



The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree

Romualdo Benigni^a, Cecilia Bossa^a, Nina Jeliazkova^b,
Tatiana Netzeva^c and Andrew Worth^c

- ^a Istituto Superiore di Sanita', Environment and Health Department, Rome, Italy**
- ^b Ideaconsult Ltd, Sofia, Bulgaria**
- ^c European Chemicals Bureau, Institute for Health & Consumer Protection, European Commission - Joint Research Centre, Ispra, Italy**

EUR 23241 EN - 2008

The mission of the IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.

The IHCP carries out research to improve the understanding of potential health risks posed by chemical, physical and biological agents from various sources to which consumers are exposed.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information

Address: Toxicology & Chemical Substances Unit, TP 582

E-mail: andrew.worth@ec.europa.eu

Tel.: +39 0332 789566

Fax: +39 0332 786717

[http:// http://ecb.jrc.it/QSAR](http://http://ecb.jrc.it/QSAR)

<http://www.jrc.cec.eu.int>

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

A great deal of additional information on the European Union is available on the Internet.

It can be accessed through the Europa server

<http://europa.eu/>

JRC 43517

EUR 23241 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 2008

Reproduction is authorised provided the source is acknowledged

Printed in Italy

Quality control insert

	<i>Name</i>	<i>Signature</i>	<i>Date</i>
Report Prepared by:	Tatiana Netzeva	<i>T. Netzeva</i>	13/01/08
Reviewed by: (Scientific level)	Andrew Worth	<i>APWorth</i>	31/01/08
Approved by: (Head of Unit)	Steven Eisenreich p.o. Ana Paya Perez		01/02/08
Final approval: (IHCP Director)	Elke Anklam		04/02/08

ABSTRACT

The Joint Research Centre's European Chemicals Bureau has developed a hazard estimation software called Toxtree, capable of making structure-based predictions for a number of toxicological endpoints. One of the modules developed as an extension to Toxtree is aimed at the prediction of carcinogenicity and mutagenicity. This module encodes the Benigni/Bossa rulebase for carcinogenicity and mutagenicity developed by Romualdo Benigni and Cecilia Bossa at the Istituto Superiore di Sanita', in Rome, Italy. The module was coded by the Toxtree programmer, Ideaconconsult Ltd, Bulgaria. In the Toxtree implementation of this rulebase, the processing of a query chemical gives rise to limited number of different outcomes, namely: a) no structural alerts for carcinogenicity are recognised; b) one or more structural alerts (SAs) are recognised for genotoxic or non-genotoxic carcinogenicity; c) SAs relative to aromatic amines or $\alpha\beta$ -unsaturated aldehydes are recognised, and the chemical goes through Quantitative Structure-Activity Relationship (QSAR) analysis, which may result in a negative or positive outcome. If the query chemical belongs to the classes of aromatic amines or $\alpha\beta$ -unsaturated aldehydes, the appropriate QSAR is applied and provides a more refined assessment than the SAs, and should be given higher importance in a weight-of-evidence scheme. This report gives an introduction to currently available QSARs and SAs for carcinogenicity and mutagenicity, and provides details of the Benigni/Bossa rulebase.

LIST OF ABBREVIATIONS

AD	Applicability Domain
ECB	European Chemicals Bureau
FDA	US Food and Drug Administration
ISSCAN	Istituto di Sanità database on chemical carcinogens
QSAR	Quantitative Structure-Activity Relationship
SA	Structural Alert

CONTENTS

1.	Summary of the system.....	1
2.	Carcinogenicity, mutagenicity, and structure-activity relationships.....	2
2.1	Background on the carcinogenicity and mutagenicity of chemicals.....	2
2.2	Structural alerts	3
2.3	Fine-tuned models: Quantitative Structure-Activity Relationships	4
3.	More on structural alerts	6
3.1	Structural alerts and mechanisms of action: examples	6
3.2	Structural alerts and the effects of the molecular environment: modulating factors	9
3.3	Compilations of structural alerts.....	10
3.4	Structural alerts for predictive toxicology	12
4.	Structural alerts included in the Benigni/Bossa rulebase.....	13
4.1	The performance of SAs	14
5.	Quantitative structure activity relationships in the Benigni/Bossa rulebase....	21
5.1	Mathematical models.....	21
5.2	Characterisation of the models.....	22
5.3	Applicability Domain of QSARs	23
6.	Outputs and classification schemes	24
7.	References	26
	Appendix 1: Structural alerts	30
	Appendix 2: QSAR models	63

1. Summary of the system

The Benigni/Bossa rulebase for mutagenicity and carcinogenicity was developed as a module (plug-in) to the Toxtree software (<http://ecb.jrc.it/qsar/qsar-tools/>). The module, which was programmed by the Toxtree programmer, Ideconsult Ltd, Bulgaria, provides the users with a number of models aimed at predicting the carcinogenicity and mutagenicity of chemicals, based on the knowledge of their structure. The main tool is a list of Structural Alerts (SA) for carcinogenicity. The SAs for carcinogenicity are molecular functional groups or substructures known to be linked to the carcinogenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognised, the system flags the potential carcinogenicity of the chemical. The present list of SAs refers mainly to the knowledge on the action mechanisms of genotoxic carcinogenicity (thus they apply also to the mutagenic activity in bacteria), but includes also a number of SAs flagging potential nongenotoxic carcinogens.

Because of their nature, the SAs have the role of pointing to chemicals potentially toxic, whereas no conclusions or indications about nontoxic chemicals are possible (except by exclusion). Thus the SAs are not a discriminant model on the same ground of the Quantitative Structure-Activity Relationships (QSAR) models that produce estimates for both positive and negative chemicals.

In addition to the SAs, this software includes QSAR models for: 1) the mutagenic activity of aromatic amines in the *Salmonella typhimurium* TA100 strain (Ames test); 2) the carcinogenic activity of the aromatic amines in rodents (summary activity from rats and mice); 3) the mutagenic activity of $\alpha\beta$ -unsaturated aldehydes in the *Salmonella typhimurium* TA100 strain (Ames test).

2. Carcinogenicity, mutagenicity, and structure-activity relationships

2.1 Background on the carcinogenicity and mutagenicity of chemicals

Carcinogenicity and mutagenicity are among the toxicological endpoints that pose the highest concern for human health, and are the object of recognised regulatory testing methods (see Annex V to *Directive 67/548/EEC*, <http://ecb.jrc.it/testing-methods/annex5/>).

Historically, the electrophilic theory of chemical carcinogenesis developed by James and Elizabeth Miller (Miller, Miller 1977; Miller, Miller 1981) enabled the activity of the large majority of animal carcinogens known by the 1970's to be tentatively rationalized. In the 1960's, the Millers noted the electrophilicity of several carcinogenic alkylating agents. Since then, a number of acylating agents were found to be carcinogenic, and these chemicals were also electrophilic as administered. Other observations pointed to a variety of chemical carcinogens -of rather different structures- for which metabolism to electrophilic reactants had been demonstrated. Overall, this evidence led them to suggest “that most, if not all, chemical carcinogens either are, or are converted *in vivo* to, reactive electrophilic derivatives which combine with nucleophilic groups in crucial tissue components, such as nucleic acids and proteins” (Miller, Miller 1981).

Following the seminal work of the Millers, distinguished contributions to the advancement and dissemination of the knowledge in this field came from several investigators. Bruce Ames created a series of genetically-engineered *Salmonella typhimurium* bacterial strains, each strain being specifically sensitive to a class of chemical carcinogens (e.g. alkylating, intercalating). The *Salmonella*, or Ames test is an *in vitro* model of chemical carcinogenicity, and consists of a range of bacterial strains that together are sensitive to a large array of DNA damaging agents (Ames 1984) (Maron, Ames 1983) (Zeiger 1987). Since most of the known carcinogens at that time acted through genotoxic mechanisms, the activity of chemicals as mutagens to *Salmonella* almost always seems plausible within the context of the Millers' hypothesis (Ashby, Tennant 1988).

After a number of decades, the hypothesis of the electrophilic reactivity of (many) chemical carcinogens maintains its validity, and has been incorporated into a more general theory on the chemical carcinogens. From the point of view of mechanism of action, carcinogens are classified into: a) genotoxic carcinogens, which cause damage directly to DNA. --many known mutagens are in this category, and often mutation is one of the first steps in the development of cancer (Arcos, Argus 1995); and b) epigenetic carcinogens that do not bind covalently to DNA, do not directly cause DNA damage, and are usually negative in the standard mutagenicity assays (Woo 2003). Whereas the epigenetic carcinogens act through a large variety of different and specific mechanisms, the genotoxic carcinogens have the unifying feature that they are either electrophiles per se or can be activated to electrophilic reactive intermediates, as originally postulated by the Millers.

2.2 Structural alerts

An important contribution came from John Ashby, that contributed to the definition and compilation of a list of Structural Alerts (SA) following the electrophilicity theory of the Millers (Ashby 1985) (Ashby, Tennant 1988). The SAs for carcinogenicity are defined as molecular functional groups or substructures that are linked to the carcinogenic activity of the chemicals. Thus, they identify the major chemical classes potentially able to cause cancer. Since the attack to, and the modification of DNA is the main step in the mechanism of action of many carcinogens (i.e., the so-called genotoxic carcinogens), the SAs relative to such classes of carcinogens are also valid for the mutagenicity endpoint.

Whereas the main and definitive proof that a chemical is a human carcinogen derives from observations in humans collected through epidemiological studies, the large majority of carcinogens have been identified by studies in animals. Rats and mice have been preferred experimental models because of their relatively short life span, the limited cost of their maintenance, their widespread use in pharmacological and toxicological studies, their susceptibility to tumour induction, and the availability of inbred or sufficiently characterised strains (Huff, Haseman, Rall 1991; Fung, Barrett, Huff 1995; Huff 1999). While potentially genotoxic carcinogens can –in principle- be detected by mutagenicity short-term assays, a long term carcinogenicity study has no

substitutes for detecting non-genotoxic carcinogens. Given the preponderance of studies based on experimental animals, the recognition and identification of SAs largely exploits the results of such studies.

It should be emphasized that models based on SAs hold a special place in predictive toxicology. The knowledge on the action mechanisms as exemplified by the SAs is routinely used in Structure-Activity Relationship (SAR) assessment in the regulatory context (see, for example, the mechanistically-based reasoning as presented in (Woo, Lai, McLain, Ko Manibusan, Dellarco 2002)). In addition, the SAs are at the basis of popular commercial (e.g., DEREK, by Lhasa Ltd.) and noncommercial software systems (e.g., Oncologic, by US Environmental Protection Agency <http://www.epa.gov/oppt/newchems/tools/oncologic.htm>).

In contrast to the matured and widespread use of SAs for genotoxic carcinogens, the use of SAs for identifying nongenotoxic carcinogens is limited in scope and still in the early stages of development. Nongenotoxic carcinogens act by a variety of mechanisms with no apparent unifying concept. These mechanisms may be loosely grouped into (a) receptor-mediated, (b) disturbance of homeostatic control, (c) indirect DNA damage, (d) cytotoxicity-induced compensatory cell proliferation, (e) loss of immune surveillance, and (f) loss of intercellular communication. The approaches for (Q)SAR analysis and identification of SAs differ accordingly. A number of SAs and characteristics of several types of nongenotoxic carcinogens have been summarized and discussed by Yin-Tak Woo (Woo 2003).

2.3 Fine-tuned models: Quantitative Structure-Activity Relationships

A set of chemicals characterised by the same SA constitute a family (class) of compounds that share the same mechanism of action. The reactivity of a SA can be modulated or abolished by the remaining part of the molecule in which the SA is embedded. At a coarse-grain level, the modulating effect can be represented by other molecular substructures (e.g., bulky groups *ortho* to an aromatic amine group) that are known to have an influence on the reactivity of the SA. Usually, the knowledge on the modulating substructures is quite limited for most of the SAs, thus it provides limited help in deciding which chemicals in a class of potential e.g., carcinogens will be actually toxic and, *viceversa* which will be not, or poorly toxic. A powerful

generalization is provided by Quantitative Structure-Activity Relationship (QSAR) analysis, which produces a mathematical model that links the biological activity to a limited number of physical chemical or other molecular properties (descriptors) with general relevance. Since most of the descriptors have continuous values, the QSARs provide fine-tuned models of the biological activity, and can give account of subtle differences (for general introductions on QSAR, see (Hansch, Leo 1995))(Franke 1984; Hansch, Hoekman, Leo, Weininger, Selassie 2002; Franke, Gruska 2003).

QSARs have been generated for a number of individual chemical classes of mutagens and carcinogens, including aromatic amines, nitroarenes, quinolines, triazines, polycyclic aromatic hydrocarbons, lactones, aldehydes (Benigni 2005). Some QSARs describe the gradation of potency of active compounds, whereas others are aimed at discriminating between active and inactive compounds. A recent survey on the QSARs for mutagens and carcinogens, performed as a collaboration between the European Chemicals Bureau and the Istituto Superiore di Sanita', has indicated that the models for the potency have a limited reliability, whereas a satisfactory predictivity is shown by the QSARs for discriminating between inactive and active chemicals (Benigni, Bossa, Netzeva, Worth 2007) (http://ecb.jrc.it/documents/QSAR/EUR_22772_EN.pdf), (Benigni, Bossa 2007) .

