Characterizing the Domain of QSAR Models Predicting Genotoxicity

Kevin P. Cross1, Lidiya Stavitskaya2, R. Daniel Benz2*, Naomi L. Kruhlak2

1 Leadscope, Inc., Columbus, OH; 2 FDA Center for Drug Evaluation and Research, Silver Spring, MD; * Retired

[The findings and conclusions in this presentation have not been formally disseminated by the FDA and should not be construed to represent any agency determination or policy.]

Abstract:

- The current draft of the International Conference on Harmonisation (ICH) M7 guidance describes the use of in silico models to qualify genotoxic impurities during the drug safety evaluation and approval process.
- When performing QSAR predictions for regulatory purposes, questions have arisen regarding the out-of-domain predictions. Specifically, why are some test compounds predicted as out-of-domain? What classes of drugs are affected by this? How many out-of-domain predictions can be expected? How does the usage of a domain of applicability quantitatively improve the accuracy of QSAR predictions?
- Domain of applicability is defined differently for different QSAR modeling applications. Leadscope restricts the domain of applicability of its QSAR models to cover test compounds with a minimum distance to the training set of 0.7 and with at least one significant structural feature in the model (in addition to having all proper descriptors). To further ensure optimal performance, predictions with positive predictive probabilities of 0.4 - 0.6 are often considered equivocal.
- The domain of the Leadscope Salmonella mutagenicity QSAR model has improved significantly with the latest model version (v3.3) due to an increase in the breadth of the training set. Coverage measurements using three different, large collections of compounds ranged from 87% for the Hansen benchmark data set (3700 compounds), to 89% for a collection of drug products (1600 compounds), to 88% for a collection of biologically active compounds (5081 compounds).
- When using the Hansen set for external validation, the relationship between accuracy and predicted probability values was much stronger than for accuracy and distances of test compounds from the training set, one of the parameters currently defining domain of applicability. Accuracies of positive and negative predicted values were 85% - 90% at positive predictive probabilities 20.9 and 90.1, respectively, yet only showed 50% - 61% accuracy at probabilities between 0.4 and 0.6. In contrast, accuracies of predictions for test compounds with a minimum distance of 0.1 or less from the training set were 83% while those with a minimum distance of 0.7 or greater were 70% accurate -- a much shallower drop-off in accuracy.
- Focusing specifically on out-of-domain predictions for the Hansen set, the 146 predictions were 74% accurate when the domain requirements were lifted, as compared to 76% accurate for the remainder of the set. When only the feature count domain restriction and probability restriction (i.e., excluding probabilities of 0.4 - 0.6) were applied, accuracy increased to 86% and almost half of those previously out-of-domain compounds (96/146) were predicted, even without consideration of the minimum distance of the test set compounds to the training set.

Results and Discussion

- Leadscope Domain Analysis and Mutagenicity Prediction Qualifications

Leadscope restricts the domain of applicability of its QSAR models to predict test compounds with:
1. A maximum Tanimoto distance to the training set of 0.7 and
2. At least one significant structural feature in the model (in addition to having all proper descriptors).

To further ensure optimal performance, predictions with positive predictive probabilities in the range 0.4 - 0.6 are considered equivocal for regulatory purposes when considering QSAR Salmonella predictions under ICH M7 guidance.

Table 1. Domain Coverage using Several Validation Sets

<table>
<thead>
<tr>
<th>Model</th>
<th>Validation Set</th>
<th>Set Size</th>
<th>Out-of-Domain</th>
<th>Equivocal</th>
<th>Total</th>
<th>Unpredicted</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hansen Validation Set</td>
<td>3700</td>
<td>3.9%</td>
<td>8.6%</td>
<td>12.5%</td>
<td>87.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadscope Drug Products Set</td>
<td>1606</td>
<td>7.1%</td>
<td>3.5%</td>
<td>10.6%</td>
<td>89.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadscope Biologically Active Compounds</td>
<td>5081</td>
<td>8.3%</td>
<td>4.2%</td>
<td>12.5%</td>
<td>87.5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Model coverage was high for all external sets and highest against a set of drug only compounds.
- There were fewer equivocal predictions for more drug-like compounds than those in the Hansen set (which are less drug-like).

Figure 1. Toxicophore Characterization of the QSAR Model Training Set

- Many (56) different toxicophores are present with varying mean levels of activity.

Figure 2. Model Predictivity using the Hansen Validation Set

- Salmonella QSAR Predictivity Distribution

- Model performance is best where the predictive probabilities are high (>0.90 for positive predictions and <0.10 for negative predictions).
- Reported probabilities correlate with reality, i.e., individual prediction probabilities accurately reflect the probability of a correct prediction.
- The frequency of poor predictions (i.e., 0.40 - 0.60 probabilities) is relatively low (9% overall).

Figure 3. Model Predictivity using the Hansen Validation Set

- Incorrect Predictions by Minimum Training Set Distance

The percentage of incorrect predictions improves modestly from 32% to 17% as the test-to-training set minimum global distance decreases from 0.7 to 0.1. However, it drops off sharply from 0.7 to 1.0, supporting the use of a 0.7 cutoff as one of the in-domain criteria.

Figure 4. Accuracy of Unpredicted Hansen Set Compounds

- All Unpredicted Compounds 63% (474)
- Enforcing Model Features 70% (422)
- Enforcing Testing Set 60% (373)
- 59% (326)
- 73% (48)
- Excluding Equivocal Predictions 82% (124)

- All unpredicted compounds partitioned by rule with accuracy and additional predicted population using only one rule or combinations
- If domain restrictions were relaxed, the most accurate predictions (86%) result with 89.1% coverage when keeping structures having at least one model feature and a predictive probability that is not equivocal.

Conclusions

1. Robust QSAR models define a domain of applicability which limits their ability to predict properties for all test compounds.
2. Model domains may be characterized by assessing the toxicophore space covered by the training set using public toxicophores.
3. Regarding domain analysis, the accuracy of Leadscope Salmonella QSAR predictions is more strongly dependent on the presence of model features in a test compound than its global distance from the training set.
4. When considering overall accuracy of Leadscope Salmonella QSAR predictions, the probability of the predictions provides a better measure of accuracy than using domain restrictions; particularly the distance of the test compound to the training set.

Acknowledgements

This project was supported in part by an appointment to the Oak Ridge Institute for Science and Education (ORISE) Research Participation Program at U.S. FDA CDER administered through an agreement between the U.S. Department of Energy and FDA CDER.

References