Review DOI: 10.1515/aiht-2016-67-2801

Mutagenic and carcinogenic structural alerts and their mechanisms of action

Alja Plošnik[1], Marjan Vračko[1], and Marija Sollner Dolenc[2]

National Institute of Chemistry Slovenia[1], Faculty of Pharmacy, University of Ljubljana[2], Ljubljana, Slovenia

[Received in March 2016; CrossChecked in March 2016; Accepted in September 2016]

Knowing the mutagenic and carcinogenic properties of chemicals is very important for their hazard (and risk) assessment. One of the crucial events that trigger genotoxic and sometimes carcinogenic effects is the forming of adducts between chemical compounds and nucleic acids and histones. This review takes a look at the mechanisms related to specific functional groups (structural alerts or toxicophores) that may trigger genotoxic or epigenetic effects in the cells. We present up-to-date information about defined structural alerts with their mechanisms and the software based on this knowledge (QSAR models and classification schemes).

KEY WORDS: *carcinogenicity; mutagenicity; software prediction; toxicophores*

Carcinogenicity is an important toxicity endpoint in assessing chemical risk and hazards. The human population is exposed to various chemical agents that may promote one of the three stages of cancer development (initiation, promotion, progression) (1). Usually, the information about carcinogenic chemicals is gathered from animal or epidemiological studies (2). According to the mechanism of action carcinogens can be divided into two major groups: a) genotoxic carcinogens, which directly interact with and damage DNA by changing its structure and b) epigenetic carcinogens, which do not directly damage DNA (through covalent bonds) but affect its expression or make the cell more sensitive to other agents. Epigenetic carcinogens act in a wide range of mechanisms, while genotoxic carcinogens have a quite similar mode of action. These compounds are usually highly reactive electrophilic molecules that interact with the nucleophilic site in DNA. They can be electrophilic *per se* or metabolised to reactive electrophilic intermediates by several cellular processes (3, 4).

In the following sections we give an overview of some known mechanisms of carcinogenic action and of the software used to predict the mutagenicity and carcinogenicity of chemical compounds based on their structure. This software is a valuable tool in identifying and regulating potentially toxic chemicals.

With the efforts made to minimise animal testing (such as the EU ban to test animals for chemicals used in cosmetics industry), researchers have been looking for new, alternative methods to evaluate the toxic properties of molecules for specific endpoints such as carcinogenicity, mutagenicity, reproductive toxicity, developmental toxicity, skin sensitivity, and hepatotoxicity (2).

The review is intended to assist everyone involved in chemical regulation who intend to use *in silico* models for hazard communication, regulatory compliance, and sustainable lifecycle management.

Correspondence to: Professor Marija Sollner Dolenc, Ljubljana University Faculty of Pharmacy, Department of Pharmaceutical Chemistry, Aškerčeva 7, 1000 Ljubljana, Slovenia, E-mail: *marija.sollner@ffa.uni-lj.si*

CARCINOGENESIS

Carcinogenesis is the result of a number of complex, sequential processes within cells and tissues triggered by a variety of molecular and cellular changes. If induced by chemical compounds, carcinogenesis has three stages: initiation, promotion, and progression (Figure 1).

Carcinogenesis begins with a mutation, a stable, heritable change of a genetic material that has escaped DNA repair mechanisms during cell proliferation. In this initiation stage, mutations accumulate because they promote uncontrolled expression of proto-oncogenes, which control the cell cycle, including apoptosis and/or inactivation of tumour-suppressor genes (such as *p53*), which in turn encode enzymes for DNA damage repair. Initiation is a rapid, irreversible process in a number of mutational events triggered by chemical or physical agents (known as initiating agents or genotoxic agents).

The second stage is promotion. Under the influence of other endogenous or exogenous chemical compounds (growth stimuli) the initiated cells are subject to clonal growth, which promotes the tumour. This is why these exogenous and endogenous compounds are called tumour promoters. They are not mutagenic by themselves but trigger other mechanisms, such as changes in gene expression that are continued in all subsequent daughter

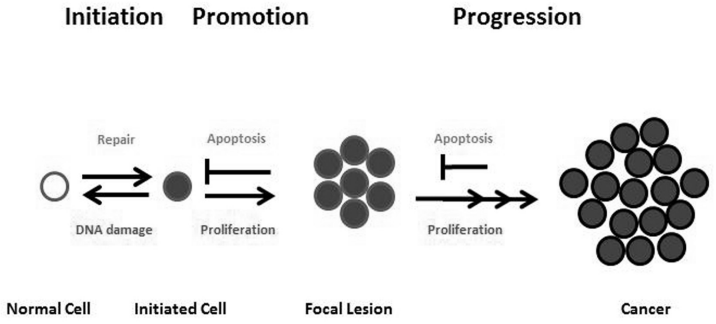


Figure 1 *Multistage carcinogenesis (3)*

cells. Cell proliferation rate increases and apoptotic cell death rate decreases. Promotion is a reversible process and only works in initiated cells. Well known promoters are phenobarbital, benzene, asbestos, and arsenic.

The last stage of carcinogenesis is progression, which involves additional genotoxic events (chromosomal aberrations and translocations). Progression is an irreversible process leading to the formation of neoplasms, benign and malignant alike (5-7).

Chemical carcinogens

Many genetic changes can occur spontaneously due to the presence of rare tautomers in nucleotide bases (keto/enol form) and errors associated with the malfunctioning of DNA polymerases and oxidation of DNA induced by reactive oxygen species (ROS) because of respiratory chain and oxidative enzyme reactions (7). However, more and more studies claim that the large increase in cancer incidence is associated with exposure to chemical carcinogens or with factors such as age, diet, hormonal balance, or environment (8). About 80% of tumours in humans are triggered by exogenous chemical agents and are not necessarily associated with direct exposure to them but may also arise from normal metabolism, oxidative stress, or chronic inflammation. Chemical genotoxic carcinogens are divided into two main groups: direct-acting carcinogens and indirect-acting carcinogens. Direct-acting carcinogens cause cancer without metabolic activation or chemical modification (activation-independent), as they damage DNA from within. These chemicals are also known as parent compounds or ultimate carcinogens (9). The most common are epoxides, imines, and alkyl and sulphate esters. Indirect-acting carcinogens become carcinogenic after metabolic activation. Typical indirect carcinogens are polycyclic aromatic hydrocarbons (PAHs, benzo[a]pyrene in particular), nitrosamines, nitrosoureas, and aromatic amines (10-12).

Photogenotoxicity

One of the ways for chemical compounds to become genotoxic intermediates is through light activation (phototoxicity/photogenotoxicity). These compounds absorb light (UV, visible, and IR) and convert to another form (photomodification), degrade under the influence of light (photodegradation) (13), or reach an excited state (photoexcitation). Some compounds such as psoralens and phenothiazines affect DNA directly through photoexcitation, some such as porphyrins and riboflavins excite the surrounding molecules (such as chromophores), and some (such as furocumarin hydroperoxides and peroxy esters) react with DNA via ROS (14). Whichever the photogenotoxic mechanism, the compound must be excited close to the target (DNA) (13, 14).

