Cardiac Toxicity – hERG Potassium Channel Blocking Affinity

The hERG (human Ether-à-go-go Related Gene) encodes potassium channels that are responsible for the normal repolarization of the cardiac action potential. Blockage or any other impairment of these channels in heart cells can lead to fatal cardiac arrhythmias associated with prolongation of the QT interval. Drug-induced blockage of potassium channels has been a major concern for the pharmaceutical industry, contributing to half of the safety-related recalls of drugs from the market since 1997 and impacting all therapeutic drug classes. Testing for hERG blockage, EADs and *Torsades de Pointes* are now mandatory for pharmaceutical discovery and development groups.

Two artificial neural network ensemble (ANNE) models (classification (hERG_Filter) and regression (hERG_pIC50)) are provided in ADMET Predictor's Toxicity module to assess the likelihood that a compound will bind to the hERG channel. The classification model indicates whether the compound is likely to have problematic affinity for the hERG K+ channel. Thus, it provides a "Yes" or "No" answer to the question, "Is this compound a potential hERG blocker?" The hERG regression model quantitatively predicts the pIC50 (-log(IC_{50} [M]) inhibition of the hERG channel. Larger pIC50 values indicate higher inhibition and larger safety concern.

hERG_Filter Model

There are 1,024 records in the data set used to build the hERG_Filter model. For compounds with experimental hERG IC_{50} values, a threshold of 10 µM was used to classify "nonbinders" from "binders", i.e., an IC_{50} greater than 10 µM was categorized as a "nonbinder", while those with an IC_{50} less than or equal to 10 µM were categorized as a "binder". Overall, there are 420 "nonbinders" and 604 "binders" in the data set. The molecular weight (MWt) values in the data set range from 78 to 880 amu and the S+logP values range from -2.5 to 10.7. In the MWt distribution plot below, the "binders" appear to be slightly skewed towards higher molecular weight. The median MWt is 422 for "binders" and 415 for "nonbinders".

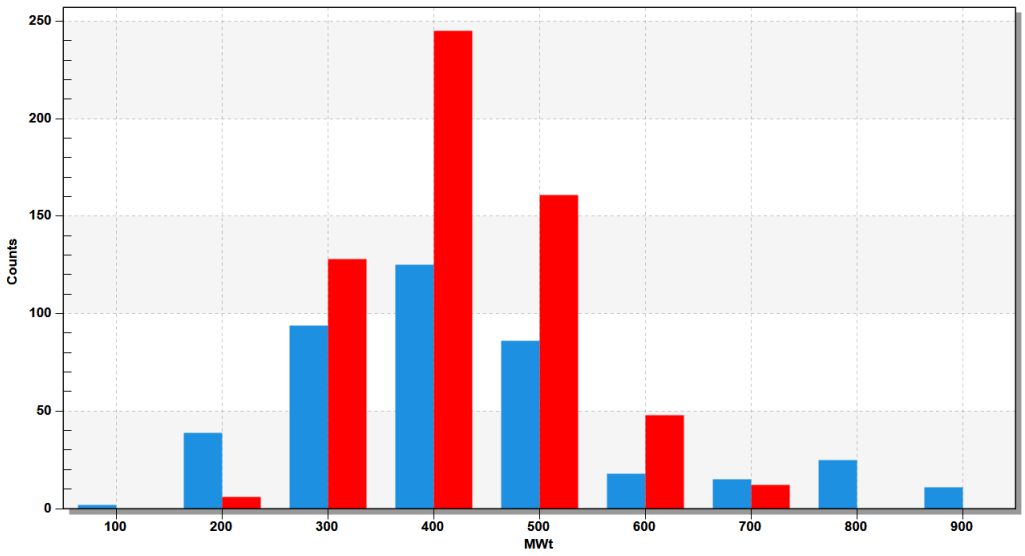


Figure 1 – MWt distribution of the hERG classification data set. The blue bars are "nonbinders" and the red bars are "binders".

The S+logP distribution below shows a much larger discrepancy between "nonbinders" and "binders". The hERG positive compounds are skewed towards much higher S+logP. Here, the median value for "nonbinders" is 2.6 whereas the median value for "binders" is 4.0.

