In silico mutagenicity assessment methodologies relating to E&L strategies

Glenn J. Myatt (gmyatt@leadscope.com)
Leadscope, Inc.

May 12-14 2015
Extractables & Leachables USA 2015
Overview

• Describe an *in silico* mutagenicity assessment protocol for extractables and leachables based on the recently issued ICH M7 guidance

• Outline the components of the two recommended *in silico* methodologies and how the results are combined as part of a consensus call

• Review the components of an expert review to ensure the results are well documented, consistently generated and traceable
ICH M7 Guidance

• In the ICH M7 guidance “Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk” the purpose is stated as:
  – “… to provide a practical framework that is applicable to the identification, categorization, qualification, and control of these mutagenic impurities to limit potential carcinogenic risk.” [1]

• The hazard assessment assigns the impurity to one of 5 classes:
  – **Class 1**: Known mutagenic carcinogens
  – **Class 2**: Known mutagens with unknown carcinogenic potential
  – **Class 3**: Alerting structure, unrelated to the structure of the drug substance; no mutagenicity data
  – **Class 4**: Alerting structure, same alert in drug substance or compounds related to the drug substance (e.g., process intermediates) which have been tested and are non-mutagenic
  – **Class 5**: No structural alerts, or alerting structure with sufficient data to demonstrate lack of mutagenicity or carcinogenicity

ICH M7 and *in silico* analysis

- In the absence of available laboratory data (e.g. rodent carcinogenicity study or a bacterial mutagenesis study) an *in silico analysis* is permitted to assign compounds to classes 3, 4, or 5.
- These results can be supplemented with an expert opinion.
- A risk assessment for the individual substance is then made, with:
  - Class 1 to be controlled “*at or below compound-specific acceptable limit*”
  - Class 2 or 3 impurities should be controlled “… *at or below acceptable limits (appropriate TTC)*”
  - Classes 4 and 5 can be treated as non-mutagenic.
- This *in silico* approach to predicting mutagenicity is equally applicable to extractables and leachables.
In silico analysis methodologies

• The ICH M7 guidance permits the use of (Q)SAR approaches
  – “…bacterial mutagenicity predictions…” which covers both A:T and G:C mutations

• Two methodologies that should be utilized:
  – “One methodology should be expert rule-based and the second methodology should be statistical-based.”
  – “…(Q)SAR models … should follow the general validation principles set forth by the Organisation for Economic Co-operation and Development (OECD).”

• The guidance also identifies how to use the results from these in silico systems to assign the impurity to class 5:
  – “The absence of structural alerts from two complementary (Q)SAR methodologies (expert rule-based and statistical) is sufficient to conclude that the impurity is of no mutagenic concern…”
Expert rule-based methodology

• **Building**
  – Extract alerts from literature and databases
  – Capture reasons for activation and/or deactivation
  – Verify against large databases (reference set)
  – Document mechanistic rationale

• **Applying**
  – Unambiguously determines the compound’s applicability domain
  – Identifies alerting structure (considering activation and deactivation)
  – Generate a confidence based on historical data

• **Explaining**
  – Identifies structural basis for the prediction
  – Identifies a mechanistic rationale
  – Identify examples with test results
Expert rule-based example

**Expert Alerts: Bacterial Mutation**
**Prediction methodology:** Genetic Toxicity Bacterial Mutation Alerts v1 (System: Leadscope Model Applier v1.9.1.14)
**Prediction results:** Positive
**Laboratory data:** Positive
**Alerts fired:** 24: aziridine (0.97)
**Alerts precision:** 0.9706

**24: aziridine (Bacterial Mutation)**

**Alert Definition**
*Description*
The alert was cited in the following: "aromatic and aliphatic aziridines" [Ashby, 1998]; "aziridine" [Kazus, 2005]; "isocyanates and aziridines" [Bailey, 2005]; "N-7 epoxides and aziridines" [Bailey, 2005].

