Chapter 21

The Consultancy Activity on In Silico Models for Genotoxic Prediction of Pharmaceutical Impurities

Manuela Pavan, Simona Kovarich, Arianna Bassan, Lorenza Broccardo, Chihae Yang, and Elena Fioravanzo

Abstract

The toxicological assessment of DNA-reactive/mutagenic or clastogenic impurities plays an important role in the regulatory process for pharmaceuticals; in this context, in silico structure-based approaches are applied as primary tools for the evaluation of the mutagenic potential of the drug impurities. The general recommendations regarding such use of in silico methods are provided in the recent ICH M7 guideline stating that computational (in silico) toxicology assessment should be performed using two (Q)SAR prediction methodologies complementing each other: a statistical-based method and an expert rule-based method.

Based on our consultant experience, we describe here a framework for in silico assessment of mutagenic potential of drug impurities. Two main applications of in silico methods are presented: (1) support and optimization of drug synthesis processes by providing early indication of potential genotoxic impurities and (2) regulatory evaluation of genotoxic potential of impurities in compliance with the ICH M7 guideline. Some critical case studies are also discussed.

Key words Genotoxic impurities, In silico methods, (Q)SAR, Statistical-based methods, Expert rule-based methods, ICH M7

1 Introduction

In silico modeling, such as (quantitative) structure-activity relationships ((Q)SARs) and molecular modeling, have been widely used in drug discovery, drug development, and regulatory purposes. In the current chapter, the focus will be primarily on the use of (Q)SARs for the evaluation of the genotoxic potential of drug impurities.

Drug impurities are defined as any component of the drug substance or drug product that is not the drug substance or an excipient (i.e., inactive constituent) and that can arise from drug synthesis or subsequent degradation, as well as from external

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_21,

contamination. In the regulatory framework for pharmaceuticals, specific guidelines exist for the qualification and control of the majority of the impurities, e.g., the International Conference on Harmonisation (ICH) Quality Guidelines Q3A ("Impurities in New Drug Substances") [1] and Q3B ("Impurities in New Drug Products") [2] and the ICH Multidisciplinary Guideline M3 ("Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorizations for Pharmaceuticals") [3]. Recently, a new guideline (ICH M7) was introduced for the identification, categorization, qualification, and control of DNA-reactive (mutagenic) impurities to limit the potential carcinogenic risk of drugs [4]. The ICH M7 guideline outlines recommendations on the use of in silico structure-based methods for genotoxicity assessment of drug impurities. According to ICH M7, computational (in silico) toxicology assessment should be performed using two (Q)SAR prediction methodologies complementing each other: a statistical-based method and an expert rule-based method. The employed (Q)SAR models should follow the internationally recognized principles for QSAR validation as defined by the Organisation for Economic Co-operation and Development (OECD) [5, 6]. According to the OECD principles, a QSAR model should (1) provide predictions for a defined endpoint; (2) be based on an unambiguous algorithm; (3) have a defined domain of applicability; (4) be internally and externally validated by applying appropriate measures of goodness of fit, robustness, and predictivity; and (5) provide a mechanistic interpretation of the prediction, when possible. The guideline recommendations state also that the outcome of any computer system-based analysis should be reviewed with the use of expert knowledge in order to provide additional supportive evidence on relevance of any positive or negative prediction and to elucidate underlying reasons in case of conflicting results. The crucial role of the expert in the final assessment is also highlighted in the literature [7–9].

In the present chapter, a practical approach for in silico assessment of mutagenic potential of drug impurities is described. The focus is on two main applications: (1) support and optimization of drug synthesis processes by providing early indication of potential genotoxic impurities and (2) regulatory evaluation of genotoxic potential of impurities in compliance with the ICH M7 guideline. Different approaches are proposed according to the specific application of the in silico assessment, and some critical case studies are discussed based on our experience.

2 Materials

In the toxicity framework, in silico predictions can be obtained by three main computational approaches: QSAR statistical-based methodologies, (Q)SAR expert rule-based methodologies, and

grouping approaches, which include read-across and chemical category formation. A brief description of the three approaches, including the underlying theory and examples of tools implementing these methods, is described in the following paragraphs.

2.1 QSAR Statistical-Based Methodology

The statistical-based QSAR method is a quantitative (mathematical) relationship between a numerical representation of the chemical structure (i.e., molecular descriptors) and a biological activity, physicochemical or fate property. Statistical-based QSARs are models based on experimental data, which extract the knowledge directly through a process of data mining and knowledge engineering. Thousands of molecular descriptors encoding for mono-, bi-, or tridimensional structural features (e.g., atom counters, topological descriptors, symmetry and steric descriptors) or chemical properties (e.g., LogP or electronic properties) have been proposed and derived from different theories and approaches, with the aim to provide an "exhaustive" description of the chemical structure. At the same time, a wide range of algorithms are now available to identify the quantitative relationship between the structure and the studied property/activity and to build statistically robust and predictive QSAR models (e.g., multiple linear regression (MLR), partial least squares (PLS) regression, artificial neural networks (ANN), etc.). It follows that the majority of statistical-based QSARs are characterized by robust validation techniques and high predicting performances, and can provide predictions also when the mechanism of action is unknown. Additionally, several mathematical/chemometrical metrics have been developed to define model applicability domain and to measure the level of extrapolation. On the other hand, in some cases, their predictions could miss a mechanistic reasoning and a clear interpretation, especially when based on complex algorithms and molecular descriptors, thus resulting "nontransparent" to the end user.

Nowadays, several tools (both commercial and freeware) are available coding QSAR statistical models for the prediction of mutagenic/genotoxic potential [10–14]. We routinely use an array of commercial and freely available tools in a weight of evidence approach. All the predictors we use fulfill the OECD principles for QSAR validation and are characterized by (1) wide and heterogeneous training set collected from valid sources (e.g., FDA—US Food and Drug Administration), (2) high robustness and external predictivity, (3) wide applicability domain, and (4) defined parameters for reliability assessment. Additionally they allow the user to visualize structure and experimental data of structural analogues, thus providing supporting information to further assess the prediction. A brief description of these tools is as follows:

- *ACD/Percepta Impurity Profiling* [15, 16] provides a battery of in silico models to accurately assess the genotoxic

and carcinogenic potential of chemicals. The impurity profiling module is a result of the collaboration between ACD/Labs and FDA Center for Food Safety and Applied Nutrition (CFSAN). This module includes probabilistic predictive models for 21 different endpoints that cover various mechanisms of hazardous activity (including mutagenicity, clastogenicity, DNA damage mechanisms, carcinogenicity, and endocrine disruption mechanisms) and that are based on experimental data obtained from FDA. Probabilistic predictive models were developed using GALAS modeling methodology [17]. Each GALAS model consists of two parts: (1) a global (baseline) model, built using binomial PLS method based on fragmental descriptors, that reflects a "cumulative" mutagenicity potential, and (2) local corrections that are applied to baseline predictions using a special similarity-based routine, after performing an analysis for the most similar compounds used in the training set. The reliability of prediction is assessed in terms of reliability index (RI), which ranges from 0 to 1 and takes into account the similarity of the target with the training set compounds and the consistency of experimental values for similar compounds. A "positive" or "negative" call is then provided if the compound can be reliably classified on the basis of p-value (i.e., probability that a compound will result in a positive test in the respective assay) and RI values ("undefined" otherwise).

- *ChemTunes Studio* is a knowledge base software consisting of experimental in vitro and in vivo toxicity information (QC'ed by experts) and in silico models for a series of human health toxicity endpoints, comprising the key genetic toxicity endpoints (i.e., Ames mutagenicity, chromosome aberration, and in vivo micronucleus). The software is made of multiple components, including genotoxic chemotypes (structural alerts); mechanistically informed (mode-of-action driven) QSAR models, i.e., an approach used at US FDA CERES (Chemical Evaluation and Risk Estimation System) [18, 19]; and comparison of the prediction results to structural analogues. A mathematically rigorous and quantitative weight of evidence (WoE) decision theory approach is used to obtain the final overall assessment and to provide a quantitative estimation of the uncertainty associated with the prediction. All ChemTunes Studio QSAR models consist of chemical mode-of-action category models as well as a general global model. The computational modeling approach is a hybrid of partial least squares (PLS)/ordinal logistic regression methods. For model building, global molecular and shape descrip-

tors (from CORINA Symphony [20]) and quantum-mechanic parameters are used. The models return probabilistic predictions (positive and negative probabilities plus a quantitative estimate of the associated uncertainty) and an overall prediction (positive/negative/equivocal). Applicability domain analysis reports whether the target compound is out of domain. QSARs for bacterial reverse mutagenesis (Ames mutagenicity) are based on selected studies for more than 2200 structures, compiled from various sources, and including *S. typhimurium* and *E. coli* strains with and without metabolic activation.

- *Leadscope Model Applier/Genetox QSAR Statistical Suite* [21] is a chemoinformatic platform that provides QSARs for the prediction of potential toxicity and adverse human clinical effects, including the microbial in vitro *Salmonella* mutagenicity model that is used by the US FDA (Food and Drug Administration) in their testing under the ICH M7 Guidance for impurities [22–24]. The in vitro *Salmonella* mutagenicity QSAR model was constructed by the FDA scientists based on a training set of over 3500 compounds (including both proprietary and nonproprietary data). The model is based on a wide set of molecular descriptors, including 369 substructural features and seven calculated properties, and on partial logistic regression (PLS) modeling technique. Model predictions consist of four possible results, i.e., "positive," "negative," "indeterminate," or "not in domain," and probability of a positive result. Predictions are provided together with several parameters, which can be used to assess the prediction in terms of applicability domain (e.g., the presence in the target compound of model training set structural features and the presence of structural analogues in the training set).
- *VEGA/CAESAR Mutagenicity* model is a QSAR model predicting mutagenicity developed under the EU project CAESAR [25] and implemented in the VEGA platform [26]. The QSAR model is based on a dataset of 4225 compounds and consists of an integrated model made of two complementary techniques: a machine learning algorithm (SVM), to build an early model with the best statistical accuracy, equipped with an expert facility for false negative removal based on known structural alerts, to refine its predictions. The reliability of predictions is assessed using an Applicability Domain Index (ADI) that ranges from 0 to 1 and is calculated by grouping several other indices, each one taking into account a particular issue of the applicability domain (i.e., the presence of similar compounds in the training set, the consistency of their experimental data and

their prediction accuracy, the presence in the target of structural fragments possessed by training set compounds, and the range of values of modeling descriptors).

2.2 (Q)SAR Expert Rule-Based Methodology

The (Q)SAR expert rule-based (or knowledge-based) method relies on rules derived from toxicological knowledge, which are likely to have strong mechanistic basis, used to make predictions about a defined adverse effect. In the expert rule-based systems, human experts identify structural fragments related to the studied effect. The examination of a series of chemicals sharing the same fragment ("structural alert"—SA) is used to detect the toxic effect (e.g., genotoxic or not); the chemical information is simply the fragment and the algorithm is, in this case, the rule. The expert rule-based systems have several advantages, e.g., they are mechanistically connected to the predicted activity, provide reasoning for the predictions, and in many cases support the prediction with literature references or expert knowledge. On the other side, applicability domain measures for expert systems are not well defined [27], and usually it is not possible to discriminate active from inactive chemicals bearing the same structural alert. The accuracy in prediction is mostly comparable to statistical-based QSARs; however, expert systems tend to exhibit a higher sensitivity at the cost of a lower specificity (SAs are conservative), whereas the statistical-based QSARs show the opposite behavior [28].

Several tools (both commercial and freeware) are now available coding expert rule-based systems [10–14]. In some tools, expert systems are combined with statistical-based models (the so-called hybrid systems), in order to provide supporting knowledge-based evidence to QSAR predictions. For our consultant activities, we routinely use an array of commercial and freely available tools in a weight of evidence approach. The predictors in use are based on wide sets of chemicals and alerts and provide means to assess the reliability of predictions. A brief description of these tools is as follows:

- *ACD/Percepta Impurity Profiling* [15, 16] is supplemented with a knowledge-based expert system that identifies potentially hazardous structural fragments that could be responsible for genotoxic and/or carcinogenic activity of the compound of interest. The expert system contains a list of 70 alerting groups of toxicophores, of which 33 represent mutagens, 24 clastogens, and 13 epigenetic carcinogens (androgens, peroxisome proliferators, etc.). The alert list is not limited to directly acting substructures, such as planar polycyclic arenes, aromatic amines, quinones, and N-nitro and N-nitroso groups, but also includes various fragments that may undergo biotransformation to reactive intermediates. Each hazardous fragment is provided with a

description of its mechanism of action, literature references, and *z*-scores. *z*-Scores show whether the presence of the fragment leads to a statistically significant increase in the proportion of compounds with a positive test result for a particular assay. The identified alerting groups are highlighted on the structure of the molecule and the five most structurally similar structures from the training set, along with experimental results, are shown.

- *ChemTunes Studio* includes, in addition to QSAR statistical-based models, genotoxic chemotypes (structural alerts), developed from mechanistic hypothesis; each alert is provided with likelihood prioritization, so that alerts can be used when combining the different information at the WoE stage. A knowledgebase was built and curated for a large dataset (over 8000 compounds) of Ames mutagenicity data from public sources. The reliability of each alert is determined by exploring the ability of the alert to hit positive compounds in a large training set. Different training sets were used for the QSAR models and the alerts, so that predictions from these are independent.
- *Leadscope Model Applier/Genetox Expert Alerts Suite* is implemented as part of the Leadscope Model Applier (in addition to the existing statistical-based QSAR model) [21]. To develop this system, an initial library of mutagenicity structural alerts was identified from the literature. Information on plausible mechanisms was collected as well as the structural definitions. Factors that deactivate the alerts were also identified from the literature and through an analysis of the corresponding data using the Leadscope data mining software. Over 200 distinct alerts are encoded in the system. These alerts were further validated against a reference database of over 7000 chemicals with known bacterial mutagenesis results. A confidence score based upon information collected for each alert is provided alongside the positive or negative call. Up to ten structurally similar structures from the alert reference set, along with experimental results, are provided.
- *Toxtree* [29] is a flexible and user-friendly open-source application that places chemicals into categories and predicts various kinds of toxic effects by applying decision tree approaches. The decision tree for estimating mutagenicity is based on discriminant analysis and structural rules as described in Benigni et al. [30]. It estimates in vitro (Ames test) mutagenicity, based on a list of 30 structural alerts (SAs). As one or more SAs embedded in a molecular structure are recognized, the system flags the potential mutagenicity of the chemical. The use of Toxtree Benigni-Bossa

decision tree implemented in VEGA platform [26] allows the user to assess the reliability of predictions by means of the Applicability Domain Index (ADI) calculated in VEGA and to visualize chemical structure and experimental data for the most similar structures in Toxtree alert training set.

2.3 Grouping Approaches: Read-Across Methodology

Chemical grouping approaches are based on the formation of chemical "categories" or "analogues," composed by groups of chemicals whose physicochemical, (eco-)toxicological, and/or environmental fate properties are likely to be similar or follow a regular pattern. This can be the result of a structural similarity or other similarity characteristics (e.g., common mechanism of action). In principle, the chemical category is composed by several members, enabling the detection of trends across endpoints, while the grouping by analogue approach is based on a limited number of chemicals, where trends in properties are not apparent [31]. In these cases, predictions are generated by applying the "read-across" method. In the read-across technique, the endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered "similar" in some way (usually based on structural similarity). The chemical(s) being used to make an estimate is commonly referred to as a "source chemical(s)," whereas the chemical for which the endpoint is being estimated is referred to as a "target chemical." The read-across methodology is currently accepted to fill data gaps in the regulatory framework, basically for the transparency and interpretability of the approach and of the final outcome. However, read-across is not a formalized approach (i.e., it is not based on a defined and reproducible algorithm), and the obtained predictions strongly depend on the expert judgment. For these reasons, specific guidelines on how to perform a read-across study in order to be accepted for regulatory purposes (e.g., REACH) have been developed [32]. According to this guideline, any read-across analysis should be supported by a detailed documentation to be provided according to the defined read-across reporting formats [31, 33].

The OECD QSAR Toolbox [34] is the main tool we use to perform read-across predictions [35]. It was developed by the OECD to use (Q)SAR methodologies to group chemicals into categories and to fill data gaps by read-across and trend analysis. It is currently recommended and released by the European Chemicals Agency (ECHA) in collaboration with OECD. The Toolbox incorporates information and tools from various sources into a logical workflow, which supports the user to carry out read-across studies through the identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical, the identification of other chemicals that have the same structural characteristics, and/or mechanism or mode of action and the use of existing experimental data to fill the data gaps. Another freely

available software useful to assist users for read-across evaluations is ToxRead [36]. ToxRead was recently developed by IRCCS (Istituto di Ricerche Farmacologiche Mario Negri), Politecnico di Milano, and KODE within a joint collaboration between the LIFE projects CALEIDOS and PROSIL and offers a workflow to generate read-across predictions with high reproducibility.

2.4 Weight of Evidence Approach

Any predictive model is by definition a simulation of reality, and therefore it will never be completely accurate. The same applies to (Q)SARs. As discussed in the previous paragraphs, each computational approach, i.e., statistical-based, expert rule-based, or read-across approach, has its own advantages and weaknesses. Likewise, each (Q)SAR model is characterized by distinctive predictive performances (e.g., sensitivity versus specificity) and a defined applicability domain (i.e., no QSAR model can be applied to every chemical of interest), thus providing different partial "views" of the whole picture. Thus, the most reasonable way to get the best out of several views and achieve accurate predictions is to combine predictions from different models and approaches in a weight of evidence approach [37–39]. A weight of evidence (WoE) approach involves an assessment of the values and relative weights of different pieces of available information [40]; in our case, it implies an assessment of different in silico predictions taking into account the reliability of each prediction and the concordance among different predictions. This can be achieved either in an objective way by using a formalized procedure or by using expert judgment. Some tools, such as ChemTunes and Leadscope Model Applier, provide algorithms for the calculation of WoE (or consensus) predictions based on the combination of predictions from statistical- and expert rule-based models as well as experimental data. It has been broadly demonstrated that the complementary use of statistical-based and expert-based approaches, supplemented by expert knowledge, improves prediction accuracy [8, 11, 14, 41].

3 Methods

3.1 Early Indication of Potential Genotoxic Impurities

In silico methods can be efficiently employed in the early stages of drug development for the screening and identification of potential genotoxic impurities, thus providing useful information to optimize the design of the synthesis scheme. When in silico methods are used for screening purposes, the integration of statistical-based and knowledge-based approaches is not mandatory, and a less detailed documentation of the burden of proof is required. Our procedure for an early indication, by means of in silico methods, of the potential genotoxicity of impurities is described and summarized in Fig. 1.

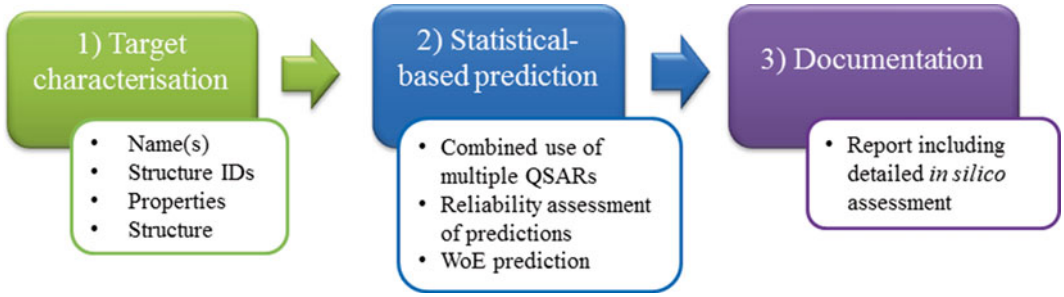

Fig. 1 Workflow for early indication of potential genotoxic impurities

1. Characterization of the target impurity by means of chemical names, registry number, structure identifiers (e.g., SMILES, InChI), chemical structure, and properties (e.g., molecular weight, molecular formula).
2. QSAR statistical-based prediction of bacterial mutagenicity:
   (a) Combined use of multiple tools for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test). For screening purposes, statistical-based QSAR models are usually preferred than knowledge-based approaches because of their higher accuracy and wider applicability [42]. Among the available predictors based on a statistical approach, we are currently using ACD/Labs Percepta, Leadscope Model Applier, and the CAESAR Mutagenicity model implemented in VEGA, while ChemTunes is going to be integrated. These predictors are particularly indicated for screening purposes since they are characterized by wide and heterogeneous training set (including drug substances), external predictivity, and wide applicability domains.
   (b) Assessment of the prediction reliability taking into account multiple issues, e.g., (i) whether the target impurity falls within the applicability domain of the model, (ii) whether and how the target impurity is represented in the training set by analyzing the structural analogues included in the training sets, (iii) prediction accuracy of the identified analogues, and (iv) consistency between the analogues' experimental test results (Ames test) and the prediction for the target impurity. Identification of the proper analogues is a critical step and depends on the methodology used to measure chemical similarity. Defining chemical similarity measures to infer mutagenic potential as well as approaches to assess the reliability of predictions is still an open challenge [43].
   (c) Generation of a WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, taking into account only

reliable predictions. If different predictors, based on different training molecules, molecular descriptors, and modeling approaches, lead to consistent results, then a higher level of confidence in the in silico prediction is achieved. If equally reliable but not consistent results are provided by different predictors, then the most conservative outcome, i.e., positive, should be concluded. Examples on how to deal with critical case studies, e.g., not consistent and/or unreliable predictions, are commented in Subheading 4 (**Notes 1–5**).

3. Documentation of the results. The predictions provided by the different tools together with the performed WoE analysis are described in a detailed report.

3.2 Regulatory Evaluation of Genotoxic Potential of Impurities (ICH M7 Guideline)

According to ICH M7 guideline, hazard assessment of genotoxic impurities first involves an analysis of actual and potential impurities, based on experimental carcinogenicity and bacterial mutagenicity data available from database and literature. If such data are not available, in silico (Q)SAR assessment of the impurities should be performed to provide predictions for bacterial mutagenicity. As a result of the hazard assessment, drug impurities are assigned to one of the five classes summarized in Fig. 2, and specific control actions are suggested [4].

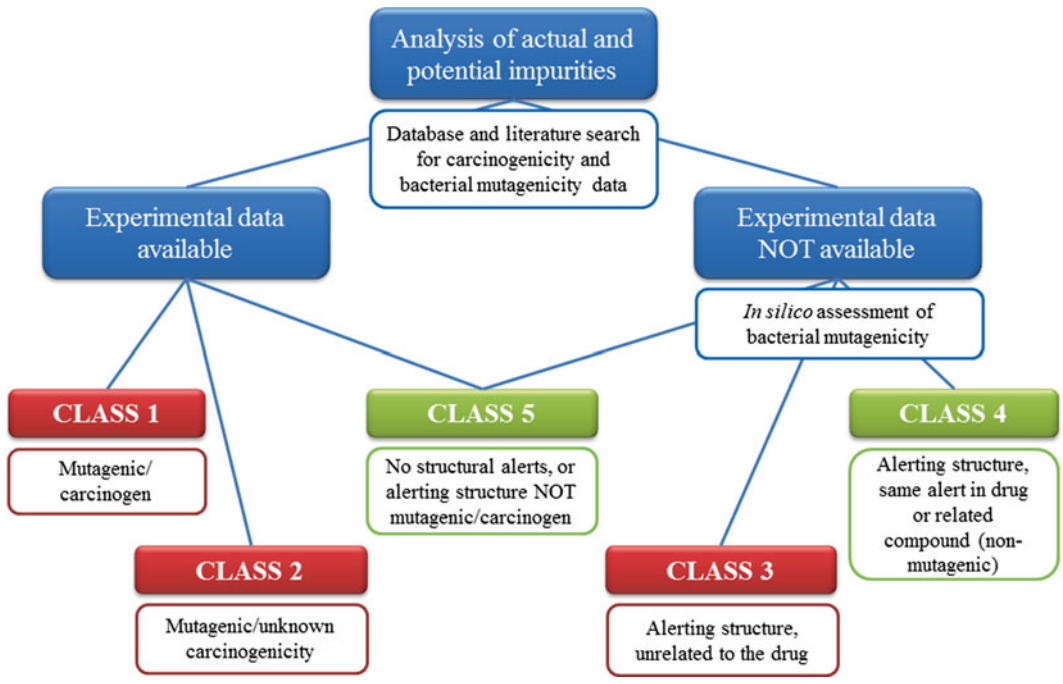

Fig. 2 Impurities classification with respect to mutagenic and carcinogenic potential

Table 1
Examples of critical case studies for in silico assessment of genotoxic impurities

No.	Statistical-based WoE	Expert rule-based WoE	Read-across study	Conclusive in silico assessment
1	NEGATIVE	OUT OF DOMAIN/ INCONCLUSIVE	NEGATIVE based on negative source chemical(s) (e.g., the API or structural related impurities)	NEGATIVE
2	OUT OF DOMAIN/ INCONCLUSIVE	NEGATIVE	NEGATIVE based on negative source chemical(s)	NEGATIVE
3	OUT OF DOMAIN/ INCONCLUSIVE	POSITIVE based on alert X	NEGATIVE based on negative source chemical(s) possessing the same alert X	NEGATIVE
4	NEGATIVE	POSITIVE based on alert X	NEGATIVE based on negative source chemical(s) possessing the same alert X	NEGATIVE
5	NEGATIVE	POSITIVE based on alert X	NOT FEASIBLE/ POSITIVE positive source chemical(s) possessing the same alert X	POSITIVE

The ICH M7 guideline states that the computational toxicology assessment should be performed by using two (Q)SAR prediction methodologies that complement each other, i.e., a statistical-based and an expert rule-based methodology. In addition, expert analysis including read-across is applied to provide additional supportive evidence on the predictions and/or to solve conflicting results. It is here described our stepwise procedure for regulatory in silico assessment of genotoxic impurities. The procedure is also summarized in the workflow of Fig. 3.

1. Characterization of the target impurity (i.e., chemical names, structure identifiers, chemical structure, and properties)
2. QSAR statistical-based prediction of bacterial mutagenicity:
   (a) Combined use of multiple statistical-based QSAR models for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test).
   (b) Assessment of the reliability of the predictions provided by the individual statistical-based tools as described in Subheading 3.1 (**step 2b**).

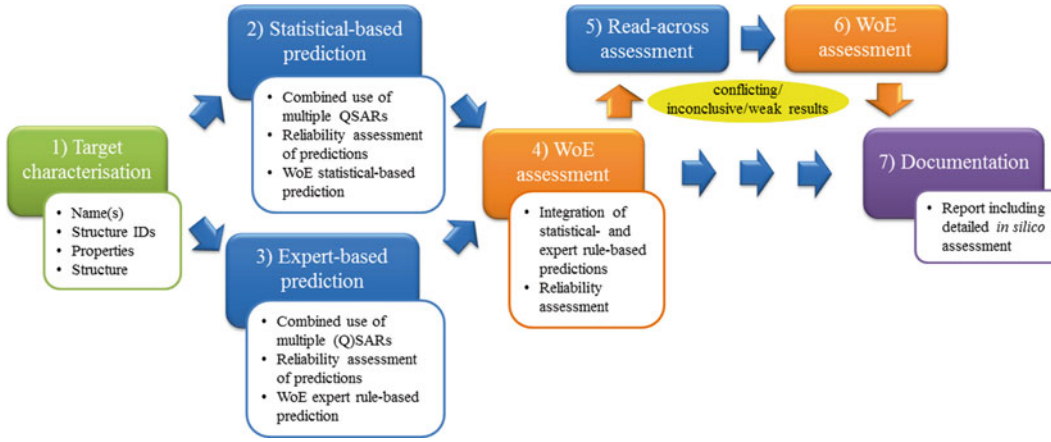

Fig. 3 Workflow for regulatory evaluation of potential genotoxic impurities

(c) Computation of the statistical-based WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based on the employed statistical-based tools. The level of confidence of the WoE prediction (e.g., unreliable, borderline, moderate, or highly reliable) is defined taking into account the reliability and consistency of the predictions obtained by the individual employed statistical-based tools.

3. (Q)SAR expert rule-based prediction of bacterial mutagenicity:

(a) Combined used of multiple expert rule-based methods for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test). Among the available knowledge-based tools, we are currently using ACD/Labs Percepta, Leadscope Model Applier, and the Toxtree in vitro mutagenicity (Benigni-Bossa) decision tree implemented in VEGA. The novel expert system implemented in ChemTunes based on genotoxic chemotypes is going to be integrated in our in silico assessment. These tools provide a positive, negative, or inconclusive prediction based on the identification of one or more structural alerts for mutagenicity, as well as the means to assess the reliability of the prediction (as discussed in the next step). Particular attention is paid to negative ("non-genotoxic") predictions based on the absence of structural alerts. In fact, the absence of any known structural alerts is NOT a sufficient evidence for a lack of effect, and there is the possibility that the target impurity may act through an unknown mechanism of action, for which structural alerts have not been developed yet.

(b) Assessment of the reliability of the predictions provided by the expert SA-based tools. Although structural alerts often lack an adequately defined applicability domain [27], the

level of confidence of the predictions can be assessed focusing on the following issues: (i) whether the target impurity is sufficiently represented in the training set, in terms of structural similarity, chemical fragments, or other structural features represented in the training set; (ii) relevance of the identified alert, i.e., the alert is characterized by a statistically significant higher frequency in genotoxic compounds compared to non-genotoxic (from the training set); (iii) precision of the identified alert, i.e., accuracy of the alert in the correctly predicted genotoxic compounds (i.e., true positive rate); and (iv) consistency between the experimental test results (Ames test) of the identified analogues (particularly those sharing the same alert(s)) and the predicted outcome of the target impurity. If no structural alerts for genotoxicity are identified, a proper reliability assessment is not applicable. In these cases, a detailed analysis of the structural analogues with no alerts and the precision of the expert system toward training compounds with no alerts is recommended [13].

(c) Generation of the expert rule-based WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based on the employed expert rule-based tools. The level of confidence of the WoE prediction is defined taking into account the reliability and consistency of the predictions obtained by individual tools.

4. Generation of the final WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based on the integration of the outcome of the statistical-based and expert rule-based WoE predictions. The level of confidence of the WoE prediction is defined taking into account the reliability and consistency of the predictions obtained by the two approaches. In case of conflicting results and/or weak WoE assessment (i.e., low reliability), either we conclude for a predicted genotoxic potential (conservative scenario) or, preferably, we integrate the in silico assessment with a read-across study (as described in **step 5**). It is important to highlight that the WoE approach is not an automatic procedure, rather an assessment based on expert judgment performed on a case-by-case analysis of the predictions. Examples on how to deal with some critical case studies, e.g., not consistent and/or unreliable predictions, are commented in Subheading 4 (**Notes 1–5**).

5. Read-across study to provide additional supportive evidence on the predictions and/or to solve conflicting results. From our consultancy experience, the source chemical(s) is often suggested by the commissioner and could be either the API (active pharmaceutical ingredient), compounds related to the drug substance (e.g., process intermediates), or structurally

related impurities, for which the commissioner already conducted an experimental Ames test. Alternatively, an extensive search in the literature and in open databases (e.g., DSSTox [44], ECHA CHEM [45], NTP [46], GENE-TOX [47], etc.) is performed to identify the most appropriate source(s) for the target impurity. The read-across study is performed and documented according to the guidance document on the grouping of chemicals (including read-across and chemical categories) [31–33]. The OECD QSAR toolbox is employed to identify the functional groups (by applying the Organic Functional Groups (OFG) system) and to profile the source and target chemicals by describing their foreseen mechanism of action relevant for mutagenic activity. Two general mechanistic profilers, namely, DNA binding by OECD and DNA binding by OASIS v.1.2, and three endpoint-specific profilers, namely, DNA alerts for AMES, MN, and CA by OASIS v.1.2, in vitro mutagenicity (Ames test), and in vivo mutagenicity (micronucleus) alerts by ISS, are used being the most meaningful profilers for genotoxicity available in the toolbox [48].

6. Conclusion from the in silico assessment on the potential genotoxicity of the target impurity, based on results of the two QSAR prediction methodologies, i.e., a statistical-based method and an expert rule-based method, and the supporting evidence coming from the read-across study.
7. Documentation of the results. The predictions provided by the different tools and approaches, together with the performed WoE analysis, are described in a detailed report.

4 Notes

The interpretation of results from a (Q)SAR assessment of genotoxic impurities is not always straightforward, and several issues are commonly encountered. Thus, the role of the expert is crucial to build up a WoE prediction by an integrated approach, which considers information gained by various techniques, to provide additional supportive evidence on relevance of any positive or negative prediction and to elucidate underlying reasons in case of conflicting or inconclusive results. Some examples of critical and real case studies are reported and illustrated in Table 1. In all cases, three statistical-based models, i.e., ACD/Percepta Impurity Profiling (in vitro *Salmonella* model), Leadscope Model Applier/Genetox QSAR Statistical Suite (microbial in vitro *Salmonella* model), and VEGA/CAESAR Mutagenicity model, were employed together with three expert rule-based systems, i.e., ACD/Percepta Impurity Profiling (in vitro *Salmonella* expert system), Leadscope Model Applier/Genetox QSAR Expert Suite (Bacterial Mutation), and

the Toxtree in vitro mutagenicity (Benigni-Bossa) decision tree implemented in VEGA platform.

1. Case study 1: The target impurity is reliably predicted as negative by the statistical-based approach, while the prediction obtained by the expert rule-based approach is not reliable ("out of domain") or inconclusive. In this case, it is not possible to derive a robust WoE prediction, since two approaches are required by the ICH M7 regulation, and the read-across approach is suggested to provide further evidence of the negative prediction.
2. Case study 2: The prediction obtained from the statistical-based approach is not reliable ("out of domain"), or inconclusive, while the outcome of the expert rule-based approach is negative, based on the absence of structural alerts for genotoxicity. Again, a read-across study is suggested to provide further evidence of the negative prediction.
3. Case study 3: The prediction obtained from the statistical-based approach is not reliable, or inconclusive, while the outcome of the expert rule-based approach is a reliable positive prediction, based on the detection of one or more structural alerts for genotoxicity. In this case, it is not possible to derive a robust WoE prediction, and the read-across approach is suggested to verify whether the presence of the alert induces (or not) a positive effect. If the identified source chemical (e.g., the API or structural related impurities) shares with the target impurity the same structural alert (e.g., same structural alert in the same position and environment in the impurity and the source) and the source chemical is non-mutagenic, then the target impurity is predicted negative by the read-across (Class 4 according to ICH M7). In this case, in agreement with the ICH M7 guideline, the read-across study overturns the expert rule-based prediction, and the final in silico assessment concludes for a negative prediction.
4. Case study 4: Conflicting predictions are obtained applying the two different methodologies, e.g., negative outcome obtained with the statistical-based approach and positive outcome obtained with the expert rule-based system. The WoE assessment, based on a precautionary approach, would conclude for a positive prediction, leading possibly to a false positive. The read-across approach is thus suggested to solve conflicting results. As discussed in case study 3, if the impurity shares with the source chemical the same structural alert and the source chemical is non-mutagenic, then the target impurity is predicted negative by the read-across (Class 4 according to ICH M7). Thus, the read-across study overturns the WoE assessment based on statistical-based and expert rule-based

predictions, and the final in silico assessment concludes for a negative prediction.

5. Case study 5: Conflicting predictions are obtained applying the two different methodologies, e.g., negative outcome obtained with the statistical-based approach and positive outcome obtained with the expert rule-based system. As discussed in case study 4, the target impurity is predicted as suspect positive following a precautionary approach, and the read-across approach is suggested. If no structural analogues justifying the read-across study can be identified or if the source chemical(s) possessing the structural alert identified in the target impurity shows positive experimental Ames test results, then the in silico assessment concludes for a positive prediction. Hence, the target impurity must be submitted for experimental assessment of mutagenicity.

References

1. ICH Harmonised Tripartite Guideline (2006) Impurities in new drug substances – Q3A(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
2. ICH Harmonised Tripartite Guideline (2006) Impurities in new drug products – Q3B(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
3. ICH Harmonised Tripartite Guideline (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals – M3(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
4. ICH Harmonised Tripartite Guideline (2014) Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk – M7. Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
5. OECD – Organization for Economic Co-operation and Development (2006) Report on the regulatory uses and applications in oecd member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. OECD Environment Health and Safety Publications, Series on testing and assessment no. 58. ENV/JM/MONO(2006)25. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)25&doclanguage=en. Accessed 11 May 2015
6. OECD – Organization for Economic Co-operation and Development (2007). Guidance document on the validation of (quantitative) structure activity relationship [(Q)SAR] Models. OECD Environment Health and Safety Publications, series on testing and assessment no. 69. ENV/JM/MONO(2007)2. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en. Accessed 11 May 2015
7. Dobo KL, Greene N, Fred C et al (2012) In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. Regul Toxicol Pharmacol 62(3):449–455
8. Sutter A, Amberg A, Boyer S et al (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul Toxicol Pharmacol 67(1):39–52
9. Powley MW (2014) (Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective on the utility of expert knowledge and data submission. Regul Toxicol Pharmacol 71(2):295–300
10. Serafimova R, Gatnik MF, Worth A (2010) Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. JRC Scientific and Technical Reports EUR 24427 EN – 2010. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_24427_EN.pdf. Accessed 6 May 2015
11. Worth AP, Lapenna S, Serafimova R (2013) QSAR and metabolic assessment tools in the assessment of genotoxicity. Methods Mol Biol 930:125–162

12. Roncaglioni A, Toropov AA, Toropova AP, Benfenati E (2013) In silico methods to predict drug toxicity. Curr Opin Pharmacol 13(5):802–806
13. Fioravanzo F, Bassan A, Pavan M et al (2012) Role of in silico genotoxicity tools in the regulatory assessment of pharmaceutical impurities. SAR QSAR Environ Res 23(3-4):257–277. doi:10.1080/1062936X.2012.657236
14. Cassano A, Raitano G, Mombelli E et al (2014) Evaluation of QSAR models for the prediction of ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 32(3):273–298
15. ACD/Percepta, release 2014. Advanced Chemistry Development, Inc., Toronto, ON, Canada. www.acdlabs.com
16. Lanevskij K, Juska L, Dapkunas J et al (2012) In silico test battery for rapid evaluation of genotoxic and carcinogenic potential of chemicals. Poster presented at 243rd ACS national meeting, San Diego, California, 25–29 Mar 2012
17. Didziapetris R, Dapkunas J, Sazonovas A, Japertas P (2010) Trainable structure-activity relationship model for virtual screening of CYP3A4 inhibition. J Comput Aided Mol Des 24(11):891–906
18. Leis M, Lidbury BA, Yang C et al (2012) Novel technologies and an overall strategy to allow hazard assessment and risk prediction of chemicals, cosmetics, and drugs with animal-free methods. Altex 29(4):373–388
19. Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
20. CORINA Symphony, Molecular Networks GmbH, Erlangen, Germany. www.molecular-networks.com. Accessed 6 May 2015
21. Leadscope Model Applier, version 2.0 (2015) Leadscope Inc., Columbus, OH. http://www.leadscope.com
22. Dearfield KL, Benz RD (1999) Can the new genetic toxicology tests be used for regulatory safety decisions? Environ Mol Mutagen 33(2):91–93
23. Matthews EJ, Kruhlak NL, Cimino MC et al (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul Toxicol Pharmacol 44(2):83–96
24. Matthews EJ, Kruhlak NL, Cimino MC et al (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regul Toxicol Pharmacol 44:97–110
25. Ferrari T, Gini G (2010) An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem Cent J 4(Suppl 1):S2
26. VegaNIC Application (Virtual Models for evaluating the properties of chemicals within a global architecture), ver 1.1.0. Laboratory of Environmental Chemistry and Toxicology of Mario Negri Institute of Pharmacological Research. http://www.vega-qsar.eu/download.html. Accessed 6 May 2015
27. Ellison CM, Sherhod R, Cronin MT et al (2011) Assessment of methods to define the applicability domain of structural alert models. Chem Inf Model 51(5):975–985. doi:10.1021/ci1000967
28. Hillebrecht A, Muster W, Brigo A et al (2011) Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations. Chem Res Toxicol 24(6):843–854
29. Toxtree (Toxic Hazard Estimation by decision tree approach), ver. 2.6.13. Ideaconsult Ltd. http://toxtree.sourceforge.net/index.html. Accessed 6 May 2015
30. Benigni R, Bossa C, Jeliazkova N et al (2008) The Benigni/Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. JRC Scientific and Technical Reports, EUR 23241 EN – 2008. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23241_EN.pdf. Accessed 6 May 2015
31. ECHA – European Chemicals Agency (2008) Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf. Accessed 6 May 2015
32. OECD – Organization for Economic Co-operation and Development (2007) Guidance on grouping of chemicals. OECD Environment Health and Safety Publications, series on testing and assessment no. 80. ENV/JM/MONO(2007)28. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)28&doclanguage=en. Accessed 11 May 2015
33. ECHA – European Chemicals Agency (2012) Practical guide 6: how to report read-across and categories. http://echa.europa.eu/documents/10162/13655/pg_report_readacross_en.pdf. Accessed 6 May 2015

34. OECD (Q)SAR Application Toolbox, version 3.3.2 (2015) Organization for economic co-operation and development. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 6 May 2015
35. OECD – Organization for Economic Co-operation and Development (2009) Guidance document for using the OECD (Q) SAR application toolbox to develop chemical categories according to the OECD guidance on grouping of chemicals. OECD Environment Health and Safety Publications, series on testing and assessment no. 102. ENV/JM/MONO(2009)5. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2009)5. Accessed 11 May 2015
36. ToxRead, v. 0.9 BETA (2015) Istituto di Ricerche Farmacologiche Mario Negri – Laboratory of Environmental Toxicology and Chemistry. http://www.toxgate.eu/. Accessed 6 May 2015
37. Judson P (2009) Combining predictions. In: Royal Society of Chemistry (ed) Knowledge-based expert systems in chemistry: not counting on computers, 1st edn. Royal Society of Chemistry, London
38. Cronin MTD (2010) Prediction of harmful human health effects of chemicals from structure. In: Puzyn T, Leszczynski J, Cronin MTD (eds) Recent advances in QSAR studies: methods and applications. Springer, New York, NY
39. Ellison CM, Madden JC, Judson P, Cronin MTD (2010) Using in silico tools in a weight of evidence approach to aid toxicological assessment. Mol Inf 26(1-2):97–110
40. ECHA – European Chemicals Agency (2010) Practical guide 2: how to report weight of evidence. https://echa.europa.eu/documents/10162/13655/pg_report_weight_of_evidence_en.pdf Accessed 8 June 2015
41. Nendza M, Gabbert S, Kühne R et al (2013) A comparative survey of chemistry-driven in silico methods to identify hazardous substances under REACH. Regul Toxicol Pharmacol 66(3):301–314
42. Roy PP, Kovarich S, Gramatica P (2011) QSAR model reproducibility and applicability: a case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles. J Comput Chem 32:2386–2396. doi:10.1002/jcc.21820
43. Naven RT, Greene N, Williams RV (2012) Latest advances in computational genotoxicity prediction. Expert Opin Drug Metab Toxicol 8(12):1579–1587
44. DSSTox (Distributed Structure-searchable Toxicity) database. U.S. Environmental Protection Agency. http://www.epa.gov/ncct/dsstox. Accessed 6 May 2015
45. ECHA CHEM database. European Chemical Agency (ECHA). http://echa.europa.eu/chem_data_en.asp. Accessed 6 May 2015
46. NTP (National Toxicology Program) database. U.S. National Institute of Environmental Health Sciences, National Institutes of Health. http://ntp.niehs.nih.gov. Accessed 6 May 2015
47. GENE-TOX (Genetic Toxicology Databank) database. U.S. National Library of Medicine, National Institutes of Health. http://toxnet.nlm.nih.gov/newtoxnet/genetox.htm Accessed 6 May 2015
48. OECD – Organization for Economic Co-operation and Development (2013), User Manual based on QSAR Toolbox version 3.0. Strategies for grouping chemicals to fill data gaps to assess genetic toxicity and genotoxic carcinogenicity. http://www.oecd.org/env/ehs/risk-assessment/genetic%20toxicity.pdf. Accessed 7 May 2015