

## Exploring QSARs of some Translocator protein (TSPO) ligands using MLR and PC-ANN techniques

Omar Deeb<sup>1</sup> and H. Baniowda<sup>1</sup>

<sup>1</sup>Faculty of Pharmacy, Al-Quds University, PO Box 20002, Jerusalem, Palestine

PO Box 20002, Jerusalem, Palestine

[deeb.omar@gmail.com](mailto:deeb.omar@gmail.com)

doi: <https://doi.org/10.37745/bjmas.2022.04903>

Published May 19, 2025

**Citation:** Deeb O., and Baniowda H. (2025) Exploring QSARs of some Translocator protein (TSPO) ligands using MLR and PC-ANN techniques, *British Journal of Multidisciplinary and Advanced Studies*,6(3)1-29

**Abstract:** *Quantitative structure–activity relationship study using principal component artificial neural network (PC-ANN) methodology was conducted to predict the inhibitory activities expressed as pIC<sub>50</sub> of 136 Translocator protein (TSPO) ligands. The results obtained by MLR was number of models, the best model was model number 24 which includes 24 descriptors, and resulted with R= 0.909, R<sup>2</sup>=.826, and R<sup>2</sup>adj.= 0.788. PCA performed to divide the data into three data sets, then the ANN performed on the chosen models (19-24) from LOO and LMO validation. The results show that model 24 has the highest correlation coefficient for the test set (0.85016) indicating its high predictive power which chosen to continue ANN to find the optimal number of hidden nodes, and in this case model 24 with 7 hidden nodes were chosen as the best model with the optimal hidden nodes. ANN resulted model were validated through randomization test, then the conditions proposed by Golbraikh and Tropsha were applied to conclude that the QSAR models have acceptable prediction power or not. However, the best ANN model with a good predictive power was model number 24.*

**Keywords:** QSAR, MLR, PC- ANN, inhibitory activity, translocator protein (TSPO) ligands.

## INTRODUCTION

Translocator protein (TSPO), was known as the peripheral benzodiazepine receptor (PBR). It is first identified in 1977 based on its distinct pharmacology with high affinity binding to benzodiazepines in peripheral tissues. The term “peripheral” was used to distinguish it from the plasma membrane “central” benzodiazepine receptor, a complex together with the  $\gamma$ -aminobutyric acid type A receptor that is important for inhibitory neurotransmission in the central nervous system. However, it became clear that its density in the brain regions can equal or exceed the density of central

benzodiazepine receptor (CBR) in the corresponding regions. TSPO is a protein of 18 kDa consisting of 169 amino acids, a five  $\alpha$ -helices composed of 21 hydrophobic residues. The N-terminus of the sequence is located in the mitochondrial domain, while the C-terminus is exposed to the cytoplasm. The transmembrane regions are connected by loops rich in hydrophilic residues. TSPO is strictly associated in a trimeric complex with the 32 kDa voltage dependent anion channel (VDAC), and 30 kDa adenine nucleotide translocase (ANT), thus forming the mitochondrial permeability transition pore (MPTP).

TSPO amino acid sequence shows conservation throughout evolution. TSPO in the photosynthetic bacteria *Rhodobacter sphaeroides* shows a 33.5% identity to human TSPO. Both human and mouse TSPO genes translate to a 169-amino acid protein with 81% sequence homology. Relatively the protein sequence of TSPO is conserved from bacteria to humans.

Expression of TSPO has been reported in different tissues including heart, brain, lung, spleen, testis, ovary, adrenal, kidney, bone marrow, salivary gland, adipose tissue, skin, and liver; and within these tissues, TSPO expression is regional and/or cell type specific. Also TSPO is expressed at low levels in other subcellular compartments such as plasma membranes and the nuclear fraction of cells [1].

Although research suggests that there exist multiple TSPO binding sites, the nature of these sites and their functional significance is poorly understood. Two ligands have been essential for characterizing the TSPO: the benzodiazepine Ro 5-4864 and the isoquinoline carboxamide PK11195, both of which are selective for the TSPO and display nanomolar binding affinity. Although these ligands exhibit saturable binding and reciprocal competition in radio ligand binding assays [2]. Furthermore, site-directed mutagenesis studies suggest certain residues in the first putative loop, TSPO are important for the binding of Ro 5-4864 but not PK11195. Thus, it is thought that PK11195 and Ro 5-4864 bind to heterogeneous sites at TSPO, either overlapping or allosterically coupled. Studies also describe PK11195 binding to multiple sites, which contradicts the initial finding that it bound to a single population of saturable sites. Scatchard analysis of 3HPK11195 binding to Ehrlich tumor cells revealed 2 independent binding sites [3].

Benzodiazepine Ro5-4864 and a nonbenzodiazepine PK11195 [an isoquinoline carboxamide derivative] were initially established as prototypical TSPO-binding chemicals, because they bind to TSPO but not to  $\gamma$ -aminobutyric acid type A receptor. Based on thermodynamic studies, and their opposing effects on neuronal seizures, PK11195 was classified as an antagonist and Ro5-4864 as an agonist. This pharmacology has been extensively used in attempts to elucidate the physiological relevance of TSPO. Although these studies did not readily reveal TSPO function, the ability of these chemicals in detecting TSPO with reasonable accuracy, and the pathological TSPO up-regulation seen at sites of inflammation led to the development

of TSPO as a diagnostic target. Radiolabeled forms of these chemicals that bind TSPO could be used to detect inflammatory lesions in vivo in a variety of human diseases using positron emission tomography. Clinical trials for different TSPO-binding agents focused on the diagnosis of various pathologies including traumatic brain injury, Alzheimer's disease, Parkinson's disease, multiple sclerosis, encephalopathy, autism, neuroinflammation, neurodegeneration, dementia, and neurocysticercosis (<https://www.clinicaltrials.gov>). Human clinical trials to detect cardiac sarcoidosis (NCT02017522), carotid atherosclerosis (NCT00547976), squamous and basal cell carcinomas (NCT01265472) remain an area of active research.

TSPO is said to be involved in a variety of biological processes including cholesterol transport, steroidogenesis, calcium homeostasis, lipid metabolism, mitochondrial oxidation, cell growth and differentiation, apoptosis induction, and regulation of immune functions [1].

Thus, TSPO can be exploited as a diagnostic marker to follow disease Progression and therapy efficacy by means of the biomedical imaging technique PET (positron emission tomography) but also as a therapeutic target. Although imaging complications have been encountered as a result of in vivo metabolism of these TSPO-binding PET tracers and aberrant signals contributing to nonspecific noise in some cases, new synthetic TSPO-binding chemicals are being developed to tackle these drawbacks. Therefore, diagnostic imaging is probably the primary clinical value that TSPO research has to offer at the present time [4].

This study aims to predict the inhibitory activity  $pIC_{50}$  of the data set in reference [5-9] as one group without splitting them into categories. This is achieved by applying ANN to develop new statistically validated QSAR models utilizing different types of descriptors. The strength and the predictive performance of the proposed models were verified using cross validation, chance correlation and external test set. Therefore, the motivation of this work is to provide QSAR models that will be used to predict inhibitory activity of unknown compounds and also these models may be used to design new drugs.

## **MATERIALS AND METHODS**

### **Software**

Geometry optimizations were performed using HyperChem (Version 7.5; Hypercube, Inc, USA, <http://www.hyper.com>) at the AM1 level of theory. An AM1 optimization was chosen because it was developed and parameterized for common organic structures. Descriptors were calculated using HyperChem and DRAGON (Milano Chemometrics and QSAR Group, USA, evaluation version 5.0, <http://www.disat.unimib.it/vhtml>) software. SPSS software (version 13.0, SPSS, Inc.)

was used for the simple MLR analysis while ANN analysis was performed using MATLAB (Version 7.0.1 (R14), <http://www.mathworks.com> ).

### **Chemical data and descriptors**

A data set of 136 Translocator protein (TSPO) ligands and their inhibitory activity ( $pIC_{50}$ ) were obtained from reference [5-9] and used in this study. They shared the same method of determination of ligands-Receptor activities using rat cortex membrane. The 136 ligands as eighteen chemical structure cores and their inhibitory activities are included in Table S1 in the supplementary materials.

The structures of the compounds are drawn by hyperchem software. The resultant structures are 2D then we convert them to 3D. HyperChem software was used to optimize the different compound structures using AM1 semi-empirical level. The optimization was preceded by the Polak-Rebiere algorithm. To be sure that we reached global minima, geometry optimization was run multiple times with different starting points for each molecule.

In this study, a pool of 1235 descriptors classified into 18 different groups was calculated using Dragon software. The constant or nearly constant descriptors for all the 136 compounds were discarded from further analysis. Furthermore, chemical descriptors such as HOMO, LUMO and polarizability were calculated using HyperChem software. Depending on the HOMO and LUMO values, electrophilicity, electronegativity, hardness, and softness descriptors were calculated. Other descriptors such as surface area approximate, surface area grid, volume, mass, polarizability, hydration energy, octanol-water partition coefficient ( $\log P$ ), and refractivity were calculated. Discarding highly inter-correlated ( $r > 0.95$ ) descriptors and following the procedure described in the next section, this number of descriptors was declined to 24 descriptors in the "final" MLR regression model (model 24 in Table 1).