**Substructure definition**
azonine

**Mechanisms**
**Endpoint:** Bacterial Mutation  
**Category:** Alkylating, Direct Acting Agents

**Description**
... alert to DNA-reactivity, [Ashby, 1998]
... electrophilic, alkylation substructures that possess significant intrinsic reactivity... [Kazus, 2005]
"Natural electrophiles" [Bailey, 2005].

Discussed in Section 4.2.5 of [Benigni, 11]: "Aziridines are extremely reactive alkylating agents..." "...activity of these compounds depends on their ability to act as DNA crosslinkers/intermolecular opening of the aziridine moiety by N7 positions of purines."

**References**
Statistical-based methodology

• Building
  – QSAR models are automatically built from large databases of historical bacterial mutagenicity data (training set)
  – The chemicals are described using descriptors (substructure fragments, molecular properties (e.g. logP))
  – Mathematical models are built to encode the relationships between the presence/absence or values of the descriptors and bacterial mutagenicity endpoints

• Applying
  – Unambiguously determines the compound’s applicability domain
  – Calculates a prediction based on the descriptors that are common with the model

• Explaining
  – Identifies structural basis for the prediction
  – Identify examples with test results
Statistical-based example

<table>
<thead>
<tr>
<th>Feature</th>
<th>Chemical Structure</th>
<th>Predicted Value</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrile</td>
<td><img src="image1" alt="Nitrile Structure" /></td>
<td>0.79</td>
<td>0.8993</td>
</tr>
<tr>
<td>KeyP</td>
<td><img src="image2" alt="KeyP Structure" /></td>
<td>1.457</td>
<td>0.5322</td>
</tr>
<tr>
<td>Paired Molecular</td>
<td><img src="image3" alt="Paired Molecular Structure" /></td>
<td>-0.0751</td>
<td>-0.0768</td>
</tr>
<tr>
<td>Weight</td>
<td><img src="image4" alt="Weight Structure" /></td>
<td>2262</td>
<td>2712</td>
</tr>
<tr>
<td>Training Negative</td>
<td><img src="image5" alt="Training Negative Structure" /></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Training Positive</td>
<td><img src="image6" alt="Training Positive Structure" /></td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Prediction methodology: Genetic Toxicity Salmonella Model v3 (System: Leadscope Model Applier v1.9.1.14)

Prediction results: Positive

Prediction probability: 0.933

Laboratory data: Positive
## Why two methodologies?

<table>
<thead>
<tr>
<th></th>
<th>Expert Alert Prediction</th>
<th>QSAR Statistical Prediction</th>
<th>M7 Consensus</th>
<th>Difference from QSAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td>82%</td>
<td>76%</td>
<td>81%</td>
<td>+5%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>90%</td>
<td>82%</td>
<td>98%</td>
<td>+16%</td>
</tr>
<tr>
<td>Specificity</td>
<td>71%</td>
<td>68%</td>
<td>59%</td>
<td>-9%</td>
</tr>
<tr>
<td>Positive Predictivity</td>
<td>81%</td>
<td>78%</td>
<td>76%</td>
<td>-2%</td>
</tr>
<tr>
<td>Negative Predictivity</td>
<td>83%</td>
<td>73%</td>
<td>95%</td>
<td>+22%</td>
</tr>
<tr>
<td>Coverage</td>
<td>97%</td>
<td>87%</td>
<td>100%</td>
<td>+13%</td>
</tr>
</tbody>
</table>

Results generated using the Leadscope Model Applier V2.0 using the Hansen dataset
**In silico analysis overview**

**Collect available laboratory data**
- Endpoints: Rodent carc, Bact. Mut, Other in vivo
- Outcomes: Positive, Negative, Indeterminate, Irrelevant, None

