QSAR-based Prediction of Ames Mutagenicity for ICH M7 Submissions

Michael Lawless, Vijay K. Gombar, and Robert D. Clark

Simulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93534, USA (www.simulations-plus.com)

Abstract Number 1099

Introduction

The "Ames test", originally developed by Bruce Ames and his group, is a way to measure the mutagenic potential of chemicals.[1] It uses strains of *Salmonella typhimurium* and *Escherichia coli* as an alternative to chronic dosing in rodents. It is a short-term bacterial reverse mutation assay that detects the type of chemically-induced genetic damage that could be carcinogenic in humans. For impurities below the ICH[2] qualification thresholds, and lacking data for classification with respect to mutagenic and carcinogenic potential, the regulatory bodies of the European Union, Japan, and the USA accept genotoxicity evaluation by (Q)SAR approaches. Here we present a set of statistically-based QSAR models that predict outcomes of the bacterial reverse mutation assay.

Ten models were created from data on five individual strains of *S. typhimurium* or *E. Coli,* with and without rat liver S9 metabolic activation. Artificial neural network ensemble (ANNE) classification methodology was used to train the models.[3] Based on predictions from these models, rules were developed to assess mutagenicity risk (MUT_Risk) using a focused subset of 2,270 compounds from the World Drug Index[4] (WDI). For a subset of the Hansen's benchmark test set[5] that was not included in the training sets of our models, it was observed that 94.2% (114/121) of the compounds with a MUT_Risk of 4 were, in fact, mutagenic.

Methodology

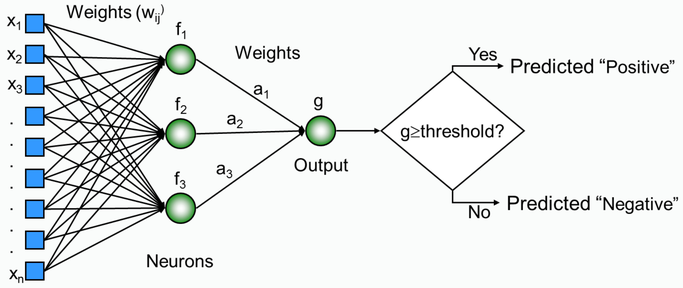

Figure 1 – ANNE classification methodology. Experimental negatives and positives are assigned values of 0 and 1, respectively. x_i are atomic and molecular descriptors. The descriptors are multiplied by the weights (w_{ij}) and summed in the nodes (f_i). The nodes are multiplied by the coefficients (a_i) and summed in the output node (g). The weights (w_{ij} and a_i) are optimized during model training to minimize the difference between experimental and predicted values. The threshold is selected as shown in Figure 2.

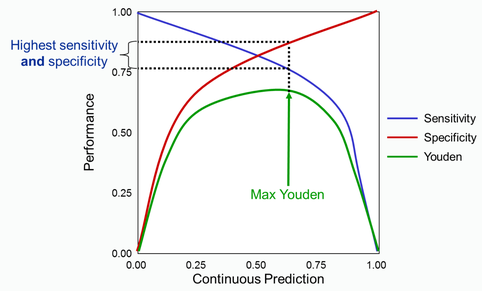

Figure 2 – Plot of performance versus "continuous prediction" (the output of the artificial neural network). The Youden index (*J*) is "1 – Specificity + Sensitivity". The maximum Youden value in the above plot represents the point with the best trade-off between Specificity and Sensitivity.

Results

Table 1 – Model performance on *E. Coli* and *S. typhimurium* strains. Models starting with the letter "M" (e.g., M98) reflect metabolic activation by a rat liver S-9 fraction. For the external test sets, the specificities range between 82.5 and 93.0%, the sensitivities between 77.3 and 100%, and the concordances between 80.8 and 93.7%.

Model		Negatives	Positives	Total	Correct	Concordance	Sensitivity	Specificity
98	Training	2213	679	2892	2608	90.2%	85.0%	91.8%
	Test	561	162	723	640	88.5%	83.3%	90.0%
M98	Training	1969	671	2640	2337	88.5%	85.2%	89.6%
	Test	394	313	707	612	86.6%	80.2%	91.6%
100	Training	2565	868	3433	2992	87.2%	84.2%	88.1%
	Test	253	129	382	332	86.9%	89.1%	85.8%
M100	Training	1889	867	2756	2303	83.6%	82.0%	84.3%
	Test	473	216	689	557	80.8%	77.3%	82.5%
97+1537	Training	1829	299	2128	1901	89.3%	80.6%	90.8%
	Test	202	34	236	215	91.1%	82.4%	92.6%
M97+1537	Training	1728	261	1989	1795	90.2%	79.7%	91.8%
	Test	174	47	221	198	89.6%	78.7%	92.5%
1535	Training	1741	230	1971	1807	91.7%	85.2%	92.5%
	Test	185	34	219	202	92.2%	88.2%	93.0%
M1535	Training	1411	231	1642	1432	87.2%	83.1%	87.9%
	Test	352	59	411	367	89.3%	83.1%	90.3%
102+wp2	Training	669	181	850	794	93.4%	90.6%	94.2%
	Test	78	17	95	89	93.7%	100.0%	92.3%
M102+wp2	Training	608	115	723	645	89.2%	83.5%	90.3%
	Test	71	9	80	71	88.8%	77.8%	90.1%

MUT_Risk is a qualitative estimate of the overall mutagenicity of a compound. It combines the individual "Positive" predictions from each model. TA98 and TA100 results overlap mechanistically[6] and so are merged into one rule. Each individual "Positive" prediction, for models built on data without metabolic activation, contributes one "vote" to the score. Note that rules involving rat liver S-9 activation models only contribute a "vote" if the corresponding model *without* activation is "Negative". The table below shows the code (MUT_Code) and criteria for each rule. About 16% of the commercial drugs in the focused subset of WDI get a MUT_Risk score of above 1 and 4% have a score above 2.