### **Multiple linear regression (MLR) analysis**

Multiple linear regression analysis with stepwise selection and elimination of variables was employed to model the inhibitory activity ( $pIC_{50}$ ) relationships with each group of descriptors separately.  $\log 1/IC_{50}$  is the dependent variable and the set of descriptors as independent variables. Then, the "optimal" descriptors for each group were selected and gathered in one group to perform final MLR analysis.

### **Principal components analysis (PCA)**

Collinear descriptors add redundancy to the input data matrix and consequently the performances of the models obtained by using these descriptors would be degraded. PCA and more specifically factor analysis, groups together variables that are collinear to form a composite indicator capable of capturing as much of common information of those indicators as possible. Each factor reveals the set of variables with the highest

relationship. The idea under this approach is to explain the highest possible variation in the indicators set using the smallest possible number of factors. Consequently, the index no longer depends upon the dimensionality of the data set but it is rather based on the 'statistical' dimensions of the data. Application of PCA on a descriptor data matrix results in a loading matrix containing factors or PCs, which are orthogonal and therefore have no correlation with each other.

The PC's were calculated by singular value decomposition (SVD) method in MATLAB environment (MathWork Inc. Version 7.0.1 (R14)). Due to the quality of data, a previous treatment of the data is essential before applying the multivariate analysis methods. Scaling and centering is one of the pre-processing methods needed before performing the regression methods joint with feature extraction. Projection methods results depend on the normalization of the data. Descriptors with small absolute values have a small contribution to overall variances leading to biased PC's caused by the presence of other descriptors with higher values. In order to have the focus on the important variables in the model, equal weights are assigned to each descriptor, with appropriate scaling. Furthermore, descriptors were standardized to unit variance and zero mean (autoscaling) to give all variables the same importance. Then, the data matrix containing the entire set of descriptors and activity were simultaneously subjected to PCA.

### **Principal component-artificial neural network (PC-ANN) analysis**

ANNs are computer-based models in which a number of nodes, also called neurons are interconnected by links forming netlike structure "layers." A variable value is assigned to every neuron.

There are three kinds of neurons: (a) the input neurons which receive their values from independent variables and constitute the input layer, (b) the hidden neurons which collect values from other neurons, giving a result that is passed to a successor neuron, (c) the output neurons which take values from other units and correspond to different dependent variables, forming the output layer. In this sense, network architecture is commonly represented as I-H-O, where I, H, and O are the number of neurons in the input, hidden, and output layers, respectively.

The weights are links between units that condition the values assigned to the neurons. The weights are adjusted through a training process in order to minimize network error. For this, a non-linear transfer function relates the input parameters with the outputs. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output.

In PC-ANN analysis, as a preliminary treatment, the input data (i.e., molecular descriptors) were normalized to have zero mean and unity variance, and then were subjected to PCA before being introduced into the neural network. It should be illustrated that for each MLR resulted model, separate ANN models were developed so that the input's descriptors were the subsets selected by the stepwise MLR



## Results and Discussion

### MLR analysis

In continuation to recent QSAR studies [12-16] done using similar methods, we developed an ANN-QSAR model that describes the inhibitory activity of a series of compounds using large number of different descriptors. MLR were performed on each one of the groups of descriptors individually (individual approach described in Ref. [17] by Deeb) where  $pIC_{50}$  is the dependent variable. Stepwise method is used to develop multilinear equation by correlating dependent variable (activity) and the best independent variables.

Next, a new or “final” MLR analysis was performed by correlating the dependent variable (activity) and the optimal descriptors selected from the individual MLR models. Table 1 shows the regression models suggested from the “final” MLR analysis. The number of descriptors in these models is varied between 12 and 24. The highest coefficient of determination ( $R^2$ ) obtained, is 0.826 for a regression model with 24 descriptors (model **24**). Table 2 shows a key for the different descriptors used in the final MLR model.

**Table 1:** MLR Models resulted from all the groups of descriptors together

Model No.	No. of descriptors	R	$R^2$	$R^2_{adj}$	Selected descriptors
12	12	0.787	0.620	0.583	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4
13	13	0.803	0.645	0.607	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e
14	14	0.815	0.665	0.626	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3

15	15	0.828	0.686	0.647	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon
16	16	0.842	0.709	0.670	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003
17	17	0.851	0.725	0.685	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e
18	18	0.862	0.742	0.703	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4
19	19	0.875	0.765	0.727	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol)
20	20	0.886	0.784	0.747	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol), G(O..Cl)

21	21	0.893	0.797	0.760	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol), G(O..Cl), electrophilicity
22	22	0.897	0.805	0.767	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol), G(O..Cl), electrophilicity, Mor22e
23	23	0.902	0.813	0.775	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol), G(O..Cl), electrophilicity, Mor22e, Mor08m
24	24	0.909	0.826	0.788	JGI2, Mor10u, C-005, R8e+, nN-N, nR10, Mor19p, RDF035m, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, BELm3, nHDon, C-003, MATS4e, BEHm4, Hydration Energy (kcal/mol), G(O..Cl), electrophilicity, Mor22e, Mor08m, Mor11e

The following equation represents the best MLR model (model 24) with 24 descriptors:

$$\begin{aligned}
 \text{pIC50} = & 32.804 (\pm 5.902) - 8.386 (\pm 6.190) \text{ JGI2} - 0.159 (\pm 0.122) \text{ Mor10u} + 0.249 \\
 & (\pm 0.120) \text{ C-005} - 6.653 (\pm 3.382) \text{ R8e}^+ - 2.222 (\pm 0.258) \text{ nN-N} + 0.661 \\
 & (\pm 0.192) \text{ nR10} + 0.537 (\pm 0.191) \text{ Mor19p} + 0.106 (\pm 0.022) \text{ RDF035m} - 0.043 \\
 & (\pm 0.031) \text{ RDF030m} - 1.288 (\pm 0.348) \text{ nCONHRPh} - 205.791 (\pm 27.444) \text{ X4Av} \\
 & - 12.548 (\pm 1.485) \text{ BEHe4} - 36.414 (\pm 15.927) \text{ G1e} + 8.988 (\pm 1.817) \text{ BELm3} - \\
 & 0.106 (\pm 0.0947) \text{ nHDon} - 2.671 (\pm 0.382) \text{ C-003} - 4.090 (\pm 0.913) \text{ MATS4e} + \\
 & 5.928 (\pm 1.338) \text{ BEHm4} + 0.255 (\pm 0.057) \text{ Hydration Energy} + 0.032
 \end{aligned}$$

( $\pm 0.007$ ) **G(O..Cl)** - 1.366 ( $\pm 0.471$ ) **electrophilicity** - 0.525 ( $\pm 0.181$ ) **Mor22e** - 0.360 ( $\pm 0.102$ ) **Mor08m** - 0.288 ( $\pm 0.103$ ) **Mor11e**.

Where  $R=0.909$ ,  $R^2 = 0.826$ ,  $R^2_{adj} = 0.788$ , and the STD error of the estimate = 0.5700.

Based on the equation of the best MLR model, the following descriptors have a positive effect on the compounds activity:

C-005, nR10, Mor19p, RDF035m, BELm3, BEHm4, Hydration Energy, G(O..Cl).

While the below descriptors have a negative effect on the compounds activity,

JGI2, Mor10u, R8e<sup>+</sup>, nN-N, RDF030m, nCONHRPh, X4Av, BEHe4, G1e, nHDon, C-003, MATS4e, electrophilicity, Mor22e, Mor08m, Mor11e.

**Table 2:** Brief description of the descriptors in the best MLR model equation.

Name	Description	Block (group)
JGI2	Mean topological charge index of order 2	Galvez topol. Charge indices
Mor10u	Signal 10 / unweighted	3D-MoRSE descriptors
C-005	CH3X	Atom-centred fragments
R8e+	R maximal autocorrelation of lag 8 / weighted by Sanderson electronegativity	GETAWAY descriptors
nN-N	Number of N hydrazines	Functional group counts
nR10	Number of 10-membered rings	Ring descriptors
Mor19p	Signal 19 / weighted by polarizability	3D-MoRSE descriptors
RDF035m	Radial Distribution Function - 035 / weighted by mass	RDF descriptors
RDF030m	Radial Distribution Function - 030 / weighted by mass	RDF descriptors
nCONHRPh	Number of secondary amides (aromatic)	Functional group counts
X4Av	Average valence connectivity index of order 4	Connectivity indices
BEHe4	Highest eigenvalue n. 4 of Burden matrix / weighted by atomic Sanderson electronegativities	BCUT
G1e	1st component symmetry directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
BELm3	Lowest eigenvalue n. 3 of Burden matrix / weighted by atomic masses	BCUT

nHDon	Number of donor atoms for H-bonds (N and O)	Functional group counts
C-003	CHR3	Atom-centred fragments
MATS4e	Moran autocorrelation of lag 4 weighted by Sanderson electronegativity	2D autocorrelations
BEHm4	Highest eigenvalue n. 4 of Burden matrix / weighted by atomic masses	BCUT
Hydration Energy (kcal/mol)	Hydration Energy (kcal/mol)	G16-quantum-chemical
G(O..Cl)	Sum of geometrical distances between O..Cl	Geometrical descriptors
Electrophilicity	Electrophilicity	G16-quantum-chemical
Mor22e	Signal 22 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor08m	Signal 08 / weighted by mass	3D-MoRSE descriptors
Mor11e	Signal 11 / weighted by Sanderson electronegativity	3D-MoRSE descriptors

Then, leave one out (LOO) and leave many out (LMO) cross validation was performed on models **12-24** since these models have coefficients of determination larger than 0.6 [18]. The results of cross validation LOO and LMO are summarized in table S2 and S3 in the supplementary materials respectively. Table S2 and S3 show a good predictive power for models 19-24 because of having high  $R^2_{CV}$  and PRESS/SST less than 0.4. Thus, **models 19-24** were chosen for ANN analysis.

