



CODEN (USA): IAJPBB

ISSN: 2349-7750

**INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES**Available online at: <http://www.iajps.com>

Research Article

**PREDICTIVE MODELS FOR SYSTEM X_c^- ANTIPORTER
INHIBITION BASED ON STRUCTURALLY DIVERSE
MOLECULES**Dhaval Patel¹, Mukesh Nandave¹ and Prashant S. Kharkar^{2*}¹Department of Pharmacology SPP School of Pharmacy and Technology Management, SVKM's NMIMS, Mumbai, India²Department of Pharmaceutical Chemistry, SPP School of Pharmacy and Technology Management, SVKM's NMIMS, Mumbai, India**Abstract:**

The morbidity and mortality throughout the world is increasing day by day due to cancer. Several molecular targets have been identified and being targeted for treatment of cancer cells. System x_c^- , an amino acid antiporter, is one such potential target. With the uptake of one molecule of cystine and release of one molecule glutamate, over expressed system x_c^- manipulates the redox status within cancer cells and protects them. Simultaneously, released glutamate helps in growth and metastasis of cancer cells. Few researches have synthesized and screened structurally diverse molecules against system x_c^- antiporter. Amongst these, few molecules like erastin analogues, amino acid analogues, iso-oxazole analogues, hydantoin analogues and sulfasalazine analogues exhibited potent inhibitory activity. It is possible to identify desirable molecular properties required for system x_c^- inhibition using information from above mentioned molecules. In context, we developed different predictive models using above mentioned analogues using SAS software using regression and decision tree analysis mainly. The score ranking overlay plots showed moderate to good fit of data for the training and validation data sets. These predictive models may further be used for the design and development of potent system x_c^- inhibitors.

Key words: System x_c^- antiporter, cystine, Glutamate, SAS software, Predictive modelling**Corresponding author:****Dr. Prashant S. Kharkar,**SPP School of Pharmacy and Technology Management,
SVKM's NMIMS, Mumbai, India

Phone: +91-22-42332000 Ext. 2028

E-mail: prashant.kharkar@nmims.edu

Please cite this article in press as Patel et al, *Predictive Models for System X_c^- Antiporter Inhibition Based On Structurally Diverse*, Indo Am. J. P. Sci, 2017; 4(05).

INTRODUCTION:

System x_c^- (Sx_c^-) antiporter is an amino acid transporter localised on several cell types of the human body [1]. The basic function of Sx_c^- (as observed in several *in vitro* experiments) is to release glutamate from the cells with uptake of cystine simultaneously in equimolar concentration [2]. By doing so, it contributes in preserving redox status of the cell through the production of glutathione (GSH) [3]. Moreover, in brain, it participates in maintaining neural plasticity through release of an excitatory neurotransmitter glutamate [4]. Structurally, Sx_c^- consists of two polypeptide chains, the light chain called xCT and the heavy chain called 4F2hc. The key role of Sx_c^- is performed by the light chain xCT while heavy chain helps in trafficking of light chain [5]. However, it was observed in one of the *in vitro* experiments that the activity of xCT is lost if heavy chain is removed [6]. Since the redox status of cancer cells is manipulated by producing GSH in abundant quantity to protect the cancer cell, the involvement of Sx_c^- in such phenomenon can be outlined in cancer pathophysiology. It is also evident through several *in vitro* experiments that, Sx_c^- is over-expressed in many cancer cell lines including glial [7], head and neck [8], lung [9], breast [10], gastrointestinal [11], pancreatic [12], ovarian [13] and colon [11] cancers. Moreover, Sx_c^- is also being explored for its role in CNS disorders like addiction, depression, epilepsy, schizophrenia, Parkinson's disease etc. [1]. Overall, Sx_c^- is a promising target for the treatment of several ailments.

Pharmacological inhibition of Sx_c^- may result in disturbed redox status of cancer cells with increased reactive oxygen species (ROS) (produced through metabolism) inside the cell causing cell death [2]. In glioblastoma, the overexpressed Sx_c^- may release elevated glutamate causing excitotoxicity followed by the death of neurons [7]. From the studies it was observed that, two FDA approved drugs sulfasalazine (SSZ) and sorafenib were potent inhibitors of Sx_c^- . Additionally, erastin also exhibited potent activity against Sx_c^- [14]. However, the enzymatic breakdown of SSZ in stomach, safety issues associated with sorafenib and unavailability of *in vivo* data for erastin have created hurdles for these potent inhibitors to move into higher phases of drug development. Indeed, few researches have synthesized analogues of erastin [14] and SSZ [15] and evaluated them against Sx_c^- which showed equivalent or superior potency. The metabolic liability in SSZ, i. e., the diazo group, was mitigated by the relatively stable alkyne group. In addition, suitable alternative fragments for the distal sulfapyridine end were discovered to reduce the toxicity of SSZ [15]. On the similar lines, extensive structure-activity relationship (SAR) of erastin led to the discovery of very potent Sx_c^- inhibitors [14]. Erastin and its analogues, representing altogether a different chemotype (quinazolin-4-one), are more drug-like than SSZ. These analogues are likely to penetrate blood-brain barrier (BBB) unlike SSZ.

Overall, erastin series looks more promising over SSZ series as Sx_c^- inhibitors.

Several other groups synthesized and evaluated isoxazole (ISO) [16-18], hydantoin [19] and amino acid [20] analogues for Sx_c^- inhibition and reported promising inhibitory activities for these molecules. The initial work on ISO analogues was based on amino methylisoxazole propionic acid (AMPA) and amino-3-carboxy-5-methylisoxazole propionic acid (ACPA). These molecules were competitive inhibitors of Sx_c^- [16]. The homology model of Sx_c^- and the computational analysis of the binding mode of previously reported isoxazole hydrazide inhibitors was elucidated in a subsequent paper [17]. In a related paper, Newell, et al. reported diaryl isoxazoles as noncompetitive inhibitors of Sx_c^- [18]. This was an extension of the previous work based on ACPA analogues. The lead molecules were the most potent inhibitors of Sx_c^- discovered so far.

In an attempt to discover inhibitors for vesicular glutamate transporter (VGLUT), Ahmed, et al. tested hydantoin analogues for VGLUT and obligate Sx_c^- inhibition. Majority of the inhibitors were potent VGLUT inhibitors whereas only few compounds exhibited Sx_c^- inhibition slightly over 50%. A similar study, Etoga et al., reported conformationally constrained amino acid analogues bearing distal sulfonic acid reduced glutamate uptake at Sx_c^- by 70-75% with no effect on glutamate uptake at VGLUT. These structurally diverse series of Sx_c^- inhibitors can greatly aid the design and development of more potent and selective analogues with the help of conventional medicinal and computational chemistry approaches such as quantitative structure-activity relationship (QSAR) and others.

The QSAR is a tool to predict the desirable molecular properties of molecules required to exhibit biological response. Different variables are used in QSAR for the prediction of biological activity for the designed molecules. This potentially helps in rationalizing the synthesis of new molecules. Working on the similar lines, we began this study with the aim of understanding the structural features and molecular properties required for Sx_c^- inhibition. Such insights are of potential benefit for the design and development of newer agents devoid of the problems associated with the earlier molecules. We hereby present our efforts in this direction to develop a robust QSAR model using structurally diverse group of structurally diverse Sx_c^- inhibitors.

MATERIALS AND METHODS:**Data set selection**

The data set used in the present study was prepared from the published literature on Sx_c^- inhibitors. All the inhibitors belonging to five series - SSZ analogues [14], erastin derivatives [15], isoxazole analogues [16-18], hydantoin derivatives [19] and amino acid analogues [20] were compiled (Tables 1-5) for the QSAR study. The inhibitory activities of all included compounds were measured in terms of their ability to

inhibit cystine/glutamate uptake at Sx_c^- . In case of erastin derivatives (Table 1) and SSZ analogues (Table 5), 50% of inhibitory concentrations (IC_{50}) of cystine/glutamate uptake at Sx_c^- in the presence of respective inhibitors have been reported. While the inhibitory activity of isooxazole analogues (Table 4), hydantoin derivatives (Table 3) and amino acid analogues (Table 2) were reported as % inhibition of uptake of cystine/glutamate at Sx_c^- considering 100% uptake in untreated controls. The respective biological activities were transformed into its appropriate form where higher value indicate more activity, viz., pIC_{50} for SSZ and erastin inhibitors and logit transforms for isoxazole, hydantoin and amino acid analogues for QSAR study. Full data set includes total 92 Sx_c^- inhibitors belonging to five chemotypes. A separate model was developed for each series containing analogous structures.

Ligand preparation

The molecular structures of the data set molecules (Tables 1-5) were drawn using ChemBioOffice suite Ultra v14.0 software (CambridgeSoft Corp., UK). Ligand preparation was performed by Schrödinger Release 2016-1: LigPrep [21] to assign appropriate ionization states, ring conformations, and stereoisomers, wherever appropriate, for all inhibitors. This was followed by the energy minimization. This set of prepared ligands (exported as .sdf file) was used for further calculation of molecular descriptors.

Calculation of descriptors

A total of 5253 molecular descriptors belonging to various blocks – constitutional, ring, topological, connectivity, information, 2D matrix-based, and many others – were calculated using Dragon 7.0 software [22]. Understanding the calculation and the interpretation of these descriptors can be challenging. An in-depth reference for understanding the descriptors is provided for the reader [23]. The descriptor list for each chemical class (Tables 1-5) along with appropriately transformed activity values (pIC_{50} , logit, etc.) was exported as .xlsx files.

Model development

SAS® Enterprise Miner™ 13.2 was used for developing various predictive models – regression and decision trees, among others [24]. The .xlsx file (containing transformed activity values and descriptors) for each chemical class was used as a separate data source in SAS® Enterprise Miner™ 13.2. All the model development activities were carried out on each of these data sources separately. Prior to model development, the data sources were examined carefully for any inconsistency with the help of ‘StatExplore’ node. Descriptors with missing or incomplete values were removed, wherever encountered. This was necessary since the data errors have significant implication on the model development. Variables (here descriptors) with missing or incomplete values can compromise the predictive power of the developed model.

Various predictive modelling methods were tried for each of the data sources. Since the target (here activity) is interval target (e. g., pIC_{50} , logit, etc.), appropriate settings were used for the corresponding nodes in SAS® Enterprise Miner™ 13.2. Further details can be found in the *Results and Discussion* section. The model development process involved mostly trial-and-error with respect to node settings. Figure 1 depicts this process for erastin analogues. Similar diagrams for other classes are given in *Supplementary Information* section (*Figures IS-4S*). The forward arrows indicate the process flow. Once the nodes are connected, the process is ‘Run’ generating the textual output along with several plots and associated statistics. Careful examination of the plots, statistics and output indicated the model quality – good/acceptable/poor. Further trial-and-error with node settings either improved or deteriorated the model quality. The good/acceptable model resulting from each node (decision tree and regression) were compared and the model descriptor relevance discussed to understand the factors governing activity trends.

RESULTS AND DISCUSSION:

The QSAR study utilized five chemical series exhibiting Sx_c^- inhibition. The first series erastin analogues included 19 compounds. The most potent compound, ERA17 (Table 1), is structurally very similar to ERA (Table 1), yet two log units away in activity. The authors could systematically explore the SAR to arrive at >100-fold gain in potency [14]. Strategic placement of aromatic rings and a methyl group on the arylalkyl ether substructure in vicinity of the quinazolinone core yielded the desired gain in potency for Sx_c^- . The dataset included pIC_{50} as interval target (activity) along with 5247 interval and 6 nominal inputs (variables). Since the number of observations (compounds) was 19, the initial investigations were carried out using all of them as (DATA Role = Train). This was followed by the dataset portioning into training and validation sets and further generational of validation statistic. This process was followed for the remaining series of molecules.

The first model was based on ‘Regression’ node with *Stepwise Linear Regression* model. Appropriate settings were selected in the *Properties* panel in SAS® Enterprise Miner™ 13.2. The selected model at the end of Step 6 included ATS2m (Broto-Moreau autocorrelation of lag 2 (log function) weighted by mass, Type: 2D autocorrelation), ATS4m (Broto-Moreau autocorrelation of lag 4 (log function) weighted by mass, Type: 2D autocorrelation), Eig12_EA_bo_ (eigenvalue No. 12 from edge adjacency matrix weighted by bond order, Type: Edge adjacency indices), GATS4s (Geary autocorrelation of lag 4 weighted by I-state, Type: 2D autocorrelation), JGI5 (mean topological charge index of order 5, Type: 2D autocorrelation), SsCl (Sum of sCl E-states, Type: Atom-type E-state indices) descriptors.

The final stepwise linear regression model was –

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-27.3702	5.0624	-5.41	<.0001
ATS2m	18.8791	2.4601	7.67	<.0001
ATS4m	-13.2387	2.3454	-5.64	<.0001
Eig12_EA_bo	2.2390	0.8210	2.73	0.0164
GATS4s	-2.2594	0.6607	-3.42	0.0041
JGI5	331.3	43.3430	7.64	<.0001
SsCl	0.0882	0.0219	4.04	0.0012

The model statistics was good with $r^2 = 0.969$, $F = 72.86$ and average squared error (ASE) = 0.0306. The Score Rankings Overlay: Activity plot (Figure 2a) showed a good agreement between the mean predicted and mean target, indicating a good fit of the data. The Effects plot (Figure 2b) listed relative contributions of the input variables to the model; JGI5 exhibited the largest absolute coefficient compared to all other descriptors present in the model.

