[Frontiers in Bioscience 19, 649-661, January 1, 2014]

In silico predictions of genotoxicity for aromatic amines

Jorg Bentzien, Ingo Muegge

Boehringer Ingelheim Pharmaceuticals, Department of Medicinal Chemistry, Structural Research, 900 Ridgebury Road, Ridgefield, CT 06877-0368, U.S.A.

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Structural alerts and expert system approaches
4. Regressive modeling approaches
5. Ab inito approaches
6. Discussions and perspectives
7. References

1. ABSTRACT

Insufficient drug safety is one of the major reasons for failure of drug candidates in Phase II and Phase III clinical trials. Determining toxicity early during the drug discovery process can help lower the attrition rate in clinical trials and lead to significant cost savings. *In silico* approaches can help to prioritize large numbers of compounds quickly and cost effectively in the early phase of drug discovery. One form of toxicity is genotoxicity due to mutagenicity. In this paper different *in silico* approaches for predicting mutagenicity, in particular in primary aromatic amines, are reviewed.

2. INTRODUCTION

Analyses of attrition rates in clinical trials show that drug safety is currently one of the major reasons for drug candidates to fail in late stage development. Drug safety has been reported to be the reason for 19% of drug candidates failing in Phase II clinical trials (1) and 21% of drug candidates failing in Phase III clinical trials (2). An analysis for the underlying causes of attrition in clinical trials for the period from 1991 to 2000 showed efficacy and safety/toxicology as the two major causes, each contributing about 30% in 2000 (3). The analysis also showed a major shift in the causes for attrition reported

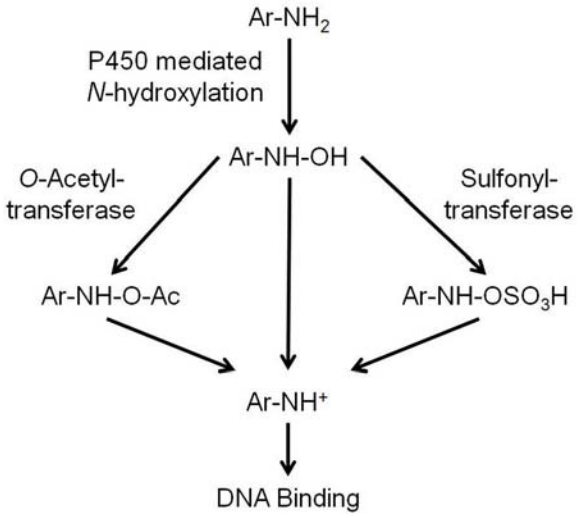


Figure 1. Mechanism by which aromatic amines elicit their mutagenic potential.

during this time period. Whereas pharmacokinetics/bioavailability was the reason for drug failure in 40% of the cases in 1991, in 2000 it only contributed in 10% of the cases to drug failure. This indicates that the pharmaceutical industry learned how to identify compounds with poor pharmacokinetic properties early. Better assays and better *in silico* models likely contributed to lowering the attrition rate due to poor pharmacokinetics/bioavailability. It can be hypothesized that a better understanding of toxicology will lead to more predictive assays and more predictive *in silico* models and that over time these would lead to lower attrition rates due to safety/toxicology. Even with the availability of cheaper and more predictive experimental assays predictive *in silico* tools will remain important in virtual high throughput screening and the prioritization of virtual compounds.

One form of toxicity is genotoxicity. Here we are particularly interested in genotoxins that have the potential to cause mutations in the cell's DNA that increase the risk of cancer. A class of compounds that is particularly prone to elicit mutagenicity is that of aromatic amines. Though the mechanism by which aromatic amines cause mutagenicity is well understood (4), predicting the mutagenicity for this class of compounds by *in silico* approaches is not easy. Experimentally mutagenicity can be assessed *in vitro* using the Ames test (5-7). A strong correlation between compounds that test positive in the Ames test and mutagenicity has been established (4). However, a negative Ames test result cannot exonerate a compound from potential mutagenicity. Therefore, the Ames test of compounds, or more often, that of their relevant metabolites, is being used to exclude compounds from further development. In other words, passing the Ames test is only a necessary but not a sufficient criterion to advance development compounds. Nevertheless, the Ames test is by far the most common mutagenicity test used and a requirement by regulatory agencies. Other experimental *in vitro* assays used to assess genotoxicity are the mouse lymphoma assay, chromosome aberration, and *in vitro* micronucleus tests.

The mechanism by which aromatic amines elicit their mutagenic potential is summarized in Figure 1. In a first step aromatic amines undergo cytochrome P450 mediated activation through *N*-oxidation to form *N*-hydroxylamines. These *N*-hydroxylamines can then form a nitrenium ion either i) directly via N-O bond cleavage, or ii) after going through the formation of an acetate or sulfate conjugate.

Typically, it takes between 50mg and 100mg of highly pure material to conduct the Ames test of a small molecule. To reduce cost and time associated with the synthesis of compounds and with conducting the Ames test and also to reduce the use of animals, as demanded for instance by the European REACH legislation (8), a large number of *in silico* approaches has been developed over the years to predict the mutagenic potential of chemical substances and in particular to predict Ames activity. These methods can be grouped into four main classes: i) structural alerts ii) expert system approaches, iii) regressive modeling approaches, and iv) *ab initio* approaches and are summarized in Table 1.

3. STRUCTURAL ALERTS AND EXPERT SYSTEM APPROACHES

Building on the electrophilic theory by Miller and Miller (9) several structural alerts have been developed to identify mutagenic chemicals (10-11). A deeper understanding of the mechanisms of carcinogenicity and mutagenicity has led to a refinement of structural alerts (12) and their incorporation into commercial (13-14) and public (15-16) expert systems.

Structural alerts can be used as an early, coarse grained, qualitative indicator for mutagenicity. A list of prominent structural alerts for aromatic amines is shown in Figure 2. When using structural alerts it is important to note that the absence of a structural alert does not necessarily indicate that the compound is safe. The success of structural alerts in predicting mutagenicity is evident by their incorporation into commercial expert systems routinely used in the pharmaceutical industry and by regulatory agencies. On a set of 564 marketed drugs with genotoxicity data Snyder reports accuracy values (rate of Ames positives and negatives correctly identified) of 86% for DEREK and 92.5% for MC4PC although the sensitivities (percentage of Ames positive compounds correctly identified) are comparatively low, ranging from 61.5% for DEREK to 44.7% for MC4PC (18). For the Toxtree structural alerts Benigni *et al.* report an overall accuracy of 79% (19). To put these numbers into perspective it is important to note that the inter-laboratory reproducibility of the Ames test is reported to be 80 – 85% (19). For the Toxtree structural alert for primary aromatic amines, aromatic hydroxylamines, esters, or amine generating groups, Benigni and Bossa report a positive predictive value, PPV, (ratio of actual Ames positive compounds and compounds predicted to be Ames positive) of 81% using the ISSCAN database of chemical

Table 1. Summary of the different approaches mentioned in this review

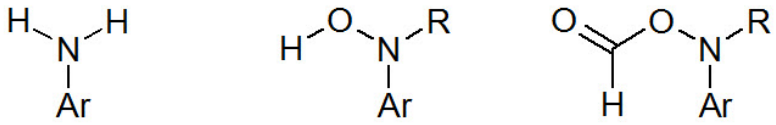
Category	Basic Technology	Webpage/Comment	References
Structural Alerts			
	Electrophilic theory	No commercial solution	9, 10, 11
	Structural Alerts		12
	Toxicophores		17
Expert Systems			
	DEREK Nexus	http://www.lhasalimited.org/products/derek-nexus.htm	13, 18, 20, 21
	MC4PC	http://www.multicase.com/products/prod01.htm	14, 18, 20, 21, 40
	Oncologic	http://www.epa.gov/oppt/sf/pubs/oncologic.htm	15
	Toxtree	http://toxtree.sourceforge.net/index.html	16, 12, 19, 20
	LeadScope	http://www.leadscope.com/	20
	TOPKAT	http://accelrys.com/products/discovery-studio/admet.html	21, 39, 62
	ToxCheck	Novartis in-house system	61
Regressive Modeling			
	QSAR (Hammett Equation)	No commercial solution. Multiple different approaches using different machine learning techniques and different descriptor engines.	22
	QSAR (logP, HOMO,LUMO)		24
	QSAR (HOMO, LUMO, Molecular Refractivity)		25, 26, 27
	AtomPairs, Topological Indices, Principal Component Analysis		28
	1D-, 2D-, 3D-Descriptors using Support Vector Machines, Neural Networks, Genetic Algorithms		29
	E-States		31
	Linear Multivariate Regression		31 - 33
	Partial Least Square		34
	Multiple QSAR Approaches		23, 35, 36, 37, 38, 43, 44, 45
	LAZAR	http://www.in-silico.de/ http://lazar.in-silico.ch/predict	41, 42
***Ab initio* Approaches**			
	Nitrenium Hypothesis (AM1)	No commercial solution. Approaches use quantum mechanical calculations with various setup and complexity.	46, 47
	Nitrenium Hypothesis (DFT/B3LYP/6-31+G(d)		48 - 53
	Nitrenium Hpothesis (AM1, HF, DFT)		54
	Nitrenium Hypothesis (DFT/B3LYP/6-31G*)		56, 57, 58
	Nitrenium Hypothesis (various DFT)		59, 60
Crowd Sourcing			
	Multiple QSAR Approaches	http://www.kaggle.com/	66

carcinogens (12). For aromatic mono and dialkylamines this number is 67% and for aromatic *N*-acylamines 76% (12). One could argue that not so much the recall of all Ames positive compounds (sensitivity) but rather the confidence in an Ames positive prediction (PPV) is more useful when predicting Ames test outcomes. Nevertheless, attempts have been made to also improve the sensitivity of predictions with varied success. Hillebrecht *et al.* recently compared the performance of DEREK, Toxtree, MC4PC, and Leadscope in predicting mutagenicity on three different datasets (20). These authors observed a satisfactory performance on public data with accuracies ranging from 66.4 to 75.4%, sensitivities between 65.2 to 85.2%, and specificities (percentage of Ames negative compounds correctly identified) between 53.1 to 82.9%. However, there was a significant deterioration in sensitivity when these methods were applied to the Roche data set. Though the accuracies (73.1 to 85.5%) and specificities (77.5 to 93.9%) for the Roche data were as good as or better than for the public data set, the sensitivities were only between 17.4 and 43.4%. The authors contribute the difference in performance to differences in the chemical composition of the data sets. Whereas the public data sets are richer in commonly known mutagenic structures, such as polycyclic aromatic and nitro compounds, this is not the case for the Roche data set. Sensitivities, correctly identifying Ames positives, have been reported as generally low for prominent software tools such as DEREK, TOPKAT, and MCASE (21). The related general question of the value in predicting Ames with high confidence will be picked up again below.

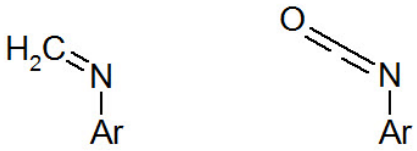
Kazius *et al.* constructed a data set of 4337 compounds with curated experimental Ames data (17). Their data set had 2401 compounds classified as mutagens and 1936 non-mutagens. From this data set the authors derived 29 toxicophores, structural alerts that are related to genotoxicity. For their data set a generic aromatic amine structural alert showed an accuracy of 79% when compounds were considered that had additional alerts besides the aromatic amine functionality and 67% when considering compounds that only had the primary aromatic amine alert. The accuracy could be significantly improved when a specific aromatic amine toxicophore was considered. This approach excludes structures where the aromatic ring carrying the amine functionality has a substituent consisting of a trifluoromethyl, sulfonamide, sulfonic acid, arylsulfonyl, or carboxylic acid or its ester derivates (17). Many of these substituents have strong electron withdrawing character, which is also reflected in their sigma-constants of the Hammett equation (22) and may not be able to help in the stabilization of the positive charge on the nitrenium ion.

Toxtree Structural Alerts (12):

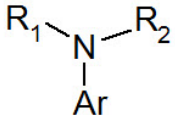
SA_28: Primary aromatic amine, hydroxyl amine and its derived esters



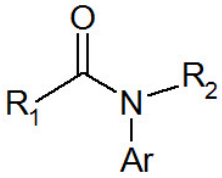
or amine generating group



SA_28bis: Aromatic mono- and dialkylamine

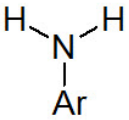


SA_28ter: Aromatic *N*-acylamine



Kazius Toxicophores (17):

Specific aromatic amine



a[NH2]

Excluded SMARTS patterns:

[NH2]a[$(a[$(C((F)F)F),$(S(=O)=O),$(C(=O)O)]),
$(aa[$(C((F)F)F),$(S(=O)=O),$(C(=O)O)]),
$(aaa[$(C((F)F)F),$(S(=O)=O),$(C(=O)O)])]

Figure 2. List of structural alerts for aromatic amines in Toxtree (12) and the Kazius' toxicophores (17).

As structural alerts provide only a coarse grained, qualitative indication for mutagenicity, often all primary aromatic amines are flagged as potential mutagens. This makes it difficult to get a more nuanced view for this class of molecules. To go beyond structural alerts Quantitative Structure Activity Relationship (QSAR) models have been developed to predict mutagenicity.

4. REGRESSIVE MODELING APPROACHES

The increased availability of bacterial mutagenicity assay data (Ames test) prompted QSAR studies to be conducted for more than 20 years. Predictive QSAR models have been reported for a series of chemical classes of carcinogens and mutagens including aromatic amines, nitroarenes, quinolines, triazenes, polycyclic aromatic hydrocarbons, lactones, and aldehydes (23). The interpretability of predictive models as present in all early QSAR models, such as pioneered by Corwin Hansch (22), has been emphasized for the prediction of mutagenicity as well. Early observations revealed that the mutagenicity of aromatic amines depends on factors such as the number and nature of aromatic rings, the position of the amine, the type and position of other ring substituents, and the size, shape, and polarity of the molecules (4). For instance, it was found that the force of conjugation increases from phenyl toward higher aryl groups (aniline is a weaker carcinogen than benzidine or naphthylamine). This finding prompted the use of a "number of rings" descriptor in early QSAR models. Also, due to steric effects, larger non-amino ring substituents, especially at the *ortho* position, decrease the mutagenicity of aromatic amines.

Early QSAR attempts with the goal of building interpretable models using limited numbers of molecules (<100) focused on logP, HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) energies, and ring substitution patterns. It was found that lipophilicity measured by logP, especially for compounds requiring metabolic activation (logP >1), appeared to be an important descriptor, accounting for up to 50% of Ames activity or mouse carcinogenicity (24). HOMO and LUMO energies were calculated with semi-empirical methods, mostly AM1. A positive correlation

between the HOMO energy and mutagenicity found in QSAR models could be rationalized by the easier activation of compounds with higher HOMO energies. An inverse correlation between mutagenicity and the LUMO energy on the other hand is less well understood.

One prominent QSAR model for the prediction of mutagenicity of aromatic amines based on the TA100 strain of *Salmonella typhimurium* (+S9) is also available from the Organization for the Economic Co-operation and Development (OECD) QSAR Toolbox (http://www.oecd.org/chemicalsafety/theoecdqsartoolbox.htm) and will therefore be outlined here in more detail. Benigni *et al.* derived a series of QSAR models for carcinogenicity and mutagenicity related to T98 and T100 strains of *Salmonella typhimurium* (16, 25). Because the performance on an internal test set was particularly strong for the TA100 model obtained by applying a Canonical Discriminant Analysis, we will focus here on this model as well. The discriminant derived by Benigni *et al* reads:

$$\omega = -2.85\ HOMO + 1.84\ LUMO + 0.70\ MR_2 + 0.69\ MR_3 + 1.90\ MR_6 + 3.36\ Idist \quad (1)$$

where HOMO and LUMO descriptors were calculated as reported previously by Debnath *et al.* (24). Molar Refractivity (MR) contributions of ring substituents in positions 2, 3, and 6 to the amino group were used. Idist is set to 1 for compounds with crowded substituents on positions 3'-, 4'-, and 5'- of 4-aminobiphenyl and 0 otherwise. The model was derived from a set of 47 nonmutagens (ω1) and 64 mutagens (ω2). The discriminant threshold was found to be 25.04. The mean calculated ω values for the nonmutagens and mutagens were 26.09 and 23.99, respectively. Applying an external test set of 22 nonmutagens and 25 mutagens yielded an accuracy of 81%, a sensitivity of 86% and a specificity of 72%. Recently, Benigni *et al.* reported attempts to enhance selectivity and specificity (26). Particularly noteworthy is the finding that hydrophobicity alone, as measured by logP, had no discriminating power for mutagens and nonmutagens in TA100 and TA98. In contrast, hydrophobicity was observed as an important factor in modulating potency for biologically actives as well as for describing carcinogenicity in mice.

In addition to predicting mutagenicity of aromatic amines as measured by *salmonella*, models for the prediction of mouse carcinogenicity have been reported. Recently, Franke *et al.* outlined a model for aromatic amines that predicts carcinogenicity in mice as defined by (27):

$$BRM = \log (MW/TD_{50}) + 3 \quad (2)$$

where MW is the molecular weight and TD_{50} is the dose producing tumors in 50% of the test animals in mg/kg body weight. Using Discriminant Analysis, the authors divided the data set into active carcinogens (BRM > 2.5) and inactive or weak carcinogens (BRM <2.5). A simple discrimination function was found to separate carcinogens from non-carcinogens:

$$\omega = -1.2\ logP + 0.96\ MR_2 - 1.17\ HOMO. \quad (3)$$

The model derived from 33 carcinogens and 41 non-carcinogens classified 86% of the compounds correctly in cross validation. The equation teaches that the risk of carcinogenicity increases with hydrophobicity, HOMO energy, and decreases with the size of *ortho* ring substitutions.

In addition to using few interpretable descriptors to build QSAR models, early authors started to build black box models using topological descriptors such as atom pairs and topological indices in combination with principal component analyses (28). More recently, Leong *et al.* reported a study of modeling TA98 +S9 mutagenicity of a combination of data sets of aromatic amines previously published, thereby attempting to increase the structural diversity among the data set (29). Leong *et al.* used a set of more than 200 1D, 2D, and 3D descriptors along with a variety of predictive modeling techniques including support vector machines, hierarchical support vector regression (HSVR), a neural network, and genetic function algorithm. The authors used variable selection techniques such as Genetic PLS and recursive feature elimination, an algorithm we have recently used in our own work as well (30), reducing the number of descriptors used in the most predictive HSVR approach to between 6 and 9 descriptors per model. It is particularly noteworthy, and speaks for the power of the variable selection algorithm, that a series of descriptors used in prominent mutagenicity models mentioned above are preferred by the automated variable selection method as well, among them the HOMO energy, logP, molar refractivity of *ortho* and *meta* position substituents of aromatic amines, and the presence or absence of three or more fused rings. Leong at al. demonstrated that both for the training set of 97 aromatic amines and the test set of 25 aromatic amines the HSVR approach performed better than a series of previously attempted QSAR models of aromatic amine mutagenicity including principal component analysis and E-state descriptors (31), Linear Multivariate Regression (31-33), Partial Least Square (PLS) (34), and others (35-38). Predicting mutagenicity for sets of noncongeneric samples of compounds that would be suitable for predicting the mutagenicity of any class of compounds has been incorporated in several popular commercial (TOPKAT (39), Multicase (40)) and freely available (LAZAR) systems (41, 42).

A large number of studies attempting to predict Ames activity of compounds including but not limited to aromatic amines has been published in recent years. A publicly available benchmark data set for *in silico* prediction of Ames mutagenicity comprising ~6500 compounds was introduced in 2009 (43). This data set has prompted an increased number of predictive modeling studies mostly deriving black box models of Ames activity (20, 44). Particularly noteworthy is an analysis by Sushko *et al.* that analyzed the influence of applicability domain (AD) and distance to model (DM) on the accuracy of mutagenicity predictions of aromatic amines (45). For a set of 30 different Ames models it was demonstrated that the

application of AD and DM concepts help significantly to increase the discrimination between reliable and unreliable predictions. The authors showed also that the uncertainties of *in silico* predictions are related to the variability of experimental measurements. The best Ames predictive *in silico* models analyzed by Sushko *et al.* (45) exhibited a performance that mirrored the inter-laboratory accuracy of the Ames test.

5. *AB INITIO* APPROACHES

Besides the QSAR approaches described above a series of *ab initio* approaches to predict the Ames activity of aromatic amines have been developed. The term *ab initio* in this context means that the mutagenic potential of a substance is assessed from first principals using the underlying mechanism that leads to the mutagenic behavior. The advantage of these types of methods is that, at least in principal, they do not need training on a set of compounds with associated mutagenicity reads, but rather calculate the mutagenic potential directly. Based on Scheme 1, Ford and Griffin used the semi-empirical AM1 method to calculate the energies for the different chemical entities (46-47). For a small set of food compounds the authors were able to develop a model that showed that the stability of the nitrenium ion of a primary aromatic amine correlates with the Ames activity. These authors showed that only the formation of the nitrenium ion, as described in Equation (4), varied significantly with the nature of the aryl moiety. Other steps such as the formation of hydroxylamine (Equation 5) and its hydrolysis (Equation 6) can be neglected.

$$ArNH_2 + PhN^+H \rightarrow ArN^+H + PhNH_2 \quad (4)$$

$$ArNH_2 + PhNHOH \rightarrow ArNHOH + PhNH_2 \quad (5)$$

$$ArNHOH + PhN^+H \rightarrow ArN^+H + PhNHOH \quad (6)$$

$$\Delta E = E_{ArNH^+} + E_{PhNH_2} - E_{ArNH_2} - E_{PhNH^+} \quad (7)$$

By calculating enthalpic energy differences relative to baseline aniline (Equation 7) Ford and Griffin were able to develop a predictive model for the Ames activity of primary aromatic amines showing a correlation between the Ames activity and the stability of the nitrenium ion.

Building on the basic idea that the stability of nitrenium ions can be used as a predictor of mutagenicity for aromatic amines, others have since developed more refined methods in particular by applying higher levels of *ab initio* quantum mechanical theory than the originally used semi-empirical AM1 function and by considering other intermediates like the acetyl sulfonyl esters. Using the Density Functional Theory (DFT) employing the B3LYP hybrid functional with the 6-31+G(d) basis set Borosky calculated N-O bond dissociation energies for the *N*-hydroxy, *N*-acetoxy and *N*-sulfate derivatives of aniline and the *N*-acetoxy esters for a set of 17 aromatic and heteroaromatic amines. The mutagenic activity was found to increase when a more negative charge developed at the exocyclic nitrogen of the nitrenium ion and when the nitrenium ion stability increased (48). The correlations between experimental mutagenicities and calculated nitrinium ion stabilities were generally better within a series of closely related compounds. A larger study of 43 aromatic and heteroaromatic amines confirmed these earlier findings (49). In a series of papers Borosky *et al.* showed that DFT calculated relevant reactive ion stabilities can be used to correlate the mutagenic potential of other compound classes like polyaromatic hydrocarbons (50-53).

Our previous work (54) validated the nitrenium hypothesis on a subset of 257 primary aromatic amines taken from the larger set assembled by Kazius *et al.* (17). We calculated the relative stability differences of the nitrenium ions for a series of Ames positive and Ames negative aromatic amines using different levels of quantum mechanical theory, AM1, *ab initio* Hartree Fock, and *ab initio* DFT/B3LYP. Already at the AM1 level of theory an overall accuracy of 85.6% was achieved. The performance was comparable to that achieved by DEREK, which showed an overall accuracy of 84.0%, and higher than that of MC4PC with 73.3%. For this data set MC4PC showed a much lower sensitivity of 66.0% compared to 91.3% for the nitrenium hypothesis. Using a consensus approach, combining the classification results from the nitrenium stability calculations with those from DEREK and the toxicophore approach by Kazius *et al.* (17), resulted in an improvement in accuracy to 95.8%, while maintaining a coverage of 74.3% of the data set. Figure 3 illustrates that a good separation between the Ames positive and Ames negative classes can be achieved by calculating the ΔΔE value from Equation 7 at the AM1 level of theory.

A detailed analysis of the structural diversity within the data set of 257 primary aromatic amines showed that the nitrenium hypothesis performed very well for aromatic amines with fused aromatic ring systems. All 15 compounds in the data set containing four or five fused ring systems were correctly predicted to be Ames positive and out of the 80 compounds with three fused rings only seven were incorrectly predicted, three false negatives and four false positives. The observation that four of the incorrectly classified compounds were anthraquinones could point to the fact that for this class issues other than the stability of the nitrenium ion lead to the observed Ames activity.

One of the appealing characteristics of the nitrenium hypothesis is that the ΔΔE value (Equation 7) provides a continuous metric, not just a classification, which allows for identifying SAR trends. In the series of *p*-halogenated anilines, the mutagenic potential is predicted to increase from Ames negative *p*-bromoaniline, ΔΔE = +4.2 kcal/mol, to the Ames positive *p*-fluoroaniline, ΔΔE = -0.6kcal/mol with the *p*-chloroaniline having a value in between, ΔΔE = +0.6 kcal/mol. These trends are not captured by DEREK, MC4PC, or the toxicophore model of Kazius *et al.* (17). The ΔΔE value correlates with the ability of the substituent to stabilize the positive charge of the nitrenium ion. Thus, *p*-*N,N*-dimethylaniline, *N*-(*p*-aminophenyl)-acetamide, *p*-methoxy-phenylamine, and *p*-phenylaniline are all predicted to be Ames positive, whereas *p*-aminobenzamide, *p*-aminobenzoic acid, *p*-aminobenzenesulfonamide, and *p*-

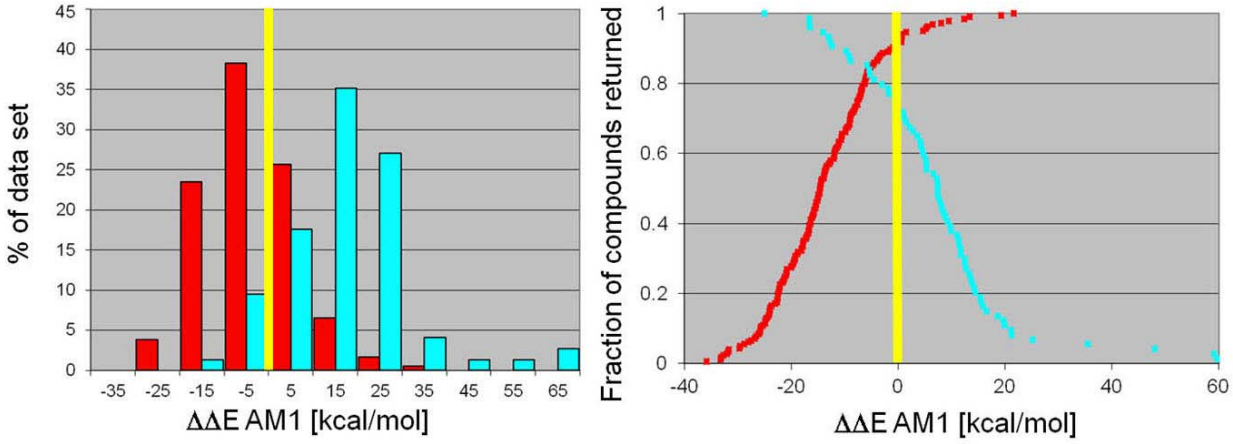


Figure 3. Performance of the nitrenium hypothesis on a dataset of 257 primary aromatic amines (54). Left: Distribution of AM1 calculated $\Delta\Delta E$ values for Ames positive (red bars) and Ames negative compounds (light blue bars). The yellow line marks the border between predicted Ames positive compounds (negative $\Delta\Delta E$ values) and predicted Ames negative compounds (positive $\Delta\Delta E$ values). Right**:** Fraction of returned Ames positive (red) and Ames negative (light blue) compounds at a given AM1 $\Delta\Delta E$ value. At $\Delta\Delta E$ = 0.0 kcal/mol 91% of all Ames positive compounds have a lower $\Delta\Delta E$ and 73% of all Ames negative compounds have a higher $\Delta\Delta E$. Both graphs demonstrate that the nitrenium stability energy separates the Ames positive from the Ames negative compounds.

aminobenzenesulfonicacid, which would not stabilize the positive charge of the nitrenium ion, are predicted to be Ames negative (Table 2).

In a detailed study on designing safe 4-aminobiphenyls Birch *et al.* (56) discuss the substituent effects for a series of anilines on the stability of nitrenium ions for anilines using the Hammett parameter sigma (22). These authors show that a phenyl ring with a σ_p value of -0.01 stabilizes the positive charge of the nitrenium ion, whereas a fluorine atom in the *meta* position, with σ_m = 0.34, is strongly electron withdrawing and thus has a destabilizing effect on the positive charge of the nitrenium ion. This finding correlates with their observation that many *meta* fluorinated 4-aminobiphenyls are Ames negative (56). We showed in our own work (54) that introducing a methyl substituent on the aniline has a stabilizing effect on the nitrenium ion, leading to an increased potential to show activity in the Ames assay. Thus, 2-methylaniline, 2,5-dimethylaniline, 3,4-dimethylaniline, and 2,4,6-trimethylaniline are all predicted to be Ames positive in agreement with the experimental results. Interestingly, DEREK predicts only the two dimethylanilines to be Ames positive. In contrast to methylation, chlorination on the aniline has a destabilizing effect. Thus 4-chloroaniline, 2-chloroanline, 3,5-dichloroaniline, 2,5-dichloroaniline, 3,4-dichloroaniline, 3,4,5-trichloroaniline, 2,4,6-trichloroaniline, and 2,3,4-trichloroanline are all predicted by the nitrenium hypothesis to be Ames negative, which with the exception for 4-choloroaniline is in agreement with the experiment (54). These trends are missed by DEREK and the Kazius toxicophore. The opposing effects of methylation and chlorination on the Ames activity of anilines correlate with their different σ values. The methyl group has σ_m and σ_p values of -0.07 and -0.17, respectively, whereas a chlorine exhibits a σ_m = +0.37 and σ_p = +0.23, in agreement with finding the 3,4-dimethylaniline to be Ames positive, and the 3,4-dichloroaniline to be Ames negative (54). We also showed that the substituent effects on the nitrenium ion stabilities are likely additive (54). For the 2-methyl-4-chloroaniline we reported an AM1 calculated ΔΔE = -3.7 kcal/mol, which is nearly the sum of the calculated ΔΔE values for 2-methylaniline, ΔΔE = -4.6 kcal/mol, and for 4-chloroaniline, ΔΔE = 0.6 kcal/mol.

In a variant of the nitrenium hypothesis Leach *et al.* (57) showed that a strong discrimination between Ames positive and Ames negative compounds can be achieved by calculating energy differences according to the following two equations:

$$ArNHOH + H_3O^+ \rightarrow ArNH^+ + 2H_2O \quad (8)$$

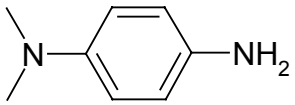
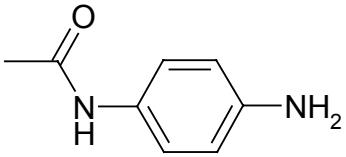
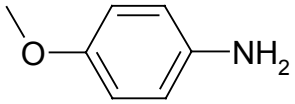
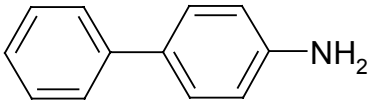
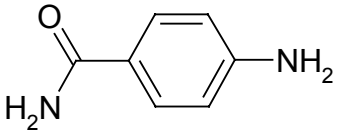
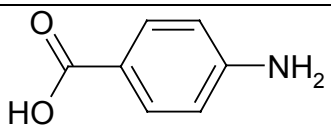
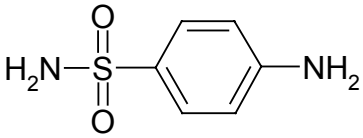
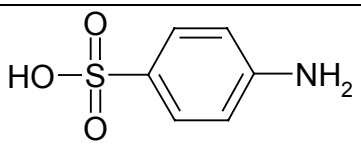
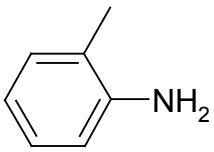
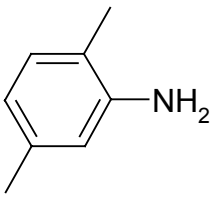
$$ArNHOAc \rightarrow ArNH^+ + AcO^- \quad (9)$$

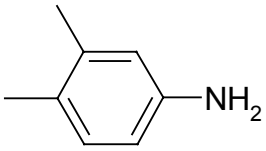
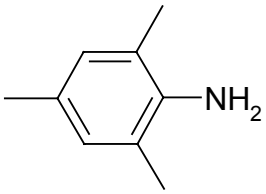
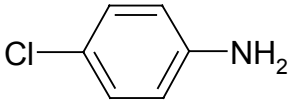
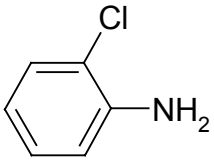
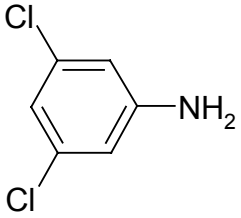
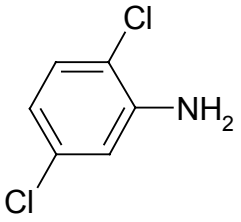
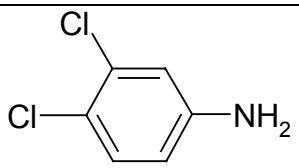
The discrimination is greater than that achieved by other descriptor based models used by these authors including, lipophilicity, and HOMO/LUMO energies. Instead of using the energy differences derived from Equations 8 and 9 directly, in subsequent publications this group used the probability to be Ames positive derived from Equation 10 (56, 58).

$$P(active) = \frac{100}{1 + e^{(0.056\Delta E - 2.012)}} \quad (10)$$

For the classes of 4-aminobiphenyls (56) and aminopyrazoles (58) the authors showed an excellent agreement between the *in silico* predicted property to be Ames positive and the experimental Ames findings. The nitrenium hypothesis is likely best suited to predict the mutagenicity for a series of congeneric compounds as pointed out above when we discussed Borosky's work (49).

Table 2. ΔΔE values for primary aromatic amines calculated at the AM1 level of theory using Equation 7

Compound Name	**Compound Structure**	**ΔΔE kcal/mol AM1**	**Predicted Ames Classification**	**Experimental Classification[1]**	**Hammett σ-value[2]**
p-N,N-dimethylaniline		-33.3	positive	positive	σ_p -0.83
N-(p-amino phenyl)-acet amide		-21.4	positive	positive	σ_p 0.00
p-methoxy-phenylamine		-15.8	positive	positive	σ_p -0.27
p-phenylaniline		-13.9	positive	positive	σ_p -0.01
p-aminobenz amide		+8.1	negative	negative	σ_p 0.36
p-aminobenzoic acid		+13.2	negative	negative	σ_p 0.45
p-amino benzene sulfonamide		+16.5	negative	negative	σ_p 0.57
p-amino benzenesulfonicacid		+25.1	negative	negative	σ_p 0.09 (for SO_3^-)
2-methylaniline		-4.6	positive	positive	σ_o -0.17
2,5-dimethyl aniline		-6.2	positive	positive	σ_o -0.17 σ_m -0.07

3,4-dimethyl aniline		-7.0	positive	positive	σ_p -0.17 σ_m -0.07
2,4,6-trimethyl aniline		-14.4	positive	positive	σ_o -0.17 σ_p -0.17
4-chloroaniline		+0.6	negative	positive	σ_p 0.23
2-chloroanline		+1.4	negative	negative	σ_0 0.20
3,5-dichloro aniline		+11.2	negative	negative	σ_p 0.37
2,5-dichloro aniline		+6.5	negative	negative	σ_o 0.20 σ_p 0.37
3,4-dichloro aniline		+5.1	negative	negative	σ_m 0.23 σ_p 0.37

The usefulness of the nitrenium hypothesis to predict the Ames activity was demonstrated by yet another group at Novartis (59-60). McCarren *et al.* found the same two equations, Equation 8 and 9, to give the best discrimination between Ames positive and Ames negative compounds. An analysis of the influence of the molecular weight on the discriminative power of the nitrenium ion formation energy showed that the separation between Ames positive and Ames negative compounds is larger for lower molecular weights, MW < 250, than for compounds with higher molecular weights, MW >500. Depending on the data set used for validation, these authors showed that overall accuracies between 70% and 88% could be achieved for aromatic amines with low molecular weight, MW <250, and 55% to 85% for those with molecular weights between 250 and 500 (60). McCarren *et al.* (60) also compared the performance of the nitrenium hypothesis to other methods such as ToxCheck (61), a modified version of DEREK, TOPKAT (62), and Toxtree (16, 25). For the two larger test sets analyzed, ToxCheck, TOPKAT, and Toxtree achieve accuracies between 73% and 87% for low molecular weight aromatic amines, MW < 250, which is similar to what is achieved with the nitrenium hypothesis. ToxCheck and TOPKAT have a much higher sensitivity, between 94% and 98% compared to 78% to 84% for the nitrenium hypothesis. In another paper

McCarren *et al.* identified that besides the nitrenium ion formation energy, the AM1 calculated HOMO energy shows a good separation between Ames positive and Ames negative compounds (59). The nitrenium natural charge and the anion formation energy lead to much lower separation between the Ames positive and Ames negative classes. In this paper the authors also compare the performance of the nitrenium ion formation to PLS and Random Forest models. They conclude that the best-validated and best-performing prediction available for low molecular weight aryl amines is the quantum-mechanics calculated reaction energy representing the formation of the nitrenium ion (59).

A potential difficulty with *ab initio in silico* predictions of the nitrenium ion stability is that they depend on the correct identification of the minimum energy conformation for each compound. To verify that the calculated geometries represent a minimum energy conformation McCarren *et al.* calculated the vibrational frequencies for the compounds in one of their example data sets (60). These calculations require significant additional computer time, and they do not guarantee that the geometry reflects the global minimum energy conformation for a given compound. For instance, in our own work we observed a ring open conformation for the nitrenium ion of 2-aminothiazole and bond breaking conformations for other sulfur containing molecules (54). Tautomers can also pose a challenge, and would need to be considered separately. Differences of several kcal/mol in the $\Delta\Delta E$ value have been observed for different tautomers. Though we did not find a case in our data set where these energy differences between tautomers of a molecule resulted in a different overall Ames classification, such instances cannot be ruled out (54).

6. DISCUSSIONS AND PERSPECTIVES

For the pharmaceutical industry it is not only important to correctly assess the mutagenic potential of the final compound but also that of pharmaceutical impurities (63). A recent survey of eight pharmaceutical companies indicated that a structural assessment is sufficient to conclude that an impurity is non-mutagenic (64). To be confident in such assessment based on computational analyses alone, a high negative predictive value (NPV), i.e. a high confidence that a compound that is predicted to be Ames negative will also be non-mutagenic experimentally, is of particular importance. The survey by Dobo *et al.* showed a NPV of 94% for the use of *in silico* tools alone (64). This number increased to 99% when the *in silico* result was coupled with expert evaluation. Interestingly, the NPV value did not change significantly depending on whether a company used a single *in silico* tool or multiple tools in combination (64). A very recent survey among five pharmaceutical companies further showed the importance of coupling an *in silico* tool with additional expert knowledge to reach a high NPV, as it is recommended that impurities not showing a structural alert will not need to be tested for mutagenicity (65).

While NPV is a most relevant measure for compounds that will not be tested as a result of an *in silico* analysis, there are other uses of mutagenicity prediction tools where the confidence in predicting that a compound is Ames positive (PPV) might be more important. A NPV or PPV maximizing model should be chosen depending on whether an advancement or an elimination decision will be made for compounds. Let's consider how the 'do no harm' principle, that is often used to assess to what degree *in silico* predictions are applicable, would apply to Ames predictions. In case a decision is to be made on advancing a single compound or making a choice among a small number of compounds in a later stage of drug discovery research, predicting Ames negativity well (high NPV) would be of little interest because the advancing compound would be tested in Ames anyway for regulatory reasons alone. On the other hand, if in such case a compound is predicted Ames positive, the confidence in such prediction (PPV) must be very high to allow the user to make the decision to eliminate the compound from further advancement. A false positive prediction would do the most harm in such case because it would potentially exclude a 'good' compound from further consideration. In another example, where a user prioritizes among a large number of compounds in early drug discovery, let's say an HTS hit set, a high PPV is of less use. In this case it is more important to have a high NPV because here decisions about advancing the best compounds are being made and one would like to be as confident as possible in the Ames negative prediction of advancing compounds. A possible harm could come later in form of positive Ames tests of compounds that have consumed many resources in the following development phase but have to be discarded later due to positive Ames test outcomes. An interesting discussion on the related topic of sensitivity and specificity of Ames predictions was provided by Snyder and Smith (21). The authors lament the lack of sensitivity (recall of Ames positives) observed from using software tools such as DEREK, MCASE, and TOPKAT, especially among compounds without structural alerts. They concluded that generally improvements are necessary for both predictive values, sensitivity and selectivity, to make these *in silico* tools more useful for Ames predictions, especially for compounds without structural alerts and those where the Ames activity is due to other mechanisms, such as non-covalent DNA interactions.

In this review we showed that for the class of primary aromatic amines a mechanism based *ab initio* approach, the nitrenium hypothesis, can be used to successfully classify Ames positive and Ames negative compounds. The fact that major pharmaceutical companies have published on this approach (54, 56-60) is evidence that this technology has come a long way since its early publications by Ford and Griffin (46-47). The appealing characteristic of this *first principal* approach is that, unlike QSAR methods, it does not rely on the identification of a training set. The successful application also demonstrates that a deeper understanding of the underlying chemical mechanism has been reached. Development of more mechanism based, *ab initio* approaches as alternatives to QSAR approaches, for predicting the mutagenic potential of other compound classes may be desirable.

In terms of QSAR modeling there is an emerging new trend towards using crowd computing. We recently reported (66) on the successful use of this approach with the Kaggle platform for modeling Ames activities using the literature set by Hansen *et al.* (43) as a test case. The top ranked models in the competition had improved performance over the benchmarks when applying a logloss metric. A detailed analysis of the model performances with different metrics showed that the ranking of the models is metric dependent and that differences are overall small. The large response and short time frame in which models were developed indicate that competitive crowd sourced approaches can be an alternative way towards modeling. Merck has since used the Kaggle platform for predicting molecular activities of chemical compounds (67-68). Another example of the successful application of prize-based contests for gene sequence annotation has been run on the TopCoder.com platform (69). It will be interesting to see how these innovative, dynamic, crowd-sourced approaches will impact modeling in the future.

7. REFERENCES

1. J. Arrowsmith: Trial watch: Phase II failures: 2008-2010. *Nat Rev Drug Discov* 10, 328-329 (2011)

2. J. Arrowsmith: Trial watch: phase III and submission failures: 2007-2010. *Nat Rev Drug Discov* 10, 87 (2011)

3. I. Kola, J. Landis: Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3, 711-715 (2004)

4. R. Benigni, A. Giuliani, R. Franke, A. Gruska: Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines. *Chem Rev* 100, 3697-3714 (2000)

5. B.N. Ames, F.D. Lee, W.E. Durston: An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc Natl Acad Sci USA* 70, 782-786 (1973)

6. J. McCann, E. Choi, E. Yamasaki, B.N. Ames: Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay of 300 chemicals. *Proc Natl Acad Sci USA* 72, 5135-5139 (1975)

7. J. McCann, B.N. Ames: Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay of 300 chemicals: discussion. *Proc Natl Acad Sci USA* 73, 950-954 (1976)

8. G. Schoeters: The REACH perspective: toward a new concept of toxicity testing. *J Toxicol Environ Health B Crit Rev* 13, 232-241 (2010)

9. J.A. Miller, E.C. Miller: Ultimate chemical carcinogens as reactive mutagenic electrophiles. In: Origins of Human Cancer. Eds: H.H. Hiatt, J.D. Watson, J.A. Winsten Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 605-627 (1977)

10. J. Ashby: Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environ Mutagen* 7, 919-921 (1985)

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

12. R. Benigni, C. Bossa: Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chem Rev* 111, 2507-2536 (2011)

13. DEREK Nexus 3.0.1 Lhasa Ltd.Leeds, UK (2012)

14. MC4PC 2.0.0.95 MultiCASE Inc. Beachwood, OH (2008)

15. Y.T. Woo, D.Y. Lai: Oncologic: A mechanism based expert system for predicting the carcinogenic potential of chemicals. In: Predictive Toxicology. Ed: C. Helma, Taylor and Francis, Boca Raton, FL 385-418 (2005)

16. Toxtree, 1.60 IdeaConsult, Ltd. Sofia, Bulgaria (2009)

17. J. Kazius, R. McGuire, R. Bursi: Derivation and validation of toxicophores for mutagenicity prediction. *J Med Chem* 48, 312-320 (2005)

18. R.D. Snyder: An update on the genotoxicity and carcinogenicity of marketed pharmaceuticals with reference to *in silico* predictivity. *Environ Mol Mutagen* 50, 435-450 (2009)

19. R. Benigni, C.L. Battistelli, C. Bossa, M. Colafranceschi, O. Tcheremenskaia: Mutagenicity, carcinogenicity, and other end points. *Methods Mol Biol* 930, 67-98 (2013)

20. A. Hillebrecht, W. Muster, A. Brigo, M. Kansy, T. Weiser, T. Singer: Comparative evaluation of *in silico* systems for Ames test mutagenicity prediction: scope and limitations. *Chem Res Toxicol* 24, 843-854 (2011)

21. R.D. Snyder, M.D. Smith: Computational prediction of genotoxicity: room for improvement. *Drug Discov Today* 10, 1119-1124 (2005)

22. C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani, E.J, Lien: Aromatic substituent constants for structure-activity correlations. *J Med Chem* 16, 1207-1216 (1973)

23. R. Benigni: Structure-activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches. *Chem Rev* 105, 1767-1800 (2005)

24. A.K. Debnath, G. Debnath, A.J. Shusterman, C. Hansch: A QSAR investigation of the role of

hydrophobicity in regulating mutagenicity in the Ames test: 1. Mutagenicity of aromatic and heteroaromatic amines in *Salmonella typhimurium* TA98 and TA100. *Environ Mol Mutagen* 19, 37-52 (1992)

25. R. Benigni, C. Bossa, T. Netzeva, A. Rodomonte, I. Tsakovska: Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environ Mol Mutagen* 48, 754-771 (2007)

26. R. Benigni, C. Bossa: Flexible use of QSAR models in predictive toxicology: a case study on aromatic amines. *Environ Mol Mutagen* 53, 62-69 (2012)

27. R. Franke, A. Gruska, C. Bossa, R. Benigni: QSARs of aromatic amines: identification of potent carcinogens. *Mutat Res* 691, 27-40 (2010)

28. S.C. Basak, G.D. Grunwald: Predicting mutagenicity of chemicals using topological and quantum chemical parameters: a similarity based study. *Chemosphere* 31, 2529-2546 (1995)

29. M.K. Leong, S.W. Lin, H.B. Chen, F.Y. Tsai: Predicting mutagenicity of aromatic amines by various machine learning approaches. *Toxicol Sci* 116, 498-513 (2010)

30. O. Demir-Kavuk, J. Bentzien, I. Muegge, E.W. Knapp: DemQSAR: predicting human volume of distribution and clearance of drugs. *J Comput Aided Mol Des 25*, 1121-1133 (2011)

31. G.G. Cash: Prediction of the genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. *Mutat Res* 491, 31-37 (2001)

32. U. Maran, M. Karelson, A.R. Katritzky: A comprehensive QSAR treatment of genotoxicity of heteroaromatic and aromatic amines. *Quant Struct -Activ Relat* 18, 3-10 (1999)

33. P. Gramatica, V. Consonni, M. Pavan: Prediction of aromatic amines mutagenicity from theoretical molecular descriptors. *SAR QSAR Environ Res* 14, 237-250 (2003)

34. S.C. Basak, D.R. Mills, A.T. Balaban, B.D. Gute: Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: a hierarchical QSAR approach. *J. Chem. Inf. Comput. Sci.* 41, 671-678 (2001)

35. M. Vracko: QSAR approch in study of mutagenicity of aromatic and heteroaromatic amines. *Top Heterocyclic Chemistry* 4, 85-106 (2006)

36. M. Casalegno, E. Benfenati, G. Sello: Application of a fragment-based model to the prediction of the genotoxicity of aromatic amines. *Internet Electron J Mol Des* 5, 431-446 (2006)

37. A.A. Toropov, A.P. Toropova: Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atoomic orbital graphs of local invariants. *Theochem-J Mol Struct* 538, 287-293 (2013)

38. A.A. Toropov, A.P. Toropova, E. Benfenati: QSAR modelling for mutagenic potency of heteroaromatic amines by optimal SMILES-based descriptors. *Chem Biol Drug Des* 73, 301-312 (2009)

39. K. Enslein, V.K. Gombar, B.W. Blake: International Commission for Protection Against Environmental Mutagens and Carcinogens. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. *Mutat Res 305*, 47-61 (1994)

40. G. Klopman, H.S. Rosenkranz: International Commission for Protection Against Environmental Mutagens and Carcinogens. Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE. *Mutat Res* 305, 33-46, (1994)

41. C. Helma: Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and *Salmonella* mutagenicity. *Mol Divers* 10, 147-158 (2006)

42. A. Maunz, M. Gütlein, M. Rautenberg, D. Vorgrimmler, D. Gebele, C. Helma: Lazar: A modular predictive toxicology framework. *Frontiers in Pharmacology* 4, 38 (2013)

43. K. Hansen, S. Mika, T. Schroeter, A. Sutter, A. ter Laak, T. Steger-Hartmann, N. Heinrich, K.R. Müller, Benchmark data set for *in silico* prediction of Ames mutagenicity. *J Chem Inf Model* 49, 2077-2081 (2009)

44. E.A. Helgee, L. Carlsson, S. Boyer: A method for automated molecular optimization applied to Ames mutagenicity data. *J Chem Inf Model* 49, 2559-2563 (2009)

45. I. Sushko, S. Novotarskyi, R. Korner, A.K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K.R. Müller, L. Xi, H. Liu, X. Yao, T. Oberg, F. Hormozdiari, P. Dao, C. Sahinalp, R. Todeschini, P. Polishchuk, A. Artemenko, V. Kuz'min, T.M. Martin, D.M. Young, D. Fourches, E. Muratov, A. Tropsha, I. Baskin, D. Horvath, G. Marcou, C. Muller, A. Varnek, V.V. Prokopenko, I. V. Tetko: Applicability domains for classification problems: Benchmarking of distance to models for Ames mutagenicity set. *J Chem Inf Model* 50, 2094-2111 (2010)

46. G.P. Ford, G.R. Griffin: Relative stabilities of nitrenium ions derived from heterocyclic amine food carcinogens: relationship to mutagenicity. *Chem Biol Interact* 81, 19-33 (1992)

47. G.P. Ford, P.S. Herman: Relative stabilities of nitrenium ions derived from polycyclic aromatic amines. Relationship to mutagenicity. *Chem Biol Interact.* 81, 1-18 (1992)

48. G.L. Borosky: Ultimate carcinogenic metabolites from aromatic and heterocyclic aromatic amines: a computational study in relation to their mutagenic potency. *Chem Res Toxicol* 20, 171-180 (2007)

49. G.L. Borosky: Carcinogenic carbocyclic and heterocyclic aromatic amines: a DFT study concerning their mutagenic potency. *J Mol Graph Model* 27, 459-465 (2008)

50. G.L. Borosky, K.K. Laali: Theoretical study of aza-polycyclic aromatic hydrocarbons (aza-PAHs), modelling carbocations from oxidized metabolites and their covalent adducts with representative nucleophiles. *Org Biomol Chem* 3, 1180-1188 (2005)

51. G.L. Borosky: Theoretical study concerning the reactivity of imine derivatives of polycyclic aromatic hydrocarbons. *J Comput Chem* 24, 601-608 (2003)

52. G.L. Borosky, K.K. Laali: Oxidized metabolites from benzo[a]pyrene, benzo[e]pyrene, and aza-benzo[a]pyrenes. A computational study of their carbocations formed by epoxide ring opening reactions. *Org Biomol Chem* 5, 2234-2242 (2007)

53. G.L. Borosky, K.K. Laali: Carbocations from oxidized metabolites of benzo[a]anthracene: a computational study of their methylated and fluorinated derivatives and guanine adducts. *Chem Res Toxicol* 19, 899-907 (2006)

54. J. Bentzien, E.R. Hickey, R.A. Kemper, M.L. Brewer, J.D. Dyekjaer, S.P. East, M. Whittaker: An *in silico* method for predicting Ames activities of primary aromatic amines by calculating the stabilities of nitrenium ions. *J Chem Inf Model* 50, 274-297 (2010)

55. Y. Takahata, D.P. Chong: Estimation of Hammett Sigma Constants of Substituted Benzenes Through Accurate Density-Functional Calculation of Core-Electron Binding Energy Shits. *Int J Quantum Chem* 103, 509-515 (2005)

56. A.M. Birch, S. Groombridge, R. Law, A.G. Leach, C.D. Mee, C. Schramm: Rationally designing safer anilines: the challenging case of 4-aminobiphenyls. *J Med Chem* 55, 3923-3933 (2012)

57. A.G. Leach, R. Cann, S. Tomasi: Reaction energies computed with density functional theory correspond with a whole organism effect; modelling the Ames test for mutagenicity. *Chem Commun (Camb)* 1094-1096 (2009)

58. A.G. Leach, W. McCoull, A. Bailey, P. Barton, C. Mee, E. Rosevere: Experimental Testing of Quantum Mechanical Predictions of Mutagenicity: Aminopyrazoles. *in press Chem Res Toxicol* (2013)

59. P. McCarren, C. Springer, L. Whitehead: An investigation into pharmaceutically relevant mutagenicity data and the influence on Ames predictive potential. *J Cheminform* 3, 51 (2011)

60. P. McCarren, G.R. Bebernitz, P. Gedeck, S. Glowienke, M.S. Grondine, L.C. Kirman, J. Klickstein, H.F. Schuster, L. Whitehead: Avoidance of the Ames test liability for aryl-amines via computation. *Bioorg Med Chem* 19, 3173-3182 (2011)

61. S Glowienke: S18: *In silico* assessment of safety concerns esp. of carcinogenic potential. *Exp Toxicol Pathol* 61, 264-265 (2009)

62. TOPKAT, v 6.2. Accelrys, San Diego, CA (2004)

63. K.L. Dobo, N. Greene, M.O. Cyr, S. Caron, W.W. Ku: The application of structure-based assessment to support safety and chemistry diligence to manage genotoxic impurities in active pharmaceutical ingredients during drug development. *Regul Toxicol Pharmacol* 44, 282-293 (2006)

64. K.L. Dobo, N. Greene, C. Fred, S. Glowienke, J.S. Harvey, C. Hasselgren, R. Jolly, M.O. Kenyon, J.B. Munzner, W. Muster, R. Neft, M.V. Reddy, A.T. White, S. Weiner: *In silico* methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. *Regul Toxicol Pharmacol* 62, 449-455 (2012)

65.A. Sutter, A. Amberg, S. Boyer, A. Brigo, J.F. Contrera, L.L. Custer, K.L. Dobo, V. Gervais, S. Glowienke, J. van Compel, N. Greene, W. Muster, J. Nicolette, M.V. Reddy, V. Thybaud, E. Vock, A.T. White, L. Müller: Use of *in silico* systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. *Regul Toxicol Pharmacol* 67, 39-52 (2013)

66. J. Bentzien, I. Muegge, B. Hamner, D.C. Thompson: Crowd computing: using competitive dynamics to develop and refine highly predictive models. *Drug Discov Today* 18, 472-478 (2013)

67. Kaggle: Merck Molecular Activity Challenge. http://www.kaggle.com/c/MerckActivity (2012)

68. J. Markoff: Scientists see promise in deep-learning programs. *The New York Times* (2012)

69. K.R. Lakhani, K.J. Boudreau, P.R. Loh, L. Backstrom, C. Baldwin, E. Lonstein, M. Lydon, A. MacCormack, R.A. Arnaout, E.C. Guinan: Prize-based contests can provide solutions to computational biology problems. *Nat Biotechnol* 31, 108-111 (2013)

Key Words: *In silico* modeling, Ames, Toxicology, Genotoxicity, Mutagenicity, Aromatic Amines, Nitrenium Ion, Crowd-Sourcing, Review

Send correspondence to: Jorg Bentzien, Department of Medicinal Chemistry, Structural Research, Boehringer Ingelheim Pharmaceuticals, 900 Ridgebury Road, Ridgefield, CT, 06877-0365, U.S.A., Tel: 203-791-6185, Fax: 203-791-6072, E-mail: jorg.bentzien@boehringer-ingelheim.com