MUT_Code	Rule
S1	TOX_MUT_97+1537 = Positive
m1	TOX_MUT_m97+1537 = Positive AND NOT TOX_MUT_97+1537 = Positive
S2	TOX_MUT_98 = Positive AND TOX_MUT_100 = Positive
m2	(TOX_MUT_m98 = Positive AND TOX_MUT_m100 = Positive) AND NOT (TOX_MUT_98 = Positive AND TOX_MUT_100 = Positive)
S3	TOX_MUT_102+wp2 = Positive
m3	TOX_MUT_m102+wp2 = Positive AND NOT TOX_MUT_102+wp2 = Positive
S4	TOX_MUT_1535 = Positive
m4	TOX_MUT_m1535 = Positive AND NOT TOX_MUT_1535 = Positive

The 6,506 records in the supplemental data of Hansen's journal article[5] were curated by standardizing functional groups, neutralizing acids and bases, removing compounds containing atoms (e.g., Si) that aren't parameterized in ADMET Predictor and standardizing tautomers. 6,490 compounds were retained.

The records from our 10 individual models were combined. This resulted in 5,072 unique records. The two sets were combined and duplicate structures (ignoring chirality) were identified. This resulted in 1,907 compounds that are present in the Hansen data set but absent in our training sets.

Results

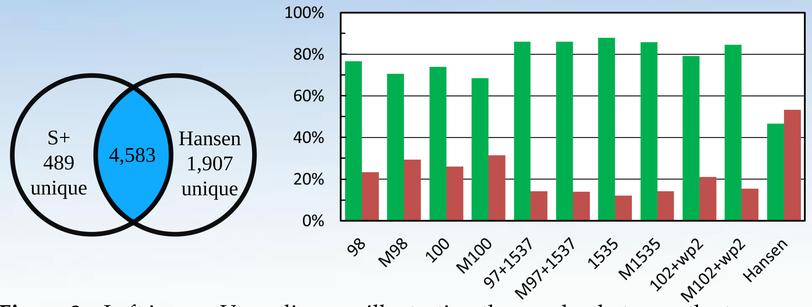

Figure 3 – Left image, Venn diagram illustrating the overlap between the two data sets. The image on the right is the percent of "Negative" (green) and "Positive" (red) in each data set. Note that the percent of positives and negatives is closer for the Hansen set. This is because a compound is classified as a positive if it is a positive in at least one out of the five strains in the presence or absence of rat S9 liver fraction.

MUT_Risk values were computed for the 1,907 unique compounds. Performance statistics were generated based on four thresholds of MUT_Risk, i.e., MUT_Risk 1 or higher considered positive, MUT_Risk 2 or higher considered positive, etc. (Figure 4). A MUT_Risk of 2 or higher results in the best balance between concordance, sensitivity and specificity. The specificity increases with increasing MUT_Risk threshold. This results in a lower percent of false positives and when a MUT_Risk of 4 is considered positive, then 114 of the 121 compounds (94.2%) predicted to be positive were in fact mutagenic in the laboratory.

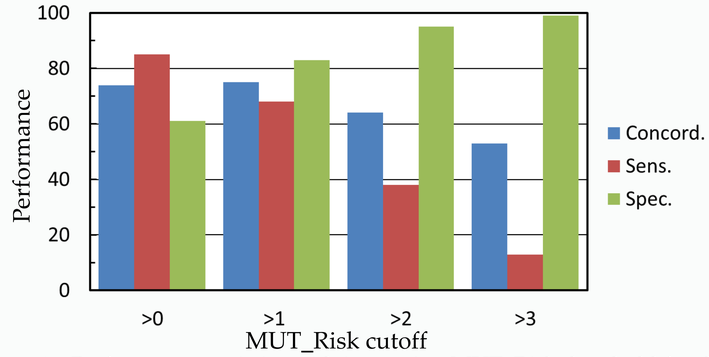

Figure 4 – Performance statistics with increasing MUT_Risk threshold for the 1,907 set of compounds in Hansen's data set.

Conclusions

ANNE classification models were created to predict Ames mutagenicity for individual strains of *S. typhimurium* and *E. Coli.* A MUT_Risk score was defined based on model predictions for a subset of the WDI. MUT_Risk was then applied to a set of compounds from Hansen that were not part of the our data set. The highest specificity and sensitivity corresponded to a MUT_Risk threshold of 2. The higher the MUT_Risk threshold, the higher the specificity. Of the 121 compounds with MUT_Risk = 4, 114 (94.2%) were experimental positives.

References

[1]Ames BN et al. *PNAS*, **70**, 8, 2281-2285.
[2]International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use.
[3]ADMET Modeler™ was used to build the models (www.simulations-plus.com).
[4]World Drug Index (http://thomsonreuters.com/en/products-services/pharma-life-sciences/life-science-research/world-drug-index.html).
[5]Hansen K et al. *J. Chem. Inf. Model*. **2009**, 49, 2077-2081.
[6]Mortelmans K and Ziegler E. *Mutation Research* **2000**, 455, 29-60.

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS