



## ORIGINAL RESEARCH

# Testing the Mutagenicity Potential of Chemicals

Geert R Verheyen\*, Koen Van Deun and Sabine Van Miert

Thomas More University College, Belgium

\*Corresponding author: Geert R Verheyen, Radius, Thomas More University College Campus Geel, Kleinhoefstraat 4, 2440 Geel, Belgium, Tel: +32-0-14-562310, E-mail: [geert.verheyen@thomasmore.be](mailto:geert.verheyen@thomasmore.be)

### Abstract

All information for the proper development, functioning and reproduction of organisms is coded in the sequence of matched base-pairs of DNA. DNA mutations can result in harmful effects and play a role in genetic disorders and cancer. As mutations can arise through exposure to chemical substances, testing needs to be done on substances that humans and animals can be exposed to. This review focuses on the different testing strategies for risk assessment of chemicals for genotoxicity and carcinogenicity. This review is not meant to cover all testing methodologies, but rather to give an overview of the main methodologies that are used in a regulatory context. *In silico* and *in vitro* tools are playing an increasingly important role, in order to reduce *in vivo* animal testing. Especially *in silico* tools like (Q)SARs are promising as they are non-testing methods and thereby also reduce the costs and time that are needed to perform *in vitro* assays. (Q)SARs are not sufficiently reliable to be used as stand-alone tools, and should be associated with expert judgment or used in combination with other Weight of Evidence such as *in vitro* and *in vivo* data and data derived from other similar molecules (read-across). For some product (sub) categories, testing (*in vitro/in vivo*) will be unavoidable; however the future may bring reliable approaches for ensuring human safety.

### Keywords

Mutagenicity, *In vivo*, *In vitro*, QSAR, Hazard identification, Risk assessment

## Introduction

The genetic make-up of all organisms is contained within the double helix of DNA. All information for the development, functioning and reproduction of organisms is coded within the sequence of matched base-pairs and it is of crucial importance that this information is passed on to progeny without or with the least amount of error possible. Thus at the cellular level it is important that the replication of DNA and the segregation of the

fully replicated DNA molecules to the 2 daughter cells occurs accurately. However, DNA replication is not a faultless process and mutations, which can be defined as various types of permanent changes in DNA, do occur in all organisms. Mechanisms underlying mutations have been well studied and are described in many textbooks [1]. Whether mutations will affect gene function depends on where they occur within a gene or whether they affect levels of gene expression, mRNA splicing or protein composition. About 70% of the mutations that result in an amino acid change in a protein is estimated to be harmful [2].

A first type of mutation occurs due to the incorporation of a wrong base during replication. This can occur spontaneously or can be chemically induced and will result in a mismatch between base pairs. If this is not repaired, it will lead to a mutation in the next round of replication. Misincorporations lead to transitions in which a pyrimidine is changed into the other pyrimidine or a purine into the other purine and to transversions in which a pyrimidine is changed into a purine or vice versa.

Other types of mutations include insertions or deletions of 1 or more base pairs into the DNA sequence. Mutations can also involve amplifications or deletions of large chromosomal areas leading to multiple copies or loss of the genes located within this region, respectively [3]. These complex mutations are caused by chromosomal translocations, interstitial deletions and chromosomal inversions and can also result in juxtaposition of previously separate pieces of DNA, potentially even forming functionally distinct fusion genes, with potential harmful effect.

Finally, during mitosis or meiosis, large chromosomal

pieces or even whole chromosomes may be lost during the formation of daughter cells due to missegregation events or chromosome breakage. This may also result in the loss or increase in number of alleles of genes, with deleterious effects for the organism. These types of genotoxicants are non-mutagenic and they may have threshold mechanisms. Typical examples are topoisomerase II poisons or mechanisms based on reactive oxygen species; other mechanisms may lead to aneuploidy [4-6].

DNA damage introduced by endogenous and exogenous sources continuously threatens the structure and integrity of DNA molecules. Due to the deleterious effects that mutations can have on the functioning of gene products, organisms have evolved complex DNA repair mechanisms that are aimed to prevent or correct mutations by reverting the mutated sequence back to its original state before they are fixed as permanent genetic changes [7]. If DNA lesions remain unrepaired or if they are erroneously repaired, they will most probably hamper crucial cellular processes like transcription and replication, which may result in cell cycle arrest and possibly (programmed) cell death. However, potentially the mutation may become fixed. At the level of the organism, mutations in germ cells can result in genetically inherited diseases, while accumulation of mutations in somatic cells is associated with the initiation and progression of cancer. DNA damage is also believed to be involved in aging [8].

## Mutations and Pathogenesis

In this short overview paper we focus on the implications of mutations for humans. The importance of DNA and the impact of mutations and chromosomal aberrations for human health are clear from their observed roles in numerous genetic diseases and cancer. Therefore potential for mutations in both somatic and germs cells need to be taken into account when an overall risk estimate resulting from mutations is concerned [9].

Many genetic disorders (e.g. cystic fibrosis) are recessive in nature, indicating that the underlying mutations are expressed when an affected individual inherits the mutated gene from both parents. Disorders showing a dominant inheritance pattern require only 1 copy of the mutated gene in order to become expressed in the affected individual. For dominant genetic disorders a single mutation is sufficient for disease expression, and therefore new mutations make a larger contribution to the incidence of dominant diseases [9]. However, the occurrence of a mutation can be obscured by incomplete penetrance of the affected phenotype.

Structural chromosome aberrations can cause miscarriage and fetal death or serious abnormalities [10]. Aneuploidy, which is defined as the gain or loss of one or more chromosomes contributes to perinatal death, but also causes disorders such as trisomy 21. The major-

ity of chromosomal abnormalities result from trisomies and occur prenatally [11]. As exposure of the developing infants to specific chemicals has been shown to induce birth defects, it is important to sufficiently screens new compounds for potential teratogenic effects prior to release to the market [12].

The correlation between the mutagenicity and carcinogenicity of chemicals has long been evident [13]. Another demonstration of this association is presented by the correlation of human chromosome instability syndromes and DNA repair deficiencies, with increased cancer risks [14]. Molecular studies of so-called proto-oncogenes and tumor suppressor genes have also shown that mutations play an important role in the processes leading to cancer. These genes encode for proteins that typically help to control cell growth and proliferation. Proto-oncogenes (e.g. *ras*) can be converted into active oncogenes by point mutations [15] or chromosomal aberrations, and subsequently stimulate the transformation of normal cells into cancer cells [16]. A tumor suppressor gene (e.g. *p53*) typically protects a cell by inhibiting cell proliferation and tumor development, and mutational inactivation or deletion of these genes have been implicated in human tumors [17]. Many cancers involve both activation of oncogenes and inactivation of tumor suppressor genes, as well as other possible events. This implies that carcinogenesis is a multistep process resulting from an accumulation of genetic alterations [18-20].