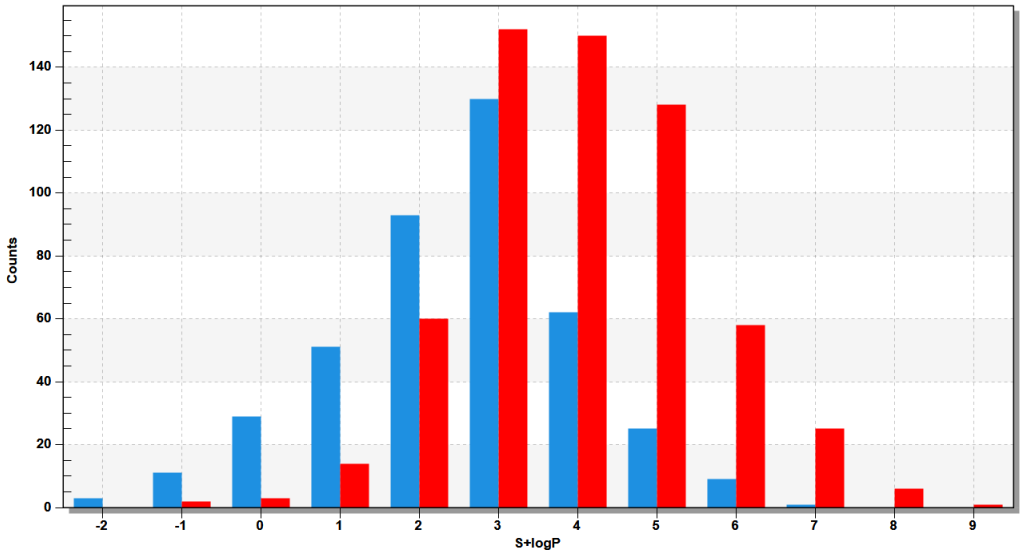


Figure 2 - S+logP distribution of the hERG classification data set. The blue bars are "nonbinders" and the red bars are "binders".

The classic pharmacophore for the hERG channel consists of a protonatable amine and one or more aromatic rings (Cavalli 2002). FCation is the fraction of cationic species of a molecule at pH 7.4. The distribution plot of FCation is shown below. A large portion of the compounds with FCation > 0.6 are hERG "binders". However, there are also a substantial number of "binders" that have an FCation value of zero. The hERG positive compounds have a median FCation value of 0.79 whereas the "nonbinders" have a median FCation value of 0.03.

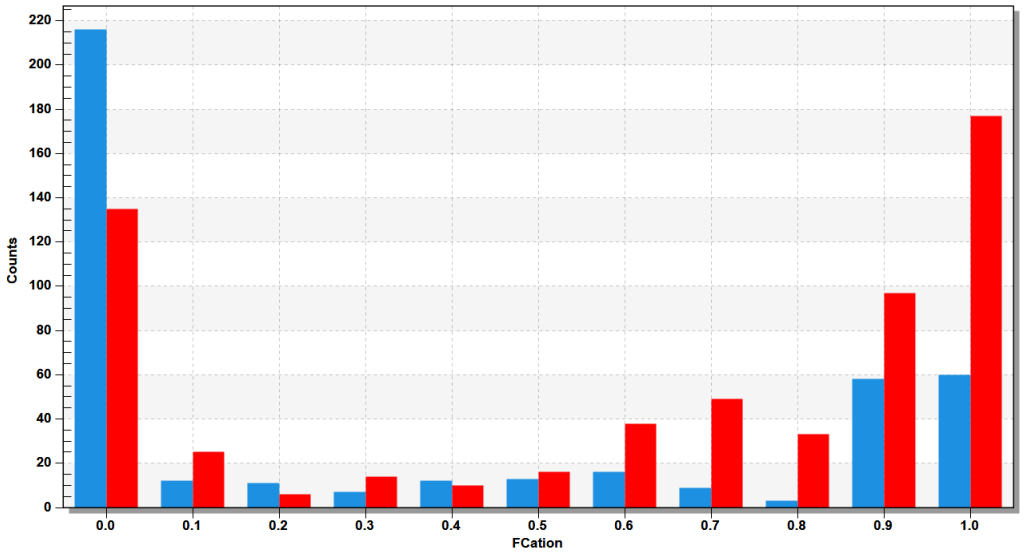


Figure 3 - FCation distribution of the hERG classification data set. The blue bars are "nonbinders" and the red bars are "binders".

