A knowledge-based expert rule system for predicting mutagenicity (Ames test) of aromatic amines and azo compounds

Domenico Gadaleta\textsuperscript{a,*}, Serena Manganelli\textsuperscript{a}, Alberto Manganaro\textsuperscript{b}, Nicola Porta\textsuperscript{a}, Emilio Benfenati\textsuperscript{b}

\textsuperscript{a} Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS—Istituto di Ricerche Farmacologiche Mario Negri, Via Giuseppe La Maia 19, 20156 Milan, Italy
\textsuperscript{b} Kode s.r.l., Via Nino Pisano 14, 56122 Pisa, Italy

\textbf{Article info}

\textbf{Article history:}
Received 29 July 2016
Received in revised form 14 September 2016
Accepted 15 September 2016
Available online 16 September 2016

\textbf{Keywords:}
Ames test
Aromatic amines
Azo dyes
Expert system
\textit{In silico}

\textbf{Abstract}

Cancer is one of the main causes of death in Western countries, and a major issue for human health. Prolonged exposure to a number of chemicals was observed to be one of the primary causes of cancer in occupationally exposed persons. Thus, the development of tools for identifying hazardous chemicals and the increase of mechanistic understanding of their toxicity is a major goal for scientific research. We constructed a new knowledge-based expert system accounting the effect of different substituents for the prediction of mutagenicity (Ames test) of aromatic amines, a class of compounds of major concern because of their widespread application in industry. The herein presented model implements a series of user-defined structural rules extracted from a database of 616 primary aromatic amines, with their Ames test outcomes, aimed at identifying mutagenic and non-mutagenic chemicals. The chemical rationale behind such rules is discussed. Besides assessing the model’s ability to correctly classify aromatic amines, its predictivity was further evaluated on a second database of 354 azo dyes, another class of chemicals of major concern, whose toxicity has been predicted on the basis of the toxicity of aromatic amines potentially generated from the metabolic reduction of the azo bond. Good performance in classification on both the amine (MCC, Matthews Correlation Coefficient = 0.743) and the azo dye (MCC = 0.584) datasets confirmed the predictive power of the model, and its suitability for use on a wide range of chemicals. Finally, the model was compared with a series of well-known mutagenicity predicting software. The good performance of our model compared with other mutagenicity models, especially in predicting azo dyes, confirmed the usefulness of this expert system as a reliable support to \textit{in vitro} mutagenicity assays for screening and prioritization purposes. The model has been fully implemented as a KNIME workflow and is freely available for downstream users.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

1.1. Background

Cancer is one of the main causes of death in Western countries. In 2012 there were about 14.1 million new cases, with globally about 8.2 million deaths (14.6% of total human deaths) [Stewart and Wild, 2014]. At present it is widely accepted that, together with the increased life expectancy, prolonged exposure to a number of synthetic and natural chemicals in the environment is a primary cause of cancer onset (Albrecht et al., 2008). Cancer prevention is today therefore a critical health issue.

Because of its serious social impact, great efforts have been made in the last few decades to understand and prevent the causes of cancer induced by exposure to chemicals. As a consequence,
carcinogenicity has been the subject of a long series of mechanistic investigations (Benigni and Bossa, 2011). Carcinogenesis is a pathological process, caused by permanent damages to genetic material of cells. Carcinogenicity in vivo assays cannot be performed for a high number of substances because of time, cost and ethical issues. Mutagenicity is widely recognized as a valid surrogate of carcinogenicity and its assessment is explicitly required by several regulations in the field of chemical safety (FDA, 2013; EC, 2006, 2009). The term mutagenicity refers to the ability of a chemical to induce genetic damages, that may occur by several mechanisms involving interactions with the DNA (i.e., formation of adducts, base substitutions, frame-shift deletions, intercalations) or with both the DNA and other cellular targets, e.g., proteins (i.e., chromosomal aberrations, and changes in the number of chromosomes). (Benigni and Bossa, 2011). Mutagenicity assessment is a suitable first step of a tiered strategy for the identification of potential hazardous compounds in large screening programs, because it is simpler than carcinogenicity assessment.

Indeed, a relative simple and rapid in vitro assay was proposed by Bruce Ames for detecting DNA mutations induced by chemicals (Ames et al., 1975; Ames, 1979). In the Ames test, frame-shift mutations or base-pair substitutions are detected by exposure of histidine-dependent genetically engineered strains of Salmonella typhimurium to the chemical to be tested. When these strains are exposed to a mutagen, reverse mutations restore the bacteria’s ability to synthesize histidine and thus to grow on a medium deficient in this amino acid (Hansen et al., 2009). Application of the Ames test to large numbers of chemicals has shown that this assay has high positive predictivity for DNA-reactive chemical carcinogens, confirming a causal relationship between genetic damage and cancer insurgence (Ziegler et al., 1990). Today the Ames test is by far the most commonly used, long-established in vitro test for chemical mutagenicity screening (OECD, 1997).

Aromatic amines are a class of chemicals traditionally recognized as of high concern for human health. They find applications in several chemical industry manufacturing sectors such as oil refining, production of synthetic polymers, adhesives and rubbers, pharmaceuticals, pesticides and explosives (Snyder-wine et al., 2002). They may also be generated through the combustion of organic materials, such as emissions of tobacco smoke (Platzek, 2009). Epidemiological studies have confirmed that some of these chemicals induce bladder cancer in occupationally exposed persons (Skipper et al., 2010).

Prolonged exposure of humans to carcinogenic aromatic amines is a primary issue in the dye manufacturing industry. Another class of chemicals related to the aromatic amines, i.e., azo compounds, is widely employed as industrial dyes. Azo dyes were detected as potential carcinogens as early as the 1930’s, when Kinosita (1936) reported that N,N-dimethyl-4-aminoazobenzene, commercially known as “butter yellow”, induced liver tumors in rats. The toxicity of azo dyes can be explained by the generation of carcinogenic aromatic amines after reductive cleavage of the azo bond. This can occur in a variety of conditions, including those encountered in the digestive tract of mammals (Pinheiro et al., 2004; Øllgaard et al., 1998; Weber and Adams, 1995). Currently, more than 3000 individual azo colorants are in use, accounting for 60–70% of all dyes used (ETAD, 2003). For this reason, they are a major subject of attention in carcinogenicity studies and occupational health preventive actions (Pinheiro et al., 2004).