Hazard assessment for genetic toxicity is done by means of testing and non-testing methods as will be discussed in this overview. Also available human data is important for hazard assessment. This is closely related to the hazard assessment for carcinogenicity, at least for genotoxic carcinogens.

Non-genotoxic carcinogens, will not be positive in genotoxicity studies because they act via other mechanisms (e.g. peroxisome proliferator-induced hepatocarcinogenesis in rodents; [21] and thus need other testing methods. However this is out of scope of this review.

Risk assessment for genotoxicity is not as easily performed as for carcinogenicity. The usual approach is to combine rodent germ and somatic cell data for induced genetic alterations with human somatic cell data (if available) in order to estimate the frequency of mutations in human cells. For a complete estimate of genetic risk, it is then necessary to obtain an estimate of the frequency of genetic alterations that are transmitted to the offspring [22].

## Chemical Mutagens and Classification

Reactive chemicals that can induce covalent bond formation with endogenous molecules such as DNA can act as a rate determining step in mutagenesis. These types of reactive chemicals can sometimes also affect other human health endpoints such as skin sensitiza-

tion and target organ toxicity. Reactive chemicals are characterized by the presence of electrophilic groups, that can react with nucleophilic sites in DNA or proteins. Some examples of typical substances causing DNA binding are alpha, beta unsaturated aldehydes, hydrazines, nitrogen mustard and polyaromatic hydrocarbons [23]. The representative nucleophilic sites in DNA are the bases guanine, adenine, thymine, and cytosine, whereas in proteins the amino acids, lysine, histidine, and tyrosine show most nucleophilic activity [24]. Schultz, et al. [25] described a conceptual framework for predicting the toxicity of reactive chemicals for which the initiating events were based on covalent reactions with nucleophiles in proteins or DNA and that could lead to a variety of different adverse outcomes, including mutagenicity and sensitisation, but also target organ toxicity such as hepatocyte cytotoxicity or respiratory toxicity. These measures of intrinsic reactivity are suitable to be modelled as endpoints for QSARs rather than the toxicity endpoints that they reflect [25].

Classification of chemicals for germ cell mutagenicity can be done according to the latest 'CLP' European classification system [26]. Based on the classification, manufacturers have to label their substances and take appropriate technical and personal protective measures for their workers. Classification for germ cell mutagenicity may also trigger the Regulatory Authorities such as ECHA and Member States to list and evaluate the substance as a very high concern substance. 'Category 1' mutagens are substances that are known to induce heritable mutations or that are to be regarded as if they can induce heritable mutations in human germ cells. These substances can be further classified into sub-class 1A (classification based on positive human epidemiological studies) or 1B, if the classification is based on positive results from testing. The positive test results can be from (1) *In vivo* heritable germ cell mutagenicity tests in mammals; (2) *In vivo* somatic cell mutagenicity tests in mammals, in combination with other evidence that the substance has potential to cause mutations to germ cells; or (3) From data showing mutagenic effects in human germ cells, without demonstration of transmission to progeny (e.g. increased aneuploidy in sperm cells of exposed people).

Substances are classified as 'Category 2' mutagens if there is concern that the substance may induce heritable mutations in human germ cells. The concern can be based on positive evidence that is obtained from animal experiments and/or sometimes from *in vitro* experiments in somatic cells. Substances that are positive in *in vitro* mammalian mutagenicity assays, and that also show chemical structure activity relationships to known germ cell mutagens, will also be considered for classification as Category 2 mutagens.

According to latest Guidance on classification of chemicals [27], the classification of individual substanc-

es shall be based on the total weight of evidence available, using expert judgement. A single, well-conducted test may be used for classification if it gives clear and unambiguously positive results. Newly developed and validated tests may also be used in the total weight of evidence. The relevance of the route of exposure used in the study compared to the most likely route for human exposure is also taken into account [27]. When there is only evidence of somatic cell genotoxicity, it is also warranted that the substance is also regarded as suspected germ cell mutagen. Classification as a suspected germ cell mutagen may also have implications for potential carcinogenicity classification. This also holds true for genotoxicants which are incapable of causing heritable mutations, because they act locally and cannot reach the germ cells. This implies that if positive results *in vitro* are supported by at least one positive local *in vivo*, somatic cell test, that this is considered as enough evidence to lead to Category 2 classification. If there is also negative or equivocal data available, a weight of evidence approach using expert judgement has to be applied [27].

### Testing for Mutagenicity Potential According to Regulatory Context

For regulatory purpose, the mutagenicity potential of a chemical has mainly been evaluated using *in vitro* assays, in particular the Bacterial Reverse Mutation Test [28] and the Mammalian Cell Gene Mutation Assay [29], as described in Table 1.

The Ames test is a short-term bacterial reverse mutation assay that is designed to detect a wide range of chemical substances that can produce genetic damage that leads to gene mutations. In this assay, mutations present in bacterial test strains are reverted in the presence of mutagenic agents, thereby restoring the functional capability of the bacteria to synthesize an essential amino acid (histidine) and to grow colonies [30]. This assay detects point mutations, including substitution, addition or deletion of one or a few DNA base. Different *Salmonella* strains are used that have different mutations in the histidine operon, each of which is designed to be responsive to mutagens that act via different mechanisms [28]. Over the years, several modifications of the test have been done in order to increase the number of applications possible and in order to allow a testing of a wider range of chemicals (including gases,..). The assay is rapid, inexpensive and relatively easy to perform. However the test utilises prokaryotic cells, which differ from mammalian cells in such factors as uptake, metabolism, chromosome structure and DNA repair processes. An exogenous source of metabolic activation (e.g. Arochlor 1254 induced rat liver S-9 fraction) is therefore commonly used as well, but such systems cannot entirely mimic the mammalian *in vivo* conditions. The test is typically employed as an initial screen for genotoxic activity and, in particular, for point mutation-inducing