As seen from the above model, the dominant contribution came from 2D autocorrelation descriptors. For chemical compounds, autocorrelation descriptors are calculated with respect to various molecular properties that can be represented at the atomic level, molecular surface level, etc. ATS2m and ATS4m are Moreau–Broto autocorrelation descriptors weighted by atomic masses. These descriptors define a set of autocorrelation terms defined by topological distances for each atom pair two and four atoms apart (ATS2 and ATS4, respectively) weighted by mass. In other words, all the atom pairs at two and four atoms apart along with the atomic masses of the constituent atoms contributed positively (ATS2m) and negatively (ATS4m) to the activity (pIC_{50}). Increase in ATS2m and decrease in ATS4m is desirable for increasing pIC_{50} . Erastin and ERA 1 differ only in having ethyl versus isopropyl ether moieties and presence and absence of CH_3 group on the linker joining quinazoline and piperazine groups, respectively (Table 1). This will definitely affect both the descriptors. ERA1 was at least two-fold more potent than erastin (Table 1).

The third descriptor Eig12_EA_bo_ belongs to Edge Adjacency Index category. These descriptors are based on a square and symmetric matrix, where the rows and column correspond to edges (bonds) of the molecular graph. Eig12_EA_bo_, weighted by bond order calculated by quantum chemical methods, contributed positively to pIC_{50} . Next two descriptors - GATS4s and JGI5 were 2D autocorrelation descriptors which contributed negatively and positively, respectively to the activity. Topological charge indices potentially evaluate the charge transfer between pairs of atoms, thereby global charge transfer in a molecule. Given the fact that Sx_c^- is an anion transporter, it is no surprise that the global charge transfer, represented

by JGI5, is one of the most important descriptors contributing to the activity. The last descriptor SsCl is an atom-type E-state index which encodes topological and electronic information related to particular atom types; here it is Cl atom type. ERA14, lacking aromatic -Cl substituent, was completely inactive as Sx_c^- inhibitor (Table 1).

The regression model was followed by decision tree model using 'Decision Tree' node. The splitting rule – *Interval Target Criterion: ProbF* was used. This setting specifies the method of searching for and evaluating candidate splitting rules in the presence of an interval target. For constructing subtree, *Assessment* method was chosen with ASE as the assessment measure. These settings decide how to construct the sub-tree in terms of selection methods. The *Assessment* method chooses the smallest subtree with the best assessment value, i.e. ASE, in the present selection. The final tree is shown in Figure 3. The tree contained seven nodes and three splitting rules based on descriptors – R8m (R autocorrelation of lag 8 / weighted by atomic masses, Type: GETAWAY descriptors), SpMax6_Bh_m_ (largest eigenvalue No. 6 of Burden matrix weighted by mass, Type: Burden eigenvalues) and Eig07_AEA_dm_ (eigenvalue No. 7 from augmented edge adjacency matrix weighted by dipole moment, Type: Edge adjacency indices). The splitting rules could clearly segregate the high-, moderate and less potent compounds (Figure 3).

The data was fitted well by the model as seen from the Score Rankings Overlay: Activity plot (Figure 5Sa, *Supplementary Information* section). The ASE for the decision tree model was found to be 0.089. The relative importance of the model descriptors was found to be R8m = 1.00, SpMax6_Bh_m_ = 0.6498 and Eig07_AEA_dm_ = 0.2366. Leaf Statistics plot depicting the average target value for each node is shown in Figure 5Sb (*Supplementary Information* section). The first split was based on R8m, a GETAWAY (GEometry, Topology and Atomic Weights Assembly) descriptor calculated from the leverage matrix obtained by the centered atomic coordinates. The R-GETAWAY is based on combination of the molecular influence matrix and the geometry matrix to give the influence/distance matrix. The observations with R8m <12.5 were separated from the observations (more active compounds) have value ≥ 12.5 . The second split is

based on SpMax6_Bh_m_, a BCUT descriptor. These descriptors are the eigenvalues of a modified connectivity matrix, based on the assumption that the lowest eigenvalues contain contributions from all the atoms and thus reflect topology of the molecule. The splitting value 3.425 correctly classified more potent compounds (higher target average) from the relatively less potent ones. Further spilt of this set of observations was based on Edge Adjacency Index, Eig07_AEA_dm_. The overall performance of the tree was good.

Although the 'Regression' and 'Decision Tree' models were capable of pruning the descriptors from a larger set, another route to the model development via variable selection was attempted. The 'Variable Selection' node with default settings was used for this purpose (Figure 1). The end result was inclusion of 16 (out of 5270) variables as 'Input' while the rest were rejected. The resulting 'regression model' contained ATS2m, ATS4m, JGI5, SsCl and VE2_H2 descriptors with $r^2=0.95$, $F= 57.33$ and $ASE = 0.049$. The Score Rankings Overlay: Activity plot exhibited good overlap between the Mean predicted and Mean Target, indicating good fit (data not shown). The decision tree had 5 nodes and the final model was based on two splitting rules involving variables – ATS4m and ATS2m. This model classified the dataset into high, moderate and low potencies (data not shown) fairly well. Overall, the performance of both the models with or without variable selection was good. Majority of the important variables belonged to type 2D autocorrelation.

For model validations and related statistical checks on the model quality, the dataset were partitioned into training, validation and test sets followed by decision tree and regression analysis. The resulting decision tree contained five nodes and two splitting rules based on CMBL (conjugated maximum bond length, Type: Geometrical descriptors) and BID (Balaban ID number, Type: Walk and path counts)

(Figure 4). The relevant plots - Score Rankings Overlay: Activity and Leaf Statistics – conveyed goodness of fit of the training/validation data with $ASE = 0.0570$ (train), 0.4658 (validation) and 0.5431 (test) (*Supplementary Information section, Figure 6S*). The first split was based on CMBL descriptor, emphasizing the influence of geometrical arrangement of different atoms of molecules on biological activity. The tree shows clear separation of highly active compounds ($CMBL < 1.411$) from less active compounds ($CMBL \geq 1.411$). Further split with BID indicated clear progression of biological activity.

Similar to decision trees, the regression node was used for further model development. The selected model was trained in the last step (Step 17). It consists of the following variables namely: Eig02_AEA_bo_ (Eigenvalue n. 2 from augmented edge adjacency mat. weighted by bond order, Type: Edge adjacency indices), JGI5 (mean topological charge index of order 5, Type: 2D autocorrelations), Mor25v (signal 25 / weighted by van der Waals volume, Type: 3D-MoRSE descriptors), RDF125i (Radial Distribution Function - 120 / weighted by ionization potential, Type: RDF descriptors), SM13_AEA_dm_ (spectral moment of order 13 from augmented edge adjacency mat. weighted by dipole moment, Type: Edge adjacency indices), SM3_B_m_ (spectral moment of order 3 from Burden matrix weighted by mass, Type: 2D matrix-based descriptors), SpAD_B_s_ (spectral absolute deviation from Burden matrix weighted by I-State, Type: 2D matrix-based descriptors), SpMax_B_e_ (leading eigenvalue from Burden matrix weighted by Sanderson electronegativity, Type: 2D matrix-based descriptors) and TDB09r (3D Topological distance based descriptors - lag 9 weighted by covalent radius, Type: 3D autocorrelations) with $r^2 = 9993$, $F = 798.31$ and $ASE = 0.0006$. The final step wise regression model was

Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	-57.2796	4.6975	-12.19	<.0001
Eig02_AEA_bo_	1	-11.5239	0.7926	-14.54	<.0001
JGI5	1	106.4	18.1563	5.86	0.0020
Mor25v	1	-1.5539	0.2142	-7.26	0.0008
RDF125i	1	-0.00999	0.00226	-4.42	0.0069
SM13_AEA_dm_	1	-0.3223	0.0989	-3.26	0.0225
SM3_B_m_	1	9.7648	0.5439	17.95	<.0001
SpAD_B_s_	1	0.1632	0.00974	16.77	<.0001
SpMax_B_e_	1	10.5497	2.0345	5.19	0.0035
TDB09r	1	7.6354	1.7013	4.49	0.0065

The Score Rankings Overlay: Activity plot (Figure 5a) exhibited good agreement between the mean predicted and mean target, showing good fit of data. The Effects plot (Figure 5b) showed the relative contributions of the input variables to the model; JGi5 exhibited the highest absolute coefficient amongst all other descriptors in the model (106.4).

From total ten descriptors, five belonged to 2D matrix-based descriptors and 2 belong to Edge Adjacency matrix indicating various molecular structural aspects crucial for biological activity. Eig02_AEA_bo_, Mor25v, RDF125i and SM13_AEA_dm_ contributed negatively while others descriptors of model contributed positively.

The second series included amino acid derivatives. Most of these compounds were α -amino acids. The most active compound AA13 possessed a sulfonic acid group at one side and amino acid functionality on the other. This series has a steep SAR with respect to the stereochemistry (Table 2), emphasizing the importance of stricter stereochemical features for transporter inhibition. The activity was expressed as logit transform of the % inhibition of Glu uptake compared to control at 250 μ M concentration. A total of 5247 interval and 6 nominal inputs were used for n (number of observations) = 17. No validation and/or test data sets were used for the analyses.

The first run utilized 'Regression' node and Stepwise Linear Regression model. The node was run with default settings for other parameters. The resulting model was trained in the two steps and included Mor28v (Signal 28 / weighted by van der Waals volume, Type: 3D-MorSE descriptors) and RDF065s (Radial Distribution Function - 065 / weighted by I-state, Type: RDF descriptors) descriptors. The final stepwise linear regression model was –

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	1.2932	0.1838	7.04	<0.0001
Mor28v	-7.9449	1.8694	-4.25	0.0008
RDF065s	-0.00807	0.00365	-2.21	0.0439

The model statistics was acceptable with $r^2 = 0.711$, $F = 17.21$ and $ASE = 0.107$. The Score Rankings Overlay: Activity plot (Figure 6a) showed a poor agreement between the mean predicted and mean target, indicating a poor fit of the data. This was not surprising given the fact that the SAR of this series was so dependent on the stereochemistry, i.e., there was astounding difference between the uptake inhibitory activities of the enantiomers.

The Effects plot showed the relative contributions of the descriptors (input variables) to the model statistics. Here 1 = Mor28v and 2 = intercept and 3 = RDF065s.

The descriptor Mor28v belonged to a category of geometrical descriptors called '3D-MorSE', under the main category 'High Conformational Dependence (HCD) descriptors'. The values of these variables are very sensitive to any conformational change in the molecule. The application of these and other related descriptors can be found in the literature [25]. The appearance of Mor28v in the stepwise linear regression model was in line with the steep SAR associated with this series of molecules. The negative sign of the parameter estimate was indicative of increase in activity due to decrease in Mor28v. The second variable with little contribution to the transporter inhibitory activity (as seen from Figure 6b) - RDF065s – was a 3D Radial Distribution Function (RDF) descriptor. Various advantages associated with RDF descriptors include their independent behaviour with respect to molecular size, invariance against the rotation or translation of the entire molecule and simplicity in interpretation. The smaller value of parameter estimate showed its little contribution to the biological activity. Given the steep-SAR of amino acid analogues, no further optimization of this model was tried.

Next a 'Decision Tree' node was used. ProbF was used as the splitting rule criterion for the interval target. Other settings were used as previously mentioned for the erastin analogues. The resulting tree contained seven nodes and three splitting rules based on SpMAD_B_p_ (Spectral mean absolute deviation from Burden matrix weighted by polarizability, Type: 2D matrix-based descriptors), RDF040s (Radial Distribution Function - 040 / weighted by I-state, Type: RDF descriptors) and J_RG (Balaban-like index from reciprocal squared geometrical matrix, Type: 3D matrix-based descriptors) (Figure 7). The relevant plots - Score Rankings Overlay: Activity and Leaf Statistics – conveyed goodness of fit of the training data with $ASE = 0.0329$ (*Supplementary Information* section, Figure 7S). The relative importance of the descriptors was found to be SpMAD_B_p_ = 1.00, RDF040s = 0.6059 and J_RG = 0.4102.

The first split was based on SpMAD_B_p variable (RDF category). Like the regression model, the decision tree model also embraced the RDF descriptors, re-emphasizing the importance of molecular conformation for the transporter inhibitory activity. The observations with SpMAD_B_p < 1.523 (highly active compounds) were clearly separated from low activity compound collection with values ≥ 1.523 . Both these nodes were further split into leaf nodes based on RDF040s and J_RG descriptors yielding four leaf nodes with clear gradation of biological activity. In summary, the 'Decision Tree' model was really good with $ASE = 0.0329$.

Further, model development using 'Variable Selection' node followed by 'Regression' node yielded similar model with 'Regression' node alone. This is not discussed further. Attempting 'Polynomial Regression' as well as 'Regression' using two-factor interactions could not improve model performance. Inclusion of

more compounds to fill 'SAR holes' may lead to further improvement in the model statistics.

Further, similar to erastin analogues, data partition was performed followed by regression analysis. The selected model is trained in the last step (Step 1) which consists of the Mor24s (signal 24 / weighted by I-state, Type: 3D-MoRSE descriptors) descriptor with $r^2 = 0.7856$, $F = 25.65$ and $ASE = 0.0446$. The final step wise regression model was

Parameter	D F	Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.4493	0.1034	4.34	0.0034
Mor24s	1	-0.3486	0.0688	-5.06	0.0015

The Score Ranking Overlay: Activity plot (Figure 8a) showed moderate agreement between the mean predicted and mean target, showing moderate poor fit of data. The effect plot (Figure 8b) shows contribution of Mor24s in model. This descriptor belongs to 3D-MoRSE category indicating the relationship between 3D molecular representations of structure based on electron diffraction with biological activity. Mor24s contributed negatively to the model. The decision tree model could not progress beyond root node. Hence, it is not discussed here.

Hydantoin analogues (series 3) were similarly used for QSAR model development process. A variety of lipophilic groups, viz., (hetero)aromatic rings bearing polar as well as lipophilic substituents, were placed at position 5. The N3-H provided the acidic feature required for mimicking the acidic groups in substrate structure. Since these molecules are quite lipophilic, they are likely to cross the BBB and reach where Sx_c^- is located. Majority of the compounds were not so effective in inhibiting Glu uptake compared to the control. Compound HYD10 (Table 3) showed the highest Glu uptake inhibition in the series (51%). Moreover, this series has a chiral centre on the hydantoin nucleus. The original authors evaluated these compounds as racemic mixtures, and not optically pure substances. Such a study would have truly helped uncovering the activity of the stereoisomers, subsequent SAR and design of potent analogues. Another important consideration was the lack of truly acidic group which could be attributed to this moderate to less potency. Nonetheless, this series was important solely from its BBB crossing potential.

The stepwise linear regression model based on the Glu uptake as target variable was trained in Step 6 and contained six inputs namely, Mor07u (signal 07 / unweighted, Type: 3D-MoRSE descriptors), Mor13i (signal 13 / weighted by ionization potential, Type: 3D-MoRSE descriptors), Mor31s (signal 31 / weighted by I-state, Type: 3D-MoRSE descriptors), P_VSA_LogP_5 (P_VSA-like on LogP, bin 5, Type: P_VSA-like descriptors), SpDiam_AEA_dm (spectral diameter from

augmented edge adjacency mat. weighted by dipole moment, Type: Edge adjacency indices) and VE1sign_X (coefficient sum of the last eigenvector from chi matrix, Type: 2D matrix-based descriptors), with $r^2 = 0.9745$, $F = 133.88$ and $ASE = 0.079$. The final stepwise linear regression models was –

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-16.1405	1.0693	-15.09	<.0001
Mor07u	-0.1831	0.0318	-5.75	<.0001
Mor13i	-0.3368	0.0703	-4.79	<.0001
Mor31s	0.1508	0.0645	2.34	0.0293
P_VSA_LogP_5	0.0375	0.00201	18.63	<.0001
SpDiam_AEA_dm	2.8973	0.1828	15.85	<.0001
VE1sign_X	0.5453	0.1272	4.29	0.0003

The Score Rankings Overlay: Activity plot (Figure 9a) exhibited moderately poor agreement between the mean predicted and mean target, hinting some issues with data fitting. The Effects plot (Figure 9b) showed the relative contributions of the input variables to the model; SpDiam_AEA_dm_ exhibited the highest absolute coefficient amongst all other descriptors in the model (2.8972).

Of the six descriptors, three belonged to 3D-MoRSE category, emphasizing higher conformational dependence of the target variable. Since the original set of compounds was racemic mixtures, both the enantiomers of each compound were considered in the QSAR analyses. The enantiomers resulting from each compound were assigned the same activity as the racemic mixture. This was done to avoid any bias towards a particular enantiomer during descriptor calculations and subsequent model development process. Other three descriptors belonged to Edge Adjacency Matrix, 2D Matrix and P_VSA-like descriptors category, featuring various molecular structural aspects essential for activity. The P_VSA descriptors were used in QSAR studies of acridines previously [26]. These surface descriptors are based on atomic contributions to van der Waals surface area, logP (octanol/water), molar refractivity and partial charge and can be computed from the connection table. No 3D alignment is required for calculation of these descriptors. Their utility lies in efficiently describing ligand-receptor interactions. P_VSA_LogP_5 contributed positively to the target variable (as seen from its estimate). Other two descriptors - SpDiam_AEA_dm_ and VE1sign_X also contributed positively to the Glu uptake inhibition activity (target variable).

Further analysis using 'Decision Tree' node yielded a tree with nine nodes (Figure 10) based on four descriptors - SM09_EA_ri_ (spectral moment of order 9 from edge adjacency matrix weighted by resonance integral, Type: Edge adjacency indices), SpMAD_D_Dt (spectral mean absolute deviation from distance/detour

matrix, Type: 2D matrix-based descriptors), BIC0 (Bond Information Content index (neighbourhood symmetry of 0-order, Type: Information indices) and GATS5m (Geary autocorrelation of lag 5 weighted by mass, Type: 2D autocorrelations). As seen from the leaf nodes, the descriptors could separate active compounds from less active or inactive compounds impressively.

The decision tree model could fit the data well as seen from the Score Rankings Overlay: Activity plot (Figure 8Sa, *Supplementary Information* section). The ASE for this model was 0.0076 based on all the molecules. The relative importance of the model descriptors was found to be SM09_EA_ri_ = 1.00, SpMAD_D_Dt = 0.4412, GATS5m = 0.2961 and BIC0 = 0.0796. Leaf Statistics plot depicting the average target value for each node is shown in Figure 8Sb (*Supplementary Information* section). The first split was based on SM09_EA_ri_, an Edge Adjacency Index, which exhibited cutoff at 9.843; molecules with SM09_EA_ri_ < 9.843 were least active (average target value 0.122) while others were highly active (average target value 0.8436 (Figure 10). Further splits based on other three descriptors yielded crystal clear separation of various target levels.

For model validation, the data were partitioned into training and validation set followed by decision tree, regression and polynomial regression analysis. In case of decision tree model (Figure 11), it contained five nodes and two splitting rules based on SM10_EA_ri_ (spectral moment of order 10 from edge adjacency mat. weighted by resonance integral, Type: Edge Adjacency Descriptors) and E1m (1st component accessibility directional WHIM index / weighted by mass, Type: WHIM descriptors). The relevant plots – Score Rankings Overlay: Activity and Leaf Statistics – exhibited best fit of test/validation data with ASE = 0.0150 (training)/ 0.0365 (validation) (*Supplementary Information Section*, Figure 9s). The first split was based on SM10_EA_ri_ descriptor, emphasizing the influence of square and symmetric matrix of the molecules on biological activity. The tree showed clear separation of highly active compounds (SM10_EA_ri_ ≥ 10.953) from least active compounds (SM10_EA_ri_ < 10.953). Further split with E1m showed clear graduation of activity.

Further, Regression node was used for the partitioned data. The selected model was trained in last step (Step 5). It consists of following variables: Intercept, ATSC3p (Centred Broto-Moreau autocorrelation of lag 3 weighted by polarizability, Type: 2D autocorrelations), P_VSA_LogP_5 (P_VSA-like on LogP, bin 5, Type: P_VSA-like descriptors), R7i_ (R autocorrelation of lag 7 / weighted by ionization potential, Type: GETAWAY descriptors), RDF070s (Radial Distribution Function - 070 / weighted by I-state, Type: RDF descriptors) and RDF080v (Radial Distribution Function - 080 / weighted by van der Waals volume, Type: RDF descriptors) with $r^2 = 9773$, $F = 128.89$ and $ASE = 0.2437$. The final step wise regression model was

Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.1803	0.2171	0.83	0.4193
ATSC3p	1	0.2150	0.0332	6.48	<.0001
P_VSA_Lo gP_5	1	0.0613	0.00443	13.82	<.0001
R7i_	1	-17.5993	1.8506	-9.51	<.0001
RDF070s	1	0.00552	0.00174	3.17	0.0063
RDF080v	1	-0.3252	0.0345	-9.43	<.0001

The Score Ranking Overlay: Activity plot (Figure 12a) exhibited moderate agreement between the mean predicted and mean target, showing moderate fit of data. The Effects plot (Figure 12b) shows the relative contribution of the input variables to the model: R7i_ exhibited the highest absolute contribution amongst all other descriptors in the model (-17.5993).

Out of five variables, two belonged to RDF descriptors category which is restricted to distance ranges or specific atom types to signify specific information in a certain 3D structure space, for example, to explain sterical hindrance or structure/activity properties of a molecule. R7i_ belongs to GETAWAY descriptors which are derived from Molecular influence matrix (matrix representation of molecule). Therefore, it may be said that specific activity demands upon influence of molecular matrix. It was observed from the above model that, R7i_ and RDF080v contributed negatively while other descriptors contributed positively.

The fourth series comprised of isoxazole analogues, representing the relatively lipophilic versions of the sulfasalazine and other acidic compounds. The imine functionality mimicked the azo linkage present in SSZ. Other pharmacophoric points such as aromatic ring feature and H-bond acceptor at the proximal end (-COOH group) represented by the isoxazole N and O compensated the possible interactions SSZ might have with the antiporter. The glutamate uptake inhibitory activity was well spread from highly potent compounds to moderately and less potent compounds (Table 4).

The first model was built using 'Regression' node (stepwise linear regression with default settings). The final model was trained in step No. 6 and consisted of CATS2D_08_AL (CATS2D Acceptor-Lipophilic at lag 08, Type: CATS 2D), Eig05_EA_ed_ (eigenvalue n. 5 from edge adjacency mat. weighted by edge degree, Type: Edge adjacency indices), GATS1i (Geary autocorrelation of lag 1 weighted by ionization potential, Type: 2D autocorrelations),

Gu (total symmetry index / unweighted, Type: WHIM descriptors), Mor05u (signal 05 / unweighted, Type: 3D-MoRSE descriptors) and R8m (R autocorrelation of lag 8 / weighted by mass, Type: GETAWAY descriptors). The model statistics were $r^2 = 0.9149$, $F = 50.16$ and $ASE = 0.04699$.

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-6.0378	1.3102	-4.61	<.0001
CATS_08_AL	-0.1536	0.0171	-9.00	<.0001
Eig05_EA_ed	-0.3492	0.0728	-4.80	<.0001
GATS1i	2.1024	0.6000	3.50	0.0016
Gu14.2005	3.9080	3.63	0.0011	
Mor05u-0.6855	0.0868	-7.90	<.0001	
R8m	1.4926	0.5740	2.60	0.0147

The score ranking overlay: activity plot exhibited moderate agreement between mean predicted and mean target, indicating average fit of data (Figure 13a). Moreover, the effects plot (Figure 13b) listed relative contribution of input variables to the model where Gu showed highest contribution amongst other descriptors present in the model. The descriptor Gu belongs to WHIM symmetry/WHIM index under the main category of WHIM descriptors. The values for this descriptor is an indicator of geometric mean of the directional summaries (Value 1 indicates central symmetry). As per the model statistic, Gu contributed positively to the glutamate uptake inhibition activity. Next descriptor is GATS1i which belongs to 2D autocorrelations category of descriptors. In this descriptor, the Geary coefficient is a distance type function which exhibits different physico-chemical properties calculated for each atom. Here, in this case this property is ionisation potential. GATS1i displays a positive sign in model statistic indicating decrease in ionisation potential causes increased inhibition of glutamate uptake by Sx_c^- . The next descriptor is R8m from GATWAY descriptor category. As mentioned earlier in the text that the value of this descriptor is based on combination of molecular influences matrix and the geometry matrix to give influence/distance matrix. R8m displayed a positive sign indicating increase in influence/distance matrix results in improvement in inhibition of Sx_c^- by isooxazole derivatives. Another descriptor with comparatively less contribution in model generation is Mor05u from 3D MorSE descriptors under the main category of geometrical descriptors. These descriptors provide information by means of 3D atomic coordinates. In this case, Mor05u unweighted 3D MorSE descriptor showing negative sign in model and thus indicates high inhibition of glutamate uptake by Sx_c^- with increase in Mor05u values. Eig05_EA_ed belongs to Edge Adjacency index category. As mentioned earlier, these descriptors are based on a square and symmetric matrix, where the rows and column correspond to edges (bonds) of the molecular graphs. Eig05_EA_ed, weighted by edge degree contributed negatively to inhibition of glutamate uptake by Sx_c^- . The last descriptor from this model is

CATS2D_08_AL belongs to CATS2D descriptors. These descriptors give information about topological distance between any pair of pharmacophore point types used. In this case, the estimate value of CATS2D_08_AL is negative indicating decrease in the value of this descriptor is associated with improved Sx_c^- inhibition potential. The 'Decision Tree' model using default settings contained a total of 9 nodes (5 leaf nodes) based on four descriptors, viz., CATS2D_08_AL (CATS2D Acceptor-Lipophilic at lag 08, Type: CATS2D), VR2_G (normalized Randic-like eigenvector-based index from geometrical matrix, Type: 3D matrix-based descriptors), E1e (1st component accessibility directional WHIM index /weighted by Sanderson electronegativity, Type: WHIM descriptors) and RDF105s (Radial Distribution Function - 105 / weighted by I-state, Type: RDF descriptors) (Figure 14). Similar to previous decision tree models, this model classified different grades of uptake inhibitory activities.

The decision tree model exhibited good fit of data as depicted in Score Rankings Overlay: activity plot (*supplementary information section*, Figure10Sa). The model ASE was found to be 0.06931. Leaf statistic plot depicting the average target value for each node is shown in figure 10Sb (*supplementary information section*). The first split was based on CATS2D_08_AL with the cut off at 7.5; molecules with $CATS2D_08_AL < 7.5$ were highly active while others were least active. Both this nodes were again split into leaf nodes based on VR2_G and E1e. Molecules with $VR2_G \geq 1.0135$ were found to be potent compared to other molecules. Moreover, compounds with $E1e < 0.571$ were highly active compared to remaining molecules. VR2_G was further split in to two leaf nodes based on RDF105s. The Overall performance of tree was good.

Next, the data were portioned in to training and validation set followed by decision tree, regression and polynomial regression analysis. In case of decision tree, it contained five nodes and two splitting rules based on CATS2D_07_AL (CATS2D Acceptor-Lipophilic at lag 07, Type = CATS2D) and RDF025p (Radial Distribution Function - 025 / weighted by polarizability, Type: RDF descriptor) (Figure 15). The relevant plots – Score Rankings Overlay: Activity and Leaf Statistics – exhibited best fit of test/validation data with ASE = 0.0958 (training)/ 0.3812 (validation) (*Supplementary Information Section* Figure 11s). The first split was based on CATS2D_07_AL descriptors emphasizing effect of topological distance between any pair of pharmacophore point types on biological activity. The tree showed clear separation of highly active compounds ($CATS2D_07_AL < 3.5$) from least active compounds ($CATS2D_07_AL \geq 3.5$). Further split was based on RDF025p descriptors showed good performance of overall tree.

The regression model was also developed for the isooxazole series using partitioned data. The selected model is the model trained in the last step (Step 7). It consists of the following descriptors: CATS2D_08_AL

(CATS2D Acceptor-Lipophilic at lag 08, Type = CATS2D), CATS3D_07_LL (CATS3D Lipophilic-Lipophilic at lag 0, Type = CATS 3D), Gu (total symmetry index / unweighted, Type: WHIM descriptors), RDF090s (Radial Distribution Function - 090 / weighted by I-state, Type: RDF descriptors), SM4_B_m_ (spectral moment of order 4 from Burden matrix weighted by mass, Type: 2D matrix-based descriptors), TDB01u (3D Topological distance based descriptors - lag 1 unweighted, Type: 3D autocorrelations) and VE1_Dz_p_ (coefficient sum of the last eigenvector from Barysz matrix weighted by polarizability, Type: 2D matrix-based descriptors) with $r^2 = 9831$, $F = 141.62$ and $ASE = 0.0099$ (training) and 0.5082 (Validation).

The final step wise regression model was

Parameter	D F	Estimate	Standard Error	t Value	Pr > t
Intercept	1	-1.4399	1.7963	-0.80	0.4339
CATS2D_08_AL	1	-0.1564	0.0152	-10.26	<.0001
CATS3D_07_LL	1	0.0803	0.0207	3.89	0.0012
Gu	1	-14.2022	2.9574	-4.80	0.0002
RDF090s	1	-0.0134	0.00210	-6.39	<.0001
SM4_B_m_	1	1.8088	0.3236	5.59	<.0001
TDB01u	1	-26.4533	4.7804	-5.53	<.0001

VE1_Dz_p_	1	-0.9867	0.1918	-5.14	<.0001
-----------	---	---------	--------	-------	--------

The Score Ranking Overlay: Activity plot showed good (Figure 16a) agreement between the mean predicted and mean target, showing good fit of data. The effects plot (Figure 16b) shows the relative contribution of input variables to the model: TDB01u showed the highest absolute contribution amongst all other descriptors in model (-26.4533).

TDB01u and Gu contributed most in this model which belongs to 3D autocorrelations and WHIM descriptors category respectively. TDB01u is 3D autocorrelation descriptors emphasizing on influence of topological and geometric distances on biological activity. WHIM descriptors explain impact of molecular 3D information regarding molecular size, shape, symmetry, and atom distribution on biological activity. Both the above descriptor contributed negatively to the model. Moreover, CATS2D_08_AL, RDF090s and VE1_Dz_p_ also contributed negatively. CATS3D_07_LL and SM4_b_m_ contributed positively.

Sulfasalazine analogues were also tried for the model development. However, due to lack of number of molecules no acceptable model could be developed and hence not reported here.

CONCLUSION:

In conclusion, our work reports of predictive model development and validation using SAS. Regression and decision tree generated models for all the series showed good predictive power as confirmed from validation data. Hence, above models may be used in future for design of novel system x_c^- inhibitors for the treatment of variety of cancer cells.

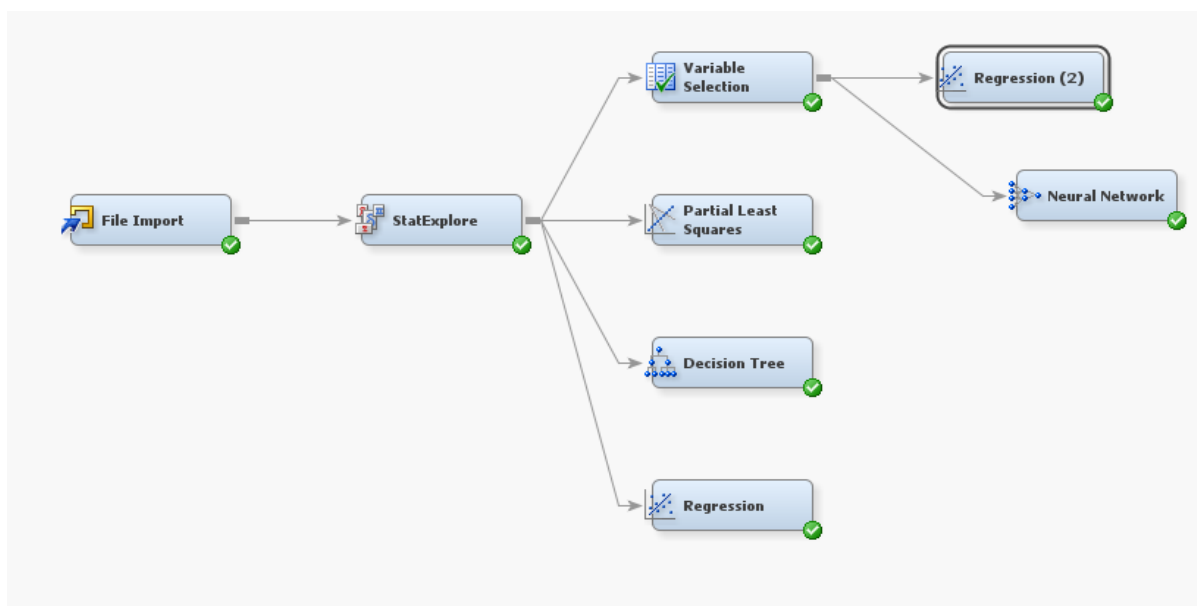


Figure 1. SAS® Enterprise Miner™ 13.2 Diagram depicting the model development for erastin analogues. Each box represents a 'node' and the 'arrows' connect various nodes completing the process flow for a

particular task, e.g., Regression (2) model development contains File Import, StatExplore, Variable Selection and Regression (2) nodes.



Figure 2. Plots resulting from stepwise linear regression model for erastin analogues. a) Score Rankings Overlay: Activity plot demonstrating good agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = JGI5, 2 = intercept, 3 = ATS2m, 4 = ATS4m, 5= GATS4s, 6 = Eig12_EA_bo_ and 7 = SsCl.

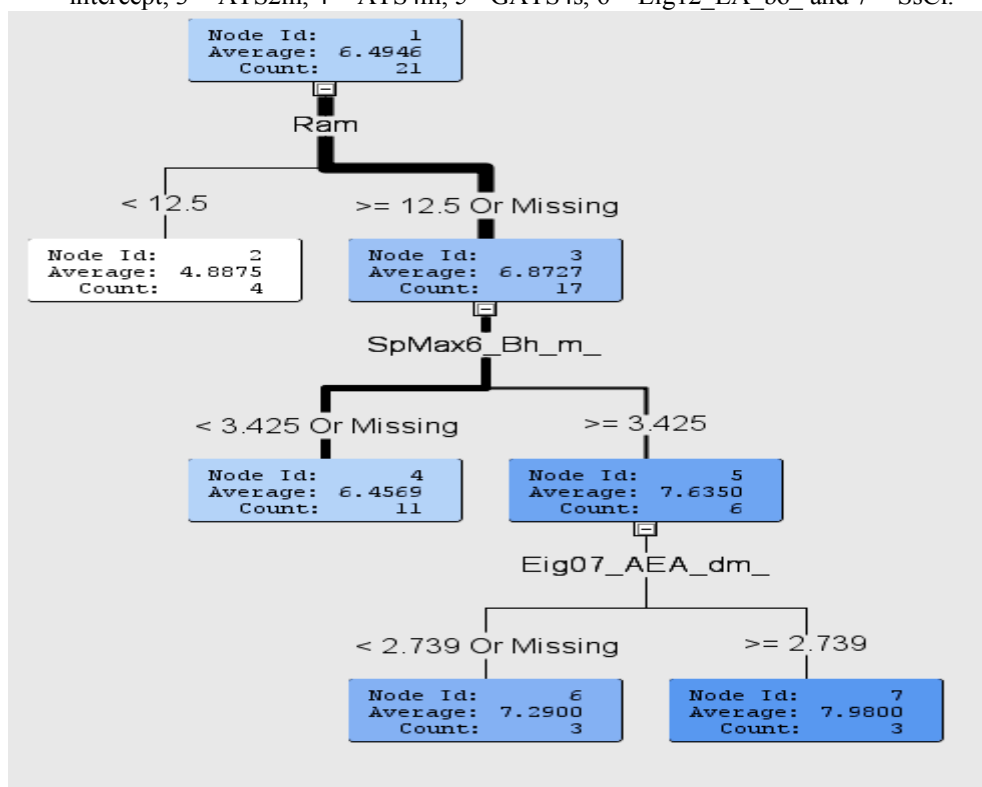


Figure 3. Decision tree model for erastin analogues with seven nodes. The splitting rules are based on three descriptors R8m, SpMax6_Bh_m_ and Eig7_AEA_dm_.

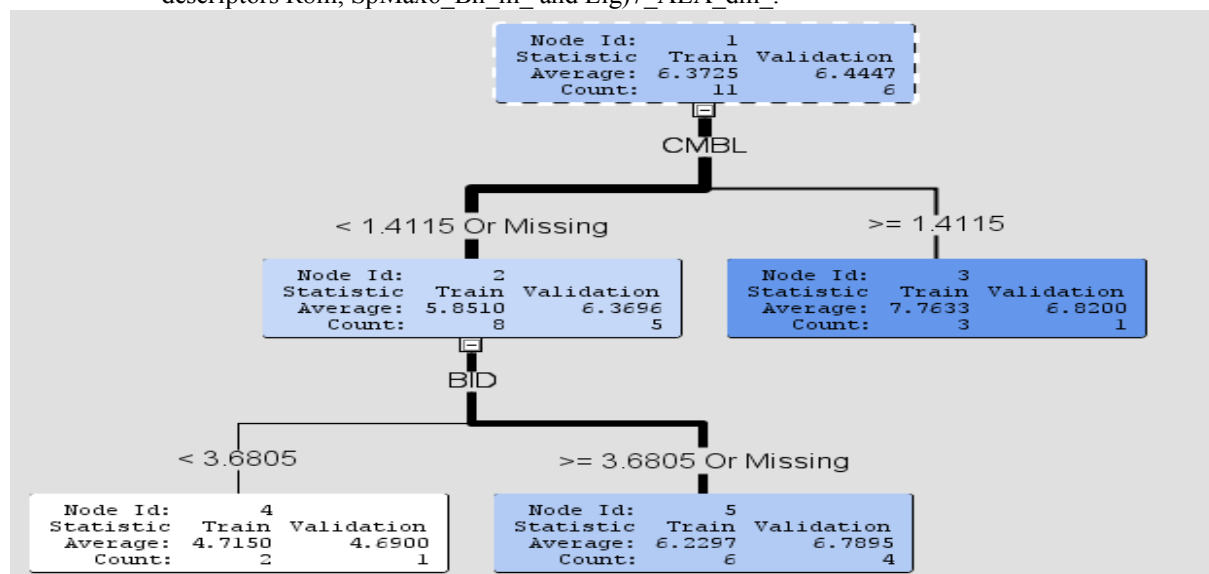
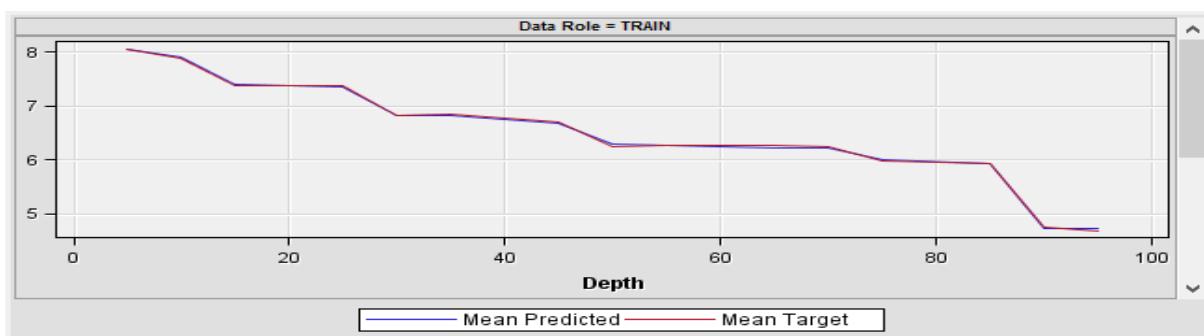


Figure 4: Decision tree model validation for erastin analogues with five nodes. The splitting rules are based on two descriptors CMBL and BID.



b)

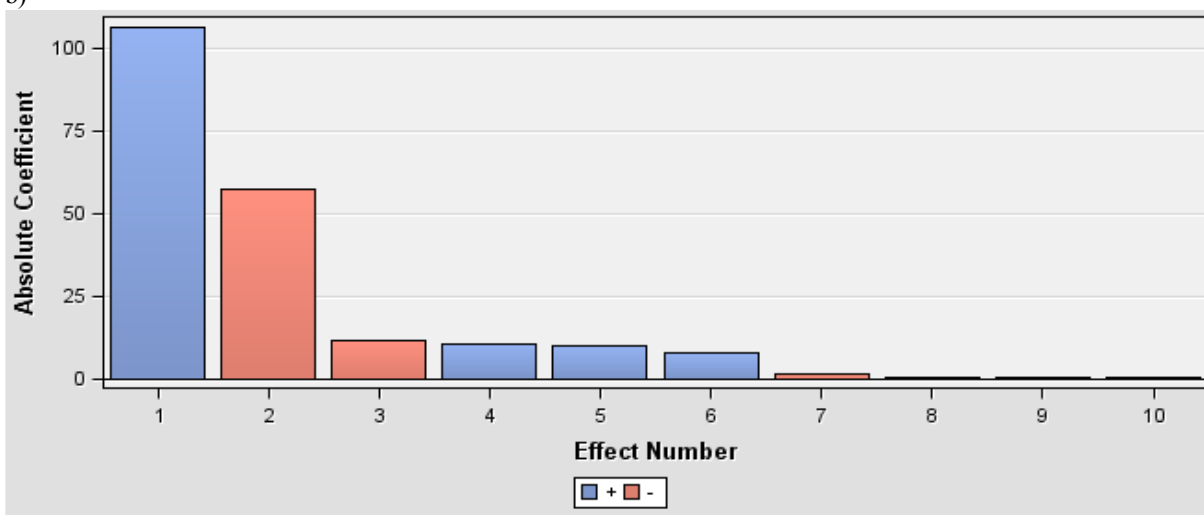


Figure 5: Plots resulting from regression validation model for erastin analogues. a) Score Rankings Overlay: Activity plot demonstrating moderately good agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = GI5, 2 = Intercept, 3 = Eig02_AEA_bo_, 4 = SpMax_B_e_, 5 = SM3_B_m_, 6 = TDB09r, 7 = Mor25v, 8 = SM13_AEA_dm_, 9 = SpAD_B_s_ and 10 = RDF125i.