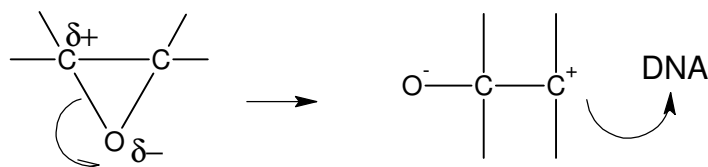
3. More on structural alerts

3.1 Structural alerts and mechanisms of action: examples

Basically, each of the SAs point to a chemical class that provokes toxic effects through one or few commonly shared mechanisms of action.

Among the major structural classes of genotoxic carcinogens are direct-acting carcinogens (including epoxides, aziridines, sulfur and nitrogen mustards, α -haloethers, and lactones). As a representative example, we will focus on the mechanism of action of epoxides. Epoxides exert their carcinogenic potential by alkylating the DNA. In fact, the strained ring system that characterises this chemical class, facilitates the generation of a carbonium ion by the opening of the ring. The carbonium ion may then react with DNA nucleophilic sites to form 2-hydroxy-2-alkyl adducts (Singer, Grunberg 1983).

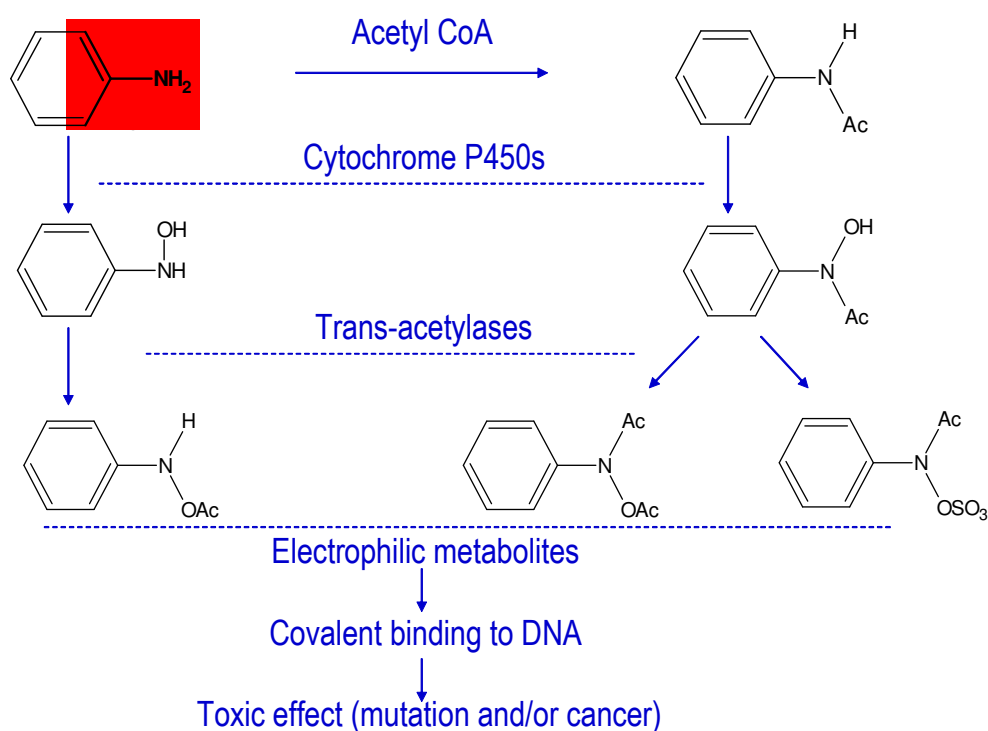
Thus, the SA “epoxide” points to a chemical class, and to a relatively simple mechanism of induction of mutations and cancer.



Other SAs point to classes of genotoxic carcinogens that are inactive as such, and become toxic after metabolic transformation. Due to the complexity of the metabolic machinery, several metabolic pathways may be working at the same time: thus one SA may point to an range of toxic final products (which are nevertheless unified by the fact that all act through genotoxic mechanisms). A widely studied example are the aromatic amines.

The aromatic amines have to be metabolized to reactive electrophiles to exert their carcinogenic potential. For aromatic amines and amides, this typically involves an initial N-oxidation to *N*-hydroxyarylamines and *N*-hydroxyarylamides, which in rat liver is mediated primarily by cytochrome P-450 isozyme *c* (BNF-B) and *d* (ISF-G). Moreover, hydroxylamino, nitro, and nitroso groups are able to generate amine groups (due to metabolic interconversion). The initial activation of nitroaromatic

hydrocarbons is likewise through the formation of an *N*-hydroxyarylamine, a reduction catalyzed by both microsomal and cytosolic enzymes. Microsomal nitroreduction too appears to depend on cytochrome P-450 complex, in particular rat liver isozymes *c*, *d*, *b* (PB-B) and *e* (PB-D). Cytosolic nitroreductase activity is associated with a number of enzymes, including DT-diaphorase, xanthine oxidase, aldehyde oxidase, and alcohol dehydrogenase. In addition to the reactions of nitrogen oxidation and reduction (main activation pathways), certain aromatic amines and nitroaromatic hydrocarbons are converted into electrophilic derivatives through ring-oxidation pathways. *N*-Hydroxyarylamines, iminoquinones, and epoxide derivatives are directly electrophilic metabolites, while *N*-hydroxy arylamides require esterification before becoming capable of reacting with DNA (Benigni 2005) (see below).

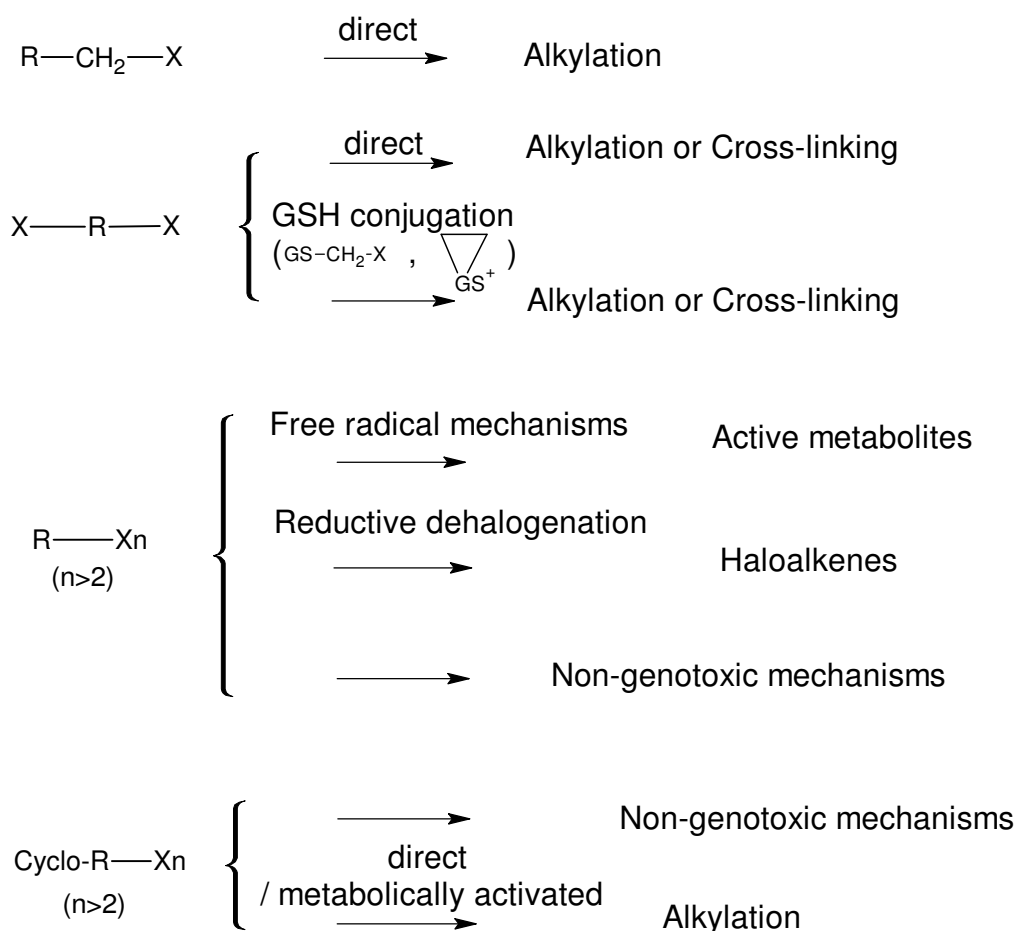


Some chemical classes, like the aliphatic halogens, act by more complicated mechanisms, and consequently are more difficult to be coded through the SAs. In fact, the action mechanisms of aliphatic halogens tend to shift from genotoxic to

epigenetic, with increasing degree of halogenation and depending on the carbon skeleton (linear chains or cyclic structures).

Short-chain monohalogenated alkanes (and alkenes) are potential direct-acting alkylating agents; dihalogenated alkanes are also potential alkylating or cross-linking agents (either directly or after GSH conjugation). Polyhaloalkanes act by free radical or nongenotoxic mechanisms, or may undergo reductive dehalogenation to yield haloalkenes.

For what concerns halogenated cycloalkanes (and cycloalkenes), the mechanism of carcinogenic action is unclear. Several possible epigenetic mechanisms have been proposed which include (i) inhibition of intercellular communication, (ii) degranulation of the rough endoplasmic reticulum, and (iii) hormonal imbalance. In addition, genotoxic mechanisms (i.e., alkylation) are also possible for some of these compounds directly or after metabolic transformation (Woo, Lai, McLain, Ko Manibusan, Dellarco 2002). In these cases, the use of more than one SA may be appropriate.



3.2 Structural alerts and the effects of the molecular environment: modulating factors

Each of the SAs is a “code” for a well-characterised chemical class, with its own specific mechanism of action. However, there are also general factors that may influence the potential reactivity of a chemical, i.e., one could expect to observe compounds with structurally alerting features but which are biologically inactive because of a number of reasons. Among the physicochemical factors that modulate and may hinder the potential biological activity of the chemicals with SAs are: 1) Molecular Weight (MW): chemicals with very high MW and size have little chance of being absorbed in significant amounts; 2) physical state, which influences the capability of the compounds to reach critical targets; 3) solubility: in general highly hydrophilic

compounds are poorly absorbed and, if absorbed, are readily excreted; 4) chemical reactivity: compounds which are “too reactive” may not be carcinogenic because they hydrolyze or polymerize spontaneously, or react with noncritical cellular constituents before they can reach critical targets in cells. Another critical factor is the geometry of the chemical compounds: many potent carcinogens and mutagens (e.g. polycyclic aromatic hydrocarbons, aflatoxin B1, etc...) are planar molecules, with an electrophilic functional group and favorable size, so that they can intercalate properly into DNA.

A practical approach is to consider structural motifs that can code for (at least some of) the above modulating factors, and that diminish or rule out the effect of a SA on the activity of the molecule. Some examples are the following: 1) For the aromatic amino, substituted amino, and nitro compounds, *ortho*-di-substitution or a carboxylic acid *ortho* to the nitrogen substituent are expected to hinder metabolic activation of the adjacent nitrogen substituent; 2) For the substituted aromatic amines –NR₂, R = C₃ or greater or extensive steric crowding of the substituents have a detrimental effect on the ability of a chemical to be metabolically activated (Ashby, Tennant 1988). The above modulating factors are likely to completely abolish the toxic effect of a SA. In principle, it is also possible to list substructures that enhance or diminish the toxic potency of the active chemicals: coding these finer effects via substructures is however more difficult than coding large yes/no effects on the activity.

3.3 Compilations of structural alerts

In the literature, a number of different lists of SAs have been reported. These were originally created as compilations of the scientific knowledge on the mechanisms of chemical carcinogenicity, without any use of statistics. With the availability of approaches for treating large databases and for manipulating chemical structure with computers, refinements have been attempted with the support of more formal approaches (e.g., statistics / artificial intelligence).

The following are the main literature sources on SAs, used by us as a basis for the development of the present expert system.

Special relevance has the compilation of SAs by John Ashby, that was used by subsequent investigators as a starting point for refinements / adjustments (Ashby 1985; Ashby, Tennant 1988). The latter reference includes additional SAs in respect to the classical poly-carcinogen presented earlier, as well as some detoxifying chemical functionalities (e.g., sulfonic groups on azo-dyes, sterically hindering groups on the aromatic amino nitrogen). This model has a total of 19 SAs.

The compilation of SAs by Bailey et al. (Bailey, Chanderbhan, Collazo-Braier, Cheeseman, Twaroski 2005) was generated for being used in the regulatory context of the newly implemented Food and Contact Notification program of the U.S. Food and Drug Administration (FDA) Office for Food Additive Safety. The list of SAs is based on the Ashby's SAs, and on a related list compiled by Munro (Munro, Ford, Kennepohl, Sprenger 1996). It consists of 33 SAs.

Kazius et al. (Kazius, McGuire, Bursi 2005) produced another list of SAs (29 in total), based on a computerized data mining analysis whose results were "supervised" with an eye to the expert knowledge formalized by John Ashby. As noted above, the Ashby SAs are tailored on the mechanistic knowledge on chemical carcinogens, mainly restricted to the genotoxic (DNA reactive) carcinogens. The exercise by Kazius et al. 2005 used a mutagenicity database (4337 mutagens and nonmutagens from the Toxnet database; <http://toxnet.nlm.nih.gov/>). Thus, the resulting SAs are typical of *Salmonella* mutagens, and for this reason they are rigorously restricted to the genotoxic carcinogens.

The fourth set of SAs was generated by Kazius et al. (Kazius, Nijssen, Kok, Back, Ijzerman 2006) in an exercise aimed at experimenting a new way of representing the chemicals (hierarchical graphs) and a new searching algorithm (called Gaston). The goal was to generate automatically SAs through artificial intelligence methods solely. This effort resulted in 6 "complex" SAs.