**Table 1:** Major *in vitro* and *in vivo* testing models for mutation.

Test model	Test guidance (Adopted)	Design (taken from the OECD test guidance)
Bacterial reverse mutation test	OECD TG 471 [30] - <i>in vitro</i>	Bacterial cell suspensions are exposed to the test substance (liquid or solid) in the presence and in the absence of an exogenous metabolic activation system. At least five different analysable concentrations of the test substance should be used. The recommended maximum test concentration for soluble non-cytotoxic substances is 5 mg/plate or 5 ml/plate. There are two methods: the plate incorporation method and the preincubation method. For both techniques, after two or three days of incubation at 37 °C, revertant colonies are counted and compared to the number of spontaneous revertant colonies on solvent control plates.
Mammalian cell gene mutation test	OECD TG 476 [72] - <i>in vitro</i>	Genetic endpoints measure mutation at Hypoxanthine-Guanine Phosphoribosyl Transferase (HPRT), and at a transgene of Xanthineguanine Phosphoribosyl Transferase (XPRT). The HPRT and XPRT mutation tests detect different spectra of genetic events. Cells in suspension or monolayer culture are exposed to, at least four analysable concentrations of the test substance, both with and without metabolic activation. They are subcultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection. Mutant frequency is determined by seeding known numbers of cells in medium containing the selective agent to detect mutant cells, and in medium without selective agent to determine the cloning efficiency (viability). After a suitable incubation time, colonies are counted.
	OECD TG 490 [73] - <i>in vitro</i>	This TG includes two distinct <i>in vitro</i> mammalian gene mutation assays requiring two specific TK heterozygous cells lines: L5178Y TK ± 3.7.2 C cells for the Mouse Lymphoma Assay (MLA) and TK6 TK ± cells for the TK6 assay. Genetic events detected using the TK locus include both gene mutations and chromosomal events. Cells in suspension or monolayer culture are exposed to, at least four analysable concentrations of the test substance, both with and without metabolic activation. They are subcultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection. Cytotoxicity is usually determined by measuring the relative cloning efficiency (survival) or relative total growth of the cultures after the treatment period. The treated cultures are maintained in growth medium for a sufficient period of time, characteristic of each selected locus and cell type, to allow near-optimal phenotypic expression of induced mutations. Mutant frequency is determined by seeding known numbers of cells in medium containing the selective agent to detect mutant cells, and in medium without selective agent to determine the cloning efficiency (viability). After a suitable incubation time, colonies are counted.
Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays	OECD TG 488 [74] - <i>in vivo</i>	This TG detects gene mutations in transgenic rats or mice that contain multiple copies of chromosomally integrated plasmid or phage shuttles. The transgenes contain reporter genes for the detection of various types of mutations induced by test substances. Administration is done 28 consecutive days followed by a 3-day recovery period, during which unrepaired DNA lesions are fixed into stable mutations. Genomic DNA is isolated from the tissue (s), and mutations are scored by recovering the transgene and analysing the phenotype of the reporter gene in a bacterial host deficient for the reporter gene. Mutant frequency is calculated by dividing the number of plaques/plasmids containing mutations in the transgene by the total number of plaques/plasmids recovered from the same DNA sample.

activity [31,32]; Gatehouse, et al., 1994) [31-33]. It is widely accepted that positive results in the Ames test correlate well with carcinogenic potential, at least in rodents (70%-90% predictivity; [28]. In certain situations the mutagenic response may be specific to the bacterial test and therefore be false positive [34,35].

The mammalian cell gene mutation test applies mammalian cell lines such as (L5178Y) mouse lymphoma cells, CHO AS52 and V79 lines of Chinese hamster cells, and TK6 human lymphoblastoid cells. In these cell lines 'forward' mutations are typically measured at the Thymidine Kinase (TK) locus and Hypoxanthine-Guanine Phosphoribosyl Transferase (HPRT) locus, and at a transgene of Xanthineguanine Phosphoribosyl Transferase (XPRT) locus. The TK, HPRT and XPRT mutation tests detect different spectra of genetic events, because the autosomal location of TK and XPRT may allow the detec-

tion of genetic events (e.g. large deletions) that are not detected at the X-linked HPRT locus [36]. These tests are used to screen for possible mammalian mutagens and carcinogens and thus complement the bacterial mutation assays. Many compounds that are positive in these tests are mammalian carcinogens, however, also in this case, the correlation is not perfect [37].

If there is a positive result in any of the *in vitro* genotoxicity studies and there are no results available from an *in vivo* study, an appropriate *in vivo* somatic cell genotoxicity study should be performed to exclude or confirm the potential mutagenicity for humans. For chemicals, these information requirements are described [38]. If there is a positive result from an *in vivo* somatic cell study available, the potential for germ cell mutagenicity should be further investigated.

**Table 2:** Major *in vivo* and *in vitro* testing models for chromosome damage.

Test model	Test guidance (Adopted)	Design (taken from the OECD test guidance)
Mammalian chromosome aberration test	OECD TG 475 [39] - <i>in vivo</i>	Animals are exposed to the test substance (liquid or solid) by an appropriate route of exposure and are sacrificed at appropriate times after treatment. Prior to sacrifice, animals are treated with a metaphase-arresting agent. Chromosome preparations are then made from the bone marrow cells and stained, and metaphase cells are analysed for chromosome aberrations. Each treated and control group must include at least 5 analysable animals per sex. The limit dose is 2000 mg/kg/body weight/day for treatment up to 14 days, and 1000 mg/kg/body weight/day for treatment longer than 14 days.
	OECD TG 473 [44] - <i>in vitro</i>	The <i>in vitro</i> chromosome aberration test may employ cultures of established cell lines, cell strains or primary cell cultures. Cell cultures are exposed to the test substance (liquid or solid) both with and without metabolic activation during about 1.5 normal cell cycle lengths. At least three analysable concentrations of the test substance should be used. At each concentration duplicate cultures should normally be used. At predetermined intervals after exposure of cell cultures to the test substance, the cells are treated with a metaphase-arresting substance, harvested, stained. Metaphase cells are analysed microscopically for the presence of chromosome aberrations.
Mammalian micronucleus test	OECD TG 474 [46] - <i>in vivo</i>	Animals are exposed to the test substance by an appropriate route. Bone marrow and/or blood cells are collected, prepared and stained. Preparations are analyzed for the presence of micronuclei. Each treated and control group must include at least 5 analysable animals per sex. Administration of the treatments consists of a single dose of test substance or two daily doses (or more). The limit dose is 2000 mg/kg/body weight/day for treatment up to 14 days, and 1000 mg/kg/body weight/day for treatment longer than 14 days.
	OECD TG 487 [48] - <i>in vitro</i>	Cell cultures of human or other mammalian origin are exposed to the test chemical both with and without an exogenous source of metabolic activation. During or after exposure to the test chemical, the cells are grown for a period sufficient to allow chromosome damage or other effects on cell cycle/cell division to lead to the formation of micronuclei in interphase cells. For induction of aneuploidy, the test chemical should ordinarily be present during mitosis. Harvested and stained interphase cells are analysed for the presence of micronuclei. Ideally, micronuclei should only be scored in those cells that have completed mitosis during exposure to the test chemical or during the post-treatment period, if one is used. For all protocols, it is important to demonstrate that cell proliferation has occurred in both the control and treated cultures, and the extent of test chemical-induced cytotoxicity or cytostasis should be assessed in all of the cultures that are scored for micronuclei.

Testing for chromosome damage potential has traditionally been evaluated using *in vivo* assays such as the chromosome aberration and micronucleus test methods. However it has been demonstrated that these *in vivo* methods can be replaced by *in vitro* assays, as described in Table 2.

The *in vivo* mammalian chromosome aberration test is used to detect structural chromosome aberrations that are induced by test compounds in bone marrow cells of rats, mice or Chinese hamsters [39]. While the majority of genotoxic chemical-induced aberrations are of the chromatid-type, chromosome-type aberrations also occur. Chromosomal damage and related events are the cause of many human genetic diseases and there is substantial evidence that, when they cause alterations in oncogenes and tumour suppressor genes, they are involved in cancer [40-43]. The *in vitro* chromosome aberration test is used to identify agents that cause structural chromosome aberrations in cultured mammalian somatic cells [44] and is usually performed with and without metabolic activation of the test substance. Both *in vivo* and *in vitro* chromosome aberration tests are not designed to measure aneuploidy.

In the micronucleus assay, chromosome fragments

(induced by chromosome breaks) and whole chromosomes (occurring due to faults in the segregation of chromosomes to daughter cells) can be detected as an extra cytoplasmic body. Subsequent to the effect induced by the substance, a chromosome fragment or a whole chromosome will become visible as a small nucleus. The micronucleus assay is thus capable of detecting clastogenic (chromosome breakage) as well as aneugenic substances. Using fluorescence *in situ* hybridisation methods or flow cytometry techniques clastogenic and aneugenic actions can be discerned. Several forms of micronucleus assays (mononucleate and binucleate assays) have been developed [45]. The *in vivo* mammalian micronucleus test is used to detect cytogenetic damage induced by test substances to the chromosomes or the mitotic apparatus of erythroblasts, sampled in bone marrow or peripheral blood of mice or rats [46]. When a bone marrow erythroblast develops into a polychromatic erythrocyte or reticulocyte, the main nucleus is extruded; however, any micronucleus that has been formed will remain behind in the cytoplasm and can be scored. An increase in the frequency of micronucleated immature erythrocytes is an indication of induced structural and/or numerical chromosomal aberrations. Visual scoring by a microscope can be easily automated, which

is also a big advantage of this assay [47]. The *in vitro* Micronucleus Test (MNvit) also detects micronuclei in the cytoplasm of interphase cells that have undergone cell division during or after exposure to a test substance [48]. Several cells such as human lymphocytes and Syrian hamster embryo cells, and cell lines (CHO, V79, L5178Y,...) are typically used. Also here different methodologies can be performed with and without metabolic activation of the test substance [6]. In the micronucleus assay, the slides can be scored relatively quickly and analysis can also be automated, which makes it practical to score thousands instead of hundreds of cells per treatment, increasing the power of the test.

Other assays are sometimes used for testing for (1) DNA damage and repair [49,50] guidelines for the Unscheduled DNA Synthesis (UDS) test *in vitro* and *in vivo*, (2) For measuring DNA strand breaks in eukaryotic cells [51] guidelines for the *in vivo* Mammalian Alkaline Comet Assay and (3) For demonstration of germ cell gene mutation (Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays). These assays have been relatively frequently been used in the past, but are nowadays typically used in situations in which an additional assay is needed to help to resolve a potential mutagenicity issue of a substance.

An *in vivo* assay with potential for regulatory safety assessment is the Pig-A gene mutation assay [52]. This assay is capable of measuring *in vivo* gene mutations at the Pig-A gene, which cause failure of 'GPI anchors', resulting in the absence of specific surface markers on the exterior of peripheral blood cells. Using flow cytometry the frequency of cells without these surface markers is easily performed, requiring only microliter volumes of blood. Other existing and modified *in vitro* assays like the more high throughput Ames II assay [53], the Vito-tox assay [54], can be used for specific purposes, however the discussion of all possible models is beyond the scope of this work.

For pharmaceutical development, an initial test battery is needed that comprises the bacterial reverse mutation (e.g. Ames) test and an *in vitro* test either for chromosomal damage or for mammalian gene mutation (e.g. Mouse Lymphoma TK assay). In addition, an *in vivo* test using rodent hematopoietic cells, either for detection of micronuclei or for chromosomal aberrations in metaphase cells is required [55]. Proposals have been made that an *in vitro* genotoxicity testing core battery should be sufficient by a combination of the Ames test and the *in vitro* Micronucleus Test (MNvit), since the latter assay detects both chromosomal aberrations and aneuploidy [56].

For chemicals, the requirements on mutagenicity are described by REACH (Annexes VI-XI) that specifies the information per tonnage band that must be submitted for registration and evaluation purposes. The bacterial reverse mutation (Ames) test is the basic requirement

from 1 tonne/year onwards, followed by an *in vitro* chromosome aberration study in mammalian cells or an *in vitro* micronucleus study for substances from 10 tonnes/year onwards. If these studies are negative, an additional *in vitro* gene mutation study in mammalian cells is required for chemicals from 10 tonnes/year onwards. Only for the higher tonnage bands of 100 and 1000 tonnes/year, an *in vivo* somatic cell genotoxicity study may have to be provided by the registrant if there is a positive *in vitro* study, to clear out the results [38].

### **In silico Assessment of Mutagenicity Potential**

A next evolution in the mutagenicity assessment of compounds is the use of *in silico* methods. These are mainly (quantitative) structure activity relationship models and expert systems that combine physico-chemical or structural properties (descriptors) and computational tools to assign a substance to a certain category or biological/toxicological activity. In other words, based on chemical class or mechanistic reactivity domain, substances can be predicted positive or negative for a certain toxicological endpoint (e.g. skin sensitisation, acute toxicity, mutagenicity,...).

The results of validated (Q)SARs and expert judgement may also be taken into account for classification and labelling. To support the self-classification process for companies to meet their obligation to self-classify the chemicals they import or produce, the Danish EPA published an advisory list for self-classification of dangerous substances [57]. The list of suggested hazard classifications was derived from (Q)SAR prediction models (obtained or developed by the Danish EPA) for various endpoints, including mutagenicity. The (Q)SAR models were used to make predictions for the numerous discrete organic substances in the EINECS, and the updated Danish Advisory List contains the results of a systematic assessment of 49292 discrete organic substances [57].

(Q)SAR models can either include expert knowledge (e.g., Derek nexus, Toxtree) or be fully based on the statistical evaluation of structure fragments determined in a model training set (e.g., CASE Ultra, VEGA, TOPKAT). There is also the OECD (Q)SAR toolbox, which provides mechanistic information along the adverse outcome pathway from which substance-specific (Q)SAR models can be created [58].

In Table 3, the two main types of models are described by means of some examples, however more models are available. Models are either mechanism- (or rule-) based or they are derived empirically using statistical approaches. Some molecules may not fall into the specified applicability domain of the (Q)SAR models. If that happens, the predictions are unreliable and other (Q)SARs or different approaches need to be applied to predict the mutagenicity, chromosome aberration or carcinogenicity potential.

**Table 3:** Examples of *in silico* models for mutagenicity.

Model	Example	Design
Knowledge- or rule-based Structure-Activity Relationship (SAR)	Toxtree (public model)	Toxtree includes modules for mutagenicity, carcinogenicity, and the <i>in vivo</i> micronucleus assay Serafimova [55]. The model includes a decision tree for assessment of mutagenicity and carcinogenicity potential by discriminant analysis and structural rules Benigni [62]. The model also includes a decision tree for the <i>in vivo</i> micronucleus assay. The accuracy of prediction is 70% for carcinogenicity, 78% for mutagenicity and 59% for the <i>in vivo</i> micronucleus assay.
	Derek Nexus (Commercial model)	This (rule-based) model contains structural alerts for mutagenicity, chromosome damage and carcinogenicity. The hazard assessment is justified with literature references. Advantages are the transparency in predictions, the rule development is peer-reviewed by a user group, and new rules can be added easily. The absence of a predicted hazard simply means that no relevant alerts were identified; it does not necessarily mean the absence of hazard Serafimova [55].
Quantitative or statically based Structure-Activity Relationship (QSAR)	CASE Ultra (Commercial model)	This model applies a statistical approach that automatically identifies molecular substructures that have a high probability of being relevant to the observed endpoint. Genotoxicity models include Ames mutagenicity, direct mutagenicity, base-pair mutagenicity, frameshift mutagenicity, chromosomal aberrations, mouse micronucleus assay, mouse sister chromosomal exchange. Carcinogenicity models include rat, mouse, female, male carcinogenicity, TD50 rat, mouse carcinogenicity Serafimova [55].
	VEGA (Public model)	A statistical model for mutagenicity that was originally developed in the frame of the EU CAESAR project. The authors reported correct classification rates of 92.3% and 83.2% for the training and test sets, respectively. The model was combined with Toxtree, showing that the number of false negatives could be reduced, but the number of the false positives increased.  The authors concluded that by using the so-called "cascade model", a classification accuracy close to the reliability of the Ames test data could be achieved Serafimova [55].  In the same project, two complementary approaches (regression and classification) were applied to develop models for carcinogenicity and an accuracy of classification of 91%-96% for the training set and 68%-74% for the test set was reported Serafimova [55].

A large experimental Ames assay mutagenicity data set, comprising about 6500 non-confidential compounds, was compiled and made publicly available [59]. A comparison was done of the predictive performance of three commercial software packages (Derek, MultiCASE, and an off-the-shelf Bayesian machine learner in Pipeline Pilot) with four non-commercial machine learning implementation tools (Support Vector Machines, Random Forests, k-Nearest Neighbours, and Gaussian Processes). In this evaluation, Pipeline Pilot, trained on the developed Ames data set, showed the best predictivity of the three commercial tools followed by MultiCASE. The expert system Derek gave the lowest sensitivity and specificity of all considered models. However, closer examination of the results revealed that the difference in sensitivity between the best commercial model and the best non-commercial machine learning approach is just a few percent, making it difficult to draw conclusions regarding robustness of the software tools. In general, machine learning algorithms are expected to perform better in cases such as this, because they draw their knowledge exclusively from the training data, as opposed to models such as Derek, which has rules derived from other datasets or based on expert knowledge. This study was very useful in terms of the mutagenicity dataset which was made publicly available, but also because it demonstrates the power of machine learning approaches for model discovery and development.

For pharmaceutical impurities, hazard assessment also involves an initial analysis of actual and potential impurities by conducting database and literature searches for carcinogenicity and bacterial mutagenicity data. If such data are not available, a (Q)SAR assessment that focuses on bacterial mutagenicity prediction should be performed. Two (Q)SAR prediction methodologies, one expert rule-based and one statistical-based, that complement each other should be applied [60]. (Q)SAR models utilizing these prediction methodologies should follow the general validation principles that are set forth by the Organisation for Economic Co-operation and Development (OECD). In the case that the two complementary (Q)SAR methodologies indicate that structural alerts for mutagenicity are absent, it is sufficient to conclude that the impurity is not of mutagenic concern, and that no further testing is needed. If warranted, the outcome of any computer system-based analysis can be supplemented with additional expert knowledge in order to increase the supportive evidence on relevance of any positive, negative, conflicting or inconclusive prediction and to provide a rationale to support the final conclusion [60]. A (Q)SAR assessment for chromosomal aberration or carcinogenicity is not required. The focus is on DNA reactive substances causing DNA damage and mutations and possibly when present at low levels. Other types of genotoxicants that are non-mutagenic typically have threshold mechanisms and usually do not

pose carcinogenic risk in humans at the level ordinarily present as impurities.

For chemicals, non-testing (i.e. not performing *in vitro* or *in vivo* experiments) information about the mutagenicity can be used as weight-of-evidence in a variety of ways, ranging from simple inspection of the chemical structure through read-across techniques, expert systems, metabolic simulators, or (Q)SARs. Generally, (Q)SAR models that contain putative mechanistic descriptors are preferred. Predictivity must be assessed case-by-case on the basis of clear documentation and a QSAR Model Reporting Format (QMRF) and a QSAR Prediction Reporting Format (QPRF) have been developed by the Joint Research Centre (JRC) of the European Commission to ensure consistency in summarising and reporting key information on (Q)SAR models and substance specific predictions generated by (Q)SAR models [38].

### Reliability of (Q)SAR Applications for Mutagenicity Prediction

Many investigations have focused on the ability of computational models to accurately predict the mutagenic effects of a chemical in the Ames assay, either individually or in combination [61]. Typically for these investigations, large numbers of compounds are evaluated with different *in silico* tools to predict their mutagenic potential, with no human interpretation being made. The use of multiple models in consensus to enhance prediction is also widely accepted [62]. However, this does not indicate how these assessments are effectively conducted in practice across the (pharmaceutical) industry.

