



# Journal of Environmental Science and Health, Part C

## Environmental Carcinogenesis and Ecotoxicology Reviews

ISSN: 1059-0501 (Print) 1532-4095 (Online) Journal homepage: <http://www.tandfonline.com/loi/lesc20>

## Integrating computational methods to predict mutagenicity of aromatic azo compounds

Domenico Gadaleta, Nicola Porta, Eleni Vrontaki, Serena Manganelli, Alberto Manganaro, Guido Sello, Masamitsu Honma & Emilio Benfenati

To cite this article: Domenico Gadaleta, Nicola Porta, Eleni Vrontaki, Serena Manganelli, Alberto Manganaro, Guido Sello, Masamitsu Honma & Emilio Benfenati (2017): Integrating computational methods to predict mutagenicity of aromatic azo compounds, Journal of Environmental Science and Health, Part C, DOI: [10.1080/10590501.2017.1391521](https://doi.org/10.1080/10590501.2017.1391521)

To link to this article: <https://doi.org/10.1080/10590501.2017.1391521>



Accepted author version posted online: 13 Oct 2017.  
Published online: 27 Nov 2017.



Submit your article to this journal [↗](#)



Article views: 26



View related articles [↗](#)



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at  
<http://www.tandfonline.com/action/journalInformation?journalCode=lesc20>



# Integrating computational methods to predict mutagenicity of aromatic azo compounds

Domenico Gadaleta<sup>a,\*</sup>, Nicola Porta<sup>a,\*</sup>, Eleni Vrontaki<sup>a,b</sup>, Serena Manganelli<sup>a</sup>, Alberto Manganaro<sup>c</sup>, Guido Sello<sup>d</sup>, Masamitsu Honma<sup>e</sup>, and Emilio Benfenati<sup>a</sup>

<sup>a</sup>Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy; <sup>b</sup>Laboratory of Organic Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Athens, Greece; <sup>c</sup>Kode s.r.l., Pisa, Italy; <sup>d</sup>Department of Chemistry, University of Milano, Milan, Italy; <sup>e</sup>Division of Genetics & Mutagenesis National Institute of Health Sciences, Setagaya-ku, Tokyo, Japan

## ABSTRACT

Azo dyes have several industrial uses. However, these azo dyes and their degradation products showed mutagenicity, inducing damage in environmental and human systems. Computational methods are proposed as cheap and rapid alternatives to predict the toxicity of azo dyes. A benchmark dataset of Ames data for 354 azo dyes was employed to develop three classification strategies using knowledge-based methods and docking simulations. Results were compared and integrated with three models from the literature, developing a series of consensus strategies. The good results confirm the usefulness of *in silico* methods as a support for experimental methods to predict the mutagenicity of azo compounds.


## KEYWORDS

Azo compounds; mutagenicity; (Q)SAR; docking; consensus model

## 1. Introduction

Azo dyes, the largest class of synthetic organic colorants, are widely used in the textile, plastics, paper, leather, cosmetics, pharmaceutical and food industries.<sup>1</sup> However, many azo dyes and their bacteria-degradation products (i.e., aromatic amines) showed mutagenicity effects, raising serious issues in environmental and human systems.<sup>2,3</sup> Most dyes are released into the environment as effluent in industrial wastewater,<sup>4</sup> and the mutagenic activity of surface waters and sediments was attributed to the presence of these compounds<sup>5,6</sup>. Moreover, mutagenic azo dyes have been found in house dust<sup>7</sup> and occupational exposure to azo compounds have been linked causally to several types of cancers in humans.<sup>8</sup>

Aromatic azo dyes have one or more azo groups in their structure. Azo bond breakage produces colorless aromatic amines, decolorizing the dye. This may occur

**CONTACT** Domenico Gadaleta  [domenico.gadaleta@marionegri.it](mailto:domenico.gadaleta@marionegri.it)  Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Via G. La Masa 19, 20156 Milano, Italy.

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/lesc](http://www.tandfonline.com/lesc).

\*These authors contributed equally to this work.

© 2017 Taylor & Francis Group, LLC

through different mechanisms, such as chemical, photochemical, or enzymatic reactions in prokaryotic and eukaryotic organisms.<sup>9</sup> Many bacteria, algae and fungi are able to degrade a wide range of azo compounds through reductive and oxidative enzymes. Some bacteria have the ability to degrade azo dyes both aerobically and anaerobically. In a subsequent step, aromatic amines are metabolically oxidized to reactive electrophilic species that covalently bind to DNA<sup>10</sup> and are further metabolized (through an  $S_N1$  mechanism) to genotoxic compounds by mammalian microsomal enzymes into the DNA reactive nitrenium ion.<sup>11,12</sup> The degree of reduction of the azo group depends on the electron density around the  $-N = N-$  bond. Therefore, the toxicity of aromatic azo compounds depends on their structural complexity, reactivity and aromatic substitution.<sup>13</sup> For example, nitro groups in the benzene ring increase the mutagenicity of azo dyes.<sup>14,15</sup>

In general, most researchers have focused on the toxicity of aromatic azo dyes, which have attracted much interest in the recent decades.<sup>16</sup> The standard *in vitro* test, i.e., the Ames test<sup>17</sup>, that is required to assess the mutagenicity of such chemicals, although effective for evaluating the toxicity of a few compounds, is not suitable for testing large amounts of chemicals (e.g., for prioritization), because of time and cost.

Recently, computational methods have been used to predict the mutagenic potential of azo dyes. Predictive models can be developed to predict the toxicity of azo compounds before their synthesis, based only on their chemical structure. (Quantitative) structure-activity relationship ((Q)SAR) methods are the most usual alternative for this purpose. To date, several *in silico* models have been developed to predict the mutagenicity of highly populated datasets of chemicals, with good predictive performances<sup>18-20</sup>.

The present study explored a range of strategies to model a family of chemical compounds that cannot be satisfactorily predicted with other general models.<sup>21,22</sup> Two new classification models addressing the mutagenic potential of azo compounds were presented: i) a fragment-based model (FBM) that simply considers structural matching of rule sets addressing toxicity; ii) a joined mechanistic model (JMM), i.e. an expert system that takes account of a broad range of factors (i.e., redictions of other models, molecular descriptors and structural fragments).

Molecular docking simulations on the azoreductase were also used for classification purposes. Azoreductase (EC 1.7.1.6) is a class of polymeric flavin-dependent NADH and polymeric flavin-dependent NADPH enzymes responsible for azo reduction by bacteria in the gut.<sup>23</sup> The first step in the biodegradation of azo dyes is the conversion to aromatic amines by azoreductase-catalyzed cleavage, with NADH and/or NADPH as electron donor. Azoreductases reduce azo compounds through a 'ping pong' mechanism.<sup>24</sup> As azoreductase is the main enzyme involved in azo bond reduction, its activity can be examined to understand the mutagenic process better.

Also provided is a comparison with three other existing models recently published by our group.<sup>21,25</sup> Finally, an overall consensus approach was developed that

proved useful for regulatory and scientific purposes and more suitable than single models to predict the mutagenic potential of aromatic azo compounds.

## 2. Material and methods

### 2.1. Dataset

The training set (TS) comprising 354 aromatic azo compounds described by Manganelli et al.<sup>25</sup> with related Ames mutagenicity data was employed to derive new mutagenicity models and for comparison. Recently new Ames mutagenicity data were made available by Japan's Health Ministry, within the Ames (Q)SAR project organized by the National Institute of Health Sciences Japan (<http://www.nihs.go.jp/dgm/amesqsar.html>). The Ames assays were conducted under GLP according to Industrial Safety and Health Act in Japan. We selected 33 azo compounds that were not included in the TS. These chemicals were used as validation set (VS) for evaluating the external predictivity of the models. The criteria already described by Manganelli et al.<sup>25</sup> were used for data curation and selection.

Both the TS and the VS were well-balanced in terms of the percentage of mutagenic (respectively 54.0% and 45.5% of the total for the TS and the VS) and non-mutagenic compounds. Istant JChem software (Istant JChem 15.11.16.0 ChemAxon, <http://www.chemaxon.com>) was used to handle chemical structures.

### 2.2. Model description

#### 2.2.1. Fragment Based Model (FBM)

The first model was a FBM consisting of a collection of hierarchical sets of fragments and sub-structures addressing mutagenicity and non-mutagenicity for azo compounds. The TS including 354 aromatic azo compounds encoded into SMILES notations (<http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>) was split into a sub-training set (sTS) of 280 compounds used for model derivation and a sub-validation set (sVS) of 74 compounds for model validation. For the split, TS chemicals were grouped in clusters by means of an in-house developed software. In a first step, clusters were defined in an iterative procedure to minimize the similarity values between molecules inside a cluster and maximize the similarity values between molecules of different clusters.<sup>26</sup> Second, the clusters were grouped into super-clusters, containing all clusters with an average similarity between their corresponding molecules higher than a given threshold. Then, 20% of compounds from each cluster were randomly assigned to the sVS while the remainder were assigned to the sTS, in order to obtain two structurally analogous datasets.

SARpy software<sup>27</sup> was used to sequentially extract from the sTS a series of collections of structural fragments addressing toxicity or lack of toxicity.<sup>28</sup> The software applies a fragmentation process to the input structure, searching for relationships between the fragments and the observed activity of the input molecules. Each fragment is encoded into SMARTS notations ([www.daylight.com/dayhtml/doc/](http://www.daylight.com/dayhtml/doc/)

[theory/theory.smarts.html](#)) and associated to an activity label (positive or negative) and a likelihood ratio, estimating the statistical relevance of the fragment. In the end, chemicals matching any of the extracted rules were classified according to the activity label.

Rules from a first SARpy run were not enough to predict all the compounds of the sTS. This dataset was therefore iteratively analysed by extracting new sets of rules from the chemicals that remained unpredicted from the previous steps, until a reasonably small percentage of the initial dataset remained unpredicted. This led to the extraction of three different rule-sets, two containing both fragments for mutagenicity and for non-mutagenicity (Rulesets I and III), and one including only fragments for mutagenicity (Ruleset II). These rules were hierarchically applied to give predictions. A fourth ruleset was manually compiled using unpredicted compounds from the first three rulesets. These rules were defined by human expert evaluation and are supported by the literature. They can be grouped as: 1) nitro aromatic; 2) *p*-azo-aniline; 3) charged azo heterocycles; 4) biphenyl groups.

Since some chemicals contained fragments with different activities (both mutagenic and non-mutagenic) from Rulesets I and III, a mutagenicity score (MS) was established to assess these compounds. This score was associated with each fragment of these two rulesets and was calculated as follows:

$$\text{MS} = \text{class} \times \text{weight factor} \quad (1)$$

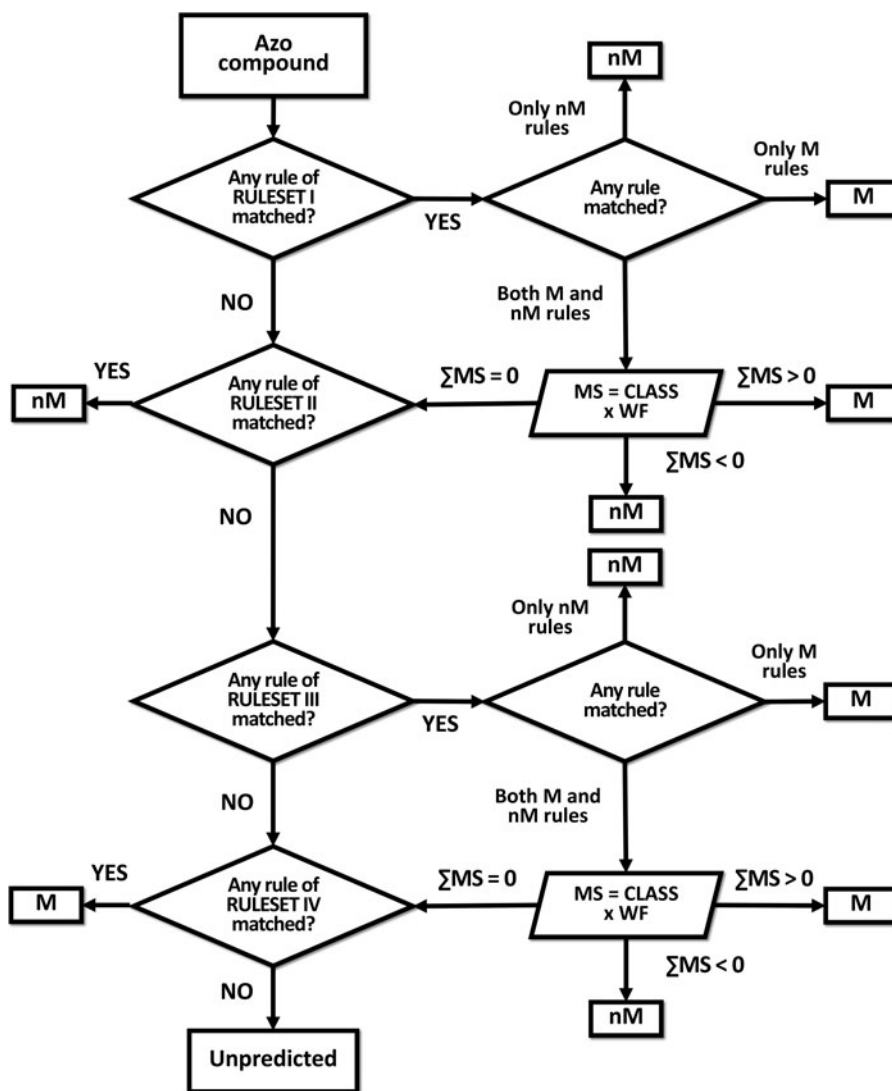
where class refers to the activity label (+1 for mutagenic, -1 for non-mutagenic), and the weight factor is a number equal to 1 or 2 depending on the likelihood ratio for each fragment (see Supporting Information for further details). Each compound was classified on the basis of the sum of the MSs for all the fragments matched; an overall positive MS implies mutagenic predictions, while an overall negative MS implies non-mutagenic predictions. Compounds with a score of zero were labeled as unpredicted and processed by the following ruleset.

The list of fragments in the four rulesets along with the MS and the settings applied to compute them is included in the Supplementary Materials, Tables S1-S4. [Figure 1](#) shows the entire prediction workflow.

### 2.2.2. *Joined Mechanistic Model (JMM)*

The second model is an expert system for classifying aromatic azo compounds on the basis of their mutagenic potential. The model is built as a hierarchical system ([Figure 2](#)) consisting of three filters, each running a test on chemicals based on a series of selected structural features and properties. From the outcome of this test, the algorithm moves to the next filter, where a new test is done, or to a final mutagenicity class (mutagenic, non-mutagenic) to which the query is assigned.

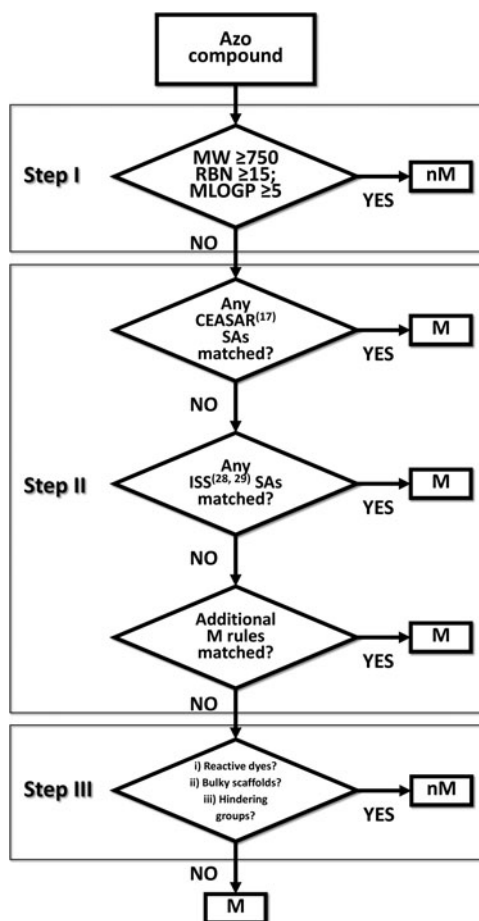
In the first part, the algorithm evaluates a series of molecular descriptors in order to identify compounds that were less likely to be mutagenic due to their unfavorable pharmacokinetic properties. Chemicals exceeding thresholds associated with at least two of those descriptors were classified as non-mutagenic, otherwise they were



**Figure 1.** Fragment-based model (FBM) for the classification of mutagenic and non-mutagenic aromatic azo compounds. M = mutagenic; nM = non-mutagenic; MS = mutagenicity score; WF = weight factor.

processed in the next filtering step. Selected descriptors and their respective thresholds were: 1) Molecular weight (MW), maximum 750 Da; 2) number of rotatable bonds (RBN), maximum 15; 3) Moriguchi's logP (MLOGP), maximum 5. Molecular descriptors were calculated with Dragon 7.<sup>29</sup>

In the second part, the algorithm matches query chemicals with a series of structural fragments addressing mutagenicity. Fragments were collected among those already included in two mutagenicity predicting models implemented within the VEGA platform (<http://www.vega-qsar.eu/>), i.e. the CAESAR model<sup>18</sup> and the ISS model that used Benigni and Bossa rules for mutagenicity.<sup>30,31</sup> SA29 flagging for aromatic azo compounds was not considered in the prediction process. Three



**Figure 2.** Joint mechanistic model (JMM) for the classification of mutagenic and non-mutagenic aromatic azo compounds. M = mutagenic; nM = non-mutagenic; MW = molecular weight; RBN = number of rotatable bonds; MLOGP = Moriguchi LogP.

fragments flagging for mutagenicity were also manually compiled and included with those described above (Table S5 of Supporting Information).

Chemicals matching at least one of the mutagenicity-flagging fragments were classified as mutagenic, otherwise they are elaborated by the last filtering step, that checks the occurrence of a series of structural rules flagging for lack of mutagenicity. These fragments were manually defined on the basis of the literature and their mechanistic rationale and are grouped into three classes: 1) reactive dyes, 2) sulfonated scaffolds, and 3) hindering groups (Table S5 of Supporting Information).

### 2.2.3. Docking-based classification

For molecular docking simulations, SMILES strings related to compounds of the TS were converted to 3D coordinates using LigPrep 3.4<sup>32</sup> available in Schrödinger Suite 2015. Hydrogen atoms were added, stereoisomers and ionization states were generated (at pH 7.4) and the molecules were geometrically refined. All the ligands were further minimized by the MacroModel 10.8<sup>31</sup> software available in

Schrödinger Suite 2015. A standard molecular mechanics energy function force field (OPLS\_2005)<sup>34</sup> and the Polak-Ribiere conjugated gradient method (500 iterations with gradient  $0.01 \text{ kcal mol}^{-1} \cdot \text{Å}^{-1}$ ) were used. Water was used as solvent.

The crystal structure of a bacterial azoreductase (AzrC) in complex with the azo dye Acid Red 88 (RE88) was retrieved from the Protein Data Bank (PDB code: 3W7A). This homodimeric protein of about 48 kDa and 211 residues was isolated from the mesophilic gram-positive bacillus sp. B29.<sup>35</sup> Each monomer forms a flavodoxin-like  $\alpha/\beta$ -fold structure and binds a molecule of FMN at the active site between the monomers of a dimer. To date this is the X-ray azoreductase structure co-crystallized with an azo compound with the highest resolution available (2.10 Å) and therefore it was chosen for docking experiments. The structure was prepared using the Protein Preparation Wizard implementation in Schrödinger Suite 2015.<sup>36</sup> Bond orders and disulfide bonds were assigned and hydrogen atoms were added. All the water molecules, the ions and the C, D chains were deleted and RE88 was kept with FMN, as default ligands. The Epik 2.4 implementation was used to predict ionization and tautomeric states of the ligands.<sup>37</sup> The hydrogen-bonding network was optimized by reorienting Asn and Gln residues, and selecting appropriate states and orientations of His residues. Finally, the protein-ligand complex was refined through a series of restrained minimizations, using the OPLS\_2005 force field and the default threshold of 0.3 Å for root mean square deviation (RMSD).

The GLIDE program implemented in Schrödinger was used for the molecular docking procedure.<sup>38</sup> The binding region was defined by a  $10 \text{ Å} \times 10 \text{ Å} \times 10 \text{ Å}$  box centered in the centroid of the ligand co-crystallized into the active site of the enzyme, and the option “dock ligands with length  $\leq 20 \text{ Å}$ ” was chosen. All other parameters were kept as default. The standard precision (SP) mode and the GlideScore scoring function were used for the flexible ligand-docking and scoring, respectively. Only the first ranked pose was retained for each compound. Considering the distribution of scores, ranging from highly negative (e.g.,  $-12.4 \text{ kcal/mol}$ ) to zeros or *near-to-zero* values, we applied a cutoff of  $-5 \text{ kcal/mol}$  to separate putative azoreductase binders from non-binders. Therefore, we derived a classification model taking account of the scores from the docking simulations. Predictions were made assuming that if an azo compound could bind the azoreductase enzyme, it is potentially mutagenic.

