Scientific Committee on Health and Environmental Risks
SCHER

Scientific Committee on Consumer Products
SCCP

Scientific Committee on Emerging and Newly Identified Health Risks
SCENIHR

Risk assessment methodologies and approaches for genotoxic and carcinogenic substances

SCHER adopted this opinion at its 27th plenary on 13 January 2009
SCENIHR adopted this opinion at its 28th plenary on 19 January 2009
SCCP adopted this opinion at its 19th plenary on 21 January 2009

Three independent non-food Scientific Committees provide the Commission with the scientific advice it needs when preparing policy and proposals relating to consumer safety, public health and the environment. The Committees also draw the Commission's attention to the new or emerging problems which may pose an actual or potential threat.
They are: the Scientific Committee on Consumer Products (SCCP), the Scientific Committee on Health and Environmental Risks (SCHER) and the Scientific Committee on Emerging and Newly-Identified Health Risks (SCENIHR) and are made up of external experts.
In addition, the Commission relies upon the work of the European Food Safety Authority (EFSA), the European Medicines Evaluation Agency (EMEA), the European Centre for Disease prevention and Control (ECDC) and the European Chemicals Agency (ECHA).

SCHER

Questions relating to examinations of the toxicity and ecotoxicity of chemicals, biochemicals and biological compound whose use may have harmful consequences for human health and the environment.
In particular, the Committee addresses questions related to new and existing chemicals, the restriction and marketing of dangerous substances, biocides, waste, environmental contaminants, plastic and other materials used for water pipe work (e.g. new organics substances), drinking water, indoor and ambient air quality. It addresses questions relating to human exposure to mixtures of chemicals, sensitisation and identification of endocrine disrupters.

Scientific Committee members

Herman Autrup, Peter Calow, Wolfgang Dekant, Helmut Greim, Hanke Wojciech, Colin Janssen, Bo Jansson, Hannu Komulainen, Ole Ladefoged, Jan Linders, Inge Mangelsdorf, Marco Nuti, Anne Steenhout, Jose Tarazona, Emanuela Testai, Marco Vighi, Matti Viluksela

SCCP

Questions concerning the safety of consumer products (non-food products intended for the consumer).
In particular, the Committee addresses questions related to the safety and allergenic properties of cosmetic products and ingredients with respect to their impact on consumer health, toys, textiles, clothing, personal care products, domestic products such as detergents and consumer services such as tattooing.

Scientific Committee members

Claire Chambers, Gisela Degen, Ruta Dubakiene, Bozena Jazwiec-Kanyion, Vassilios Kapoulas, Jean Krutmann, Carola Lidén, Jean-Paul Marty, Thomas Platzek, Suresh Chandra Rastogi, Jean Revuz, Vera Rogiers, Tore Sanner, Günter Speit, Jacqueline Van Engelen, Ian White

SCENIHR

Questions concerning emerging or newly-identified risks and on broad, complex or multidisciplinary issues requiring a comprehensive assessment of risks to consumer safety or public health and related issues not covered by other Community risk-assessment bodies
In particular, the Committee addresses questions related to potential risks associated with interaction of risk factors, synergic effects, cumulative effects, antimicrobial resistance, new technologies such as nanotechnologies, medical devices, tissue engineering, blood products, fertility reduction, cancer of endocrine organs, physical hazards such as noise and electromagnetic fields and methodologies for assessing new risks.

Scientific Committee members

Anders Ahlbom, James Bridges, Wim De Jong, Philippe Hartemann, Thomas Jung, Mats-Olof Mattsson, Jean-Marie Pagès, Konrad Rydzynski, Dorothea Stahl, Mogens Thomsen

Contacts:

European Commission
Health & Consumer Protection DG
Directorate C: Public Health and Risk Assessment
Unit C7 - Risk Assessment
Office: B232 B-1049 Brussels

SCCP: Sanco-Sc6-Secretariat@ec.europa.eu

SCHER: Sanco-Sc8-Secretariat@ec.europa.eu

SCENIHR: Sanco-Sc1-Secretariat@ec.europa.eu

ACKNOWLEDGMENTS

Members of the working group are acknowledged for their valuable contribution to this opinion:

Prof. H. Autrup - University of Aarhus – DK

Prof. J. Bridges – University of Surrey -UK

Prof. G. Degen – IfADo Technical University of Dortmund - DE

Prof. W. Dekant - University of Wurzburg – DE

Prof. H. Greim - Technische Universitat Munich - DE

Prof. W. Hanke - Nofer Institute of Occupational Medicine - PL

Dr. I. H. Mangelsdorf - Institute of Toxicology and Experimental Medicine - DE

Prof. T. Platzek - Federal Institute for Risk Assessment - DE

Prof. K. Rydzynski - Nofer Institute of Occupational Medicine - PL

Prof. T. Sanner - University of Oslo - NO

Prof. A. Steenhout - Faculté des Sciences Université Libre de Brussels - BE

Dr. M. Viluksela – National Institute for Health and Welfare- FI

External experts

Dr. I. Anundi - Medical Products Agency Sweden; (EMEA)

Prof. A. Knaap – RIVM – NL; (EFSA)

Dr. U. Andrae- Helmholtz Zentrum München- DE

Keywords: SCHER, SCCP, SCENIHR, scientific opinion, genotoxicity, carcinogenicity, MOE, T25, TTC.

Opinion to be cited as:

SCHER/SCCP/SCENIHR scientific opinion on the risk assessment methodologies and approaches for genotoxic and carcinogenic substances, January 2009

TABLE OF CONTENTS

1. BACKGROUND

Several approaches are in use to assess the risk of substances that are carcinogenic. In broad terms, these methods make a distinction between those substances that are carcinogenic via a genotoxic mechanism and those that are carcinogenic via a non genotoxic mechanism. In addition, several approaches are used to assess the risks of genotoxic substances for which the evidence of carcinogenic potential is either limited or missing altogether.

Within the scope of harmonizing different approaches in the field of food safety, the European Food Safety Authority (EFSA) has adopted an opinion on "Harmonized approach for substances that are both genotoxic and carcinogenic" on October 2005[1].

The issue of risk assessment of genotoxic and carcinogenic substances is relevant for chemicals used or present in foods, non foods, industrial applications, and is one that poses significant challenges to the risk managers. It is therefore essential that the assessment of these substances is conducted using the best science available and that efforts are made globally to avoid divergences of scientific opinions and, if possible, arrive at a harmonised approach.

2. TERMS OF REFERENCE

Based on the publicly available information on the health risk assessment of substances that are genotoxic and/or carcinogenic, SCHER/SCCP/SCENIHR are asked:

1) To critically review the available methodologies and approaches used for the risk assessment of substances that have been shown to be carcinogenic in the appropriate toxicity studies.
2) If possible, to identify a harmonised methodology/approach for the risk assessment of substances that are carcinogenic taking into account all factors that may be of relevance (e.g. differences in their mechanism of action).
3) To critically review the available methodologies and approaches used for the health risk assessment of substances that are shown to be genotoxic in the appropriate toxicity studies and for which evidence of carcinogenicity is either limited or missing.
4) If possible, to identify a harmonised methodology/approach for the risk assessment of substances that are genotoxic and for which evidence of carcinogenicity is limited or missing, taking into account all factors that may be of relevance (e.g. differences in their mechanism of action).

PUBLIC CONSULTATION

A public consultation of the preliminary report of this opinion took place from 15 October to 26 November 2008. During the consultation 12 contributions were received,

In evaluating the responses from the consultation, submitted material has only been considered for revision of the opinion if

1. it is directly referring to the content of the report and relating to the issues that the report addresses,

2. it contains specific comments and suggestions on the scientific basis of the opinion,

4. it has the potential to add to the preliminary opinion of Scientific Committees

Each submission which meets these criteria has been carefully considered by the Working Group.

[1] http://www.efsa.europa.eu/etc/medialib/efsa/science/sc_commitee/sc_opinions/1201.Par.0002.File.dat/sc_op_ej282_gentox_en3.pdf

3. OPINION

3.1 General comments

Due to improved analytical techniques, extended testing and increased surveillance, genotoxic and carcinogenic chemicals may be detected in environmental media and consumer products at very low levels. Potential health risks of these exposures need to be assessed. Broadly defined, carcinogens are agents, which increase the incidence of neoplasms in experimental animals and/or in humans.

Genotoxic versus non-genotoxic carcinogens

The distinction between carcinogens likely causing tumours by interaction with the genetic material (genotoxic) and carcinogens causing tumours by other mechanism not involving genotoxicity (non-genotoxic) is the major determinant for the selection of risk assessment methodologies. A genotoxic chemical or physical agent has the ability to induce mutations or so-called indicator effects which are mechanistically associated with the formation of mutations (e.g. induction of DNA modifications, DNA repair, or recombination). Mutations are alterations of the genetic material within living cells, which can be transmitted from one cell generation to another (somatic mutations) or to the progeny of affected individuals through germ cells (germinal mutations). Mutations include gene mutations, structural chromosome mutations, and genome mutations.

Genotoxic agents are considered not to have a threshold but induced increases in DNA damage linearly related to the administered dose. It is also theoretically assumed that even a single molecule of a genotoxic carcinogen may cause a mutation and thus result in an increased cancer incidence, although the increase in risk may be infinitesimally small.

Compounds which induce aneuploidy and which are clearly devoid of any DNA reactivity or clastogenic activity constitute a particular subgroup of genotoxins. Aneuploidy induction does not result from an interaction of the compound or its metabolites with the DNA, but with proteins, which are present in multiple copies and which, thus, represent highly redundant targets. As a consequence, a certain minimal level of protein damage is required before the process of chromosome separation and distribution to the daughter cells is impaired. This level is assumed to constitute a threshold below which no disturbance of chromosome segregation occurs.

Moreover, a number of chemicals, which are devoid of genotoxic activity (eg peroxisome proliferators, hormones and local irritants) induced tumours in animal testing. Such non-genotoxic chemicals lack genotoxicity as a primary biological activity. While these agents may yield genotoxic events as a secondary result (secondary genotoxicity) of induced toxicity, such as forced cellular growth or formation of reactive oxygen species due to inflammation, their primary action does not involve reactivity with the genetic material. There is general agreement that knowledge of the underlying non-genotoxic mechanism of such compounds justifies identification of an NOEL.

In the present opinion only compounds that are both genotoxic and carcinogenic are discussed.

Classification of carcinogens

In the current classification principles (e.g. EU, GHS, IARC), mode of action and potency of a carcinogen are either not taken into account, or at best used as supporting arguments. However, the advancing knowledge of biochemical mechanisms of carcinogenesis and the different potencies of carcinogens have initiated a re-evaluation of the traditional concepts.

