Integrative Prediction Approaches in Genotoxicity / Carcinogenicity: Structure-Activity Relationships. Part II

Romualdo Benigni

Istituto Superiore di Sanita’
Rome Italy

rbenigni@iss.it
Structure-activity relationship concepts:

application to different issues, through different approaches

Coarse-grain    Structure Alerts

Fine-tuned    Quantitative Structure-Activity Relationships (QSAR)
SAs have limitations

- Only flag chemicals with potentially reactive substructures
- Poor discrimination within a chemical class
- No negative prediction (except by exclusion)

A fine-tuned generalization:

**Quantitative Structure-Activity Relationships** (QSAR)

- Use a few physical chemical properties
  - hydrophobic, electronic (reactivity) and steric effects
- Finely scaled parameters to distinguish subtle differences
- Predictions for both positives and negatives
Corwin Hansch brought together physical organic chemistry and the study of chemical-biological interactions.

Hansch approach

\[ \log k = f (\text{steric, electronic, hydrophobic}) \]

Out of qualitative Structure-Activity Relationships (SAR):

- Hammett equation (electronic factors)
  \[ \log k = a \sigma + \text{constant} \]
- Taft equation (steric factors)
  \[ \Delta \log k = a E_s + \text{constant} \]

Quantitative Structure-Activity Relationships (QSAR)
QSAR approach: widest application reach, e.g.:

- **Ionization of phenols in aqueous solution**
  \[
  \log K = 2.01 \sigma_- + 1.94 F_2 - 9.86
  \]
  
  \( N = 23, r^2 = 0.979, s = 0.146, q^2 = 0.966 \)

- **Elevation of serum alanine transaminase in mice due to hepatic toxicity by X-C6H4CHdCH2**
  \[
  \log 1/C = -0.46 \sigma+ + 3.22
  \]
  
  \( n = 6, r^2 = 0.862, s = 0.118, q^2 = 0.738 \)

- **LD100 for humans** (human kill by miscellaneous poisons)
  \[
  \log 1/C = 1.17 \log P + 1.70
  \]
  
  \( n = 12, r^2 = 0.869, s = 0.498, q^2 = 0.825 \)

*C. Hansch, Chem. Rev. 2002, 102, 783-812*
Contribution of Corwin Hansch to the foundation of the QSAR:

• recognition of the role of the hydrophobic interaction, and its parameterization

• adoption of quantitative models and statistics

• multiparametric models and multivariate statistics
Hansch approach to QSAR

- **Congeneric** sets of chemicals:
  - Same structure
  - Different substituents
  - Same mechanisms of action
  - Same rate limiting step

- Modeling the *potency of the active* compounds (however, it can be extended to yes/no activity)

- **Statistics** (training set) for the construction of the QSAR model
The **Hansch approach** widely modified and extended:

- new molecular parameters
- chemical substructures and indicator variables
- quantum mechanical calculations
- molecular mechanics
- computer graphics
- a wider array of statistical tools

However, the QSAR area still maintains a fundamental unity founded on:

- systematic use of mathematical models
- multivariate point of view
Building a QSAR: mutagenicity of αβ-unsaturated aldehydes

Training set

Biological data (measured):
• logTA100: log (revertants/μmol)
• Activity: 1 = negative; 3 = positive

Chemical descriptors (calculated):
• logP: hydrophobicity
• MR: Molar Refractivity (bulkiness)
• C_β: charge on β-carbon
• C_carb: charge on carbonilic carbon
• HOMO: energy of the Highest Occupied Molecular Orbital
• LUMO: energy of the Lowest Unoccupied Molecular Orbital
## Mutagenicity and parameters: αβ–unsaturated aldehydes

<table>
<thead>
<tr>
<th>Chemical</th>
<th>CAS</th>
<th>logTA100</th>
<th>logP</th>
<th>MR</th>
<th>C_b</th>
<th>C_carb</th>
<th>HOMO</th>
<th>LUMO</th>
<th>Act.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrolein</td>
<td>107–02–8</td>
<td>3.38</td>
<td>-0.01</td>
<td>1.655</td>
<td>-0.0645</td>
<td>0.307</td>
<td>-10.684</td>
<td>-0.15105</td>
<td>3</td>
</tr>
<tr>
<td>2-Chloroacrolein</td>
<td>683–51–2</td>
<td>6.12</td>
<td>0.644</td>
<td>2.147</td>
<td>-0.0755</td>
<td>0.3165</td>
<td>-10.039</td>
<td>-0.4395</td>
<td>3</td>
</tr>
<tr>
<td>2-Bromoacrolein</td>
<td>14925–39–4</td>
<td>5.86</td>
<td>0.734</td>
<td>2.432</td>
<td>-0.044</td>
<td>0.323</td>
<td>-10.6451</td>
<td>-0.5065</td>
<td>3</td>
</tr>
<tr>
<td>Crotonaldehyde</td>
<td>4170–30–3</td>
<td>3.29</td>
<td>0.52</td>
<td>2.119</td>
<td>-0.047</td>
<td>0.312</td>
<td>-10.4528</td>
<td>-0.13665</td>
<td>3</td>
</tr>
<tr>
<td>2,4-Hexadienal</td>
<td>142–83–6</td>
<td>3.27</td>
<td>1.154</td>
<td>3.325</td>
<td>-0.0105</td>
<td>0.313</td>
<td>-9.47035</td>
<td>-0.60528</td>
<td>3</td>
</tr>
<tr>
<td>Cinnamaldehyde</td>
<td>104–55–2</td>
<td>1.9</td>
<td>4.394</td>
<td>0.005</td>
<td>0.314</td>
<td>-9.44035</td>
<td>-0.80765</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Citral</td>
<td>5392–40–5</td>
<td>2.95</td>
<td>4.876</td>
<td>-0.01175</td>
<td>0.31625</td>
<td>-9.53225</td>
<td>-0.16113</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>α-Methoxycinnamaldehyde</td>
<td>60125–24–8</td>
<td>1.818</td>
<td>5.011</td>
<td>0.018</td>
<td>0.316</td>
<td>-9.0849</td>
<td>-0.72165</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2-Dichloromethyl-3,3-dichloroacrolein</td>
<td>14109–84–1</td>
<td>3.77</td>
<td>2.132</td>
<td>4.085</td>
<td>-0.149</td>
<td>0.3155</td>
<td>-10.2067</td>
<td>-1.15895</td>
<td>3</td>
</tr>
<tr>
<td>2-Methylacrolein</td>
<td>78–85–3</td>
<td>2.66</td>
<td>0.3</td>
<td>2.119</td>
<td>-0.086</td>
<td>0.319</td>
<td>-10.458</td>
<td>-0.11045</td>
<td>3</td>
</tr>
<tr>
<td>2-Chloro-3-methylacrolein</td>
<td>53175–28–3</td>
<td>5.31</td>
<td>1.173</td>
<td>2.61</td>
<td>-0.0525</td>
<td>0.32</td>
<td>-9.73915</td>
<td>-0.39375</td>
<td>3</td>
</tr>
<tr>
<td>2-Chloro-3,3-dimethylacrolein</td>
<td>7364–09–2</td>
<td>4.63</td>
<td>1.572</td>
<td>3.074</td>
<td>-0.0115</td>
<td>0.321</td>
<td>-9.5561</td>
<td>-0.41</td>
<td>3</td>
</tr>
<tr>
<td>2-Ethylacrolein</td>
<td>922–63–4</td>
<td>2.38</td>
<td>1.24</td>
<td>2.583</td>
<td>-0.075</td>
<td>0.318</td>
<td>-10.7713</td>
<td>-0.10135</td>
<td>3</td>
</tr>
<tr>
<td>2-Propylacrolein</td>
<td>1070–13–9</td>
<td>2.25</td>
<td>1.358</td>
<td>3.047</td>
<td>-0.075</td>
<td>0.3175</td>
<td>-10.5589</td>
<td>-0.1028</td>
<td>3</td>
</tr>
<tr>
<td>2-Butylacrolein</td>
<td>1070–66–2</td>
<td>2.62</td>
<td>1.887</td>
<td>3.51</td>
<td>-0.0745</td>
<td>0.317</td>
<td>-10.5622</td>
<td>-0.1101</td>
<td>3</td>
</tr>
<tr>
<td>2-Pentenal</td>
<td>1576–87–0</td>
<td>3.14</td>
<td>1.049</td>
<td>2.583</td>
<td>-0.0495</td>
<td>0.312</td>
<td>-10.4708</td>
<td>-0.12915</td>
<td>3</td>
</tr>
<tr>
<td>2-Hexenal</td>
<td>6728–26–3</td>
<td>2.99</td>
<td>1.578</td>
<td>3.047</td>
<td>-0.0485</td>
<td>0.312</td>
<td>-10.4601</td>
<td>-0.1335</td>
<td>3</td>
</tr>
<tr>
<td>2-Heptenal</td>
<td>18829–55–5</td>
<td>2.67</td>
<td>2.107</td>
<td>3.51</td>
<td>-0.0485</td>
<td>0.312</td>
<td>-10.4632</td>
<td>-0.13485</td>
<td>3</td>
</tr>
<tr>
<td>2-Chlorocinnamaldehyde</td>
<td>18365–42–9</td>
<td>3.78</td>
<td>2.602</td>
<td>4.886</td>
<td>-0.018</td>
<td>0.321667</td>
<td>-9.31563</td>
<td>-0.99007</td>
<td>3</td>
</tr>
<tr>
<td>2-Bromocinnamaldehyde</td>
<td>5443–49–2</td>
<td>5.02</td>
<td>2.742</td>
<td>5.171</td>
<td>0.036</td>
<td>0.3325</td>
<td>-9.5057</td>
<td>-1.0918</td>
<td>3</td>
</tr>
</tbody>
</table>
Building a QSAR: Application of multivariate statistics