### **2.3. Mutagenicity models for azo compounds from the literature**

Three mutagenicity models specifically tailored for azo compounds were recently published by our group.<sup>21,25</sup> Here they are briefly summarized. These three models, as well as those here presented and described, were validated on a new VS of 33 azo compounds and integrated into a series of consensus strategies (Table 1).

The CORAL model<sup>25</sup> is a (Q)SAR built using the standalone application software CORAL (CORrelations And Logic, <http://www.insilico.eu/coral>). It uses SMILES-based descriptors, calculated from structural attributes, which can be local (e.g., pairs of connected atoms, etc.) or global (e.g., co-occurrence of two elements, etc.). The correlation weights (i.e., contributions) for these attributes are computed with



the Monte Carlo optimization method, to define which ones are optimal for predicting the target property/endpoint (with the splits of the initial TS into training, calibration and validation set).

The k-NN model<sup>25</sup> is based on the *k*-nearest neighbors algorithm. The prediction for a query molecule is a weighted consensus of the experimental values of its four most similar molecules from the TS. The model uses the similarity index as implemented within the VEGA platform.<sup>21</sup> Molecules with a similarity index lower than 0.75 are excluded. If only one similar molecule is found, it was used for prediction only if the similarity was 0.85 or higher, otherwise no prediction is provided.

The Knowledge-Based Expert System (KBES)<sup>21</sup> was developed for the classification of aromatic amines on the basis of their mutagenic (Ames) potential. It has been also proved to be suitable for other compounds whose mutagenicity is related to the release of aromatic amines, such as aromatic azo compounds, assuming that all the azo bonds are cleaved. The system is a decision algorithm, developed as a KNIME workflow, implementing filters evaluating a series of molecular descriptors, scaffolds and structural rules (e.g., amine-generating groups etc.). In a conservative approach, if at least one of the amines produced is predicted as mutagenic, then also the parent compound too is flagged as mutagenic.

#### 2.4. Integrating mutagenicity models for azo compounds

A series of three integration strategies were evaluated by combining the six models described above (Table 1) for predicting the mutagenicity of aromatic azo compounds.

- 1) All six models are considered, including the docking-based approach. In this case, models are applied in parallel, integrating the independent predictions of each one. A final prediction for a chemical is produced only when four, five or all six of the models returned concordant predictions, resulting in three integration schemes. Chemicals with ambiguous predictions are considered unpredictable.
- 2) The final prediction for a given chemical is obtained excluding the docking based model and based on the other five ones. Three integration schemes are applied, as above, returning predictions only for chemicals with at least three, four or all five concordant predictions.

**Table 1.** Mutagenicity models for aromatic azo compounds in the present study.

Name of the model	Abbreviation	Training set (number of compounds)	Reference
Fragment-Based Model	FBM	sTS (280)	This paper
Joined Mechanistic Model	JMM	TS (354)	This paper
Docking-based classification		TS (354)	This paper
CORAL	CORAL	TS (354)	(23)
k-NN	k-NN	TS (354)	(23)
Knowledge-based Expert System	KBES	Amine dataset	(20)

- 3) The docking-based strategy is used as a preliminary filter. If the chemical is classified as non-binder by the docking procedure, it is labelled as not mutagenic, otherwise the consensus of the other models is used.

## 2.5. Evaluation of classification performance

The performance of the classification models was evaluated using Cooper's parameters<sup>39</sup> which are accuracy, sensitivity and specificity. These account for the number of correctly classified mutagenic (true positives, TPs) and non-mutagenic compounds (true negatives, TNs) and the number of misclassified mutagenic (false positives, FPs) and non-mutagenic ones (false negatives, FNs). Accuracy provides a measure of the total errors (Equation 2). Models with high sensitivity (Equation 3) have fewer FNs while those with high specificity (Equation 4) have fewer FPs.

$$\text{Accuracy} = \frac{TP + TN}{\text{Total number of compounds}} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

The Matthews Correlation Coefficient (MCC) was also used. MCC indicates the quality of a binary classification and is generally regarded as a balanced measure, which can be used even for classes of very different sizes. MCC ranges between  $-1$  and  $+1$ . A value of  $+1$  means a perfect classification,  $0$  no better than random and  $-1$  indicates complete misclassification<sup>40</sup> (Equation 5).

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

## 3. Results

Tables 2 and 3 compare the performance of the three strategies presented (FBM and JMM, docking-based) and of the three literature models (CORAL, k-NN and KBES) in predicting the mutagenic potential of the TS and VS chemicals. As for the FBM,

**Table 2.** Classification performance of the six mutagenicity predicting models for the TS (354 compounds).

	FBM	JMM	Docking	CORAL	k-NN	KBES
TP	145	165	167	162	158	172
FN	34	26	24	29	33	19
FP	15	49	112	24	41	60
TN	143	114	51	139	122	103
Compounds predicted	337	354	354	354	354	354
Accuracy	0.85	0.79	0.62	0.85	0.79	0.78
Specificity	0.91	0.70	0.31	0.85	0.75	0.63
Sensitivity	0.81	0.86	0.87	0.85	0.83	0.90
MCC	0.71	0.57	0.23	0.70	0.58	0.56
Coverage	0.95	1.00	1.00	1.00	1.00	1.00

**Table 3.** Classification performance of the six mutagenicity predicting models for the VS (33 compounds).

	FBM	JMM	Docking	CORAL	k-NN	KBES
TP	13	14	13	14	13	14
FN	2	1	2	1	2	1
FP	4	8	12	5	5	10
TN	14	10	6	13	12	8
Compounds predicted	33	33	33	33	32	33
Accuracy	0.82	0.73	0.58	0.82	0.78	0.67
Specificity	0.78	0.56	0.33	0.72	0.71	0.44
Sensitivity	0.87	0.93	0.87	0.93	0.87	0.93
MCC	0.64	0.52	0.23	0.66	0.58	0.42
Coverage	1.00	1.00	1.00	1.00	0.97	1.00

data used for model derivation were a subset of the entire TS (280 out of 354 compounds) (see section 2.2.1) and therefore the statistics reported in Table 2 also refer to external data used for model validation (i.e., sVS). The statistical performance of these subsets is given in Supplementary material, Table S6.

The prediction performance of the six models was solid, reaching in most cases accuracy of about 0.80 on TS and about 0.70 on the new VS, confirming their suitability for real-life purposes. The strategy giving the worst results was the docking-based method, but this was not unexpected. Indeed, the aim of this approach was not to provide an independent predicting method, but a preliminary filtering step to be applied as support for other predicting models. Azo compounds rejected by the docking procedure (i.e., non-binders) were labelled as non-mutagenic, because they cannot be converted to potentially dangerous aromatic amine metabolites. On the other hand, the successful interaction between an azo compound and the enzyme was considered a potential warning that would require further analysis of suspicious molecules using different models.

With this assumption, we integrated the results of the different models. Table 4 and Table 5 show the results with the three different integration strategies on TS and

**Table 4.** Classification performance of the three integration strategies for the TS (354 compounds). Statistics refer to three separate integration schemes and to a different number of concordant predictions among models needed to return consensus predictions. Chemicals not having the required number of concordant predictions were labelled as unpredicted.

	(a)			(b)			(c)		
	6/6	5/6	4/6	5/5	4/5	3/5	5/5	4/5	3/5
TP	117	146	168	123	154	154	117	140	140
FN	4	13	18	16	21	37	34	37	51
FP	7	18	27	7	21	21	7	18	18
TN	31	92	117	112	134	142	122	139	145
Compounds predicted	159	269	330	258	330	354	280	334	354
Accuracy	0.93	0.88	0.86	0.91	0.87	0.84	0.85	0.84	0.81
Specificity	0.82	0.84	0.81	0.94	0.86	0.87	0.95	0.89	0.89
Sensitivity	0.97	0.92	0.90	0.88	0.88	0.81	0.77	0.79	0.73
MCC	0.81	0.76	0.72	0.82	0.74	0.68	0.72	0.68	0.62
Coverage	0.45	0.76	0.93	0.73	0.93	1.00	0.79	0.94	1.00

(a) All six models were integrated in parallel

(b) The docking-based method was not considered and the other five models were integrated

(c) The docking strategy was applied as preliminary filter, then scheme (b) was used for chemicals classified as positive by docking

**Table 5.** Classification performance of the three integration strategies for the VS (33 compounds). Statistics refers to three separate integration schemes and to a different number of concordant predictions among models needed to return consensus predictions. Chemicals not having the required number of concordant predictions were labelled as unpredicted.

	(a)		(b)			(c)			
TP	6/6	5/6	4/6	5/5	4/5	3/5	5/5	4/5	3/5
FN	10	11	15	11	12	12	10	10	10
FP	0	0	0	0	0	3	2	2	5
TN	1	3	4	1	3	3	1	3	3
Compounds predicted	1	7	10	9	13	15	11	14	15
Accuracy	12	21	29	21	28	33	24	29	33
Specificity	0.92	0.86	0.86	0.95	0.89	0.82	0.88	0.83	0.76
Sensitivity	0.50	0.70	0.71	0.90	0.81	0.83	0.92	0.82	0.83
Sensitivity	1.00	1.00	1.00	1.00	1.00	0.80	0.83	0.83	0.67
MCC	0.67	0.74	0.75	0.91	0.81	0.63	0.75	0.65	0.51
Coverage	0.36	0.64	0.88	0.64	0.85	1.00	0.73	0.88	1.00

(a) All six models were integrated in parallel

(b) The docking-based method was not considered and the other five models were integrated

(c) The docking strategy was applied as preliminary filter, then scheme (b) was used for chemicals classified as positive by docking

VS. The integration schemes, described in paragraph 2.4, are: 1) the combination of the six models; 2) the combination of five models excluding the docking-based one; 3) the application of the docking-based method as filter, followed by integration of the five remaining models.

In general, statistics on both TS and VS confirm the effectiveness of integrating different models for mutagenicity prediction. Indeed, the MCC of some of the integration schemes improved those referred to the best single models (i.e., best values of respectively 0.70 and 0.66 for CORAL model on TS and VS), keeping acceptable coverage (i.e., the percentage of predicted compounds).

Unexpectedly, application of the docking-based method as filter never gave any improvement in performance on the TS compared to simply integrating of the remaining five models. This was more evident on the VS, where MCC reached best values greater than 0.80 in the case of the latter strategy, but never exceeded 0.75 with the former strategy. In the same way, using docking as a stand-alone method together with other models did not improve the performance given by completely excluding the docking-based method.

#### 4. Discussion

We present here three new methods for predicting mutagenicity of azo dyes. This study aimed to explore a range of strategies to handle a family of chemicals that cannot be satisfactorily predicted with global models (i.e., models not tailored for a specific chemical class).<sup>21,22</sup> The three models here presented (FBM, JMM and the docking-based method) and the others previously developed by our group (CORAL, k-NN, KBES) implemented orthogonal modelling methods, each having different advantages and issues. The models were compared in order to identify the

best single approaches and, more importantly, they were integrated following different consensus methods aimed at overcoming problems related to the single models. The newly derived methods are discussed more in-depth below, with particular attention to the results of their integration with existing mutagenicity models.

#### 4.1. FBM

FBM was derived by SARpy software, which is a data mining tool for extracting rules. SARpy has already been used to derive sets of rules to predict mutagenicity and other endpoints.<sup>27</sup> The model we present is the first attempt to apply this tool to a dataset focused on a single class of chemicals. SARpy was used for extracting three of the four rulesets used for mutagenicity predictions. However, because SARpy proved not suitable for extracting rules from small sets of chemicals, ruleset IV was manually extracted from the last 25 unpredicted compounds within the sTS (see section 2.2.1 and Table S4 of Supporting Information). These rules, addressing the mutagenicity of chemicals, were supported by literature findings and human expert judgement, and can be grouped in four categories. 1) Nitro aromatic and hydroxy-amino groups: these led to the release of nitrogen species highly reactive towards DNA after metabolic activation.<sup>21,41</sup> 2) *p*-amino aromatic azo compounds: the *p*-aniline species released after azo bond cleavage can undergo an amino-imino tautomeric shift giving an imino-quinone species that can form adducts with DNA and proteins.<sup>42</sup> 3) Positively charged azo heterocycles (e.g., imidazoles and thiazoles): this species can stabilize the charge by electronic delocalization within the aromatic system and act as electron-withdrawing groups, facilitating the cleavage of the azo group and the generation of potential aromatic amines. 4) Biphenyl groups: aromatic amines generated by azo compounds with more than one ring forming conjugated systems showed greater mutagenic activity than single-ring species.<sup>43,44</sup>

FBM showed the highest MCC on the TS (0.71), and one of the highest (0.64) on the VS, after the CORAL model. FBM also had the highest specificity, both on the TS (0.91) and the VS (0.78). This means that gave a relatively small number of FP predictions, compared with the other models. This can be explained by the fact that the majority of rules implemented in FBM were extracted by an automatic tool, to identify fragments related to mutagenicity and non-mutagenicity. On the other hand, some of the other models (e.g., JMM and KBES) were based on rules defined by human expert evaluation that can sometimes be “biased” towards the identification of rules for toxicity.