The International Agency for Research of Cancer (IARC 2006) and the OECD propose to use data on the carcinogenic mechanism and potency. Similarly, the General Directorate Employment of the EU is discussing application of information on carcinogenic mechanisms and potency as criteria for a revised classification. Within the EU context, the T25 dose-descriptor is in use for inclusion of potency considerations in setting specific concentration limits for carcinogens in Annex I of Directive 67/548/EEC (EC, 1999).

The Globally Harmonized System (GHS) has simplified and harmonized the classification criteria and categories of IARC, US Environmental Protection Agency (EPA), and the European Commission for carcinogens but does not include criteria for consideration of exposure and thus of carcinogenic risk.

The EPA Cancer Guidelines of 2005 (EPA 2005a) recommend that an agent's human carcinogenic potential is described in a *weight-of-evidence-narrativ*e, to characterise the full range of available evidence and the conditions associated with an agent's hazard potential. The use of mode of action data is explicitly mandated, but linear extrapolation from cancer incidences in animal studies at high doses to low doses, relevant for human exposures, using a benchmark dose as point of departure, is still required as default. However, the extrapolation process should utilize both toxicokinetic and toxicodynamic modelling to conclude on risks if substantial data support is available.

The narrative may explain that an agent appears to be carcinogenic by some routes of exposure but not others (inhalation versus ingestion) and may be attributed to specific exposures.

The MAK committee of the German Research Foundation considers mode of action to differentiate between genotoxic and non-genotoxic carcinogens and exposure for classification (Neumann et al 1998, DFG 2008). The American Conference of Governmental Industrial Hygienists has been using this concept since 1995.

Within the European Commission this concept is applied by the Scientific Committee of Occupational Exposure Limits (Bolt and Huici-Montagud 2008), the former SCF and EFSA (2005), which all differentiate between genotoxic and non-genotoxic carcinogens and consider the latter to be thresholded.

Risk assessment of Carcinogens

Risk assessment of carcinogens includes hazard identification, hazard characterisation (dose-response), and exposure assessment. These pillars are then merged into the assessment of potential risks for tumour induction in exposed populations.

Methods to identify the carcinogenic potential and potency and the approaches to assess the carcinogenic risk at a given exposure are indicated subsequently.

3.1.1 Hazard identification

Hazard identification of carcinogens usually requires a 2-year animal experiment following OECD-guidelines; however, a number of approaches using animal studies of shorter duration including endpoints related to cancer formation have been developed. These include tests in transgenic animals, induction of preneoplastic foci, genomic analyses after exposure duration of less than 2 years and may be increasingly utilized. Hazard identification can also be based on human observations on tumour incidences in exposed populations, but such data are only available for a few agents with a history of high exposures and often a very specific tumour response.

The 2-year carcinogenicity studies are typically carried out in rats and/or mice. This study type is intended for detection of potential of chemicals to induce neoplastic lesions, but also for identification of target organs, determination of dose-response relationships (establishment of NOAEL for non neoplastic effects or other points of departure), and providing data on mode of action. The study requires a control group and at least 3 dose levels of the test chemical, and at least 50 animals of each sex per dose-level. Evaluation of carcinogenic potential is primarily based on histopathological examination of all tissues. In order to maximise the ability of the study to detect carcinogens the highest dose level is selected to elicit some evidence of toxicity, such as a decrease in body weight development at most by 10% (MTD). However, a high dose level exposure to a test chemical may result in cytotoxicity followed by regenerative cell replication, which in a long-term study may eventually lead to increased incidence of neoplasms without specific carcinogenic potential of the test chemical. These types of tumours may not be relevant for human exposure levels, and therefore the evaluation of the data requires a

careful analysis of dose-response relationships of both neoplastic and non-neoplastic lesions in the target organ.

Animal experiments of shorter duration often use protocols that screen for preneoplastic effects, sometimes in a single tissue, such as altered foci in liver. They may also involve tumour induction in various organs after initiation-promotion treatment. In these types of studies the selected organ rather than the whole animal is the actual test system.

Experiments using transgenic rodents typically utilise activated oncogenes or inactivated tumour suppressor genes introduced into the test species using genetic engineering. These test systems may therefore provide specific mechanistic information on the chemical-gene interaction. Tumour development is usually faster than in standard carcinogenicity studies, and the test system makes it possible to establish a quantitative relationship between exposure and tumourigenic response. The limitations of these test systems include lack of complete development of tumours, manipulation of the carcinogenic process and limited histopathology. Therefore, at the present stage of development their role is supplementary.

Besides the observation of an increased tumour incidence in animals or humans exposed to the chemicals, the major aspect of hazard identification is to determine if an agent as parent compound or after bioactivation interacts with the genetic material of the cells; in other words, if is it genotoxic.

Hazard identification for genotoxicity mainly relies on *in vitro* studies determining mutagenicity of agents in bacteria and in mammalian cells. Indications for genotoxicity of an agent can also be derived by assessment of so-called indicator effects, e.g. DNA-damage, such as formation of strand breaks or adducts, or induction of DNA repair. *In vivo* studies further evaluate the genotoxic potential and usually are applied to confirm the observations made *in vitro* (see 3.4).

For non-genotoxic carcinogens, threshold mechanisms are accepted, when information on the underlying mechanisms indicates that non-toxic low concentrations of these substances are not tumourigenic. For receptor-mediated processes, a disproportional relationship between receptor occupancy and hormonally mediated cancer is also considered likely. For non-genotoxic compounds specific experimental approaches to delineate mode of action and its relevance are required. Specific approaches to evaluate available information on non-genotoxic modes of action and confidence in the conclusions have been published and should be rigorously applied (Sonich-Mullin et al 2001).

Since most of the genotoxic compounds require metabolic activation to produce electrophilic species that can interact with the genetic material, the differences in capacity of biotransformation enzymes and dose-dependent metabolic processes have a major influence on the target concentrations of reactive intermediates formed from carcinogens. To estimate the dose-dependent relationship between dose administered and effective dose, physiologically based toxicokinetic (PBTK) models can be applied. Such models integrate animal and human specific parameters such as blood flow to organs, enzymatic activity for activation and detoxication, and rates of elimination. They may be further enhanced by integration of DNA-repair and other toxicodynamic responses as well as by the dose-dependent balance between metabolic activation and detoxication. A guidance document is currently being prepared by IPCS (WHO 2008).

Susceptible populations

Risk assessment requires consideration of susceptible populations, because variations in toxicokinetics of chemical agents have been observed. Parameters affecting susceptibility include genetic differences, windows of susceptibility and predisposing diseases. This variation is partly due to polymorphisms of genes involved in the metabolism of chemical carcinogens, both in phase I and phase II, uptake of the toxicants, i.e. ABC transporter proteins, defences against oxidative stress induced by toxicants, and repair of the DNA damage. However, most of these polymorphisms have not been linked to a functional defect such as increased tumour rates (Garte et al, 2001).

Polymorphisms in phase I, phase II and DNA repair enzymes have been linked to increased risk of cancer as well as altered levels of biomarkers for genotoxic exposure, i.e. DNA adducts, and chromosomal damage (Autrup, 2004). However, as the metabolism of carcinogens to their ultimate carcinogenic form depends on several metabolic steps, the risk associated with a single polymorphism is limited, but persons carrying more than one risk variant allele may have a higher risk.

The interpretation of the role of the genotypes is complicated by the observation that the effect of the genotype depends on the dose. In case of vinyl chloride the expression of GSTT1 was associated with increased liver toxicity at low exposure, whereas it was protective at high exposure (Huang et al., 1997).

It has been estimated that the human variation in phase I and phase II metabolism in general is covered by the general kinetic default factor of 3.16 (Renwick et al., 2003) for inter-individual variations, the largest deviation observed was for the cytochrome CYP 2C19 and CYP2D6. Furthermore, some ethnic differences, as well as differences between elderly, children and neonates were estimated (Dorne et al., 2005).

The impact of predisposing disease on genotoxicity has not been systematically investigated. However, some diseases affect the expression of enzymes involved in the metabolism of genotoxic agents, e.g. diabetes and thus influence the risk associated with compounds metabolized by CYP 2E1 (Gonzales, 2007).

A special concern is exposure during development and early childhood. Human exposures during this period could contribute to cancer in children and young adults, and to cancer later in life. However, these two scenarios are qualitatively different. Childhood cancers, with a short latency period, most likely contain molecular lesions that may have been present at birth, whereas cancers that appear later in life, as a results of perinatal events, more likely reflects interaction with other causative factors later in life (Anderson, 2004).

Several international agencies have intensively discussed the need to apply specific uncertainty factors to protect sensitive subpopulations.

In its document "Chemical-specific adjustment factors for interspecies differences and human variability" (WHO, 2005) IPCS concludes that extra uncertainty factors to protect sensitive populations are not necessary. Instead, potentially susceptible population subgroups should be addressed separately. WHO further concluded (WHO, 2006) that representative data on toxicokinetics and toxicodynamics of the exposed population is required including an assessment of neonates, if appropriate. Furthermore polymorphism of xenobiotic biotransformation should be taken into account and in cases where the default factor will not cover the variability, the default should be modified appropriately.

In the Technical Guidance Document on Risk Assessment (TGD, 2005) it was concluded that higher intraspecies extrapolation factors from 10 to 100 should be considered when the following criteria are fulfilled:

- Specific exposure of very young children
- Indication or suspicion of effects on organ systems and functions that are especially vulnerable under development and maturation in early life
- Deficiencies in the databases on such effects in young animals.

US EPA (EPA 2005b) has evaluated several hundred references on 50 chemicals that have been shown to induce cancer following prenatal exposure. Both the acute and repeated dose studies support the concept that early-life exposure to chemicals with a genotoxic mode of action would lead to an increased tumour incidence compared to adult exposure. A practical approach was recommended that for exposures before 2 years of age a 10-fold adjustment and for exposures between 2 and 16 years of age a 3-fold adjustment should be used for risk assessment.

3.1.2 Dose-response characterisation

Dose-response characterization integrates all available toxicological, physical and chemical information on a compound for judging the likely potency of the carcinogen in humans and the shape of the dose-response curve for the carcinogenic effect in a weight of evidence approach. Dose-response characterization is also required for investigation of mode of action. A guidance document is available from WHO (2004).

The distinction between carcinogens likely causing tumours by interaction with the genetic material (genotoxic) and carcinogens causing tumours by other mechanism not involving genotoxicity (non-genotoxic) is the major determinant for the dose-response characterisation. Genotoxic chemicals exhibiting such activity can usually be identified by assays that measure induction of mutations or other genotoxic effects (indicator effects).