With the recent introduction of more stringent chemical safety regulations, in silico methods have been recognized as a valuable support and, sometimes, as an alternative to in vivo and in vitro assays. In silico methods can provide information for a very large number of substances in relatively short times and at low cost. The proven reproducibility and ease of execution of the Ames test has made an abundance of experimental data available in recent years. This has led to the development of many in silico models for the prediction of Ames mutagenicity, which has become one of the most commonly modeled endpoints, with the best results. The effectiveness of in silico methods for this endpoint is demonstrated by the fact that regulatory agencies may consider candidate genotoxic impurities (GTIs) predicted as non-mutagenic by validated in silico models equivalent to Ames negative ones (FDA, 2008, 2013).

1.2. In silico models for predicting the mutagenicity of aromatic amines: state of the art

Numerous Quantitative Structure-Activity Relationship (QSAR) approaches for the prediction of aromatic amine mutagenicity have been proposed in the last decades. Debnath et al. (1992) compiled a large database of chemicals with various chemical scaffolds (e.g., aniline, biphenyl, anthracene, pyrene, quinoline, carbazole), with quantitative mutagenicity data determined in experiments on Salmonella TA98 and TA100 strains, with S9 metabolic activation. Debnath et al. found a quantitative correlation between mutagenicity and the electrostatic and hydrophobic properties of these chemicals, expressed respectively by the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) energies, and log P. Since then, several studies were performed on the dataset compiled by Debnath et al. Maran et al. (1999) built a six descriptor model starting from a large pool of constitutional, geometrical, topological, electrostatic, and quantum chemical descriptors. Gramatica et al. (2003) derived Genetic Algorithm – Multiple Linear Regression (GA-MLR) models based on theoretical descriptors available from DRAGON software (Todeschini and Consonni, 2008). Basak and Gruenwald (1995) derived k-Nearest Neighbor (k-NN) models comparing the effectiveness of atom-pair counts, topological indices and physicochemical parameters for computing similarity between chemicals. Bhat et al. (2005) developed Artificial Neural Networks (ANNs) based on a variety of molecular descriptors calculated using quantum-chemical semiempirical methods. Several QSAR studies were proposed by Hatch and coworkers on the mutagenic potency (frame-shift mutations in TA98 or TA1538 Salmonella strains) of aromatic amines. A first study (Hatch et al., 1991) suggested a correlation between the mutagenic potency of a series of aminoimidazo-azaarene and aminocarboline and a series of relevant structural features (e.g., number of fused rings, number of heteroatoms in rings, and methyl substitution on ring atoms). Further studies on aminoimidazoazaarenes (Hatch et al., 1996) and aromatic and heteroaromatic amines (Hatch and Colvin, 1997; Hatch et al., 2001) also highlighted the role of electronic properties, such as LUMO energies, in the modulation of the mutagenic potency of these chemicals. The key role of electronic properties, particularly HOMO and LUMO energies, was confirmed by several other computational studies (Zhang et al., 1993; Lewis et al., 1995; Felton et al., 1999).

While QSARs for aromatic amines were effective for modeling mutagenicity, they were not suitable for distinguishing mutagenic from non-mutagenic amines (Benigni et al., 1998). As a result, fewer examples of successful classification models are reported in literature. A noteworthy attempt was proposed by Benigni et al. (2007): amines were first separated into structural subclasses, then for each class different factors were considered for classification purposes, e.g. classification of single-ring amines was mainly based on electronic factors, biphenyls on steric factors, while fused-ring amines were always classified as mutagens.

The majority of the models above have never been implemented. Hence they can be not useful for regulatory purposes. On the other hand, there are several widely used software implementing mutagenicity (Ames) models for the prediction of a wide
range of chemical classes. However, these software often fail in predicting certain classes of chemicals, e.g. aromatic amines and azo compounds (see Section 4.3).

Here we propose a new knowledge-based expert system for the classification of primary aromatic amines on the basis of their mutagenic potential (i.e., mutagen, non-mutagen). The system was developed as a decision algorithm, implementing a series of hierarchically applied rules to distinguish toxic from non-toxic aromatic amines. The electrophilic theory of chemical carcinogenesis (Miller and Miller, 1981) states that the majority of mutagens, including aromatic amines, either are electrophiles, or are converted in vivo to reactive electrophilic derivatives. These derivatives exert their toxicity by binding covalently to nucleophilic centers of macromolecules, such as the DNA. In this light, the rules represent particular functional groups/structural features able to generate electrophilic derivatives, or groups able to favor/disfavor this conversion through electronic effects, as well as influencing the reactivity of the final electrophilic intermediates.

The predictive power of the knowledge-based system was evaluated on two separate datasets of chemicals, one including primary aromatic amines and the other aromatic azo compounds. The model was implemented as a KNIME workflow, with the aim to provide a useful and freely available tool to support the risk assessment of two different classes of hazardous chemicals. The approach here presented can be considered valuable in reducing the number of experimental assays, e.g. in large screening programs, in combination with other methods as a part of weight-of-evidence strategy, or for prioritization of experimental assays. At the same time, this manuscript will provide confirmation, as well as new evidence about the chemical and mechanistic rationale behind the toxicity of aromatic amines and azo compounds.