#### 4.2. JMM

JMM implemented in a hierarchical structure a series of pharmacokinetic related rules, and fragments addressing mutagenicity (retrieved from other models) and lack of mutagenicity (supported by human expert knowledge and literature findings). The pharmacokinetic filter is conceptually similar to other well-known

examples reported in literature<sup>45,46</sup> and aims to identifying compounds likely to be non-mutagenic due to their unfavorable toxicokinetic properties, which prevent them reaching the site(s) of action (e.g., DNA inside the nucleus).

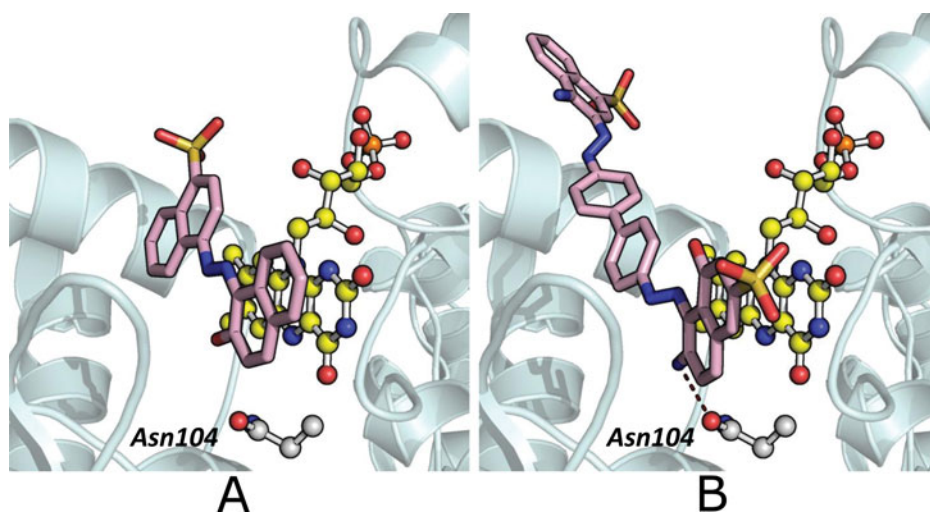
Fragments addressing lack of toxicity were manually compiled and can be grouped in four different classes. 1) Reactive dyes are azo compounds characterized by particular moieties (e.g., chlorinated triazines and vinylsulfones) that can react with fibres or proteins, forming covalent bonds. This property has been exploited in several fields, from textile and paper cellulose fibres dyeing to cell staining (reaction with cell proteins).<sup>47</sup> These moieties may also hypothetically react with the medium of the Ames test or with cell membrane proteins, resulting in possible negative experimental outcomes. 2) Bulky groups, (e.g., naphthalene-azo-naphthalene) can hypothetically reduce the passage through the cell membrane and/or hamper the interaction with the enzymatic environment due to steric hindrance. Consequently, it can hamper the azo bond cleavage, leading to non-toxicity. 3) The presence of polar/chelating groups in *ortho*- position to the azo bonds may shield the bond from reduction by directly interacting with it, or by forming complexes after incorporation of metal ions.<sup>48,49</sup>

### 4.3. Docking-based classification

GLIDE has shown good reproducibility of the co-crystal ligand conformation, which has also been recommended in the literature for its accuracy in molecular docking and scoring, considering disparate target proteins.<sup>50</sup>

The crystal structure of the azoreductase consists of two homodimeric proteins (A, B, C and D chains) with a molecular weight of about 96 kDa. For docking simulations, only one homodimer was used. The active site, and therefore the binding pocket of the azoreductase contained a non-covalently bound FMN and a co-crystallized ligand in each subunit, and was characterized by a hydrophobic environment with a depth of 3.3 Å.<sup>35</sup> Figure 3a shows the binding mode of co-crystallized ligand, RE88, which established pi-pi ( $\pi$ ) stacking interactions with the FMN. Single and double rings of the compounds mimicked the nicotinamide adenine of NAD(P)H, in order to be docked in the pocket.

Docking calculations were made on the aromatic azo compounds in the active site of azoreductase and their docking results, expressed as GlideScore, ranged from  $-12.433$  kcal/mol to very low negative values. The lower of the docking score, the larger the expected interaction with the azoreductase. As was expected, the bulkier compounds of the data set (MW >1000) were docked into the enzyme with a low ratio; 71 compounds were not docked, and four of the docked compounds gave too low scores (according to an arbitrary cut-off of  $-5.000$  kcal/mol). These compounds were considered non-mutagens. The most frequent interactions observed in the ligand-protein complexes with highest scores were the pi-pi ( $\pi$ ) stacking with FMN, hydrophobic interactions and hydrogen bonds with nearest residues of chains A and B (mainly with Asn104 (A)) (Figure 3b). The ligands attempted to mimic the



**Figure 3.** Binding mode of (A) the co-crystallized ligand RE88 and (B) an example of a docked mutagenic TS compound (SlideScore: -9.872) in the azoreductase active site (PDB code: 3W7A). Relevant amino acid residues and NAD(P)H cofactor are represented as balls-and-sticks and azo compounds as sticks, coloured according to the atom code (C atoms in pink, white and yellow for inhibitors, amino acid residues and cofactor, respectively).

binding mode of a co-crystal structure, while the very large compounds adopted a more closed conformation in order to fit into the binding pocket.

The docking-based strategy was applied to simulate the behaviour of the azoreductase enzyme. Chemicals not able to dock onto the enzyme might be considered safe, because they cannot be converted to potentially dangerous aromatic amine metabolites. On the other hand, the successful interaction between an azo compound and the enzyme, leading to the production of aromatic amine metabolites, was not sufficient to flag the parental compound as toxic, because its metabolites may or not be mutagens on the basis of their structure and chemical environment. These considerations justified the large number of FPs produced by the method, as documented by the low specificity for both TS (0.31) and VS (0.33). The correct assumption would be to use the docking-based method to identify those chemicals that can potentially generate toxic aromatic amine metabolites, then for suspicious compounds, to apply other models specifically tailored to predict the mutagenicity of aromatic azo compounds. This is what we tested in (c) (Tables 4 and 5). There may be other causes of mutagenicity within the molecule not related to cleavage of the azo bond and in fact this approach generated some FNs (Tables 2 and 3), but only few.

The results of the CORAL, k-NN and KBES models have been discussed in previous papers, and will be not addressed here. As already explained, CORAL proved to be the best model in terms of MCC, while the KBES gave a number of FPs, since it is based on the assumption that the azo bonds are always cleaved and all possible aromatic amines are generated.

#### 4.4. Integration of the results of the six models

One of the main conclusions on this study is that the use of single models may often fails in reaching high predictivity. However, combining more predictions from different tools for a given chemical in a weight-of-evidence approach enabled in many cases to positively revaluing the role of *in silico* methods and increasing their relevance for real-life applications.

