

ToxGPS Prediction

Expert System for Toxicity Predictions

ToxGPS® Prediction is a toxicity knowledgebase providing easy-to-use *in silico* prediction models for the major human health- and regulatory relevant endpoints. The predictions are based on mechanistically informed, probabilistic QSAR models and endpoint-specific structural knowledge (expert rules) which are combined in a final weight-of-evidence approach to minimize the uncertainty in the prediction.

Toxicity Endpoints

- Genetic toxicity
 - Bacterial reverse mutagenicity (Ames), *in vitro* chromosome aberration, *in vivo* micronucleus
- Carcinogenicity
 - Rat or mouse tumorigenicity
- Developmental and reproductive toxicity
 - Cleft palate, pregnancy loss
- Dermal toxicity
 - Skin irritation and sensitization (hazard & potency)
- Hepatotoxicity
 - Mammalian steatosis and mitochondrial toxicity, human drug-induced liver injury
- Acute toxicity (consumer products)

Training Sets

- Data preferentially collected from regulatory sources: US NTP, EPA and FDA, SCCS, HESS, ECETOC, ECHA, EFSA and open literature
- Different training sets for QSAR models and chemotype alerts (rule-based prediction), *e.g.*, for bacterial reverse mutagenicity
 - 2,864 compounds for QSAR models
 - 9,391 compounds for chemotype alerts
- Further details available in the respective ToxGPS model documentations



Prediction Components

- Combined QSAR and rule-based system with probabilistic outcome by weight-of-evidence
 - Global and MOA-based QSAR models including applicability domain check
 - Endpoint-specific chemotype alerts including severity scores (odds ratios)
- Nearest neighbors analysis of query linked to study data in ChemTunes Database

Modeling Approach

- | | |
|-------------------------------|---|
| Statistical Modelling Methods | ■ Partial least squares & logistic regression |
| | ■ <i>k</i> -nearest neighbors |
| | ■ Random forest |
| Descriptors | ■ Structural rules (chemotypes) |
| | ■ Physicochemical descriptors and QM parameters |
| | ■ Biological assays |
| Training Set | ■ Curated structures and data |
| | ■ Diversity for broad applicability domain |
| | ■ Mechanistic/MOA categories |