2. Methods

2.1. Datasets

2.1.1. Mutagenicity data

Two datasets comprising 616 primary aromatic amines and 354 azo compounds (Manganelli et al., 2016) were extracted from a larger dataset of about 10,000 organic compounds with their mutagenic activities, from different sources:

1) A freely available Benchmark dataset for in silico predictions of Ames mutagenicity (Hansen et al., 2009), already used in the European Commission funded project ANTARES (IRFMN, 2011) to evaluate the performance of eight QSAR models. These data were derived from several well-known sources such as Chemical Carcinogenesis Research Information (NCI, 1985–2011), Helmia et al. (2004), Kazius et al. (2005), Feng et al. (2003), VITIC (Marchant et al., 2008), and the GeneTox databases (Judson et al., 2005);

2) Data provided by the Japan’s Ministry of Health (data produced within the Ames (Q)SAR project organized by National Institute of Health Sciences of Japan). The Ames assays were conducted under GLP according to the Industrial Safety and Health Act in Japan;

3) The Scientific Committee on Consumer Safety dossier on hair dyes (Ates et al., 2016);

4) A database assembled by Italian dye companies from different sources, mainly US and Canadian authorities (Kulkarni and Barton-Maclaren, 2014).

The datasets contain: i) chemical structure represented as Simplified Molecular Input Line Entry System (SMILES) strings where available, ii) CAS number and/or common name as identifiers of each compound, and iii) experimental mutagenicity activity from the Ames assay (Ames et al., 1975; Ames, 1979). As for negative compounds, we kept only those for which all the Ames assays had been conducted by Good Laboratory Practice (GLP) laboratories according to the OECD TG 471 (OECD, 1997) guideline with and without metabolic activation. In the dataset of 616 aromatic amines, mutagenic (about 74%) and non-mutagenic (about 26%) compounds are unevenly distributed, while the distribution for the 354 azo compounds is more balanced (54% mutagenic, 46% non-mutagenic).

2.1.2. Carcinogenicity data

Experimental rodent carcinogenicity data relative to a series of aromatic amines were collected from Franke et al. (2001) and used to verify the correlation between carcinogenicity data and mutagenicity predictions generated by the knowledge-based expert system. Franke et al. classified chemicals in three groups on the basis of experimental evidence supporting their carcinogenicity, i.e. group 1 (inactive), group 2 (equivocal or borderline) and group 3 (actives). In the present study, only chemicals included in groups 1 and 3, and only primary aromatic amines were considered. A final dataset including 55 primary aromatic amines and their in vivo rodent carcinogenicity test results (actives/inactives) was obtained.

2.2. A rule-based expert system for predicting mutagenicity of aromatic amines

A knowledge-based expert system was developed for the classification of aromatic amines on the basis of their mutagenic potential. The system was built in the form of a decision algorithm comprising a number of filters hierarchically connected, each running a test on a query chemical based on a series of selected structural features. From the outcome of this test, the algorithm moves to the next filter, where a new test is done, or to a leaf node that represents a class (mutagen, non-mutagen) to which the query is assigned. Structural rules for the mutagenic potential were formulated manually on the basis of the experimental activities of the 616 aromatic amines in the first dataset, as well as on literature findings.

The system can be divided in three logical blocks, each considering a different kind of structural filter, as follows:

<table>
<thead>
<tr>
<th>Class</th>
<th>Aromatic atoms count</th>
<th>Number of ring systems</th>
<th>Ring bridge count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Monocycles</td>
<td>5 or 6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2. Bicycles, non-fused</td>
<td>10, 11 or 12</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3. Three or more cycles, non-fused</td>
<td>9</td>
<td>3 or more</td>
<td>0</td>
</tr>
<tr>
<td>4. Polycycles, 5+6 terms</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5. Polycycles, 6+6 terms</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6. Polycycles, more than three fused rings</td>
<td>–</td>
<td>1</td>
<td>2 or more</td>
</tr>
</tbody>
</table>
1) A preliminary filter based on the size of the chemicals was implemented, with the aim to exclude bulky chemicals and those characterized by an unfavorable geometry (e.g., non-planar chemicals) that were less likely to be toxic because of their Adsorption, Distribution, Metabolism Excretion (ADME) behavior (Benigni, 2008). This filter was based on the evaluation of a series of properties included in the “Molecular Properties” node implemented within the Chemistry Development Kit (CDK) freely available extension of KNIME software (Berthold et al., 2008) (www.knime.org/). A threshold was assigned to each property; chemicals exceeding at least one of the thresholds were classified as non-mutagens. Selected properties and their respective thresholds were: i) molecular weight (430 Da), ii) rotatable bonds count, non-terminal (5 bonds), iii) largest chain, estimating the length of the longest aliphatic chain in a molecule (10 atoms) and iv) sp3 character, giving the ratio between the sp3 atom count and total number of carbons (0.17).

2) In the second part of the expert system, amines were grouped in subclasses describing different aromatic systems. These subclasses (Table 1) were defined on the basis of the following descriptors: i) aromatic atoms count, ii) number of ring systems (Feher and Schmidt, 2003) and iii) ring bridge count (Lipkus, 2001). The first descriptor was calculated using the “Molecular Properties” node available in the CDK extension of KNIME; the others were included in the “Ring descriptors” block calculated with Dragon 7 software (Kode Chemoinformatics S.r.l., 2016) and were implemented ad hoc within the KNIME workflow as Java nodes. The chemicals not included in these classes were not processed in the next steps so no final predictions were returned.

3) The last part of the model checked the occurrence of a series of substructures affecting the mutagenic potential of aromatic amines. These features were grouped in four classes: i) amine-generating groups (AGGs) (including additional primary amino groups), ii) deactivating groups, iii) activating groups, iv) conjugated structures, and v) aromatic nitrogen atoms. Classes i, iii and iv included fragments and substructures characteristic of aromatic mutagenic amines, while fragments in classes ii and v were observed in non-mutagenic chemicals. A fuller explanation of the role of each fragment in the mutagenicity of aromatic amines is proposed in Section 4. Fragments in each group were encoded into Smiles Arbitrary Target Specification (SMARTS) notation (Daylight Chemical Information Systems Inc., 2011) (Table 2). The “SMARTS Query” node available from the CDK extension of KNIME was applied to verify the occurrence of SMARTS in each test chemical.

The entire knowledge-based classification system was developed as a KNIME workflow (Fig. 1). It is freely available from https://chm.kode-solutions.net/research/Supporting-Materials-Gadaleta-2016.zip.

2.3. Generation of aromatic amines from azo dyes

The toxicity of azo dyes is related to the potential cleavage of the azo bonds and the generation of amine mutagenic products (Ølgaard et al., 1998; Weber and Adams, 1995). We developed an in house application in Java language, based on the CDK and VEGA libraries, to generate aromatic amines from azo compounds by replicating the reductive cleavage of azo bonds. This tool fragments each azo bond in parent compounds, giving a set of all possible primary amine products. The software requires a list of SMILES encoding chemical structures of azo compounds as input file. We used the dataset of 354 azo compounds as input and, by splitting, we obtained 825 aromatic primary amines (about 380 without considering replicates). The original azo compounds were classified by the knowledge-based system presented here, considering the predicted activity of theoretically generated amines. If at least one of the generated amines was predicted as mutagenic, then the original azo compound was classified as a mutagen; otherwise it was classified as a non-mutagen. Two different prediction strategies were used if at least one of the generated amine was unpredicted: 1) in the first strategy, unpredicted amines were ignored and the parent azo compound was predicted on the basis of remaining amines; 2) in the second strategy, azo compounds generating at least one unpredicted amine were left unpredicted. The inhouse application was included as a Java node within the KNIME workflow (https://chmkode-solutions.net/research/Supporting-Materials-Gadaleta-2016.zip).

2.4. Evaluation of classification performance

Performance in classification of the model above described was evaluated by Cooper statistics (Cooper et al., 1979). Predictions given by the classification system were defined as: i) True Positives (TPs), i.e. correctly classified positive compounds (mutagens); ii) True Negatives (TNs), i.e. correctly classified negative compounds (non-mutagens); iii) False Negatives (FNs), i.e. misclassified positive compounds; iv) False Positives (FPs), i.e., misclassified

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>SMARTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Amine generating groups</td>
<td>Amine</td>
<td>a[N][R2]</td>
</tr>
<tr>
<td></td>
<td>N-methylamine</td>
<td>a[NH][CH3]</td>
</tr>
<tr>
<td></td>
<td>N,N-dimethylamine</td>
<td>a<a href="CH3">N</a>[CH3]</td>
</tr>
<tr>
<td></td>
<td>N-ethylamine</td>
<td>a[NH][CH2][CH3]</td>
</tr>
<tr>
<td></td>
<td>N,N-diethylamine</td>
<td>a<a href="CH2">N</a>[CH3][CH3]</td>
</tr>
<tr>
<td></td>
<td>Hydroxyamine, Nitro, Nitroso, Hydrazine, etc.</td>
<td>aNH</td>
</tr>
<tr>
<td></td>
<td>N-acrylamide</td>
<td>aNH</td>
</tr>
<tr>
<td>2. Deactivating groups</td>
<td>C=O (X is a halogen)</td>
<td>aC=O</td>
</tr>
<tr>
<td></td>
<td>Nitrile</td>
<td>aC=O</td>
</tr>
<tr>
<td></td>
<td>Sulfonyl</td>
<td>aS</td>
</tr>
<tr>
<td>3. Activating groups</td>
<td>Hydroxyl</td>
<td>a[OH]</td>
</tr>
<tr>
<td></td>
<td>Methyl</td>
<td>a[CH3]</td>
</tr>
<tr>
<td></td>
<td>Ethyl</td>
<td>a[CH3][CH3]</td>
</tr>
<tr>
<td></td>
<td>Methoxyl</td>
<td>aO[CH3]</td>
</tr>
<tr>
<td></td>
<td>Ethoxyl</td>
<td>aO[CH2][CH3]</td>
</tr>
<tr>
<td>4. Conjugated structures</td>
<td>Biphenyl-like</td>
<td>a-a</td>
</tr>
<tr>
<td></td>
<td>Stilbene-like</td>
<td>a-C-C-a</td>
</tr>
<tr>
<td>5. Aromatic nitrogen atoms</td>
<td>Aromatic nitrogen atoms</td>
<td>a-n</td>
</tr>
</tbody>
</table>
negative compounds. Cooper statistics used for evaluating the classification are:

1) **Accuracy**
   
   Accuracy = \( \frac{TP + TN}{TP + TN + FN + FP} \)

2) **Sensitivity** is defined as the ratio of TPs to the total of the experimentally positive compounds;

   Sensitivity = \( \frac{TP}{TP + FN} \)

3) **Specificity** is equal to the ratio of TNs to the total of the experimentally negative compounds.

   Specificity = \( \frac{TN}{TN + FP} \)

Cooper statistics are often not suitable for assessing the goodness of classification when the populations of the two classes are numerically unbalanced. In this case, a useful parameter is the MCC. This ranges between –1 and +1, a value of +1 indicating an optimal classification, a 0 a random classification, and –1 optimal misclassification (Matthews, 1975).

\[
MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\]

### 2.5. Mutagenicity predicting models

We compared our model with five widely used mutagenicity (Ames) predicting models, four implemented within the VEGA platform (www.vega-qsar.eu) (i.e., CAESAR, ISS, SARpy/IRFMN, and k-NN models) and another one developed by Environmental Protection Agency’s (EPA), i.e. T.E.S.T. (www.epa.gov/chemical-research/toxicity-estimation-software-tool-test). These models are publicly available, allowing straightforward comparisons of performance and reliability.