The most common changes affecting DNA are pyrimidine dimers, covalent adducts, base modifications generated by oxidation, single-strand breaks, and base loss. Under the influence of light, polycyclic aromatic hydrocarbons (PAHs) tend to form adducts that covalently bind to DNA or cause DNA strand breaks (13). In these cases DNA repair mechanisms often fail, which can lead to photomutagenesis or even the initiation of photocarcinogenesis. Photogenotoxicity can be experimentally studied with different techniques such as HPLC coupled with tandem mass spectrophotometry (HPLC/MS-MS) or HPLC coupled with an electrochemical detector (HPLC/ECD) combined with the *in vitro* Comet assay or similar *in vitro* methods (15).

Electrophiles as carcinogens

In their ultimate form, direct or indirect-acting chemical carcinogens (after metabolic activation) work as reactive electrophiles (11). These compounds form covalent adducts with most of the cellular informational macromolecules (nucleophiles) such as DNA, RNA, or proteins. Nucleophiles contain nucleophilic sites that typically include electron-rich (unpaired electrons) heteroatoms such as S (side chain of cysteine residues - thiol groups, S-atoms of methionine), O (alcohol group of the phenolic amino acid of tyrosine,

serine, and threonine residues), and N (primary amino-groups of lysine or arginine, secondary amino-group of histidine) (16). The most vulnerable cellular nucleophiles are nucleic acids, DNA bases guanine, adenine, cytosine, and thymine in particular. The most vulnerable to adduct forming are the following base sites: N^1, N^3, and N^7 of adenine, N^2, N^7, and O^6 of guanine, N^3 and O^2 of cytosine, and N^3, O^2, and O^4 of thymine (17, 18). The most common reaction between electrophiles and nucleophiles is alkylation, especially of purine at the N^7 site of guanine. Alkylation may also occur at the O^6 site of guanine and other bases, but more slowly. The O^6 site is important for mutagenicity. The newly formed ether covalent bond is known to change the electronic distribution around the base, which leads to deprotonation at site N^1. This changes the pattern of hydrogen bonds and causes misconnections between the bases (16, 19). The targets of alkylation are not only nitrogen bases but also phosphodiester bonds (phosphate alkylation).

Each carcinogenic electrophile has a specific activation target, specific metabolism, and specific binding site on the nucleophile. Strong electrophiles are small, poorly polarised molecules, and their electron deficiency shows as positive electron charge. Soft electrophiles are usually large, highly polarised molecules, and their electron deficiency spreads all over the molecule (20, 21). Strong (hard reactive) electrophiles include nitroso compounds, epoxides, α,β-unsaturated aledhydes, *N*-sulphonyloxy-*N*-methyl-4-aminoazobenzene, *N*-sulphonyloxy-*N*-acetyl-2-aminofluorene, and *N*-hydroxy-2-aminofluorene. Soft electrophiles include safrole, estragole, *N*-methylol-4 aminoazobenzene (17, 22).

Another important chemical event in neoplasm formation is the hydroxylation of DNA bases as a result of interaction between a hydroxyl radical ($OH^{\cdot}$) and base. The resulting products include 8-hydroxyguanine (8-OH-dG), 8-hydroxyadenine (8-OH-dA), 5-hydroxyuracil (5-OH-dU), 5-hydroxycytosine (5-OH-dC), thymine glycol, and uracil glycol (23).

Structural alerts for genotoxicity

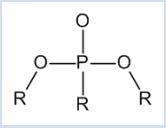
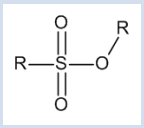
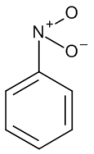
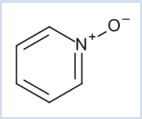
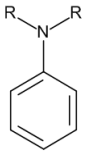
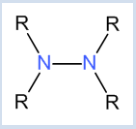
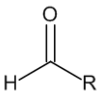
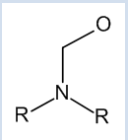
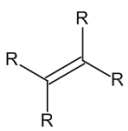
In 1985, John Ashby introduced structural alerts as a way to predict genotoxicity (Figure 2) (1). Benigni and Bossa (24) summarised structural alerts as follows: "The Structural Alerts are molecular substructures or reactive groups that are related to the carcinogenic and mutagenic properties of the chemicals, and represent a sort of 'codification' of a long series of studies aimed at highlighting the mechanisms of action of the mutagenic and carcinogenic chemicals". Structural alerts are very helpful not only in the classification of potential carcinogens, but are also important in understanding the mechanisms of genotoxicity (24-28).

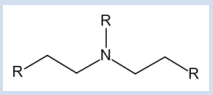
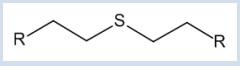
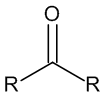
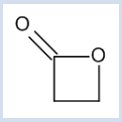
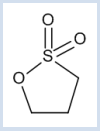
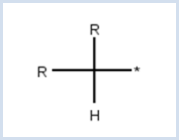
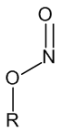
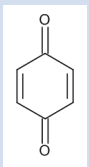
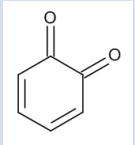
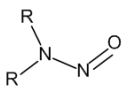
Electrophilic chemical reaction mechanisms forming adducts with DNA

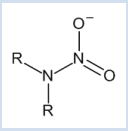
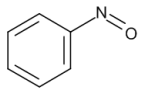
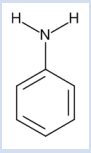
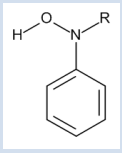
The understanding of the mechanisms of action through which electrophiles react with biological nucleophiles is based on the classical chemical reactions: conjugation, substitution, and addition, in which the electron-rich component interacts with the electron-deficient one (29). We know of about fifty mechanisms of covalent binding, but only six can lead to cancer: S_N, S_N2, acylation, Schiff base formation, Michael addition, and S_NAr (Figure 3). These mechanisms enable us to classify different electrophiles into appropriate mechanistic domain (Table 1)(30). S_N1 is a first-order, two-stage nucleophilic substitution. In the first stage, the leaving group is branched off, and in the second stage, the resulting carbenium ion reacts with a nucleophile. S_N2 is a second-order nucleophilic substitution where the leaving group disconnects and attacks a nucleophile in a single step (31). Acylation is a reaction where reactive tetrahedral intermediate is formed with a nucleophile, and the leaving group is released. Schiff base formation is a mechanism where electrophilic carbon atoms of aldehydes or ketones are attacked by amines and the C=O double bond is replaced by the C=N double bond (32). Michael addition is a two-step nucleophilic addition, where nucleophile attacks a double bond and forms two new bonds. S_NAr is electrophilic aromatic substitution in which nucleophile attacks the leaving group and the aromatic structure stays unchanged (33, 34).