Various strategies are proposed for integrating the results of (Q)SAR models.<sup>51</sup> Here we explored a series of straightforward majority vote approaches, as described in paragraph 2.4. As shown in Tables 4 and 5, the classification performance was best when excluding the docking-based strategy, with MCC of 0.82 on the TS and 0.91 on the VS. These were higher than the results from any single model. This demonstrated once more that the consensus strategy in many cases improves the results from single models, particularly when models based on different strategies are integrated.<sup>52–53</sup> These values referred to complete agreement between all five (Q)SAR models. In this regard, it was to be expected that a lower, but still acceptable percentage of compounds was predicted, ranging from 0.64 for the VS to 0.73 for the TS. The application of a majority vote scheme accepting predictions that were concordant for at least four out of five models increased the coverage to 0.93 for the TS and 0.85 for the VS. In this case, the MCC (0.74 for TS and 0.81 for VS) were still better than those for the best single models (Tables 2 and 3).

When predictions concordant for at least three out of five models were considered, the MCC were still good, from 0.68 for the TS to 0.63 for the VS, in both cases with full coverage. However, this last strategy did not give any improvement in performance over the best single models.

The inclusion of the docking-based strategy did not improve the results in terms of coverage and predictive performance, compared to complete exclusion of the docking model. The same was true for the sequential strategy (Tables 4 and 5).

Observing the results of the different strategies, we can conclude that integration of the five (Q)SAR models (FBM, JMM, CORAL, k-NN and KBES) with the exclusion of the docking-based strategy proved adequate for predicting the mutagenicity of aromatic azo compounds. The five models, when integrated, were already able to cover the mechanistic information provided by the docking, i.e. the first reduction step involved in the overall mutagenicity process of azo compounds (also coping with the issues of structural moieties hindering the azo-bond cleavage). It can also be concluded that, for this specific case, 3-D techniques do not improve the results given by simpler, straightforward 2-D methods.<sup>55</sup>

## Conclusions

We used a dataset of 354 azo compounds to develop three new computational tools (FBM, JMM and docking-based) to predict the mutagenicity of azo dyes. The three methods were compared and integrated with three other models (CORAL, k-NN and KBES) described in the literature, to provide more reliable predicting tools.



Statistics were good in terms of both Cooper parameters and MCC, the latter in the range of 0.53 – 0.69, depending on the integration strategy. This confirms that the integration of these models offers a more robust method for assessing the mutagenicity of this class of chemicals than single models. This study also confirmed that the integration of multiple strategies and a weigh-of-evidence approach might overcome limitations inherent in single models.<sup>22,56,57</sup> The models presented and the integrated approaches described will be implemented in the freely available VEGA platform (<http://www.vega-qsar.eu/>).

## Acknowledgements

This work was supported by the LIFE Programme, for the project: PROSIL (LIFE12 ENV/IT/000154).

## References

1. Raffi F, Hall JD, Cerniglia CE. Mutagenicity of azo dyes used in foods, drugs and cosmetics before and after reduction by *Clostridium* species from the human intestinal tract. *Food Chem. Toxicol.* 1997;35:897–901.
2. Tsuboy MS, Angeli JPF, Mantovani MS, Knasmüllerc S, Umbuzeirod GA, Ribeiro LR. Genotoxic, mutagenic and cytotoxic effects of the commercial dye CI disperse blue 291 in the human hepatic cell line HepG2. *Toxicology In Vitro* 2007;21:1650–1655.
3. Chung KT. Azo dyes and human health: A review. *Journal of Environmental Science and Health, Part C.* 2016;34:233–261.
4. Saratale RG, Saratale GD, Kalyani DC, Chang JS, Govindwar SP. Enhanced decolorization and biodegradation of textile azo dye Scarlet R by using developed microbial consortium-GR. *Bioresource Technology.* 2009;100:2493–2500.
5. Oliveira GAR, Ferraz ERA, Chequer FMD, Grando MD, Angeli JPF, Tsuboy MS, Marcarini JC, Mantovani MS, Osugi ME, Lizier TM, Zanoni MVB, Oliveira DP. (2010). Chlorination treatment of aqueous samples reduces, but does not eliminate, the mutagenic effect of the azo dyes Disperse Red 1, Disperse Red 13 and Disperse Orange 1. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis.* 2010;703(2):200–208.
6. Umbuzeiro G, Freeman HS, Warren SH, De Oliveira DP, Terao Y, Watanabe T, Claxton LD. The contribution of azo dyes to the mutagenic activity of the Cristais River. *Chemosphere.* 2005;60(1):55–64.
7. Peng H, Saunders DMV, Sun J, Jones PD, Wong CKC, Liu H, Giesy JP. Mutagenic azo dyes, rather than flame retardants, are the predominant brominated compounds in house dust. *Environmental Science and Technology.* 2016;50:12669–12677.
8. Golka K, Kopps S, Myslak ZW. Carcinogenicity of azo colorants: influence of solubility and bioavailability. *Toxicology Letters.* 2004;151:203–210.
9. Nakayama T, Kimura T, Kodama M, Nagata C. Generation of hydrogen peroxide and superoxide anion from active metabolites of naphthylamines and amino azo dyes: its possible role in carcinogenesis. *Carcinogenesis.* 1983;4:765–769.
10. Brown MA, De Vito SC. Predicting azo dye toxicity. *Crit. Rev. Env. Sci. Technol.* 1993;23: 249–324.
11. Møller P, Wallin H. Genotoxic hazards of azo pigments and other colorants related to 1-phenylazo-2-hydroxynaphthalene. *Mutation Research/Reviews in Mutation Research.* 2000;462:13–30.

12. Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J, Harriman SP. A comprehensive listing of bioactivation pathways of organic functional groups. *Current Drug Metabolism*. 2005;6:161–225.
13. Maran U, Sild S. QSAR modeling of genotoxicity on noncongeneric sets of organic compounds. *Artificial Intelligence Review*. 2003;20:13–38.
14. Brown JP, Roehm GW, Brown RJ. Mutagenicity testing of certified food colours and related azo, xanthene and triphenylmethane dyes with the Salmonella/microsome system. *Mutation Research*. 1978;56:249–271.
15. Nestmann ER, Kowbel DJ, Wheat JA. Mutagenicity in Salmonella of dyes used by defence personnel for detection of liquid chemical warfare agents. *Carcinogenesis*. 1981;2:879–883.
16. Liu ZJ, Chen H, Shaw N, Hopper SL, Chen L, Chen S, Wang BC. Crystal structure of an aerobic FMN-dependent azoreductase (AzoA) from *Enterococcus faecalis*. *Archives of Biochemistry and Biophysics*. 2007;463:68–77.
17. Ames BN, McCann J, Yamasaki E. Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. *Mutation Research/Environmental Mutagenesis and Related Subjects* 1975;31:347–363.
18. Ferrari T, Gini G. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. *Chemistry Central Journal*. 2010;4 (Suppl 1):S2.
19. Benigni R, Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutation Research/Reviews in Mutation Research*. 2008;659:248–261.
20. Liu R, Wallqvist A. Merging applicability domains for *in silico* assessment of chemical mutagenicity. *Journal of Chemical Information and Modeling*. 2014;54:793–800.
21. Gadaleta D, Manganelli S, Manganaro A, Porta N, Benfenati E. A knowledge-based expert rule system for predicting mutagenicity (Ames test) of aromatic amines and azo compounds. *Toxicology*. 2016;370:20–30.
22. Kulkarni SA, Barton-Maclaren TS. Performance of (Q)SAR models for predicting Ames mutagenicity of aryl azo and benzidine based compounds. *Journal of Environmental Science and Health, Part C*. 2014;32:46–82.
23. Chen H. Recent advances in azo dye degrading enzyme research. *Current Protein & Peptide Science*. 2006;7:101–111.
24. Chan-Ju W, Nicola L, Areej A. Role of tyrosine 131 in the active site of pAzoR1, an azoreductase with specificity for the inflammatory bowel disease prodrug balsalazide. *Acta Crystallographica*. 2010;66(Pt 1):2–7.
25. Manganelli S, Benfenati E, Manganaro E, Kulkarni S, Barton-Maclaren TS, and Honma M. New quantitative structure–activity relationship models improve predictability of Ames mutagenicity for aromatic azo compounds. *Toxicological Sciences*. 2016;153(2): 316–326.
26. Floris M, Manganaro A, Nicolotti O, Medda R, Mangiatordi GF, Benfenati E. A generalizable definition of chemical similarity for read-across. *Journal of Cheminformatics*. 2014;6(1):1–7.
27. Ferrari T, Cattaneo D, Gini G, Golbamaki Bakhtyari N, Manganaro A, Benfenati E. Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. *SAR & QSAR in Environmental Research*. 2013;24:365–383.
28. Lombardo A, Pizzo F, Benfenati E, Ferrari T, Gini G. A new *in silico* classification model for ready biodegradability, based on molecular fragments. *Chemosphere*. 2014;108:10–6.
29. Kode srl. Dragon version 7.0.4, 2016, <https://chm.kode-solutions.net>
30. Benigni R, Bossa C. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chemical Reviews*. 2011;111:2507–2536.
31. Benigni R, Bossa C, Tcheremenskaia O. In vitro cell transformation assays for an integrated, alternative assessment of carcinogenicity: a databased analysis. *Mutagenesis* 2013;28: 107–16.
32. *LigPrep, version 3.4*. New York, NY: Schrödinger, LLC; 2015

33. *MacroModel, version 10.8*. New York, NY: Schrödinger; LLC; 2015
34. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*. 2001;105:6474–6487.
35. Yu J, Ogata D, Gai Z, Taguchi S, Tanaka I, Ooi T, Yao M. Structures of AzrA and of AzrC complexed with substrate or inhibitor: insight into substrate specificity and catalytic mechanism. *Acta Crystallographica Section D Biological Crystallography*. 2014;70(Pt 2): 553–564.
36. *Protein Preparation Wizard 2015-2; Epik version 2.4*. New York, NY: Schrödinger, LLC; 2015  
*Impact version 5.9*. New York, NY: Schrödinger, LLC; 2015  
*Prime version 3.2*, New York, NY: Schrödinger, LLC; 2015
37. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design*. 2007;21:681–691.
38. *Glide, version 6.7*, New York, NY: Schrödinger, LLC; 2015
39. Cooper JA, Saracci R, Cole P. Describing the validity of carcinogen screening tests. *British Journal of Cancer*. 1979;39:87–89.
40. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*. 1975;405:442–451.
41. Benigni R, Bossa C, Netzeva T, Rodomonte A, Tsakovska I. Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environmental and Molecular Mutagenesis*. 2007;48:754–771.
42. Smith, GF, 2011. Designing drugs to avoid toxicity. *Prog. Med. Chem.* 50, 1–47.
43. Hooberman BH, Brezzell MD, Das SK, You Z, Sinsheimer JE. Substituent effects on the genotoxicity of 4-nitrostilbene derivatives. *Mutation Research/Genetic Toxicology*. 1994;341: 57–69.
44. Lai DY, Woo YT, Argus MF, Arcos JC. Cancer risk reduction through mechanism-based molecular design of chemicals. *Designing Safer Chemicals*. 1996;640:62–73.
45. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*. 1997;23:3–25.
46. Congreve M, Carr R, Murray C, Jhoti H. A ‘rule of three’ for fragment-based lead discovery?. *Drug Discovery Today*. 2003;8(19):876–877.
47. Kater SB, Nicholson C. *Intracellular staining in neurobiology*. New York, NY: Springer Science & Business Media press; 2012
48. Freeman HS, Hinks D, Esancy J. Genotoxicity of azo dyes: bases and implications. In AT Peters & HS Freeman, eds. *Physico-Chemical Principles of Color Chemistry. Advances in Color Chemistry Series, vol. 4. (Advances in Color Chemistry Series)*. Glasgow, UK: Blackie Academic & Professional; 1996:254–292.
49. Øllgaard H, Frost L, Galster J, Hansen OC. *Survey of azo-colorants in Denmark: Consumption, use, health and environmental aspects*. Copenhagen, DK: Ministry of Environment and Energy, Danish Environmental Protection Agency; 1998
50. Kontoyianni M, McClellan LM, Sokol GS. Evaluation of docking performance: comparative data on docking algorithms. *Journal of Medicinal Chemistry* 2004;47:558–565.
51. Dudek AZ, Arodz T, Galvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial Chemistry & High Throughput Screening* 2006;9(3):213–228.
52. Amaury N, Benfenati E, Bumbaru S, Chana A, Craciun M, Chretien J R, Gini G, Guo G, Lemke F, Minzu V, Mueller J A, Neagu C D, Pintore M, Stroia S A, Trundle P. Hybrid systems. In: E. Benfenati, ed. *Quantitative Structure-Activity Relationships*

- (QSAR) for Pesticide Regulatory Purposes. Amsterdam, NL: Elsevier Science Ltd.; 2007: 149–183.
53. Cassano A, Raitano G, Mombelli E, Fernández A, Cester J, Roncaglioni A, Benfenati E. Evaluation of QSAR Models for the Prediction of Ames Genotoxicity: A Retrospective Exercise on the Chemical Substances Registered Under the EU REACH Regulation. *Journal of Environmental Science and Health, Part C: Environmental Carcinogenesis and Ecotoxicology Reviews* 2014;32:273–298.
  54. Kulkarni SA, Benfenati E, Barton-Maclaren T. Improving confidence in (Q)SAR predictions under Canada's Chemicals Management Plan – a chemical space approach. *SAR and QSAR in Environmental Research* 2016;27:851–863.
  55. Cherkasov A., Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J Richard A, Tropsha A. QSAR modeling: where have you been? Where are you going to?. *Journal of Medicinal Chemistry* 2014;57(12):4977–5010.
  56. Fernández A, Lombardo A, Rallo R, Roncaglioni A, Giralt F Benfenati E. Quantitative consensus of bioaccumulation models for integrated testing strategies. *Environment International* 2012;45:51–58.
  57. Mazzatorta P, Tran L, Schilter B, Grigorov M. Integration of Structure-Activity Relationship and Artificial Intelligence Systems To Improve *in Silico* Prediction of Ames Test Mutagenicity. *Journal of Chemical Information and Modeling* 2007;47:34–38.