Here is a brief description of each model:

1. **VEGA CAESAR model** (Ferrari and Gini, 2010) is a Support Vector Machine (SVM) classifier, implementing Structural Alerts (SAs) from Benigni and Bossa (2008) for FNs detection and exclusion.
2. **VEGA SARpy/IRFMN** (Ferrari et al., 2013) is a rule-based model developed by extracting from training compounds two sets of rules for mutagenicity (112 rules) and non-mutagenicity (93 rules), by the SARpy software. If at least one mutagenicity rule matches, a mutagen prediction is given; if one or more non-mutagenicity rule is matching, a “possible non-mutagen” prediction is given.
3. **VEGA ISS model** applied Benigni and Bossa rules for mutagenicity (Benigni and Bossa, 2011) as implemented by ToxTree software (v. 2.6; http://toxtree.sourceforge.net). If at least one mutagenicity rule matches with the given compound, it is classified as mutagenic, otherwise, it is non-mutagenic.
4. **VEGA k-NN model** implemented in VEGA employs the similarity index described by Floris et al. (2014) searching the four compounds from the dataset that are most similar to the chemical to be predicted (Pizzo et al., 2016). Compounds with similarity lower than 0.7 are discarded. The prediction for a given compound is the class most represented among its similar compounds.
5. **EPA’s T.E.S.T.** employs different algorithms for mutagenicity (hierarchical clustering, single model, group contribution, FDA method, nearest neighbor). It calculates an average of the predicted values using a consensus model.

---

Fig. 1. Knowledge-based expert rule system for the classification of mutagenic and non-mutagenic aromatic amines. AGGs: additional primary amines/amine generating groups; DG: deactivating groups; AG: activating groups; Nar. aromatic nitrogen atoms; MW: molecular weight; RTB: rotatable bonds; LC: largest chain; sp3: sp3 character.
3. Results

Chemicals classified by the knowledge-based system were arranged in a confusion matrix (Kohavi and Provost, 1998). Table 3 shows the confusion matrices for the classification results for amines (Table 3a) and azo dyes. Statistics on the second class of compounds where evaluated both ignoring generated amines that were unpredicted (Table 3b) and rejecting azo compounds generating at least one unpredicted amines (Table 3c). The compounds correctly classified in each class are reported on the main diagonal, and misclassifications on the non-diagonal cells. Table 4 shows the Cooper statistics and MCC for the primary amines and the azo dyes databases.

The classification system predicted 94% of the aromatic amines with good statistical performance. Accuracy, sensitivity and specificity were well-balanced and greater than 0.880. MCC was 0.743. The statistics for azo compounds were lower, but still acceptable, with MCC of 0.532 and accuracy of 0.777 when ignoring unpredicted amines. Slightly better performance was obtained by excluding azo compounds generating unpredicted amines (MCC was 0.584 and accuracy was 0.791). The loss of performance was due to the high percentage of FPs among azo compounds, leading to lower specificity compared to aromatic amines. On the other hand, sensitivity values were comparable for the two datasets.

In order to reduce FPs, the percentage of mutagenic amines generated by each azo compound was used as an indicator of the reliability of mutagenic predictions, being those positive predicted compounds characterized by a lower percentage more likely to result in FPs. Table 5 shows the increase in classification performance by excluding azo compounds with a percentage of mutagenic amines lower than a given threshold. As expected, the application of greater thresholds led to the exclusion of the majority of FPs, at the cost of a loss of total predictions and of a slightly reduction of sensitivity. The application of a threshold equal to 0.500 allowed to reach maximum MCC and accuracy values of 0.751 and 0.876, respectively, for about the 35% of the initial dataset.

To understand the real predictive power and its applicability for real-life purposes, performance of the model here described were compared with those of a series of widely used and freely available mutagenicity predicting models. Comparison was relative on both the dataset of 616 aromatic amines and the dataset on 354 azo compounds. As shown in Table 6, none of the models used for comparison achieved performance comparable with the model presented here on the aromatic amine dataset, with only the CAESAR and k-NN models implemented in VEGA, and EPA’s T.E.S.T. model reaching acceptable MCC values of 0.549, 0.669, and 0.557, respectively. The performance of existing models was also disappointing for azo compounds, with only EPA’s T.E.S.T. model returning a near-acceptable MCC of 0.490.

4. Discussion

4.1. Structural rules

Lai et al. (1996) summarized the structural requirements for the mutagenic potential of aromatic amines. The basic requirements are: i) an amino group or a group that can generate an amine after metabolic interconversion (i.e., an AGG); ii) the nature, number, and position of other ring substituents; iii) one or more rings forming a conjugated system, fused or non-fused; and iv) the size and shape of the molecule. Such structural features were employed for developing the expert system presented here for the prediction of mutagenicity of aromatic amines. This model can easily be adapted to other chemical species whose toxicity is closely related to those of amines, such as azo dyes. The ability of the model to predict different classes of chemicals, and its predicitvity evaluated on two different datasets, confirmed its utility for regulatory purposes and real-life applications. Here an attempt to explain a mechanistic relationship between the structural features and the variation of the mutagenic potential of aromatic amines was proposed.

4.1.1. Amine-generating groups

Aromatic amines exert their mutagenicity after metabolic interconversion to highly reactive electrophiles that bind covalently to DNA (Benigni and Bossa, 2011). The main metabolic pathway consists of a series of oxidation steps (Fig. 2), initially involving N-oxidation of the aniline to hydroxylamine, mediated by cytochrome P450 (Cramer et al., 1960). The N-hydroxylamine undergoes enzymatic acetylation or sulfonation to deliver a good leaving group which leads to the formation of a nitrenium ion (Smith, 2011; Beland and Kadlubar, 1990; Benigni et al., 2000). This highly reactive electrophilic specie can bind covalently to biomacromolecules, generating aminooxy derivatives (Benigni and Bossa, 2011). N-hydroxylamines can also be metabolically converted to nitroso species that are reactive on their own. Redox cycling between nitroso and nitro species leads to the formation of Reactive Oxygen Species (ROS), leading to oxidative damage.

