Ensuring Regulatory Acceptable (Q)SAR Models and Expert Alerts for ICH M7 Reflect Proprietary Chemical Space

Kevin P. Cross[1], Naomi L. Kruhlak[2], Glenn J. Myatt[1], Lidiya Stavitskaya[2], Angela White[3]

[1]Leadscope, Inc., Columbus, OH; [2]FDA Center for Drug Evaluation and Research (CDER), Silver Spring, MD; [3]GlaxoSmithKline, Ware, UK

[The findings and conclusions in this presentation reflect the views of the authors and should not be construed to represent FDA's views or policies. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.]

Abstract

The International Conference on Harmonisation (ICH) M7 guideline permits the use of two *in silico* methodologies to qualify any actual or potential drug impurities as non-mutagenic. Statistical models and expert alerts built from public domain knowledge and data are used for ICH M7 assessments as they provide the necessary transparency and are sufficiently predictive for this purpose. Knowledge from proprietary data increases the specificity for selected chemical classes as well as expands the (Q)SAR models' applicability domain. This leads to less laboratory testing and synthesis of the impurity for the sponsor. We have investigated the use of proprietary data made available through confidentiality agreements directly with pharmaceutical sponsors to identify potential solutions to specific model predictivity issues, and are working towards the release of appropriate knowledge, compounds, and experimental data necessary to solve the (Q)SAR issue while preserving model transparency. The end result will be a regular transfer of knowledge for use in the development of (Q)SAR statistical models and incorporation of those data into the Leadscope alert-based expert system. This poster outlines the collaborative program for sharing data and knowledge for inclusion in regulatory acceptable public models without releasing any confidential business information. The process for knowledge sharing through the use of structural fingerprints for several compound classes will be discussed. A case study will be presented where, as a result of the incorporation of knowledge derived from proprietary data, the specificity of a model to predict primary aromatic amines was increased by 14% with no decrease in sensitivity.

ICH M7 and *in silico* Hazard Assessment

The International Conference on Harmonisation (ICH) recently issued a guidance document entitled "Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk"[1]. It is currently in the implementation phase (step 5) and is being adopted by the regulatory bodies of the European Union, Japan and USA. The document outlines the steps to identify, categorize, qualify and control DNA reactive impurities to ensure any impurity poses a negligible risk of carcinogenicity or other toxic effects. As part of hazard assessment the ICH M7 guidance permits the use of (Q)SAR models for predictions of bacterial mutagenicity to be used in place of an experimental assay[1]. As part of the submission, the guidance recommends the use of two (Q)SAR prediction methodologies, one expert rule-based and one statistical-based, which can be supplemented by an expert opinion. It also states that the two methodologies should adhere to the general principles set forth by the Organisation for Economic Co-operation and Development (OECD) in the development of the models[2]. One of these principles is "a defined domain of applicability", meaning that each model must initially check to see if it is capable of making a prediction for the type of test chemical. When a test chemical is not in the domain of applicability of the model, no prediction is made and the result is shown as "out of domain".

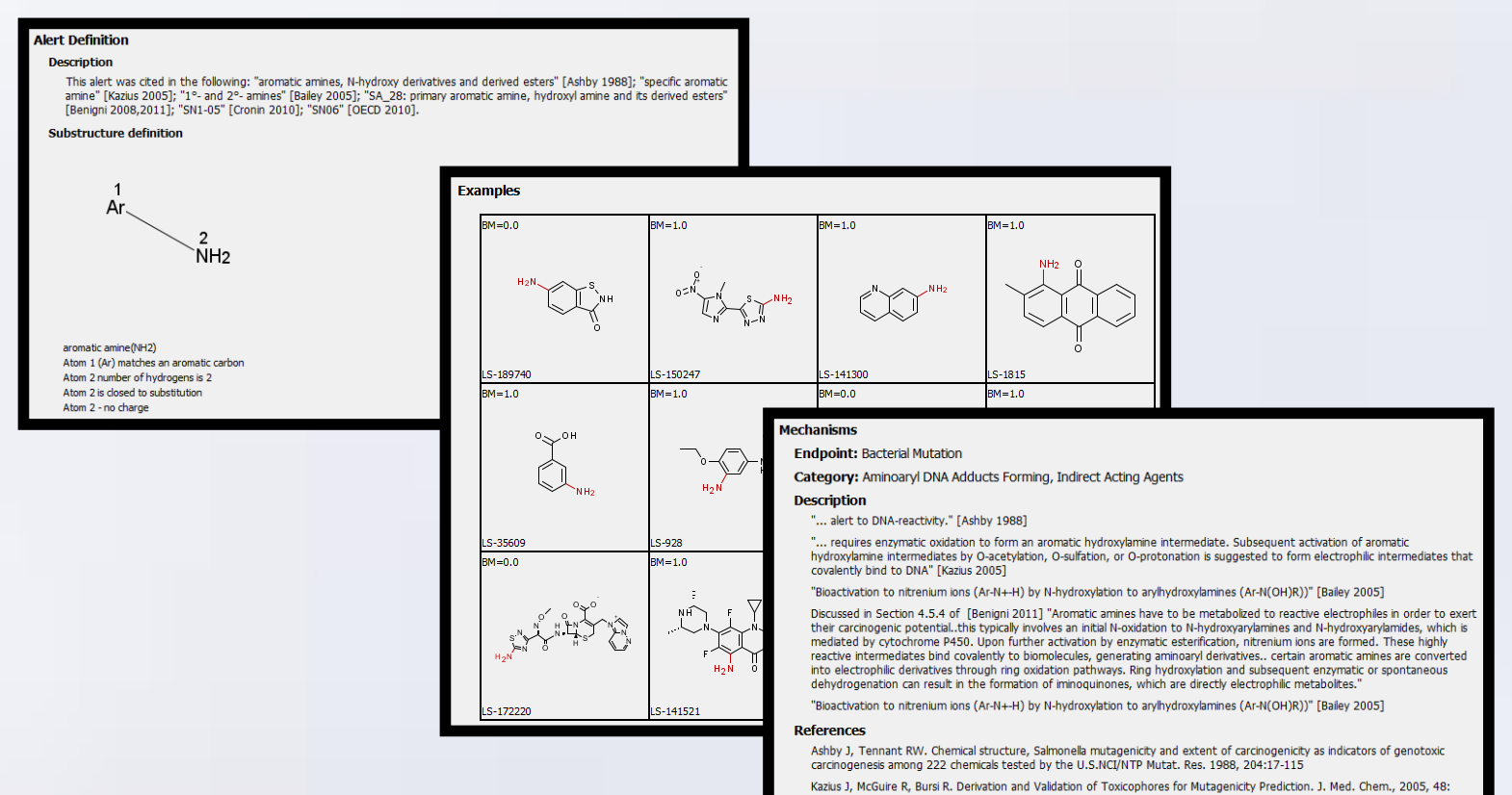
Any (Q)SAR model is limited by the knowledge from which it is built. Since today's commercial (Q)SAR models are primarily constructed from public information, several issues arise when these models are applied to proprietary chemical space.

First, prediction calls may not be computed for all test compounds where the proprietary space is not adequately reflected in the (Q)SAR models. Test compounds may be out of the domain of the model. In addition, some prediction results may be categorized as equivocal or indeterminate as a result of conflicting evidence.

Second, the performance of these models for specific classes may have low specificity, since the underlying structural reasons for deactivation may not be fully enumerated, especially using only public domain information. These performance issues directly affect submissions and drug safety, since any false positives result in unnecessary bacterial mutation testing and false negatives may compromise clinical safety.

Primary Aromatic Amines

Primary aromatic amines (PAAs) are commonly occurring functional groups found in many starting materials, industrial chemicals and drug impurities. This structural class also represents a mutagenicity structural alert reported in the literature[3-11]. The principal mechanism for primary aromatic amine mutagenicity involves metabolic activation of PAAs by the enzyme CYP1A2, resulting in a hydroxylamine and nitrenium ions that form covalent adducts with DNA. The alert is qualified by empirical data showing an association between compounds that contain a primary aromatic amine and positive mutagenicity findings. However, predictions based on *only* the presence of a primary aromatic amine, without consideration of other factors, will result in some false positive classifications.



Alert Definition

Description

This alert was cited in the following: "aromatic amines, N-hydroxy derivatives and derived esters" [Ashby 1988]; "specific aromatic amine" [Kazius 2005]; "1°- and 2°- amines" [Bailey 2005]; "SA_28: primary aromatic amine, hydroxyl amine and its derived esters" [Benigni 2008,2011]; "SN1-05" [Cronin 2010]; "SN06" [OECD 2010].

Substructure definition

aromatic amine(NH2)
Atom 1 (Ar) matches an aromatic carbon
Atom 2 number of hydrogens is 2
Atom 2 is closed to substitution
Atom 2 - no charge

Examples

BM=0.0 LS-189740; BM=1.0 LS-150247; BM=1.0 LS-141300; BM=1.0 LS-1815; BM=1.0 LS-35609; BM=1.0 LS-928; BM=0.0; BM=1.0; BM=0.0 LS-172220; BM=1.0 LS-141521

Mechanisms

Endpoint: Bacterial Mutation

Category: Aminoaryl DNA Adducts Forming, Indirect Acting Agents

Description

"... alert to DNA-reactivity." [Ashby 1988]

"... requires enzymatic oxidation to form an aromatic hydroxylamine intermediate. Subsequent activation of aromatic hydroxylamine intermediates by O-acetylation, O-sulfation, or O-protonation is suggested to form electrophilic intermediates that covalently bind to DNA" [Kazius 2005]

"Bioactivation to nitrenium ions (Ar-N+-H) by N-hydroxylation to arylhydroxylamines (Ar-N(OH)R))" [Bailey 2005]

Discussed in Section 4.5.4 of [Benigni 2011] "Aromatic amines have to be metabolized to reactive electrophiles in order to exert their carcinogenic potential..this typically involves an initial N-oxidation to N-hydroxyarylamines and N-hydroxyarylamides, which is mediated by cytochrome P450. Upon further activation by enzymatic esterification, nitrenium ions are formed. These highly reactive intermediates bind covalently to biomolecules, generating aminoaryl derivatives.. certain aromatic amines are converted into electrophilic derivatives through ring oxidation pathways. Ring hydroxylation and subsequent enzymatic or spontaneous dehydrogenation can result in the formation of iminoquinones, which are directly electrophilic metabolites."

"Bioactivation to nitrenium ions (Ar-N+-H) by N-hydroxylation to arylhydroxylamines (Ar-N(OH)R))" [Bailey 2005]

References

Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested by the U.S.NCI/NTP Mutat. Res. 1988, 204:17-115

Kazius J, McGuire R, Bursi R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 2005, 48: 312-320

Bailey AB, Chanderbhan N, Collazo-Braier N, Cheeseman MA, Twaroski ML. The use of structure-activity relationship analysis in the food contact notification program. Regulat Pharmacol Toxicol 2005; 42:225-235.

Benigni R & Bossa C, Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology, Chem Rev. 2011 Apr 13;111(4):2507-36

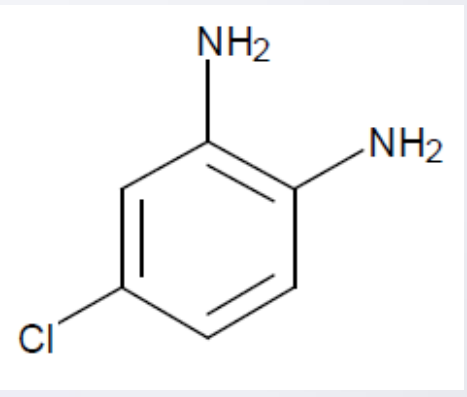
Expert Alert for Primary Aromatic Amine

The objective of this analysis is to better understand the structure-activity relationships among compounds containing PAAs to avoid inaccurate classifications for mutagenicity (false positives and false negatives) by identifying classes of PAAs representing the highest and lowest safety concern. For this exercise, two terms of art are defined: (1) active subclasses and (2) deactivating fragments. An active subclass represents a specific group of aromatic amines that are generally positive, whereas deactivating fragments represent classes of aromatic amines that are generally negative. These two classes will be identified and refined through an assessment of public and proprietary data from pharmaceutical sponsors. The analysis of proprietary data has been performed in a manner that is sensitive to intellectual property concerns of commercial organizations.

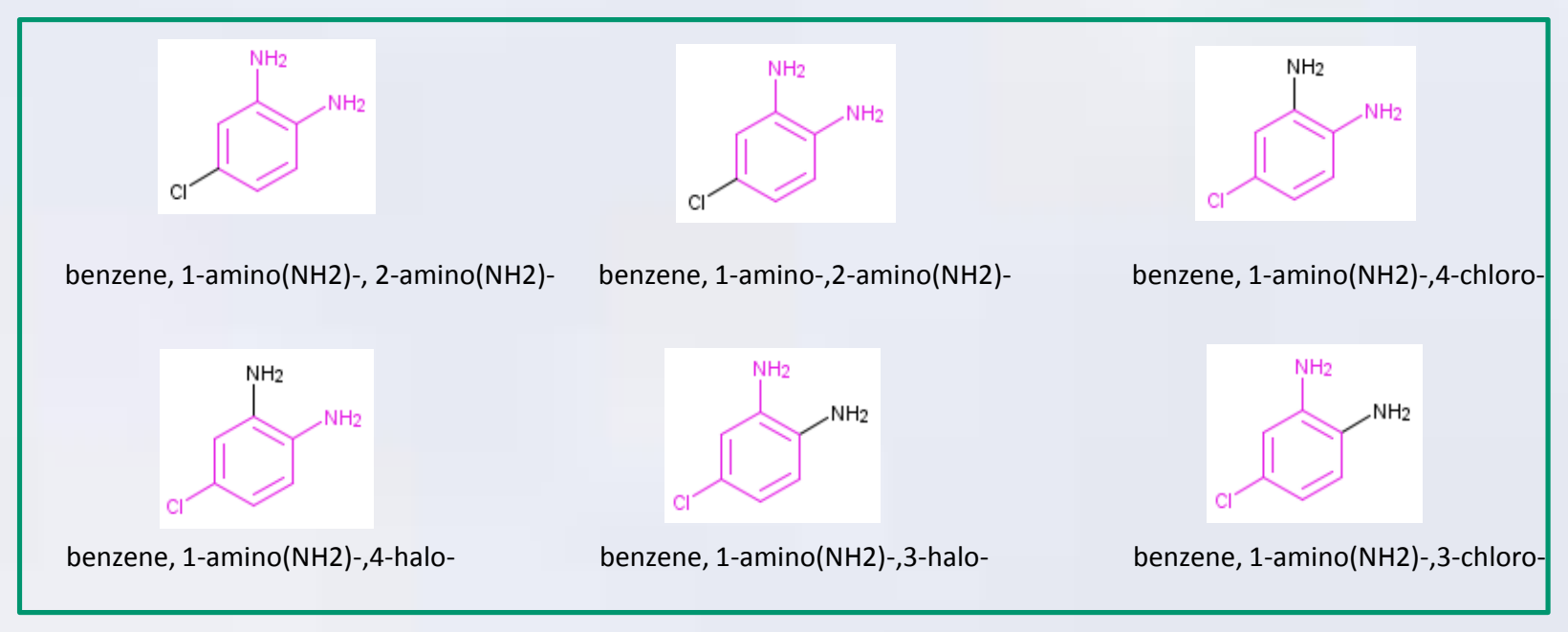
Two high quality databases, each comprised of data from numerous sources, containing primary aromatic amines were used: (1) The non-proprietary training sets used to build (Q)SAR models at the US FDA with Research Collaboration Agreement (RCA) partners[12,13], and (2) the Leadscope SAR 2014 database[14]. The RCA-QSAR database is comprised of 3,979 chemicals, with data on overall *Salmonella* strains as well as 1,198 chemicals with a composite TA102 / *E. coli* strain call. These datasets contain non-proprietary data harvested from FDA approval packages and the published literature. The Leadscope 2014 SAR Genetox Database is a collection of genetic toxicology studies (including bacterial mutagenesis, chromosome aberration, mammalian mutagenesis, and in vivo micronucleus) and contains 6,805 chemicals with graded bacterial mutagenesis calls. Information was collected from electronic sources or manually harvested to create the data set. The following sources were used: (1) The US Food and Drug Administration (FDA) Center for Food Safety and Applied Nutrition (CFSAN) Food Additive Regulatory Management system (FARM) and Priority-based Assessment of Food Additives (PAFA); (2) the US FDA's Center for Drug Evaluation and Research (CDER) Pharmacology Reviews based on the new drug approval (NDA) documents; (3) the Chemical Carcinogenicity Research Information System (CCRIS); (4) the National Toxicology Programs (NTP) genetic toxicology database; (5) the Tokyo-Eiken database; (6) and other publications.

Methodology – Using Chemical Fingerprints

A chemical fingerprint is derived using a set of pre-defined substructure search queries. Each substructure is named and applied to each chemical to determine whether it is present or not in the test chemical, as illustrated here:

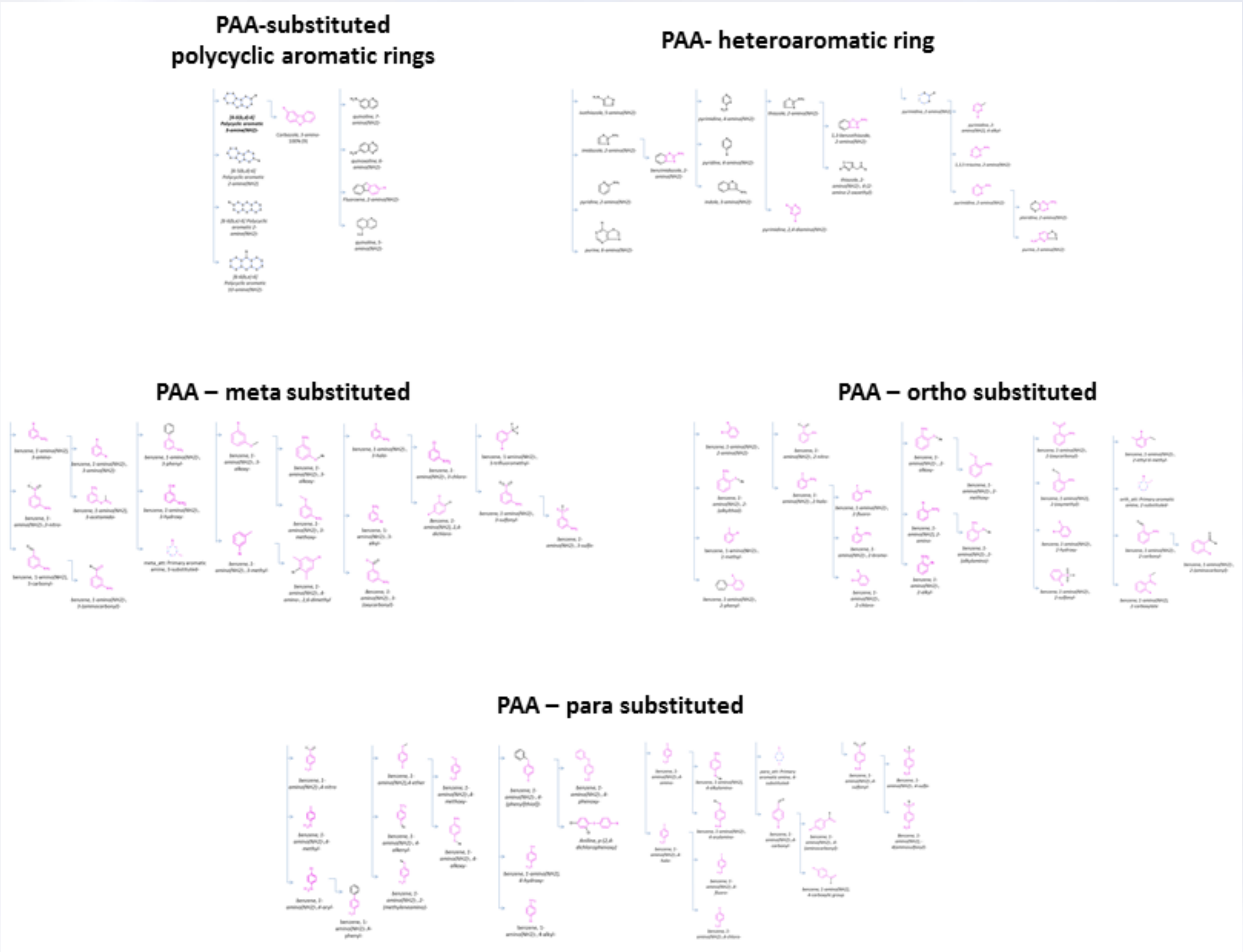


Substructures present
benzene, 1-amino(NH2)-, 2-amino(NH2)-
benzene, 1-amino-,2-amino(NH2)-
benzene, 1-amino(NH2)-,4-chloro-
benzene, 1-amino(NH2)-,4-halo-
benzene, 1-amino(NH2)-,3-halo-benzene
1-amino(NH2)-,3-chloro-


benzene, 1-amino(NH2)-, 2-amino(NH2)-
benzene, 1-amino-,2-amino(NH2)-
benzene, 1-amino(NH2)-,4-chloro-
benzene, 1-amino(NH2)-,4-halo-
benzene, 1-amino(NH2)-,3-halo-
benzene, 1-amino(NH2)-,3-chloro-

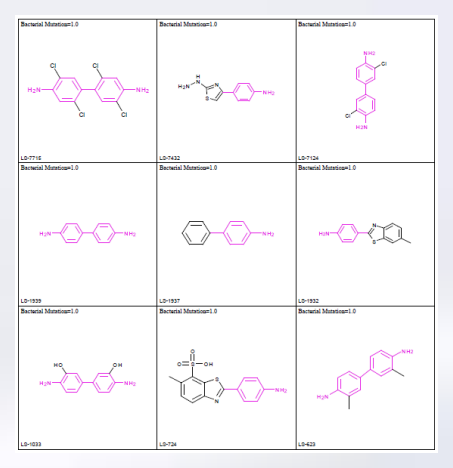
Substructure matching the target compound

A fingerprint containing a variety of primary aromatic amines has been developed to help understand the structure-activity relationships for this class. These structural features were based on both the Leadscope fragment hierarchy, data analysis of the reference set and external knowledge. The list of substructures includes meta-, para-, ortho-, hetero-substituted, polycyclic, as well as more complex substitution patterns. A list of 591 unique PAA substructures was used to define the fingerprint, as illustrated here.



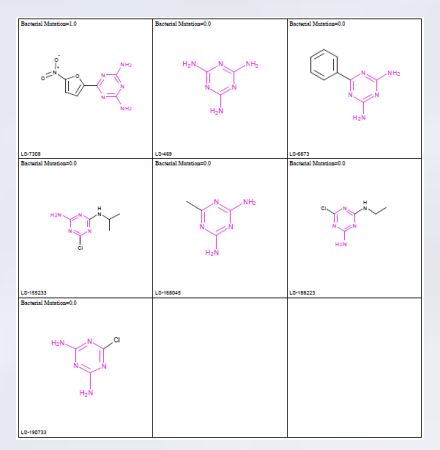
This fingerprint was applied to both the public data set as well as proprietary collections. The result is a listing of named substructures that are present along with the number of positive and negative bacterial mutagenicity examples (as illustrated here for the public data sources):

	Bacterial mutagenicity data	
Name	Number of positives	Number of negatives
benzene, 1-amino(NH2)-,4-aryl-	37	5



9 (out of 42) matching Database Structures

	Bacterial mutagenicity data	
Name	Number of positives	Number of negatives
1,3,5-triazine, 2-amino(NH2)-	1	6



7 matching Database Structures

It is possible to apply this fingerprint over a proprietary database without revealing confidential information about individual compounds or their data as it only summarizes results across collections of chemicals for the pre-defined substructures in the fingerprint.

Results

The fingerprint was applied to both the public data as well as a series of proprietary databases (without exchanging information on compounds or bacterial mutagenicity results). Through this analysis it was possible to identify both new active subclasses as well as new deactivating fragments. A number of these classes could not have been identified from public data alone. The following table illustrates the results for 4 out of the 591 substructures. The gray shaded boxes hide the individual values from the proprietary data sets.

Name	Positive (public data)	Negative (public data)	Positive (proprietary #1)	Negative (proprietary #1)	Positive (proprietary #2)	Negative (proprietary #2)	Positive (proprietary #3)	Negative (proprietary #3)	Positive (proprietary #4)	Negative (proprietary #4)	Total Positive	Total Negative
....												
Carbazole, 3-amino-	9	0									9	0
benzene, 1-amino(NH2)-,3-hydroxy-	5	3									14	3
....												
3,4-dihaloaniline	0	3									0	10
Aniline, 2-methyl-6-ethyl-	1	6									1	6
....												

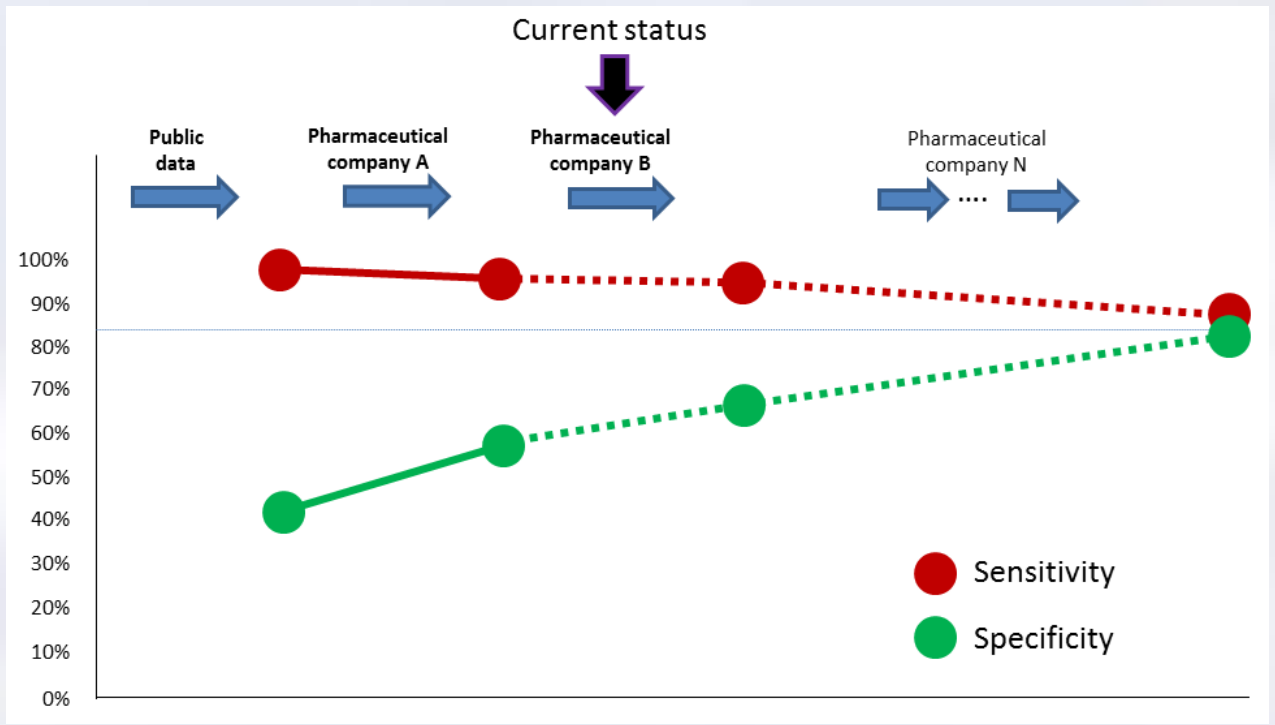
Confidential

As a proof of concept, active subclasses and deactivating fragments were identified from (1) public information only and (2) public plus proprietary information from a single pharmaceutical company. These classes were selected from the list using criteria that included a statistically significant association and a plausible mechanistic rationale. This SAR knowledge was subsequently incorporated into a rule-based expert alert system and compared against a test set containing 913 primary aromatic amines. The inclusion of the proprietary knowledge increased the prediction specificity by 14% since the number of false positives decreases from 203 to 159 as a result of the newly identified deactivating fragments. This is illustrated in the following table showing an increased number of positive and negative predictions made with confidence, while the number of less certain positive assignments (with no active subclass or deactivating fragment) decreases.

	Concordance	Sensitivity	Specificity	Positive (containing an active subclass)	Negative (containing a deactivating fragment)	Positive (no active subclass or deactivating fragment)
Using public only knowledge	75%	95%	40%	113	163	348
Using public and proprietary knowledge	78%	93%	54%	137	223	264

Discussion and Conclusion

Sharing data and knowledge with the developers of (Q)SAR models will improve the performance of these models as well as increase the number of compounds that can be predicted. It is possible to implement this transfer of knowledge while still protecting the intellectual property associated with individual compounds and data using a chemical substructure fingerprint. Improving the specificity of specific classes of compounds such as primary aromatic amine, alkyl halides or aryl boronic acids by reducing false positives means that sponsors will need to perform less laboratory testing while still protecting patient safety. This process is illustrated showing the expected outcome from combining multiple companies' fingerprint data for primary aromatic amines.



Acknowledgement

The authors wish to acknowledge and thank Ronald Snyder and Errol Zeiger for their advice on this project as well as Robert Jolly from Lilly Research Laboratories and Rosemary Hicks from GlaxoSmithKline.

References

[1] http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_Step_4.pdf

[2] OECD (2007). Guidance document on the validation of (Quantitative) structure activity relationships [(Q)SAR] models.(http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/background/oecd-principles)

[3] Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested by the U.S.NCI/NTP Mutat. Res. 1988, 204:17-115

[4] Bailey AB, Chanderbhan N, Collazo-Braier N, Cheeseman MA, Twaroski ML. The use of structure-activity relationship analysis in the food contact notification program. Regulat Pharmacol Toxicol 2005; 42:225-235.

[5] Benigni R, Bossa C, Jeliazkova N, Worth A. The Benigni / Bossa Rulebase for Mutagenicity and Carcinogenicity - A Module of Toxtree - EUR 23241 EN 2008

[6] Benigni R & Bossa C, Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology, Chem Rev. 2011 Apr 13;111(4):2507-36

[7] Enoch SJ & Cronin MTD, A review of electrophilic reaction chemistry involved in covalent DNA binding, Critical Reviews in Toxicology, 2010, 40, 728-748

[8] S. Enoch, M.T.D. Cronin Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding , Mutation Research 743 (2012) 10-19

[9] Shamovsky I et. al. *J Am Chem Soc*. 2011 Oct 12;133(40):16168-85

[10] Galloway S.M. et al. *Regulatory Toxicology and Pharmacology* 66 (2013) 326–335

[11] Kazius J, McGuire R, Bursi R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 2005, 48: 312-320

[12] L. Stavitskaya, B. L. Minnier, R. D. Benz, N. L. Kruhlak, *FDA Center for Drug Evaluation and Research, SOT 2013 "* Development of Improved Salmonella Mutagenicity QSAR Models Using Structural Fingerprints of Known Toxicophores"

[13] L. Stavitskaya, B. L. Minnier, R. D. Benz, N. L. Kruhlak, FDA Center for Drug Evaluation and Research, "Development of Improved QSAR Models for Predicting A-T Base Pair Mutations ",GTA 2013b poster

[14] Leadscope SAR Genetox Database 2014 User Guide