Toxicology 370 (2016) 20–30

Contents lists available at ScienceDirect

Toxicology

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

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

Domenico Gadaleta[a,*], Serena Manganelli[a], Alberto Manganaro[b], Nicola Porta[a], Emilio Benfenati[a]

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

ARTICLE INFO

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

Keywords:
Ames test
Aromatic amines
Azo dyes
Expert system
In silico

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 *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.

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 (Albreht 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,

Abbreviations: ADME, adsorption distribution metabolism excretion; AGG, amine-generating group; CDK, chemistry development kit; EDG, electron donating group; EPA, Environmental Protection Agency; EWG, electron-withdrawing group; FN, false negative; FP, false positive; GA-MLR, genetic algorithm-multiple linear regression; GLP, good laboratory practice; GTI, genotoxic impurity; HOMO, highest occupied molecular orbital; k-NN, k-Nearest Neighbor; LUMO, lowest unoccupied molecular orbital; MCC, Matthews correlation coefficient; QSAR, quantitative structure-activity relationship; ROS, reactive oxygen species; SA, Structural alert; SVM, Support Vector Machine; SMARTS, Smiles arbitrary target specification; SMILES, simplified molecular input line entry system; TN, true negative; TP, true positive.

* Corresponding author.
E-mail address: domenico.gadaleta@marionegri.it (D. Gadaleta).

http://dx.doi.org/10.1016/j.tox.2016.09.008
0300-483X/

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 (Zeiger 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 (Snyderwine 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 by 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 Grunwald (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 aminoimidazo-azaarenes (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