Another source of information on SAs is provided by the Oncologic expert system. Oncologic is a noncommercial software created by the US Environmental Protection Agency, that can be freely downloaded (<http://www.epa.gov/oppt/newchems/tools/oncologic.htm>). It consists of separate modules, one of which performs predictions on the carcinogenicity of chemicals using a database of SAs and accompanying modulating factors. Oncologic follows a

mechanism-based expert reasoning, and provides a final semi-quantitative assessment (low, marginal, low-moderate, moderate, high-moderate, high). Whereas the modulating factors considered by the authors of the above lists of SAs are used to “cancel” the relevance of the SAs in a yes/no fashion, Oncologic transforms them into a probabilistic scale of gravity.

3.4 Structural alerts for predictive toxicology

Recently, the ability of the SAs to uncover carcinogens/mutagens in large databases of chemicals experimentally tested has been compared (Benigni, Netzeva, Benfenati et al. 2007) (see also the report (Benigni, Bossa, Netzeva, Worth 2007)). The sets of SAs considered are those by Ashby, Bailey et al., and the two by Kazius et al. A chemical containing a SA was considered to be predicted as positive, whereas a chemical without any known SA was predicted, by exclusion, as negative. When a SA was accompanied by the presence of modulating factors supposed to annihilate the SA-related activity, the chemical was classified as negative.

Overall, the four SA models did not differ to a large extent in their performance. In databases including chemicals from diverse chemical classes, the SA models appear to agree around 65% with rodent carcinogenicity data, and 75% with Salmonella mutagenicity data. As an exception, the Bailey SAs exhibit lesser specificity (higher false positives) than the Ashby SAs, without a comparable increase in sensitivity. In addition, the SA models do not work equally efficiently in the discrimination between active and inactive chemicals within individual chemical classes: their poorer performance can be ascribed to the fact that the SA models considered lack sub-rules detailed enough as to be able to describe how each alert is modulated by the different molecular environments.

The above measures can be considered as the “average” accuracy of the SAs for the “known” universe of chemicals in the public domain. It is emphasised that these SAs do not consider all possible chemicals in the universe (e.g., chemicals that will be synthesized in the future for new commercial applications), and chemicals from proprietary studies. Within the above limits, the SAs have a unique role for: a) description of sets of chemicals; b) preliminary hazard characterisation; c) formation

of categories for e.g., regulatory purposes; d) generation of subsets of congeneric chemicals to be analyzed subsequently with QSAR methods; e) priority setting.

4. Structural alerts included in the Benigni/Bossa rulebase

The SAs included in Toxtree are 33; out of them, five SAs refer to nongenotoxic mechanisms of action. Appendix 1 provides the structure of the SAs, together with a number of representative toxic chemicals for each of them.

The SAs derive from an analysis of several literature sources. The main references are: a) (Ashby 1985; Ashby, Tennant 1988) ; b) (Bailey, Chanderbhan, Collazo-Braier, Cheeseman, Twaroski 2005); c) (Kazius, McGuire, Bursi 2005); d) the Oncologic expert system. The evidence from these sources has been combined in such a way as to be as exhaustive and non-redundant as possible, and at the same time as to be suitable for the software implementation.

Only SAs actually present in major databases of chemicals have been accepted. The databases were ISSCAN (Benigni, Bossa 2006b) (Benigni, Bossa, Richard, Yang 2008), CPDB (Benigni, Bossa 2006a), and Toxnet-Kazius (Kazius, McGuire, Bursi 2005). The frequency varies from 1 up to around 100.

Several SAs have accompanying modulating factors. Previous papers (Benigni, Andreoli, Giuliani 1994; Benigni, Bossa, Netzeva, Rodomonte, Tsakovska 2007), and some studies preliminary to this expert system (Benigni and Franke, unpublished) have shown that structural effects on potency should be distinguished from effects on yes/no activity. Taking the aromatic amines as an example, Cl and NH₂ groups ortho or meta to the main –NH₂ functionality strongly decrease the carcinogenic potency; however, such effect on the potency of the carcinogens may not abolish their carcinogenic activity. Since a) the main goal of the SAs is preliminary or large-scale screenings (Benigni et al. 2007), and b) the knowledge on modulating factors for most chemical classes is not available, the accepted modulating factors in this expert system are only the structural motifs that have a high probability of abolishing the effects of the SAs.

No attempt was done to define the Applicability Domain (AD) of the SAs in a rigorous way. The concept of AD applies to the structural or physical chemical characteristics of the set of chemicals used as training set in the derivation of a (Q)SAR model; it is understood that the model cannot be applied to new chemicals that do not obey to such characteristics (Netzeva, Worth, Aldenberg et al. 2005). Since most of the knowledge on SAs derives from a complex body of mechanistic observations and concepts with different origins and not from a formal analysis of experimental data, such a strict definition cannot be provided. It can be assumed that each SA is ruling the biological activity of a molecule as far as its reactivity is not seriously hampered by other groups or substructures present in the same molecule. In this sense, the definition of modulating factors (when known) for a SA is like in signification to the definition of its AD.

4.1 The performance of SAs

The agreement of the present list of SAs with the carcinogenicity and mutagenicity of chemicals in the ISSCAN database has been tested. ISSCAN has been used because of the quality of its data, and because previous work has shown that it is representative of the performance of the alerts in other large public databases (Benigni, Bossa, Netzeva, Worth 2007).

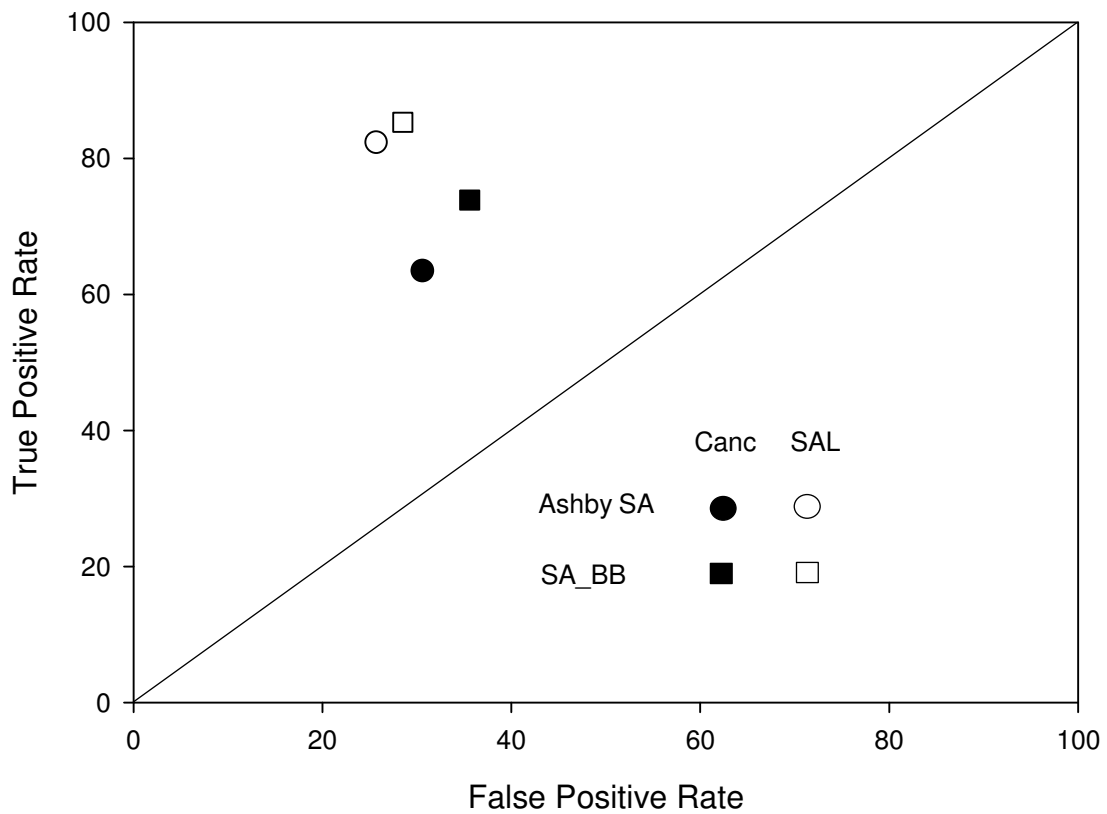
Table I displays the sensitivity, specificity and accuracy of the SAs implemented in this system (SA_BB), and reports for a comparison the performance of the Ashby SAs in the same database (results in (Benigni, Bossa, Netzeva, Worth 2007)). Obviously, the alerts for nongenotoxic effects in this system have not been considered when assessing the performance in respect to mutagenicity.

Table I

		Sensitivity	Specificity	Accuracy
SA_BB	Canc	0,74	0,64	0,70
	Mut	0,85	0,72	0,78
Ashby SA	Canc	0,64	0,69	0,65
	Mut	0,82	0,74	0,78

For an easy visual comparison, the results are also expressed as Receive Operating Characteristics graphs (Figure 1). A ROC graph reports true positive rate (sensitivity) on the Y-axis, and false positive rate (1 - specificity) on the X-axis. In a ROC graph, perfect performance is located at the left upper corner; the diagonal line represents random results (Provost, Fawcett 2001).

Figure 1



The analysis shows that the present list of alerts has increased sensitivity and accuracy in respect to the Ashby alerts, at the cost of a diminished specificity. Thus an overall increase in performance is apparent.

Table II displays the numerical presence of the various alerts for carcinogenicity in the ISSCAN database, together with the percentage of true positive chemicals (carcinogens) in each category.

Table II

STRUCTURAL ALERT	N. of chemicals fired	N. of carcinogens	True Positives Rate
SA_1: Acyl halides	1	1	100%
SA_2: alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid	12	10	83.33%
SA_3: N-methylol derivatives	2	2	100%
SA_4: Monohaloalkene	6	6	100%
SA_5: S or N mustard	10	10	100%
SA_6 Propiolactones or propiosultones	4	4	100%
SA_7: Epoxides and aziridines	22	18	81.82%
SA_8: Aliphatic halogens	66	49	74.24%
SA_9: Alkyl nitrite	1	1	100%
SA_10: α , β unsaturated carbonyls	38	29	76.32%
SA_11: Simple aldehyde	8	7	87.5%
SA_12: Quinones	12	10	83.33%
SA_13: Hydrazine	53	51	96.23%
SA_14: Aliphatic azo and azoxy	7	7	100%
SA_15: : isocyanate and isothiocyanate groups	3	3	100%
SA_16: alkyl carbamate and thiocarbamate	6	6	100%
SA_17: Thiocarbonyl (nongenotoxic)	19	13	68.42%
SA_18: Polycyclic Aromatic Hydrocarbons	12	9	75%
SA_19: Heterocyclic Polycyclic Aromatic Hydrocarbons	12	11	91.67%
SA_20: (Poly) Halogenated Cycloalkanes (nongenotoxic)	17	14	82.35%
SA_21: alkyl and aryl N-nitroso groups	79	78	98.73%
SA_22: azide and triazene groups	5	3	60%

SA_23: aliphatic N-nitro group	4	4	100%
SA_24: α , β unsaturated aliphatic alkoxy group	2	2	100%
SA_25: aromatic nitroso group	3	3	100%
SA_26: aromatic ring N-oxide	3	2	66.67%
SA_27: Nitro-aromatic	74	56	75.68%
SA_28: primary aromatic amine, hydroxyl amine and its derived esters	93	78	83.87%
SA_28bis: Aromatic mono- and dialkylamine	11	9	81.82%
SA_28ter: aromatic N-acyl amine	16	12	75%
SA_29: Aromatic diazo	20	16	80%
SA_30: Coumarins and Furocoumarins	6	5	83.33%
SA_31a: Halogenated benzene (nongenotoxic)	11	4	36.36%
SA_31b: Halogenated PAH (nongenotoxic)	9	8	88.89%
SA_31c: Halogenated dibenzodioxins (nongenotoxic)	4	2	50%

The inspection of Table II shows that most of the alerts implemented are highly selective (i.e., the number of non carcinogens erroneously flagged as carcinogens is relatively low), with the exception of few alerts (e.g., SA_31a, SA_31c). The latter alerts are mainly related to nongenotoxic mechanisms of action. It appears that more work is necessary to define the appropriate modulating factors that are able to diminish or destroy the carcinogenicity potential of these alerts.

Table III provides statistics limited to the alerts for genotoxic carcinogenicity in the ISSCAN database, and compares them with the Salmonella mutagenicity results. Thus, this table refers to the predictivity (selectivity) for mutagenicity.

Table III

STRUCTURAL ALERT	N. of chemicals fired	N. of mutagens	True Positives Rate
SA_1: Acyl halides	1	1	100%
SA_2: alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid	11	7	63.64%
SA_3: N-methylol derivatives	1	0	0%
SA_4: Monohaloalkene	5	4	80%
SA_5: S or N mustard	8	7	87.5%
SA_6 Propiolactones or propiosultones	4	4	100%
SA_7: Epoxides and aziridines	18	15	83.33%
SA_8: Aliphatic halogens	56	36	64.29%
SA_9: Alkyl nitrite	1	1	100%
SA_10: α , β unsaturated carbonyls	26	8	30.77%
SA_11: Simple aldehyde	6	2	33.33%
SA_12: Quinones	11	11	100%
SA_13: Hydrazine	29	22	75.86%
SA_14: Aliphatic azo and azoxy	4	3	75%
SA_15: : isocyanate and isothiocyanate groups	3	3	100%
SA_16: alkyl carbamate and thiocarbamate	5	3	60%
SA_18: Polycyclic Aromatic Hydrocarbons	9	8	88.89%
SA_19: Heterocyclic Polycyclic Aromatic Hydrocarbons	10	9	90%
SA_21: alkyl and aryl N-nitroso groups	45	42	93.33%
SA_22: azide and triazene groups	5	5	100%
SA_23: aliphatic N-nitro group	4	4	100%
SA_24: α , β unsaturated aliphatic alkoxy group	2	1	50%

SA_25: aromatic nitroso group	1	1	100%
SA_26: aromatic ring N-oxide	1	1	100%
SA_27: Nitro-aromatic	62	56	90.32%
SA_28: primary aromatic amine, hydroxyl amine and its derived esters	83	66	79.52%
SA_28bis: Aromatic mono- and dialkylamine	11	6	54,55%
SA_28ter: aromatic N-acyl amine	14	12	85.71%
SA_29: Aromatic diazo	19	13	68.42%
SA_30: Coumarins and Furocoumarins	5	4	80%

Table III shows that the selectivity of the alerts is quite high also for mutagenicity.

The low selectivity of the exceptions (e.g., SA_10, SA_11) provides very useful evidence, since it points to SAs whose modulating factors have to be studied further.

5. Quantitative structure activity relationships in the Benigni/Bossa rulebase

Based on a recent survey and subsequent refinements, three QSARs for discriminating between inactive and active chemicals have been identified as particularly promising (Benigni, Bossa, Netzeva, Worth 2007) (Benigni, Bossa, Netzeva, Rodomonte, Tsakovska 2007) and were selected for inclusion in the Benigni/Bossa rulebase. They are models for: 1) the mutagenic activity of aromatic amines in the *Salmonella typhimurium* TA100 strain (Ames test); 2) the carcinogenic activity of the aromatic amines in rodents (summary activity from rats and mice); 3) the mutagenic activity of $\alpha\beta$ -unsaturated aldehydes in the *Salmonella typhimurium* TA100 strain (Ames test).

These QSARs are meant to provide more finely-tuned estimations for chemicals belonging to the two chemical classes above: at odds with the SAs, the QSARs generate both negative and positive predictions. The QSARs are applied when query chemicals with the appropriate SAs are recognised.

Details on the individual models are in Appendix 2.

5.1 Mathematical models

The QSAR models were obtained through Canonical Discriminant Analysis (Franke 1984; Franke, Gruska 2003). Shortly, the so-called discriminant function, w , is based on the descriptor variables supposed to be related to the distribution of compounds over the classes of actives and inactives, and has the general (linear) form

$$\text{Activity} = a_0 + a_1x_1^{n1} + \dots + a_ix_i^{ni} + \dots + a_nx_n^{nn}$$

The coefficients a_i are so determined that the separation of classes is optimal. This is done by solving a special eigenvalue problem. In a two-class case, the discriminant function w can be visualized as the axis of an one-dimensional coordinate system with the two classes occupying the opposite ends. The further these regions are apart, the better is the separation of classes achieved by the respective discriminant function.

Based on the distribution of w values for the two classes of actives and inactives, a threshold that best separates the two classes is decided. Once a discriminant function is known, a compound can be classified by computing the value of w for this compound from inserting the values of the respective descriptor variables into the discriminant function. The chemical is assigned to one class or another, based on its position in respect to the established threshold between classes.

5.2 Characterisation of the models

The model statistics include: accuracy, sensitivity, and specificity, together with the Squared Canonical Correlation.

Accuracy is the percentage of all chemicals correctly identified by the model. Sensitivity is the percentage of biologically active (positive) chemicals correctly identified (calculated out of the total number of positives). Specificity is the percentage of biologically inactive (negative) chemicals correctly identified (calculated out of the total number of negatives).

The Squared Canonical Correlation is a measure of the correlation between the biological activity variable, and the linear combination of descriptor variables that best separates the negatives from the positives.

Validation of QSAR model performance is an important consideration. It is generally accepted that the gold standard is an external validation test that employs a robust and diversified set of chemical structures not used for the derivation of the model. Thus, the model is applied to the external test set, and the concordance between the experimental data and the activity estimated through the QSAR is calculated.

Due to limitations of external experimental data, a number of statistical techniques aimed to simulate the above procedure have been devised. In practice, many investigators use internal cross-validation procedures to generate artificial test sets by splitting the training set of chemicals into two or more test sets, and regarding one as training and the other one as test set. There is evidence that the internal cross-validation procedures are useful tools in the phase of the model construction (assessment of statistical consistency) and concur to better characterise the data in the training set, whereas external validation may better assess the confidence one can

have in the predictions of the model itself, if a sufficiently large and diversified set of chemical structures not considered in the model can be used for such a purpose (Kubinyi 2005; Benigni, Bossa 2007).

For the QSARs in this expert system, characterisation through both cross-validation methods, and external test sets is reported. For cross-validation, three leave-many-out procedures were considered, leaving out: a) 10%; b) 25%; and c) 50% of the chemicals in the data set: the model was re-calculated on the remaining chemicals, and then applied to predict the activity of the chemicals left out. Each procedure was applied ten times (by random selection of excluded chemicals, in such a way as to maintain the proportion between negatives and positives in the overall data set).

5.3 Applicability Domain of QSARs

The QSAR models are derived empirically from the analysis of a training set of chemicals, whose biological activity is known. The QSAR analysis is aimed at discovering the properties, or features of the molecules that correlate with the biological activity. In order to attain the best results, a QSAR analysis should focus on a well defined set of congeneric chemicals, i.e., chemicals with similar structure that act through the same mechanism of action (Franke 1984; Hansch, Leo 1995). Thus when the QSAR model is applied to new chemicals to predict their biological activity, it is crucial that the chemicals to be predicted have the same characteristics of the training set. These characteristics are called Applicability Domain of the model, and are typical of each individual model.

The Applicability Domain of the models contained in this expert system are defined in terms of structural characteristics of the chemical classes to which they apply. This expert system applies the models only to chemicals that respect such constraints. The constraints are presented in the description of each model (Appendix 2).

6. Outputs and classification schemes

The processing of a query chemical by the software can give rise to a limited number of different outcomes, namely: a) no presence of SAs for carcinogenicity; b) one or more SAs are recognised; c) SAs relative to aromatic amines or $\alpha\beta$ -unsaturated aldehydes are recognised, and the chemical goes through QSAR analysis, which may result in a negative or positive outcome. The system flags either outcome through one, or a combination, of a few labels, as follows:

- No alerts for cancerogenic activity

No SAs have been recognised by the system.

- Structural Alert for genotoxic carcinogenicity
- Structural Alert for nongenotoxic carcinogenicity

The system recognises the presence of one or more SAs, and specifies the genotoxic or nongenotoxic mechanism.

Potential *S. typhimurium* TA100 mutagen based on QSAR

Unlikely to be a *S. typhimurium* TA100 mutagen based on QSAR

- Potential carcinogen based on QSAR
- Unlikely to be a carcinogen based on QSAR

If the query chemical belongs to the classes of aromatic amines or $\alpha\beta$ -unsaturated aldehydes, the appropriate QSAR is applied. A QSAR provides an assessment more refined in respect to the SAs, and should be given higher importance in a weight-of-evidence scheme. Thus, a QSAR analysis might point to an estimated lack of toxic effects, in spite of the presence of SAs.

A special, seemingly contradictory case is when the system flags potential mutagenicity or carcinogenicity based on a QSAR, but no SA. This logical incongruity is solved by the fact that some SAs may not fire in the presence of modulating factors (e.g., because of large substituents in the vicinity of the main functional group); nevertheless a finely-tuned QSAR analysis may still suggest potential toxicity.

7. References

Ames BN. The detection of environmental mutagens and potential carcinogens. *Cancer* 1984;53:2030-2040.

Arcos JC, Argus MF. Multifactor interaction network of carcinogenesis - a "Tour Guide". In Arcos JC, Argus MF, Woo YT, eds. *Chemical induction of cancer. Modulation and combination effects*. Boston: Birkhauser, 1-20.

Ashby J. Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environ Mutagen* 1985;7:919-21.

Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested by the U.S.NCI/NTP. *Mutat Res* 1988;204:17-115.

Bailey AB, Chanderbhan N, Collazo-Braier N, Cheeseman MA, Twaroski ML. The use of structure-activity relationship analysis in the food contact notification program. *Regulat Pharmacol Toxicol* 2005;42:225-35.

Benigni R. Structure-activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches. *Chem Revs* 2005;105:1767-800.

Benigni R, Andreoli C, Giuliani A. QSAR models for both mutagenic potency and activity: application to nitroarenes and aromatic amines. *Environ Mol Mutagen* 1994;24:208-19.

Benigni R, Bossa C. Structural alerts of mutagens and carcinogens. *Curr Comput -Aid Drug Des* 2006a;2:169-76.

Benigni R, Bossa C. Structure-activity models of chemical carcinogens: state of the art, and new directions. *Ann Ist Super Sanità* 2006b;42:118-26.

Benigni R, Bossa C. Predictivity and reliability of QSAR models: the case of mutagens and carcinogens. *Toxicol Mechanisms Meth* 2007;accepted.

Benigni R, Bossa C, Netzeva TI, Rodomonte A, Tsakovska I. Mechanistic Qsar Of Aromatic Amines: New Models For Discriminating Between Homocyclic Mutagens And Nonmutagens, And Validation Of Models For Carcinogens. *Environ Mol Mutagen* 2007;48:754-71.

Benigni, R., Bossa, C., Netzeva, T. I., and Worth, A. P. Collection and evaluation of (Q)SAR models for mutagenicity and carcinogenicity. EUR 22772 EN. 2007. Luxembourg, Office for the Official Publications of the European Communities. EUR - Scientific and Technical Research Series. 12-7-2007.

Benigni R, Bossa C, Richard AM, Yang C. A novel approach: chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity. *Ann Ist Super Sanità* 2008;to be published.

Benigni R, Netzeva TI, Benfenati E et al. The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens. *J Environ Sci Health C Environ Carcinog Ecotoxicol Revs* 2007;25:53-97.

Franke R. *Theoretical drug design methods*. Amsterdam: Elsevier.

Franke R, Gruska A. General introduction to QSAR. In Benigni R, ed. *Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens*. Boca Raton: CRC Press, 1-40.

Fung VA, Barrett JC, Huff J. The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. *Environ Health Perspect* 1995;103:680-683.

Hansch C, Hoekman D, Leo A, Weininger D, Selassie CD. Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology. *Chem Revs* 2002;102:783-812.

Hansch C, Leo A. *Exploring QSAR. 1. Fundamentals and applications in chemistry and biology*. Washington, D.C.: American Chemical Society.

Huff J. Value, validity, and historical development of carcinogenesis studies for predicting and confirming carcinogenic risks to humans. In Kitchin KT, ed. *Carcinogenicity. Testing, predicting, and interpreting chemical effects*. New York: Marcel Dekker, Inc., 21-123.

Huff JE, Haseman JK, Rall DP. Scientific concepts, value, and significance of chemical carcinogenesis studies. *Ann Rev Pharmacol Toxicol* 1991;31:621-25.

Kazius J, McGuire R, Bursi R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J Med Chem* 2005;48:312-20.

Kazius J, Nijssen S, Kok J, Back T, Ijzerman AP. Substructure mining using elaborate chemical representation. *J Chem Inf Model* 2006;46:597-605.

Kubinyi H. Validation and predictivity of QSAR models. In Aki-Sener E, Yalcin I, eds. *QSAR and Molecular Modelling in rational design of bioactive molecules. The 15th European Symposium on Quantitative Structure-Activity Relationships and molecular modelling*. Ankara: CADDs, 30-33.

Maron DM, Ames BN. Revised methods for the *Salmonella* mutagenicity test. *Mutat Res* 1983;113:173-215.

Miller EC, Miller JA. Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer* 1981;47:2327-45.

Miller JA, Miller EC. Ultimate chemical carcinogens as reactive mutagenic electrophiles. In Hiatt HH, Watson JD, Winsten JA, eds. *Origins of human cancer*. Cold Spring Harbor: Cold Spring Harbor Laboratory, 605-28.

Munro IC, Ford RA, Kennepohl E, Sprenger JG. Threshold of toxicological concern based on Structure-Activity Relationships. *Drug Metab Rev* 1996;28:209-17.

Netzeva TI, Worth AP, Aldenberg T et al. Current status of methods for defining the applicability domain of (Quantitative) Structure-Activity Relationships. *ATLA-Alternatives to Laboratory Animals* 2005;33:1-19.

Provost F, Fawcett T. Robust classification for imprecise environment. *Machine Learn J* 2001;42:5-11.

Singer B, Grunberg D. *Molecular biology of mutagens and carcinogens*. New York: Plenum Press.

Woo YT. Mechanisms of action of chemical carcinogens, and their role in Structure-Activity Relationships (SAR) analysis and risk assessment. In Benigni R, ed. *Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens*. Boca Raton: CRC Press, 41-80.

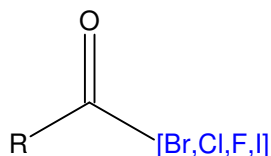
Woo YT, Lai DY, McLain JL, Ko Manibusan M, Dellarco V. Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products. *Environ Health Perspect* 2002;110:75-87.

Zeiger E. Carcinogenicity of mutagens: predictive capability of the Salmonella mutagenesis assay for rodent carcinogenicity. *Cancer Res* 1987;47:1287-96.

Appendix 1: Structural alerts

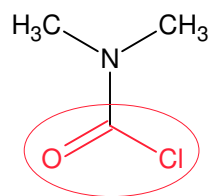
STRUCTURAL ALERT

SA_1: Acyl halides



R = any atom/group, except OH, SH

ISSCANv2a_25



ChemName: Dimethylcarbamoyl Chloride

CAS: 79-44-7

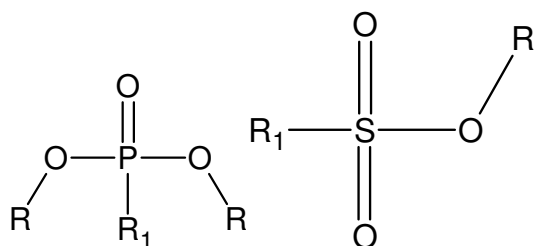
Mouse_Male : ND

Mouse_Female : 3

Rat_Male : ND

Rat_Female : ND

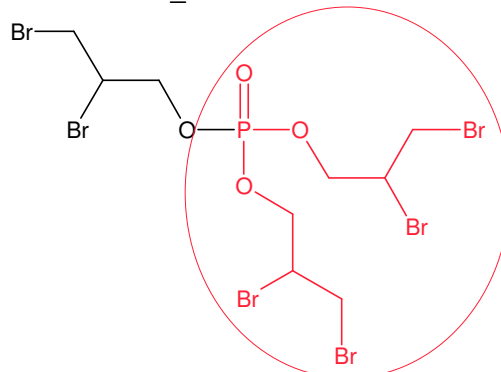
SA_2: alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid



R= Alkyl with C<5 (also substituted with halogens), or benzyl

R1= any atom/group except OH, SH, O⁻, S⁻

ISSCANv2a_237



ChemName: Tris(2,3-dibromopropyl) phosphate

CAS: 126-72-7

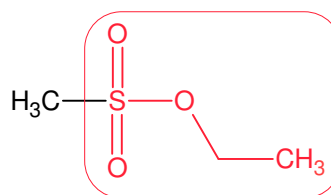
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_286



ChemName: Ethyl Methanesulfonate

CAS: 62-50-0

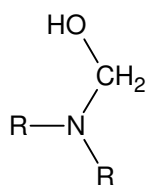
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

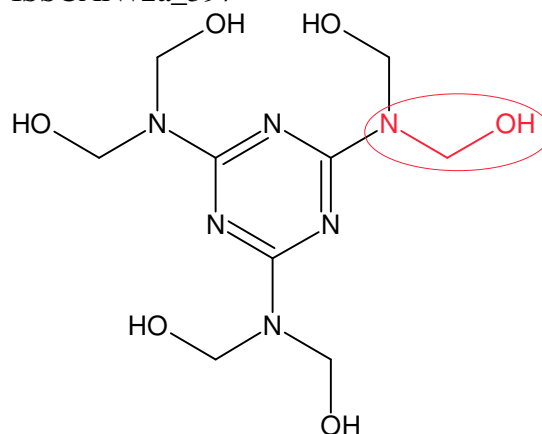
Rat_Female : 3

SA_3: N-methylol derivatives



R = any atom/group

ISSCANv2a_397



ChemName:Hexa(hydroxymethyl)melamine

CAS: 531-18-0

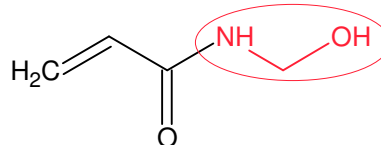
Mouse_Male : ND

Mouse_Female : 3

Rat_Male : ND

Rat_Female : ND

ISSCANv2a_616



ChemName: N-methylolacrylamide

CAS: 924-42-5

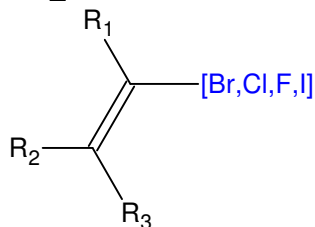
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 1

Rat_Female : 1

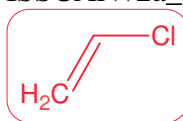
SA_4: Monohaloalkene



R₁, R₂ (or R₃) = H or Alkyl

R₃ (or R₂) = any atom/group except halogens

ISSCANv2a_253



ChemName: Vinyl Chloride

CAS: 75-01-4

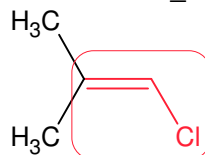
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_262



ChemName: Dimethylvinyl Chloride

CAS: 513-37-1

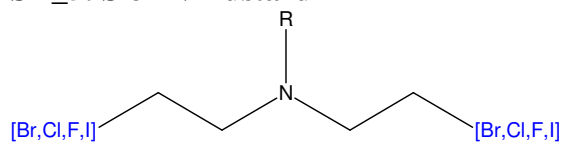
Mouse_Male : 3

Mouse_Female : 3

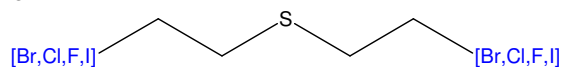
Rat_Male : 3

Rat_Female : 3

SA_5: S or N mustard

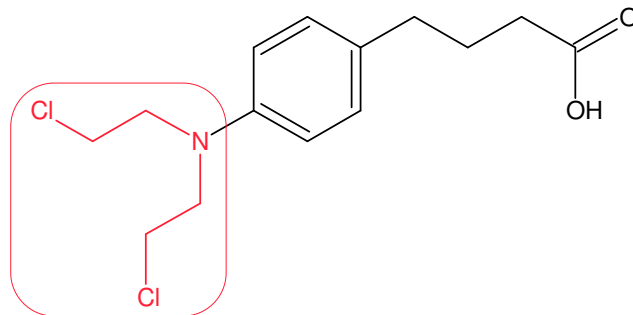


or



R = any atom/group

ISSCANv2a_405



ChemName: Chloroambucil

CAS: 305-03-3

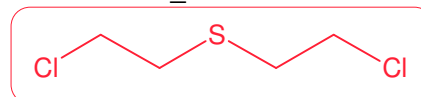
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_242



ChemName: Bis(2-chloroethyl)sulfide

CAS: 505-60-2

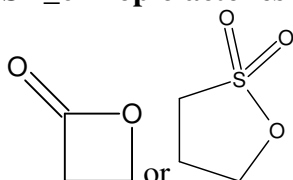
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : ND

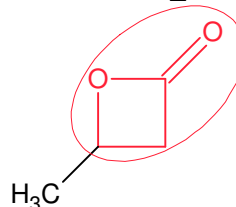
Rat_Female : ND

SA_6 Propiolactones or propiosultones



Any substance with the displayed substructures

ISSCANv2a_15



ChemName: beta-Butyrolactone

CAS: 3068-88-0

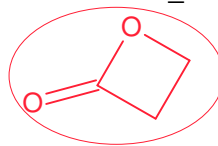
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_22



ChemName: beta-Propiolactone

CAS: 57-57-8

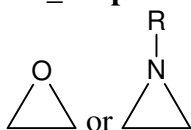
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : ND

Rat_Female : 3

SA_7:Epoxides and aziridines



R = any atom/group

ISSCANv2a_243



ChemName: Ethylene Oxide

CAS: 75-21-8

Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_463



ChemName: Ethyleneimine

CAS: 151-56-4

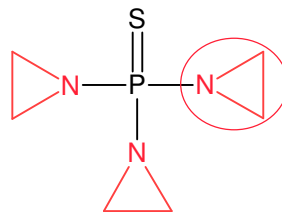
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : ND

Rat_Female : ND

ISSCANv2a_289



ChemName: Thiotepa

CAS: 52-24-4

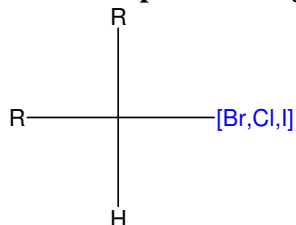
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

SA_8: Aliphatic halogens



R = any atom/group

ISSCANv2a_49



ChemName: 1,2-dibromoethane

CAS: 106-93-4

Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_252



ChemName: 1,2-dichloroethane

CAS: 107-06-2

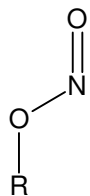
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

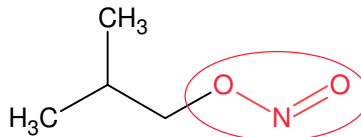
Rat_Female : 3

SA_9: Alkyl nitrite



R= any alkyl group

ISSCANv2a_656



ChemName: Isobutyl Nitrite

CAS: 542-56-3

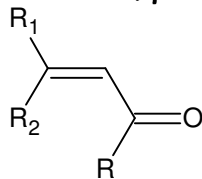
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

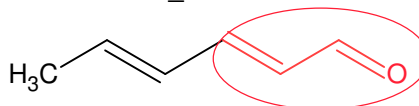
SA_10: α , β unsaturated carbonyls



R1 and R2 = any atom/group, except alkyl chains with C>5 or aromatic rings.

R= any atom/group, except OH, O⁻

ISSCANv2a_773



ChemName: 2,4-Hexadienal

CAS: 142-83-6

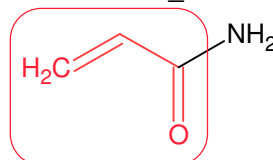
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_282



ChemName: Acrylamide

CAS: 79-06-1

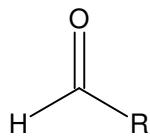
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

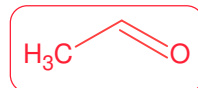
Rat_Female : 3

SA_11: Simple aldehyde



R= aliphatic or aromatic carbon
 α, β unsaturated aldehydes are excluded

ISSCANv2a_321



ChemName: Acetaldehyde

CAS: 75-07-0

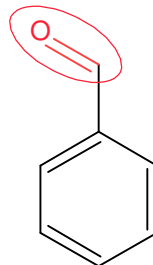
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_661



ChemName: Benzaldehyde

CAS: 100-52-7

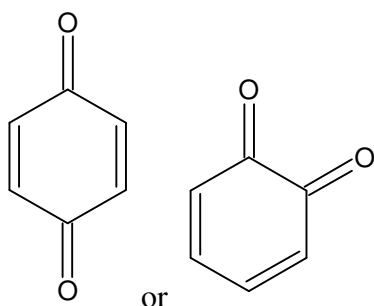
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 1

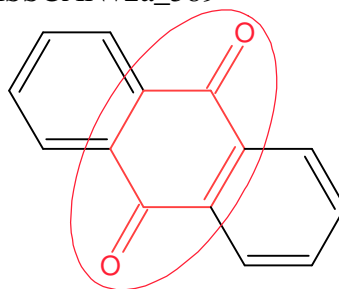
Rat_Female : 1

SA_12: Quinones



Any substance with the displayed substructures

ISSCANv2a_589



ChemName: 9,10-Anthraquinone

CAS: 84-65-1

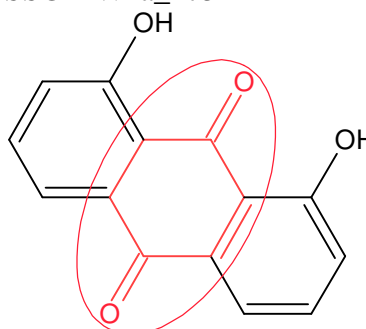
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_475



ChemName: Chryszin

CAS: 117-10-2

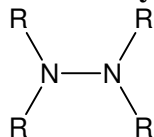
Mouse_Male : 3

Mouse_Female : ND

Rat_Male : 3

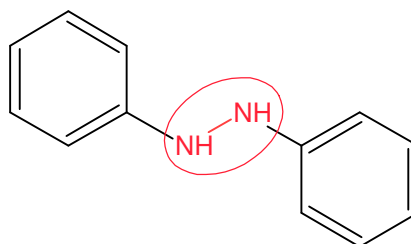
Rat_Female : ND

SA_13: **Hydrazine**



R= any atom/group

ISSCANv2a_197



ChemName: Hydrazobenzene

CAS: 122-66-7

Mouse_Male : 1

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_464



ChemName: Hydrazine

CAS: 302-01-2

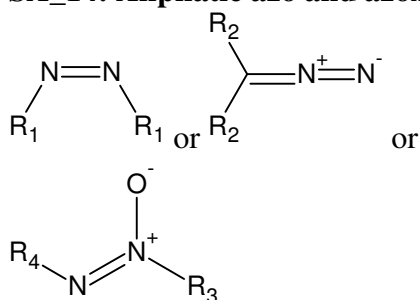
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

SA_14: Aliphatic azo and azoxy

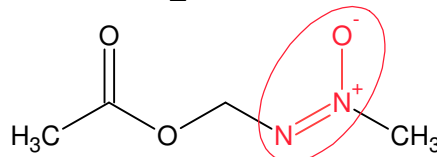


R1= Aliphatic carbon or hydrogen

R2, R3 = Any atom/group

R4 = Aliphatic carbon

ISSCANv2a_29



ChemName: Methylazoxymethanol Acetate

CAS: 592-62-1

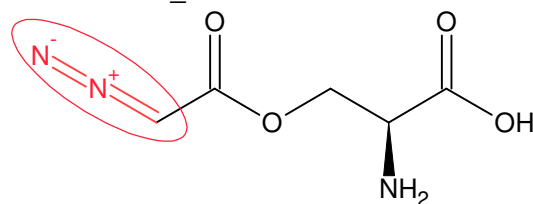
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_443



ChemName: Azaserine

CAS: 115-02-6

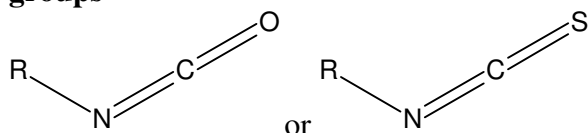
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

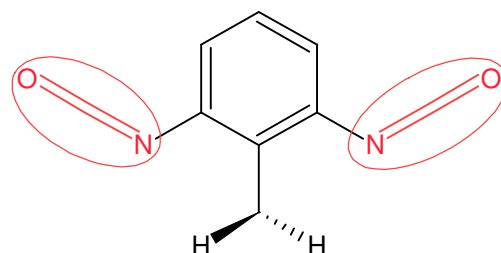
Rat_Female : 3

SA_15: : isocyanate and isothiocyanate groups



R= any atom/group

ISSCANv2a_67



ChemName: Toluene Diisocyanate

CAS: 26471-62-5

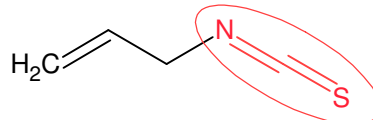
Mouse_Male : 1

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_34



ChemName: Allyl Isothiocyanate

CAS: 57-06-7

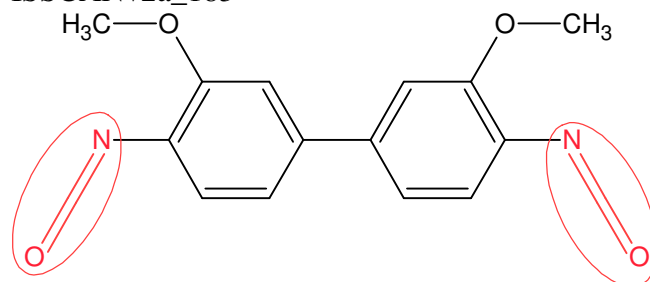
Mouse_Male : 1

Mouse_Female : 1

Rat_Male : 3

Rat_Female : 2

ISSCANv2a_183



ChemName: 3,3'-dimethoxy-4,4'-biphenylene diisocyanate

CAS: 91-93-0

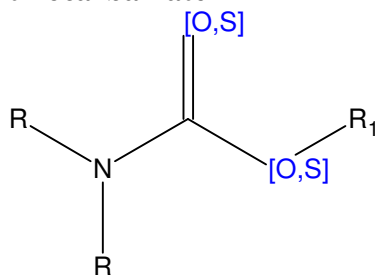
Mouse_Male : 1

Mouse_Female : 1

Rat_Male : 3

Rat_Female : 3

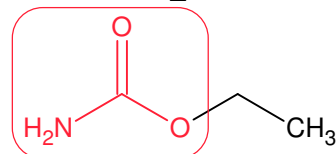
SA_16: alkyl carbamate and thiocarbamate



R = Aliphatic carbon or hydrogen

R1 = Aliphatic carbon

ISSCANv2a_24



ChemName: Urethane

CAS: 51-79-6

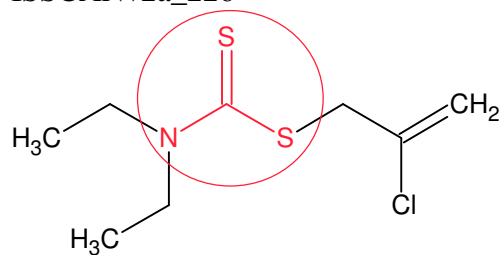
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_226



ChemName: Sulfallate

CAS: 95-06-7

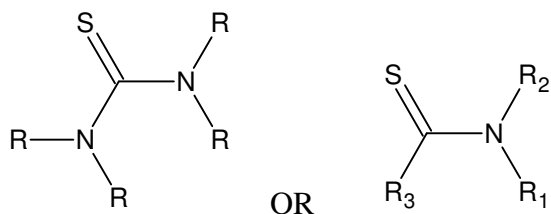
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

SA_17: Thiocarbonyl (nongenotoxic)

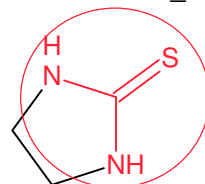


R, R₁, R₂ = Any atom/group

R₃ = Any atom/group except OH, SH, O⁻, S⁻;

Thiocarbamates are excluded.

ISSCANv2a_14



ChemName: Ethylenethiourea

CAS: 96-45-7

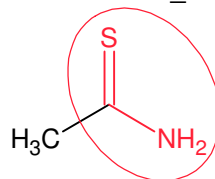
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_625



ChemName: Thioacetamide

CAS: 62-55-5

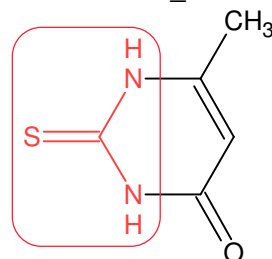
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : ND

ISSCANv2a_356



ChemName: Methylthiouracil

CAS: 56-04-2

Mouse_Male : 3

Mouse_Female : 3

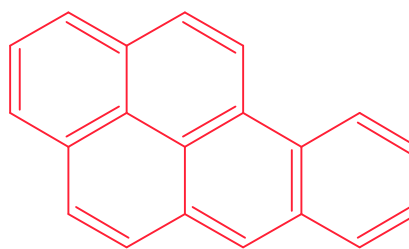
Rat_Male : 3

Rat_Female : 3

SA_18: Polycyclic Aromatic Hydrocarbons

Three or more fused rings, not heteroaromatic

ISSCANv2a_10



ChemName: Benzo(a)pyrene

CAS: 50-32-8

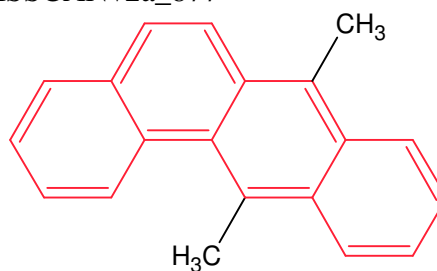
Mouse_Male : 3

Mouse_Female : 1

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_877



ChemName: 7,12-Dimethylbenz(a)anthracene

CAS: 57-97-6

Mouse_Male : 3

Mouse_Female : 3

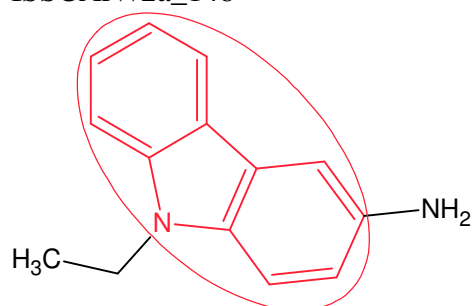
Rat_Male : 3

Rat_Female : 3

SA_19: Heterocyclic Polycyclic Aromatic Hydrocarbons

Three or more fused rings, heteroaromatic

ISSCANv2a_148



ChemName: 3-Amino-9-Ethylcarbazole

CAS: 132-32-1

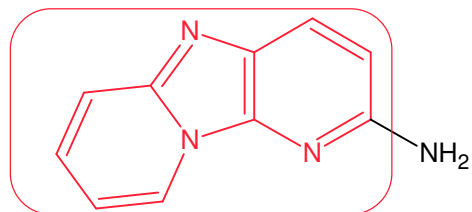
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_334



ChemName: 2-Aminodipyrido[1,2-a:3',2'-d]imidazole

CAS: 67730-10-3

Mouse_Male : 3

Mouse_Female : 3

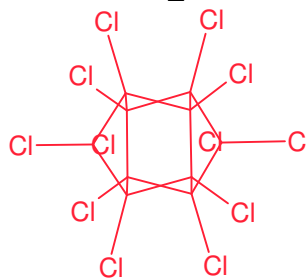
Rat_Male : 3

Rat_Female : 3

**SA_20: (Poly) Halogenated Cycloalkanes
(nongenotoxic)**

Any cycloalkane skeleton with three or more halogens directly bound to the same ring

ISSCANv2a_482



ChemName: Mirex

CAS: 2385-85-5

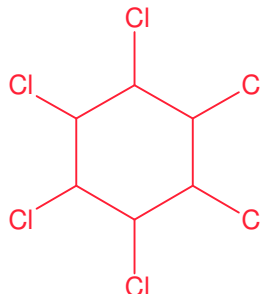
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_450



ChemName: Hexachlorocyclohexane

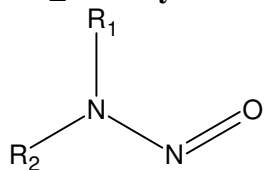
CAS: 608-73-1

Mouse_Male : 3

Mouse_Female : 3

Rat_Male : ND

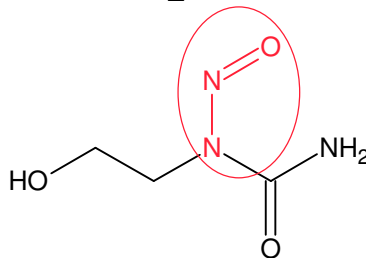
SA_21: alkyl and aryl N-nitroso groups



Rat_Female : ND

R1= Aliphatic or aromatic carbon,
R2= Any atom/group

ISSCANv2a_487



ChemName: 1-(2-Hydroxyethyl)-1-nitrosourea

CAS: 13743-07-2

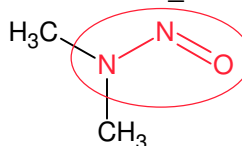
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_519



ChemName: N-nitrosodimethylamine

CAS: 62-75-9

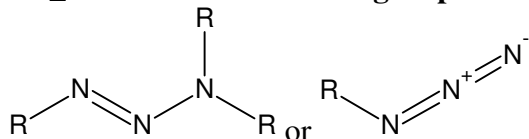
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

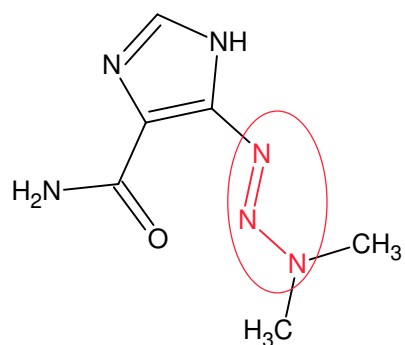
Rat_Female : 3

SA_22: azide and triazene groups



R= Any atom/group

ISSCANv2a_385



ChemName: Dacarbazine

CAS: 4342-03-4

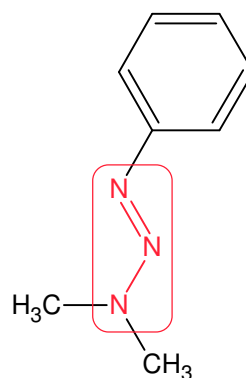
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : ND

Rat_Female : 3

ISSCANv2a_544



ChemName: 1-phenyl-3,3-dimethyltriazene

CAS: 7227-91-0

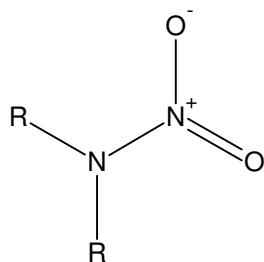
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

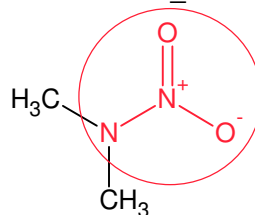
Rat_Female : 3

SA_23: aliphatic N-nitro group



R = Aliphatic Carbon or hydrogen

ISSCANv2a_403



ChemName: Dimethylnitramine

CAS: 4164-28-7

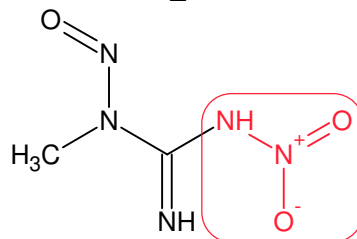
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_360



ChemName: N-methyl-N'-nitro-N-nitrosoguanidine

CAS: 70-25-7

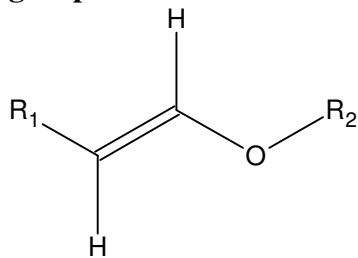
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

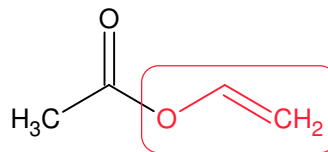
SA_24: α , β unsaturated aliphatic alkoxy group



R1= Any aliphatic Carbon

R2 = Aliphatic or aromatic carbon

ISSCANv2a_499



ChemName: Vinyl Acetate

CAS: 108-05-4

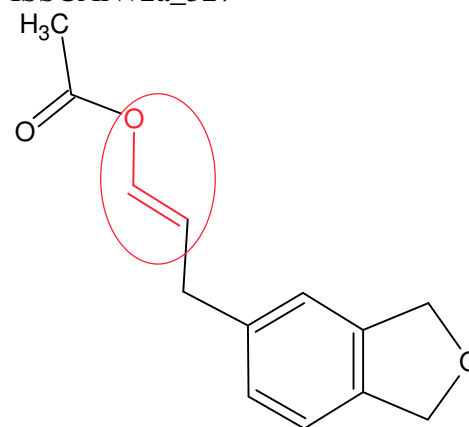
Mouse_Male : 1

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_327



ChemName: 1'-Acetoxysafrole

CAS: 34627-78-6

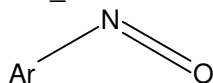
Mouse_Male : 1

Mouse_Female : ND

Rat_Male : 3

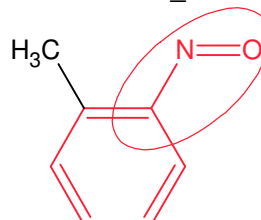
Rat_Female : ND

SA_25: aromatic nitroso group



Ar = Any aromatic/heteroaromatic ring

ISSCANv2a_533



ChemName: o-Nitrosotoluene

CAS: 611-23-4

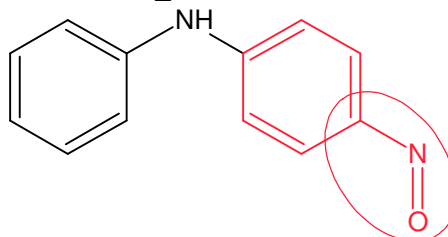
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

Rat_Female : ND

ISSCANv2a_212



ChemName: 4-Nitrosodiphenylamine

CAS: 156-10-5

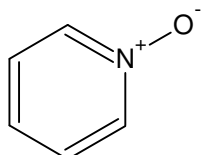
Mouse_Male : 3

Mouse_Female : 1

Rat_Male : 3

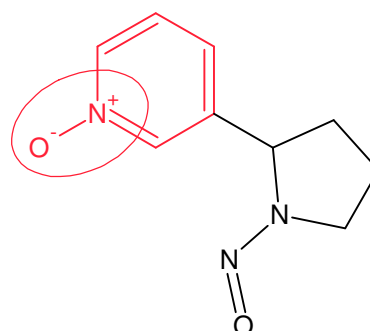
Rat_Female : 1

SA_26: aromatic ring N-oxide



Any aromatic or heteroaromatic ring

ISSCANv2a_558



ChemName: N'-nitrosornicotine-1-N-oxide

CAS: 78246-24-9

Mouse_Male : ND

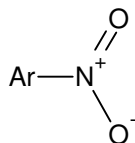
Mouse_Female : ND

Rat_Male : 3

Rat_Female : 3

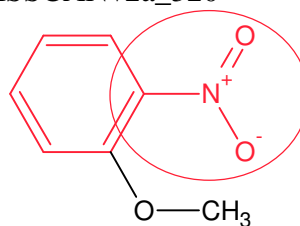
Ar = Any aromatic/heteroaromatic ring

SA_27: Nitro-aromatic



- Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (-SO₃H) on the same ring of the nitro group are excluded .

ISSCANv2a_320



ChemName: o-Nitroanisole

CAS: 91-23-6

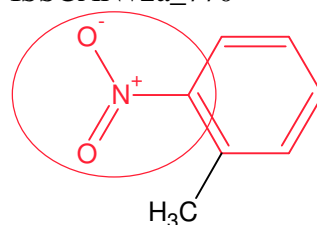
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_770



ChemName: 2-Nitrotoluene

CAS: 88-72-2

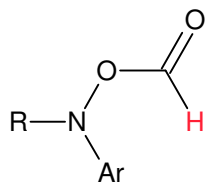
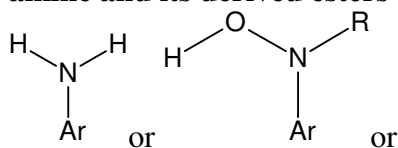
Mouse_Male : 3

Mouse_Female : 3

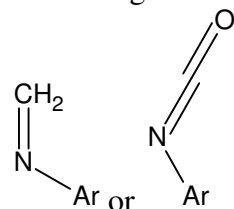
Rat_Male : 3

Rat_Female : 3

SA_28: primary aromatic amine, hydroxyl amine and its derived esters



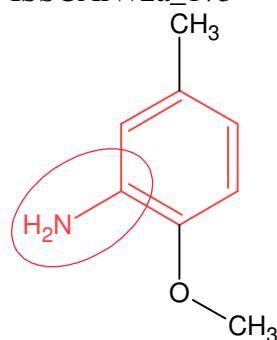
or amine generating group:



Ar = Any aromatic/heteroaromatic ring
R= Any atom/group

- Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (-SO₃H) on the same ring of the amino group are excluded .

ISSCANv2a_173



ChemName: para-Cresidine

CAS: 120-71-8

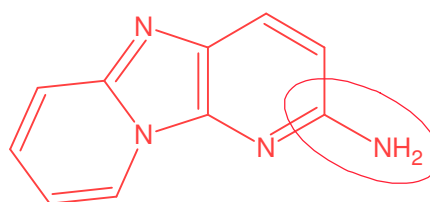
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_334



ChemName: 2-Aminodipyrido[1,2-a:3',2'-d]imidazole

CAS: 67730-10-3

Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

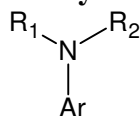
Rat_Female : 3

Ar = Any aromatic/heteroaromatic ring

R1 = Hydrogen, methyl, ethyl

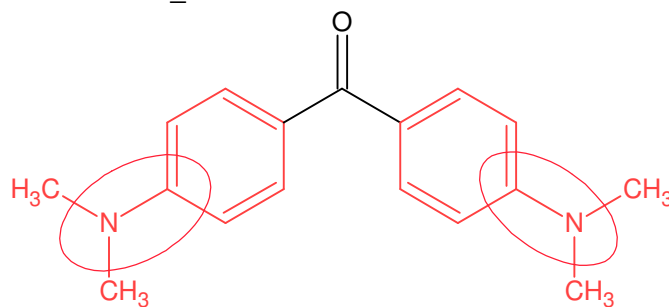
R2 = Methyl, ethyl

SA_28bis: Aromatic mono- and dialkylamine



- Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (-SO₃H) on the same ring of the amino group are excluded .

ISSCANv2a_203



ChemName: Michler's Ketone

CAS: 90-94-8

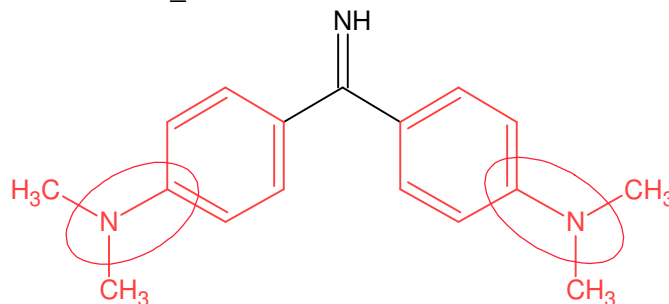
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_33



ChemName: Auramine

CAS: 492-80-8

Mouse_Male : 3

Mouse_Female : 3

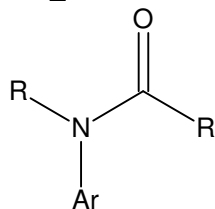
Rat_Male : 3

Rat_Female : ND

Ar = Any aromatic/heteroaromatic ring

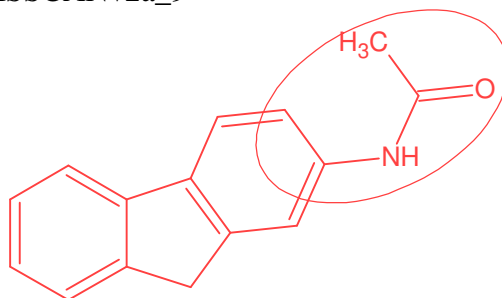
R = Hydrogen, methyl

SA_28ter: aromatic N-acyl amine



- Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (-SO₃H) on the same ring of the amino group are excluded .

ISSCANv2a_9



ChemName: 2-Acetylaminofluorene

CAS: 53-96-3

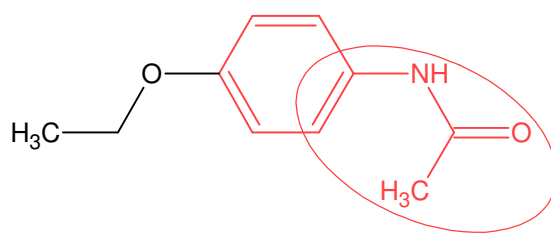
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_565



ChemName: Phenacetin

CAS: 62-44-2

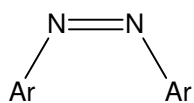
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

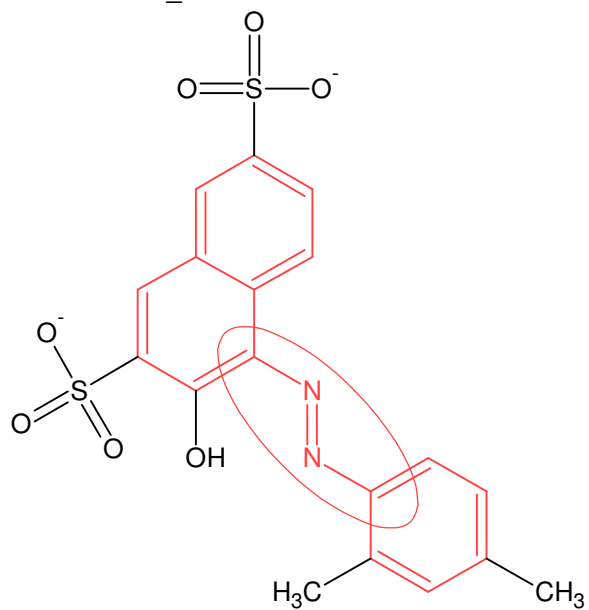
SA_29: Aromatic diazo



Ar = Any aromatic/heteroaromatic ring

- Chemicals with a sulfonic acid group (-SO₃H) on both rings linked to the diazo group are excluded.

ISSCANv2a_420



ChemName: D&C Red no. 5

CAS: 3761-53-3

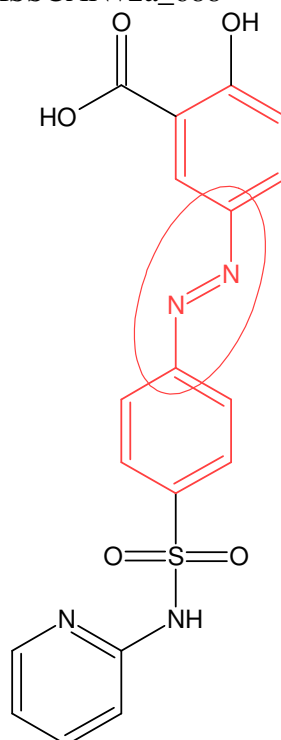
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_688



ChemName: Salicylazosulfapyridine

CAS: 599-79-1

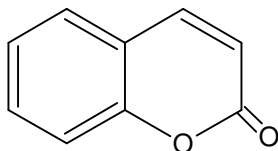
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

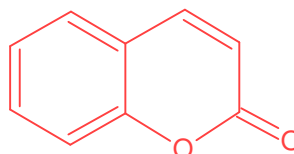
Rat_Female : 3

SA_30: Coumarins and Furocoumarins



Any substance with the displayed substructure

ISSCANv2a_318



ChemName: Coumarin

CAS: 91-64-5

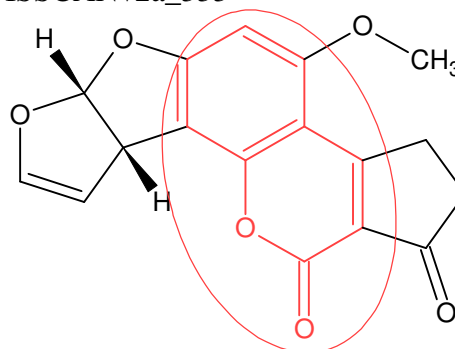
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 2

ISSCANv2a_333



ChemName: Aflatoxin B1

CAS: 1162-65-8

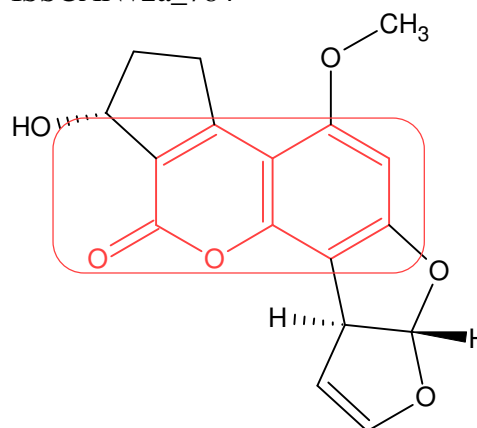
Mouse_Male : 1

Mouse_Female : 1

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_784



ChemName: Aflatoxicol

CAS: 29611-03-8

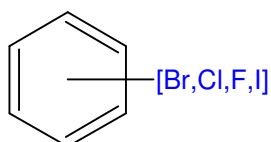
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

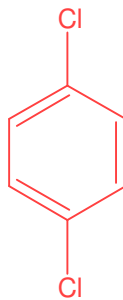
Rat_Female : ND

**SA_31a: Halogenated benzene
(nogenotoxic)**



- Chemicals with two halogens in ortho or meta are excluded.
- Chemicals with three or more hydroxyl groups are excluded.

ISSCANv2a_259



ChemName: 1,4-Dichlorobenzene

CAS: 106-46-7

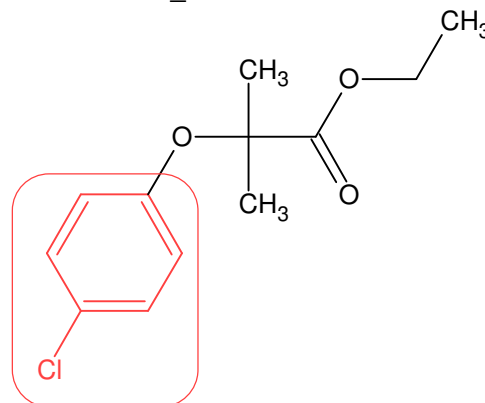
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 1

ISSCANv2a_433



ChemName: Ethyl 2-(4-chlorophenoxy)-2-methylpropionate

CAS: 637-07-0

Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

Rat_Female : 3

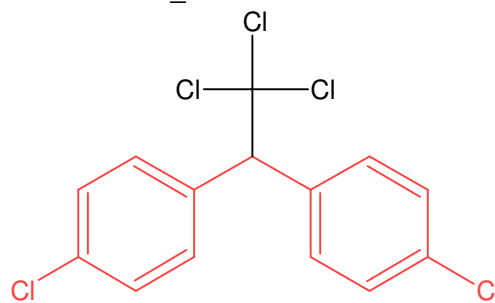
SA_31b: Halogenated PAH (nogenotoxic)

[Br,Cl,F,I]

Ar

Ar = naphthalene, biphenyl, diphenyl

ISSCANv2a_751



ChemName: DDT

CAS: 50-29-3

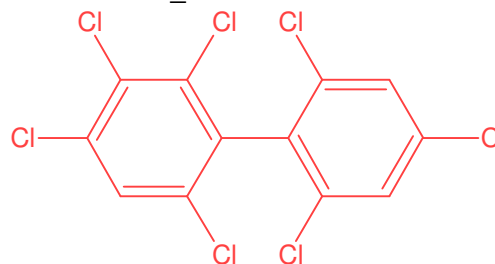
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_393



ChemName: Aroclor 1260

CAS: 11096-82-5

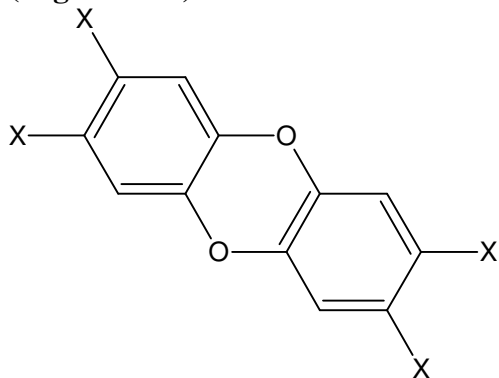
Mouse_Male : ND

Mouse_Female : ND

Rat_Male : 3

Rat_Female : 3

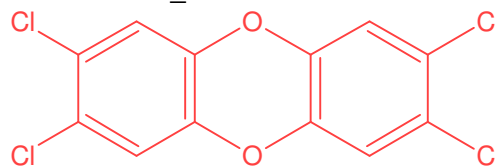
**SA_31c: Halogenated dibenzodioxins
(nogenotoxic)**



X= F, Cl, Br, I

Only chemicals with at least one halogen in one of the four lateral positions are considered.

ISSCANv2a_65



ChemName: 2,3,7,8-Tetrachlorodibenzo-p-dioxin

CAS: 1746-01-6

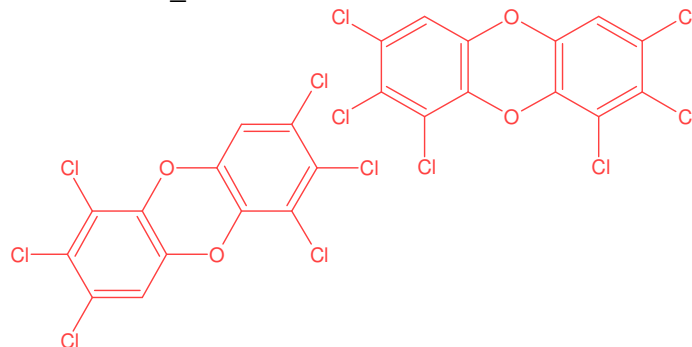
Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 3

Rat_Female : 3

ISSCANv2a_349



ChemName: HCDD mixture

CAS: 57653-85-7 & 19408-74-3 (34465-46-8)

Mouse_Male : 3

Mouse_Female : 3

Rat_Male : 2

Rat_Female : 3

Legend:

The appendix displays the Structural alerts used in this expert system.

For each alert, essential specifications with rules for non applicability (i.e., modulating factors) are given.

The selected examples shown were retrieved from the ISSCAN database, and are provided with:

Identification code in ISSCAN (e.g., ISSCANv2a_349);

ChemName: Chemical Name;

CAS: Registry Number of the Chemical Abstract Service;

Rat_Male ; Rat_Female ; Mouse_Male ; Mouse_Female : Carcinogenicity results in the four experimental groups most commonly used for the cancer bioassay, where the outcomes codes are:

1 = noncarcinogen; 2 = equivocal or borderline; 3 = carcinogen; ND: Not Done.

The ISSCAN database can be freely downloaded from:

<http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7>

or

http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html

Appendix 2: QSAR models

QSAR 6: Mutagenic activity of aromatic amines in *Salmonella thyphimurium* TA100 (with S9 metabolic activation)

$$w = - 3.14 \text{ HOMO} + 1.76 \text{ LUMO} + 0.62 \text{ MR}_2 + 0.75 \text{ MR}_3 + 1.88 \text{ MR}_6 + 3.75$$

Idist

$$w(\text{mean,Class1}) = 28.42 \quad N1 = 47 \text{ (non-mutagens)}$$

$$w(\text{mean,Class2}) = 26.44 \quad N2 = 64 \text{ (mutagens)}$$

$$\text{Threshold} = 27.43$$

Descriptors:

The PM3 molecular orbital energies for the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular orbital (LUMO) are given in eV.

MR₂, MR₃, MR₆ are the Molar Refractivity (MR) contributions of substituents in position 2, 3, and 6 to the amino group. The values are multiplied x 0.1.

The indicator variable Idist is a structural parameter coding for the presence (Idist = 1, otherwise Idist = 0) of substituents on the positions 3-, 4- and 5- of 4-aminobiphenyl (e.g.: 4'-nButyl-4-aminobiphenyl; 4'-tButyl-4-aminobiphenyl; 4'-Trifluoromethyl-4-aminobiphenyl; 3'-Trifluoromethyl-4-aminobiphenyl).

Model statistics:

The Squared Canonical Correlation of the model is 0.52. The equation correctly reclassified 87.4 % (Accuracy) of the compounds (Class1, nonmutagens, 95.7 % (Specificity); Class2, mutagens, 81.3 % (Sensitivity)).

The application of cross-validation to QSAR 6 resulted in the following accuracy values (with Standard Deviation): a) 10%-out: 87.1 (1.1); b) 25%-out: 86.5 (1.8); c) 50%-out: 87.4 (4.2).

External validation:

The model was tested for its external predictivity on a set of aromatic amines retrieved from the literature and not included in the training set, with the following results:

Accuracy = 81%; Sensitivity = 86%; Specificity = 72%

(Negatives = 18 ; Positives = 29)

Applicability Domain:

The model applies only to homocyclic amines, and excludes aromatic amines containing aromatic nitro groups as well.

This QSAR applies also to chemicals containing diazo, isocyanate and immine groups, that are considered as precursor of the corresponding aromatic amine.

Reference:

R. Benigni, C. Bossa, T. Netzeva, A. Rodomonte, and I. Tsakovska (2007) Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environ.Mol.Mutag.* 48: 754-771.

QSAR 8: Carcinogenicity of aromatic amines in rodents (mice, rats)

$$w = - 3.79 L(R) + 3.52 B_5(R) - 4.12 \text{ HOMO} + 4.41 \text{ LUMO} + 3.09 \text{ MR}_3 \\ + 2.60 \text{ MR}_5 + 4.63 \text{ MR}_6 - 3.49 \text{ I(An)} + 1.80 \text{ I(NO}_2) - 1.78 \text{ I(BiBr)}$$

$$w(\text{mean,class1}) = 27.82 \quad N1 = 12 \quad (\text{non-carcinogens})$$

$$w(\text{mean,class2}) = 30.34 \quad N2 = 52 \quad (\text{carcinogens})$$

$$\text{Threshold} = 29.08$$

Descriptors:

$L(R)$ (length) and $B_5(R)$ (maximal width) are Sterimol parameters.

The PM3 molecular orbital energies for the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular orbital (LUMO) are given in eV.

MR_3 , MR_5 , MR_6 are the Molar Refractivity contributions of substituents in position 3, 5, and 6 to the amino group. The values are multiplied x 0.1.

$I(\text{An})$, $I(\text{NO}_2)$ and $I(\text{BiBr})$ are indicator variables that take value = 1 for anilines, for the presence of a NO_2 group, and for biphenyls with a bridge between the phenyl rings, respectively.

Model statistics:

The Squared Canonical Correlation of the model is 0.50. The equation correctly reclassified 95.3 % (Accuracy) of the compounds (Class1, non-carcinogens, 100 % (Specificity); Class2, carcinogens, 94.2 % (Sensitivity)).

The application of cross-validation to QSAR 8 resulted in the following accuracy values (with Standard Deviation): a) 10%-out: 78.3 (13.0); b) 25%-out: 83.8 (7.3); c) 50%-out: 83.4 (5.7).

External validation:

The model was tested for its external predictivity on a set of aromatic amines retrieved from the literature and not included in the training set, with the following results:

Accuracy = 70%; Sensitivity = 92%; Specificity = 46%

(Negatives = 13 ; Positives = 14)

Applicability Domain:

The model applies only to homocyclic amines. The model includes also aromatic amines containing aromatic nitro groups.

This QSAR applies also to chemicals containing diazo, isocyanate and immine groups, that are considered as precursor of the corresponding aromatic amine.

References:

Franke R., A. Gruska, A. Giuliani, and R. Benigni (2001) Prediction of rodent carcinogenicity of aromatic amines: a quantitative structure-activity relationships model. *Carcinogenesis*, 22: 1561-1571.

R. Benigni, C. Bossa, T. Netzeva, A. Rodomonte, and I. Tsakovska (2007) Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environ.Mol.Mutag.* 48: 754-771.

QSAR 13: Mutagenic activity of $\alpha\beta$ -unsaturated aliphatic aldehydes in *Salmonella thyphimurium* TA100 (without S9 metabolic activation)

$$w = 0.387 \text{ MR} - 3.12 \text{ logP} + 3.23 \text{ LUMO}$$

$$w(\text{mean,class1}) = 9.69 \quad N1 = 3 \quad (\text{non-mutagens})$$

$$w(\text{mean,class2}) = 6.37 \quad N2 = 17 \quad (\text{mutagens})$$

$$\text{Threshold} = 8.03$$

Descriptors:

MR is the Molar Refractivity of the whole molecule.

LogP is the logarithm of the partition coefficient between octanol and water.

The PM3 molecular orbital energies for the Lowest Unoccupied Molecular orbital (LUMO) are given in eV.

Model statistics:

The Squared Canonical Correlation of the model is 0.61. The equation correctly reclassified 100 % of the compounds.

The application of the Leave-One-Out cross-validation to QSAR 13 resulted in 85% accuracy. Given to the small number of negatives, no other cross-validation procedures were applicable.

External validation:

The model was tested for its external predictivity on a set of $\alpha\beta$ -unsaturated aldehydes not included in the training set and tested *ad hoc* for the validation work, with the following results:

Accuracy = 100%

(Negatives = 3 ; Positives = 2)

Applicability Domain:

The QSAR applies to linear aldehydes (no unsaturated bond in a cycle). Are also excluded chemicals with additional SAs (in these cases, other reactions not modeled by this QSAR may take place).

References:

R. Benigni, L. Passerini, and A. Rodomonte (2003) Structure-activity relationships for the mutagenicity and carcinogenicity of simple and $\alpha\beta$ -unsaturated aldehydes. *Environ.Mol.Mutag.*, 42: 136-143.

R. Benigni, L. Conti, R. Crebelli, A. Rodomonte, and M. R. Vari (2005) Simple and $\alpha\beta$ -unsaturated aldehydes: correct prediction of genotoxic activity through structure-activity relationship models. *Environ.Mol.Mutag.*, 46: 268-280

European Commission

EUR 23241 EN – Joint Research Centre – Institute for Health and Consumer Protection

Title: The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree

Author(s): Romualdo Benigni , Cecilia Bossa, Nina Jeliaskova, Tatiana Netzeva and Andrew Worth

Luxembourg: Office for Official Publications of the European Communities

2008 – 68 pp. – 21 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract

The Joint Research Centre's European Chemicals Bureau has developed a hazard estimation software called Toxtree, capable of making structure-based predictions for a number of toxicological endpoints. One of the modules developed as an extension to Toxtree is aimed at the prediction of carcinogenicity and mutagenicity. This module encodes the Benigni/Bossa rulebase for carcinogenicity and mutagenicity developed by Romualdo Benigni and Cecilia Bossa at the Istituto Superiore di Sanita', in Rome, Italy. The module was coded by the Toxtree programmer, Ideaconconsult Ltd, Bulgaria. In the Toxtree implementation of this rulebase, the processing of a query chemical gives rise to limited number of different outcomes, namely: a) no structural alerts for carcinogenicity are recognised; b) one or more structural alerts (SAs) are recognised for genotoxic or non-genotoxic carcinogenicity; c) SAs relative to aromatic amines or $\alpha\beta$ -unsaturated aldehydes are recognised, and the chemical goes through Quantitative Structure-Activity Relationship (QSAR) analysis, which may result in a negative or positive outcome. If the query chemical belongs to the classes of aromatic amines or $\alpha\beta$ -unsaturated aldehydes, the appropriate QSAR is applied and provides a more refined assessment than the SAs, and should be given higher importance in a weight-of-evidence scheme. This report gives an introduction to currently available QSARs and SAs for carcinogenicity and mutagenicity, and provides details of the Benigni/Bossa rulebase.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