A matched molecular pair analysis (MMPA) was performed on the 1,024 compounds in the classification data set. The analysis criterion was somewhat strict - the maximum number of atoms and bonds that are not a match was set to 6. Also, the pairs also needed to represent a difference in their categorization, e.g., one "nonbinder" and one "binder". This resulted in 195 pairs. The pairs were then clustered to identify common transformations. The figure below shows a few examples from a cluster of 8 pairs with an aromatic carbon (the red highlighted C) to an aromatic nitrogen (the blue highlighted N) transformation. The transformation results in conversion of a "binder" to a "nonbinder". The first example converts a pyridine to a pyrimidine ring. This conversion results in a reduction in S+logP. This is consistent with the S+logP distribution above which indicates that compounds with lower logP are less likely to inhibit hERG. In fact, 7 out of the 8 pairs show a substantial decrease in S+logP. The exception is found in the second row of
Figure **4**, where conversion of imidazole to triazole results in slightly higher S+logP. Note, that the hERG IC_{50} values of each molecule in the pair could be similar i.e. the "binder" might have an IC_{50} of 9 µM and the "nonbinder" might have an IC_{50} of 11 µM. MMPA of the classification data set is helpful in identifying compounds that are difficult for the model to correctly classify. It can also identify molecules where the underlying data needs to be scrutinized.

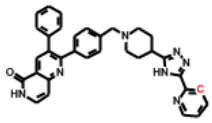
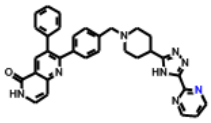
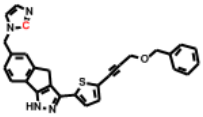
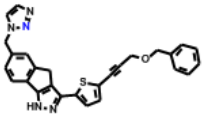
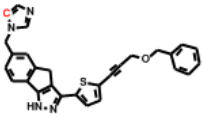
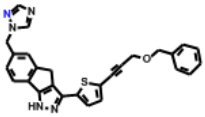
Structure 1	Structure 2	Binds 1	Binds 2	Change(Binds)	Change(S+logP)
		Toxic	Nontoxic	-1.000	-0.603
		Toxic	Nontoxic	-1.000	0.098
		Toxic	Nontoxic	-1.000	-0.669

Figure 4 - three examples from a cluster of 8 pairs with an aromatic carbon transformed into an aromatic nitrogen. Each molecule on the left is a "binder" and the molecule on the right is a "nonbinder".

The figure below shows pairs of compounds that differ by a methyl group. Three out of the four pairs go from "nonbinders" to "binders" when a methyl group is removed. This type of small structural change that results in a different category can be challenging to model. A couple of these pairs are discussed in the section on the hERG_pIC50 model.

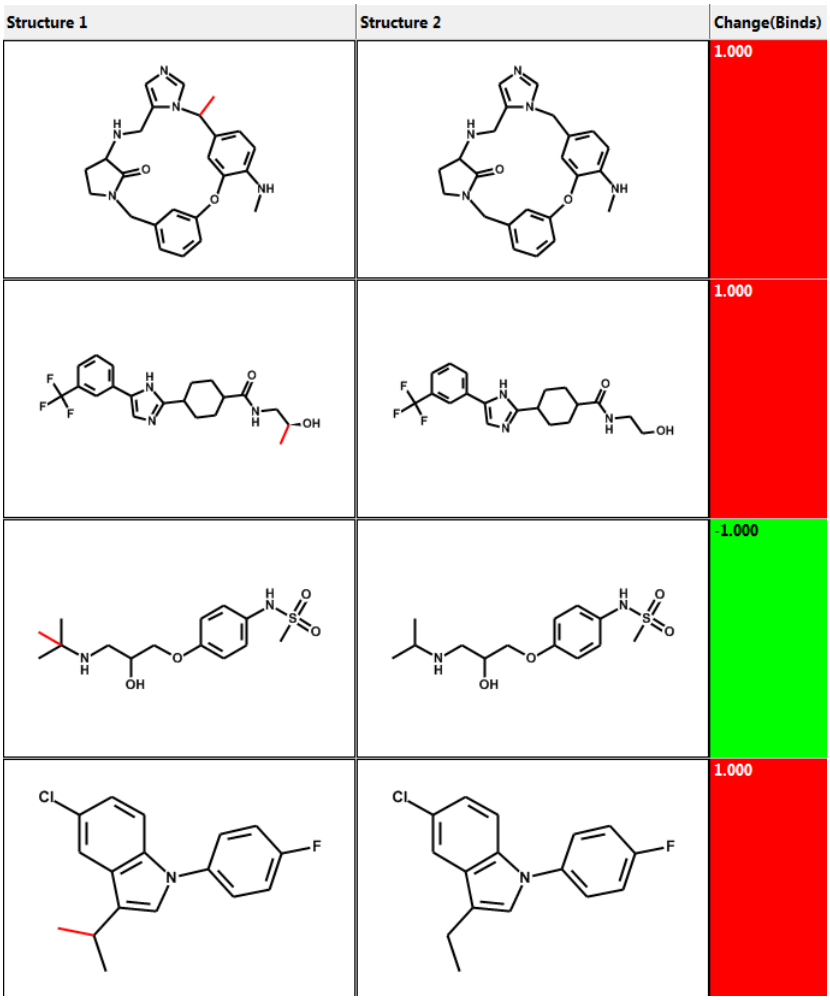


Figure 5 - Four pairs of compounds where a methyl group is removed.

The results of the classification model are shown in the table below. Approximately 20% of the data set was used as a test set. Each set has a higher percentage of "binders" than "nonbinders". The specificity, sensitivity, and concordance are all very high indicting an excellent model.

Model	Set	Negatives	Positives	Total	Correct	Concordance	Sensitivity	Specificity
hERG_Filter	Training	344	467	811	717	88.4%	88.2%	88.7%
	Test	76	137	213	184	86.4%	86.9%	85.5%

Table 1 - Performance of the hERG classification model (hERG_Filter) for the training and test sets.

How did the model perform on the compounds in Figure 5? The results are shown below in Figure 6. The model consistently predicted the same result for each compound in the pair. Thus, one prediction was incorrect. The confidence scores, shown in parenthesis, are fairly low. This reflects the difficulty in correctly classifying each compound. The exception is indole containing compounds where the confidence is >90%. How can this be explained? A similarity search of the data set, using the compound with the ethyl substituent as the query, resulted in 11 neighbors shown in Figure 7. Unfortunately, there is not a simple answer. There are six molecules categorized as "binders" and five as "nonbinders".

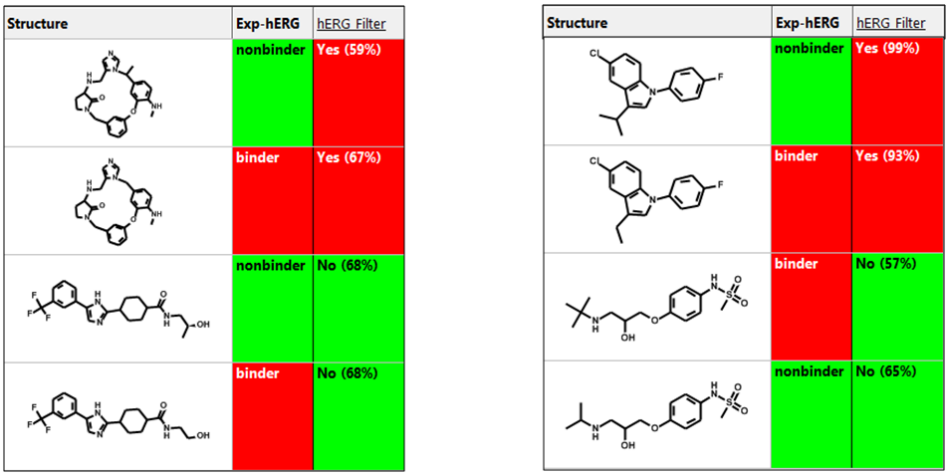


Figure 6 - Experiment and predicted hERG classification of pairs that differ by a methyl group. The confidence estimate for each prediction is in parenthesis.

