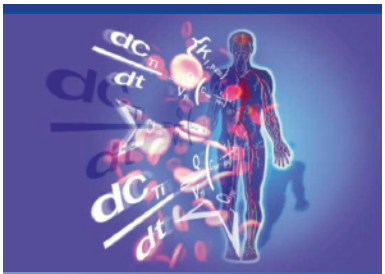
cyprotex AN EVOTEC COMPANY

In Silico Prediction

chemTox - Toxicity Prediction Directly from Chemical Structure

Background Information



'A virtual screening tool for predicting toxicity from chemical structure alone.'

- Predicts key toxicity parameters (Ames mutagenicity, rat acute dose LD_{50} following iv or po administration) and aqueous solubility directly from structure.
- No *in vitro* physicochemical or toxicity data required
- Save money and time by allowing toxicity to be assessed virtually (no synthesis required).
- Provides early stage filter for directing chemistry and prioritising screening.

chemTox Input Requirements

Chemical structure, e.g., SMILES, mol or sdf

chemTox Data Delivery

- Predicted toxicity measures: Ames mutagenicity; rat acute dose LD_{50} following iv or po administration.
- Predicted aqueous solubility

How does it work?

chemTox is implemented as a node for the KNIME analytics platform which executes a model of the workflow illustrated in Figure 1 below.

Figure 1
chemTox workflow.

‘KNIME can be downloaded easily and for free. Cyprotex then provide the bespoke toxicity nodes (chemTox) for the KNIME platform.’

Properties

The following properties are reported:

1. Ames mutagenicity classification (mutagenic/non-mutagenic).
2. Ames mutagenicity probability (probability of being a mutagen).
3. Rat LD_{50} (mmol/kg) following acute administration by intravenous route.
4. Rat LD_{50} (mmol/kg) following acute administration by oral route.
5. Aqueous solubility (mol/l).

Output is a KNIME data table facilitating saving to a file or database, or using the predictions as inputs to subsequent workflow steps.

Model Development

- Models are quantitative-structure property relationships (QSPR) that calculate the toxicity properties of interest in terms of a compound’s structural descriptors.
- Models have been trained using large, well-validated datasets.
- Training was performed using up-to-date, rigorous statistical pattern recognition methods.
- Repeated 10-fold cross-validation has been used to generate the most robust statistics for prediction performance.

Figure 2

Receiver operating characteristic (ROC) plots for classification of rat acute LD_{50} following oral administration. (a). LD_{50} < 5mg/kg, (b). LD_{50} < 50mg/kg, (c). LD_{50} < 300mg/kg.

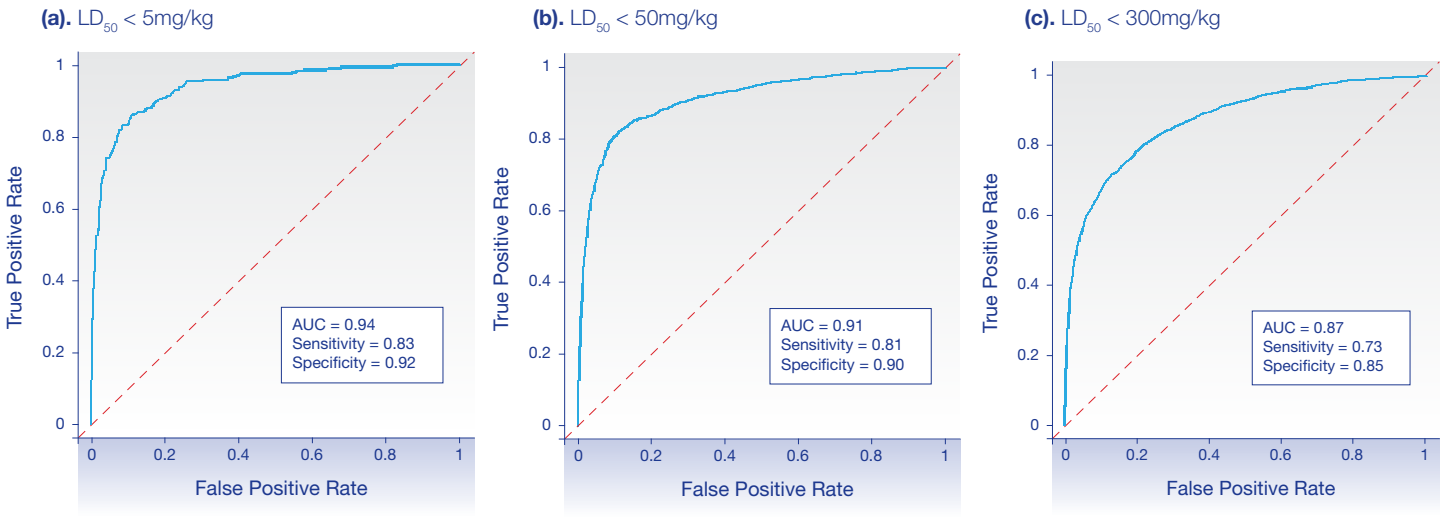


Table 1

Statistics for predicted Ames mutagenicity based on 4336 compounds (1935 non-mutagenic and 2401 mutagenic).

Statistics	
Area under ROC curve	0.905
Sensitivity	0.85
Specificity	0.82

Model variants can be generated having different balance of sensitivity and specificity to suit different screening requirements.

Contact enquiries@cyprotex.com for a demo.