

An In Silico Test Battery for Rapid Evaluation of Genotoxic and Carcinogenic Potential of Chemicals

Kiril Lanevskij¹, Liutauras Juska^{1,2},
Justas Dapkunas¹, Andrius Sazonovas¹,
Pranas Japertas¹, Remigijus Didziapetris¹

¹ ACD/Labs, Inc., A.Mickevicius g. 29,
LT-08117 Vilnius, Lithuania,
² Department of Biochemistry and Biophysics,
Vilnius University, M.K.Ciurlonio 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.

This work is an extension of our previous study¹ focusing on computational assessment of genotoxic impurities in drug products. Our new approach relies on a battery of probabilistic QSAR models supplemented by a knowledge-based expert system that identifies structural fragments, potentially responsible for hazardous activity.

The analysis was based on experimental data obtained from FDA, and involved 21 endpoints corresponding to different mechanisms of toxic action: mutagenicity, clastogenicity, carcinogenicity, etc. Probabilistic models for most endpoints were derived using GALAS (Global, Adjusted Locally According to Similarity) modeling methodology developed in our group. The updated list of alerting groups contained 70 distinct substructures. The expert system was highly sensitive, recognizing >90% of potent carcinogens, as classified by the FDA. Sensitivity of probabilistic models ranged from about 60% to more than 90%, while maintaining high (>80%) specificity of predictions for the majority of considered assays. These results show that the described computational platform ensures sufficient prediction accuracy for rapid genotoxicity/carcinogenicity profiling of various chemicals.

EXPERIMENTAL DATA

A concise list of the considered endpoints is provided in Table 1, while 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.²

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, the Carcinogenic Potency DataBase³ and other publicly available sources. Raw data was converted to a binary classification using a weight of evidence (WOE) approach². 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 original publications. 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 (%Positive)	Sensitivity	Specificity
Genetic toxicity					
Mutagenicity	Prokaryote	Salmonella	7826 (49.5%)	87.1%	81.7%
		Escherichia	1479 (26.1%)	72.5%	87.0%
	Eukaryote	Composite	2901 (54.9%)	78.1%	64.0%
		Yeast	658 (52.7%)	86.7%	80.0%
		Drosophila	600 (48.8%)	70.6%	81.8%
		MLA	1272 (60.0%)	76.2%	64.7%
CHO/CHL all loci	1229 (47.6%)	80.0%	67.5%		
Clastogenicity	Chromosome aberrations	<i>In vitro</i>	2034 (46.3%)	74.6%	71.5%
	Micronucleus test in rodents	<i>In vivo</i>	1299 (31.0%)	62.1%	69.5%
DNA damage	UDS	<i>In vivo/in vitro</i>	593 (28.0%)	66.7%	76.3%
Carcinogenicity	Rodent	Composite	2211 (30.5%)	58.8%	81.9%
		<i>(Data from rats and mice, both genders)</i>			
Reproductive toxicity	Estrogen receptor binding	LogRBA > 0	1464 (51.6%)	82.9%	97.1%
		LogRBA > -3	1464 (23.8%)	92.1%	85.6%

TABLE 1. Key bioassays considered in the study, the respective dataset sizes, and model performance.

NOTE: Sensitivity and specificity values are provided for test set compounds within the Model Applicability Domain, as indicated by Reliability Index (RI) \geq 0.3, except Rodent carcinogenicity where statistics for the entire test set are given.

METHODS

Probabilistic predictive models for genetic and reproductive toxicity endpoints were developed using GALAS modeling methodology.⁴ Each GALAS model consists of two parts:

- Global (baseline) model** that reflects general trends in the property of interest. Baseline models were built using binomial PLS method based on fragmental descriptors.
- 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 **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 groups of endpoints representing different mechanisms of genotoxic action (mutagenicity and clastogenicity). This model reflects the "cumulative" toxic potential of chemicals in these assays. Experimental values specific to each assay were used in the local model to yield the final GALAS model for individual endpoints. An outline of modeling procedures is presented in Figure 1.

In case of carcinogenicity, predictive models were derived in a slightly different manner. Due to the complexity of the system, the local correction step was only used to produce RI values, while probability estimation also took into account the output of mutagenicity and reproductive toxicity models as well as the presence of specific hazardous fragments typical to carcinogens.

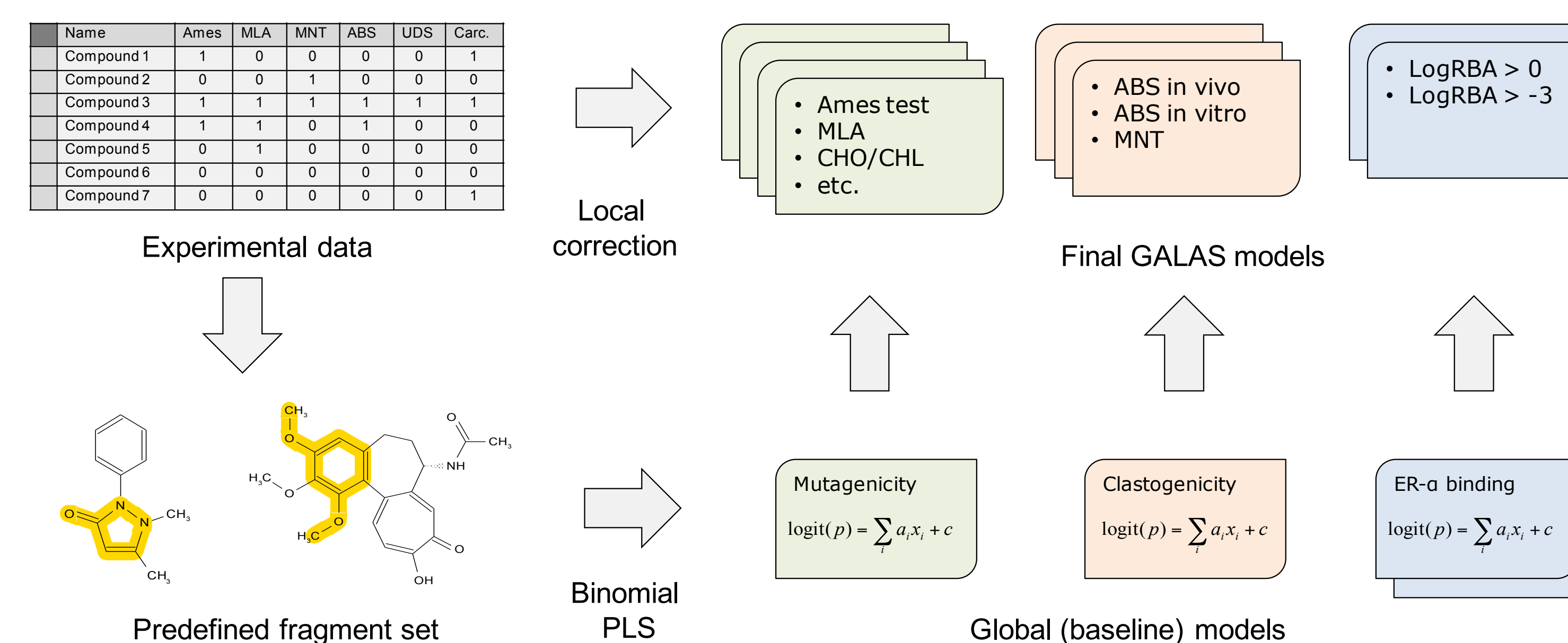


FIGURE 1. A schematic outline of the model development workflow.

GENOTOXICITY/CARCINOGENICITY HAZARDS

The knowledge-based expert system identifies structural moieties that are frequently present in compounds that tested positive in the Ames test, eucaryote gene mutation tests, and chromosomal damage assays, as well as in carcinogens acting by non-genotoxic (epigenetic) mechanisms. The final list includes 70 structural alerts, of which 33 represent mutagens, 24–clastogens, and 13–epigenetic carcinogens (androgens, peroxisome proliferators, etc.).

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, troglitazone, a thiazolidinedione class antidiabetic drug, was classified by the FDA as a potent carcinogen and has since been withdrawn from the US 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 the thiazolidinedione ring results in a reactive sulfenic acid metabolite that also contains an isocyanate moiety.⁴ As shown in Figure 2, both bioactivation pathways are predicted by the presented Hazards identification system.

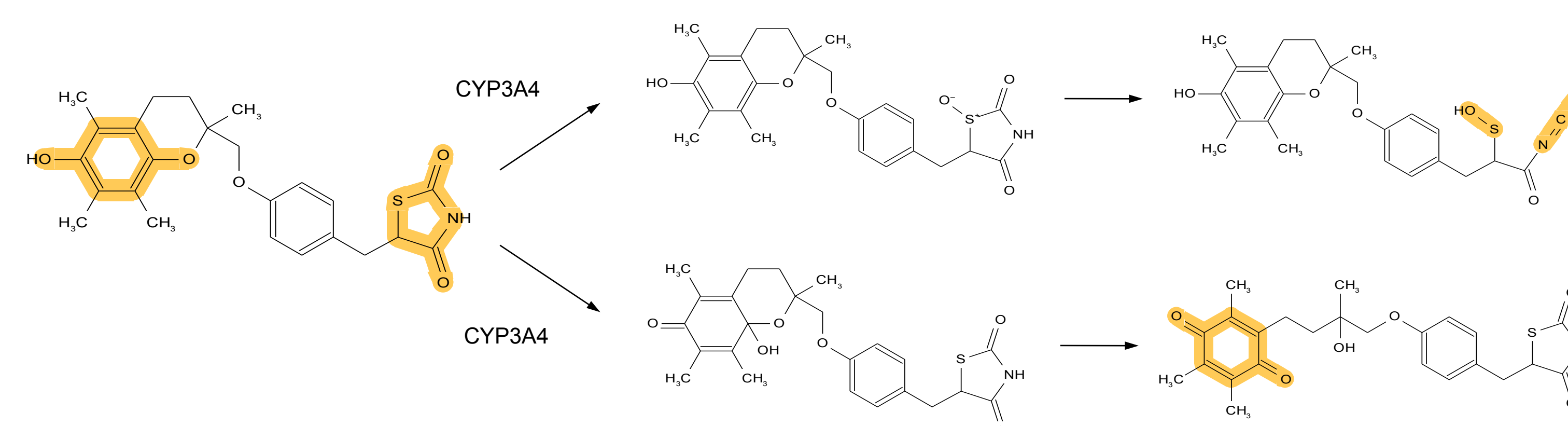


FIGURE 2. Biotransformation of troglitazone in human liver microsomes.

CLASSIFICATION RESULTS

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 the FDA (Fig. 3). The performance of probabilistic models for various endpoints is listed in Table 1. Evidently, sensitivity of predictions is very high in well studied systems, such as Ames Salmonella reverse mutation assay, while more work is needed to improve detection of active chemicals in less well understood systems (UDS) or complex multi-mechanism endpoints (carcinogenicity). Specificity is mostly close to, or exceeds, 80%. Lower values are observed in eucaryote assays, such as *in vivo* Mouse Lymphoma Assay (MLA), where hazardous chemicals may produce a negative outcome due to their inability to reach the site of action.

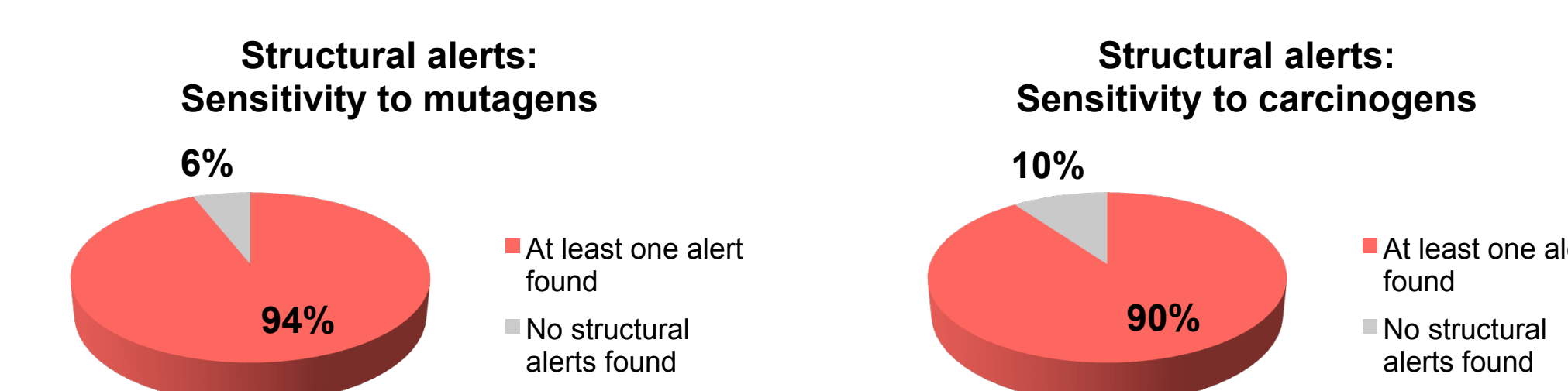


FIGURE 3. Results of genotoxicity/carcinogenicity hazard identification.

SOFTWARE PACKAGE FOR IMPURITY PROFILING

The described profiling system for impurities and degradants is available as a part of our ACD/Percepta software platform.

1. Probabilistic Predictors

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

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

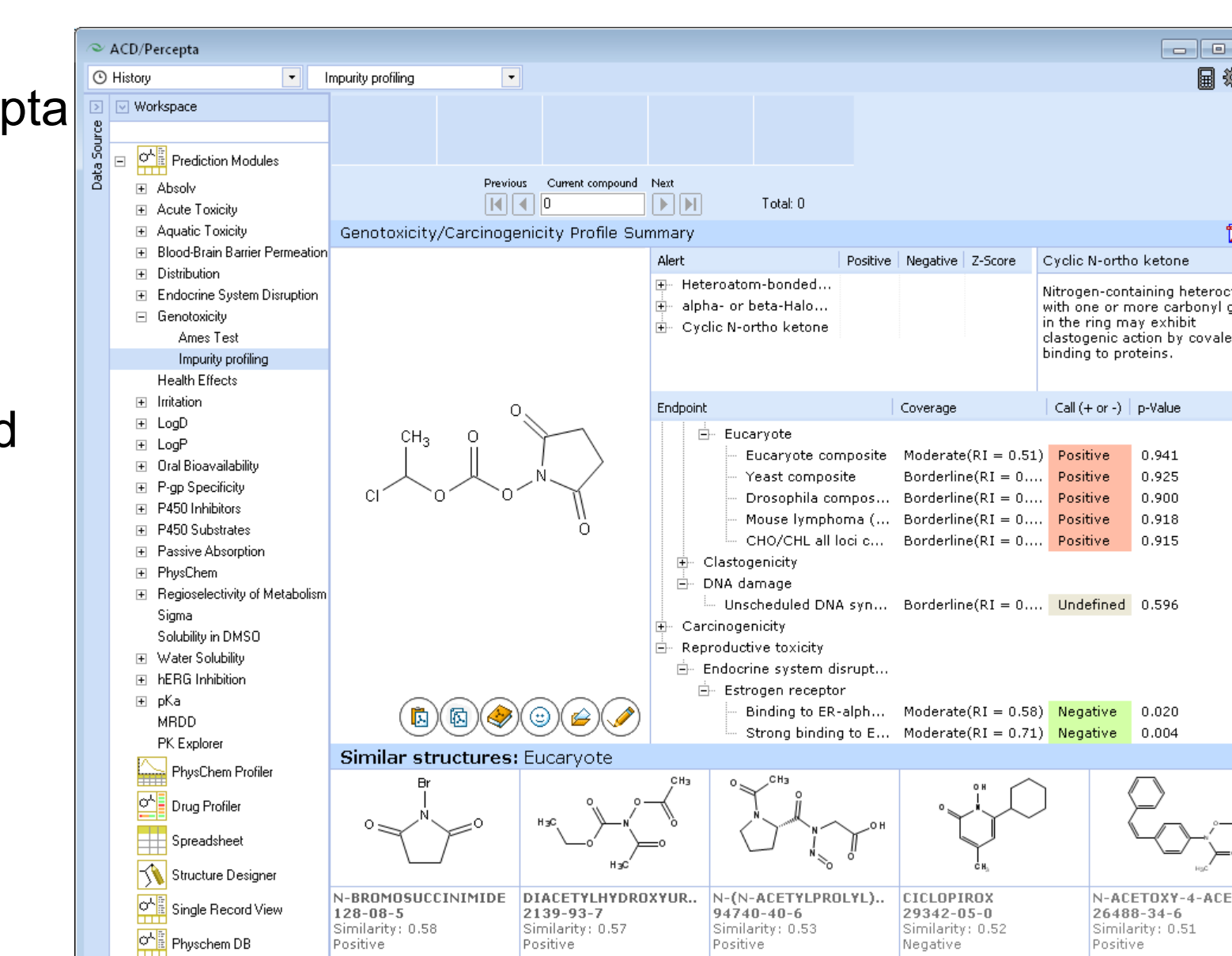


FIGURE 4. ACD/Percepta Impurity Profiling module: probabilistic predictions.

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 the 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 (Figure 5).