- **Mutagenic potency** of actives  ➤➤  Multivariate regression

- **Activity** (yes / no)  ➤➤  Canonical discriminant analysis
**QSAR of αβ–unsaturated aldehydes**

Mutagenic potency in *Salmonella typhimurium* TA100 (no S9)

<table>
<thead>
<tr>
<th>Electronic</th>
<th>Steric</th>
<th>Hydrophobic</th>
</tr>
</thead>
<tbody>
<tr>
<td>logTA100</td>
<td>-4.58</td>
<td>-3.66</td>
</tr>
<tr>
<td>LUMO</td>
<td>-</td>
<td>MR + 72.46</td>
</tr>
<tr>
<td>Ccarb</td>
<td>2.55</td>
<td>logP + 13.09442</td>
</tr>
<tr>
<td>CBβ</td>
<td>-12.61</td>
<td></td>
</tr>
</tbody>
</table>

\[logTA100 = -4.58 \text{LUMO} - 3.66 \text{MR} + 72.46 \text{Ccarb} + 2.55 \text{logP} + 13.09442 \text{CBβ} - 12.61\]

\(n = 17; \quad r^2 = 0.84; \quad \text{cross-validated } r^2 (q^2) = 0.40;\)

Benigni et al., 2003
QSAR of αβ–unsaturated aldehydes

Mutagenic activity in *Salmonella typhimurium* TA100 (no S9)

(no)activityTA100 = 3.87 MR - 3.12 logP + 3.23 LUMO

n = 20;  
w_neg = 9.69 (n=3)  
w_pos = 6.37 (n=17)  
Threshold = 8.03

Squared Canonical Correlation = 0.61  
100% correct re-classification;  
85% correct cross-validation

external prediction: 5 / 5 correct

Benigni et al., 2003  
(Revised eq.; raw.coeffs.)
QSAR of αβ–unsaturated aldehydes: activity prediction

Mutagenic activity in Salmonella typhimurium TA100 (no S9)

3,3-dimethylacrolein

\[
\text{(no)activity}_{TA100} = 3.87 \times 2.583 - 3.12 \times 0.919 + 3.23 \times (-0.1445)
\]

\[
= 6.66
\]

\[
w_{\text{neg}} = 9.69 \ (n=3)
\]

\[
w_{\text{pos}} = 6.37 \ (n=17)
\]

6.66 < threshold

Threshold = 8.03

the substance is (correctly) predicted as positive
VALIDATION OF QSAR MODELS

• Cross-validation:

  Estimate the predictive power within the compounds set
  Leave-one-out
  Leave-many-out
  Artificial split into Training set / Test set
  Necessary but not sufficient condition for high predictive power

• External validation:

  Estimate the predictive power on compounds from an external test set
  Unbiased
  Into account the changes in the chemicals in use at different times
Collection and Evaluation of (Q)SAR Models for Mutagenicity and Carcinogenicity

Romualdo Benigni, Cecilia Bossa, Tatiana Netzeva, Andrew Worth

PUBSY ID - EUR 22772 EN

ECB-ISS Project

Short listed local QSARs for congeneric classes:

- scientifically (mechanistically) interpretable
- good internal statistics
- **applicability domain** checked: functional group parameters range, chemical similarity
- **real external predictivity** tested
Regression-based QSARs for Potency (positives): fit and predictivity

<table>
<thead>
<tr>
<th>QSAR</th>
<th>rtra</th>
<th>q^2</th>
<th>q^2_10</th>
<th>lever</th>
<th>rte</th>
<th>accte</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amm TA98</td>
<td>.90</td>
<td>.78</td>
<td>.71</td>
<td>.06</td>
<td>.41</td>
<td>.36</td>
</tr>
<tr>
<td>Amm TA100</td>
<td>.88</td>
<td>.74</td>
<td>.66</td>
<td>.06</td>
<td>.68</td>
<td>.57</td>
</tr>
<tr>
<td>Amm mouse</td>
<td>.91</td>
<td>.58</td>
<td>0</td>
<td>.25</td>
<td>.56</td>
<td>.58</td>
</tr>
<tr>
<td>Amm rat</td>
<td>.93</td>
<td>.81</td>
<td>.79</td>
<td>.15</td>
<td>.48</td>
<td>.71</td>
</tr>
<tr>
<td>Nitro TA98</td>
<td>.90</td>
<td>.89</td>
<td>.80</td>
<td>.04</td>
<td>-.23</td>
<td>.43</td>
</tr>
<tr>
<td>Nitro TA100</td>
<td>.88</td>
<td>.77</td>
<td>.73</td>
<td>.05</td>
<td>.36</td>
<td>.32</td>
</tr>
</tbody>
</table>

Amm: aromatic ammines; Nitro: nitroarenes

**Training set:**
rtra: corr.coeff.; q^2: r^2 cross-val (LOO); q^2_10: q^2 L-10-O; lever: mean leverage

**Test set:**
rte: corr.coeff.; accte: accuracy (within 1 log activity unit)
## Discriminant QSARs for Activity (+/-): fit and predictivity

<table>
<thead>
<tr>
<th>QSAR</th>
<th>sqcc</th>
<th>acctra</th>
<th>acc10</th>
<th>accte</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rodent</td>
<td>0.38</td>
<td>0.88</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.94</td>
<td>0.78</td>
<td>0.70</td>
</tr>
<tr>
<td>TA98</td>
<td>0.46</td>
<td>0.89</td>
<td>0.88</td>
<td>0.69</td>
</tr>
<tr>
<td>TA100</td>
<td>0.52</td>
<td>0.87</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Ald</td>
<td>0.61</td>
<td>1.0</td>
<td>0.85</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Amm**: aromatic ammines; **Ald**: α-β unsaturated aldehydes

**Training set**: **Sqcc**: Squared Canonical Corr.; **acctra**: Accuracy; **acc10**: Accuracy L-10%-O

**Test set**: **Accte**: Accuracy
Local, mechanistically-based QSARs for congeners:

- scientifically interpretable, good internal statistics, but vary for their external predictivity

- **QSARs for potency**: predictions 30 – 70 % correct

- **QSARs for activity**: predictions 70 – 100 % correct

- Estimating intervals more reliable than estimating data points

- Internal validation measures do not correlate with external predictivity

Putting QSAR predictivity into context:

Intra-Assay agreement for the Ames test: 80 – 85%

Piegorsch and Zeiger, 1990

Local QSARs for activity: 70 – 100 % correct external predictions
• However, the number of available local QSARs for congeneric series is very limited (because of lack of experimental data, etc.)