Eight companies were surveyed for their success rate in prediction of mutagenicity with *in silico* tools [63]. The Negative Predictive Value (NPV) of the *in silico* approaches was 94%. When human interpretation of *in silico* model predictions was conducted, the NPV increased substantially to 99%. The survey clearly illustrated the importance of expert interpretation of *in silico* predictions. In addition, the survey also suggested that the use of multiple computational models is not a significant factor in the success of the *in silico* approaches with respect to NPV [63]. A new tool that can help in the (*in silico*) interpretation of mutagenicity of a chemical is the ToxRead software [64]. ToxRead attempts to standardize the read across process using ad hoc visualization and data search methods. Based on similarity measures and fragment searches, the software constructs a picture that contains all relevant information that is useful to make an assessment of the mutagenic potential of the substance. In this picture the substance under evaluation is compared to similar compounds, which are known to be mutagenic or not. Structural alerts that are linked to the evaluated chemical are also included in the picture, thereby summarizing a lot of information into a picture. By clicking structural alerts and compounds in the picture, additional information can be obtained in tabular form.

An evaluation of (Q)SAR models for the prediction of mutagenicity (and carcinogenicity) was performed by [65] for models for congeneric chemical classes and Structure Alert (SA) models. QSARs that estimated the potency of congeneric chemicals were 30% to 70% correct, when evaluating chemicals that were not considered by the developers of the models. The models that discriminated active and inactive chemicals were considerably more accurate (70% to 100%). It was shown that the internal cross-validation statistics were poorly correlated with external validation statistics using chemicals that were not part of the initial dataset used for constructing the QSAR. The genotoxic-based SA models had an accuracy of about 75% for *Salmonella* mutagens. However, the SA models were not able to discriminate active and inactive chemicals within individual chemical classes well and they seem better suited for preliminary or large-scale screenings. Overall it was concluded that the (Q)SAR-based tools are not meant to be black-box machines for predictions, but that they can play an important role. QSARs are able to significantly enrich for safer chemicals, contribute to the organization and rationalization of data, elucidate mechanisms of action, and complement data that is derived from other sources [65].

The FDA's 2008 draft guidance on genotoxic and carcinogenic impurities in drug substances and products allows the use of QSARs to identify structural alerts for known and expected impurities present at levels below qualified thresholds [66]. A Leadscope model predicting *Salmonella t.* mutagenicity (Ames assay) of drug impurities and other chemicals was described for regulatory use. A high sensitivity (81%) and high negative predictivity (81%) was obtained, based on external validation with 2368 compounds foreign to the model and having known mutagenicity. A database of drug impurities was created from proprietary FDA submissions and from the public literature, which found a significant overlap between the structural features of drug impurities and training set chemicals in the QSAR model. The predictive performance of the model was found to be acceptable for screening drug impurities for *Salmonella* mutagenicity [66].

For evaluation of impurities in drug substances and products, computational toxicology assessment should be performed using (Q)SAR methodologies that predict the outcome of a bacterial mutagenicity assay [67]. Two (Q)SAR prediction methodologies that complement each other should be applied, one methodology should be expert rule-based and the second should be statistical-based. The absence of structural alerts for mutagenicity from two complementary (Q)SAR methodologies is sufficient to conclude that the impurity is of no mutagenic concern, and that no further testing is recommended. If warranted, the outcome of any computer system-based analysis can be reviewed with the use of expert knowledge (including for example the use



of tools like ToxRead) in order to provide additional supportive evidence on relevance of any positive, negative, conflicting or inconclusive prediction and provide a rationale to support a final conclusion [60].

The performance of public (Q)SAR models predicting Ames genotoxicity was further assessed for a selection of chemicals registered under REACH regulation [68]. The Benigni/Bossa rule base originally implemented in ToxTree and re-implemented within VEGA displayed the best performance (accuracy = 92%, sensitivity = 83%, specificity = 93%, MCC = 0.68) indicating that this rule base provides a reliable tool for the identification of genotoxic chemicals.

## Further Evolutions

Various initiatives are being taken to integrate (Q)SAR methods with other Weight of Evidence. A ChemTunes prediction system was implemented by the Chemical Evaluation and Risk Estimation System (CERES) at the U.S. FDA's Office of Food Additive Safety (OFAS) for review and monitoring of food ingredients and packaging materials. The system allows access to expansive, reliable experimental data, and predicts results based on both expert knowledge and mode-of-action driven QSAR models. The prediction results from local and global models as well as chemotype rules (alerts) are combined to predict the outcome while minimizing the overall uncertainty [69]. When tested on a more extensive databased of 3950 compounds, validation results showed an average Negative Predictive Value (NPV) of 92.6% [70].

Another study was performed to combine QSAR data with literature on carcinogenic modes of action automatically generated by a text-mining tool. A group so 96 rat carcinogens pointed out that skin and lung rat carcinogens were mainly mutagenic, whereas hematopoietic and liver carcinogens included a large proportion of non-mutagens. Mechanisms such as immunosuppression and hormonal receptor-mediated effects were involved. The method can be particularly useful in increasing the understanding of structure and activity relationships for non-mutagens [71].

## Final Conclusions

Genotoxicity hazard identification and risk assessment is important to assure the safety of substances that humans can be exposed to. Several implemented testing methods that are routinely performed have been reviewed here. This review is not meant to be complete and covering all testing methodologies, but rather to give an overview of main methodologies (*in silico*, *in vitro* and *in vivo*) that are used for a regulatory context.

In order to reduce the number of animals sacrificed for testing of the numerous numbers of chemical substances that need better characterisation for their mu-

tagenic potential, *in silico* and *in vitro* are playing an increasingly important role. Especially *in silico* tools like QSARs are promising as they are non-testing methods, thereby reducing cost and time to perform *in vitro* assays. However, more work on improving reliability of these non-testing tools needs to be done in order for them to become fully accepted alternatives. (Q)SARs are not meant as stand-alone tools; they should be done either in a battery of different (rule- and statistically based) QSAR models, associated with expert judgment or in combination with other Weight of Evidence (WoE) such as *in vitro* data and *in vivo* data from other similar molecules (read-across). In this context, they can be used more and more for various product categories as an Integrated Testing Strategy (ITS) in order to avoid and replace animal testing. For some product (sub) categories such as pharmaceuticals and chemicals, testing (*in vitro/in vivo*) will be unavoidable, however future may bring more reliable approaches for ensuring human safety. For other product types, such as impurities, a combination of (Q)SARs for bacterial mutagenicity is acceptable to assess potential genotoxic impurities if no data are available. Newer models also integrate more Weight of Evidence information, either by means of literature data or mechanism of action.

## References

1. Strachan T Read, Sutter A (2010) Human Molecular Genetics. (4<sup>th</sup> edn), Use of *in silico* systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul Toxicol Pharmacol 67: 39-52.
2. Sawyer SA, Parsch J, Zhang Z, Hartl DL (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. Proc Natl Acad Sci U S A 104: 6504-6510.
3. PJ Hastings, James R Lupski, Susan M Rosenberg, Grzegorz Ira (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10: 551-564.
4. Bolt HM, Degen GH (2004) Degen Human Carcinogenic Risk Evaluation, Part II: Contributions of the EUROTOX Specialty Section for Carcinogenesis. Toxicol Sci 81: 3-6.
5. Heidi Foth, Gisela H Degen, Hermann M Bolt (2005) New aspects in the classification of carcinogens. Arh Hig Rada Toksikol 56: 167-175.
6. Kirsch-Volders M, Decordier I, Elhajouji A, Plas G, Aardeva MJ, et al. (2011) *In vitro* genotoxicity testing using the micronucleus assay in cell lines human lymphocytes and 3D human skin models. Mutagenesis 26: 177-184.
7. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem 73: 39-85.
8. Holmes GE, Berstein C, Bernstein H (1992) Oxidative and other DNA damages as the basis of aging a review. Mutat Res 275: 305-315.
9. Preston RJ, Hofmann GR (2008) Genetic Toxicology. In: C D Klaassen Casarett, Doull's, Toxicology The Basic Science of Poisons. (7<sup>th</sup> edn), McGraw Hill Medical, Kansas City, USA, 381-413.

10. Kochkar PK, Ghosh P (2013) Reproductive outcome of couples with recurrent miscarriage and balanced chromosomal abnormalities. *J Obstet Gynaecol Res* 39: 113-120.
11. NRC (National Research Council) (1990) Committee on the Biological Effects of Ionizing Radiations Health Effects of Exposure to Low Levels of Ionizing Radiation. BEIRV, Washington DC.
12. Vargesson N (2015) Thalidomide-induced teratogenesis history and mechanisms. *Birth Defects Res C Embryo Today* 105: 140-156.
13. Ashby J, Tennant RW (1991) Definitive relationships among chemical structure carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res* 257: 229-306.
14. Friedberg EC (1985) DNA Repair. New York, Freeman.
15. Barret JC (1993) Mechanisms of multistep carcinogenesis and carcinogen risk assessment. *Environ Health Perspect* 100: 9-20.
16. Bishop JM (1991) Molecular themes in oncogenesis. *Cell* 64: 235-248.
17. Feng Z, Hu W, Rajagopal R, Levine AJ (2008) The tumor suppressor p53 cancer and aging. *Cell Cycle* 7: 842-847.
18. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61: 759-767.
19. Marshall CJ (1991) Tumor suppressor genes. *Cell* 64: 313-326.
20. Carnero A, Blanco-Aparicio C, Kondoh H, Leonart ME, Martinez-Leal JF, et al. (2015) Disruptive chemicals senescence and immortality. *Carcinogenesis* 36: S19-S37.
21. MS Rao, Jk Reddy (1991) An overview of peroxisome proliferator-induced hepatocarcinogenesis. *Environ Health Perspect* 93: 205-209.
22. UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation) (2001) Report to the General Assembly, with Scientific Annex. Hereditary Effects of Radiation, United Nations, New York.
23. Poirier MC (2012) Chemical-induced DNA damage and human cancer risk. *Discov Med* 14: 283-288.
24. Mekenyan O, Patlewicz G, Dimitrova G, Kuseva C, Todorov M, et al. (2010) Use of Genotoxicity Information in the development of Integrated Testing Strategies (ITS) for Skin Sensitization. *Chem Res Toxicol* 23: 1519-1540.
25. Schultz TW, Carlson RE, Cronin MT, Hermens JL, Johnson R, et al. (2006) A conceptual framework for predicting the toxicity of reactive chemicals modeling soft electrophilicity. *SAR QSAR Environ Res* 17: 413-428.
26. CLP regulation (ec) no 1272/2008 of the European parliament and of the council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing.
27. ECHA - European Chemicals Agency (2013) Guidance on the Application of the CLP Criteria. Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures, Version 4.0.
28. Mortelmans K, Zeiger E (2000) The Ames Salmonella/microsome mutagenicity assay. *Mutat Res* 455: 29-60.
29. Johnson GE (2012) Mammalian cell HPRT gene mutation assay: test methods. *Methods Mol Biol* 817: 55-67.
30. (1997) OECD Guideline for the testing of chemicals Bacterial Reverse Mutation. Guideline for the testing of chemicals. In Vitro Mammalian Chromosome Aberration Test.
31. Ames BN, McCann J, Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the salmonella mammalian-microsome mutagenicity test. *Mutat Res* 31: 347-364.
32. Maron DM, Ames BN (1983) Revised Methods for the Salmonella mutagenicity Test. *Mutat Res* 113: 173-215.
33. Gatehouse D, Haworth S, Cebula T, Gocke E, Kier L, et al. (1994) Recommendations for the Performance of Bacterial Mutation Assays. *Mutat Res* 312: 217-233.
34. McCann J, Choi E, Yamasaki E, Ames BN (1975) Detection of carcinogens in the Salmonella microsome test assay of 300 chemicals. *Proc Natl Acad Sci USA* 72: 5135-5139.
35. Zeiger E (1998) Identification of rodent carcinogens and noncarcinogens using genotoxicity tests premises promises and performance. *Regul Toxicol Pharmacol* 28: 85-95.
36. Moore MM, DeMarini DM, DeSerres FJ, Tindall KR (1997) Banbury Report Mammalian Cell Mutagenesis Cold Spring Harbor Laboratory, New York.
37. Aaron CS, Bolcsfoldi G, Glatt HR, Moore M, Nishi Y, et al. (1994) Mammalian Cell Gene Mutation Assays Working Group Report. *Mutation Res* 312: 235-239.
38. ECHA - European Chemicals Agency (2015) Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. Version 3.1.
39. (2014) OECD Guideline for the testing of chemicals Mammalian Bone Marrow Chromosomal Aberration.
40. Adler ID (1984) Cytogenetic Tests in Mammals. In: Venittand S, Parry JM, Mutagenicity Testing: A Practical Approach. Washington DC, IRL Press, 275-306.
41. Preston RJ, Dean BJ, Galloway S, Holden H, McFee AF, et al. (1987) Mammalian in vivo cytogenetic assays Analysis of chromosome aberrations in bone marrow cells. *Mutat Res* 189: 157-165.
42. Richold M (1990) In Vivo Cytogenetics Assays. Basic Mutagenicity Tests: UKEMS Recommended Procedures: UKEMS Subcommittee on Guidelines for Mutagenicity Testing: Report, Part I revised. In: Kirkland DJ, Cambridge University Press, Cambridge, 115-141.
43. Tice RR, Hayashi M, MacGregor JT, Anderson D, Blakey DH, et al. (1994) Report from the working group on the in vivo mammalian bone marrow chromosomal aberration test. *Mutat Res* 312: 305-312.
44. (2014) OECD Guideline for the testing of chemicals. In vitro Mammalian Chromosome Aberration Test.
45. Doherty AT (2012) The in vitro micronucleus assay. *Methods Mol Biol* 817: 121-141.
46. (2014) OECD Guideline for the testing of chemicals. Mammalian Erythrocyte Micronucleus Test.
47. Hayashi M, MacGregor JT, Gatehouse DG, Blakey DH, Derfingher SD, et al. (2007) In vivo erythrocyte micronucleus assay III. Validation and regulatory acceptance of automated scoring and the use of rat peripheral blood reticulocytes, with discussion of non-hematopoietic target cells and a single dose-level limit test. *Mutat Res* 627: 10-30.
48. (2014) OECD Guideline for the testing of chemicals. In vitro Mammalian Cell Micronucleus Test.
49. (2014) OECD Guideline for the testing of chemicals Genetic Toxicology: DNA Damage and Repair Unscheduled DNA Synthesis in Mammalian Cells in vitro.

50. (1997) OECD Guideline for the testing of chemicals. Unscheduled DNA Synthesis (UDS) Test with mammalian Liver Cells in vivo.
51. (2016) OECD Guideline for the testing of chemicals. In Vivo Mammalian Alkaline Comet Assay.
52. Dobrovolsky VN, Miura D, Heflich RH, Dertinger SD (2010) The in vivo Pig-a gene mutation assay a potential tool for regulatory safety assessment. *Environ Mol Mutagen* 51: 825-835.
53. Flückinger-Isler S, Baumeister M, Braun K, Gervais V, Hasler-Nguyen N, et al. (2004) Assessment of the performance of the Ames II assay: a collaborative study with 19 coded compounds. *Mutation Research Genetic Toxicology and Environmental Mutagenesis* 558: 181-197.
54. Van Der Lelie D, Verschaeve L, Regniers L (2000) Corbisier P Use of bacterial tests (the VITOTOX® genotoxicity test and the BIOMET heavy metal test) to analyse chemicals and environmental samples. In: *New MicroBiotests for Routine Toxicity Screening and Biomonitoring*. 197-207.
55. ICH S2 (R2) (2011) Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use. Step 5: EU: Adopted by CHMP, December 2011, issued as EMA/CHMP/ICH/126642/2008; MHLW, Federal Register, 77: 33748-33749.
56. Kirkland D, Reeve L, Gatehouse D, Vanparys P (2011) A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. *Mutat Res* 721: 27-73.
57. Niemelä JR, Wedebye EB, Nikolov NG, Jensen GE, Ringsted T, et al. (2010) The Advisory list for self-classification of dangerous substances. The Danish Environmental Protection Agency.
58. Serafimova R, Gatnick F, Worth A (2010) Review of QSAR Models and Software Tools for Predicting of Genotoxicity and Carcinogenicity. Publications Office of the European Union.
59. Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, et al. (2009) Benchmark data set for in silico prediction of Ames mutagenicity. *J Chem Inf Model* 49: 2077-2081.
60. ICH M7 Genotoxic impurities (2014) Assessment and Control of DNA reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk.
61. White AC, Mueller RA, Gallavan RH, Aaron S, Wilson AG (2003) A multiple in silico program approach for the prediction of mutagenicity from chemical structure. *Mutat Res* 539: 77-89.
62. Jolly R, Ahmed KB, Zwickl C, Watson I, Gombar V (2015) An evaluation of in-house and off-the-shelf in silico models implications on guidance for mutagenicity assessment. *Regul Toxicol Pharmacol* 71: 388-397.
63. Dobo KL, Greene N, Fred C, Glowienke S, Harvey JS, et al. (2012) In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities an industry survey. *Regul Toxicol Pharmacol* 62: 449-455.
64. Gini G, Franchi AM, Manganaro A, Golbamaki A, Benfenati E (2014) ToxRead: a tool to assist in read across and its use to assess mutagenicity of chemicals. *SAR QSAR Environ Res* 25: 999-1011.
65. Benigni R, Bossa C (2008) Predictivity and reliability of QSAR Models the case of mutagens and carcinogens. *Toxicol Mec Methods* 18: 137-147.
66. Valerio LG, Cross KP (2012) Characterization and validation of an in silico toxicology model to predict the mutagenic potential of drug impurities. *Toxicol Appl Pharmacol* 260: 209-221.
67. Sutter A, Amberg A, Boyer S, Brigo A, Contrera JF, et al. (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. *Regul Toxicol Pharmacol* 67: 39-52.
68. Cassano A, Raitano G, Mombelli E, Fernandez A, Cester J, 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: 273-298.
69. Arvidson K, Rathman J, Volarath P, Mostrag A, Tarkhov A (2014) Evaluation of the chemical inventories in the US FDA's Office of Food Additive Safety for human health endpoints using a toxicity prediction system. U.S. FDA CFSAN Altamira LLC Molecular Networks GmbH Ohio State University, USA.
70. Schwab CH, Rathman JF, Maruszczyk J, Mostrag A, Bienfait B, et al. (2016) A reliable workflow for in silico assessment of genetic toxicity and application to pharmaceutical genotoxic impurities. *SOT Abstract/Toxicology Letters* 258: S54-S61.
71. Papamokos G, Silins I (2016) Combining QSAR Modeling and Text-Mining Techniques to Link Chemical Structures and Carcinogenic Modes of Action. *Front Pharmacol* 7: 284.
72. (2015) OECD Guideline for the testing of chemicals. In vitro Mammalian Gene Mutation Test using the Hprt and xprt genes.
73. (2015) OECD Guideline for the testing of chemicals. In vitro Mammalian Gene Mutation Test using the Thymidine Kinase Gene.
74. (2013) OECD Guideline for the testing of chemicals. Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays.