Toxicology and Applied Pharmacology 260 (2012) 209–221

Contents lists available at SciVerse ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

# Characterization and validation of an in silico toxicology model to predict the mutagenic potential of drug impurities*☆

Luis G. Valerio, Jr. [a,*], Kevin P. Cross [b]

[a] Science and Research Staff, Office of Pharmaceutical Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993–0002 USA
[b] Leadscope, Inc., 1393 Dublin Road, Columbus, OH, 43215–1084 USA

ARTICLE INFO

*Article history:*
Received 22 December 2011
Revised 24 February 2012
Accepted 2 March 2012
Available online 9 March 2012

*Keywords:*
Safety
Mutagenicity
QSAR
Modeling

ABSTRACT

Control and minimization of human exposure to potential genotoxic impurities found in drug substances and products is an important part of preclinical safety assessments of new drug products. The FDA's 2008 draft guidance on genotoxic and carcinogenic impurities in drug substances and products allows use of computational quantitative structure–activity relationships (QSAR) to identify structural alerts for known and expected impurities present at levels below qualified thresholds. This study provides the information necessary to establish the practical use of a new *in silico* toxicology model for predicting *Salmonella t.* mutagenicity (Ames assay outcome) of drug impurities and other chemicals. We describe the model's chemical content and toxicity fingerprint in terms of compound space, molecular and structural toxicophores, and have rigorously tested its predictive power using both cross-validation and external validation experiments, as well as case studies. Consistent with desired regulatory use, the model performs with high sensitivity (81%) and high negative predictivity (81%) based on external validation with 2368 compounds foreign to the model and having known mutagenicity. A database of drug impurities was created from proprietary FDA submissions and the public literature which found significant overlap between the structural features of drug impurities and training set chemicals in the QSAR model. Overall, the model's predictive performance was found to be acceptable for screening drug impurities for *Salmonella* mutagenicity.

Published by Elsevier Inc.

## Introduction

Control and minimization of human exposure to potential genotoxic impurities found in drug substances and products is an important part of preclinical safety assessments of new drug products. Organic drug impurities that arise during development can include starting materials, by-products, synthesis intermediates, degradation products and reagents. Generally impurities convey no patient benefit, and rationale for their reporting and qualification of safety to protect public health includes the consideration of genotoxic and carcinogenic potential. Therefore, efforts are made to achieve the lowest level technically feasible that would not convey significant health risks to humans.

Current regulatory guidance on identification, qualification, and reporting of impurities in drug substances, and degradants in drug products is addressed by the International Conference on Harmonisation (ICH) Q3A(R) and Q3B(R2), respectively (FDA, 2006, 2008a). ICH Q3C (FDA, 1997) addresses limits of residual solvents in pharmaceuticals. In addition, the FDA's 2008 draft guidance on genotoxic and carcinogenic impurities in drug substances and products describes ways to reduce patient exposure to potential genotoxic and carcinogenic impurities and provides ways for use of computational toxicology assessments including quantitative structure–activity relationships (QSAR) to identify structural alerts for known and expected impurities present at levels below ICH qualification thresholds. According to the FDA draft guidance, a computational assessment of drug impurities for potential genotoxicity and carcinogenicity may be conducted to help control low level drug impurities (FDA, 2008b). The importance of assessing bacterial mutagenicity of drug-related molecules in drug products and substance is further underscored in that the *Salmonella t.* (Ames assay) mutagenicity test is widely recognized as part of the strategy for the risk assessment of genotoxic impurities. Moreover, a number of studies have examined and shown how chemical structural alerts relate to Ames assay outcome to predict non-threshold genotoxic

☆ The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This paper reflects the current thinking and experience of the authors.

* Corresponding author at: Science and Research Staff, Office of Pharmaceutical Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, White Oak 51 Room 4128, 10903 New Hampshire Ave., Silver Spring, MD 20993–0002, USA. Fax: +1 301 796 9997.
*E-mail address:* luis.valerio@fda.hhs.gov (L.G. Valerio,).

0041-008X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.taap.2012.03.001

carcinogens (Ashby et al., 1989; Tennant and Ashby, 1991; Dobo et al., 2006; Benigni and Bossa, 2008), and how identification, prioritization, control and measurement of genotoxic impurities in drug products can be managed with the aid of computational (*in silico*) models for genotoxicity (Contrera et al., 2008; Contrera, 2011). Recently, a public workshop was held by the Drug Information Agency (DIA) and chaired by FDA/CDER to discuss what constitutes an acceptable computational assessment using (Q)SAR for assessing impurities for potential genotoxicity (DIA, 2011). Several considerations were raised as important criteria to judge the appropriate construction and use of a (Q)SAR model for mutagenicity including, chemical space, type of compounds covered, an emphasis on reliable negative predictions, and establishing scientific validity of a model through proper validation.

To address these considerations, the goal of this investigation was to characterize and validate a new *in silico* toxicology model of the Ames assay to predict the mutagenic potential of drug impurities. We use a statistical QSAR approach and describe the model in terms of its content, structural features and chemical space overlap in over a thousand drug impurities from FDA/CDER drug applications and public sources. Importantly we conduct a thorough validation of the model describing its predictivity statistics (sensitivity, specificity, coverage, etc.) using both internal cross-validation and rigorous external validation techniques. Given the historical application of SAR and structural alerts to the prediction of genotoxicity and carcinogenicity as exemplified by the classic "Ashby-Tennant alerts" and that such human expertise has served as one basis for development of new (Q)SAR approaches, we examine our data sets for these and other well known human knowledge-based alerts (*e.g.*, Bailey, Cronin, Cramer, Benigni) for comparative performance to our computational QSAR model to predict mutagenicity.

## Materials and methods

*Computational software.* Computational modeling and analysis of chemical structural information was performed using the Leadscope Enterprise software version 3.0 (www.leadscope.com, Leadscope Inc, Dublin, OH). QSAR models were run using Leadscope Model Applier software (version 1.3) and these technologies have been described elsewhere (Valerio et al., 2010). The software was made available to the FDA/CDER through an agency-approved Research Collaboration Agreement.

*QSAR training data set.* A data set of 3575 chemical structures tested in the *Salmonella t.* (Ames) reverse mutation assay was employed as the QSAR training set. The development of this training set has been previously described (Contrera et al., 2005). Briefly, the data set contains 1591 Ames assay positive (44%) and 1984 Ames assay (56%) negative chemicals. The training set contains approximately 26% pharmaceuticals (Contrera et al., 2005) but 94% of the compounds have drug-like properties (i.e. one or fewer Lipinski score violations). The study data are derived from Ames assays using ±S9 and the following tester strains; TA100, TA1535, TA1537, TA98, TA97, TA1538, TA1536. Data sources for these Ames assay studies originate from publically available databases including Drugs@FDA, the U.S. FDA Center for Food Safety Applied Nutrition's Priority based Assessment of Food Additive database, NIH/NLM Gentox database, National Institute of Environmental Health Genetox summary reports, public data from the U.S. Environmental Protection Agency Office of Pesticide Programs, and public human expert alerts assembled by Leadscope Inc. An extensive description of the data sources and transformation of assay results is provided previously (Matthews et al., 2006a,b).

*Assessment of chemical space* The Leadscope Enterprise software was deployed to enable a thorough examination and comparison of the chemical space of all data sets (training set of the model, external validation test set of the model, and drug impurities). One important assessment was to characterize the chemical space of these sets in terms of domain of applicability (DoA). Without establishing a DoA, there is no means of determining whether a compound was within or outside of the chemical space predicted using the rules addressed by the model (Mostrag-Szlichtyng et al., 2010b). DoA affects the reliability of both positively and negatively predicted compounds. Leadscope DoA measurement involves two steps that a test compound must pass in order to be predicted. First the structural features in a test compound must be at least 30% similar to that of at least one training set compound in the model. The Leadscope fingerprint of 27,142 structural features is used as the descriptors in the calculation which is performed using a Tanimoto similarity equation (Willett, 2006). After passing this global similarity test, at least one structural feature in the model must be present in the test compound. This is in addition to the calculated properties also used in the model and provides a local measure of similarity addressing relevance of the model features. Another measurement of chemical space used was the Leadscope fingerprint to globally compare structures in the test and training sets independently of model-only features. Fingerprints were used in an agglomerative clustering algorithm to produce clusters which were combined using the average linkage joining method. Each cluster was described by a substructural signature representing the cluster. Training and test set clusters were compared by looking for common members and the resulting overlap between the clusters was represented as a Venn diagram. In our figures we only reveal the signatures from public compounds in the cluster set. Positive and negative z-score features were calculated to identify mutagenic structural features and features which mitigate mutagenicity.

*Hierarchical assessment of mutagenic structural features.* The relative importance of individual features in the model was illustrated by creation of a hierarchical feature tree in Leadscope containing all model features. Major positive structural features (i.e. with positive z-scores for mutagenicity activity) from the model were organized hierarchically in order to visualize the parent/child relationships of features as nodes. For each feature node, we calculated the frequency of structures in the model matching the feature, the z-score representing the activity of the structure set matching the feature, the percentage of mutagenic structures in the entire data set that the feature matches (node sensitivity), and the mean activity of the structures matching the feature (node precision).

*External validation data set.* A total of 2368 chemicals from the public domain with known outcomes from testing in the *Salmonella* mutagenicity assay were obtained by Leadscope which comprises the external validation data set. The composition of the external validation test set included 1110 (47%) mutagenic chemicals, and 1258 (53%) non-mutagenic chemicals. These were assembled through analysis of the Leadscope Level-II Genetox database of high-quality, curated compound structures and overall *Salmonella* calls that are supported by multiple studies from reliable sources (Leadscope, 2010). None of the compounds in the model training set were included in the validation set. These chemicals served as an independent external validation test set to determine the predictive performance of the new Ames assay QSAR model model. The validation set was not constructed until after the QSAR model was created. Consequently, no knowledge from the validation set (*e.g.*, compound descriptors) was included in building the model. It is generally accepted that external validation is the best and most rigorous method for testing the accuracy of a predictive model since the chemicals used as the test set