The additional amino groups increase the mutagenic potential of the chemical, because there are more theoretically reactive centers. As described above, N-hydroxyamine, nitroso and nitro groups are able to generate amino groups (due to metabolic interconversion) (Benigni et al., 2007). Therefore these groups can be considered potential toxicophores in the same way as amino groups, and are in fact termed AGGs (Lai et al., 1996). Monoalkylamino/dialkylamino groups are also AGGs since metabolic N-dealkylation readily occurs in vivo for small alkyl groups. On the other hand, N-dealkylation and bioactivation does not occur for dialkyl anilines with bulkier or longer N-alkyl substitution (Gorrod and Damani, 1979) so their mutagenicity is attenuated. The same oxidation pathway was observed for N-acetyl aniline species (Smith, 2011; Brodie and Axelrod, 1948; Veronese and McLean, 1991). This is consistent with the evidence that acetanilide shows mutagenic activity similar to aniline (Sicardi et al., 1991). The presence of additional amines or AGGs increase the mutagenic
potential. Out of 616 aromatic amines in the dataset, 244 includes AGGs. 209 of these compounds (i.e., 86%) are experimentally mutagens.

### 4.1.2. Ring substituents

Some ring substituents other than amino or AGGs influence the mutagenic potential of aromatic amines, on the basis of their steric and electronic properties. The mutagenicity of aromatic amines is mainly related to i) the degree of oxidizability of the amino group and ii) the stability of the generated nitrenium ion (Loew et al., 1978; Smith, 2011).

In the classification model here presented, ring substituents exerting an effect on the mutagenicity of amines have been grouped in two classes, i.e. deactivating and activating groups. For deactivating groups, i) nitriles, ii) sulfuryl group (including sulfones and sulfonic acids) and iii) trihalo methyl groups (such as trichloro methyl or trifluoro methyl) were considered. The mitigating effect on mutagenicity of these groups can be explained by their electron-withdrawing nature. This reduces the electron availability of the lone pair of the amine nitrogen, reducing its reactivity and making it less susceptible to oxidation to N-hydroxylamine. Several studies (Ford and Herman, 1992; Sabbioni and Wild, 1992; Wild and Dirr, 1989) have noted a correlation between aromatic amine mutagenicity and the stability of the nitrenium ion: the more stable the nitrenium ions, the more mutagenic the corresponding aromatic amines. Electron-withdrawing groups (EWG) destabilize the nitrenium ion, reducing the mutagenic potential of the parent amine. As a simple example, trifluoromethyl aromatic compounds were historically used as intermediates for the synthesis of a number of dyes, including azo, anthraquinone, and triphenylmethane-based dyes (Siegemund et al., 2000). Trifluoro methyl substituted anilines are usually negative in the Ames test because of higher nitrenium formation energies (McCarren et al., 2011) and because of the strongly destabilizing effect of fluoride atoms on the nitrenium cationic specie (Chambers, 2004). The electron-withdrawing properties of the sulfuryl groups also decrease the mutagenic potential of aromatic amines. The increase in solubility related to the presence of sulfonic acids may reduce the potential biological activity of aromatic amines (including those released from highly sulfonatedazo dyes) (Brantom, 2005) because they are rapidly absorbed, modified by the liver and excreted in the bile and urine (Parkinson and Brown, 1981) before they can exert their toxicity. Out of 616 aromatic amines in the dataset, 45 includes deactivating groups. 30 of those compounds (i.e., 66%) are experimentally non-mutagens.

Aromatic nitrogen atoms are also listed as negative modulators in the model. These atoms in nitrogen-based heterocyclic compounds act as EWG because of their electro-negativity. Electron density is not evenly distributed over hetero-aromatic rings, reflecting the negative inductive effect of nitrogen atoms (Joule and Mills, 2010).

For activating groups, 1) small alkyl groups (i.e., methyl and ethyl), 2) small alkyl ethers (i.e., methoxy and ethoxy) and 3) aromatic hydroxyls were listed as positive modulators of amine mutagenic potential. Similarly to what we have already described, these electron donating groups (EDG) might explain the increased toxicity, related to the enhanced reactivity of the aromatic amine in the first oxidation step and to the greater stability of the nitrenium ion. This is consistent with Marques et al. (1997) who described a reduction of the half-wave oxidation potentials of a series of methyl, dimethyl- and ethyl-anilines compared to unsubstituted aniline; moreover, the half-wave oxidation potentials of these alkanilines were inversely correlated with the stability of the corresponding nitrenium ions, calculated by semiempirical methods. Not unexpectedly, Sabbioni (1993) found that enthalpies of formation for nitrenium ions derived from para- substituted alkanilines were lower than the meta- substituted analogues. This again can be explained by the greater electron-donating ability of para- alkyl substituents that will contribute to stabilizing

---

**Table 4** Classification by the knowledge-based expert classification system on the datasets of 616 aromatic amines and 354 azo compounds. Statistics on azo compounds were evaluated a) ignoring unpredicted amines and b) excluding compounds generating at least one unpredicted aromatic amine.

<table>
<thead>
<tr>
<th></th>
<th>Aromatic amines</th>
<th>Azo compounds (a)</th>
<th>Azo compounds (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted compounds</td>
<td>580/616</td>
<td>354/354</td>
<td>326/354</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.893</td>
<td>0.777</td>
<td>0.791</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.897</td>
<td>0.901</td>
<td>0.905</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.882</td>
<td>0.632</td>
<td>0.653</td>
</tr>
<tr>
<td>MCC</td>
<td>0.743</td>
<td>0.532</td>
<td>0.584</td>
</tr>
</tbody>
</table>

---

**Table 5** Classification by the knowledge-based expert classification system on the datasets of 354 azo compounds after excluding compounds predicted as actives with a percentage of mutagenic amines lower or equal than a given threshold. Azo compounds generating at least one unpredicted aromatic amine were also rejected.

<table>
<thead>
<tr>
<th></th>
<th>Percentage of aromatic amines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.500</td>
</tr>
<tr>
<td>Predicted compounds</td>
<td>186/354</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.882</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.800</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.950</td>
</tr>
<tr>
<td>MCC</td>
<td>0.766</td>
</tr>
<tr>
<td>TPs</td>
<td>68</td>
</tr>
<tr>
<td>FPs</td>
<td>5</td>
</tr>
</tbody>
</table>

---

**Table 6** Comparison of the knowledge-based expert system here presented and others mutagenicity models. The comparison is relative to Cooper’s statistics and MCC and refers to the whole dataset.