• Thus, room for SAs or noncongeneric QSARs
Another class of models:

(Q)SARs for noncongeneric chemicals

- Large sets of chemicals from different classes, acting through different mechanisms

- No simple model possible
Another class of models:

(Q)SARs for noncongeneric chemicals

Tendency to sort out models by:

• Starting from thousands of descriptors
  (e.g., Dragon software http://michem.disat.unimib.it/chm/)

• Streamlining with sophisticated statistics to avoid chance correlations
Another class of models:

(Q)SARs for noncongeneric chemicals

- Aimed at modelling simultaneously all (??) chemical classes
- Many commercial systems
- Often non-mechanistically based descriptors
- Often no mechanistic interpretation
- Mostly validated through internal statistics only
• **External predictivity** of models for noncongeneric chemicals
External validation exercises

Prospective prediction exercises held under the aegis of the US National Toxicology Program (NTP):

- Predictions on the *rodent carcinogenicity* (two studies) and *S. typhimurium mutagenicity* (one study) of chemicals that were in the process of being assayed by the NTP

- Experimental results unknown at the time of prediction !!!!
### NTP mutagenicity prediction exercise

**(100 chemicals)**

<table>
<thead>
<tr>
<th>System</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA</td>
<td>0.72</td>
</tr>
<tr>
<td>SA+</td>
<td>0.76</td>
</tr>
<tr>
<td>TOPKAT</td>
<td>0.74</td>
</tr>
<tr>
<td>CASE/n</td>
<td>0.76</td>
</tr>
<tr>
<td>CASE/e</td>
<td>0.76</td>
</tr>
<tr>
<td>Ke</td>
<td>0.60</td>
</tr>
<tr>
<td>Ke-b</td>
<td>0.61</td>
</tr>
</tbody>
</table>

**Diagram:**

- **False positive rate** vs. **True positive rate**
- **Systems:**
  - SA (Ashby)
  - SA+
  - TOPKAT
  - CASE/n
  - CASE/e
  - Ke
  - Ke-b
NTP-1 carcinogenicity prediction exercise
(44 chemicals)

<table>
<thead>
<tr>
<th>System</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tennant/Ashby</td>
<td>0.75</td>
</tr>
<tr>
<td>RASH</td>
<td>0.68</td>
</tr>
<tr>
<td>Weisburger</td>
<td>0.65</td>
</tr>
<tr>
<td>Ke</td>
<td>0.65</td>
</tr>
<tr>
<td>DEREK</td>
<td>0.59</td>
</tr>
<tr>
<td>TOPKAT</td>
<td>0.57</td>
</tr>
<tr>
<td>Benigni</td>
<td>0.57</td>
</tr>
<tr>
<td>S. typhimurium</td>
<td>0.57</td>
</tr>
<tr>
<td>DEREKh</td>
<td>0.56</td>
</tr>
<tr>
<td>Lijinsky</td>
<td>0.55</td>
</tr>
<tr>
<td>COMPACT</td>
<td>0.53</td>
</tr>
<tr>
<td>CASE/MCASE</td>
<td>0.49</td>
</tr>
</tbody>
</table>
NTP-2 carcinogenicity prediction exercise
(30 chemicals)

False positive rate

True positive rate

Table:

<table>
<thead>
<tr>
<th>System</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>OncoLogic</td>
<td>0.65</td>
</tr>
<tr>
<td>SHE</td>
<td>0.65</td>
</tr>
<tr>
<td>R1</td>
<td>0.64</td>
</tr>
<tr>
<td>Huff et al.</td>
<td>0.62</td>
</tr>
<tr>
<td>R2</td>
<td>0.61</td>
</tr>
<tr>
<td>Benigni et al.</td>
<td>0.61</td>
</tr>
<tr>
<td>Tennant et al.</td>
<td>0.60</td>
</tr>
<tr>
<td>Ashby</td>
<td>0.57</td>
</tr>
<tr>
<td>Bootman</td>
<td>0.53</td>
</tr>
<tr>
<td>FALS</td>
<td>0.50</td>
</tr>
<tr>
<td>RASH</td>
<td>0.45</td>
</tr>
<tr>
<td>COMPACT</td>
<td>0.43</td>
</tr>
<tr>
<td>DEREK</td>
<td>0.43</td>
</tr>
<tr>
<td>S. typhimurium</td>
<td>0.33</td>
</tr>
<tr>
<td>Purdy</td>
<td>0.32</td>
</tr>
<tr>
<td>Progol</td>
<td>0.29</td>
</tr>
<tr>
<td>MULTICASE</td>
<td>0.25</td>
</tr>
</tbody>
</table>
• **External predictivity** of some popular commercial models for noncongeneric chemicals
Predictive Toxicology - DS TOPKAT

A QSAR-based system generates and validates accurate, rapid assessments of chemical toxicity solely from a chemical's molecular structure. Unique among SAR-based technologies, DS TOPKAT uses robust, cross-validated models based on experimental data of highly consistent protocol. The models are subjected to extensive diagnostics for accuracy and validity. And only DS TOPKAT uses patented Optimum Prediction Space (OPS) technology to assure that the compounds under investigation are well represented in the models. Included within DS TOPKAT are tools that allow you to easily build molecules or queries from available fragment libraries. DS TOPKAT can be used for tests including physical/chemical, environmental fate, ecotoxicity, toxicity, mutagenicity, and subchronic reproductive/developmental. DS TOPKAT is fast, cost-effective, and proven. DS TOPKAT Models include:

- Rodent Carcinogenicity
- Ames Mutagenicity
- Rat Oral LD50
- Rat Chronic LOAEL
- Developmental Toxicity Potential
- Skin Sensitization
- Fathead Minnow LC50
- Daphnia Magna EC50
- Weight of Evidence Rodent Carcinogenicity
- Rat Maximum Tolerated Dose
- Aerobic Biodegradability
- Eye Irritancy
- Log P
- Rabbit Skin Irritancy
- Rat Inhalation Toxicity LC50
- Rat Maximum Tolerated Dose

Accelrys has recently signed collaborations with the Bioinformatics and Molecular Design Research Center at Caltech, the California Institute of Technology, that will lead to the development of an integrated computational and experimental platform for the rational design and development of safer drugs.
TOPKAT

http://accelrys.com/products/discovery-studio/toxicology

• QSAR-like model for noncongeneric chemicals

• A range of toxicity end-points (mutagenicity, carcinogenicity, oral toxicity, skin and eye irritation, etc…)

• Descriptors: pre-defined list of substructures, and continuous descriptors (topological shape and symmetry, electrotopological e-state, etc…)

• Check for the Optimum Prediction Space
TOPKAT: External Validation

Mutagenicity

Carcinogenicity

False positive rate

True positive rate

PEARL 2001a
PEARL 2001b
CARIELLO 2002
MUELLER 2000
SNYDER 2004
NTP

PEARL 2001a
PEARL 2001b
NTP1
PRIVAL 2001

False positive rate

True positive rate

TOPKAT
Mutagenicity External Validation
False Positive Rate
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
True Positive Rate
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Pearl 2001a
Pearl 2001b
NTP

Carcinogenicity External Validation
False Positive Rate
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
True Positive Rate
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Pearl 2001a
Pearl 2001b
NTP

Prival 2001
Welcome to MultiCASE Inc., the leading provider of artificial intelligence based bioactivity software for the Chemical and Pharmaceutical Industries.

Our Mission

We develop and license computer programs designed to help a user to assess the potential pharmacological activity, toxicity and metabolic transformation of new chemicals.