REFERENCES

- Japertas P et al. *Toxicol Lett.* **2011**, 205, S95.
- Matthews EJ et al. *Regul Toxicol Pharmacol.* **2007**, 47, 115.
- Gold LS et al. *Toxicol Sci.* **2005**, 85, 747.
- Didziapetris R et al. *J Comput Aided Mol Des.* **2010**, 24, 891.
- Mansuy D & Dansette PM. *Arch Biochem Biophys.* **2011**, 507, 1745.

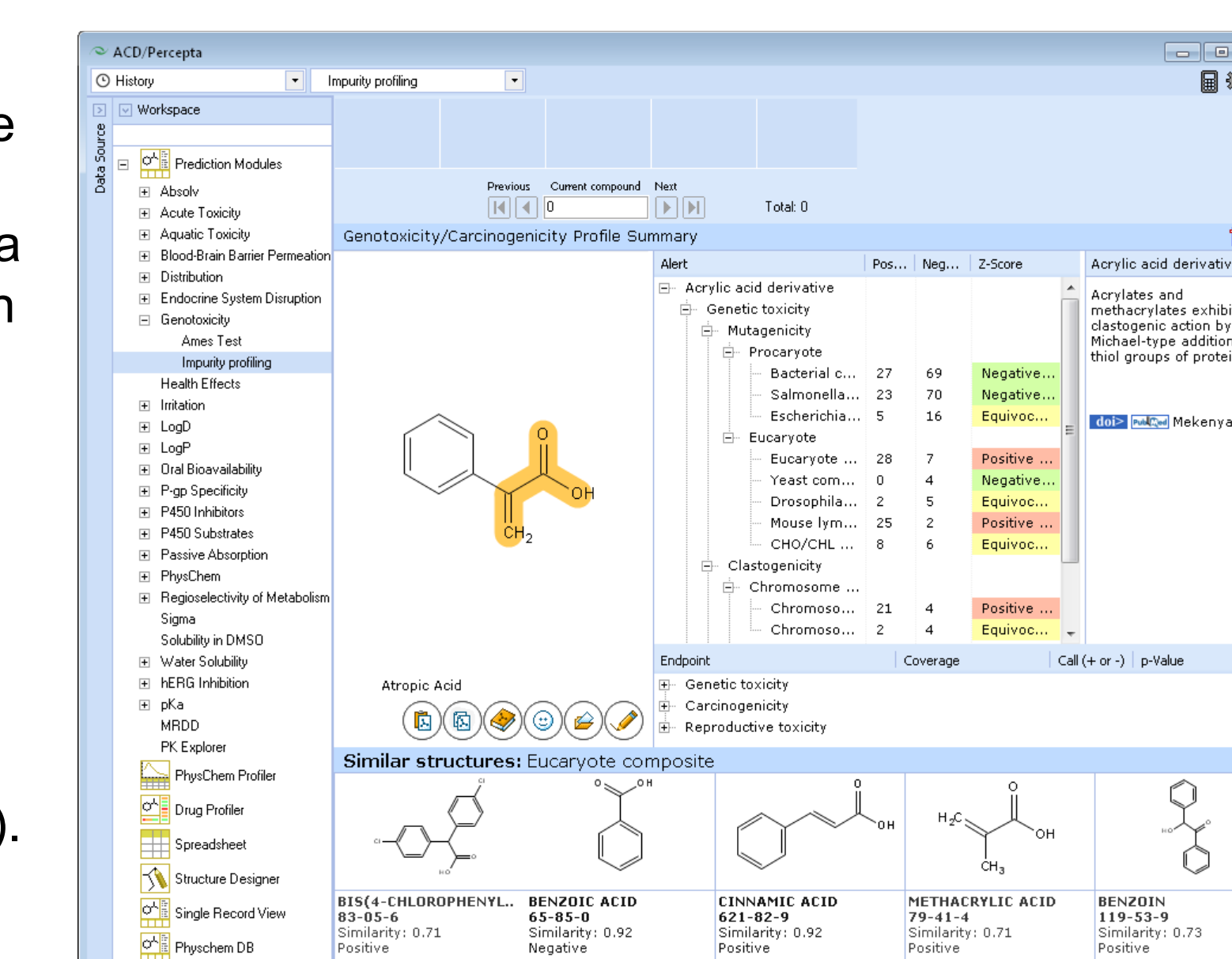


FIGURE 5. ACD/Percepta Impurity Profiling module: structural alerts.