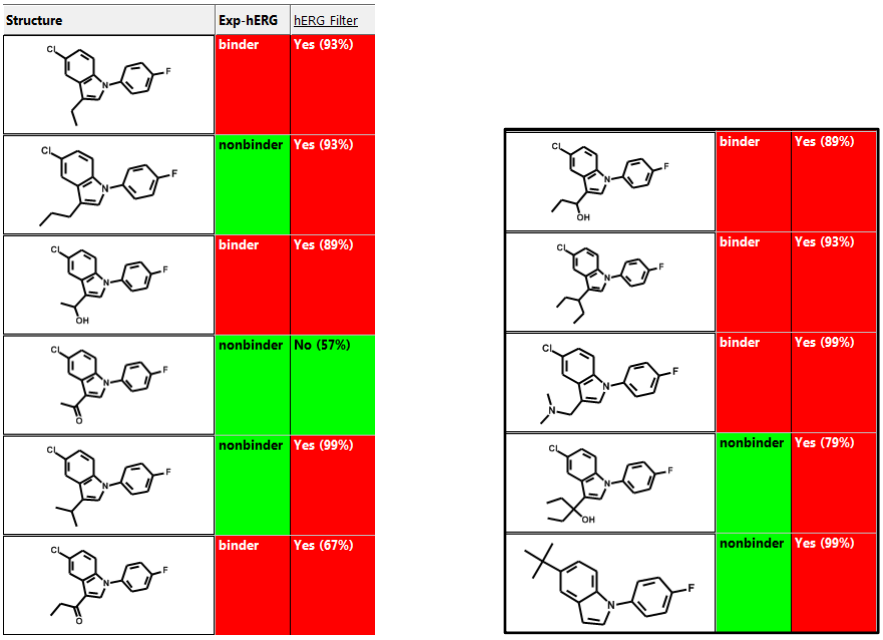


Figure 7 - Similarity search results. The query molecule is the first row in the image on the left.

hERG pIC50 Model

The hERG_pIC50 model predicts the pIC50 value of a compound where IC_{50} is given in molar (M) units of concentration. A combination of patch-clamp and ligand displacement data obtained from mammalian cell lines were used, the majority of measurements having been made on human embryonic kidney [HEK] or Chinese hamster ovary [CHO] cells stably transfected with hERG. The ligand displacement data was taken from studies similar to those of (Diaz 2004) – i.e., protocols that provide IC_{50}s commensurate with

those obtained from patch-clamp experiments. Similar models were built using individual cell lines and combinations thereof, but a model combining data from all the mammalian cell lines performed best when evaluated against an external proprietary test set of over 300 compounds evaluated by a major pharmaceutical company.

The data set contains 746 records that are a subset of the hERG classification data set. A distribution of the experimental pIC50 values is shown in Figure **8**. The most active compound is astemizole with a pIC50 of 9.046. Note the protonatable amine (S+pKa = 7.97) and the aromatic rings. The least potent compound is mannitol with a pIC50 of 0.967. The molecular weight (MWt) values in the data set range from 94 to 880 amu and the S+logP values range from -2.5 to 10.7. A plot of hERG pIC50 versus S+logP is shown in Figure 9. The red points represent "binders" from the hERG classification model. One can clearly see the pIC50 threshold at 5. There is a trend of higher pIC50 with increasing S+logP.

