A Comprehensive Approach for *in Silico* Risk Assessment of Impurities and Degradants in Drug Products

Pranas Japertas[1], Kiril Lanevskij[1,2], Liutauras Juska[1,2], Justas Dapkūnas[1,2], Andrius Sazonovas[1], Remigijus Didziapetris[1]

[1] ACD/Labs, Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania,

[2] Department of Biochemistry and Biophysics, Vilnius University, M.K.Ciurlionio g. 21/27, LT-03101 Vilnius, Lithuania.

INTRODUCTION

According to the FDA Guidance for Industry, impurities identified below the ICH qualification thresholds may be evaluated for genotoxicity and carcinogenicity based on structural activity relationship (SAR) assessments using computational software. The aim of this study was to develop a comprehensive *in silico* approach to aid this assessment.

An *in silico* package for impurity profiling, presented here, is a result of the collaboration between ACD/Labs and FDA Center for Food Safety and Nutrition (CFSAN). Evaluation of genotoxic and/or carcinogenic potential is based on a battery of probabilistic models for bioassays reflecting different mechanisms of hazardous activity. A knowledge-based expert system identifies potentially hazardous structural fragments that could be responsible for carcinogenic activity of the test molecule.

EXPERIMENTAL DATA

A complete list of modeled endpoints is provided in Table 1, while the data sources are briefly described below.

Genetic toxicity: data sets for standard assays reflecting different mechanisms of genetic damage were obtained from the FDA. Gene mutation tests and techniques detecting clastogenic/aneugenic effects are included. Data was collected from EPA GENE-TOX database and scientific literature [1].

Carcinogenicity: results of chronic (two-year term) carcinogenicity studies in rats and mice were received from FDA. This data was based on NTP technical reports, IARC monographs, Carcinogenic Potency DataBase [2] and other publicly available sources. Raw data was converted to binary classification using a weight of evidence (WOE) approach [1]. Classification using the WOE threshold corresponding to "potent carcinogens" was used to build the models in the current study.

Reproductive toxicity: experimental data characterizing the potential for endocrine system disruption due to Estrogen receptor α binding were acquired from ChEMBL database [3]. Compounds were classified as binders/non-binders on the basis of their relative binding affinities (RBA) compared to reference ligand estradiol. Two cut-offs were used: LogRBA > -3 ("general binding"), and LogRBA > 0 ("strong binding").

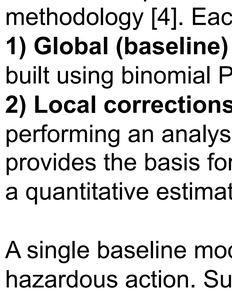
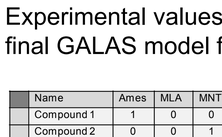
Mechanism	Test system	Endpoint	N	% Positives
Genetic toxicity				
Mutagenicity 	Prokaryote	Composite	7953	4003 (50.3%)
		Salmonella	7826	3875 (49.5%)
		Escherichia	1479	386 (26.1%)
	Eukaryote	Composite	2901	1592 (54.9%)
		Yeast	658	347 (52.7%)
		Drosophila	600	293 (48.8%)
		MLA	1272	763 (60.0%)
		CHO/CHL all loci	1229	585 (47.6%)
Clastogenicity 	Chromosome aberrations	*In vitro*	2034	941 (46.3%)
		In vivo	441	133 (30.2%)
	Micronucleus test in rodents	*In vivo*	1299	403 (31.0%)
DNA damage	UDS	*In vivo/in vitro*	593	166 (28.0%)
Carcinogenicity	Rodent	Composite	2211	674 (30.5%)
	Rat	Male	1818	647 (35.6%)
		Female	1793	635 (35.4%)
	Mouse	Male	1669	556 (33.3%)
		Female	1727	561 (32.5%)
Reproductive toxicity	Estrogen receptor binding	LogRBA > 0	3423	1488 (43.5%)
		LogRBA > -3	3423	2549 (74.5%)

TABLE 1. A list of bioassays considered and dataset sizes in the study.

METHODS

Probabilistic predictive models for all considered endpoints were developed using GALAS modeling methodology [4]. Each GALAS model consists of two parts:
1) Global (baseline) model that reflects general trends in the property of interest. Baseline models were built using binomial PLS method based on fragmental descriptors.
2) Local corrections were applied to baseline predictions using a special similarity-based routine, after performing an analysis for the most similar compounds used in the training set. The local part of the model provides the basis for the calculation of the **Reliability index (*RI*)**, a value ranging from 0 to 1 that provides a quantitative estimate of prediction accuracy.

A single baseline model was derived for each group of endpoints representing the same mechanism of hazardous action. Such model reflects a "cumulative" toxicity potential of chemicals in these assays. Experimental values specific for a particular assay were used during the local part of the modeling to yield final GALAS model for that endpoint. An outline of the modeling procedures is presented in Fig. 1.

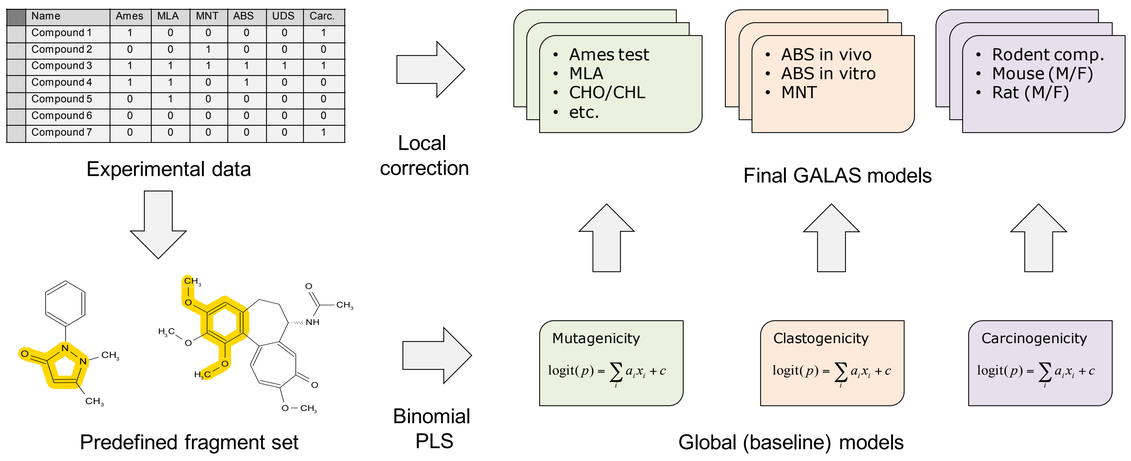


FIGURE 1. A schematic outline of the model development workflow.

GENOTOXICITY/CARCINOGENICITY HAZARDS

The knowledge-based expert system that identifies structural fragments potentially responsible for genotoxic effect is an extension of the previously described Ames mutagenicity hazards system [5]. The list of alerting groups was augmented with structural moieties that are frequently present in compounds tested positive in chromosomal damage assays, eucaryote gene mutation tests, as well as in carcinogens acting by non-genotoxic (epigenetic) mechanisms. The final list included 67 structural alerts, 14 of which represent epigenetic carcinogens (androgens, peroxisome proliferators, etc.).

Overall, the expert system was able to detect 94% of mutagens in the Ames test DB and 90% of compounds labeled as potent carcinogens by FDA (Fig. 2).

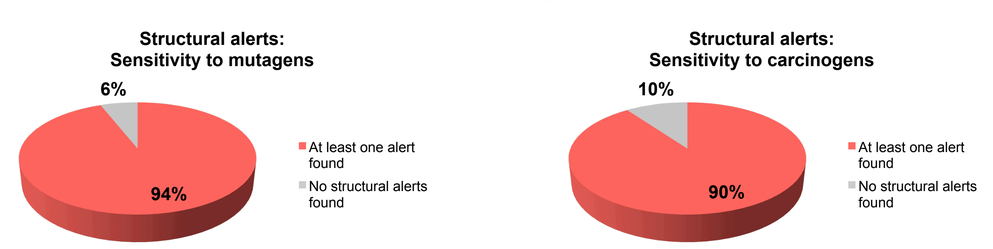


FIGURE 2. Results of genotoxicity/carcinogenicity hazard identification using the presented expert system.

The alert list is not limited to directly acting substructures, such as planar polycyclic arenes, aromatic amines, quinones, N-nitro and N-nitroso groups, but also includes various fragments that may undergo biotransformation to reactive intermediates. As an example, troglizatone, a thiazolidinedione class antidiabetic drug, was classified by the FDA as a potent carcinogen and has since been withdrawn from the USA market. The carcinogenic effect of this drug is mediated by several reactive metabolites. In human liver microsomes, the chromane ring of troglitazone is metabolized by CYP3A4 to form quinone and quinone-methide products. Furthermore, oxidative cleavage of thiazolidinedione ring results in a reactive sulfenic acid metabolite that also contains an isocyanate moiety [6]. As shown in Fig. 3, both bioactivation pathways are predicted by the Hazards identification system presented here.

