,20].

To do this, we used multiple linear regression (MLR) analysis and artificial neural networks (ANN). The predictive capacity of the established model was tested by external validation based on the criteria of Golbraikh and Tropsha and Y-randomization methods [21]. In addition, we performed molecular docking of two compounds, 12 and NTBC, with HPPD. Finally, we also predicted molecules with very strong activity against this disease.
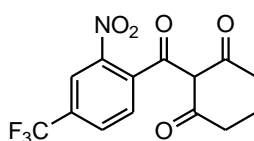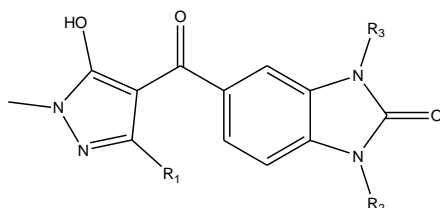

Figure 1. Chemical structure of NTBC.

2. Materials and Methods
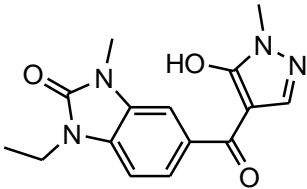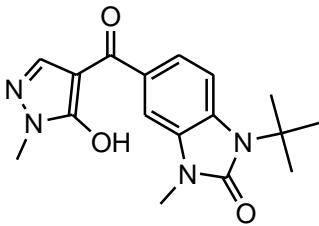
2.1. Experimental data.
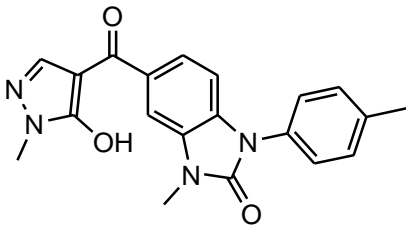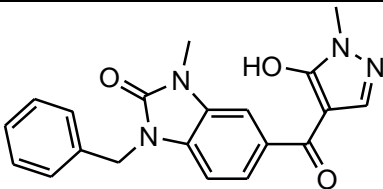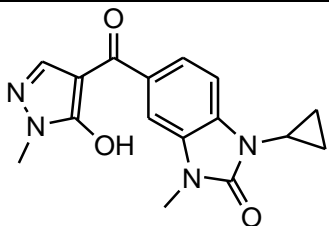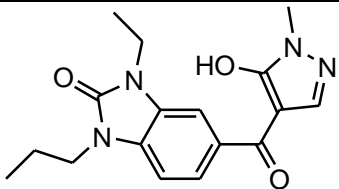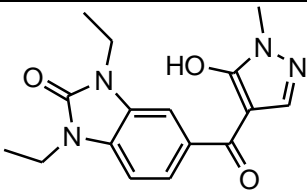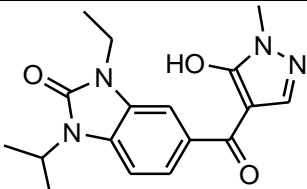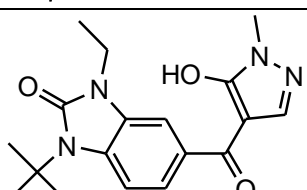
The study involved the use of 33 benzimidazol derivatives (Figure 1); table 1 shows the activity observed for each compound on a logarithmic scale.
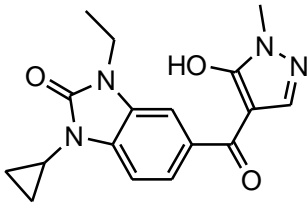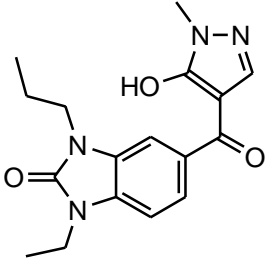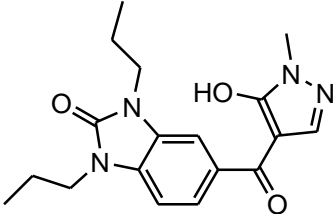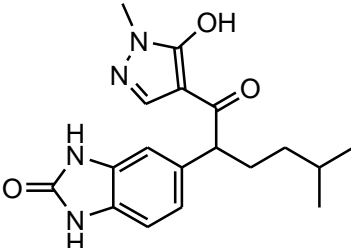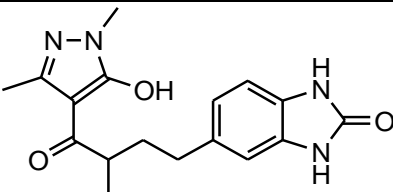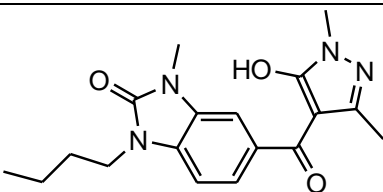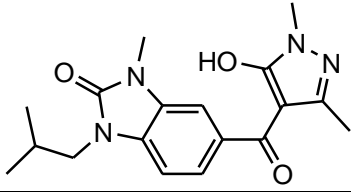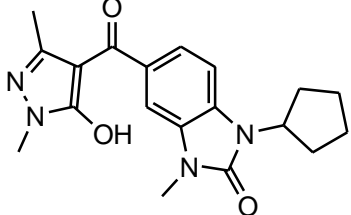
Figure 2. Basic molecule of benzimidazole derivative.
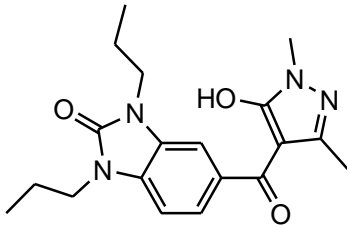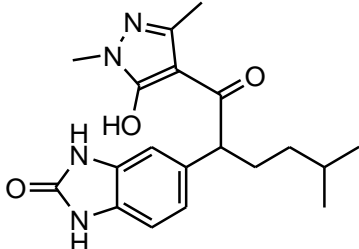
Table 1. Chemical structures of benzimidazole compounds and their biological activity against recombinant.