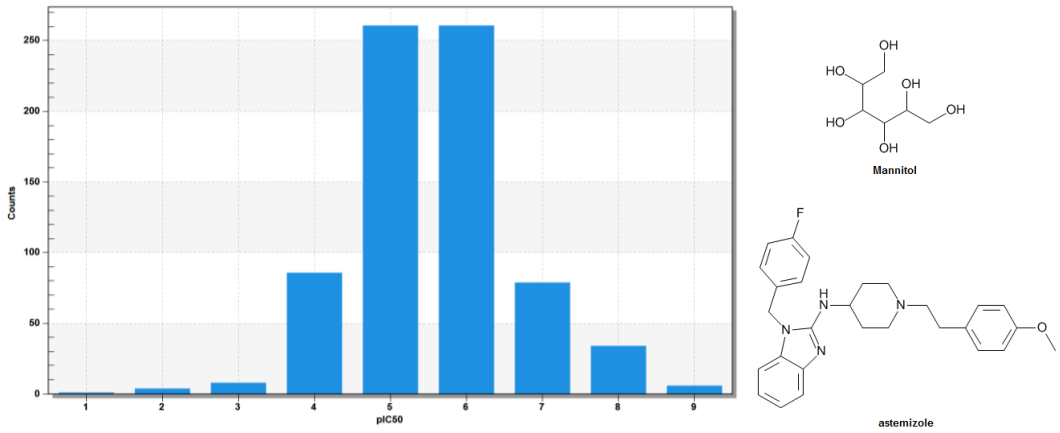


Figure 8 - Distribution of experimental hERG pIC50 values. Structures of the most potent compound, astemizole, and the least potent, mannitol are displayed.

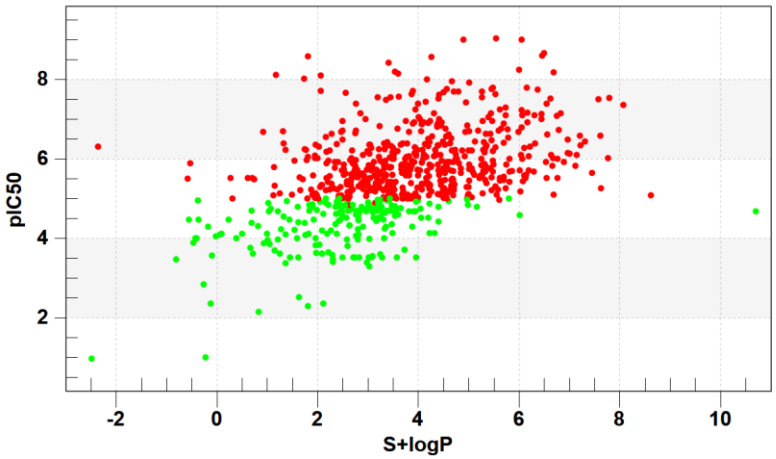


Figure 9 - hERG pIC50 versus S+logP. The red points are categorized as "binders" in the hERG classification model.

Matched Molecular Pair Analysis

Two of the four pairs in Figure 5, above, are in the regression data set. The pairs are shown below along with their hERG pIC50 values and their corresponding differences. All of the compounds have pIC50 values around the threshold value of 5, i.e., hERG IC_{50} = 10 µM. It is usually difficult for classification models to correctly predict the observed category when the molecule is close to the threshold. This "difficulty" is typically reflected in a low confidence estimate.

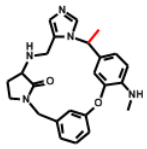
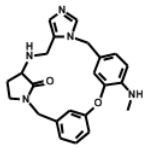
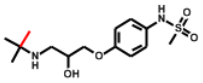
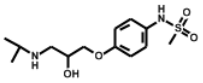
Structure 1	Structure 2	Exp-hERG 1	Exp-hERG 2	pIC50 1	pIC50 2	Change(pIC50)
		nonbinder	binder	4.848	5.105	0.257
		binder	nonbinder	5.523	4.481	-1.041

Figure 10 - Two pairs of compounds where adding or removing a methyl group changes the category.

Three matched molecular pairs where identified using a strict matching criteria of only 6 nonmatching atoms or bonds and a pIC50 change of at least 3 (3 orders of magnitude difference in hERG IC_{50}). The pairs are shown in Figure 11. The first and last rows indicate that Zwitterionic species (the piperazine ring nitrogen is protonatable in all four compounds) do not typically inhibit hERG. Conversion of the acid into an ester (first row) or fluorine atom (last row) results in high hERG activity. The pair shown in the second row illustrates the importance of a protonatable amine, since the compound without the tertiary nitrogen in basically inactive.