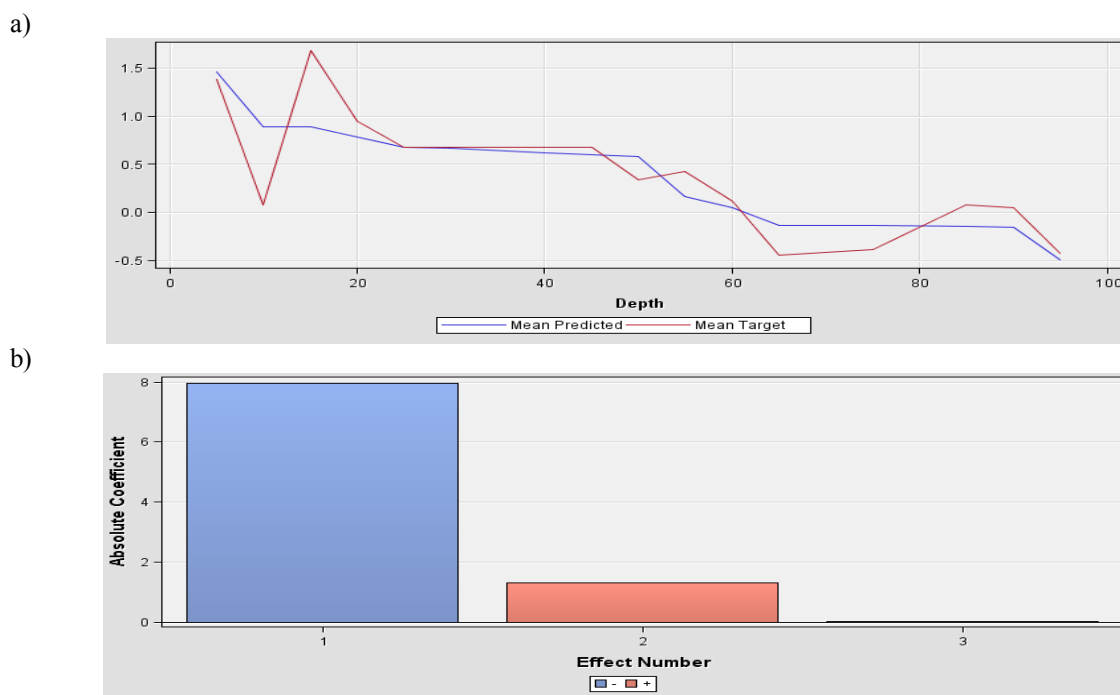


Figure 6. Plots resulting from stepwise linear regression model for amino acid analogues. a) Score Rankings Overlay: Activity plot demonstrating good agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = Mor28v, 2 = intercept, 3 = RDF065s.

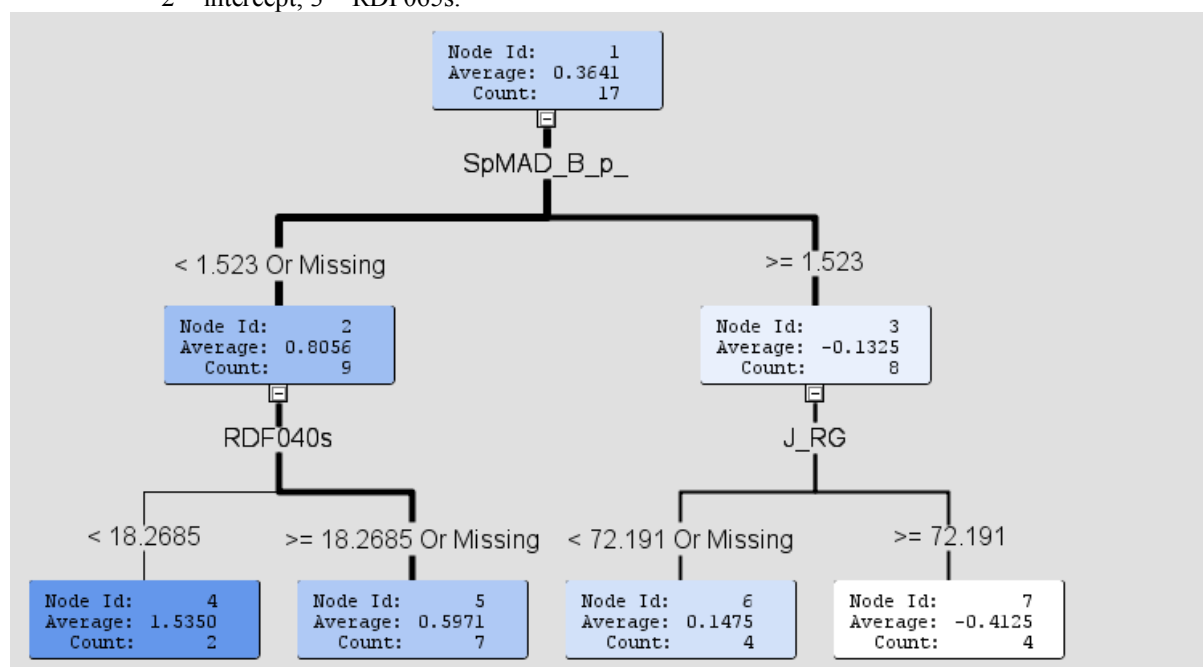


Figure 7. Decision tree model for amino acid analogues with seven nodes. The splitting rules are based on three descriptors SpMAD_B_p, RDF040s and J_RG.

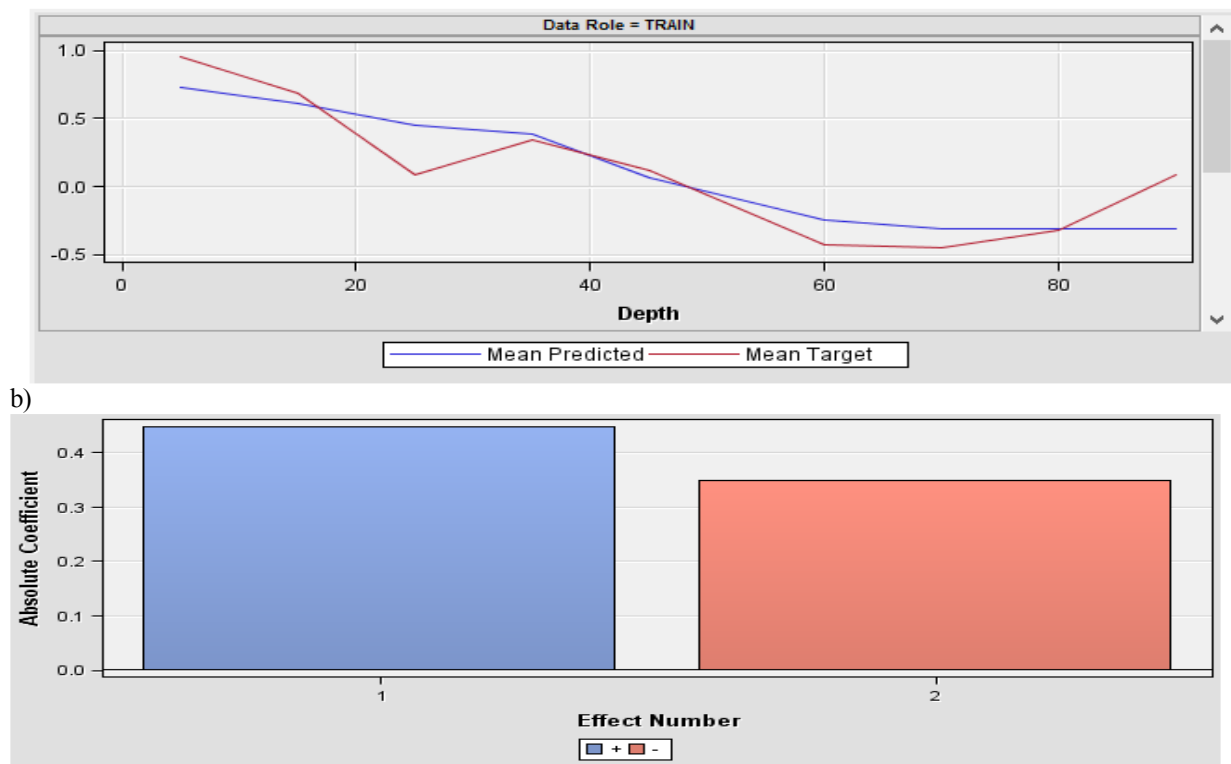


Figure 8: Plots resulting from regression validation model for Amino acid analogues. a) Score Rankings Overlay: Activity plot demonstrating moderately moderate agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = Intercept and 2 = Mor24s.

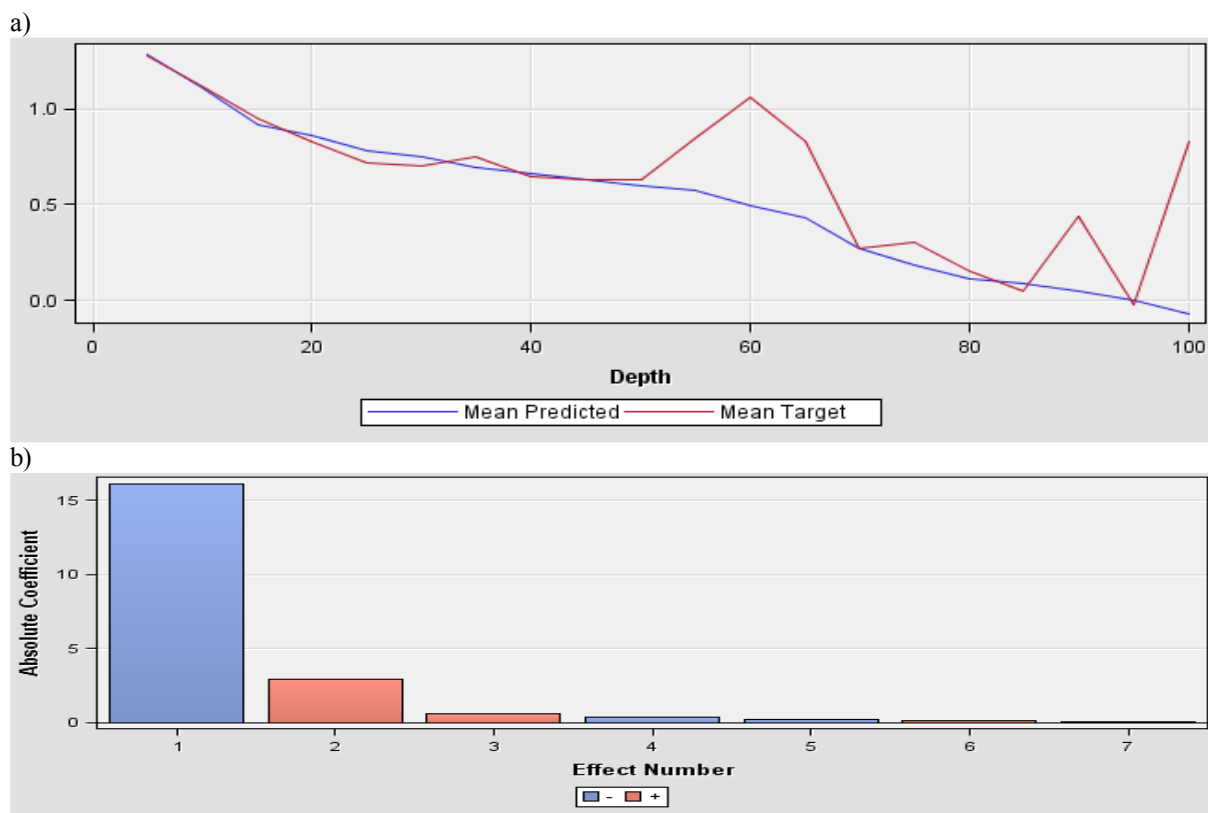


Figure 9: Plots resulting from stepwise linear regression model for hydantoin analogues. a) Score Rankings Overlay: Activity plot demonstrating moderately poor agreement between mean predicted and

mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = intercept, 2 = SpDiam_AEA_dm, 3 = VE1sign_X, 4 = Mor13i, 5 = Mor07u, 6 = Mor31s and 7 = P_VSA_LogP_5