Compound	Structure	Activity
NTBC		7,1674
Compound 1		7,2291
Compound 2		6,7695
Compound 3		6,8538
Compound 4		6,8860
Compound 5		7,2146

Compound	Structure	Activity
Compound 6		7,6575
Compound 7		7,0000
Compound 8		6,6020
Compound 9		6,8860
Compound 10		7,0000
Compound 11		7,6575
Compound 12		7,6777
Compound 13		7,1938
Compound 14		6,8239

Compound	Structure	Activity
Compound 15		6,6575
Compound 16		7,4089
Compound 17		7,3979
Compound 18		7,1135
Compound 19		5,9871
Compound 20		6,1938
Compound 21		5,5670
Compound 22		6,0315

Compound	Structure	Activity
Compound 23		6,0268
Compound 24		6,0132
Compound 25		6,0043
Compound 26		6,0043
Compound 27		6,1739
Compound 28		6,0555
Compound 29		6,0268
Compound 30		6,0132
Compound 31		6,1307

Compound	Structure	Activity
Compound 32		6,1426
Compound 33		5,9244

2.2. Molecular descriptor selection.

In the interest of accurately modeling and predicting inhibitor activity, 16 descriptors listed in Table 1 have been introduced. The lipophilic, geometric, physicochemical, and steric descriptors were calculated with the MM2 method using the ACD/ChemSketch program [22] [23] and ChemBioOffice [24]. While the electronic descriptors were calculated using the DFT method [25], using the Gaussian03 quantum chemistry software [26]. The optimization of the compounds was performed by the DFT method using Becke's three-parameter hybrid function (B3LYP) [27] with a 6-31G base to calculate electronic descriptors [28]. All of the descriptors used in this study are presented in Table 2.

Table 2. The calculated descriptors used in this study.

Category of descriptors		
	Quantum chemical descriptors	**Steric, thermodynamic, and topologic chemical**
Name of the descriptors	Repulsion Energy (RE)	Cluster Count (Cls)
	Total dipole moment (μ)	Molecular Weight(MW)
	Electron affinity (EA)	Cluster Count (ClsC)
	Ionization potential (IP)	Wiener Index (Windx)
	Lowest Unoccupied Molecular Orbital (E_{LUM})	Parachor (Par)
	Highest occupied molecular orbital energy(E_{HOMO})	Balban Index (BIndx)
	Energy gap ($E_{HOMO}-E_{LUMO}$)	Molecular Topological Index (TIndx)
	Mulliken Electronegativity (χ)	Lypophilie (LogP)
	Hardness (η)	Molecular Refraction(RM)
	Electrophilicity (ω)	Henry's law constant (H)
	Softness (S)	Heat of Formation (HF)
	Electrophilicity index (ω_i)	VDW Energy (E_{VDW})
		Torsion Energy(Et)

The Pearson correlation matrix was used to select the 25 most relevant descriptors influencing activity, eliminate descriptors with low correlation with activity, and randomly select among descriptors with high correlation with each other and activity.

2.3. Statistical methods.

A multiple linear regression (MLR) analysis [29,30] with the descending selection method was used to select the most appropriate descriptors. This mathematical technique studies the relationship between a dependent variable (Activity) and several independent variables (Descriptors). This method reduces the divergence between experimental and predicted values of a given set of biological activities.

The regression method focuses on three criteria: the determination correlation (R^2), the value of the Fisher ratio (F), and the root mean square error (RMSE). The MLR model has been generated using XLSTAT [31]. Confirming the model's reliability required an external validation that Golbraikh and Tropsha proposed [32].

In this work, we selected a set of 33 compounds mentioned above. The complete set was randomly divided into two subsets, a training set (27 compounds) to build the model and a test set (6 compounds) to evaluate the reliability of the established model. Note that MLR was used to select the descriptors used as input parameters in the artificial neural network (ANN) [33,34], ANN analysis is performed using SAS JMP (v8,0, SAS Institute Inc, Cary, NC, USA). The neural networks are organized into three layers: the input layer contains six neurons representing the relevant descriptors obtained with the MLR technique, the output layer contains one neuron representing the activity values calculated log IC_{50}, and the hidden layer is composed of 3 neurons determined by ρ=(number of weights)/(number of connections).

2.4. Docking molecular.

Molecular docking often studies ligand-protein interactions, which focus on the relationship between the ligand and the amino acids forming the active site responsible for the biological activity [35].

The result of molecular docking is in the form of several conformations of different energy levels, hence choosing the weakest [36]. Molecular docking of the selected compounds as ligands and the target protein was performed in AUTODOCK software [37].

For the preparation of the protein, the 3D crystal structure of the protein (PDB 6J36) [38] with a resolution of 2,6 Å and 432 amino acids was downloaded from the protein database. The pretreatment of the protein was carried out by adding the missing hydrogens, the water molecules above 5 Å of the hetero group were removed, the optimization of the protonated residues was carried out at a pH of 7,0.

Using the AUTOGRID algorithm, and to evaluate the interaction energy between the ligand and the protein, the 3D grid was produced to simplify the docking analysis [39]. The grids were constructed using (x,y,z)=(60,60,60) with a grid point spacing of 0.375A. The center grid is 29.39A, 5.56A, and 52.49A, depending on the location of the ligand in the complex.

3. Results and Discussion

In our work, the set of molecules was divided into two subsets, one of which includes 27 random active molecules; this is the training set, while the second test set includes six compounds. The values of the selected descriptors and the predicted values of this activity obtained by MLR and ANN methods are presented in Table 3.

Table 3. The values of selected descriptors and observed/predicted activity (logIC50).