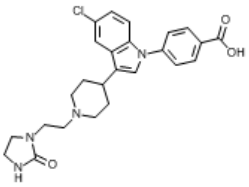
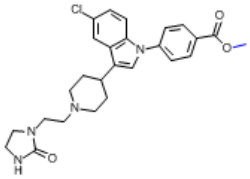
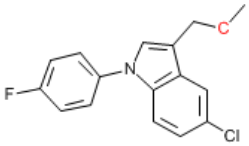
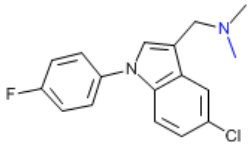
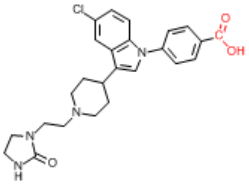
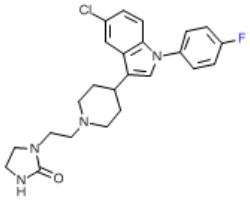
	Structure 1	Structure 2	pIC50 1	pIC50 2	Change(pIC50)
1			4.125	7.444	3.319
2			4.585	7.959	3.374
3			4.125	8.569	4.444

Figure 11 - MMPA analysis of hERG IC_{50} data set.

The hERG_pIC50 model was build using 582 compounds in the training set and 164 in the test set. A graph of the observed pIC50 values versus predicted values is shown below. The training and test set have similar statistics.

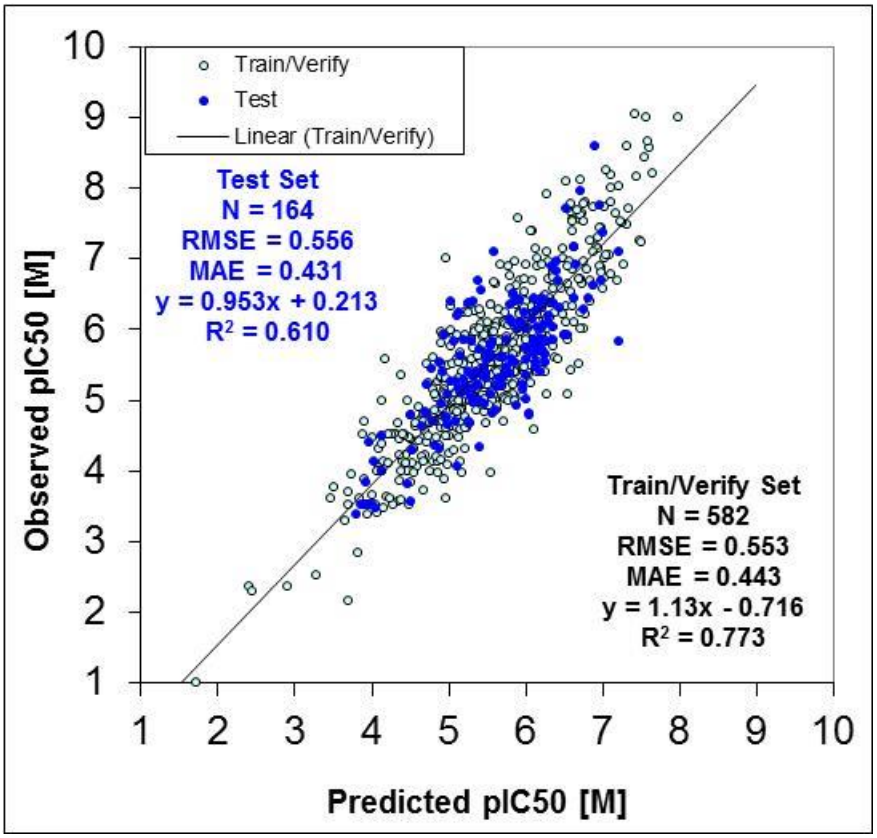


Figure 12 - Plot of experimental vs. predicted pIC50 (-logIC50) values for inhibition of hERG K+ channels. The blue dots denote predictions for the external test.

The TOX_Risk code for hERG liability is "hERG" and is set when hERG_FILTER = Yes and hERG_pIC50 is > 5.5. The full penalty is accessed for compounds with hERG_pIC50 > 6.

References

Cavalli A, Poluzzi E, De Ponti F, Recanatini M. "Towards a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of HERG K^+ Channel Blockers". *J. Med. Chem.* **2002**, *45*, 3844-3853.

Diaz GJ, Daniell K, Leitza ST, Martin RL, Su Z, McDermott JS, Cox BF, Gintant GA. "The [3H]dofetilide binding assay is a predictive screening tool for hERG blockade and proarrhythmia: Comparison of intact cell and membrane preparations and effects of altering [K+]." *J Pharm Tox Meth* **2004**; 50:187-199.