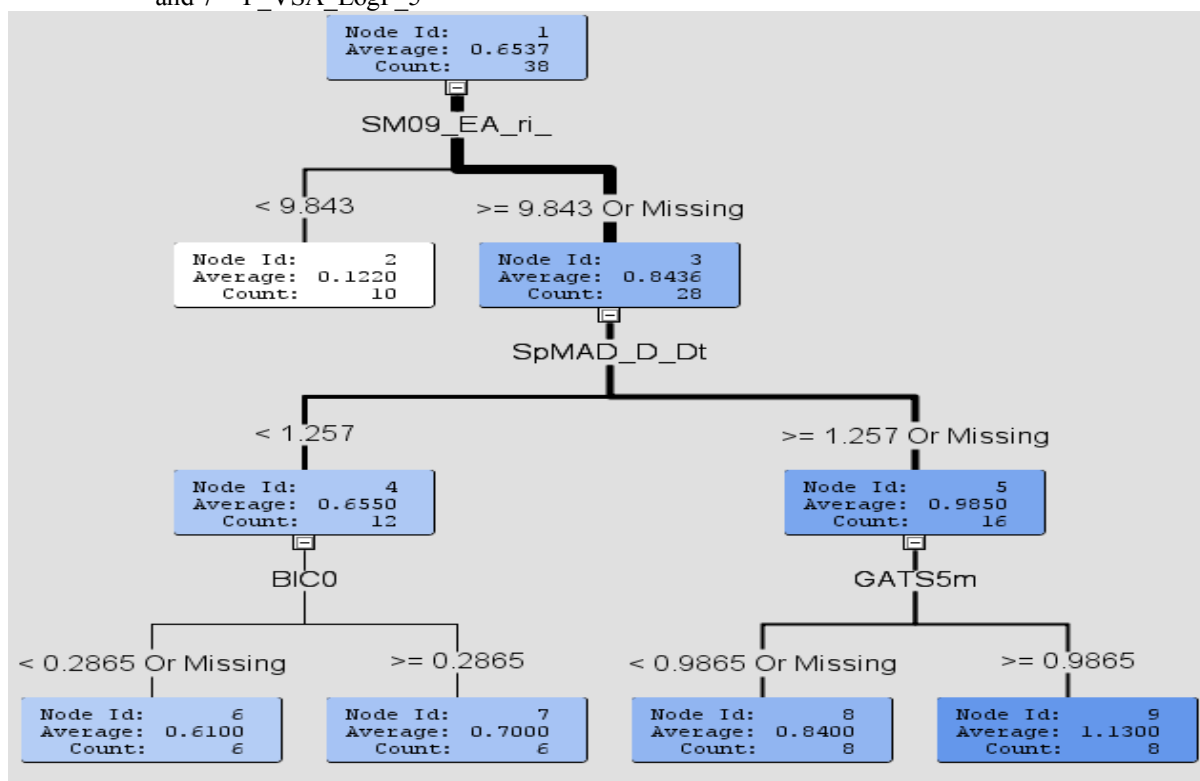


Figure 10. Decision tree model for hydantoin analogues with nine nodes. The splitting rules are based on four descriptors SM09_EA_ri_, SpMAD_D_Dt, BIC0 and GATS5m

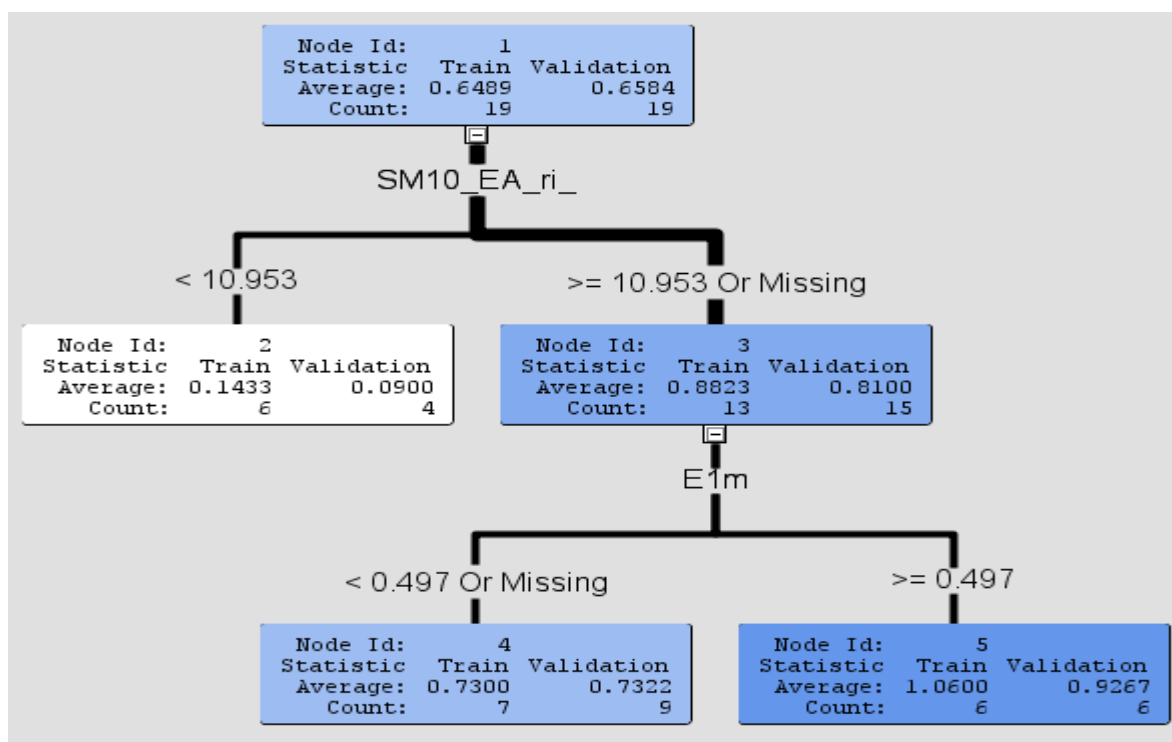


Figure 11: Decision tree model validation for hydantoin analogues with five nodes. The splitting rules are based on two descriptors SM10_EA_ri_ and E1m

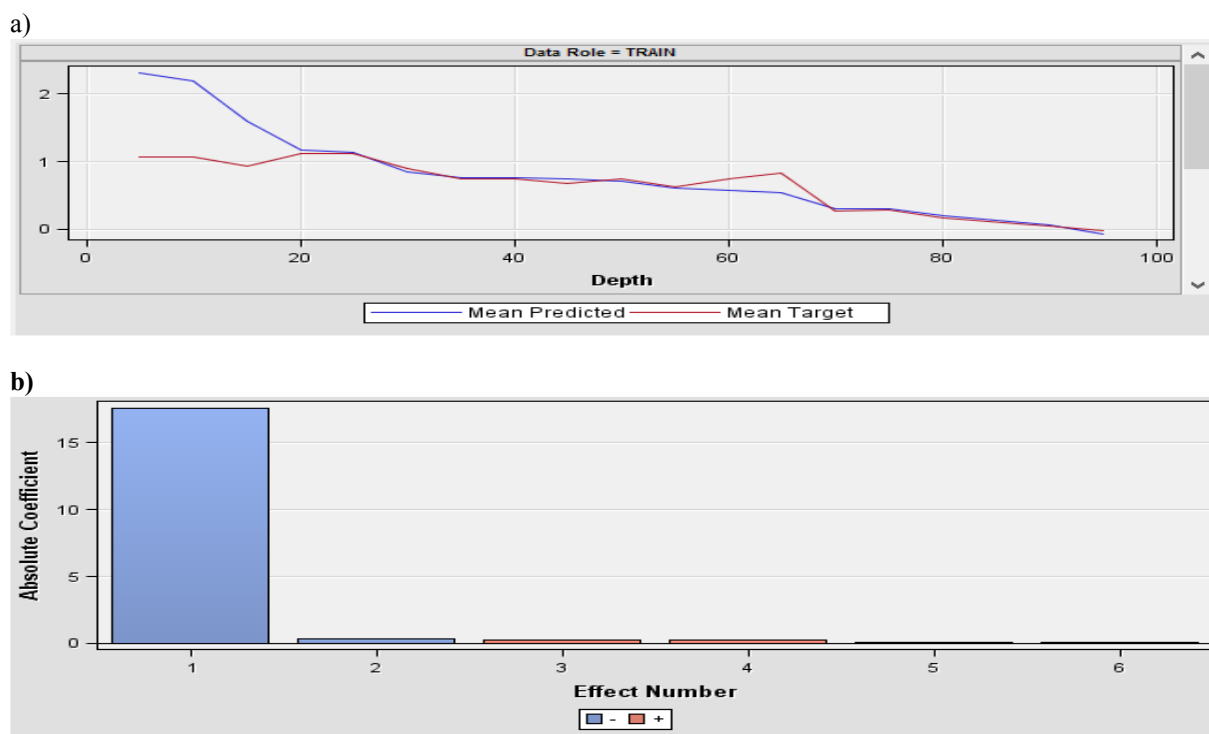


Figure 12: Plots resulting from regression validation model for hydantoin analogues. a) Score Rankings Overlay: Activity plot demonstrating moderately moderate agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 = R7i_, 2 = RDF080v, 3 = ATSC3P, 4 = Intercept, 5 = P_VSA_LOGP_5 and 6 = RDF070v.

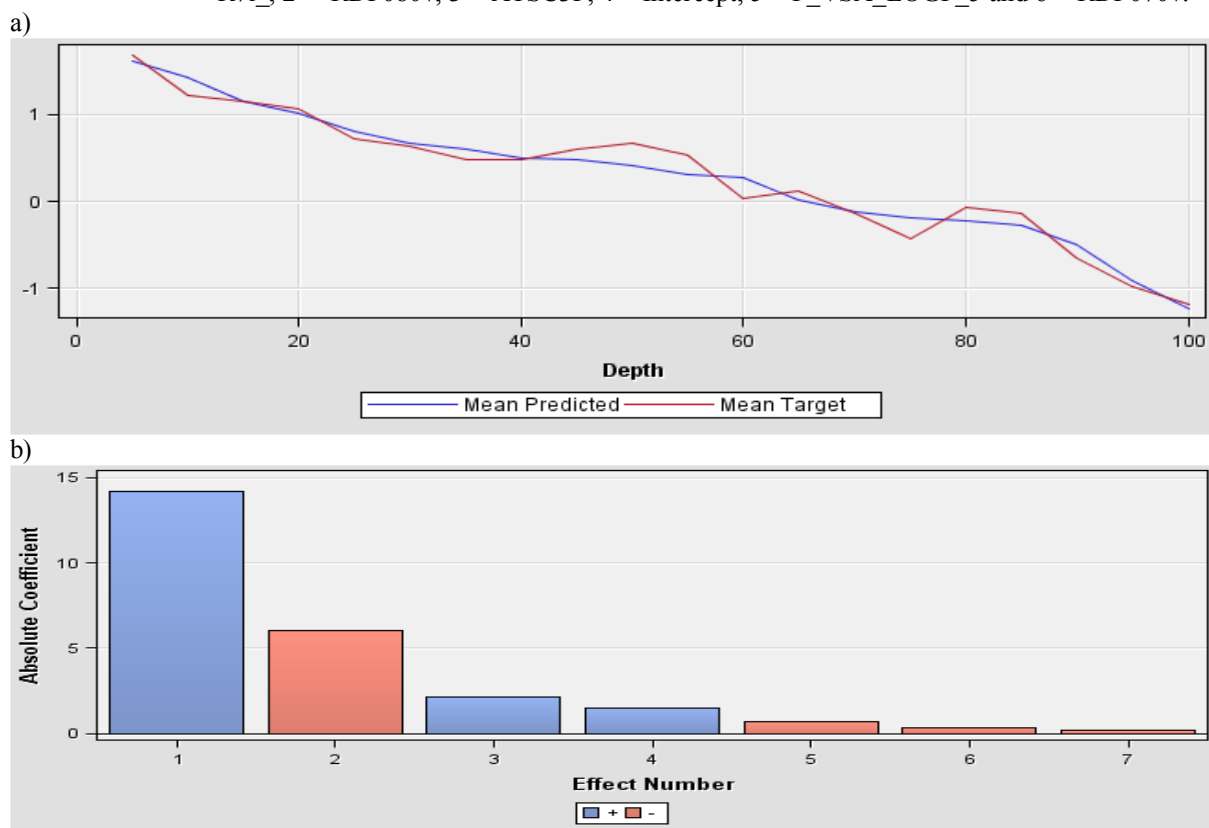


Figure 13: Plots resulting from stepwise linear regression model for isooazole analogues. a) Score Rankings Overlay: Activity plot demonstrating moderately poor agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model;

Here 1 = G_u , 2 = Intercept, 3 = GATS1i, 4 = R8m, 5 = Mor05u, 6 = Eig05_EA_ed_ and 7 = CATS2D_08_AL.