Compound	Dp	Stretch-Bend	LogP	Ovality	PIC50	MLR	ANN	CV(LOO)
compound 1	3,1830	0,2279	1,120	1,5131	7,2291	7,1928	7,1869	7,2029
compound 2	3,2197	0,2495	1,537	1,5392	6,7695	7,1629	6,9417	6,958
compound 3	3,8697	0,1892	1,518	1,5060	6,8538	7,0916	6,9567	6,9674
compound 5	3,0806	0,2557	0,952	1,4859	7,2146	7,0259	7,1898	7,1434
compound 6	4,2356	0,2009	0,633	1,4862	7,6575	7,896	7,6235	7,7257
compound 7	3,0745	0,4896	1,170	1,4789	7,0000	7,1194	7,0427	7,054
compound 8	3,3268	0,1436	2,446	1,5723	6,6020	6,6077	6,5331	6,5809
compound 9	3,9243	0,1075	2,028	1,5510	6,8860	7,0347	6,8323	6,9177
compound 11	3,0751	0,3470	1,458	1,5263	7,6575	7,1619	7,4875	7,4356
compound 12	3,0910	0,3188	0,972	1,4996	7,6777	7,2663	7,6283	7,5241

Compound	Dp	Stretch-Bend	LogP	Ovality	PIC50	MLR	ANN	CV(LOO)
compound 13	2,9638	0,3712	1,290	1,5002	7,1938	6,9964	7,2032	7,1311
compound 14	2,9464	0,6163	1,508	1,4942	6,8239	7,1125	6,8328	6,9231
compound 16	3,0761	0,3453	1,458	1,5263	7,4089	7,1594	7,4807	7,3497
compound 17	3,0946	0,3739	1,944	1,5522	7,3979	7,066	7,3732	7,279
compound 18	4,4642	0,0929	1,916	1,4688	7,1135	6,5113	7,1019	6,9089
compound 19	2,7768	0,0840	1,874	1,5322	5,9871	6,2646	6,0547	6,1021
compound 20	2,9579	0,2781	2,241	1,5264	6,1938	6,2821	6,0478	6,1746
compound 21	3,2706	0,2166	2,223	1,4942	5,5670	6,0237	5,5017	5,6975
compound 23	2,7911	0,2849	1,656	1,4731	6,0268	6,1304	5,9947	6,0506
compound 24	2,9587	0,2287	1,338	1,4716	6,0132	6,4075	6,0575	6,1594
compound 25	2,5311	0,5196	1,874	1,4680	6,0043	6,0995	6,0593	6,0544
compound 26	3,2395	0,1697	3,150	1,5593	6,0043	5,8218	6,1001	5,9754
compound 27	3,2300	0,3791	2,732	1,5391	6,1739	6,2989	6,0841	6,1856
compound 28	3,1288	0,3739	2,162	1,5138	6,0555	6,4641	6,2287	6,2494
compound 30	1,9779	0,3973	1,994	1,4885	6,0132	5,7063	5,9904	5,9033
compound 31	2,1918	0,3709	2,162	1,5138	6,1307	5,9183	6,1284	6,0591
compound 33	4,1878	0,1771	2,621	1,4604	5,9244	5,7581	5,9181	5,8669

3.1. Multiple linear regression.

The antityrosemic activity is explained by the electronic structure of benzimidazole. The model QSAR (Eq1) contains four descriptors determined from the MLR method; figure 2 shows the normalization coefficients of each of them, while the correlation between the predicted activity resulting from the model and the observed one is well illustrated in figure 3. Equation 1:

pIC_{50} = -10, 79969+0, 57740*Dp+1, 58336*Stretch-Bend-0, 89854*LogP+11, 10284*Ovality

N= 27   R = 0, 85 R^2= 0, 71   F = 8, 84   RMSE = 0, 29

N: Number of compounds in the training set

R^2: Square of the regression correlation coefficient for the training and test sets
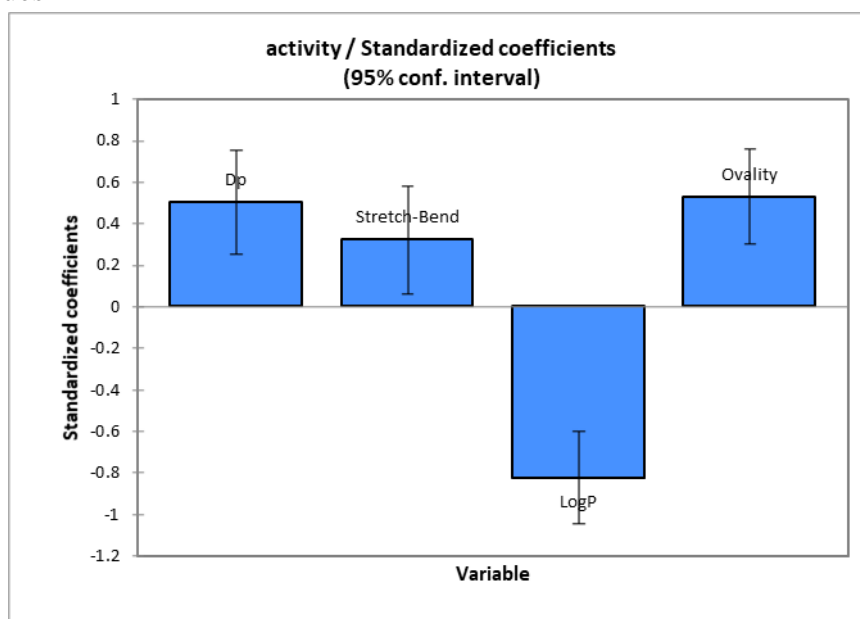
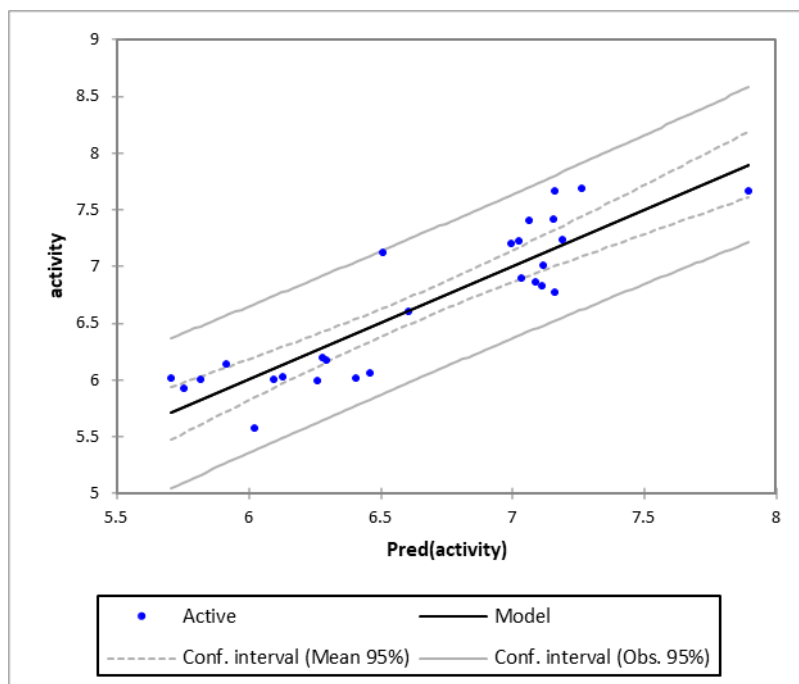S: Regression standard deviation

F : Fischer's test

p-value: Statistical confidence level

q^2: Square of the crossover-validated (LOO coefficient)

SPRESS: Standard deviation of the sum of squares of the difference between the predicted and observed values

Figure 3. Modeling characterization by the normalized coefficients.