### PCA

The **PCA** was performed to divide the molecules into training, validation, and prediction (test) sets. Performing PCA on the whole data of 136 compounds, 24 descriptors and plotting the first and second principals, first and third principals, and second and third principals. The data division into 60% training, 20% test and 20% validation, should be in equal manner in which picking one compound from each zone to each set.

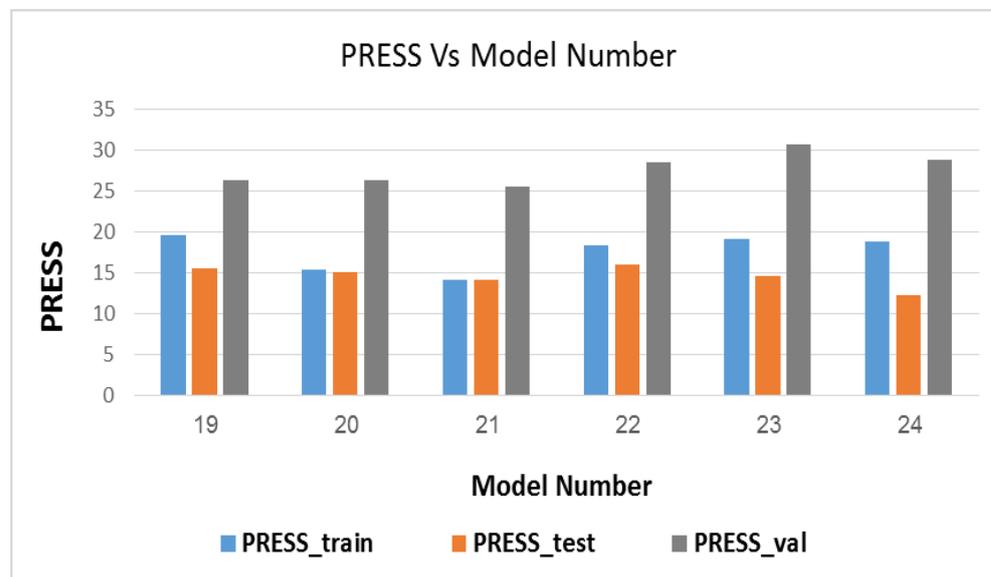
The first and second principals and first and third principals plots were having a condensed data towards the X axis, however second and third principals plot have the data distributed in a good way in comparison with the other plots. Therefore, relying on the second and third principals plot, it shows compounds **13**, **22** and **86** as outliers (Figure 1). Although these three compounds don't differ structurally in comparison with other compounds. But they behave in a different manner, therefore these compounds removed from the data in the next analysis. And so the data divided after

removing the outliers into 60% (81 compounds) training group, 20% (26 compounds) of each test and validation groups.

## ANN

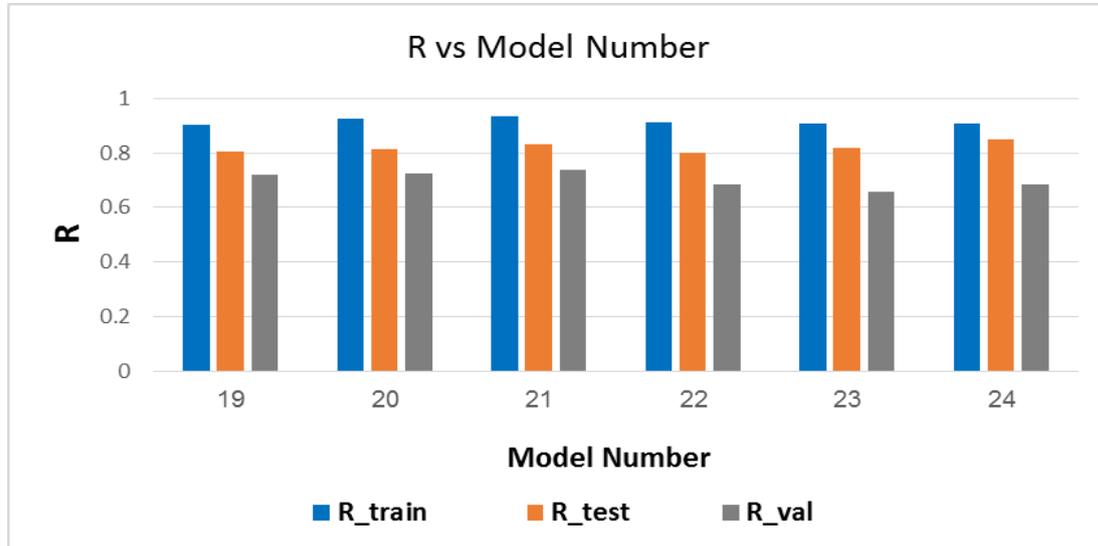
In this study, a three-layered feed-forward ANN model with back propagation learning algorithm [19] was employed. At first, non-linear relationship between the subset of descriptors selected by stepwise selection-based MLR and TSPO inhibitory activity was preceded by ANN models with similar structure. The number of hidden layer's nodes was set to 7 for all models, and the number of nodes in the input layer was the number of descriptors.

The results of first ANN is in table S4 in the supporting information, the table shows that model **24** has the highest correlation coefficient for the test set (0.85016) indicating its high predictive power and the one after it is model **21**. Figure 2 shows the relation of PRESS values for the training, test and validation sets versus model number. This figure shows that the minimum PRESS of the training set is obtained for model **21** the one after it is model **20**. While the minimum PRESS of the test sets is obtained for model **24** the one after it is model **21** then **23**.



**Figure 2:** Plots of ANN Predictive Residual Sum of Squares (PRESS) values for the training, test and validation sets versus model number.

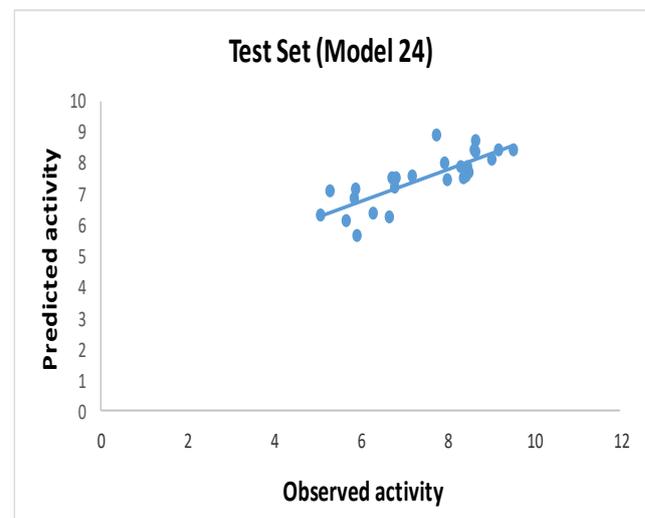
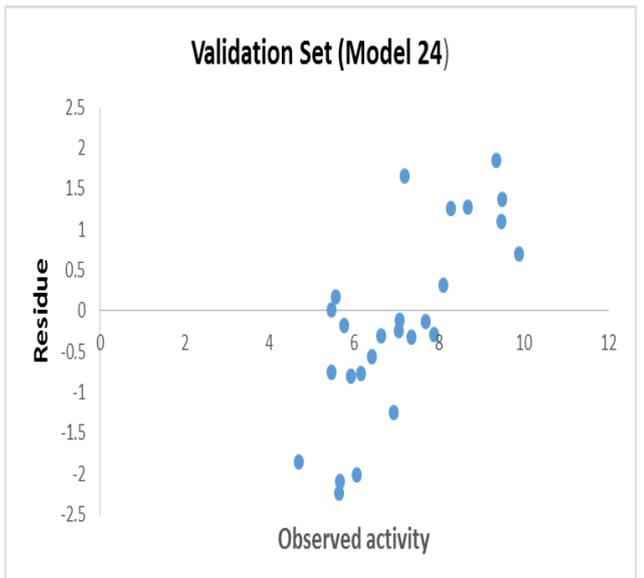
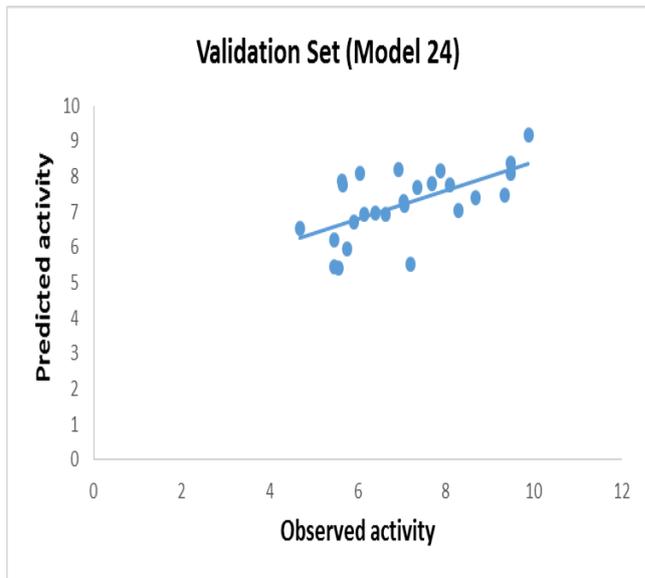
Figure 3 shows the relation of correlation coefficient (R) values for the training, test and validation sets versus model number. This figure shows that the highest (R) value of the training set is obtained for model **21** then **20**. While the highest (R) value of the test set is obtained for model **24** then model **21** then model **23**.