Consequences of endogenous DNA adduct formation

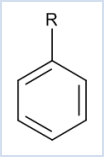
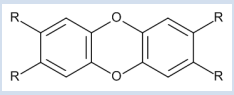
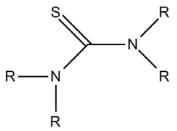
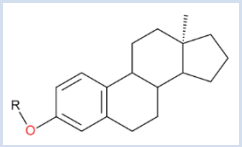
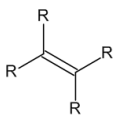
DNA adducts may trigger different structural changes in genetic material (1, 3, 12). Different chemicals react with different DNA bases on different DNA positions (4). The most common types of damage caused by DNA adducts are base oxidation, ethenobases, alkylation (usually methylation), and base hydrolysis (deamination or depurination). In deamination ammonia is released, and cytosine (adenine) is transformed into uracil (hypoxanthine), which causes the binding of adenine (cytosine) instead of guanine (thymine) when DNA is replicated. Depurination is the cleaving of the *N*-glycoside bond between the purine base and deoxyribose in DNA, leading to the formation of an apurinic site and structural change. Depurination of bases on a single-stranded DNA during replication can lead to mutation, because an incorrect base is added to the apurinic site (35). This can lead to transversion or transition (36). Electrophilic PAH metabolites are well-known examples of adducts that promote depurination. They are also capable of forming bulky DNA adducts. These PAH-DNA adducts can trigger nucleotide excision repair or affect important regulatory genes such as *Ras* and *p53* (37, 38). Bulky adducts can form a stable intercalation between the bases. Intercalators are planar molecules that intercalate between the base pairs in the double-stranded DNA. Intercalation changes the shape of the double helix and can cause

Structural Alerts	Examples
Alkyl esters of either phosphonic or sulphonic acids  or 	Tris(2,3-dibromopropyl) phosphate Ethyl methanesulphonate
Aromatic nitro groups 	*O*-nitroanisole 2-nitrotoluene
Aromatic *N*-oxides 	*N*'-nitrosonornicotine-1-*N*-oxide
Aromatic mono- and dialkylamino groups 	Michler's ketone Auramine
Alkyl hydrazines 	
Simple aldehydes 	Formaldehyde Acetaldehyde Benzaldehyde
N-methylol derivates 	Hexamethylolmelamine *N*-methylolacrylamide
Monohaloalkenes  R=[Br, Cl, F, I]	Vinyl chloride Dimethylvinyl chloride

S- or N-mustards  **or**  R=[Br, Cl, F, I]	Chloroambucil Bis(2-chloroethyl)sulphide
Acyl halides  R=[Br, Cl, F, I]	Dimethylcarbamoyl chloride
Propiolactones and propiolsultones  **or** 	Beta-butyrolactone Beta-propiolactone
Epoxides and aziridines  or. 	Ethylene oxide Ethyleneimine Thiotepa
Aliphatic halogens  *=[Br, Cl, F, I]	1,2-dibromoethane 1,2-dichloroethane
Alkyl nitrite 	Isobutyl nitrite
Quinones  **or** 	9,10-anthraquinone Chrysazin
Alkyl and aryl *N*-nitroso group 	1-(2-hydroxyethyl)-1- Nitrosourea *N*-nitrosodimethylamine

Aliphatic *N*-nitro group 	Dimethylnitramine *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine
Aromatic nitroso group 	*o*-nitrosotoluene 4-nitrosodiphenylamine
Aromatic amines and hydroxylamine  or 	para-Cresidine 2-aminodipyrido[1,2-a:3',2'-d]imidazole

Nongenotoxic

Structural alert	Example
Halogenated benzene  R=[Br, Cl, F, I]	1,4-Dichlorobenzene Ethyl 2-(4-chlorophenoxy)-2-methylpropionate
Halogenated polycyclic aromatic hydrocarbon (PAH)  R=[Br, Cl, F, I]	Benzo[a]pyrene DDT Aroclor 1260 Dihydrodiol epoxides
Halogenated dibenzodioxins  R=[Br, Cl, F, I]	2,3,7,8-tetrachlorodibenzo-*p*-dioxin 1,2,3,6,7,8-hexachlorodibenzodioxin
Thiocarbonyls 	Thiosemicarbazide 2-methyl-3-thiosemicarbazide
Steroidal oestrogens 	Catechols 2-methoxyestrone 4-hydroxyestradiol 2-hydroxyestrone
Trichloro (/fluoro) or. Tetrachloro (/fluoro) ethylene  R=[Cl, F]	Chloroethylene Tetrafluoroethylene

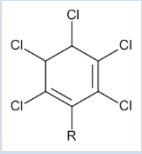
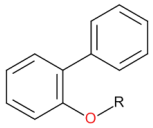
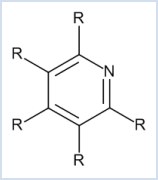
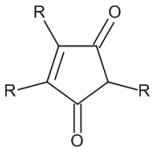
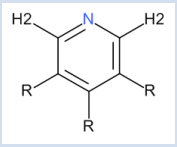
Pentachlorophenol 	Pentachlorophenol (PCP) Chloranil
o-Phenylphenol 	Sodium ortho-phenylphenate
Imidazole 	4-methylimidazole 2,5-diarylimidazoles Imidazole carboxamide
Dicarboximide 	*N*-octyl bicycloheptene dicarboximide Perylene-3,4-dicarboximide
Dimethylpyridine 	3,5-benzoyl-4-(3-chlorophenyl)-1,4-dihydro-2,6-dimethylpyridine

Figure 2 *Upgrade of Ashby's structural alerts with genotoxic potential and their examples (1, 2, 24-28)*

mutations (insertions and deletions), block replication and transcription, or affect DNA repair mechanisms (39).

All of these events can lead to DNA base changes (5-hydroxyuracil, 5-hydroxycytosine, 7,8-dihydro-8-oxoguanine, 7,8-dihydro-8-oxoadenine, 2,6-diamino-4-hydroxy-5-formamidopyrimidine, or 5-hydroxypyrimidines), base loss, base substitutions (mostly G>T transversions), frameshift mutations, insertion, sequence amplification, generation of single- and double-strand breaks, sister chromatid exchanges, DNA-protein cross-links, and pyrimidine dimer formations (40-45).

Epigenetic mechanisms of carcinogenic molecules

Exposure to some chemical carcinogens may cause cancer without changes in the nucleotide sequences. Epigenetic factors are common in cells that are constantly under stress. Such chemicals do not form DNA adducts nor do they alter DNA but affect the expression of certain genes (28). All epigenetic factors (physical, chemical, and biological) mainly operate in two ways: either via methylation or via post-translational modifications of histones (acetylation). DNA methylation occurs at the promoter region, which contains CpG islands (cytosine and guanine nucleotides linked with a phosphodiester bond) (46) and results in the conversion of cytosine to 5-methylcytosine, which has a much higher mutagenic potential. There are two mechanisms of methylation: hypermethylation and hypomethylation. Hypermethylation usually occurs at CpG islands and may affect genes involved in the cell cycle, DNA repair mechanisms, intercellular interactions, and apoptosis. Hypermethylation may also increase deamination of 5-methylcytosine to thymine, leading to C to T conversion. In contrast, hypomethylation at CpG sites may lead to the overexpression of oncogenes, chromosome instability, and metastases (47).

Acetylation of histones is regulated by histone acetyl transferases (HATs), which play an important role in chromatin transformation and in the regulation of gene transcription (48). Acetylation of lysine residues neutralises

histones in the nucleosome and thus reduces their affinity to negatively charged DNA. Reduced affinity leads to the decondensation of DNA and eventually to transcription (49). Histone acetylation does not affect histone affinity to DNA alone but also affects the interaction between histones and histone interaction with other regulatory proteins. In other words, histone acetylation may affect processes such as replication, formation of nucleosomes, and chromatin packaging (50).

Cell exposure to oxidative stress and DNA damage

When we are talking about chemical carcinogenesis we cannot skip oxidative stress caused by reactive oxygen/nitrogen species (ROS/RNS) (46). The reactive species most often involved in carcinogenesis are superoxide anion radical $O_2^{\cdot-}$, hydroxyl radical ($HO^{\cdot}$), nitric oxide (NO), nitrous acid (HNO_2), peroxynitrite ($ONOO^-$), hydroperoxyl radical ($HO_2^{\cdot}$), ozone (O_3), and hydrogen peroxide (H_2O_2). These compounds can be generated by inflammation, radiation, interruption of mitochondrial oxidative phosphorylation, or xenobiotic metabolism, but most of them are generated by redox cycling induced by chemical carcinogens that contain structural alerts, such as halogenated compounds, aromatic hydrocarbons, aromatic *N*-oxides, quinones, aromatic nitro compounds, conjugated imines, heterocyclic amines, and pyridyl compounds (5, 30). Oxidative stress arises when the redox balance is disrupted and the number of ROS/RNS exceeds the number of natural cell defence molecules (antioxidants), which leads to DNA, protein, and lipid damage (51). Directly acting oxidative stress can cause structural DNA changes such as base pair substitution, deletion, insertion, base oxidation, guanine-to-cytosine or thymine-to-adenine transversion, double- or single strand-breaks, deamination of guanine and adenine, nitration of guanine, or modification of purine/pyrimidine nucleosides (30). Indirectly acting oxidative stress can change the membrane, cytoplasmic, and nuclear signal transduction pathways, modulate genes that increase cell proliferation or differentiation, and inhibit programmed cell death (apoptosis). Not only can ROS generate mutations but can also interfere with DNA repair (4).

The role or endocrine disruptors in carcinogenicity

In recent years the role of endocrine disruptors in carcinogenesis has received a lot of attention. Endocrine disrupting chemicals (EDC) can interfere with and affect the endocrine system, which can lead to hormone-related cancers (breast, testicular, prostate, or leukaemia). The main mechanism of EDC action is that they bind to the active sites of oestrogen, androgen, and thyroid receptors (52). They can trigger the same response as natural hormones, agonistic or antagonistic alike. Ligand-receptor interactions most often result in changes in the transcription genes (53), which, in turn, changes cell activities such as regulation/stimulation of cell proliferation, regulation of gene expression, gene signalling, and hormone metabolism, biosynthesis, bioactivation, and degradation. Changes in hormonal levels affect DNA methylation, histone modifications, or apoptosis (54). EDCs show some structural similarity to non-genotoxic compounds such as phenolic compounds, PAHs, isoflavonoids, stilbenes, and compounds with steroid structure (54-56).

Other factors determining the carcinogenic potential of chemical compounds

Carcinogenicity and mutagenicity are not related only to structural alerts. Certain compounds may contain all of the structural alerts, but are not metabolically active inside the cell. Among the physiochemical factors that may hinder the functioning of these toxic molecules are: i) *molecular weight* and the size of chemicals: the higher the molecular weight of compounds, the lesser the chance they will be absorbed in significant amounts; ii) *state of matter*: affects the ability of a compound to reach the critical point; iii) *solubility*: highly hydrophilic compounds are poorly absorbed by the cell membrane, and, if absorbed, readily excreted; iv) *geometry of chemicals*: planar shape of a molecule has the highest carcinogenic potential; and v) *chemical reactivity*: chemicals that are highly reactive have lower toxic potential because they hydrolyse or polymerise spontaneously or react with non-critical cellular components before they reach the target (29). Other factors include the stability of a compound, transport through the membrane, and half-life (2, 57).

Software packages for mutagenicity and carcinogenicity predictions

Due to ethical reasons, reduced resources, and time savings, toxicity testing of chemicals in animals is getting more and more restrictive. This is why chemical and related industries have started to adopt the "3R" (replacement, reduction, and refinement) principle. Two major alternatives to *in vivo* animal testing are *in vitro* techniques and *in silico* computer simulation.

In the last decade, a number of computer programs have been developed to assess the mutagenicity and carcinogenicity based solely on chemical structures as input. Below we present some of the most common software packages for predicting mutagenicity and carcinogenicity of chemical compounds.

VEGA platform and CAESAR

The VEGA platform serves to access a number of QSAR models for predicting mutagenicity and carcinogenicity such as the Computer-Assisted Evaluation of Industrial Chemical Substances According to Regulation (CAESAR). CAESAR is a software tool which was specifically dedicated to develop QSAR mutagenicity models for the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation in collaboration with the

Table 1 *Structural alerts which belong to certain mechanistic domain (30, 32, 33)*

Mechanistic domain	Structural alerts
S_N2	Alkyl esters of either phosphonic or sulphonic acids Monohaloalkenes S- or N-mustards Propiolactones and propiolsultones Epoxides and aziridines Aliphatic halogens Alkyl nitrite
S_N1	Aromatic nitro groups Alkyl hydrazines Alkyl and aryl *N*-nitroso groups Aliphatic *N*-nitro group Aromatic nitroso group Aromatic amines and hydroxylamine Halogenated polycyclic aromatic hydrocarbon (PAH) Halogenated dibenzodioxins
Acylation	Aromatic diazo groups Acyl halides
Schiff Base Formation	Simple aldehydes *N*-methylol derivates
Michael addition	Quinones
S_NAr	Aromatic *N*-oxides, Aromatic mono- and dialkylamino groups Halogenated benzene

United States Environmental Protection Agency (more information is available at: http://www.caesar-project.eu/). Models for predicting mutagenicity are based on a set of 4225 molecules tested with the Ames bacterial test. Models for carcinogenicity were built on a set of 805 chemicals from the Carcinogenic Potency Database (CPDBAS) (58). CAESAR meets all five principles of the Organisation for Economic Co-operation and Development (OECD). It has good predictive capabilities for mutagenicity and carcinogenicity but, unfortunately, it does not include prediction models for genotoxicity (59).

DEREK

The Deductive Estimation of Risk from Existing Knowledge (DEREK), developed by LHASA Limited, (Leeds, Great Britain) is a software package that predicts whether a particular substance triggers toxic response based on structural similarity with known toxic compounds and their structural alerts associated with specific endpoints (genotoxicity, mutagenicity, and carcinogenicity). DEREK contains over 75 rules for the Ames mutagenicity endpoint predictions, which are based on empirical relationships and mechanisms of action. The model includes 89 structural alerts for mutagenicity, 77 for chromosome aberrations, and 91 for carcinogenicity. Structural alerts causing genotoxicity are composed of mutagenicity and structural alerts based on data from several *in vitro* and *in vivo* mutagenicity tests and other genotoxicity data. The software was developed for research and industry users in collaboration with industry, academia, and regulatory authorities (2, 60).

TOPKAT

The Toxicity Predictions by Komputer Assisted Technology (TOPKAT) is an expert system developed by Accelrys, Inc. (now Biovia, San Diego, CA, USA). Unlike the two above mentioned software tools, TOPKAT is entirely based on two-dimensional electrotopological descriptors but it also relies on the QSAR model. TOPKAT can predict a wide range of toxicological endpoints, including mutagenicity and carcinogenicity. The results of the program are given numerically (0 - inactive compound; 1 - active compound) based on structural similarity with known toxic and nontoxic compounds. Mutagenicity and carcinogenicity models include data derived from bacterial mutagenicity and rodent carcinogenicity tests. The mutagenicity model is based on data for 393 chemicals from the US EPA GeneTox protocol (60).

MultiCASE

This expert system is based on the US FDA and EPA for genotoxicity and carcinogenicity endpoints (MultiCASE Inc., Cleveland, OH, USA). It is often used for pharmaceutical toxicity screening of drug candidates with potential for development. It automatically identifies structural alerts with a potential to initiate high biological activity (toxic response) and analyses statistical parameters to get the final

predictions. Its mutagenicity and genotoxicity models are based on the Ames mutagenicity, direct mutagenicity, base-pair mutagenicity, frameshift mutagenicity, chromosomal aberrations, and sister chromatid exchange data. The carcinogenicity model includes different rodent assays (rate, mouse, male, female, and TD_{50} rats) and human epigenetic studies. All models use the statistical approach with the exception of the rule-based model for the Ames mutagenicity. The models are developed according to the OECD rules, the ICH M7 guidelines for impurities in pharmaceuticals, and the REACH guidelines, and therefore targets pharmaceutical and chemical industry in particular (2, 61, 62).

Another approach for detecting genotoxic compounds that covalently bind to DNA are tools that group similar chemicals into appropriate classes, according to the same structural alerts. Below we present some of the tools that are based on these principles.

QSAR Toolbox

QSAR TOOLBOX is a software application for grouping chemicals into categories and assessing the potential adverse effects of chemicals in cooperation with the European Chemical Agency (ECHA), according to

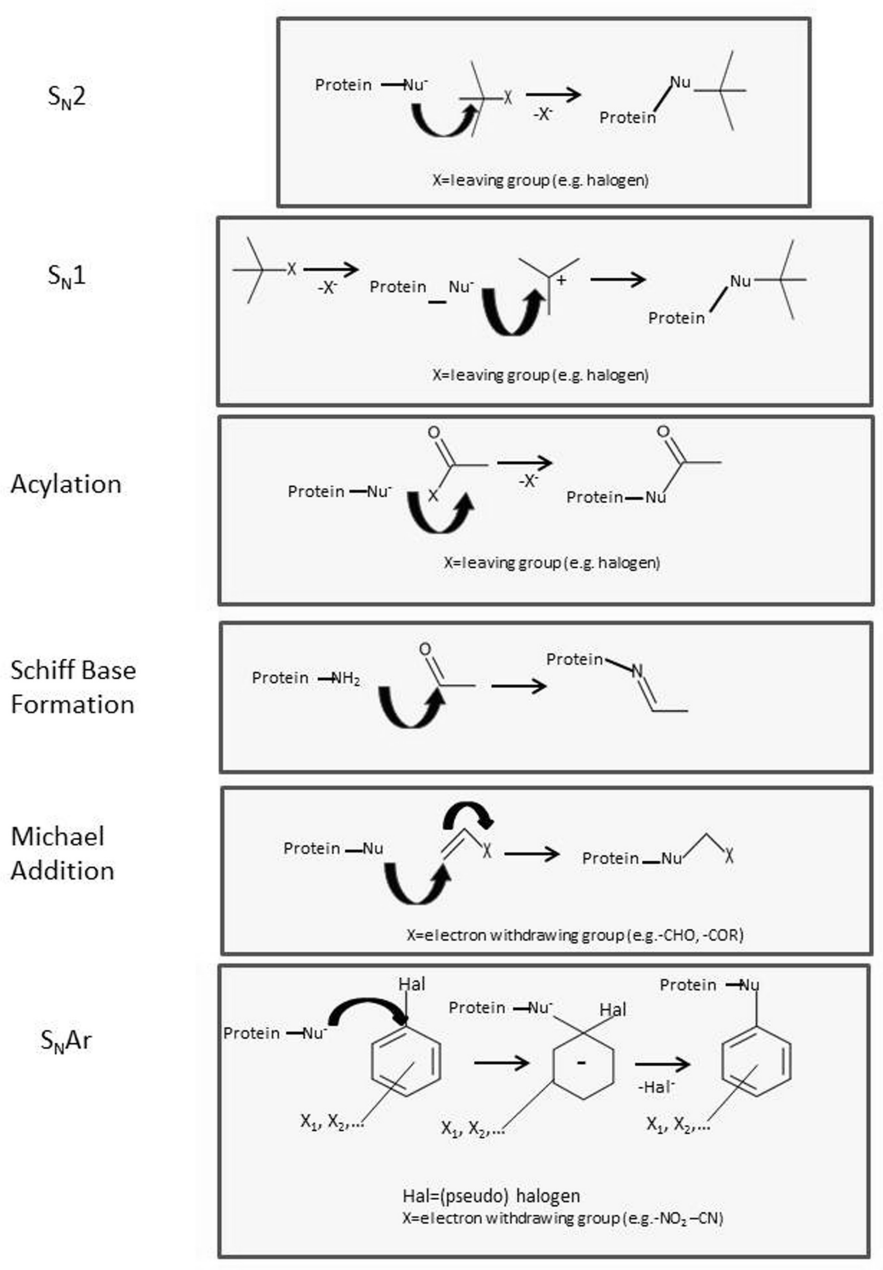


Figure 3 *Mechanisms of covalent binding to cellular nucleophiles (DNA, proteins)*
**Nu- nucleophilic site of molecule*

OECD principles (more information is available at: http://www.qsartoolbox.org/). It is intended for government agencies, chemical industry, and public organisations to provide missing (eco)toxicological information necessary to assess chemical hazard. The QSAR Toolbox systematically groups chemicals into categories according to their structural, physicochemical, and toxicological properties. The program identifies structural characteristics and modes of action for a specific target based on experimental information. It allows quick evaluation of chemicals for common mechanisms or modes of action as well as for common toxicological behaviour or consistent trends among results related to regulatory endpoints. The experimental information has been pooled from three databases: ISSSTY (bacterial mutagenicity database), CPDB (Carcinogenicity Potency Database for carcinogenicity), and ISSCAN (carcinogenicity and mutagenicity database (57, 61, 63).

Toxtree

Toxtree is a user-friendly, open-source application that predicts various kinds of toxic effects using decision trees to place chemicals into appropriate categories. It estimates mutagenicity and carcinogenicity potentials based on the Benigni-Bossa rules for mutagenicity and carcinogenicity, structural alerts for identification of Michael acceptors, and structural alerts confirmed by positive *in vivo* micronucleus tests. If a structural alert is present in the molecule, the program recognises the mutagenic and carcinogenic potential of the compound. Input can be entered using the simplified molecular-input line-entry system (SMILES) or 2D structure. The results are colour-coded: green highlight for class I - inactive, yellow highlight for class II - weak activity, and red highlight for class III - active. The program was developed by Ideaconsult Ltd. (Sofia, Bulgaria), for researchers and other stakeholders (especially industry) (1, 61, 64).

LAZAR

Lazy Structure-Activity Relationships (LAZAR) is an open-source tool for the prediction of complex toxicological endpoints such as carcinogenicity (female/male, hamster/mouse/rat/rodent) and *Salmonella* mutagenicity. Unlike other software, LAZAR creates local endpoint QSAR models based on a training set (only nearest neighbours) for each compound separately. It first calculates the molecular descriptors and determines chemical similarity (alerts and compounds) and then it builds a local QSAR model based on a database of experimental toxicity data. Carcinogenicity models are based on CPDB, while the *Salmonella* mutagenicity model uses a dataset of 3895 compounds determined *in vitro*. This program meets all five OECD principles (65, 66).

ACD/Tox Suite

This industry-leading software package was developed to predict various toxicity endpoints such as genotoxicity, carcinogenicity, cytochrome P450 (CYP3A4) inhibition, oestrogen receptor (ER) binding affinity, irritation, rodent acute lethal toxicity (LD_{50}), aquatic toxicity, and organ-specific health effects. The predictions are made on the basis of validated QSAR models in combination with expert knowledge of organic chemistry and toxicology. It is primarily intended for ICH M7 submissions in pharmaceutical industry. The software can determine and visualise which parts of molecules, i.e., structural alerts (toxicophores) are responsible for toxic responses and can identify analogues from the training set. The training set is based on compounds that are genotoxic in the Ames test and are taken from the Chemical Carcinogenesis Research Information (CCRIS) and Genetic Toxicology Data Bank (GENE-TOX). Input can be entered as a 2D structure or a SMILES string, and the program yields predictions and up to five similar compounds from the training set (67).

Leadscope Model Applier

The Model Applier developed by Leadscope, Inc. (Columbus, OH, USA) uses QSAR models for the following endpoints: *Salmonella* mutagenicity, *E. coli* mutagenicity, mouse lymphoma, *in vitro* chromosome aberrations, and *in vivo* micronuclei. It was developed according to the ICH M7 guideline for impurities and is basically intended for the pharmaceutical industry (68).

CONCLUSION

In silico methods have become an acceptable alternative to animal testing to fill data gaps and improve the management of chemicals (e.g. the UN Globally Harmonized System of Classification and Labelling of Chemicals, GHS). They are run to submit data for regulatory purposes (e. g. REACH) and to obtain marketing authorisation for pharmaceuticals. In the past decade, several software applications have been designed to use data systems involving molecular descriptors specific to toxic endpoints and reference databases with the aim to predict the properties of new compounds.

All studies so far have shown that mutagenicity and carcinogenicity are parallel, although carcinogenicity is much more complex than mutagenicity. Understanding the mechanisms of covalent binding of chemicals with nucleophilic cellular targets enable us to group chemical compounds with similar mechanisms of action and similar toxic (mutagenic and carcinogenic) effects on the cell (11). Knowledge of the interconnections between the structure and mechanisms of action of potentially carcinogenic compounds helps us to understand these events.

Acknowledgments

We wish to thank the Agency of Research of the Republic of Slovenia for financial support under the contract P-01 017 and young researcher grant No. 1000-12-0104.

Conflicts of interest

The authors declare no conflicts of interests.

REFERENCES

1. Benigni R, Bossa C. Structure alerts for carcinogenicity, and the *Salmonella* assay system: A novel insight through the chemical relational database technology. Mutat Res 2008;659:248-61. doi: 10.1016/j.mrrev.2008.05.003
2. Benigni R, Bossa C. Structural Alerts of mutagens and cacinogens. Curr Comput - Aid Drug Design 2006;2:169-76. doi: 10.2174/157340906777441663
3. Farmer PB. DNA and protein adducts as markers of genotoxicity. Toxicol Lett 2004;149:3-9. doi: 10.1016/j.toxlet.2003.12.015
4. Powell CL, Swenberg JA, Rusyn I. Expression of base excision DNA repair genes as a biomarker of oxidative DNA damage. Cancer Lett 2005;229:1-11. doi: 10.1016/j.canlet.2004.12.002
5. Klaunig J, Wang Z, Pu X, Zhou S. Oxidative stress and oxidative damage in chemical carcinogensesis. Toxicol Appl Phramacol 2011;254:86-99. doi: 10.1016/j.taap.2009.11.028
6. Pulliero A, Godschalk R, Andreassi MG, Curfs D, Van Schooten FJ, Izzotti A. Environmental carcinogens and mutational pathways in atherosclerosis. Int J Hyg Environ Health 2015;218:293-312. doi: 10.1016/j.ijheh.2015.01.007
7. Tanaka T, Shimizu M, Kochi T, Moriwaki H. Chemical-induced carcinogenesis. J Exp Clin Med 2013;5:203-9. doi: 10.1016/j.jecm.2013.10.009
8. Helguera AM, Cordeiro N, Perez MA. Combes R, Perez Gonzales M. QSAR modeling of the rodent carcinogenicity of nitrocompounds. Bioorg Med Chem 2008;16:3395-407. doi: 10.1016/j.bmc.2007.11.029
9. Gonzales-Perez M, Gomez-Bombarelli R, Arenas-Valganon J, Perez-Prior T, Garcia-Santos P, Calle E, Casado J. Connecting the chemical and biological reactivity of epoxides. Chem Res Toxicol 2012;25:2755-62. doi: 10.1021/tx300389z
10. Zhang S, Chen K, Aliaga C, Sun Y, Lin J, Sharma A, Amin S, El-Bayoumy K. Identification and quantification of DNA adducts in the oral tissues of mice treated with the environmental carcinogen dibenzo[a,l]pyrene by HPLC-MS/MS. Chem Res Toxicol 2011;24:1297-303. doi: 10.1021/tx200188j
11. Miller E, Miller C. Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. Cancer 1981;47:2327-45. PMID: 7272889
12. La DK, Swenberg JA. DNA adducts: biological markers of exposure and potential applications to risk assesment. Mutat Res 1996;365:129-46. doi: 10.1016/S0165-1110(96)90017-2
13. Schwaab S, Czich A, Epe B, Kaina B, Müller L, Pollet D, Utesch D. Photochemical genotoxicity: principles and test methods: Report of a GUM task force. Mutat Res 2004;566:65-91. doi: 10.1016/S1383-5742(03)00052-8
14. Epe B. DNA damage spectra induced by photosensitization. Photochem Photobiol Sci 2012;11:98-106. doi: 10.1039/c1pp05190c
15. Spielmann H, Lovell W, Hoelzle E. *In vitro* phototoxicity testing. The report and recommendations of ECVAM workshop 2. ATLA 1994;22:314-48.
16. Dennehy M, Richards K, Wernke G, Shyr Y, Liebler D. Cytosolic and nuclear protein targets of thiol-reactive electrophiles. Chem Res Toxicol 2006;19:20-9. doi: 10.1021/tx050312l
17. Schultz W, Yarbrough J, Hunter R, Aptula A. Verification of the structural alerts for Michael acceptors. Chem Res Toxicol 2007;20:1359-63. doi: 10.1021/tx700212u
18. Delaney J, Essgmann J. Biological properties of single chemical-DNA adducts: a twenty year perspective. Chem Res Toxicol 2008;21:232-52. doi: 10.1021/tx700292a
19. Drinkwater NR, Miller JA, Miller EC, Yang N-C. Covalent intercalate binding to DNA in relation to the mutagenicity of hydrocarbon epoxides and N-acetoxy-2-acetylaminofluorene. Cancer Res 1978;38:3247-55. PMID: 356963
20. Ketterer B, Coles B, Meyer DJ. The role of glutathione in detoxication. Environ Healt Persp 1983;49:59-69. PMCID: PMC1569131
21. Yan Z, Maher N, Torres R, Huebert N. Use of a trapping agent for simultaneous capturing and high-throughput screening of both "soft" and "hard" reactive metabolites. Anal Chem 2007;79:4206-14. doi: 10.1021/ac0701029
22. Zang H, Gates KS. Sequence specificity of DNA alkylation by the antitumor natural product leinamycin. Chem Res Toxicol 2003;16:1539-46. doi: 10.1021/tx0341658
23. Allgayer H, Kolb M, Stuber V, Kruis W. Modulation of base hydroxylation by bile acids and salicylate s in a model of human colonic mucosal DNA putative implications in colonic cancer. Dig Dis Sci 1999;44:761-7. doi: 10.1023/A:1026670027150
24. Benigni R, Bossa C, Jeliazkova N, Netzeva T, Worth A. The Benigni/Bossa rulebase for mutagenicity and carcinogenicity-a module of Toxtree, 2008 [displayed 28 July 2016]. Available at https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23241_EN.pdf
25. Snodin D. Genotoxic impurities: from structural alerts to qualification. Org Proc Res Develop 2010;14:960-76. doi: 10.1021/op100118e
26. Ellison CM, Sherhod R, Cronin MTD, Enoch SJ. Madden JC, Judson PN. Assessment of methods to define the applicability domain of structural alert models. J Chem Inf Model 2011;51:975-85. doi: 10.1021/ci1000967
27. Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 2005;48:312-20. doi: 10.1021/jm040835a
28. Benigni R, Bossa C, Tcheremenskaia O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. Chem Rev 2013;113:2940-57. doi: 10.1021/cr300206t
29. Goetz ME, Luch A. Reactive species: A cell damaging rout assisting to chemical carcinogens. Cancer Lett 2008;266:73-83. doi: 10.1016/j.canlet.2008.02.035
30. Enoch SJ, Cronin M. Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA

binding. Mutat Res 2012;743:10-9. doi: 10.1016/j.mrgentox.2011.12.029
31. Schwöbel JA, Koleva Y, Enoch S, Bajot F, Hewitt M, Madden JC, Roberts DW, Schultz TW, Cronin MT. Measurement and estimation of electrophilic reactivity for predictive toxicology. Chem Rev 2011;111:2562-96. doi: 10.1021/cr100098n
32. Nash HM, Rongzhen L, Lane WS, Verdinel GL. The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution. Chem Biol 1997;4:693-702. doi: 10.1016/S1074-5521(97)90225-8
33. Enoch SJ, Cronin MTD, Ellison CM. The use of a chemistry-based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity. ATLA 2011;39:131-45. PMID: 21639678
34. Enoch SJ, Cronin MTD. A review of the electrophilic reaction chemistry involved in covalent DNA binding. Crit Rev Toxicol 2010;40:728-48. doi: 10.3109/10408444.2010.494175
35. Kolšek K, Sollner Dolenc M, Mavri J. Computational study of the reactivity of bisphenol A-3,4-quinone with deoxyadenosine and glutathione. Chem Res Toxicol 2013;26:106-11. doi: 10.1021/tx300411d
36. Garner RC. The role of DNA adducts in chemical carcinogenesis. Mutat Res 1988;402:67-75. doi: 10.1016/S0027-5107(97)00283-2
37. Henkler F, Stolpmann K, Luch A. Exposure to polycyclic aromatic hydrocarbons: Bulky DNA adducts and cellular responses. Mol Clin Environ Toxicol 2012;101:107-31. doi: 10.1007/978-3-7643-8340-4_5
38. Veglia F, Matullo G, Vineis P. Bulky DNA adducts and risk of cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2003;12:157-60. PMID: 12582026
39. Mukherjee A, Lavery R, Bagchi B, Hyne JT. On the molecular mechanism of drug intercalation into DNA: A simulation study of the intercalation pathway, free energy, and DNA structural changes. J Am Chem Soc 2008;130:9747-55. doi: 10.1021/ja8001666
40. Jałoszyński P, Jaruga P, Oliński R, Biczysko W, Szyfter W, Nagy E, Möller L, Szyfter K. Oxidative DNA base modifications and polycyclic aromatic hydrocarbon DNA adducts in squamous cell carcinoma of larynx. Free Radic Res 2003;37:231-40. doi: 10.1080/1071576021000041014
41. Lenne-Samuel N, Janel-Bintz R, Kolbanovskiy A, Geacintov NE, Fuchs RP. The processing of a Benzo(a)pyrene adduct into a frameshift or a base substitution mutation requires a different set of genes in *Escherichia coli*. Mol Microbiol 2000;38:299-307. doi: 10.1046/j.1365-2958.2000.02116.x
42. Heidrun Ellinger-Ziegelbauer H, Stuart B, Wahle B, Werner B, Juergen AH. Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. Mutat Res 2005;575:61-84. doi: 10.1016/j.mrfmmm.2005.02.004
43. Butterworth EB. A classification framework and practical guidance for establishing a mode of action for chemical carcinogens. Regul Toxicol Pharmacol 2006;45:9-23. doi: 10.1016/j.yrtph.2006.01.011
44. Marnett LJ, Burcham PC. Endogenous DNA adducts: potential and paradox. Chem Res Toxicol 1993;6:771-85. doi: 10.1021/tx00036a005
45. Wiseman H, Kaur H, Halliwell B. DNA damage and cancer: measurement and mechanism. Cancer Lett 1995;93:113-20. doi: 10.1016/0304-3835(95)03792-U
46. Kulis M, Esteller M. DNA methylation and cancer. Adv Genet 2010;70:27-56. doi: 10.1016/B978-0-12-380866-0.60002-2
47. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene 2010;21:5400-13. doi: 10.1038/sj.onc.1205651
48. Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. Mol Oncol 2007;1:19-25. doi: 10.1016/j.molonc.2007.01.001
49. Polo SE. Almouzni G. Histone metabolic pathways and chromatin assembly factors as proliferation markers. Cancer Lett 2005:220:1-9. doi: 10.1016/j.canlet.2004.08.024
50. Momparler RL. Cancer epigenetics. Oncogene 2003;22:6479-83. doi: 10.1038/sj.onc.1206774
51. Klaunig JE, Xu Y, Isenberg JS, Bachowski S, Kolaja KL, Jiang J, Stevenson DE, Walborg EF. The role of oxidative stress in chemical carcinogenesis. Environ Health Perspect 1998;106:289-95. doi: 10.2307/3433929
52. Crews D, McLachlan JA. Epigenetic, evolution, endocrine disruption, health, and disease. Endocrinology 2006;147(6 Suppl):S4-10. doi: 10.1210/en.2005-1122
53. Darbre PD. Environmental oestrogens, cosmetics and breast cancer. Best Pract Res Clin Endocrinol Metab 2006;20:121-43. doi: 10.1016/j.beem.2005.09.007
54. Birnbaum LS, Fenton SE. Cancer and developmental exposure to endocrine disruptors. Environ Health Perspect 2003;111:389-94. PMCID: PMC1241417
55. Choi SM, Yoo SD, Lee BM. Toxicological characterisitics of endocrine-disrupting chemicals: developmental toxicity, carcinogenicity, and mutagenicity. J Toxicol Environ Health 2004;7:1-23. doi: 10.1080/10937400490253229
56. Harvey PW, Johnson I. Approaches to the assessment of toxicology data with endpoints related to endocrine disruption. J Appl Toxicol 2002;22:241-7. doi: 10.1002/jat.854
57. Zhong M, Nie X, Yan A, Yuan Q. (2013) Carcinogenicity prediction of noncongeneric chemicals by a support vector machine. Chem Res Toxicol 2013;26:741-9. doi: 10.1021/tx4000182
58. Plošnik A, Zupan J, Vračko M. Evaluation of toxic endpoints for a set of cosmetic ingredients with CAESAR models. Chemosphere 2015;120:492-9. doi: 10.1016/j.chemosphere.2014.09.013
59. Ferrari T, Gini G. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem Cent J 2010;4:1-6. doi: 10.1186/1752-153X-4-S1-S2
60. Cariello N F, Wilson JD, Britt BH, Wedd DJ, Burlinson B, Gombar V. Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Mutagenesis 2002;17:321-9. doi: 10.1093/mutage/17.4.321
61. Serafimova R, Gatnik FM, Worth A. Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity, 2010 [displayed 29 July 2016]. Available at https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_24427_EN.pdf
62. Klopman G. The MultiCASE program II. Baseline activity identification algorithm (BAIA). J Chem Inf Comput Sci 1998;38:78-81. doi: 10.1021/ci9700790

63. QSAR Toolbox User Manual [displayed 17 April 2015]. Available at http://www.oecd.org/chemicalsafety/risk-assessment/TB3%200_GettingStarted_rev2.pdF
64. Ideaconsult. Toxtree User Manual [displayed 17 April 2015]. Available at https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/Toxtree_user_manual.pdf
65. Helma C. Lazy structure-activity relationships (LAZAR) for the prediction of rodent carcinogenicity and *Salmonella* mutagenicity. Mol Divers 2006;10:147-58. doi: 10.1007/s11030-005-9001-5
66. Maunz A, Gütlein M, Rautenberg M, Vorgrimmler D, Gebele D, Helma C. LAZAR: a modular predictive toxicology framework. Front Pharmacol 2013;4:1-8. doi: 10.3389/fphar.2013.00038
67. Worth A, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A, Serafimova R. A framework for assessing *in silico* toxicity predictions: case studies with selected pesticides. JRC report EUR 24705 EN, 2011.
68. Leadscope webpage [displayed 19 February 2016]. Available at http://www.leadscope.com

Mehanizmi delovanja tveganih kemijskih struktur za mutagenost in kancerogenost

Poznavanje mutagenih in rakotvornih lastnosti kemikalij je zelo pomembno za njihovo oceno tveganja in nevarnosti. Eden od ključnih dogodkov, ki sprožijo genotoksični in včasih kancerogeni učinkek je tvorba aduktov med kemikalijami in nukleinskimi kislinami ter histoni. Ta članek povzema pregled mehanizmov povezanih s specifičnimi funkcionalnimi skupinami (strukturnimi alerti ali tveganimi kemijskimi strukturami), ki lahko sprožijo genotoksične ali epigenetske učinke v celicah. Predstavlja aktualne informacije o poznanih strukturnih alertih, njihove mehanizme interakcij z genetskim materialom in programsko opremo, ki na osnovi poznavanja teh mehanizmov z uporabo QSAR modelov in klasifikacijskih shem omogočajo napovedovanje genostoksičnosti še nepoznanih kemikalij.

KLJUČNE BESEDE: *mutagenost, kancerogenost, računalniški programi za napoved učinka; strukturni alerti*