Company Profile

MultiCASE Inc. (formerly BIOSOFT Inc.) is a Software company started in Cleveland Ohio in 1996. The principal is Gilles Kropman, the Charles F. Mabery Professor of Chemistry at Case Western Reserve University (CWRU). Its business is to develop and license programs for use by chemical and pharmaceutical companies to design new useful molecules and to evaluate their impact on the environment at an early stage of development.

Latest News

- We have acquired a new set of FDA ICSAS teratogenicity modules.
- List of modules was updated to include all newly acquired and updated in 2006 modules.
- MCWeb - an online version of MC4PC and CaseTox was launched.
- ACS Expo in Atlanta. See details.
- ToxExpo2006 in San Diego. See details.
- ToxExpo2005 in New Orleans. See details.
- Fifth contributor joined our Mutagenicity Data Sharing consortium. See details.
- Renewed CRADA with the FDA...See details.
- First release of ToxLite - the most affordable version of CASETOX. See details.
- Three more participants joined the Mutagenicity Data Sharing Consortium. See details.
- ToxExpo 2004 in Baltimore. See details.
- MultiCASE Inc was recognized by the Weatherhead 100 as one of the fastest growing companies. See details.
MULTICASE

http://www.multicase.com/

• QSAR-like model for noncongeneric chemicals

• Open list of toxicity end-points (mutagenicity, carcinogenicity, oral toxicity, aquatic toxicity, teratogenicity, etc…)

• Descriptors: computer-automated generation of substructures, plus general descriptors (e.g., lipophilicity, electronic)

• Applicability Domain checked through similarity criteria
MULTICASE: External Validation

**Mutagenicity**

- True positive rate vs. False positive rate

**Carcinogenicity**

- True positive rate vs. False positive rate

Data points include:
- **Pearl 2001a**
- **Pearl 2001b**
- **Novartis**
- **amines**
- **Snyder 2004**
- **NTPn**
- **NTPe**
- **NTP2**
- **NTP1**
General Information

Derek for Windows is an expert knowledge base system that predicts whether a chemical is toxic in humans, other mammals and bacteria. The application is a high throughput screen for the endpoints listed below.

What is an Expert Knowledge Base System?

Expert knowledge base systems are sometimes confused with database management systems such as Vitic. In fact the two types of applications are very different.

In toxicology:
A database is a large set of structured toxicology data and a database management system is a computer program designed to manage a database.

An expert knowledge base system is a computer program that contains expert knowledge rules in toxicology and applies the rules to make predictions about the toxicity of chemicals, usually when no experimental data is available.

Endpoints:

The program indicates potential toxicity for many toxicological endpoints, including:

- Carcinogenicity
- Mutagenicity
- Genotoxicity
- Skin Sensitisation
- Teratogenicity
- Irritancy
- Respiratory Sensitisation
- Hepatotoxicity
- Neurotoxicity
- Ocular Toxicity
DEREK

http://www.lhasalimited.org/

• Expert system (If … then … rules)

• Descriptors: Structure Alerts plus Modifying Factors
  Mechanistic explanation

• End-points: Mutagenicity, Carcinogenicity, Genotoxicity, Skin Sensitization, Teratogenicity, Irritancy, etc…

• Applicability Domain checked through similarity
Large performance variability in different regions of the chemical space

- Insufficient rules for Applicability domain
- Insufficient representation of mechanisms in the training chemicals
(Q)SAR does not replace reality; however, it provides powerful scientific support

“...As the drug discovery process is of a very complex nature, effective drug design requires an entire spectrum of techniques in which QSAR methods still play an important role. …

The real power of drug design methods is to extract and synthesize information from data to obtain hypotheses that can be put to experimental test. No dramatic overnight discoveries of wonder drug will result, but an increase in the chance of success due to indications of promising directions is a realistic expectation....”

General References on QSAR


References on Carcinogenicity / Mutagenicity


References on QSAR and Regulatory Issues: Carcinogenicity / Mutagenicity


References on QSAR and Regulatory Issues: Phys-Chem and Toxicity (general)