**Figure 3:** Plots of ANN correlation coefficient (R) values for the training, test and validation sets versus model number.

Randomization test is performed to investigate the probability of chance correlation for the optimal model 24 with 7 hidden nodes. Chance correlation was done using the same configuration parameters and the same activation functions of all our ANN models. The results of chance correlation for model 24 with 7 hidden nodes are summarized in **Table S5** in the supporting information. This table shows that the Correlation coefficients obtained by chance are low in general while PRESS values are high. This indicates that model 24 which obtained from PCA-ANN are better than those obtained by chance and it is not due to chance.

Figure 4 shows regressions between observed and predicted activity as well as their residuals for the training, validation, and test sets for model 24.



**Figure 4:** Plot of the predicted activity against observed one as well as their residues for model 24 using 7 hidden nodes. Training set, validation set, and external test set.

### Comparison with other QSAR studies

Few QSAR studies were done related to TSPO. Recently [20], checked the structure - based drug design for TSPO, challenges and opportunities and the others conclude that there are several limitations associated with currently used computational methodologies for modeling this protein. Another recent study [21] performed a comprehensive analysis involving molecular docking, QSAR modeling in which a mathematical model was established to correlate and evaluate the binding affinity of a series of 33 TSPO ligands.

More studies, Kunal and Sengupta in 2002 performed QSAR study for the binding affinities of 31 compounds of [2-phenylimidazo[1,2-a]pyridin derivatives with central benzodiazepine and peripheral benzodiazepine (TSPO) receptors using physico-chemical parameters. Attempt has been made to explore the structural and/or physico-chemical requirements of the compounds that are responsible for the selective action against peripheral benzodiazepine receptors over central ones [22].

Dalai, Leonard & Kunal [23] performed a QSAR for TSPO binding affinity in 2006, with 35 compounds of 2-phenylpyrazolo(1,5-a)pyrimidin-3-yl-acetamides using topological and physicochemical descriptors and resulted with sex models with average  $R^2=0.7$ . The calculated hydrophobicity,  $\log P_{\text{calc}}$ , shows a parabolic relation with the TSPO receptor binding affinity, which suggests that the binding affinity increases with the increase in the partition coefficient of the compounds until it reaches the critical value after which the affinity decreases. The range of the optimum values of  $\log P_{\text{calc}}$  is between 5.423-5.819 as found from different equations.

Roy Kunal and Dalai performed a QSAR study in 2007 to explore the structural and physicochemical requirements of ligands N, N-dialkyl-2-phenylindol-3-yl-glyoxylamides for binding with peripheral benzodiazepine receptor (TSPO) by using 27 compounds. The calculated partition coefficient values show parabolic relations with the TSPO binding affinity, suggesting that the binding affinity increases with increase in the partition coefficient of the compounds until it reaches the critical value after which the affinity decreases. The critical value of  $\log P$  is within range of 6.052-6.410 [24].

### CONCLUSIONS

A quantitative-structural activity relationship analysis has been conducted on the activity of a set of 136 ligand for Translocator protein (TSPO), by using MLR and principal component-artificial neural networks (PC-ANN) modeling methods, where

the strength and the predictive performance of the proposed models was verified using internal (cross-validation and Y-scrambling).

The results obtained by MLR was a number of models (Models 12- 24) which have a good predictive power ( $R^2$ ) > 0.6, the best model was model number 24 which includes 24 descriptors, and resulted with  $R= 0.909$ ,  $R^2=.826$ , and  $R^2_{adj.}= 0.788$ . The results show that model 24 has the highest correlation coefficient for the test set (0.85016) indicating its high predictive power, which chosen to continue ANN to find the optimal number of hidden nodes for each one of these models

According to the results; model 24 with 7 hidden nodes were chosen as the best models with the optimal hidden nodes because they have high prediction power (R), minimum PRESS value of the test set, and minimum number of hidden nodes. ANN resulted model were validated through randomization test, then the conditions proposed by Golbraikh and Tropsha were applied to conclude that the QSAR models have acceptable prediction power or not. However, the best ANN model with a good predictive power was model no. 24.

## REFERENCES

- [1] Selvaraj, V., D.M. Stocco, and L.N. Tu, Minireview: Translocator Protein (TSPO) and Steroidogenesis: A Reappraisal. *Molecular Endocrinology*, 2015
- [2] Scarf, A.M., L.M. Ittner, and M. Kassiou, The translocator protein (18 kDa): central nervous system disease and drug design. *J Med Chem*, 2009. **52**(3): p. 581-592.
- [3] Sakai, M., et al., Translocator protein (18kDa) mediates the pro-growth effects of diazepam on Ehrlich tumor cells in vivo. *European journal of pharmacology*, 2010. **626**(2): p. 131-138.
- [4] Arbo, B., et al., Therapeutic actions of translocator protein (18kDa) ligands in experimental models of psychiatric disorders and neurodegenerative diseases. *The Journal of steroid biochemistry and molecular biology*, 2015, **154**: p. 68-74.
- [5] Cappelli, A., et al., Structure-activity relationships in carboxamide derivatives based on the targeted delivery of radionuclides and boron atoms by means of peripheral benzodiazepine receptor ligands. *J Med Chem*, 2003. **46**(17): p. 3568-71.
- [6] Trapani, G., et al., Structure-activity relationships and effects on neuroactive steroid synthesis in a series of 2-phenylimidazo[1,2-a]pyridineacetamide peripheral benzodiazepine receptors ligands. *J Med Chem*, 2005. **48**(1): p. 292-305.
- [7] Cappelli, A., et al., Synthesis and structure-activity relationship studies in translocator protein ligands based on a pyrazolo[3,4-b]quinoline scaffold. *J Med Chem*, 2011. **54**(20): p. 7165-75.

- [8] Cappelli, A., et al., Synthesis and structure-activity relationship studies in peripheral benzodiazepine receptor ligands related to alpidem. *Bioorg Med Chem*, 2008. **16**(6): p. 3428-37.
- [9] Anzini, M., et al., Mapping and fitting the peripheral benzodiazepine receptor binding site by carboxamide derivatives. Comparison of different approaches to quantitative ligand-receptor interaction modeling. *J Med Chem*, 2001. **44**(8): p. 1134-50.
- [10] O. Deeb, B. Hemmateenejad. "ANN-QSAR model of drug-binding to human serum albumin", *Chem. Biol. Drug Des.* (2007), **70**: 19–29.
- [11] B. Hemmateenejad, M. A. Safarpour, R. Miri, N. Nesari, " Toward an optimal procedure for PC-ANN model building: prediction of the carcinogenic activity of a large set of drugs", *J. Chem. Inf. Model.* (2005), **45**: 190–199.
- [12] Omar Deeb, Manal Muhtaseb, Basheerulla Shaik (2024), "Exploring QSARs for inhibiting activity of a set of EGFR tyrosine kinase inhibitors by GA-MLR and molecular Docking simulations, BJMAS- British Journal of Multidisciplinary and Advanced Studies: Health and Medical Sciences, 2024, **5** (2), 12-40.
- [13] O. Deeb and M. Drabh, "Exploring QSARs of Some Analgesic Compounds by PC-ANN", *Chem Biol Drug Des*; (2010), **76**: 255–262.
- [14] P. V. Khadikar, O. Deeb, A. Jaber, J. Singh, V. K. Agrawal, S. Singh and M. Lakhwani. "Development of Quantitative Structure-Activity Relationship for a set of Carbonic Anhydrase Inhibitors : Use of Quantum and Chemical Descriptors". *Letters in Drug Design & Discovery*; 2006, **3**(9): 622-635
- [15] O. Deeb, B. Hemmateenejad , A. Jaber, R. Garduno-Juarez and R. Miri. "Effect of the electronic and physicochemical parameters on the carcinogenesis activity of some sulfa drugs using QSAR analysis based on genetic-MLR and genetic PLS". *Chemosphere* (2007), **67**(11): 2122-2130
- [16] O. Deeb, K. M. Youssef and B. Hemmateenejad, "QSAR of Novel Hydroxyphenylureas as Antioxidant Agents". *QSAR and Combinatorial Sciences*; 2008, **27**(4): 417-424.
- [17] O. Deeb, "Correlation ranking and stepwise regression procedures in PC-ANN modeling and application to predict the toxic activity and HSA binding affinity". *Chemometrics and Intelegent Laboratory Systems.*; 201), **104**: 181-194.
- [18] A. Golbraikh, A. Tropsha. "Beware of q2!". *J Mol Graph Model*; 2002, **20**: 269–276.
- [19] D. E. Rumelhart, G. E. Hinton, R. J. Williams. "Learning representations by back-propagating errors". *Nature*; 1986, **323**: 33–536.
- [20] M. Gladi, A.P. Montgomery, M. Kassiou et al. "Structure-based drug design for TSPO: Challenges and opportunities" *Biochimi*, 2024, **224**: 41-50.
- [21] Singh, V.K., Azad, P. & Tiwari, A.K. Computational analysis of new generation TSPO ligands. *Discov.* 2025, 2(3), <https://doi.org/10.1007/s44345-025-00009-9>

- [22] Roy, K., A.U. De, and C. Sengupta, QSAR of peripheral benzodiazepine receptor ligand 2-phenylimidazo-[1,2-a]pyridine derivatives with physico-chemical parameters. *Indian J Biochem Biophys*, 2003. **40**(3): p. 203-8.
- [23] Dalai, M.K., J.T. Leonard, and K. Roy Exploring QSAR of peripheral benzodiazepine receptor binding affinity of 2-phenylpyrazolo[1,5-a]pyrimidin-3-yl-acetamides using topological and physicochemical descriptors. 2006. **45B**.
- [24] Roy, K. and M.K. Dalai, Exploring QSAR of peripheral benzodiazepine receptor binding affinity of N,N-dialkyl-2-phenylindol-3-yl-glyoxylamides using physico-chemical descriptors. *Indian J Biochem Biophys*, 2007. **44**(2): p. 114-21.

### Supporting material

Table S1: Translocator protein (TSPO) ligands and their inhibitory activity ( $pIC_{50}$ )

Table S2. LOO cross validation results for models 12-14.

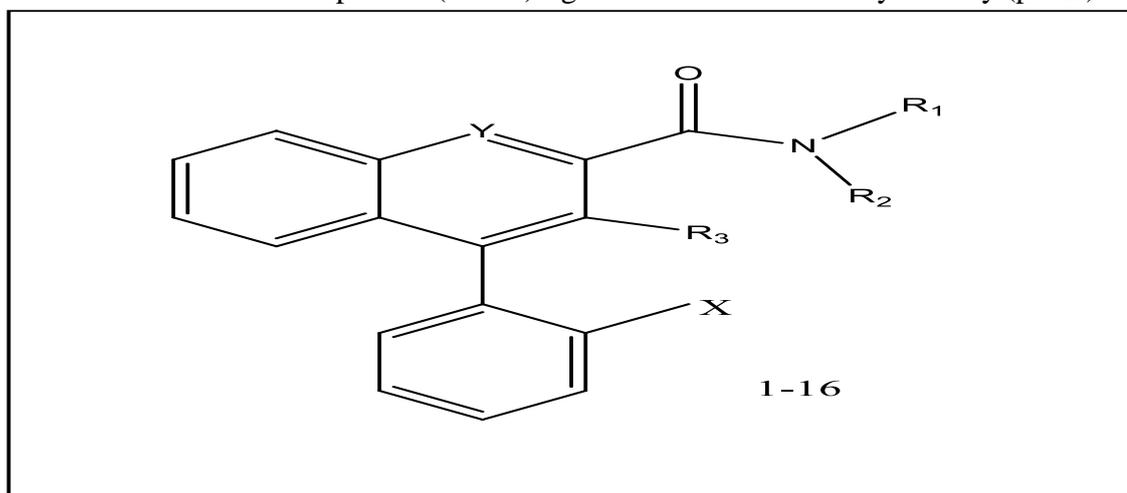
Table S3. LMO cross validation results for models 12-14.

Table S4. Correlation Coefficient and Cross Validation Parameters for ANN Models 19-24.

Table S5. Chance Correlation of Model 24 with 7 Hidden Nodes.

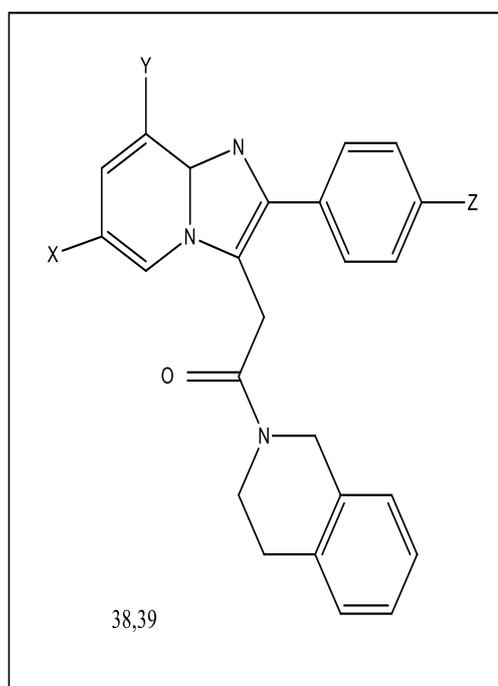
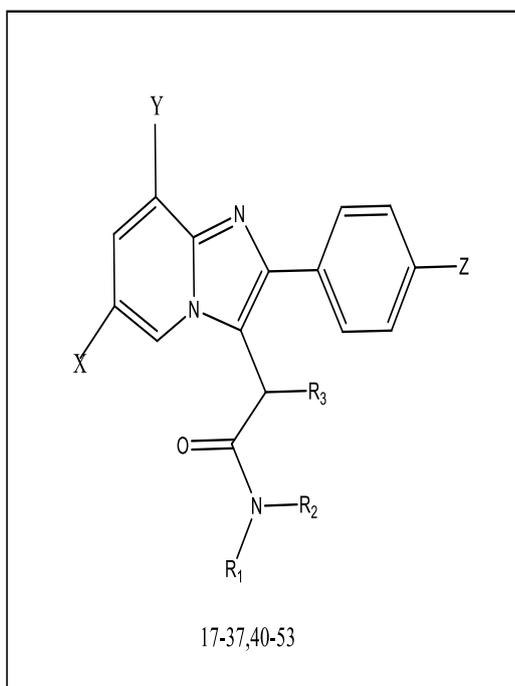
### Supplementary material

Table S1. Translocator protein (TSPO) ligands and their inhibitory activity ( $pIC_{50}$ )



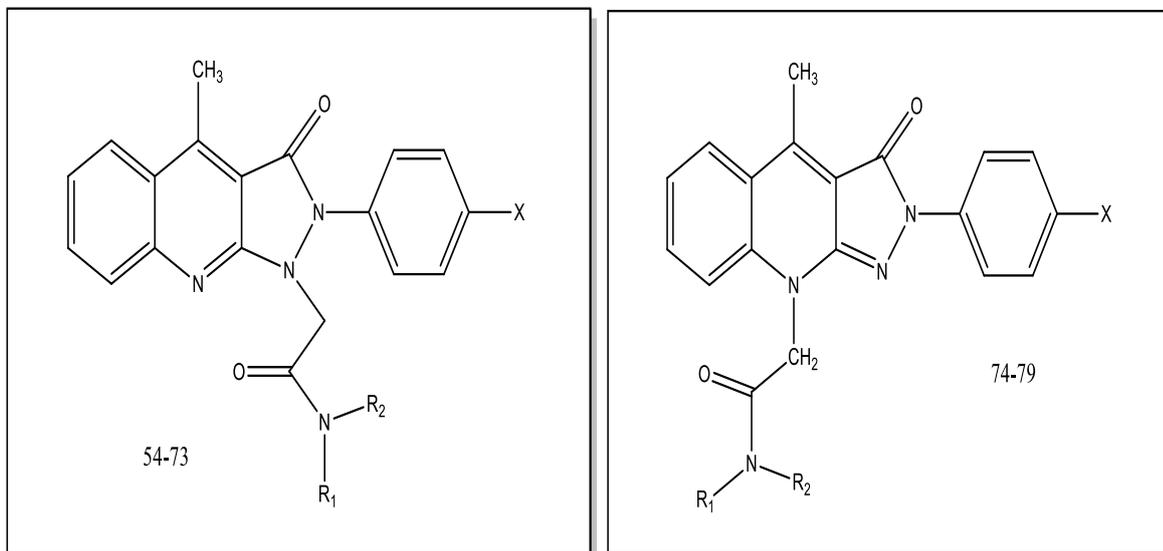
Compound s Number	Index *	X	Y	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pIC <sub>50</sub>
001	3a	H	CH	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	H	5.56863
002	3b	H	CH	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	H	7.19382
003	3c	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	H	7.19382
004	3d	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	H	7.42021
005	3e	H	CH	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	CH <sub>3</sub>	5.18045
006	3f	H	CH	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>3</sub>	8.00877
007	3g	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	CH <sub>3</sub>	4.98842
008	3h	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>3</sub>	8.33724
009	3i	F	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	CH <sub>3</sub>	5.82681
010	3j	F	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>3</sub>	8.65757
011	3k	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	7.95860
012	3l	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>2</sub> OH	8.06048
013	3m	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>2</sub> CL	9.34678
014	3n	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>2</sub> N(C <sub>2</sub> H 5) <sub>2</sub>	7.92081
015	3o	H	N	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>2</sub> N(C <sub>2</sub> H 5)CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	7.88605
016	3p	H	N	CH <sub>2</sub> CCH	CH <sub>3</sub>	CH <sub>3</sub>	7.49485

\*: Reference [5]



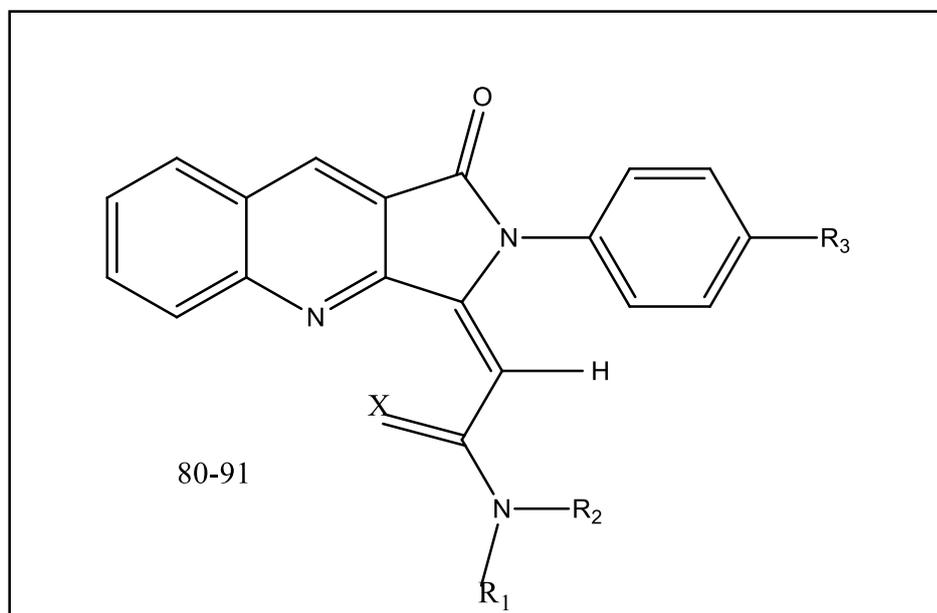
Compounds Number	Index *	X	Y	Z	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pIC <sub>50</sub>
017	1	H	H	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	H	8.230
018	2	H	Cl	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	H	8.104
019	3	Cl	Cl	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	H	8.284
020	4	Cl	Cl	Cl	<i>n</i> -C <sub>6</sub> H <sub>13</sub>	<i>n</i> -C <sub>6</sub> H <sub>13</sub>	H	6.424
021	5	Cl	H	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	H	8.485
022	6	Cl	H	Cl	<i>n</i> -C <sub>6</sub> H <sub>13</sub>	<i>n</i> -C <sub>6</sub> H <sub>13</sub>	H	8.292
023	7	Cl	H	H	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	C <sub>6</sub> H <sub>5</sub>	H	7.939
024	8	Cl	CL	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	C <sub>6</sub> H <sub>5</sub>	H	7.876
025	9	Cl	H	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	C <sub>6</sub> H <sub>5</sub>	H	8.824
026	10	Cl	CL	H	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	CH <sub>2</sub> C <sub>6</sub> H <sub>6</sub>	H	7.616
027	11	Cl	CL	Cl	<i>tert</i> -C <sub>4</sub> H <sub>9</sub>	CH <sub>2</sub> C <sub>6</sub> H <sub>6</sub>	H	5.464
028	12	Cl	CL	Cl	<i>n</i> -C <sub>3</sub> H <sub>7</sub>	4-NO <sub>2</sub> -CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	7.566
029	13	Cl	CL	Cl	C <sub>6</sub> H <sub>5</sub>	H	H	7.701
030	14	Cl	CL	Cl	CH <sub>2</sub> CHC H <sub>2</sub>	CH <sub>2</sub> CHCH <sub>2</sub>	H	8.092
031	15	Cl	CL	Cl	-(CH <sub>2</sub> ) <sub>4</sub> -		H	6.668
032	16	Cl	CL	H	-(CH <sub>2</sub> ) <sub>4</sub> -		H	5.907
033	17	Cl	Cl	H	-(CH <sub>2</sub> ) <sub>5</sub> -		H	6.804
034	18	Cl	Cl	Cl	-(CH <sub>2</sub> ) <sub>5</sub> -		H	8.301
035	19	Cl	H	Cl	-CH <sub>2</sub> CH(COOC <sub>2</sub> H <sub>5</sub> )(CH <sub>2</sub> ) <sub>3</sub> -		H	7.454
036	20	Cl	Cl	Cl	-CH <sub>2</sub> CH(COOC <sub>2</sub> H <sub>5</sub> )(CH <sub>2</sub> ) <sub>3</sub> -		H	6.845
037	21	Cl	Cl	Cl	-(CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub> )(CH <sub>2</sub> ) <sub>2</sub> -		H	4.682
038	22	Cl	Cl	H	-	-	-	7.412
039	23	Cl	Cl	Cl	-	-	-	8.313
040	24	Cl	CL	H	2- pyridylethyl	CH <sub>3</sub>	H	5.663
041	25	Cl	Cl	Cl	2- pyridylethyl	CH <sub>3</sub>	H	6.046
042	26	Cl	CL	H	2-pyridyl	H	H	5.677
043	27	Cl	CL	CL	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	H	H	6.409
044	28	Cl	Cl	Cl	C <sub>6</sub> H <sub>11</sub>	H	H	6.640
045	29	Cl	Cl	H	C <sub>6</sub> H <sub>11</sub>	H	H	5.878
046	30	Cl	Cl	Cl	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	H	H	6.772
047	31	Cl	Cl	Cl	<i>n</i> -C <sub>3</sub> H <sub>7</sub>	<i>n</i> -C <sub>3</sub> H <sub>7</sub>	CH <sub>3</sub>	5.920
048	32	Cl	Cl	Cl	C <sub>6</sub> H <sub>11</sub>	CH <sub>3</sub>	CH <sub>3</sub>	5.288
049	33	Cl	Cl	Cl	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	CH <sub>3</sub>	5.005
050	34	Cl	Cl	Cl	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	CH <sub>3</sub>	H	9.347
051	35	Cl	Cl	H	<i>n</i> -C <sub>4</sub> H <sub>9</sub>	CH <sub>3</sub>	H	8.456
052	36	Cl	Cl	Cl	C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	H	9.481
053	37	Cl	Cl	Cl	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	H	8.623

\*: Reference [6].



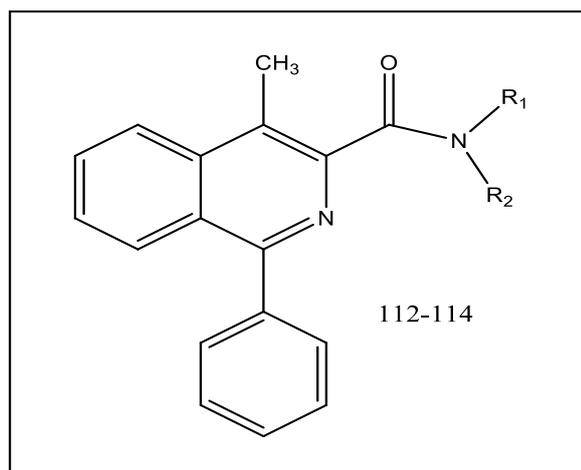
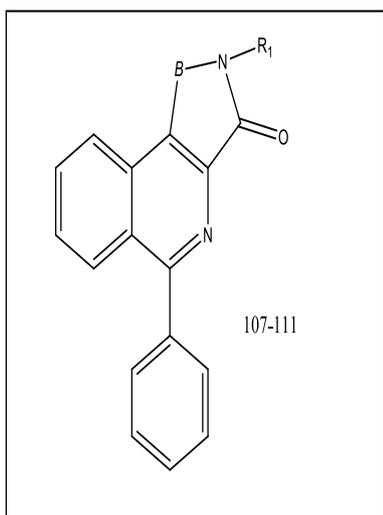
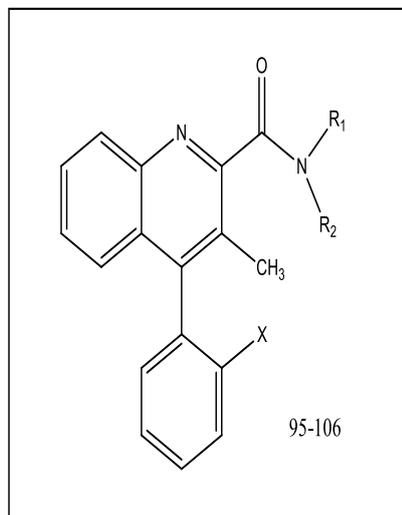
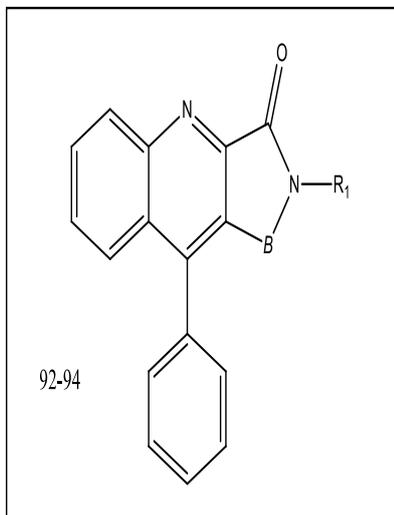
Compounds Number	Index *	R1		IC50
		CH3	(CH2)3CH3	.18708
		C2H5	C2H5	.74472
		CH(CH3)2	CH(CH3)2	.85636
		(CH2)2CH3	(CH2)2CH3	.04575
		(CH2)3CH3	(CH2)3CH3	.81815
		CH3	C6H5	.93181
		CH3	p-Cl-C6H4	.74472
		CH3	p-CH3OC6H4	.76955
		CH3	CH2C6H5	.05502
		CH2C6H5	C2H5	.65757
		CH2C6H5	CH(CH3)2	.51712
		CH2C6H5	(CH2)3CH3	.15676
		CH2C6H5	CH2C6H5	.46042
		CH3	(CH2)3CH3	.95860
		C2H5	C2H5	.64378
		(CH2)2CH3	(CH2)2CH3	.06048
		CH3	p-Cl-C6H4	.16749
		CH3	(CH2)3CH3	.84163
		(CH2)2CH3	(CH2)2CH3	.04575
		CH3	p-Cl-C6H4	.79588
		CH3	(CH2)3CH3	.85387
		C2H5	C2H5	.08092
		(CH2)3CH3	(CH2)3CH3	.46852
		CH3	p-Cl-C6H4	.88605
		CH3	(CH2)3CH3	.88605
		CH3	p-Cl-C6H4	.55284

\*: Reference [7].



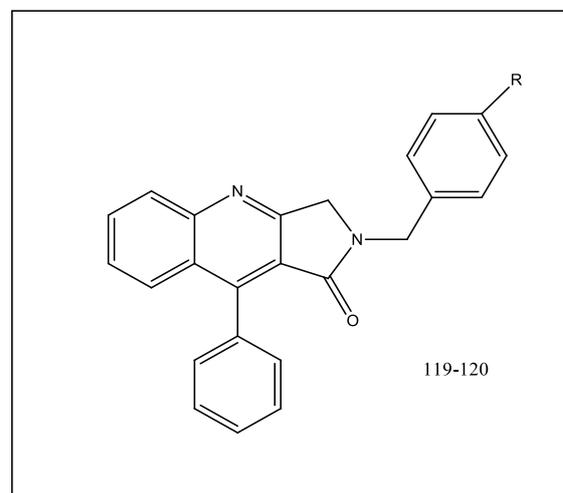
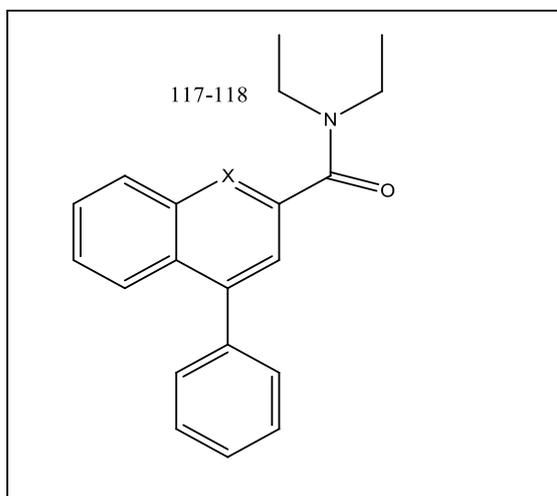
Compounds Number	Index *	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X	pIC <sub>50</sub>
080	8a	4-(CH <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	Cl	=O	9.04575
081	8b	4-ClC <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	Cl	=O	9.20760
082	8c	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	Cl	=O	7.58502
083	8d	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	Cl	=O	7.82390
084	8e	(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	Cl	=O	7.67778
085	9a	C <sub>6</sub> H <sub>5</sub>	CH <sub>3</sub>	Cl	=O	7.35654
086	9b	4-(CH <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	H	Cl	=O	7.11918
087	9c	4-(OH) C <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	Cl	=O	7.13076
088	9d	CH <sub>2</sub> CCH	CH <sub>3</sub>	Cl	=O	6.49485
089	9e	4-(CH <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	H	=O	7.92081
090	9f	4-ClC <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	H	=O	8
091	9g	4-(CH <sub>3</sub> O)C <sub>6</sub> H <sub>4</sub>	CH <sub>3</sub>	Cl	=H <sub>2</sub>	5.62708

\*: Reference [8].



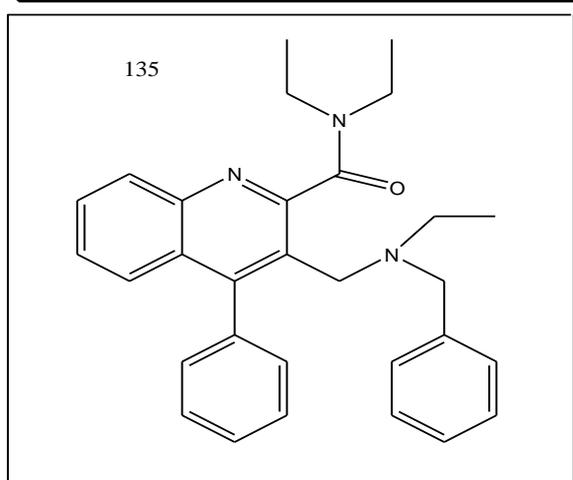
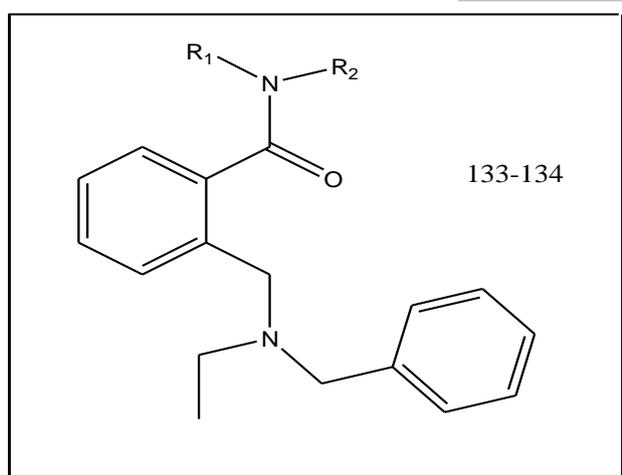
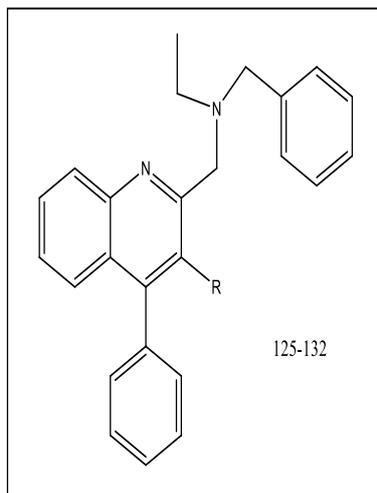
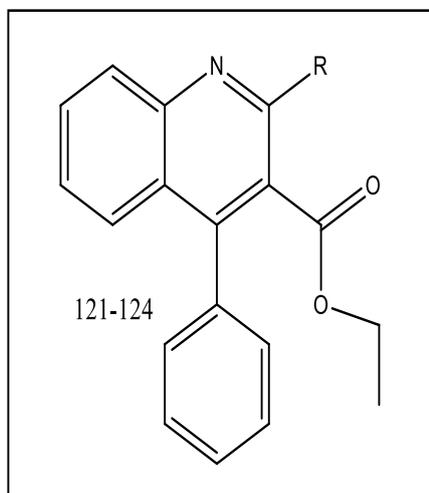
Compounds Number	Index *	Bridge (B)	X	R <sub>1</sub>	R <sub>2</sub>	pIC <sub>50</sub>
092	7a	<i>n</i> -Bu-CH	-	Benzyl	-	6.09151
093	7b	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub>	-	<i>s</i> -Bu	-	5.92081
094	7c	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub>	-	Benzyl	-	6.76955
095	8a	-	H	<i>s</i> -Bu	H	6.63827
096	8b	-	F	<i>s</i> -Bu	H	7.88605
097	8c	-	H	Benzyl	H	5.92081
098	8d	-	H	4-Cl-Benzyl	H	5.76955
099	8e	-	F	4-Cl-Benzyl	H	6.56863
100	8f	-	H	<i>s</i> -Bu	Me	8.67778
101	8g	-	F	<i>s</i> -Bu	Me	8.53760
102	8h	-	H	Benzyl	Me	8.67778
103	8i	-	H	4-Cl-Benzyl	Me	8.00877
104	8j	-	F	4-Cl-Benzyl	Me	8.46852
105	8k	-	H	4-Cl-Ph	Me	8.19382
106	8l	-	H	4-MeO-Ph	Me	8.05551
107	9a	CH <sub>2</sub>	-	<i>s</i> -Bu	-	6.20760
108	9b	CH <sub>2</sub>	-	Benzyl	-	6.31875
109	9c	CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub>	-	<i>s</i> -Bu	-	6.10790
110	9d	CH=CH-CH <sub>2</sub>	-	<i>s</i> -Bu	-	6.30980
111	9e	CH=CH-CH <sub>2</sub>	-	Benzyl	-	7.34678
112	10a	-	-	<i>s</i> -Bu	H	6.25963
113	10b	-	-	<i>s</i> -Bu	Me	7.95860
114	10c	-	-	Benzyl	Me	8.50863
115	11a	CH <sub>2</sub> -CH <sub>2</sub>	-	-	-	8.05060
116	11b	O-CH <sub>2</sub> -CH <sub>2</sub>	-	-	-	8

\*: Reference [9]



Compounds Number	Index *	X	R	pIC50
117	4a	CH	-	8.30102
118	4b	N	-	7.56863
119	5a	-	Me	6.38721
120	5b	-	CL	6.67778

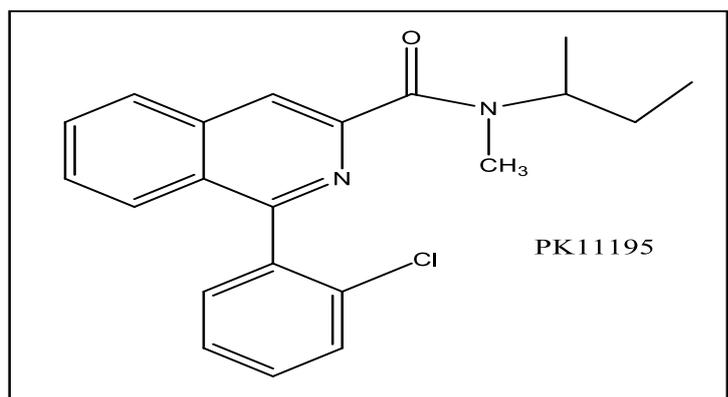
\*: Reference [9]



Compounds Number	Index *	R	R <sub>1</sub>	R <sub>2</sub>	pIC <sub>50</sub>
121	6a	CH <sub>2</sub> N(Et)Bn	-	-	6.15490
122	6b	N(Et)Bn	-	-	6.26760

123	12a	Cl	-	-	5.43179
124	12b	CH <sub>2</sub> THIQ	-	-	5.92628
125	13a	CONMe <sub>2</sub>	-	-	6.64016
126	13b	CONEt <sub>2</sub>	-	-	6.42829
127	13c	CON( <i>n</i> -Pr) <sub>2</sub>	-	-	6.34103
128	13d	CON(Me)Ph	-	-	5.88041
129	13e	CON(Me) <sub>4</sub> -Cl-Ph	-	-	5.60572
130	13f	CON(H) <i>n</i> -Pr	-	-	6.05948
131	13g	CON(H)Bn	-	-	5.53835
132	13h	H	-	-	5.46852
133	15b	-	Et	Et	5.08570
134	15c	-	<i>n</i> -Pr	<i>n</i> -Pr	5.05222
135	16	-	-	-	8.38721

\*: Reference [9]



p. ber	reference compound	0	0	0	0	0
	1195	572	502	505	778	757

**Table S2:** LOO cross validation results.

model	No. desc.	PRESS	SPRESS	SST	R <sup>2</sup> <sub>cv</sub>	PRESS/SST	PSE	RSEP
12	12	87.824	0.845	125.942	0.302	0.697	0.803	11.004
13	13	87.431	0.846	125.303	0.302	0.697	0.801	10.980
14	14	78.757	0.806	132.741	0.406	0.593	0.760	10.421
15	15	74.705	0.789	136.882	0.454	0.545	0.741	10.149
16	16	68.680	0.759	142.558	0.518	0.481	0.710	9.731

17	17	67.408	0.755	144.202	0.532	0.467	0.704	9.641
18	18	63.928	0.739	147.714	0.567	0.432	0.685	9.389
19	19	56.146	0.695	154.679	0.637	0.362	0.642	8.799
20	20	54.708	0.689	156.560	0.650	0.349	0.634	8.685
21	21	52.413	0.678	158.942	0.670	0.329	0.620	8.5015
22	22	50.644	0.669	160.676	0.684	0.315	0.610	8.356
23	23	49.162	0.662	162.211	0.696	0.303	0.601	8.233
24	24	46.108	0.644	165.143	0.720	0.279	0.5822	7.973

**Table S3:** LMO cross validation results.

Model	No. desc.	PRESS	SPRESS	SST	R <sup>2</sup> <sub>cv</sub>	PRESS/SST	PSE	RSEP
12	12	102.289	0.911	145.424	0.296	0.703	0.867	11.795
13	13	95.909	0.886	152.582	0.371	0.628	0.839	11.421
14	14	89.930	0.862	154.158	0.416	0.583	0.813	11.060
15	15	84.282	0.838	157.975	0.466	0.533	0.787	10.707
16	16	83.151	0.835	162.574	0.488	0.511	0.781	10.635
17	17	82.525	0.836	172.536	0.521	0.478	0.779	10.595
18	18	77.411	0.813	174.643	0.556	0.443	0.754	10.261
19	19	69.566	0.774	177.321	0.607	0.392	0.715	9.727
20	20	64.308	0.747	179.401	0.641	0.358	0.687	9.352
21	21	62.574	0.740	189.644	0.67	0.33	0.678	9.225
22	22	59.976	0.728	191.991	0.687	0.312	0.664	9.032
23	23	58.036	0.719	195.521	0.703	0.296	0.653	8.885
24	24	52.387	0.687	195.586	0.732	0.267	0.620	8.441

**PRESS** Predictive residual sum of squares which also called **SSE** (Error sum of squares).

**SST** Total sum of squares

**R<sup>2</sup><sub>cv</sub> or Q<sup>2</sup>** Cross validated correlation coefficient

**SPRESS** uncertainty of prediction

**PSE** Predictive Square Errors which also called **RMSE** (Root Mean Square Error)

**RSEP** Relative Standard Error of Prediction

**Table S4:** Correlation Coefficient and Cross Validation Parameters for ANN Models 19-24.

Mo.#	hn	nPCs	R_tr	PRESS_tr	R <sup>2</sup> cv_tr	R_test	PRESS_test	R <sup>2</sup> cv_test	R_val	PRESS_val	R <sup>2</sup> CV_val
19	7	7	0.90502	19.60468	0.75342	0.80464	15.51120	0.40055	0.71777	26.23444	0.21058
20	7	7	0.92656	15.31905	0.82410	0.81178	15.06395	0.43651	0.72596	26.25901	0.29857
21	7	7	0.93237	14.08497	0.84020	0.83188	14.03396	0.60656	0.73968	25.52826	0.35811
22	7	6	0.91075	18.29186	0.78943	0.80201	15.93371	0.33536	0.68588	28.44290	-0.04749
23	7	6	0.90644	19.11755	0.78280	0.81751	14.52566	0.50213	0.65518	30.73424	-0.13270
24	7	6	0.90862	18.70121	0.78523	0.85016	12.20571	0.57830	0.68529	28.85221	0.10824

**Table S5:** Chance Correlation of Model 24 with 7 Hidden Nodes

Trial No.	nPCs	R_tr	PRESS_tr	R <sup>2</sup> cv_tr	R_test	PRESS_test	R <sup>2</sup> cv_test	R_val	PRESS_val	R <sup>2</sup> cv_val
1	6	0.224	7183.46	-6.709	0.189	12.548	9-270.10	7-0.25	11.440	-69.747
2	6	0.292	4175.06	-10.017	4-0.45	14.204	-1085.815	8-0.10	413.20	-292.240
3	6	70.07	5195.87	-12.219	-0.358	913.87	-63.759	0.354	3.999	-23.686
4	6	7-0.27	229.782	7-16.54	1-0.38	613.32	3-502.79	20.65	93.75	-51.454
5	6	60.07	9200.85	-13.897	5-0.34	911.92	-301.221	-0.808	11.332	8-74.11
6	6	10.06	5223.05	7-4.67	1-0.16	11.580	-19590.4	0.639	9.807	-1928.422
7	6	-0.173	283.432	8-3.88	0.365	10.382	5-27.96	-0.728	519.07	5-8.39
8	6	0.047	257.630	-2.261	-0.454	913.18	1-35.86	0.699	13.07	9-6.05
9	6	-0.179	4268.56	-4.842	-0.428	16.450	-19.304	30.44	3.942	2-5.32
10	6	0.167	4203.66	6-3.47	0.009	811.52	-367.780	-0.692	84.97	-832.566