Figure 4. The correlation between the observed and the predicted activities.

The model was constructed with 33 compounds, 6 used as a test set. Statistical analysis showed a strong correlation between activity and the five descriptors composing the model with high R2, q2, and F values and low values of s and SPRESS. In addition, this model passed all predictive ability tests, demonstrating that it is very robust and can provide reliable results (Table 3).

The most important parameter of the model is the ovality parameter. Thus, this descriptor is related to the molecular surface area and the minimum surface area corresponding to the van der Waals volume. The ovality index is unitary for spherical molecules and increases with the molecule's increasing linearity (elongated shape).

In this model, the ovalization has a positive sign, so a positive correlation is marked between this index and the pIC50; this is the case for the compounds with the largest values of ovalization (12, 11, and 6) that have the greatest activity.

The Stretch-Bend energy is a thermodynamic parameter that deals with the conformational flexibility of the molecule. This descriptor has a positive coefficient, indicating that substituents that increase the flexibility of benzimidazole derivatives will improve activity.

On the other hand, a negative correlation is recorded between activity and the third most important descriptor of the model, the Log P.

Lipophilia is affected by the structure of the molecule and the presence of functional groups, unsaturated bonds, and molecular weight. The importance of dipole moment in modulating activity against Type I tyrosinemia may be due to the presence of carbonyl group, where permanent polarization is seen due to electronegativity difference between the atoms. The carbonyl oxygen of substituted benzimidazoles may involve fruitful binding interactions with amino acids present at the target site through hydrogen bonding. The positive coefficient of Dp in the equation indicates a positive correlation between substituted benzimidazoles and the dipole moment.

This is evidenced by the activity data of the substituted benzimidazoles (Table 1) and their Dp values (Table 3). Compounds 6 and 18 with maximum Dp values of 2.31 and 1.09

(Table 3), respectively, have maximum activity (compound 6, PIC50=7.6575 and compound 18, PIC_{50}=7.1135, (Table 3).

Statistical analysis proved the significant effectiveness of the QSAR model (R = 0.84 and RMSE=0.23) hence the good prediction generated.

3.2. Artificial neural networks.

The neural network has a three-dimensional architecture; it contains three layers; the input layer is composed of 6 neurons, and the output layer contains one neuron, which is the activity. The third layer is called hidden. It is composed of several neurons (ρ) varying from 1 to 3; this interval is obtained from the consideration parameter ρ=(number of weight)/(number of connection) [40-42].

Using the descriptors obtained from MLR the statistical analysis showed an improvement in the quality of the QSAR model (R^2=0.984 and RMSE=0.026) over the previous one (Figure 4 and 5), hence the relevance of NN.
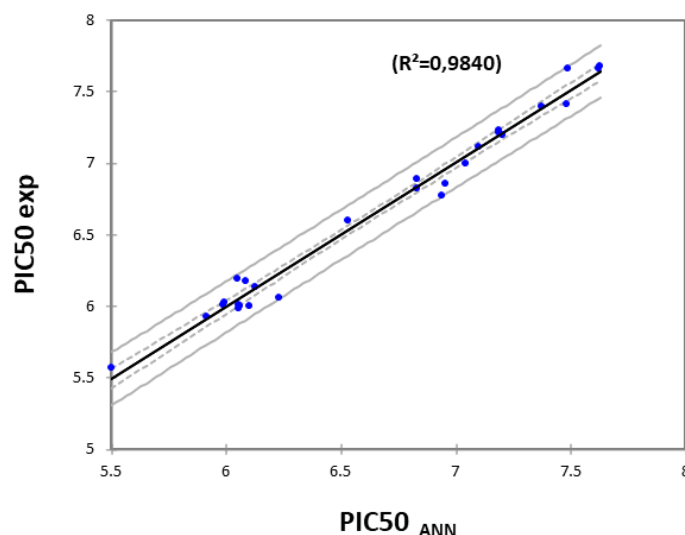
Figure 5. The correlation between the observed and the predicted activities established by ANN.

3.3. Cross-validation.

Cross-validation is a method to validate the performance of a prediction model [43].
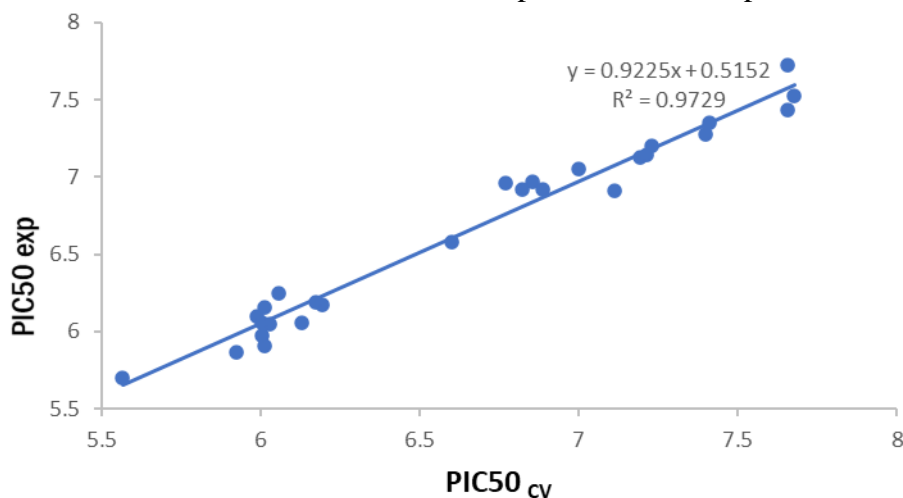
Figure 6. Correlation of observed and predicted activities calculated using Cross-validation.

In our case, the results generated by this method showed that the predicted model is stable and robust; this is reinforced by the statistical parameters (R^2=0.78 and RMSE=0.19). However, according to Golbraikh and Tropsha, the internal validation methods are insufficient to confirm the reliability of the constructed QSAR models. Therefore, external validations are hardly necessary to build a reliable QSAR model.

3.4. Extern validation.

The external validation of the predicted model was based on the criteria pre-established by Golbraikh and Tropsha (Table 4). This method consists of reconstructing the QSAR model based on 27 compounds only in contrast to the above, which included the whole molecules. Note that the 6 excluded compounds are reserved for the external test. The new model and the results obtained are represented by equation 2 and table 5.

pIC50 = -10,04789+0,60897*Dp+1,57224*Stretch-Bend 0,88345*LogP+10,98987*Ovality

N =27   R =0,83 R^2 =0,68   F = 8,14   RMSE =0,27

From the results obtained (Table 5), the predicted QSAR model is in perfect agreement with the validation methods used, and thus that the experimental activity could be accurately predicted using the established QSAR model.

Table 4. Values obtained after external validation by MLR, NN, and CV methods.

Compounds	PIC_{50}	Pred PIC_{50}	Compounds	PIC_{50}	Pred PIC_{50}	Compounds	PIC_{50}	Pred PIC_{50}
1	7,8885	7,2291	11	7,8565	7,6575	4	7,1371	6,8860
2	7,8628	6,7695	12	7,9574	7,6777	15	7,4859	6,6575
3	7,8162	6,8538	13	7,6876	7,1938	22	6,9438	6,0315
5	7,7186	7,2146	14	7,8044	6,8239	29	7,7047	6,0268
6	8,6209	7,6575	16	7,8540	7,4089	32	8,0198	6,1426
7	7,8134	7,0000	17	7,7653	7,3979	28	7,1721	6,0555
8	7,3222	6,6020	18	7,2660	7,1135	30	6,3780	6,0132
9	7,7646	6,8860	19	6,9583	5,9871	31	6,5968	6,1307
24	7,1041	6,0132	20	6,9855	6,1938	10	8,3961	7,0000
25	6,7878	6,0043	21	6,7410	5,5670	33	6,5147	5,9244
26	6,5453	6,0043	23	6,8257	6,0268	27	7,0158	6,1739

Table 5. Golbraikh and Tropsha criteria.

|	Formula	Threshold	Modelscore
$R^2_{ANN\,ext}$	$R^2_{ext} = 1 - \dfrac{\sum(Y_{pred(test)} - Y_{(test)})^2}{\sum(Y_{(test)} - \bar{Y}_{tr})^2}$	$R^2_{ext}>0,6$	
r^2	Coefficient of determination for the plot of predicted versus observed for test set by MLR	$r^2>0,6$	0,69
r_0^2	r^2 at zero intercept		0,62
$r_0'^2$	r2 for the plot of observed versus predicted activity for the test set at zero intercept		0,64
$|r_0^2 - r_0'^2|$		$|r_0^2 - r_0'^2| < 0,3$	0,02
k	Slope of the plot of predicted versus observed activity for test set at zero intercept	0,85<k<1,15	0,996
$\dfrac{r^2 - r_0^2}{r^2}$		$\dfrac{r^2 - r_0^2}{r^2} < 0,1$	0,1
k'	Slope of the plot of observed versus predicted activity at zero intercept	0,85<k'<1,15	0,999
$\dfrac{r^2 - r_0'^2}{r^2}$		$\dfrac{r^2 - r_0'^2}{r^2} < 0,1$	0,07

3.5. Molecular docking study.

The NTBC compound interacted with the active site by generating several hydrophobic bonds with a series of amino acids (Figure 7). The vdw bonds are the dominant ones; they characterize the majority of the amino acids citing VAL 228, VAL 269 PRO 280 ASN 282

SER 267 LEU 265 LYS 421 GLN 293 GLN 307 GLN 379 ALA 310 PHE 319 GLY 420 LEU 368 ASN 423, LEU 427, PHE 392 and PHE 381 constructed alkyl pi bonds, while a Pi-Pi STACKET bond was generated between PHE 424 and the compound. On the other hand, we note that the amino acid MET 335 has built two bonds, one Pi sulfur with its sulfur atom and another pi alkyl. Finally, three metal acceptor-type bonds were made between the Cobalt atom existing at the NTBC and the three amino acids; HIS 308, HIS 226, and GLU 394.
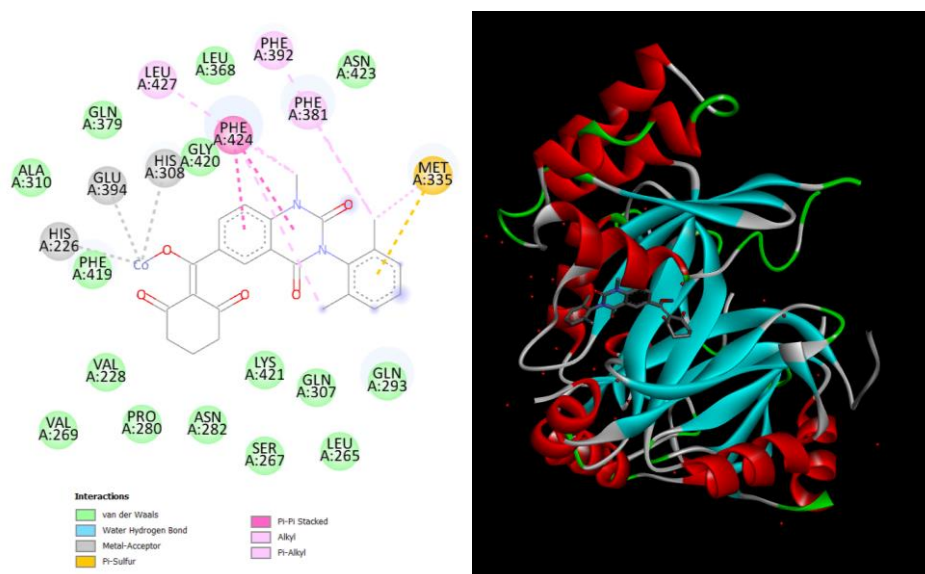
Figure 7. 2D and 3D docking poses showing interactions of NTBC in the binding sites HPPD (binding energy 10.3 kcal/mol).
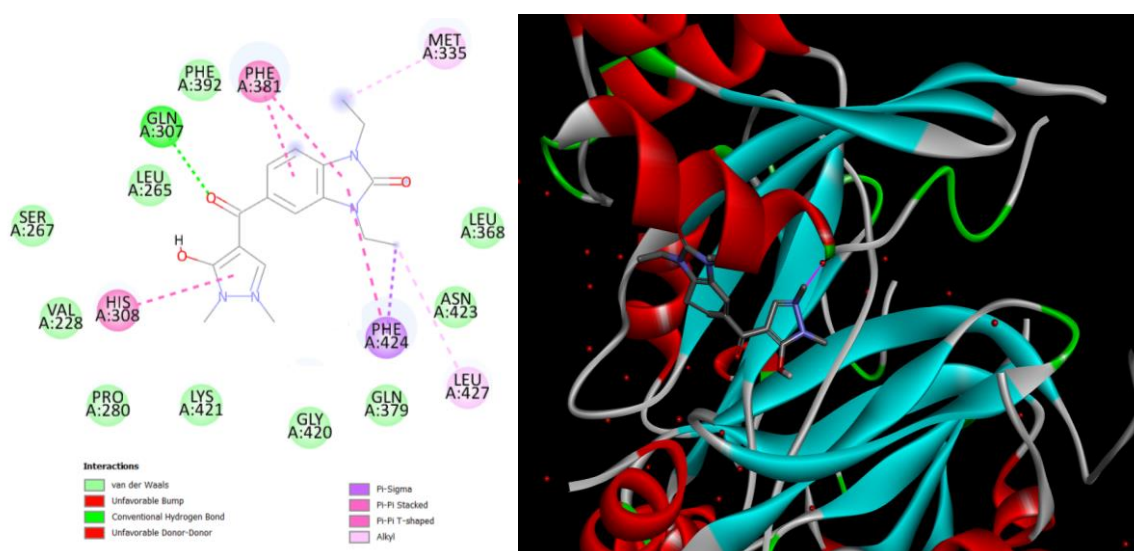
Figure 8. 2D and 3D docking poses showing interactions of compounds 12 in the binding sites (binding energy 10.6 kcal/mol).

Comparing our compound 12 with NTBC, we find that 60% of the amino acids that showed VDW binding with NTBC compound kept the same binding with our compound; these are VAL 228, PRO 280 SER 267 LEU 265 LYS 421 GLN 379 GLY 420, and LEU 368 (Figure 8). On the other hand, new amino acids have interacted with compound 12 by making the same type of linkage: ASN 423 and PHE 392. Note that the latter changed its type of binding from PI alkyl to VDW, PHE 424 kept its Pi-STACKED binding with our compound 12, showing a new PI sigma interaction. Concerning MET 335, the two bonds made with NTBC were gone in our case, and an alkyl bond was generated.

The amino acid GLN 307 changed its bond from VDW with NTBC to a conventional hydrogen bond with compound 12; this last interaction was made with oxygen at pyrazole. The results showed that most of the amino acids at the active site persist during the interaction with compound 12 with more or less similar bonds, yet the latter remains economically cheaper.

4. Conclusions

To produce new drugs that inhibit type 1 HPPD, a QSAR model was developed using *in vitro* data reported by a series of novel pyrazole-benzimidazolone hybrids. These compounds were designed and evaluated as potent inhibitors of human HPPD. Most of the compounds showed significant inhibitory activity against the recombinant human protein. Some showed higher activity than NTBC, particularly compound 12, characterized by a pyrazole moiety. Pyrazole, linked to a substituted benzimidazolone, was identified as the most potent inhibitor of human HPPD with an IC_{50} value of 21 nM; it is 3 times more potent than NTBC. Molecular modeling based on QSAR and molecular docking proved the efficacy of benzimidazolone derivatives, including compound 12, it showed a strong interaction with the majority of amino acids existing at the active site of the protein, yet compound 12 remains economically cheaper, hence the possibility of its use as an effective drug against tyrosinemia type 1 disease.

Funding

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kvittingen EA. - Abstract - Europe PMC https://europepmc.org/article/med/3296130 (accessed Aug 10, **2021**).
2. Morrow, G.; Tanguay, R.M. Biochemical and Clinical Aspects of Hereditary Tyrosinemia Type 1. *Advances in experimental medicine and biology* **2017**, *959*, 9-21, https://doi.org/10.1007/978-3-319-55780-9_2.
3. Li, N.; Gou, S.; Wang, J.; Zhang, Q.; Huang, X.; Xie, J.; Li, L.; Jin, Q.; Ouyang, Z.; Chen, F.; Ge, W.; Shi, H.; Liang, Y.; Zhuang, Z.; Zhao, X.; Lian, M.; Ye, Y.; Quan, L.; Wu, H.; Lai, L.; Wang, K. CRISPR/Cas9-Mediated Gene Correction in Newborn Rabbits with Hereditary Tyrosinemia Type I. *Molecular Therapy* **2021**, *29*, 1001-1015, https://doi.org/10.1016/j.ymthe.2020.11.023.
4. Lindblad, B.; Lindstedt, S.; Steen, G. On the enzymic defects in hereditary tyrosinemia. *Proceedings of the National Academy of Sciences* **1977**, *74*, 4641–4645, https://doi.org/10.1073/pnas.74.10.4641.
5. Sassa, S.; Kappas, A. Hereditary Tyrosinemia and the Heme Biosynthetic Pathway. Profound Inhibition Of δ-Aminolevulinic Acid Dehydratase Activity By Succinylacetone. *The Journal of Clinical Investigation* **1983**, *71*, 625-634, https://doi.org/10.1172/jci110809.
6. Menon, J.; Shanmugam, N.; Valamparampil, J.J.; Hakeem, A.; Vij, M.; Jalan, A.; Reddy, M.S.; Rela, M. Liver Transplantation: A Safe and Definitive Alternative to Lifelong Nitisinone for Tyrosinemia Type 1. *Indian Journal of Pediatrics* **2021**, https://doi.org/10.1007/s12098-021-03826-1.
7. Chinsky, J.M.; Singh, R.; Ficicioglu, C.; van Karnebeek, C.D.M.; Grompe, M.; Mitchell, G.; Waisbren, S.E.; Gucsavas-Calikoglu, M.; Wasserstein, M.P.; Coakley, K.; Scott, C.R. Diagnosis and treatment of tyrosinemia

type I: a US and Canadian consensus group review and recommendations. *Genetics in Medicine* **2017**, *19*, 1380-1380, https://doi.org/10.1038/gim.2017.101.