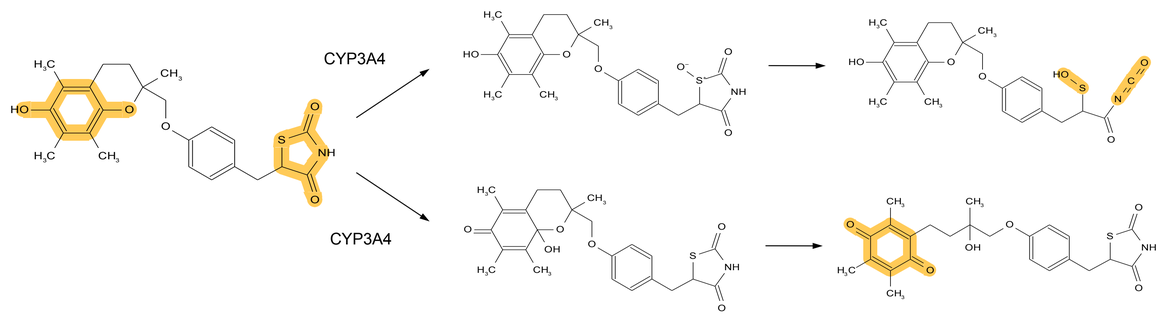


FIGURE 3. Biotransformation of troglitazone in human liver microsomes.

SOFTWARE PACKAGE FOR IMPURITY PROFILING

The Profiling system for impurities and degradants described here is a part of ACD/Tox Suite (info@acdlabs.com).

1. Probabilistic predictors

The output of probabilistic models for all considered endpoints consists of the following parts (Fig.4):

- ***p*-value** – probability that a compound will result in a positive test in the respective assay
- **Coverage** – an indication whether the compound belongs to Model Applicability Domain according to calculated *RI* value
- **Call** – (+ or –) if the compound can be reliably classified on the basis of *p* and *RI* values, "Undefined" otherwise.

2. Hazard identification system

Each hazardous fragment is provided with a short description of its mechanism of action, literature references, and z-scores. *Z*-scores show whether the presence of the fragment leads to a statistically significant increase in proportion of compounds with a positive test result for a particular assay. This information provides further evidence regarding the possible mechanisms of action.

For example, acrylic acid derivatives do not cause direct DNA damage but are primarily reactive towards sulfhydryl groups of proteins (including those involved in DNA replication/maintenance). As a result, most acrylates are negative in reverse point-mutation assays, such as in the Ames test. Yet, they cause chromosomal aberrations and produce positive results in forward mutation tests, such as Mouse Lymphoma Assay (Fig. 5).

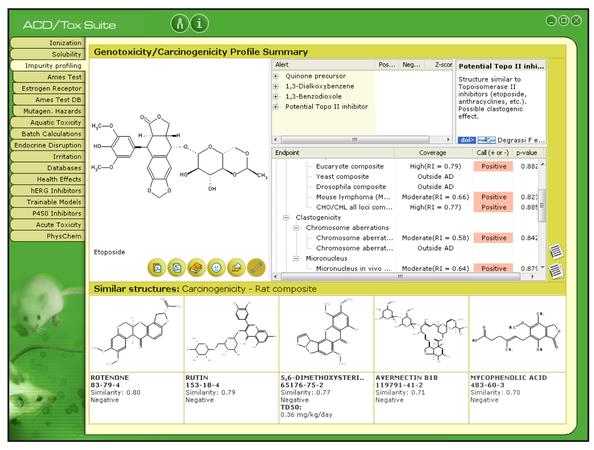


FIGURE 4. Package for Toxicity Screening of Impurities in ACD/Tox Suite: Probabilistic predictions.

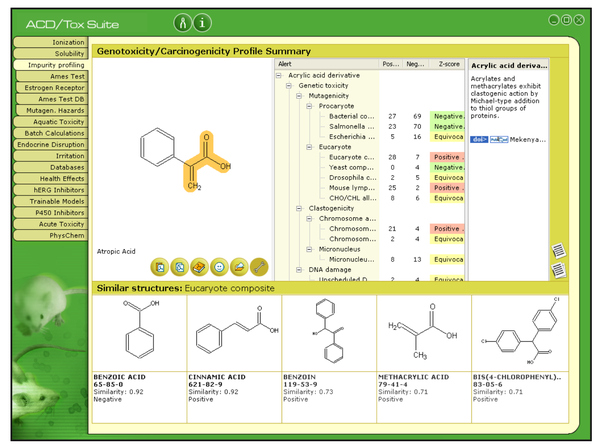


FIGURE 5. Package for Toxicity Screening of Impurities in ACD/Tox Suite: Structural alerts.

REFERENCES

[1] Matthews EJ et al. *Regul Toxicol Pharmacol*. **2007**, 47, 115.
[2] Gold LS et al. *Toxicol Sci*. **2005**, 85, 747.
[3] ChEMBL DB (https://www.ebi.ac.uk/chembldb). Target ID 206.
[4] Didziapetris R et al. *J Comput Aided Mol Des*. **2010**, 24, 891.
[5] Didziapetris R et al. *Toxicol Lett.* **2008,** 180, S152.
[6] Mansuy D & Dansette PM. *Arch Biochem Biophys*. **2011**, 507, 1745.

ACD/Labs
Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com