**Assess available data**
- Positive
- Negative
- None, indeterminate, irrelevant

**Perform in silico**
- Endpoints: GC QSAR, AT QSAR, Expert alerts
- Outcomes: Positive, Negative, Indeterminate, Out-of-domain, Positive lab data, Negative lab data

**Assess in silico**
- Clear positive
- Consensus
- Clear negative
- All outcomes

**Generate opinion?**
- Accepted positive (including indeterminates and out-of-domains)
- Refuted negative
- Accepted negative
- Refuted positive, indeterminates, and out-of-domain

**Class 1,2,3**
- Control/establish limit

**Class 4,5**

*Leadscope®*
**In silico** analysis overview

- **Collect available laboratory data**
  - **Endpoints:** Rodent carc. Bact. Mut. Other in vivo
  - **Outcomes:** Positive, Negative, Indeterminate, Irrelevant, None

- **Assess available data**
  - Positive
  - None, indeterminate, irrelevant
  - Negative

- **Perform in silico**
  - **Endpoints:** GC QSAR, AT QSAR, Expert alerts
  - **Outcomes:** Positive, Negative, Indeterminate, Out-of-domain, Positive lab data, Negative lab data

- **Assess in silico**
  - Clear positive
  - Consensus
  - All outcomes
  - Clear negative

- **Generate opinion?**
  - Accepted positive (including indeterminates and out-of-domains)
  - Refuted negative
  - Accepted negative
  - Refuted positive, indeterminates, and out-of-domain

- **Class 1,2,3**
  - Control/establish limit

- **Class 4,5**
**In silico consensus**

**Perform in silico**

**Endpoints:**
- GC QSAR
- AT QSAR
- Expert alerts

**Outcomes:**
- Positive
- Negative
- Indeterminate
- Out-of-domain
- Positive lab data
- Negative lab data

**Consensus**

**Assess in silico**
ICH M7 consensus rules

The rules are based on the general principle that any positive prediction (from any of the methodologies) is positive and laboratory training set data is used over predictions when available.
**In silico analysis overview**

**Collect available laboratory data**
- **Endpoints:** Rodent carc. Bact. Mut. Other *in vivo*
- **Outcomes:** Positive Negative Indeterminate Irrelevant None

**Assess available data**
- Positive
- Negative
- None, indeterminate, irrelevant

**Perform in silico**
- **Endpoints:** GC QSAR AT QSAR Expert alerts
- **Outcomes:** Positive Negative Indeterminate Out-of-domain Positive lab data Negative lab data

**Assess in silico**
- Clear positive
- Consensus
- Clear negative
- All outcomes

**Generate opinion?**
- Accepted positive (including indeterminates and out-of-domains) Refuted negative
- Accepted negative Refuted positive, indeterminates, and out-of-domain

**Class 1,2,3**
- Control/establish limit

**Class 4,5**
In silico analysis overview

Collect available laboratory data
- **Endpoints**: Rodent carc. Bact. Mut. Other in vivo
- **Outcomes**: Positive Negative Indeterminate Irrelevant None

Assess available data
- Positive
- None, indeterminate, irrelevant
- Negative

Perform in silico
- **Endpoints**: GC QSAR AT QSAR Expert alerts
- **Outcomes**: Positive Negative Indeterminate Out-of-domain Positive lab data Negative lab data

Assess in silico
- Clear positive
- Consensus
- Clear negative
- All outcomes

Generate opinion?
- Accepted positive (including indeterminates and out-of-domains)
- Refuted negative
- Accepted negative
- Refuted positive, indeterminates, and out-of-domain

Class 1,2,3
- Control/establish limit

Class 4,5

Controls/establish limit
When to generate an expert opinion?

Assess *in silico*

- All outcomes

Class 1, 2, 3

- Control/ establish limit

- Clear positive
- Accepted positive (including indeterminates and out-of-domains)
- Refuted negative

- Clear negative

Class 4, 5

- Accepted negative
- Refuted positive, indeterminates, and out-of-domain

Generate opinion?
Clear negative example

- Negative predictions with low positive probability/precision scores in all methodologies
- Good structural coverage of features by the QSAR models
- The lack of any alerting substructures

<table>
<thead>
<tr>
<th>#</th>
<th>Impurities</th>
<th>Laboratory Data</th>
<th>Expert rule-based</th>
<th>Matching Alerts</th>
<th>QSAR Salmonella</th>
<th>Salmonella Probability</th>
<th>QSAR E.coli/TA102</th>
<th>E.coli/TA102 Probability</th>
<th>In silico consensus assessment</th>
<th>N7 Class assignment</th>
<th>Additional supportive evidence and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>[Image of impurity]</td>
<td>None</td>
<td>Negative</td>
<td>No Alerts</td>
<td>Negative</td>
<td>0.0243</td>
<td>Negative</td>
<td>0.134</td>
<td>Negative</td>
<td>5</td>
<td>Accepted negative in silico result. The impurity lacks obvious reactive potential.</td>
</tr>
</tbody>
</table>
Clear positive example

- All three models used were positive, with high precision and positive probability scores.
- The structural basis or alert identified in the three systems is the primary aromatic amine; however, the presence of the phenyl substituent para to the amine is highlighted as having an activating effect, resulting in a particularly strong positive prediction.

<table>
<thead>
<tr>
<th>#</th>
<th>Impurities</th>
<th>Laboratory Data</th>
<th>Expert rule-based</th>
<th>Matching Alerts</th>
<th>QSAR Salmonella</th>
<th>Salmonella Probability</th>
<th>QSAR E.coli/TA102</th>
<th>E.coli/TA102 Probability</th>
<th>In silico consensus assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>\includegraphics[width=0.5cm]{impurity_d}</td>
<td>None</td>
<td>Positive</td>
<td>2657: aromatic amine [NH2] (strong activating amines) (0.89)</td>
<td>Positive</td>
<td>0.945</td>
<td>Positive</td>
<td>0.609</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Impurity D [453927]
Decision to generate an opinion for a negative

- The overall consensus was negative
  - no alerts identified
  - negative Salmonella QSAR result
  - indeterminate E.coli QSAR result
- The Salmonella QSAR probability is marginal*
- Highlighting of the naphthalene

*just under the negative cut-off value of 0.4 and the E. coli model is just over this cut-off, but below the 0.6 criteria for a positive assignment
Negative opinion – questions to consider

• Are there any potentially reactive groups present?
  – If the QSAR models highlight a feature, is it valid (e.g. known alert)?
  – For any mitigated alerts, is it supported by a mechanistic rationale or laboratory data?
  – Is it possible to identify visually any potentially reactive groups?
  – Does any portion of the impurity (not considered by the model) contain a structural feature that has any association with bacterial mutagenicity data?
  – Are there analogs to support the negative assessment?
  – Are the rules employed to arrive at the overall consensus appropriately applied?
Negative opinion example

• In evaluating “Scaffold 9”, all examples contained another common structural alert and hence it was opined that the training set compounds that are the basis for “Scaffold 9” all contain other structural alert (e.g. aromatic nitro, polycyclic aromatic hydrocarbons, primary alkyl halide, ...)

![Chemical Structure Diagram](image)
Decision to generate an opinion for a positive

- Are there potential reasons to refute a positive prediction, e.g. similarity to a related substance?

<table>
<thead>
<tr>
<th>Impurities</th>
<th>Laboratory Data</th>
<th>Expert rule-based</th>
<th>Matching Alerts</th>
<th>QSAR Salmonella</th>
<th>Salmonella Probability</th>
<th>QSAR E.coli/TA102</th>
<th>E.coli/TA102 Probability</th>
<th>In silico consensus assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>None</td>
<td>Positive</td>
<td>140: Dibenzofuran-and-related (0.9) Positive</td>
<td>0.699</td>
<td>Out of Domain</td>
<td>0.0823</td>
<td>Positive</td>
<td></td>
</tr>
</tbody>
</table>

Impurity C [X-195763]
Options for opinions to refute a positive

- A class 4 opinion when the compound is a close analog of a related negative substance
- Explanation of mechanism
- Irrelevant QSAR features
- Chemical analogs
- Verifying no additional reactive groups
Positive opinion example

• Class 4 - close to related compound

“An impurity with a structural alert that is shared (e.g., same structural alert in the same position and chemical environment) with the drug substance or related compounds can be considered as non-mutagenic ... if the testing of such material in the bacterial mutagenicity assay was negative.”

The data used here is hypothetical to illustrate the class 4 concept.
Positive opinion example

The following impurity was predicted as negative in the rule-based expert alerts and positive by the other QSAR statistical models. The overall call was therefore positive; however, an expert opinion to refute the positive QSAR result was made, based on an understanding of the mechanism and analogs.

As illustrated in these close analogs, only unhindered epoxides are mutagenic.
Positive opinion example

• The QSAR model features may be irrelevant because coincidently features, mitigating features, limited training set examples, no significant positive model feature, irrelevant training set examples, or other reasons.

• In the following example, the E.coli QSAR model predicted the impurity to be positive with a probability just over the 0.6 cut-off for a positive call.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Laboratory Data</th>
<th>Expert rule-based</th>
<th>Matching Alerts</th>
<th>QSAR Salmonella</th>
<th>Salmonella Probability</th>
<th>QSAR E.coli TA102</th>
<th>E.coli TA102 Probability</th>
<th>In silico Consensus Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS-158915-1</td>
<td>Negative</td>
<td>Negative</td>
<td>No Alerts</td>
<td>Negative</td>
<td>0.183</td>
<td>Positive</td>
<td>0.642</td>
<td>Positive</td>
</tr>
<tr>
<td>LS-158915-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The positive training set compounds containing the carbamate or related compounds also contained other structural alerts – aziridine or acetahydroxamic acids (or related).
Reporting

Materials and methods
- Software, models and databases used, along with version numbers

Summary of the results and conclusions
- Laboratory data and/or *in silico* results (Statistical QSAR models for A:T and G:C mutations and expert rule-based alerts results for bacterial mutagenicity)
- Class 1-5 assignment
- Summary of any supporting opinions or remarks
- Summary of risk assessment/control
Reporting

**Supporting information**

- Opinion(s) to refute a positive *in silico* result, along with examples and references to illustrate

**Appendices**

- Complete *in silico* reports
- Complete study reports
Ensuring regulatory models reflect proprietary chemical space

- Improves performance and applicability domain analysis
  - Sharing data
  - Sharing knowledge

- Primary aromatic amine (PAA) fingerprinting project

<table>
<thead>
<tr>
<th>Name of Region</th>
<th>Expert Alerts (original)</th>
<th>Expert Alerts (modified)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concordance</td>
<td>63.6%</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>92.8%</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>40.1%</td>
</tr>
<tr>
<td></td>
<td>Positive predictivity</td>
<td>55.4%</td>
</tr>
<tr>
<td></td>
<td>Negative predictivity</td>
<td>87.5%</td>
</tr>
</tbody>
</table>
Conclusions

• Outlined process for *in silico* bacterial mutagenicity analysis
• Reviewed two methodologies
• Described rules to generate a consensus call
• Discussed accompanying expert opinions, including possibilities for refuting a positive prediction
• Detailed the generation of a complete report
• Summarized methodologies to ensure regulatory acceptable models reflect proprietary chemicals space
Acknowledgements

• **Leadscope**
  – Kevin Cross
  – Dave Bower
  – Scott Miller
  – Pat Quigley

• **US Food and Drug Administration**
  – Naomi L. Kruhlak
  – Lidiya Stavitskaya
  – Mark Powley