8. Wang, M.-M.; Huang, H.; Shu, L.; Liu, J.-M.; Zhang, J.-Q.; Yan, Y.-L.; Zhang, D.-Y. Synthesis and Herbicidal Activities of Aryloxyacetic Acid Derivatives as HPPD Inhibitors. *Beilstein J. Org. Chem.* **2020**, *16*, 233–247, https://doi.org/10.3762/bjoc.16.25.

9. Rudnick, D.A.; Ebach, D.R. Tyrosinemia. In: *Encyclopedia of Gastroenterology*. Johnson, L.R. Ed.; Elsevier: New York, **2004**; pp 538–541, https://doi.org/10.1016/B0-12-386860-2/00772-3.

10. Lin, Y.-L.; Wu, C.-S.; Lin, S.-W.; Yang, D.-Y. SAR studies of 2-o-substituted-benzoyl- and 2-alkanoyl-cyclohexane-1,3-diones as inhibitors of 4-hydroxyphenylpyruvate dioxygenase. *Bioorganic & Medicinal Chemistry Letters* **2000**, *10*, 843-845, https://doi.org/10.1016/s0960-894x(00)00115-3.

11. Matthews, D.E. An overview of phenylalanine and tyrosine kinetics in humans. *J Nutr* **2007**, *137*, 1549S-1575S, https://doi.org/10.1093/jn/137.6.1549S.

12. Cansever, M.Ş.; Aktuğlu-Zeybek, A.Ç.; Erim, F.B. Determination of NTBC in serum samples from patients with hereditary tyrosinemia type I by capillary electrophoresis. *Talanta* **2010**, *80*, 1846-1848, https://doi.org/10.1016/j.talanta.2009.10.032.

13. Lock, E.A.; Gaskin, P.; Ellis, M.; Provan, W.M.; Smith, L.L. Tyrosinemia produced by 2-(2-nitro-4-trifluoromethylbenzoyl)-cyclohexane-1,3-dione (NTBC) in experimental animals and its relationship to corneal injury. *Toxicology and Applied Pharmacology* **2006**, *215*, 9-16, https://doi.org/10.1016/j.taap.2006.01.015.

14. Kavana, M.; Moran, G.R. Interaction of (4-Hydroxyphenyl)pyruvate Dioxygenase with the Specific Inhibitor 2-[2-Nitro-4-(trifluoromethyl)benzoyl]-1,3-cyclohexanedione. *Biochemistry* **2003**, *42*, 10238-10245, https://doi.org/10.1021/bi034658b.

15. Gaba, M.; Mohan, C. Development of drugs based on imidazole and benzimidazole bioactive heterocycles: recent advances and future directions. *Medicinal Chemistry Research* **2016**, *25*, 173-210, https://doi.org/10.1007/s00044-015-1495-5.

16. Ates-Alagoz, Z. Antimicrobial Activities of 1-H-Benzimidazole-based Molecules. *Current Topics in Medicinal Chemistry* **2016**, *16*, 2953-2962, https://doi.org/10.2174/1568026616666160506130226.

17. van Ginkel, W.G.; Rodenburg, I.L.; Harding, C.O.; Hollak, C.E.M.; Heiner-Fokkema, M.R.; van Spronsen, F.J. Long-Term Outcomes and Practical Considerations in the Pharmacological Management of Tyrosinemia Type 1. *Pediatric Drugs* **2019**, *21*, 413-426, https://doi.org/10.1007/s40272-019-00364-4.

18. Xu, Y.-L.; Lin, H.-Y.; Ruan, X.; Yang, S.-G.; Hao, G.-F.; Yang, W.-C.; Yang, G.-F. Synthesis and bioevaluation of pyrazole-benzimidazolone hybrids as novel human 4-Hydroxyphenylpyruvate dioxygenase inhibitors. *European Journal of Medicinal Chemistry* **2015**, *92*, 427-438, https://doi.org/10.1016/j.ejmech.2015.01.018.

19. Golbraikh, A.; Wang, R.; Alves, V.M.; Liepina, I.; Muratov, E.; Tropsha, A. Dataset Modelability by QSAR: Continuous Response Variable. In: *Practical Aspects of Computational Chemistry V*. Leszczynski, J.; Shukla, M.K. Eds.; Springer International Publishing: Cham, **2022**; pp 233–253, https://doi.org/10.1007/978-3-030-83244-5_7.

20. Kharkar, P.S.; Reith, M.E.A.; Dutta, A.K. *Structure-activity relationship 3d-qsar: Topics by WorldWideScience.org.* (accessed Jan 25, **2020**).

21. Tropsha, A.; Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Current Pharmaceutical Design* **2007**, *13*, 3494-3504, https://doi.org/10.2174/138161207782794257.

22. Sevastos, A.A.; Baker, C.M.; Taylor, P. A simple method for predicting alkane-water partition coefficients of surfactants. *Journal of Surfactants and Detergents* **2021**, *n/a*, https://doi.org/10.1002/jsde.12545.

23. Österberg, T.; Norinder, U. Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors. *European Journal of Pharmaceutical Sciences* **2001**, *12*, 327-337, https://doi.org/10.1016/S0928-0987(00)00189-5.

24. ChemBioOffice Ultra 2010 Suite|Journal of the American Chemical Society https://pubs.acs.org/doi/full/10.1021/ja1005306 (accessed Aug 10, **2021**).

25. Yu, L.-J.; Dale, S.G.; Chan, B.; Karton, A. Benchmark study of DFT and composite methods for bond dissociation energies in argon compounds. *Chemical Physics* **2020**, *531*, https://doi.org/10.1016/j.chemphys.2019.110676.

26. Appendix 1 - MATLAB Code for Generating the COSMO File as Derived from GAUSSIAN03 Package. In: *Desulphurization and Denitrification of Diesel Oil Using Ionic Liquids*. Banerjee, T.; Ramalingam, A.; Eds.; Elsevier: Amsterdam, **2015**; pp 321–324, https://doi.org/10.1016/B978-0-12-801347-2.15001-3.

27. Guibedj, D.; Bougherara, H.; Kadri, M.; Akkari, H.; Berredjem, M.; Khaled, A. Methoxyphenyl N-sulfamoyloxazolidinone Cu(II) and Co(II) complexes: Synthesis, DFT/B3LYP (B2PLYP) study and molecular docking. *Journal of Molecular Structure* **2020**, *1202*, https://doi.org/10.1016/j.molstruc.2019.127190.

28. Mitro, S.K.; Hossain, K.M.; Majumder, R.; Hasan, M.Z. Effect of the negative chemical pressure on physical properties of doped perovskite molybdates in the framework of DFT method. *Journal of Alloys and Compounds* **2021**, *854*, https://doi.org/10.1016/j.jallcom.2020.157088.

29. Jozaghi, A.; Shen, H.; Ghazvinian, M.; Seo, D.-J.; Zhang, Y.; Welles, E.; Reed, S. Multi-model streamflow prediction using conditional bias-penalized multiple linear regression. *Stochastic Environmental Research and Risk Assessment* **2021**, *35*, 2355-2373, https://doi.org/10.1007/s00477-021-02048-3.

30. Abrougui, K.; Gabsi, K.; Mercatoris, B.; Khemis, C.; Amami, R.; Chehaibi, S. Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil and Tillage Research* **2019**, *190*, 202-208, https://doi.org/10.1016/j.still.2019.01.011.

31. Vidal, N.P.; Manful, C.F.; Pham, T.H.; Stewart, P.; Keough, D.; Thomas, R. The use of XLSTAT in conducting principal component analysis (PCA) when evaluating the relationships between sensory and quality attributes in grilled foods. *MethodsX* **2020**, *7*, https://doi.org/10.1016/j.mex.2020.100835.

32. [Rosen, M.; Betz, L.T.; Schultze-Lutter, F.; Chisholm, K.; Haidl, T.K.; Kambeitz-Ilankovic, L.; Bertolino, A.; Borgwardt, S.; Brambilla, P.; Lencer, R.; Meisenzahl, E.; Ruhrmann, S.; Salokangas, R.K.R.; Upthegrove, R.; Wood, S.J.; Koutsouleris, N.; Kambeitz, J. Towards clinical application of prediction models for transition to psychosis: A systematic review and external validation study in the PRONIA sample. *Neuroscience & Biobehavioral Reviews* **2021**, *125*, 478-492, https://doi.org/10.1016/j.neubiorev.2021.02.032.

33. Maleki, A.; Elahi, M.; Assad, M.E.H.; Alhuyi Nazari, M.; Safdari Shadloo, M.; Nabipour, N. Thermal conductivity modeling of nanofluids with ZnO particles by using approaches based on artificial neural network and MARS. *Journal of Thermal Analysis and Calorimetry* **2021**, *143*, 4261-4272, https://doi.org/10.1007/s10973-020-09373-9.

34. Gong, S.; Chen, J.; Jiang, C.; Xu, S.; He, F.; Wu, Z. Prediction of solitary wave attenuation by emergent vegetation using genetic programming and artificial neural networks. *Ocean Engineering* **2021**, *234*, https://doi.org/10.1016/j.oceaneng.2021.109250.

35. Shaghaghi, N. Molecular Docking Study of Novel COVID-19 Protease with Low Risk Terpenoides Compounds of Plants. *Biological and Medicinal Chemistry* **2020**. https://doi.org/10.26434/chemrxiv.11935722.v1.

36. Arshad, M.; Ahmed, K.; Bashir, M.; Kosar, N.; Kanwal, M.; Ahmed, M.; Khan, H.U.; Khan, S.; Rauf, A.; Waseem, A.; Mahmood, T. Synthesis, structural properties and potent bioactivities supported by molecular docking and DFT studies of new hydrazones derived from 5-chloroisatin and 2-thiophenecarboxaldehyde. *Journal of Molecular Structure* **2021**, *1246*, https://doi.org/10.1016/j.molstruc.2021.131204.

37. Sriramulu, D.K.; Lee, S.-G. Effect of Molecular Properties of the Protein-Ligand Complex on the Prediction Accuracy of AutoDock. *Journal of Molecular Graphics and Modelling* **2021**, *106*, https://doi.org/10.1016/j.jmgm.2021.107921.

38. Bank, R.P.D. RCSB PDB - 6J63: Crystal structure of Arabidopsis thaliana HPPD complexed with NTBC. http://doi.org/10.2210/pdb6J63/pdb.

39. Tacke, S.; Erdmann, P.; Wang, Z.; Klumpe, S.; Grange, M.; Plitzko, J.; Raunser, S. A streamlined workflow for automated cryo focused ion beam milling. *Journal of Structural Biology* **2021**, *213*, https://doi.org/10.1016/j.jsb.2021.107743.

40. Neidle, S. *Therapeutic Applications of Quadruplex Nucleic Acids*. Academic Press, **2012,** https://doi.org/10.1016/C2009-0-30510-3.

41. Mrabti, N.; Khalid, D.; Hadni, H.; Elhallaoui, M. QSAR study and molecular docking of benzimidazole derivatives inhibitors of p38 kinase. *Moroccan Journal of Chemistry* **2018**, *6*, 511-524.

42. Mrabti, N. N.; Elhallaoui, M. QSAR Study and Molecular Docking of Benzimidazole Derivatives as Potent Activators of AMP-Activated Protein Kinase. *J. Taibah Univ. Sci.* **2017**, *11*, 18–39, https://doi.org/10.1016/j.jtusci.2016.05.004.

43. Fong, E.; Holmes, C.C. On the marginal likelihood and cross-validation. *Biometrika* **2020**, *107*, 489-496, https://doi.org/10.1093/biomet/asz077.