<table>
<thead>
<tr>
<th></th>
<th>Aromatic amines (616 cmpds)</th>
<th>Azo dyes (354 cmpds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Acc</td>
<td>Sens</td>
</tr>
<tr>
<td>Expert system</td>
<td>580</td>
<td>0.893</td>
</tr>
<tr>
<td>VEGA CAESAR</td>
<td>616</td>
<td>0.833</td>
</tr>
<tr>
<td>VEGA SARpy</td>
<td>616</td>
<td>0.703</td>
</tr>
<tr>
<td>VEGA ISS</td>
<td>616</td>
<td>0.756</td>
</tr>
<tr>
<td>VEGA kNN</td>
<td>616</td>
<td>0.880</td>
</tr>
<tr>
<td>EPA’s T.E.S.T.</td>
<td>609</td>
<td>0.823</td>
</tr>
</tbody>
</table>

---
the nitrenium ion, with a consequent increase in mutagenic activity. This is true, however, only for small groups such as methyl and ethyl. For larger substituents (especially in the ortho-position), the mutagenicity of anilines is primarily governed by steric effects (Lai et al., 1996; Benigni et al., 2007) leading to less toxic compounds than analogs with smaller substituents.

Besides the oxidation pathway leading to the formation of nitrenium ions, another mechanism of toxicity of aromatic amines is oxidation of the aromatic ring ortho- or para- to the aniline nitrogen, leading to ortho- and para- hydroxy anilines, respectively (Fig. 3). These species may undergo enzymatic or spontaneous dehydrogenation resulting in the formation of iminoquinones, which are direct electrophiles (Smith, 2011; Benigni and Bossa, 2011) and exert their adverse effects through alkylation of DNA. The presence of aromatic hydroxyls or groups able to generate hydroxyls after metabolism (e.g. dealkylation of methoxy and ethoxy groups) explains the greater toxicity than unsubstituted rings, because fewer steps are needed for the formation of iminoquinone species. The lower toxicity of nitrogen-based hetero-aromatic amines can also be explained in the light of this mechanism, because the presence of ring heteroatoms may block ring oxidation in ortho- and para- positions (Smith, 2011), preventing the formation of iminoquinone species. Out of 616 aromatic amines in the dataset, 242 includes activating groups. 197 of these compounds (i.e., 86%) are experimentally mutagens.

4.1.3. Conjugated structures

An aromatic ring system is another essential feature for mutagenicity. At least one ring is required, whereas aromatic amines with more than one ring forming a conjugated system, either fused or non-fused, showed greater mutagenic activity. Polyaromatic compounds with three or more fused aromatic or heteroaromatic rings (Kazius et al., 2006; Benigni and Bossa, 2006, 2011; Benigni, 2008) are widely recognized as chemicals of concern because of their ability to show toxicity without the need for bioactivation, by intercalating and binding reversibly to DNA, causing frame shift errors in replication (Ferguson and Denny, 2007). In addition to this direct toxicity, the presence of conjugated polynuclear moleities increases the mutagenicity of aromatic amines compared to smaller aromatic systems. This is partly explained by the higher force of conjugation of extended systems that both facilitate the departure of the acyloxy/sulfonic group during the formation of the nitrenium ion through electron shift, and stabilize the nitrenium ion through charge delocalization (Lai et al., 1996). Similarly, some non-fused moleities, such as biphenyl (Chung et al., 2000) and stilbene-like structures (Mullin et al., 1987; Hooberman et al., 1994), have been reported to increase the mutagenic potential because of their high force of conjugation. Out of 616 aromatic amines in the dataset, 99 include conjugated structures. 93 of these compounds (i.e., 94%) are experimentally mutagens.

4.2. Aromatic amines as toxic products of azo dyes

Statistical analysis of the predictions returned by the model confirmed its real-life usefulness in predicting the mutagenicity of aromatic amines. Azo dyes are another class of chemicals of high concern for human health because of their widespread industrial use. Most of the attention about the possible hazards arising from the use of azo compounds has been transferred to their metabolic products. The azo bond may undergo both biotic (mediated by azoreductase enzymes) and abiotic (thermal, photochemical) degradation (Øllgaard et al., 1998; Weber and Adams, 1995; Shu et al., 1994), leading to the release of aromatic amines.

While several attempts to predict mutagenic properties of aromatic amines were described in literature (see Paragraph 1.2), fewer local models for azo compounds are available. Manganelli et al. (2016) recently proposed two new models to predict Ames mutagenicity of azo compounds. The first model made use of descriptors based on SMILES, calculated with the CORAL software. The second model was based on a k-NN algorithm.

Because the production of toxic metabolites seems crucial for the mutagenicity outcome of azo compounds, we decided to apply the model here presented to predict the mutagenicity of the dataset of 354 azo compounds used by Manganelli et al. (2016). This is a quite innovative strategy, indeed we tried to adapt a model tailored to predict aromatic amines for the evaluation of a different class of compounds. The overall solid statistics from the validation on this dataset proved the effectiveness of the method for predicting the mutagenicity of two different classes of substances.

Fig. 2. Mechanism of toxicity of aromatic amines through the formation of a reactive electrophilic nitrenium ion. Marvin (version 5.9.2, 2012 ChemAxon, http://www.chemaxon.com) was used for drawing chemical structures.

Fig. 3. Mechanism of toxicity of aromatic amines through aromatic hydroxylation and formation of reactive para- and ortho- iminoquinones. Marvin (version 5.9.2, 2012 ChemAxon, http://www.chemaxon.com) was used for drawing chemical structures.
and gave a solid confirmation to the link between the toxicity of azo compounds and aromatic amines. The only limitation was the larger number of FPs among azo compounds than aromatic amines. This may be explained by the stringent criteria used for defining mutagenic azo compounds. Chemicals generating at least one mutagenic amine after azo bond cleavage are predicted as mutagen, hence a single misclassification among generated amines would result in a FP. For negative predictions, a single misclassified amine is often irrelevant in the final prediction if there are other amines correctly predicted as mutagens. In the light of this, positive predictions for azo compounds having a low percentage of generated mutagenic amines were considered less reliable compared to those having a higher percentage of mutagenic amines. As explained in Section 3, the exclusion of less reliable predictions led to a strong improvement of performance on the remaining fraction of the dataset.

Another possible aspect worth of note is that the in house software used for simulating the cleavage of the azo bond generated every single amine that theoretically may be produced. However, certain chemical features of azo dyes may make them less susceptible to cleavage by azoreductase, so some theoretically likely toxic products are in fact never formed. For example, sulfonation of azo dyes may inhibit the release of aromatic amines (Olggaard et al., 1998). In the same way, the use of dye intermediates capable of metal complexation after incorporation into the dye structure may shield the azo bond from reduction (Freeman et al., 1996).

The formulation of rules for such hindering features would be useful to define this issue. However, the mechanisms underlying the cleavage of azo bonds are still not fully elucidated, so this remains a challenge.

4.3. Comparison with other mutagenicity predicting models

Comparison carried out on the two datasets of aromatic amines and azo compounds revealed the superiority of the model here presented over the already existing mutagenicity models. Poor performance of existing models, especially when applied to azo dyes, can be explained by the fact that the majority of them (with the exception of EPA’s T.E.S.T.) are expert systems that identify the occurrence of fragments supposed to be toxic, such as the azo group. Thus, all the compounds with this fragment are systemati- cally classified as mutagens, leading to values of specificity that are unacceptably low, and values of sensitivity that, although very high, are simply an artificial product derived from the almost total absence of negative predictions. Conversely, the model here presented showed more balanced statistics and overall better performance, as a result of an exhaustive evaluation of chemical factors attenuating the toxicity of chemicals, and not only those flagging toxic compounds.

4.4. Exploring the correlation between mutagenicity and carcinogenicity

The Salmonella mutagenicity assay is performed as a compo- nent of safety and hazard evaluation process by industrial and regulatory organizations for the identification of potential carcinogens. In order to verify the suitability of the expert system for hazard assessment, we verified the correlation between the mutagenicity predictions and the results of in vivo rodent carcinogenicity tests (as a surrogate for human carcinogenicity) for a series of primary aromatic amines.

We used our model to obtain mutagenicity predictions for 56 primary aromatic amines whose rodent carcinogenicity class was reported by Franke et al. (2001). For 42 out of 56 amines, the Ames mutagenicity predictions were in agreement with the reported in vivo rodent carcinogenicity. However, despite a fairly acceptable accuracy (i.e., 0.765), eight FPs (i.e., predicted mutagens that are non-carcinogens) and five FNs (i.e., predicted non-mutagens that are carcinogens) resulted from this analysis. FNs can be explained by the fact that negative Ames results have the same probability of being carcinogenic or not, suggesting that also other molecular pathways not involving DNA reactivity lead to carcinogenicity (i.e., non-genotoxic carcinogens) (Zeiger, 1998). On the other hand, compounds positive in the Ames test were traditionally assumed to be in vivo carcinogens, making more difficult to explain the presence of FPs. Discrepancies are not the results of model’s error, indeed seven out of eight FPs were included in the dataset used for model derivation, and all of them were indicated as experimentally mutagens. The presence of FPs can be explained by different factors, such as metabolic differences between the in vivo carcinogenicity test and in vitro Ames assay, differences between the in vitro dose and the in vivo effective dose after chemical distribution, differences in DNA repair capacity among species, and different sensitivities of the Salmonella and rodent systems to genetic damage. Similar findings were already described by Zeiger, that reported the presence of a significant number of Ames positive, non-carcinogenic chemicals resulting from the compari- son of mutagenicity and in vivo carcinogenicity results derived from the U.S. National Cancer Institute and the National Toxicology Program (Zeiger, 1998, 2001). It was also reported that the largest number of mutagens, non-carcinogens were aromatic amines or substituted aromatic amines, suggesting the need of further studies to better clarify the behavior of this class of chemicals (Zeiger, 2001).

5. Conclusions

Here we described a knowledge-based classification algorithm, implementing a series of user-defined rules aimed at predicting the mutagenic potency of aromatic amines. The model has been fully implemented as a KNIME workflow and made freely available for use as well as for future improvement. Particular attention has been paid to explaining the chemical rationale behind the rules. They describe a series of functional groups that were observed to affect mutagenicity, e.g. through their ability to generate new putative toxic reactive centers, or influence the reactivity of existing ones. This last aspect was mainly related to the ability of such functional groups to favor or disfavor the first oxidative step that results in the formation of hazardous amine reactive products, and in the ability to stabilize or destabilize the same reactive species. The model showed good predictive power, confirming its utility for real-life applications. Furthermore, because of a mechanistic relationship between the toxicity of aromatic amines and azo compounds, the same model was applied on this second class of hazardous and widely used compounds, confirming its predictive power as well as its flexibility. Comparison of the performance of the model with other mutagenicity predicting tools showed that local models, specifically trained and conse- quently particularly suitable for predicting specific chemical classes (e.g., azo compounds) are often a solution to difficulties with global models which, although useful in predicting a wide range of chemicals, may fail in predicting certain classes with peculiar structural features. Besides the better predictions, this approach is attractive because it offers explanations for the results, in terms of chemical and biochemical mechanisms.

Competing interests

AM is employed by Kode srl, who provide chemoinformatics consultancy and software products. Dr. Gadaleta, Dr. Benfenati, Dr.
Porta and Dr. Manganelli report grants from European Commission (EC), during the conduct of the study.

Acknowledgements

We acknowledge the PROSIL project (LIFE12 ENV/IT/154) for financial support. We acknowledge Masamitsu Honma of the Division of Genetics & Mutagenesis, National Institute of Health Sciences (1-18-1 Kamiyoga, Setagaya-Ku, Tokyo 158-8501, Japan) for kindly providing us experimental mutagenicity data.

References