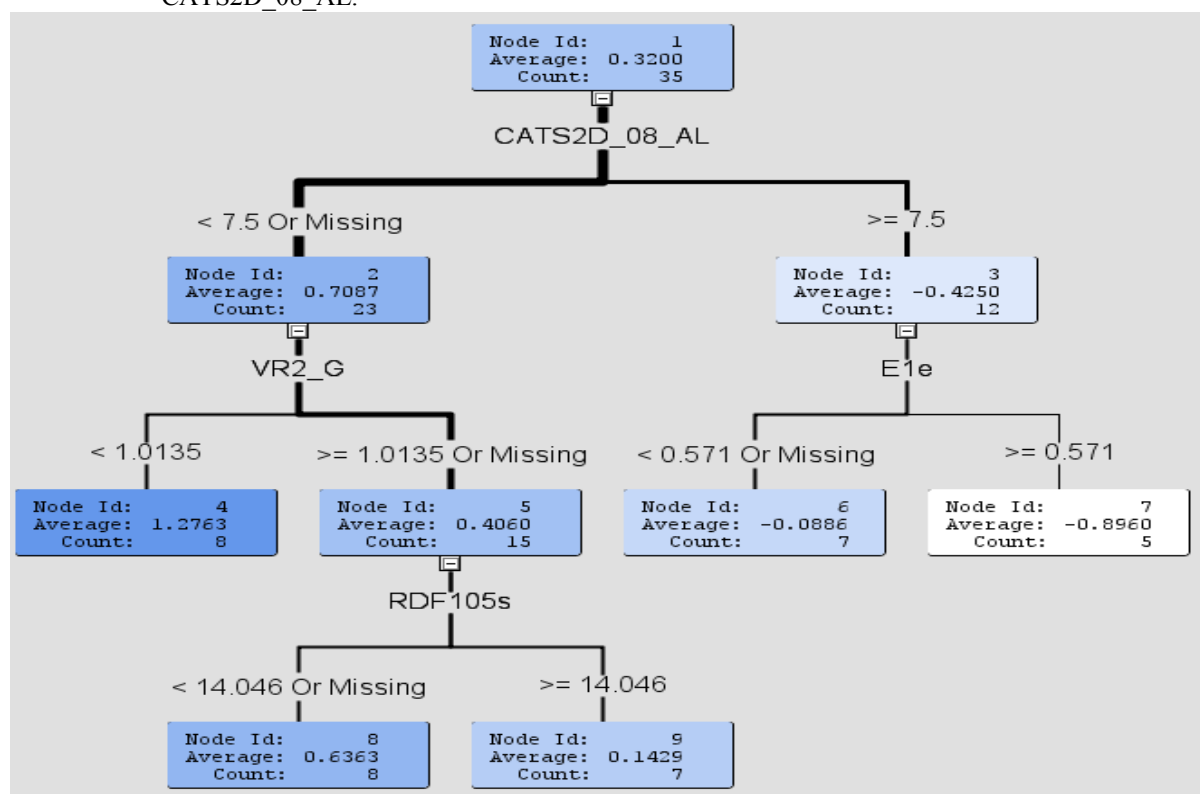


Figure 14: Decision tree model for isooxazole analogues with nine nodes. The splitting rules are based on four descriptors CATS2D_AL, VR2_G, E1e, RDF105s

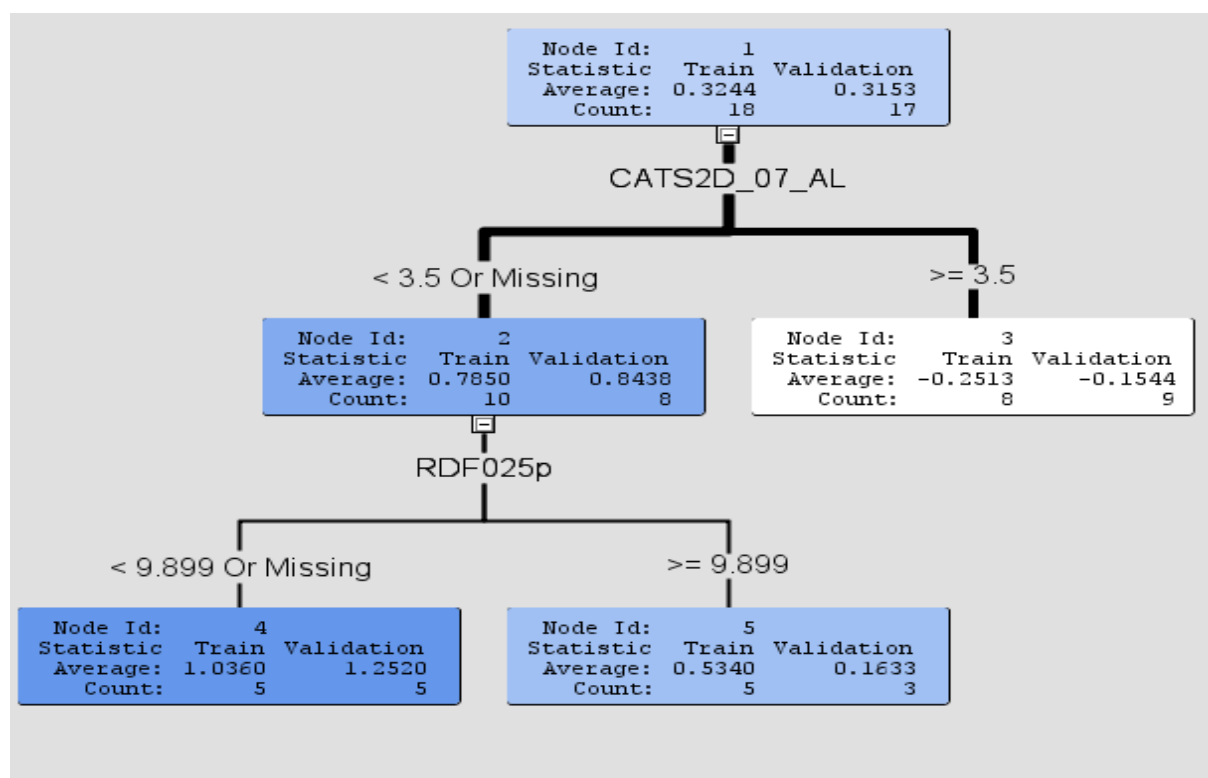


Figure 15: Decision tree model validation for isooxazole analogues with five nodes. The splitting rules are based on two descriptors CATS2D_07_AL and RDF025p.

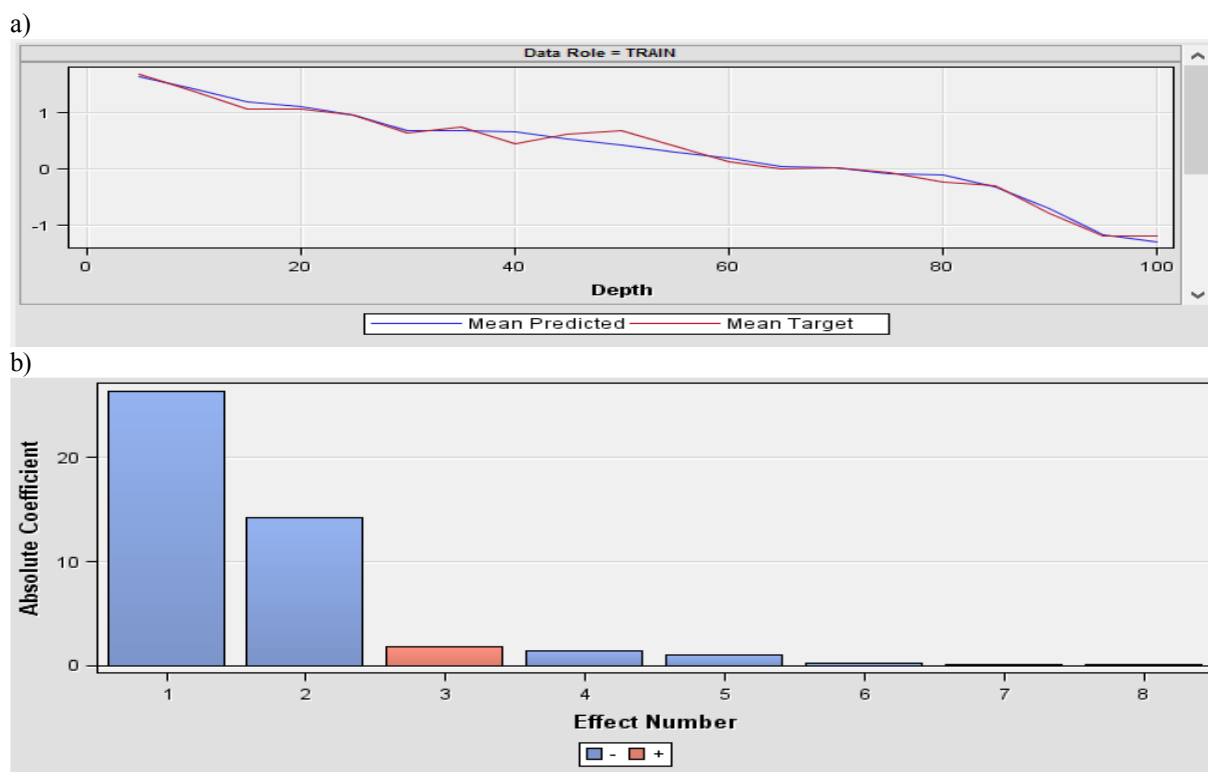


Figure 16. Plots resulting from regression validation model for isooxazole analogues. a) Score Rankings Overlay: Activity plot demonstrating moderate agreement between mean predicted and mean target and b) effects plot showing relative contributions of the input variables to the model; Here 1 =TDB01u, 2 = Gu, 3 = SM4_B_m, 4 = Intercept, 5 = CATS2D_08_AL, 6 = VE1_Dz_p_, 7 = CATS3D_07_LL and 8 = RDF090s.

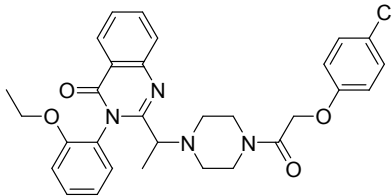
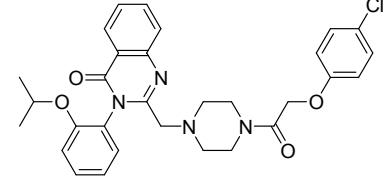
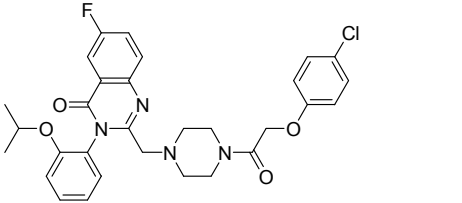
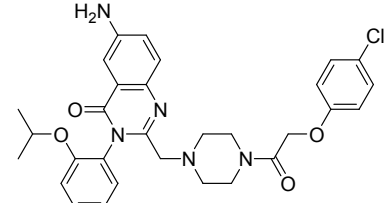
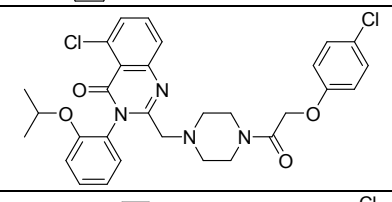
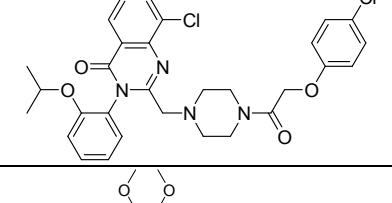
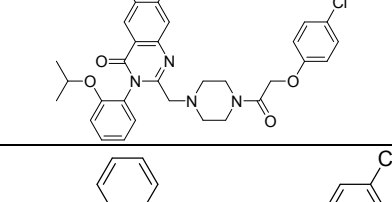
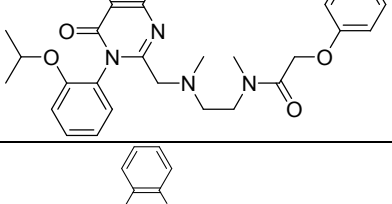
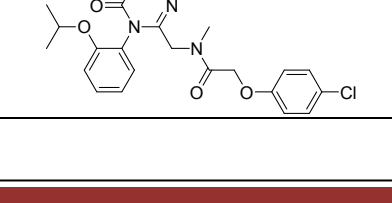
REFERENCES:

- 1.D. Patel, P.S. Kharkar PS, M. Nandave, *Mol. Membr. Biol.* 2015, **32**, 89-116.
- 2.S. Bannai, A. Takada, H. Kasuga, N. Tateishi, *Hepatology.* 1986, **6**, 1361-1368.
- 3.Y. Yang, D. Yee, *Cancer Res.* 2014, **74**, 2295-2305.
- 4.D.A. Baker, K. McFarland, R.W. Lake, H. Shen, X.C. Tang, S. Toda, P.W. Kalivas, *Nat Neurosci.* 2003, **6**, 743-749.
- 5.F. Verrey, E.I. Closs, C.A. Wagner, M. Palacin, H. Endou, Y. Kanai, *Pflugers Arch.* 2014, **447**, 532-542.
- 6.H. Sato, M. Tamba, T. Ishii, S. Bannai, *J Biol Chem.* 1999, **274**, 11455-11458.
- 7H. Sontheimer, *J. Neurochem.* 2008, **105**, 287-295.
- 8.S. Chintala, K. Tóth, M.B. Yin, A. Bhattacharya, S.B. Smith, M. S. Ola, Y.M. Rustum, *Chemotherapy.* 2010, **56**, 223-233.
- 9.J. Guan, M. Lo, P. Dockery, S. Mahon, C.M. Karp, A.R. Buckley, S. Lam, P.W. Gout, Y.Z. Wang, *Cancer Chemother Pharmacol.* 2009, **64**, 463-72.
- 10.E. Habib, K. Linher-Melville, H.X. Lin, G. Singh, *Redox Biol.* 2015, **5**, 33-42.
- 11.T. Ishimoto, O. Nagano, T. Yae, M. Tamada, T. Motohara, H. Oshima, M. Oshima, T. Ikeda, R. Asaba, H. Yagi, T. Masuko, T. Shimizu, T. Ishikawa, K. Kai, E. Takahashi, Y. Imamura, Y. Baba, M. Ohmura, M. Suematsu, H. Baba, H. Saya, *Cancer Cell.* 2011, **19**, 387-400.
- 12.M. Lo, V. Ling, Y.Z. Wang, P.W. Gout, *Br. J Cancer.* 2008, **99**, 464-72.
- 13.S. Okuno, H. Sato, K. Kuriyama-Matsumura, M. Tamba, H. Wang, S. Sohda, H. Hamada, H. Yoshikawa, T. Kondo, S. Bannai *Br. J Cancer.* 2003, **88**, 951-956.
- 14.S.J. Dixon, D.N. Patel, M. Welsch, R. Skouta, E.D. Lee, M. Hayano, A.G. Thomas, C.E. Gleason, N.P. Tatonetti, B.S. Slusher, B.R. Stockwel, *Elife.* 2014, **20**, e02523.
- 15.K. Shukla, A.G. Thomas, D.V. Ferraris, N. Hin, R. Sattler, J. Alt, C. Rojas, B.S. Slusher, T. Tsukamoto, *Bioorg Med Chem Lett.* 2011, **21**, 6184-7.
- 16.S.A. Patel, T. Rajale, E. O'Brien, D.J. Burkhart, J.K. Nelson, B. Twamley, A. Blumenfeld, M.I. Szabon-Watola, J. M. Gerdes, R.J. Bridges, N.R. Natale, *Bioorg Med Chem.* 2010, **18**, 202-13.
- 17.A.A. Matti, J. Mirzaei, J. Rudolph, S.A. Smith, J.L. Newell, S.A. Patel, M.R. Braden, R.J. Bridges,

N.R. Natale, *Bioorg Med Chem Lett.* 2013, **23**, 5931-5.
18.J.L. Newell, C.M. Keyari, S.W. McDaniel, P.J. Diaz, N.R. Natale, S.A. Patel, R.J. Bridges, *Neurochem Int.* 2014, **73**, 132-138.
19.S.K. Ahmed, J-L.G. Etoga, S.A. Patel, R.J. Bridges, C.M. Thompson, *Bioorg Med Chem Lett.* 2011, **21**, 4358-4362.
20.J-L.G. Etoga, S.K. Ahmed, S. Patel, R.J. Bridges, C.M. Thompson. *Bioorg Med Chem Lett.* 2010, **20**, 2680-2683.
21.Schrödinger Release 2016-4: LigPrep, Schrödinger, LLC, New York, NY, 2016.

22.I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko. *J Comput Aided Mol Des.* 2005, **19**, 453-463.
23.Todeschini R, Consonni V. volume 41 (2 volume set). John Wiley & Sons; 2009 Oct 30.
24.SAS Institute Inc 2013. Getting Started with SAS® Enterprise Miner™ 13.1. Cary, NC: SAS Institute Inc.
25.L. Saiz-Urra, M.P. Gonzalez, and M. Teijeira, *Bioorg. Med. Chem.* 2006, **14**, 7347-7358.
26.N.S.H.N. Moorthy, C. Karthikeyan, P. Trivedi, *Indian J. Chem.* 2007, **46**, 177-184.

Table 1: Chemical structures and Sxc¹ inhibitory activity of erastin analogues [14]

Sr. No.	Code	Structure	IC ₅₀	pIC ₅₀
1	Erastine		0.20	6.698
2	ERA 1		0.09	7.04
3	ERA 2		0.14	6.85
4	ERA 3		0.54	6.26
5	ERA 4		0.042	7.37
6	ERA 5		0.074	7.13
7	ERA 6		0.15	6.82
8	ERA 7		0.56	6.25
9	ERA 8		>10	4.69

10	ERA 9		0.57	6.24
11	ERA 10		3.7	5.43
12	ERA 11		0.53	6.27
13	ERA 12		1.2	5.92
14	ERA 13		2.6	5.58
15	ERA 14		>10	4.69
16	ERA 15		0.042	7.37
17	ERA 16		0.0094	8.02
18	ERA 17		0.0035	8.45
19	ERA 18		0.013	7.88

Table 2: Chemical structures and Sxc⁻ inhibitory activity of amino acid analogues [20]

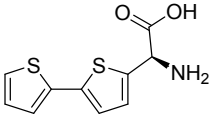
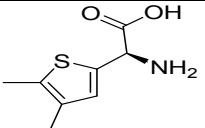
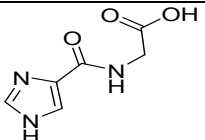
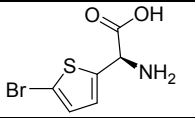
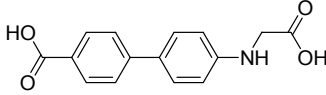
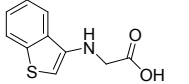
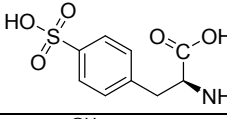
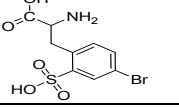
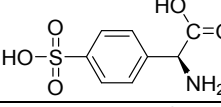
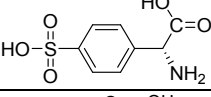
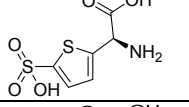
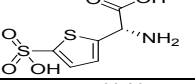
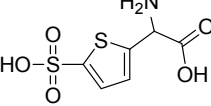
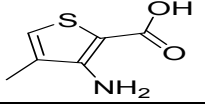
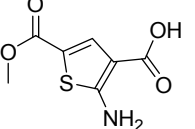
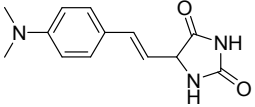
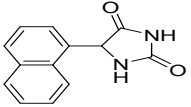
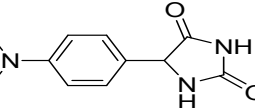
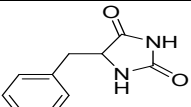
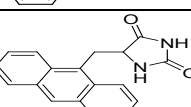
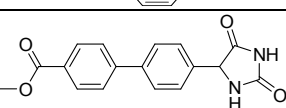
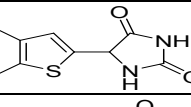
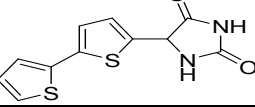
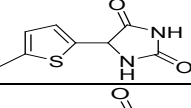
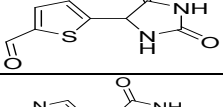
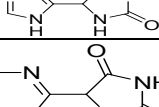
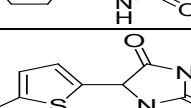
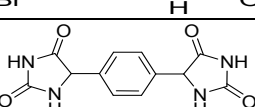
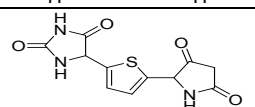

Sr. No	Code	Structure	% glutamate uptake inhibition	Logit values
1	AA 1		31	-0.347
2	AA 2		27	-0.432
3	AA 3		4	-1.380
4	AA 4		17	-0.689
5	AA 5		73	0.432
6	AA 6		47	-0.052
7	AA 7		10	-0.954
8	AA 8		17	-0.689
9	AA 9		2	-1.690
10	AA 10		45	-0.087
11	AA 11		68	0.327
12	AA 12		45	-0.087
13	AA 13		74	0.454
14	AA 14		17	-0.689
15	AA 15		43	-0.122

Table 3: Chemical structures of hydantoin analogues along with respective biological activity [19]

Hydantoin derivatives				
SR no	Code	Structure	% of glu Uptake inhibition	Logit transforms
1	HYD 1		10	-0.954
2	HYD 2		17	-0.689
3	HYD 3		13	-0.826
4	HYD 4		47	-0.052
5	HYD 5		19	-0.630
6	HYD 6		7	-1.123
7	HYD 7		13	-0.826
8	HYD 8		15	-0.753
9	HYD 9		34	-0.288
10	HYD 10		51	0.017
11	HYD 11		18	-0.659
12	HYD 12		5	-1.279
13	HYD 13		35	-0.269
14	HYD 14		15	-0.753
15	HYD 15		49	-0.017

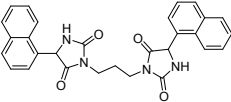
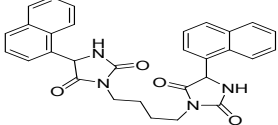
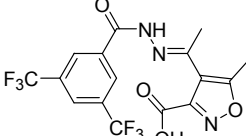
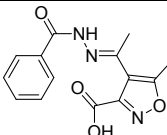
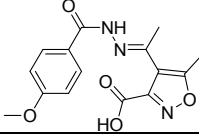
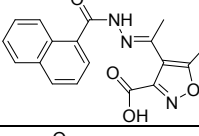
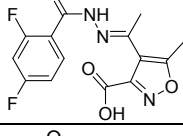
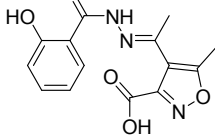
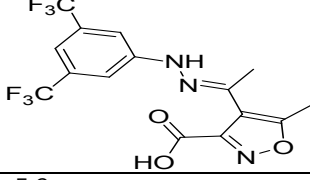
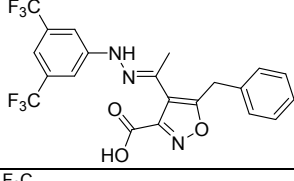
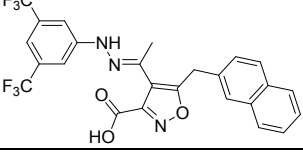
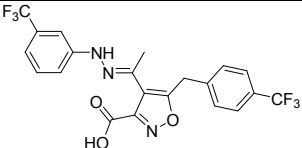
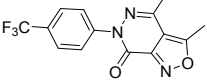
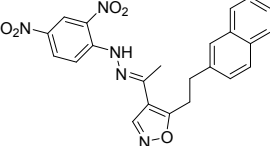
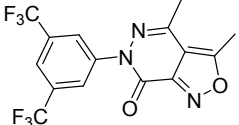
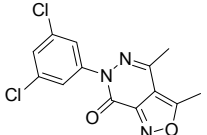
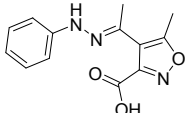
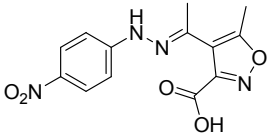
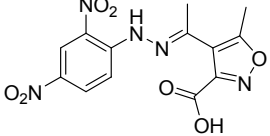
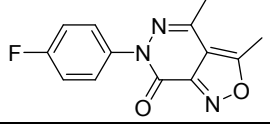
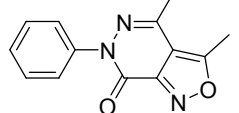
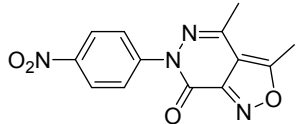
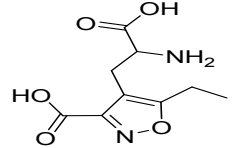
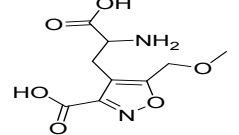
16	HYD 16		20	-0.602
17	HYD 17		8	-0.954

Table 4: Chemical structures of iso-oxazole analogues along with respective biological activity [16-19]

Sr. No.	Code	Structure	% [3H] glu uptake inhibition	Logit transforms
1	ISO 1		10	-0.954
2	ISO 2		17	-0.689
3	ISO 3		32	-0.327
4	ISO 4		32	-0.327
5	ISO 5		42	-0.140
6	ISO 6		48	-0.035
7	ISO 7		86	0.788
8	ISO 8		94	1.195
9	ISO 9		86	0.788

10	ISO 10		94	1.195
11	ISO 11		86	0.788
12	ISO 12		55	0.087
13	ISO 13		18	-0.659
14	ISO 14		11	-0.908
15	ISO 15		44	-0.105
16	ISO 16		78	0.550
17	ISO 17		55	0.087
18	ISO 18		14	-0.788
19	ISO 19		26	-0.454
20	ISO 20		15	-0.753
21	ISO 21		8	-1.061
22	ISO 22		2	-1.690

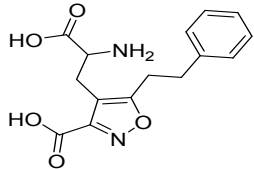
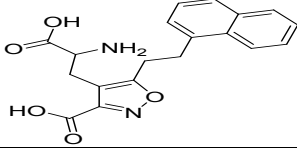
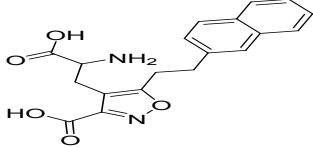
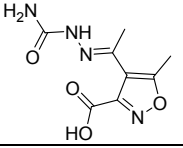
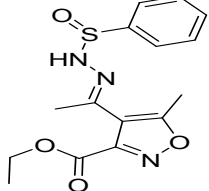
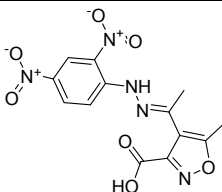
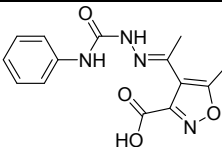
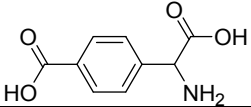
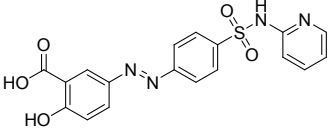
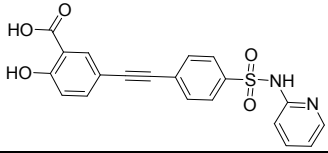
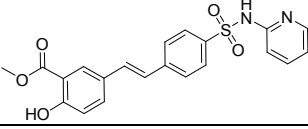
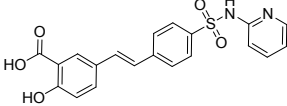
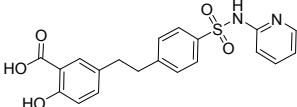
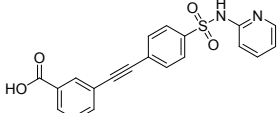
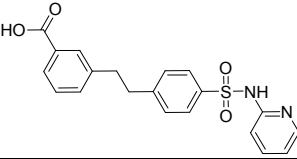
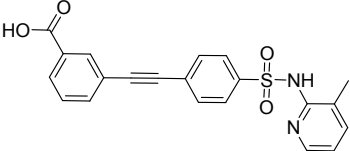
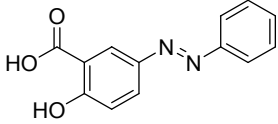
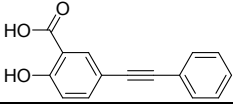
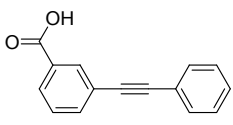
23	ISO 23		30	-0.368
24	ISO 24		53	0.052
25	ISO 25		63	0.231
26	ISO 26		9	-1.005
27	ISO 27		4	-1.380
28	ISO 28		28	-0.410
29	ISO 29		24	-0.501

Table 5: Chemical structures of sulfasalazine and its analogues along with respective biological activity [15]

Sr. No.	Code	Structure	IC ₅₀ (μM)	pIC ₅₀
1	(S)-4-carboxyphenylglycine		15	4.82
2	SSZ		30	4.52
3	SSZ 1		30	4.52
4	SSZ 2		900	3.04
5	SSZ 3		70	4.15
6	SSZ 4		700	3.15
7	SSZ 5		75	4.12
8	SSZ 6		920	3.03
9	SSZ 7		70	4.15
10	SSZ 8		200	3.69
11	SSZ 9		150	3.82
12	SSZ 10		270	3.56