For genotoxic carcinogens, dose-response characterisation usually relies on data from animal cancer bioassays. Extrapolation of such data to exposure conditions likely encountered by humans requires integration of toxicokinetic and toxicodynamic information, understanding of the mode of action of genotoxicity, influence of non-genotoxic processes such as hyperplasia on tumour incidence, and consideration of potential subpopulations with special sensitivity/susceptibility. Dose-response characterization for genotoxic agents will also need to integrate non-linear toxicokinetics and role of DNA-repair (see below).

3.1.3 Exposure assessment

The quantification of exposure, both in individuals and in populations, is a prerequisite for the quantification of risk. Reliable data on exposure are needed to recognize specific risk factors such as occupation, life-style or social status. The dimensions of exposure include intensity, frequency, route and duration of the exposure; in addition, the exposed population should be characterized. Typically, field measurements and estimations are required. The estimation of human exposure to a particular xenobiotic involves an initial estimation of the possible sources of the chemical and the characterisation of exposure.

A good inventory of sources provides important information on critical pathways of exposure, populations at particular risk and the levels of exposure. In many cases, the duration and level of exposure, especially after chronic contact, can only be estimated from ambient levels of the xenobiotic in the environment and estimations may thus be crude; due to the large numbers of potentially exposed people, only in special situations (e.g. occupational exposure, after disasters), exposure data including determination of the internal dose will be available.

Exposure assessment for the general population is often derived from the concentrations of the agent in environmental media and standardized assumptions on intake of such media. In addition, exposure from consumer products has to be taken into account. In the area of exposures from food, exposure assessment includes use patterns for the agent, food consumption data, actual measured concentrations of the agent in food, and information of migration from food contact materials.

Specific procedures to estimate "external" exposure include direct measurement of the chemical in environmental samples such as water, air and soil. Measurement of the chemical, its metabolites or products of the interaction of the chemical or its biotransformation products with cellular macromolecules (protein and/or DNA) in body fluids and tissues determines "internal exposure". Due to the development of highly sensitive analytical techniques, the use of biomarkers becomes increasingly important in exposure assessment since it provides more exact information on actual internal exposure (target dose) to an agent. (Angerer et al. 2007, Boogaard 2007, Calafat et al. 2006, Needham et al. 2007, Pirkle et al. 2005, Yang et al. 2006b). Biomarkers are frequently used in experimental studies in animals and humans to assess the individual and internal exposure as compared to external exposure and other indirect exposure assessments.

Biomarkers of genotoxic effects include: DNA-adducts including oxidative damage to the DNA, which may be detected by the comet assay, chromosomal aberrations, sister

chromatid exchanges, increased frequency of micronuclei. Recently Swenberg et al (2008) have evaluated the biological relevance of DNA modifications such as DNA breaks and their direct consequences. These are considered to be markers of exposure, versus mutations, which are considered as markers of effect. Whereas biomarkers of exposure extrapolate down to zero, biomarkers of effect can only be interpolated back to the spontaneous or background number of mutations. Thus at high exposures, the biology that results in mutagenesis is driven by DNA damage preferentially resulting from the chemical exposure. In contrast, at very low exposures, the biology that results in mutagenesis is driven by endogenous DNA damage.

There were also efforts to assess the prognostic value of biomarkers for genotoxicity for carcinogenic risk in humans. Controversial results have been obtained with respect to the predictive value of chromosomal aberrations for increased cancer risk (Hagmar et al 1998, Bonassi et al 2004, Rossner et al 2005, Norppa et al., 2006, Boffetta et al 2007). Other studies evaluated frequency of micronuclei in peripheral blood lymphocytes (Znaor et al 2003, Bonassi et al 2008), and the levels of "bulky" DNA adducts for diagnosis of lung cancer in different smoking groups (Veglia et al 2008).

Since there is exposure from different sources and co-exposure by similarly acting compounds a clear differentiation between cumulative, aggregated and combined exposure is needed. Accordingly, SCHER has proposed the following terms and definitions (SCHER 2006):

- Cumulative exposure is the total exposure to one stressor from several sources and/or via several pathways.
- Aggregate exposure is total exposure to one stressor from several sources and/or via several pathways over time.
- Combined exposure describes the exposure to several stressors giving similar effects.

Especially combined exposures should be given specific attention in the risk characterization step. Aspects of combined exposure should, however, clearly be separated from the aspect of cumulative exposure to a single chemical from different exposure routes and exposure scenarios.

Due to the time consuming and cost and labour intensive procedures required, data on the exposure to xenobiotics usually are limited. Insufficient quality of chemical-analytical procedures, difficulties in identifying concomitant exposures, interactions with other xenobiotics or activities, special risk groups such as the very old or very young and pregnant women, and patterns of exposure can result in uncertainties in the exposure assessment in human populations.

3.1.4 Risk assessment

The risk assessment comprises all information on the mode of action, including genotoxic versus non-genotoxic effects, dose response of the toxic effects, identification of the NOEL of the most sensitive endpoint, and scenarios of human exposure. The various procedures for risk assessment of genotoxic carcinogens are described subsequently.

3.2 Request 1

To critically review the available methodologies and approaches used for the risk assessment of substances that have been shown to be genotoxic and carcinogenic in the appropriate toxicity studies.

A summary of currently used methodologies for risk assessment of genotoxic and carcinogenic chemicals is presented in Table 1.

3.2.1 Risk characterization for genotoxic carcinogens

Estimation of potency from animal cancer bioassays has inherent difficulties in the translation to potency estimates for humans because of species differences in physiology and due to the application of usually very high doses in the animal testing to compensate

for the low number of animals which can be used per dose group. The issues of extrapolations from incidence data in tests with inbred animals using high doses to predict incidences in genetically diverse human population exposed to much lower doses with a potentially wide variation in exposures represent the major challenges.

At present, a number of different procedures are used in the risk characterisation for such compounds (Table 1).

3.2.2 Linear Extrapolation from High Doses in Repeated Dose Animal Studies to Low Dose Human Exposure

A number of mathematical models have been developed for extrapolation from responses at the high experimental doses generally used in animal carcinogenicity tests to those of the substantially lower exposure levels usually encountered in human situations, well outside the range of experimental observations. These models describe the relationship between the administered daily dose and resulting tumour incidence. The models are either based on tolerance distributions or mechanistic assumptions. Because of the small number of doses tested experimentally, i.e. usually only 2 or 3, almost all data sets fit equally well these various mathematical functions. Yet, in the low dose region, risk estimates by these models may differ by several orders of magnitude (Krewski et al., 1999). Moreover, for most substances it is not known, whether the model actually reflects the underlying biological process and the method implies that any low exposure is actually associated with a defined cancer incidence.

Three different methods have been used by regulatory authorities in Europe and USA. The "Linearised Multistage Model" has been used extensively by the US EPA (1986), but the "LED10 method" has later been proposed (EPA, 1996), while the "T25 method" has been used in Europe (Sanner et al., 2001) (Table 1). The two latter methods, also called the "linear" approach, are basically driven by the assumption of a linear dose response relationship between tumour formation and exposure and involve linear extrapolation from a dose descriptor. The results obtained with these three extrapolation methods are, in most cases, nearly indistinguishable (Sanner et al., 2001) and the differences are much smaller than generally found when different tumours or experiments are considered. It should be noted that, in cases where high quality epidemiology and animal carcinogenicity studies are available, a good agreement was found between hazard characterisation based on epidemiology and hazard characterisation based on animal studies using the T25 method (Sanner and Dybing, 2005a).

The linear approach is used when there is an absence of sufficient information on modes of action or when mode of action information indicates that the dose-response curve at low dose is or is expected to be linear. Although it is recognised that linear extrapolation may in some cases result in overestimation of risks at low exposures, Bolt et al (2004) pointed out that for a number of chemical carcinogens a linear non-threshold extrapolation appears appropriate and scientifically well founded. However, if the available data do indicate a deviation from linearity, these data should be taken into account resulting in a modification of the default approach.

Different dose descriptors or point of departures may be used in the linear extrapolation. In the USA, it is recommended that a **BMDL** (LED) is selected that is representative of the lower end of the observed range. In most cases that would be BMDL10 (LED10), but in some cases even BMDL01 may be used (EPA, 2005a). For the regulation of existing chemicals the Technical Guidance Document (TGD 2007) states that the **T25** should be used as the default dose-descriptor in relation to linear extrapolation. The BMD05 or BMD10 should be used in certain cases in addition to the T25 when data are adequate for modelling purposes.

In order to use the advantages that the BMD05 or BMD10 offer, the TGD (2007) requests that a good set of quantitative data is needed, i.e. a control group and at least three dose levels. The BMD05 may offer an advantage in the case of sub- and supralinear dose relationship. In cases when the model shows a good fit, it is recommended that the point estimate of the BMD05 is calculated. The justification for using a benchmark dose (BMD) rather than the BMDL, is that (1) the point estimate is

the best estimate of the response, (2) a sufficient amount of conservative assumptions are included in a linear extrapolation to the origin and (3) it is in line with the procedure used for the T25 methods which also uses the point estimate as the starting point. On the other hand the BMDL advantageously represents the uncertainty of the dose-response relationship and is considered more conservative.

The determination of lifetime cancer risks and associated conclusions for certain exposure scenarios is carried out in several distinct steps as illustrated for the T25 method in fig. 1. The same applies for other dose descriptors. After having decided what animal data set to use, the dose descriptor T25 is determined. The T25 is determined by linear extrapolation from the lowest dose giving a statistically significant increase in tumours. Subsequently, the human equivalent of the animal dose-descriptor, i.e. the HT25[2] is calculated by dividing with an assessment factor for differences in metabolic rate (allometric scaling). The human dose is determined on the basis of a relevant scenario or measurements and the lifetime cancer risk is subsequently calculated. If the exposure is less than lifetime a correction factor should be applied to the calculated chronic exposure dose[3]. Subsequently, a commentary statement (narrative) is generated whether an overall evaluation of the data available[4] indicates that the actual risk may be higher or lower than the risk calculated for a specific scenario.

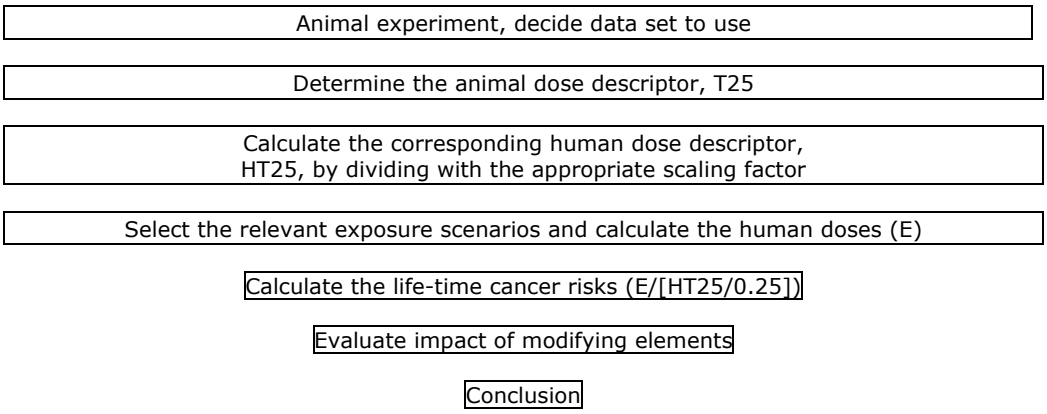

Figure 1- Outline of the procedure of the life-time cancer risk approach (After Sanner et al., 2001)

Of special importance is comparison with possible epidemiological information. Even in cases were no excess cancer risk were found it may be possible from the strength of the study to indicate at what risk the epidemiological study would be able to detect an increase and compare that with the calculated risk.

Finally, the calculated lifetime cancer risk for a certain exposure scenario will be compared to a lifetime cancer risk of little concern. Based on this comparison together with the commentary statement, conclusions with respect to the risks associated with the specific exposure situations are made.

The decision whether a calculated lifetime cancer risk is of concern or not is a political question. Some guidance may, however, be found by using the lifetime cancer risk in different countries, and if international organisations have considered to be of no or little concern. It has been found that the cancer risk decision points used for *lifetime* exposure

[2] HT25=T25/(weight human/weight animal)$^{0.25}$

[3] For workers with an exposure time of 8 hours per day, 5 days per week, 48 weeks per year for 40 years the default correction factor is 2.8 (7/5 x 52/48 x 75/40) for oral studies, whereas a factor of 1.5 (6/8 x 5/5 x 52/48 x 75/40) is used for inhalation studies (which normally involves 6 hours exposure per day, 5 days per week).

[4] Elements that may influence the value of the lifetime risk estimate are i.e. epidemiology, data-sets available, site-/species/strain/gender activity, dose-response relationships, chemical class, toxicokinetics (see Sanner et al., 2001)

of the general population are generally in the range of 10^{-5} to 10^{-6}. The decision point for 'acceptable' *lifetime* (i.e., a working life of 40 years) cancer risk levels used for workers are generally around 10^{-5} but higher or lower levels have been considered to be tolerable under certain circumstances (TGD, 2007). The Scientific Committees (SCs) note that the tolerable cancer risk values for occupational exposure are within one order of magnitude. These values should be harmonized internationally. The values for environmental exposure have a much larger variation. These values should also be harmonized globally.

In the REACH guidance document on dose-response characterisation (ECHA, 2008) linear extrapolation from T25 or the BMD to a risk of 10^{-5} is proposed to derive the so-called "derived maximum exposure level" (DMEL). For the risk characterisation this value is compared with exposure concentrations. If the DMEL is higher than the exposure level, the risk level is seen as tolerable.

3.2.3 Margin of Exposure (MOE)

EFSA recommends application of the concept of MOE for assessing the risk of genotoxic and carcinogenic substances (EFSA 2005, Barlow et al 2006, O'Brien et al 2006; see Table 1). The MOE represents the ratio between the dose descriptor for tumour formation in animals or humans and the measured or estimated human exposure. Application of the MOE-approach requires reliable animal carcinogenicity data or reliable epidemiological data including good quality exposure assessment. Comparative potency estimated is represented by the benchmark dose, or the T25 derived from animal carcinogenicity studies. Depending on the quality of the animal carcinogenicity data and the number of dose levels used in these studies, the dose-descriptors T25 or the BMD(L) are used as points of departure for risk estimation in the low dose region. EFSA (2005) concluded, "that a MOE of 10,000 and above, based on a BMDL10 from an animal study, would be a value that would indicate a low concern from a public health point of view and that might be considered a low priority for risk management actions." The proposal has been discussed and supported during a joint EFSA/WHO/ ILSI Europe conference (Barlow et al 2006). This MOE of 10,000 is based on the conclusion that for non-genotoxic substances the product of two 10-fold factors (one to allow for possible inter-species differences, the other to allow for human variability (WHO, 1987 and 1994) also apply for substances which are both genotoxic and carcinogenic and an additional factor of 100. This factor takes into account that the additional point is not equivalent to the NOEL, additional uncertainties related to human variability in cell cycle control and DNA repair, and uncertainties about the shape of the dose–response curve below the BMD and the dose level below which the cancer incidence is not increased.

Although SCHER/SCCP/SCENIHR (SCs) agree to this additional factor because of limited knowledge and the lack of experience with this approach, it notices that several investigations revealed that genetic variations in xenobiotic-metabolising enzymes have, in general, only a modest effect on the individual cancer risk associated with low level environmental exposures (Hirvonen, 1999; Taningher et al., 1999; D'Errico et al., 1999; Pavanello and Clonfero, 2000).

Moreover, although *in vitro* treatment of blood cells from healthy subjects with genotoxic agents showed a variation in response in a range of an order of magnitude (Gu et al., 1999), the contributions of individual variant alleles of DNA repair genes is less than two-fold although the impact of low penetrance polymorphisms may theoretically be barely detectable (Mohrenweiser et al., 2003).

However, the magnitude of the MOE requires adjustment to the type of point of departure since these do represent doses inducing different tumour incidences thus influencing conclusions on risk. Size of the required MOE may be influenced by more detailed information on mode of action such as genotoxicity versus non-genotoxicity and integration of all available data in a weight of evidence approach. Different MOE-values may also be derived for subpopulations with different (high or low) exposures.

As already mentioned in 3.2.2 the REACH guidance (ECHA, 2008) requires the identification of a DMEL for carcinogenic substances. A DMEL may also be obtained by

applying the MOE-approach. The T25 or BMD are divided by the assessment factors described above (10000 for the BMD10) to obtain the DMEL.

The document does not provide guidance on which method - linear extrapolation or MOE approach - is to be preferred. However, there is a comparison on the outcome of the linearized approach and the MOE approach in Appendix 8-7 of the guidance, which shows that the resulting DMEL is principally the same. As the way, how the DMEL is obtained by applying assessment factors, the MOE-approach is called "large assessment factor approach" in the REACH guidance.

3.2.4 ALARA

While not a risk assessment method, but a risk management tool, the ALARA principle is often applied by advisory bodies and regulatory agencies and requires comments in the context of risk characterisation (Table 1). The application of the ALARA principle intends to keep the exposure to carcinogenic substances at the lowest achievable level, usually limited by technological limitations or economic considerations. The advantage of the ALARA principle is that only hazard identification data is needed and that exposure to genotoxic and carcinogenic substances is limited to technically unavoidable amounts. The hazard identification data may be restricted to an *in vivo* genotoxicity test considered indicative of potential carcinogenicity, but the ALARA principle may also use animal carcinogenicity data. There is no requirement to evaluate carcinogenicity data in a quantitative manner. In addition, no information on exposure is needed. By that, the approach is qualitative and can readily be communicated to the general public.

However, the ALARA principle has a number of major disadvantages. It usually does not make use of the general toxicological database available, is not useful for risk comparison and does not take into account the carcinogenic potency and the actual (sometimes extremely low) exposures, which may result in very low risks. Although the ALARA principle is a valuable tool to generally reduce exposure, it may create problems for risk communication because it provides only qualitative information and will not communicate that certain agents may be of higher risks than others, because of their potency or of high exposures.

3.2.5 Threshold of toxicological concern (TTC)

The TTC principle is dealing with compounds of unknown toxicity. Based on the chemical structure of a compound, human exposure threshold values for the daily uptake of a compound are established (Kroes et al., 2000, 2004).

For compounds with structural alerts for genotoxicity (Ashby and Tennant, 1991) the TTC is 0.15 µg/person/day. As it is assumed, that there is no threshold for genotoxicity and carcinogenicity a risk based approach has been used for deriving this TTC. The criterion for "no appreciable risk" for carcinogenicity was a risk of 1 in 1 million, based on the TD50 values from rodent carcinogenicity studies. To establish a TTC-value for these compounds a database of rodent carcinogenicity studies with over 700 chemicals (Cheeseman et al., 1999, Kroes et al., 2004) has been analysed. Most of the carcinogens assessed have a risk level which is less then 1 in 1 million at the TTC. However some aflatoxin-like-, azoxy-, or N-nitroso-compounds give risks higher than 1 in 1 million. Therefore these structural classes were excluded from the use of the TTC principle, because compound specific risk assessments are required. Metals, metal-containing compounds, polyhalogenated dibenzodioxins, dibenzofurans or biphenyls and proteins are also excluded from the use of the TTC.

The TTC principle has been developed originally for chemicals present in food (Munro et al., 2008). It is applied by organizations such as the US Food and Drug Administration in the regulation of food contact materials (USFDA, 1995) and by EFSA and the Joint FAO /WHO Expert Committee on Food Additives (JECFA) in evaluations of flavouring substances. Further its use has been proposed for cosmetic ingredients (Kroes et al., 2007), herbal preparations (EMEA, 2007), personal and household care products (Blackburn et al., 2005), and impurities in pharmaceuticals (Müller et al., 2005). Concerning the latter, its use has been criticized as too conservative (Delaney, 2007), as

the use of a single figure infers an unwarranted level of precision (Humfrey, 2007). It has been proposed that the TTC for genotoxic compounds should be adapted for impurities in pharmaceuticals, which are designed for shorter duration of intake as compared to the life-time intake of food (Humfrey, 2007).

The TTC principle is currently being assessed by the Sanco SCs (SCCP/SCHER/SCENIHR, 2008)

3.2.6 (Q)SAR for the assessment of genotoxic agents and carcinogens

As the carcinogenicity bioassay in rodents is long and costly and requires the use of many animals, chemical carcinogenicity has been the target of numerous attempts to create alternative predictive in silico models. A number of commercial and non commercial computer programs or tools are available on genotoxicity or carcinogenicity of chemical substances e.g. COMPAQ, DEREK, HAZARDEXPERT, MCASE, MDL-SAR, Oncologic®, TOPKAT, the OECD toolbox and structural alerts.

Several external validation exercises of these (Q)SAR programs are available in the literature (Benigni and Zito, 2004, Benigni et al., 2007; Bristol et al., 1996, Contrera et al., 2007; Matthews et al., 2006; Mayer et al., 2008). They differ in numbers of chemicals assessed and in how the chemicals were selected for the prediction.

In the most recent evaluation of Mayer et al., (2008), the predictive ability of genotoxicity tests (bacterial reverse mutation, mouse lymphoma, chromosome aberration) and SAR (structural alerts, MCASE, OncoLogic®) have been compared with the results of carcinogenicity studies on 650 compounds from the TD50 carcinogenicity Potency DataBase of Gold (http://potency.berkeley.edu/cpdb.html). While the concordance for the structural alerts (70.9 %) was similar to the genotoxicity tests (62.5 -75 %), MCASE and OncoLogic® were more effective in predicting carcinogenicity (82.9-88.4 %). One reason for this difference may be that the SAR programs also detect non genotoxic carcinogens. MCASE and OncoLogic® had a similar degree of concordance (88.4 % versus 82.9 %). The better performance of OncoLogic® and MCASE compared to genotoxicity assays may at least partly be due to the inclusion of non genotoxic carcinogens, which are detected by these programs. In another analysis encompassing 1540 compounds Contrera et al. (2007) obtained 69 % concordance with results of carcinogenicity studies for MCASE and 73 % for MDL-QSAR. The prediction has been refined to 76 %, if both programs were merged. For analyses of chemicals, where the outcome of carcinogenicity studies was not yet available at the time of the analysis (Benigni and Zito, 2004, Bristol et al., 1996), the concordance of the predictions by Oncologic® as the best performing programme was 65 % (Benigni and Zito, 2004).

One major shortcoming of SAR approaches is the limited data availability in the training set of the program. If the compound analysed is not similar to any of the compounds in the training set, then the SAR programs is not able to make a significant prediction. In the analysis of Mayer et al., (2008), many compounds could not be analysed by OncoLogic®, the best performing program, because they did not fall into one of OncoLogic®´s defined structure classes. Further there are several types of compounds that cannot be predicted at all in many SAR programs, e.g. inorganics, organometallic compounds, mixtures, and silicone-based compounds.

Besides assessing carcinogenicity as a yes/no result, (Q)SAR may be useful for analysing mode of action and for guiding mechanistic studies. As an example the mode of action of aromatic amines has been explored (Benigni and Passerini, 2002). Carcinogenicity of these compounds depended on hydrophobicity as well as electronic and steric properties. Thus, the information provided by (Q)SAR can expand the knowledge on a chemical and may help to put into context available evidence.

In conclusion, given the structural diversity and possible complexity of chemicals to be analysed and the huge range and variability of possible interactions of chemicals in biological systems, it is highly unlikely that (Q)SAR models will ever achieve absolute certainty in predicting carcinogenicity. However, (Q)SAR software programs provide the possibility of cost effective screening of chemicals for possible carcinogenic effects.

Several programs may be used together in a weight of evidence approach. As they may also detect non genotoxic carcinogens they are also valuable in combination with genetic toxicity assays. Therefore SCHER recommends that SAR should be incorporated into a battery of approaches for evaluating the genotoxic and carcinogenic potential of a chemical thus adding to the overall evidence.

3.2.7 Use of epidemiological studies in cancer quantitative risk assessment

It is generally recognised that dose-response information from epidemiological studies is preferred as the starting point for quantitative risk analysis of carcinogens instead of data from experimental animal studies. The main advantage is that there is no need for usually conservative extrapolation step for species-species extrapolation and that exposure conditions are representative of those of the target population. However, the results of human studies are often hampered by insufficient exposure assessment and by other methodological shortcomings or too large random variations between study results (Brandt, 2002).

The SCs stress that exposure assessment is most critical for the establishment of exposure response relationships. Three 'exposure fields' can be defined: consumer exposure, environmental exposure and occupational exposure. The measures of exposure(s) or exposure surrogates should be: (a) conceptually relevant to the risk assessment being conducted; (b) based on principles that are biologically plausible and (c) properly quantified.

For identification of appropriate epidemiologic studies the SCS recommend to use the quality criteria developed by Swaen (2006), which include consideration of: 1) study design, 2) appropriate control group, 3) sufficient long observation period with respect to latency 4) if applicable, correction for known confounding factors, including concomitant exposures 5) sufficient statistical power, and 6) appropriate statistical analysis.

Whenever feasible, human data on biomarkers and other biological measures as well as data on individual susceptibility should be employed.

US EPA has developed an electronic database (IRIS) containing information on human health effects that may result from exposure to various substances in the environment. By the end of 2007, 24 agents have been classified according to the EPA guidelines for carcinogen risk assessment as human or probably (likely) human carcinogens. Of these agents risk estimation based on epidemiological data were available for 15 substances. From the epidemiological studies the difference in risk of a particular condition between those who are exposed and those who are not were calculated. The additional risk is an absolute measure of the excess risk attributed to the exposure. (http://cfpub.epa.gov/ncea/iris/index.cfm?fuseaction=iris.showSubstanceList).

As in quantitative risk assessment based on animal studies, the evaluation of uncertainties is essential. Main sources of uncertainty in epidemiological studies, such as bias (e.g. healthy worker effect), exposure measurement error or possible confounding can be estimated in validation study e.g. Steenland and Greenland (2004).

Until recently, transparent and generally accepted protocol for calculating cancer risk from epidemiological studies was not available. A systematic approach has been proposed (Goldbohm et al., 2006) based on the case of hexavalent chromium and lung cancer. The SCs strongly recommend this approach, which is structured, transparent and allows conducting risk assessment process in reproducible manner. The approach includes: (1) selection and evaluation of epidemiological data, (2) derivations of relative risk as a function of exposure from the selected epidemiological data and (3) calculation of excess lifetime risk for an exposed target population at risk.

In conclusion, the SCs point out that, epidemiological studies of sufficient quality are best suitable for quantitative risk assessment if they include dose response information based on which starting point for dose extrapolation can be identified. Epidemiological studies without such detailed information can support a risk assessment based on appropriate animal studies.

3.2.8 The question of threshold for genotoxic carcinogens

In its opinion the Scientific Committee of EFSA (EFSA, 2005) has concluded that based on the current understanding of cancer biology there are levels of exposure to substances, which are both genotoxic and carcinogenic, below which cancer incidence is not increased (biological thresholds in dose-response).

As mentioned above, the default assumption for risk assessment of carcinogens is that exposure to even very small amounts of a chemical may result in additional risk. While this low-dose linearity is supported by observations on the dose-response for DNA adduct formation of many chemicals, non-linearity in the low-dose range, suggesting the presence of thresholds has been observed in a number of systems regarding endpoints considered relevant in the process of tumour formation, e.g. mutations. This assumption is based on the existence of cellular defence mechanism and that at a certain dose the agent may no longer increase the spontaneous mutation rates.

However, the relative contribution of the different cellular defence mechanisms and the dose dependence of their responses are insufficiently understood. Moreover, further research is needed to elucidate the rate limiting parameters that trigger the defence mechanisms, which will permit the determination of the onset of such counterbalancing reactions. A better understanding is of utmost regulatory importance, since a scientifically defendable threshold concept for genotoxic carcinogens will allow to identify NOEL and by that to propose health based exposure limits for genotoxic carcinogens.

In its opinion "Research priorities for the 7th Framework Program - Human Health and the Environment", adopted on 30 October 2006 SCHER has addressed this problem and has recommended the following activities:

Development of tools to assess interaction of chemicals at low dose, including natural stressors, comparison of dose-response of several interrelated biomarkers (systems biology) focusing on chemical carcinogenesis.

- Shape of dose-response relationships for low-level exposures to chemicals
- Explore the possibilities to establish threshold for genotoxic carcinogens.
- Integration of physiologically based toxicokinetic models into the risk assessment process.

The working group supports this proposal and strongly recommends further research within the 7th Framework Program.

3.3 Request 2

If possible, to identify a harmonised methodology/approach for the risk assessment of substances that are genotoxic and carcinogenic taking into account all factors that may be of relevance (e.g. differences in their mechanism of action).

No international scientific consensus on what is the best approach for risk assessment of substances that are both genotoxic and carcinogenic is currently available. However, if good epidemiological data are available these should primarily be used for risk assessment.

The Technical Guidance Document Chemical Safety Assessment (REACH Implementation Project 3.2-2) recommends two default methodologies for deriving an exposure level of little or no concern (RIP, 2007). The 'linearised' approach (T25 method) results in a lifetime cancer risk considered being of low concern while the MOE approach results in values representing a low concern. If data allows, more sophisticated methodologies for deriving a no-effect level may be applied. The choice of such alternative methodologies should be justified. EFSA (2005) has recommended application of the MOE approach.

Both the linear extrapolation and the MOE approach use the same dose-descriptors, a benchmark dose or a T25, as determined from experimental studies in animals. Both also require a precise and comprehensive exposure assessment. Both approaches allow comparing the risks of exposures by considering differences in potency of the substances concerned and exposure patterns in the population. Moreover, both approaches may be

applied to individual chemical agents as well as to chemical classes (e.g.PAHs) and aggregate exposures.

Both approaches can be adjusted by information on differences in metabolic rates between the experimental animal and man. Other differences in toxicokinetics and toxicodynamics may also be included. Moreover, confirmation by epidemiological information can be performed.

The MOE represents the ratio between a dose-descriptor BMDL or T25 and the actual human exposure. Thus, the MOE approach avoids extrapolations of dose-response curves way outside of the observable range, and as a consequence the presentation of a virtual cancer risk with intrinsic uncertainties. On the other hand, only limited experience with the method including interpretation and communication of MOE is available.

The SCs conclude that evaluation of the carcinogenic risk of a genotoxic carcinogen should be done on a case-case basis. Whenever sufficient information is available, emphasis should be laid to identify the appropriate dose descriptor as a starting point for risk assessment of genotoxic carcinogens to either apply the linear extrapolation or the MOE approach. Both have their advantages and disadvantages. From a scientific point of view the MOE approach avoids the usually intrinsic uncertainties of the extrapolation models and the presentation of a virtual cancer risk of a high uncertainty. The need for a high quality exposure assessment including specific subpopulation and of high quality bioassay data and inability to serve as a quantitative estimate of risk may be considered as disadvantages of the MOE-approach. Moreover, risk comparisons for management decisions require that all MOEs or quantitative risk characterisations are based on databases with similar and high quality.

The MOE and linear extrapolation approaches combine information on human exposure and the potency of the carcinogen and can be used to indicate levels of concern and also ranking between various exposures to such agents.

With this the SCs have identified several conditions where one or the other approach may be preferable:

- For risk communication the MOE is seen preferable.
- For prioritisation of measures to reduce risk, both, the MOE approach and the linear extrapolation from a dose descriptor e.g. the BMDL10 or the T25 are applicable
- Linear extrapolation is a method that provides a quantitative expression of risk that is commensurate with cost benefit analysis It provides likely units of life under certain exposure conditions.

The SCs also recommend application of the TTC principle when appropriate. In cases where the exposures are below the relevant TTC level, non-necessary toxicological testing or regulations can be avoided.

The concept to reduce the exposure of carcinogens to a level as low as reasonably achievable (ALARA principle) is a valuable measure to minimize exposure to genotoxic and carcinogenic substances. It is not applicable for risk assessment, and it does not provide a basis for setting priorities for action.

3.4 Request 3

To critically review the available methodologies and approaches used for the health risk assessment of substances that are shown to be genotoxic in the appropriate toxicity studies and for which evidence of carcinogenicity is either limited or missing.

Genotoxic effects, i.e. toxic effects on the genome of cells, comprise both mutations and damage to the genetic material which cannot be inherited but has the potential to cause the formation of mutations after further processing. The terms "genotoxic" and "mutagenic" or "genotoxicity test" and "mutagenicity test" must, therefore, not be considered synonymous. Whereas mutations can only be identified in the subgroup of

genotoxicity tests termed "mutagenicity tests", damage to the genetic material or cellular responses specific to the presence of DNA damage can be identified in a different subgroup termed "indicator tests".
Information on the biological background of the individual assays is given in the Annex.

3.4.1 Ability of in vitro and in vivo genotoxicity tests to correctly predict the carcinogenicity of chemicals

The capability of individual genotoxicity tests or combinations of tests to correctly predict the carcinogenic properties of chemicals has been assessed in a number of studies (see, e.g., Tennant et al. 1987, Haseman et al. 1990, Zeiger et al. 1990, Zeiger, 1998). The most recent and largest of these analyses are those of Kirkland et al. (2005) and Matthews et al. (2006a) the outcomes of which will be discussed here in some detail.

A straightforward interpretation of the results of the two studies is hampered by the fact that the rodent carcinogens in the data sets analysed included both clearly genotoxic carcinogens as well as others with a presumed non-genotoxic mode of action. Kirkland et al. (2005) argued that "…there are only a few rodent carcinogens for which the non-genotoxic mechanism of action is sufficiently defined to discern whether an absence of response (…) in the in vitro battery is an accurate reflection of mechanism or a true lack of predictivity". As a consequence, the non-genotoxic compounds were not separated from the genotoxic ones for the correlation analysis. Therefore, the sensitivity of a genotoxicity assay, i.e. its ability to give a positive response with a rodent carcinogen, cannot be regarded a good measure for the utility of an assay to identify genotoxic carcinogens, as negative results must be expected for any single test system with carcinogens that are non-genotoxic. More critical parameters for the ability of the tests to correctly predict the carcinogenicity of genotoxic carcinogens are specificity, i.e. the ability to give negative results with known rodent non-carcinogens, and positive predictivity, i.e. the probability of an assay to correctly identify a rodent carcinogen from a positive result.

Kirkland et al. (2005) assessed the ability of a battery of three commonly used *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. Assays investigated were the bacterial mutagenicity test, the mouse lymphoma assay and a test for clastogenicity, the *in vitro* micronucleus or the chromosomal aberration test.

The sensitivity of the three-test battery, i.e. its ability to give positive responses with rodent carcinogens, was high. Of the 553 genotoxic and non genotoxic carcinogens for which there were valid genotoxicity data, 93% of the compounds evaluated gave positive results in at least one of the three tests. Combinations of two and three test systems had greater sensitivity than individual tests resulting in sensitivities of around 90% or more, depending on test combination.

Only 19 carcinogens out of 206 tested in all three tests gave consistently negative results in a full three-test battery. Most were either recognised as known non-genotoxic carcinogens or were extremely weak (presumed) genotoxic carcinogens. 183 chemicals were non-carcinogenic after testing in both male and female rats and mice. There were genotoxicity data on 177 of these, which were used to determine the specificity of the individual assays or their combinations. The specificity of the bacterial mutagenicity test, for which the largest database was available, was reasonable (73.9%), but all mammalian cell tests alone had low specificity (i.e. 31% for the micronucleus test, 39% for the mouse lymphoma assay and 45% for the chromosomal aberration test).

This already low specificity declined to extremely low levels in combinations of two and three test systems. When all three tests were performed, 75-95% of the non-carcinogens gave positive (i.e. false positive) results in at least one test in the battery.

These findings were largely corroborated and extended by Matthews et al. (2006a) who performed an analysis of a large set of data from numerous genotoxicity tests, reproductive and developmental toxicity studies and rodent carcinogenicity bioassays in order to identify the genotoxicity and reprotoxicity endpoints which best correlate with the results from the carcinogenicity bioassays.

The data in the ICSAS carcinogenicity database at the time of analysis contained 24708 study records. For 1112 of the 7205 organic chemicals selected, both genotoxicity data and results from carcinogenicity assays were available. Of the various statistical parameters evaluated only a new statistical parameter, the "correlation indicator" (CI), was used. The CI was defined as the average of specificity and positive predictivity.

A CI value of 73%, which was 5% less than the CI value for the *Salmonella* composite data (78%), was arbitrarily chosen as the cut-off for the discrimination between "well correlated" and "not well correlated" endpoints. Of the 21 composite endpoints composed of genotoxicity test systems, 10 had CI values below 73%. These endpoints included two composites comprising differentially evaluated mouse lymphoma assays (CI = 62.0 and 53.4%), 7 composites categorised as "Clastogenicity", which contained the *in vitro* micronucleus assay (9.3%), chromosome aberrations *in vitro* (65.2%) and plant cytogenetics, but also SCE *in vitro* and *in vivo* and fungal and *Drosophila* aneuploidy.

The 11 composites that were classified as "well correlated" comprised 8 groups categorised as "Gene mutation" which included the induction of gene mutations in *Salmonella* (CI = 78.3%), *Escherichia coli* (75.1%), all bacteria (77.3%), fungi (78.2%), plants (88.6%), *Drosophila* (80.5%), the HPRT system (78.4%) and rodent mutation *in vivo* (consisting of results from dominant lethal, heritable translocation and specific locus tests; 90.5%), two composites for *in vivo* clastogenicity which contained chromosome aberrations *in vivo* (75.2%) and micronuclei *in vivo* (73.4%), and DNA damage, i.e. the *in vitro* UDS test (86.6%). Of the individual test systems in these composites using bacteria or mammalian cells, the highest CI values of >90% were associated with UDS in human fibroblasts (95.6%) and rat hepatocytes (92.0%) and the dominant lethal assay (92.5%). The high CI values of these three assay systems were due to both specificities and positive predictivities between 91 and 96%.

It appears likely that a certain proportion of "false positive" results from *in vivo* genotoxicity tests, i.e. positive findings that were not associated with a positive result in a carcinogenicity bioassay and which impair the positive predictivities of the assays, actually reflects the genotoxicity of very weak carcinogens that were not identified in the standard bioassay due to its inherently low sensitivity.

In a further study, a good performance with respect to the prediction of rodent carcinogens was reported for various cell transformation tests, particularly the SHE cell assay (OECD, 2007). However, these assays give no information on the genotoxicity of an agent, as cell transformation can be brought about by both genotoxic and non-genotoxic mechanisms.

In conclusion, an evaluation of the results of the analyses of Kirkland et al. (2005) and Matthews et al. (2006) with respect to those test systems commonly used for genotoxicity testing suggests that the two *in vitro* test systems specific for gene mutations, the bacterial gene mutation assay and the HPRT test, show good and similar predictivity. The analyses also confirm numerous earlier reports that the predictivity of positive results from *in vitro* assays responding to the clastogenic activity of chemicals in mammalian cells, i.e., the test for chromosome aberrations, the micronucleus assay and the mouse lymphoma assay, is very limited. In particular, the low specificity and CI of the *in vitro* micronucleus test would suggest that this assay is not well suited for an inclusion in a battery of basic *in vitro* tests for genotoxicity. On the other hand, it is the only available test capable of detecting aneugenic effects, and this appears to justify its inclusion in a battery of initial tests. The particularly good predictivity of the UDS test, the performance of which was only analysed *in vitro*, indicates that this indicator test should be an excellent assay for the follow-up testing of chemicals *in vivo*, as there is no reason to believe that the predictivity of the *in vivo* version of the assay is lower than that of the *in vitro* version. The predictivity of the standard *in vivo* assay, the micronucleus test, appears good, although clearly lower than that of germ cell tests involving an examination of the progeny, such as the dominant lethal test. These assays, however, are definitely too laborious and require too many animals for being employed for routine testing.

Current Strategies for Genotoxicity Testing

As an initial step in the evaluation of the genotoxicity of an agent, the available information on its toxicity, metabolism, toxicokinetics etc. should be scrutinised, as the exact strategy employed in the subsequent experimental phase has to consider all available relevant data. This information can be supplemented by an *in silico* assessment of QSAR. This kind of initial computational toxicology may also be employed in cases where prioritisation of agents to be tested is desired. The *in silico* approach is then followed by experiments in which the genotoxicity of the agent in question is assessed *in vitro* and if need be *in vivo*.

For the experimental evaluation of the genotoxic potential of chemical compounds, several national and international institutions and scientific societies have proposed guidelines for a successive approach (Cimino 2006; Lorge et al., 2007; Tweats et al., 2007a, b; WHO, 2008). There is extensive, but not complete agreement among these proposals. Recently, a general, simplified, highly flexible strategy for genotoxicity testing has been proposed (Pfuhler et al. 2007) which is in general accordance with the majority of these proposals and which takes into account the experience with the individual test systems (see Annex, "Test systems for genotoxicity"). In the following, a schematic outline of a strategy largely based on this proposal is presented. It is important to note, however, that the presentation of the strategy is unavoidably simplified. For example, it will not take into account quantitative aspects of results obtained *in vitro*, like the magnitude of observed effects and concentration-effect relationships, which may have an impact on the overall testing strategy.

Initial *in vitro* testing

In vitro systems for the detection of genotoxicity are generally comprised of test cells in which genotoxic effects (the so-called "endpoints") can be detected, and an added artificial system for the extracellular metabolic activation of the test agent. This system (usually the so-called "S9 mix") is employed with the objective of compensating for the limitations of the capacity of the tests cells, e.g. bacteria and mammalian cell lines, to metabolically activate the tests agents.

There is agreement among the individual proposals that the first level of initial or basic testing should be comprised of at least two basic *in vitro* tests which cover the two categories of mutations any test strategy must be able to detect, namely gene mutations and chromosomal mutations, with high sensitivity. The test system of choice for the detection of gene mutations is the bacterial gene mutation assay. For the detection of chromosomal mutations, the *in vitro* chromosomal aberration test, which detects structural chromosome aberrations, but not genome mutations, has been used for a long time. It is now suggested to use the *in vitro* micronucleus assay for the detection of chromosomal effects, primarily because it provides the possibility to identify both clastogens and aneugens (Kirsch-Volders et al., 1997).

This general scheme of basic testing may not be applicable in a meaningful way if there is evidence that the S9 mix may be inappropriate for metabolic activation of the agent to be tested. In this case the use of metabolically competent cells, e.g. primary hepatocytes, in which indicator effects can be detected, or transgenic cell lines engineered to express the relevant enzymes may provide an alternative test strategy.

Follow-up testing

Following the initial *in vitro* tests, the further line of action is dependent on the outcome of these assays. Four possible cases can be envisioned:

(a) In case both tests yield unequivocally negative results, the agent can generally be regarded as non mutagenic and further confirmatory testing may be deemed unnecessary. However, if there are any specific aspects (e.g., substance class–specific factors, inappropriate representation of *in vivo* metabolism) that raise concern about the reliability of the results, further testing, possibly using a non-routine strategy, must be considered.

(b) In case only the bacterial gene mutation assay yields positive results and in the absence of evidence that a bacteria-specific metabolic pathway was responsible for the positive outcome, the capacity of the agent to induce gene mutations should be further assessed in mammalian cells *in vitro*, i.e. in the HPRT or $TK^{+/-}$ gene mutation assay. A negative result could be taken as indication that the agent does not induce mutations in mammalian cells unless there are other compound-specific reasons for concern (see (a)). A positive outcome would indicate a mutagenicity of the agent in mammalian cells *in vitro* and necessitate further studies (see (c)).

(c) In case only the in vitro micronucleus test yields positive results, the result should be confirmed in an *in vivo* assay unless it can be shown convincingly that it was caused by experimental factors specific for the *in vitro* situation.

(d) In case positive results are obtained in both test systems, the same strategy will be followed as in case (c).

Further *in vivo* testing subsequent to positive results obtained *in vitro* is generally mandatory as the *in vitro* assays, in particular the micronucleus assay and the $TK^{+/-}$ gene mutation assay, have a rather limited specificity and are, thus, prone to yield a large proportion of "false positives" (see chapter 3.4.1). The particular assay and the experimental protocol employed for the follow-up testing *in vivo* should be selected case-by-case based on the available information on the toxicokinetics of the agent, e.g. the systemic availability, the exposed organs and the pathways possibly involved in its metabolism, as well as the specific genetic endpoint probably affected. If regarded appropriate on the basis of the aforementioned considerations, the *in vivo* micronucleus assay is used to further evaluate the relevance of positive results of *in vitro* tests or, in specific cases, to add further confirmatory evidence to negative results obtained. A negative result of the micronucleus assay would only be meaningful if there is definite evidence from toxicokinetic studies that the test agent actually reaches the bone marrow in significant amounts. In specific cases, e.g., when it is known that the test compound is metabolised in the liver and the metabolites formed are too short-lived to reach the bone marrow, even the demonstrated bioavailability of the compound itself in the bone marrow will not suffice to consider bone marrow an appropriate target for an *in vivo* assay. In these situations, the use of an *in vivo* assay able to detect indicator effects in exposed tissues or organs where mutation induction is difficult or impossible to attest will represent an alternative or supplementation to the micronucleus assay. These assays include the UDS test and the Comet assay.

Consistently negative results obtained in combinatorial tests of the type discussed above for tissues or organs definitely exposed to the test agent or its reactive metabolite(s) would generally result in a classification of the compound as non-genotoxic *in vivo*. In case the compound is known to be carcinogenic, the results from the genotoxicity assays would then support the conclusion that genotoxicity is most likely not involved in tumour formation.

A positive outcome of an *in vivo* micronucleus assay, which cannot be attributed to compound-related disturbances in the physiology of the test animals, would establish the mutagenicity of the agent *in vivo*, and in the absence of further relevant data, the agent should be regarded as a potential genotoxic carcinogen and a potential germ cell mutagen. A similar conclusion should be drawn from positive results obtained with the *in vivo* UDS test. If uncertainties regarding the outcomes of these *in vivo* tests remain, further follow up-studies, such as *in vivo* mutation assays using transgenic animals, should be considered.

3.5 Request 4

If possible, to identify harmonized methodology/approach for the assessment of substances that are genotoxic and for which evidence of carcinogenicity is limited or missing, taking into account all factors that may be of relevance (e.g. differences in their mechanism of action)

- Formally, a carcinogenic effect can only be assessed in experimental studies on animals or in epidemiological investigations. If valid studies of these types are not available or if studies have yielded limited information only, *in vitro* and *in vivo* approaches can provide some information on the likelihood that a given agent is carcinogenic. It should be realised, however, that these studies cannot give a definite answer to the question about a carcinogenic potential. Moreover, due to this inherent uncertainty and the use of surrogate endpoints, i.e. genotoxicity instead of carcinogenicity, they can be used for hazard identification but, so far, not for risk assessment (i.e. derivation of numbers).

- For a compound known to be "genotoxic", the biological relevance of the observed genotoxicity will depend on the assay system(s) in which its genotoxic effects were detected. Ideally, a test strategy as described in Chapter 3.4 "Strategies for genotoxicity testing" should be adopted to accumulate as much knowledge as possible on both the potential mechanism of action and the concentration-response relationship and the dose-response relationship (*in vivo*).

- Due to the limitations of *in vitro* and *in vivo* test systems outlined in Chapter 3.4.1 "Ability of *in vitro* and *in vivo* genotoxicity tests to correctly predict the carcinogenicity of chemicals", a defendable prediction of carcinogenicity associated with an exposure to a compound for which such genotoxicity data are available cannot be made.

- In specific cases, agents may be regarded as potential genotoxic carcinogens on the basis of positive *in vivo* genotoxicity tests. Although the positive predictivity of most *in vivo* genotoxicity assays is limited, compounds shown to induce DNA strand breaks, chromosomal mutations and, particularly, unscheduled DNA synthesis or gene mutations *in vivo* in addition to positive *in vitro* data are highly suspect of being carcinogenic. These compounds may be considered potential carcinogens until results from appropriate repeated dose studies in animals or human data become available.

- In specific cases, compounds may be regarded as putative genotoxic carcinogens even in the absence of *in vivo* genotoxicity data. This may be justified, if there is convincing evidence for compounds with a closely related structure that have been demonstrated to be carcinogenic, that the compound in question will behave in a similar way and cause similar biological effects, i.e. tumour formation. Compounds of this type include certain alkylating agents, nitrosamines and certain polycyclic aromatic hydrocarbons. For every single compound, however, it is mandatory that the structure-activity relationship, on which the assessment is based, is delineated explicitly. In summary, each compound requires a case-by-case decision, taking into consideration all available information on the compound under consideration, i.e. its toxicokinetics, toxicodynamics and mode of action to evaluate the weight of evidence to justify categorisation without appropriate long term studies.

  As stated above, there is general agreement that data from *in vitro* and *in vivo* genotoxicity tests are not suitable for the quantification of the carcinogenic risk. Recently Sanner and Dybing (2005b) discussed the possibility to predict the carcinogenic potency of compounds for which genotoxic mechanisms are likely or possible from the lowest effective dose (LED) obtained from *in vivo* genotoxicity studies. On the basis of IARC monographs a linear correlation between the lowest effective dose for *in vivo* genotoxicity after oral administration and inhalation exposure and the lowest dose descriptor T25 for tumour formation in long-term studies was found. If further evaluated, this approach may become a tool for risk assessment of genotoxic carcinogens without a long-term study.

- For compounds which induce aneuploidy and which are clearly devoid of any DNA reactivity or clastogenic activity, the dose-response relationship observed *in vivo* is an important aspect that needs to be taken into consideration for risk

assessment. Aneuploidy induction does not result from an interaction of the compound or its metabolites with the DNA, but with proteins, which are present in multiple copies and which, thus, represent highly redundant targets. As a consequence, a certain minimal level of protein damage is required before the process of chromosome separation and distribution to the daughter cells is impaired. This minimal level is assumed to constitute a threshold below which no disturbance of chromosome segregation occurs. In case there is clear evidence for a certain experimental aneugen that the minimal concentration required for aneuploidy induction cannot be attained in humans, an assessment of this compound as potential human carcinogen is not warranted.

4. EXECUTIVE SUMMARY

The SCs conclude that risk assessment of compounds that are both genotoxic and carcinogenic should be done on a case-case basis. Whenever sufficient information is available, the appropriate dose descriptor as a starting point is identified to either apply the linear extrapolation or the MOE approach. Both have their advantages and disadvantages. From a scientific point of view the MOE approach avoids the usually intrinsic uncertainties of the extrapolation models and the presentation of a virtual cancer risk of a high uncertainty.

The MOE and linear extrapolation approaches combine information on human exposure and the potency of the carcinogen. The SCs have identified several conditions where one or the other approach may be preferable:

- For risk communication the MOE is seen preferable
- For prioritisation of measures to reduce risk, both the MOE approach and the linear extrapolation from a dose descriptor are applicable
- Linear extrapolation is a method that provides a quantitative expression of risk that is commensurate with cost benefit analysis. It provides likely units of life lost or gained under certain exposure conditions.

When appropriate the SCs also recommend application of the TTC.

The ALARA principle is a valuable measure to minimize exposure to genotoxic and carcinogenic substances. It is a qualitative procedure and not applicable for risk assessment.

Genotoxic effects, i.e. toxic effects on the genome of cells, comprise both heritable alterations (mutations) and damage to the genetic material which cannot be inherited but has the potential to cause the formation of mutations after further processing. Genotoxic effects can be assessed both *in vitro* and *in vivo*. For basic *in vitro* testing, assays capable of identifying both gene mutations and chromosomal mutations must be employed. Most *in vitro* assays, particularly those detecting DNA breakage and associated chromosomal effects, have low specificity and, consequently, limited predictivity with respect to the carcinogenic potential of agents. Therefore, the demonstration of a genotoxic activity of an agent *in vitro* usually entails further studies *in vivo*.

The specific assay(s) employed for the follow-up testing *in vivo* should be selected case-by-case, based on the available information on the toxicokinetics of the agent, the pathways possibly involved in its metabolism and the specific genetic endpoint probably affected. In general, the micronucleus assay is recommended as first *in vivo* test, as it shows good positive predictivity for rodent carcinogenicity and as it can detect both chromosome and genome mutations. In specific cases, however, e.g. when organs and tissues other than the bone marrow are regarded as potential targets of genotoxicity, other tests able to detect local genotoxic effects should be employed. These assays include tests for the induction of DNA repair synthesis (UDS) or DNA strand breaks.

In summary, each compound requires a case-by-case approach, taking into consideration all available information on the toxicokinetics, toxicodynamics and mode of action for the

evaluation of the weight of evidence, which may justify categorisation without appropriate long-term studies. A quantification of the carcinogenic risk associated with the exposure to an *in vivo* genotoxin is not possible solely on the basis of data from *in vitro* and *in vivo* genotoxicity tests; it requires appropriate repeated dose studies in animals.

5. TABLE 1 - Summary on the available risk assessment methodologies: advantages and disadvantages

Method	Note	Advantages	Disadvantages
Quantitative risk characterisation by linear extrapolation	Linear extrapolation is performed from animal carcinogenicity data to the low doses usually encountered for human exposures. Corrections are made for differences in metabolic rate of the animal used and humans.	• Provides a quantitative risk estimate • Applicable to cumulative and aggregate exposures • Can be used for prioritisation of measures to reduce risk • Can be used in cost-benefit analyses	• Requires determination of a dose-descriptor e.g. T25 from animal data • Requires linear extrapolations of dose-response curves outside of the observable range. However, in cases where good epidemiology data are available a good agreement with the epidemiology data have been found • Requires human exposure data
Margin of exposure (MOE)	A ratio between a dose resulting a defined carcinogenic response and measured or estimated human exposure. Used by EFSA.	• Provides a risk estimate • Applicable to aggregate exposures • Avoids extrapolations of dose-responses outside the observable range • Can be used for prioritisation	• Requires quantitative potency data on animal or human carcinogenicity • Requires human exposure data • Requires adjustments based on the type of the dose-response reference point
As low as reasonably achievable (ALARA)	Aims to keep the exposure to carcinogenic chemicals at the lowest achievable level. A tool for risk management.	• Requires only hazard identification • No exposure data needed • No quantitative risk estimate needed • Easy to communicate	• Cannot be used for risk comparison or prioritization • Carcinogenic potency not considered • Exposure not considered
Threshold of toxicological concern (TTC)	Human exposure threshold values for daily uptake of a chemical, below which there is no appreciable risk to human health. Developed originally for chemicals in food. Applicability for other chemical groups under	• Simple and easy to use • Does not require potency data • Can be used for prioritization	• Based on worst case assumptions • requires human exposure data • Assumes no threshold for carcinogenicity • Certain structural classes of chemicals excluded due to insufficient data • Databases need further development and validation

	evaluation.		
Structure-activity relationships (SARs)	*In silico* method that utilises chemical structure, physico-chemical properties and structural alerts for effect and potency predictions.	• No experimental studies needed, quick and cheap • Suitable for screening and prioritization of structurally related chemicals • May be useful for analysis of mode of action • May be useful in combination with experimental test systems	• Poor predictivity beyond structurally and mechanistically similar chemicals • Very limited use as a stand-alone method
Epidemiological data for quantitative risk characterisation	The relevant study types are case-control studies and cohort studies.	• No species extrapolation needed • Realistic exposure levels • Suitable for calculating excess lifetime risks	• Require good data at several exposure levels • Limited availability of appropriate studies • Usually poor exposure assessment • Subject for confounding and bias • Long follow up needed in cohort studies
No-observed-adverse-effect –level (NOAEL)	The highest dose-level used in experimental studies that is not associated with adverse effects.	• Simple to use	• Dependent on number of test animal, choice of doses and spacing between doses • Ignores the shape of dose-response curve
T25	The chronic daily dose which will induce 25% of the animals tumours at a specific tissue site after correction for spontaneous incidence, within the standard life span of that species. Assumes linearity. Currently used in the EU.	• A simplified and robust index of carcinogenic potency • No computer program needed • Suitable for carcinogen potency classification and labelling • Can be used for quantitative risk characterisation and MOE calculations	• Potency estimates not very accurate, but generally in reasonable agreement with BMDL determinations • Sensitive to experimental design differences

Benchmark dose lower confidence limit (BMDL)	The dose associated with the lower 95% confidence limit for a predetermined change in response. Based on dose-response modelling. Recommended by the EFSA Scientific Committee.	• Takes into account the shape and other information of the dose-response curve • Corresponds to an explicit response level • Use of lower bound confidence limit appropriately reflects the sample size and uncertainty • Can be used for quantitative risk characterisation and MOE calculations	• Requires at least 3 dose-levels • Currently used guideline study designs are not optimal for BMD modelling (too few doses)
Short-term genotoxicity tests	*In vitro* and short-term *in vivo* studies as part of genotoxicity test batteries, mechanistic studies for specific purposes. Form the basis for the current genotoxicity classification.	• Suitable for hazard identification, prioritization and clarification of mechanisms • Reasonable predictivity for rodent carcinogens • Provide mechanistic information • Quick and cheap	• Do not provide quantitative risk estimate • Limited use as stand-alone methods • Usually require high exposure levels
Long-term carcinogenicity bioassays *in vivo*	Carcinogencity bioassays in rodents, initiation / promotion studies, carcinogenicity studies in transgenic animals.	• Suitable for hazard identification and hazard characterization • Provide quantitative risk estimate	• Require animal experimentation • Time consuming, laborious and expensive • Usually require high dose-levels that may involve toxicity mechanisms not relevant for human low dose exposures

6. REFERENCES

Anderson LM (2004). Introduction and overview. Perinatal carcinogenesis: growing a node for epidemiology, risk management, and animal studies. Tox Appl Pharmacol 199: 85-90.

Angerer J, Ewers U, Wilhelm M (2007) Human biomonitoring: state of the art Int J Hyg Environ Health. 210(3-4):201-28.

Ashby J, Tennant RW (1991) Definitive relationships among chemical structure, carcinogenicity, and mutagenicity for 301 chemicals tested by the US NTP. Mutation Research 257, 229-306

Autrup H (2004). Gene-Environment interaction in environmental carcinogenesis, In: Cancer as an Environmental Disease (Eds: Nicolopou-Stamati et al) Kluwer Academic Publishers, pp 71-94.

Barlow S, Renwick AG, Kleiner J, Bridges JW, Busk L, Dybing E, Edler L, et al. (2006): Risk assessment of substances that are both genotoxic and carcinogenic - Report of an international conference organised by EFSA and WHO with support of ILSI Europe. Food Chem Toxicol 44, 1636–1650

Benigni R, Bossa C, Netzeva TI, Rodomonte A & Tsakovska I (2007) Mechanistic QSAR of Aromatic Amines: New Models for Discriminating Between Homocyclic Mutagens and Nonmutagens, and Validation of Models for Carcinogens. Environmental and Molecular Mutagenesis 48, 754-771.

Benigni R, Zito R (2004) The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutat. Res. 566, 49-63

Benigni R, Passerini L (2002) Carcinogenicity of the aromatic amines: from structure-activity relationships to mechanisms of action and risk assessment. Mutation Research 511, 191-206

Blackburn K., J. A. Stickney, H. L. Carlson-Lynch, P. M. McGinnis, L. Chappell, S. P. Felter (2005) Application of the threshold of toxicological concern approach to ingredients in personal and household care products Regulatory Toxicology and Pharmacology 43 249–259

Boffetta P., van der Hel O, Norppa H, Fabianova E, Fucic A, et al (2007): Chromosomal Aberrations and Cancer Risk: Results of a Cohort Study from Central Europe Am J Epidemiol; 165(1): 36-43

Bolt HM, Foth H, Hengstler JG, Degen GH. (2004) Carcinogenicity categorization of chemicals – new aspects to be considered in a European perspective Toxicol Letters 151: 29-41

Bolt HM, Huici-Montagud A (2008) Strategy of the scientific committee on occupational exposure limits (SCOEL) in the derivation of occupational exposure limits for carcinogens and mutagens. Arch Toxicol 82, 61-64

Bonassi S, Znaor A, Ceppi M, Lando C, Chang WP, et al (2004) Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective Cytogenet Genome Res, 104, 376-82

Bonassi S, Norppa H, Ceppi M et al (2008) Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: results from a pooled cohort study of 22348 subjects in 11 countries. Carcinogenesis 29: 1178-1183.

Boogaard PJ. (2007) Human biomonitoring activities programmes by industry. Int J Hyg Environ Health. 210, 259-61

Brandt PA, Voorrips L, Hertz-Picciotto I, Shuker D, Boeing H, Speijers G, Guittard C, Kleiner J, Knowles M, Wolk A, Goldbohm A (2002) The contribution of epidemiology to risk assessment of chemicals in food and diet. Food Chem Toxicol, 40, 387-424

Bristol DW, WAchsman JT, Greenwell A (1996) The NIEHS predictive Toxicology evaluation project. Environ Health Perpect 104,: 1001-1010

Calafat AM, Ye X, Silva MJ, Kuklenyik Z, Needham LL. (2006) Human exposure assessment to environmental chemicals using biomonitoring. Int J AndrolFeb;29(1):166-71

Cheeseman MA, Machuga EJ, Bailey AB (1999) A tiered approach to threshold of regulation. Food and Chemical Toxicology 37, 387-412.

Cimino MC (2006) Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environmental and Molecular Mutagenesis, 47: 362–390

Contrera JF, Kruhlak NL, Matthews EJ, Benz RD (2007) Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. Regul Toxicol Pharmacol. 49(3):172-82.

D'Errico, A.; Malats, N.; Vineis, P. Boffetta, P. (1999) Review of studies of selected polymorphisms and cancer. In: W. Ryder (ed.) metabolic polymorphisms and susceptibility to cancer. IARC Scientific Publication n. 148, Lyon, International Agency for Research on Cancer 1999, pp. 323-393.

Delaney E. J. (2007) An impact analysis of the application of the threshold of toxicological concern concept to pharmaceuticals Regulatory Toxicology and Pharmacology 49 107–124

Dorne JLCM, Walton K, Renwick AG (2005) Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. Food Chem Toxicol 43: 203-216.

Dybing E, Sanner T, Roelfzema H, Kroese D, Tennant RW. (1997) T25: A simplified carcinogenic potency index: Description of the system and study of correlations between carcinogenic potency and species/site specificity and mutagenicity. Pharmacol Toxicol 80: 272–279

EC (1999). Guidelines for setting specific concentration limits for carcinogens in Annex 1 of directive 67/548/EEC. Inclusion of potency considerations. Commission working group on the classification and labelling of dangerous substances. Brussels

ECHA (2008) Guidance on information requirements and chemical safety assessment Chapter R.8: Characterisation of dose [concentration]-response for human health

EFSA (2005) Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. The EFSA Journal 282, 1-31. http://www.efsa.eu.int/science/sc_commitee/sc_opinions/1201/sc_op_ej282_gentox_summary_en1.pdf

EMEA committee for medicinal products for human use: Guidelines on the assessment of genotoxic constituents in herbal substances/preparations 31 October 2007

EPA (US Environmental Protection Agency; 1986). Guidelines for carcinogen risk assessment. Fed. Reg. 51: 33992–34003

EPA (2005a) US Environmental Protection Agency Guidelines for Carcinogen Risk Assessment EPA/630/P-03/001F Washington DC.

EPA (2005b) US Environmental Protection Agency Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens, EPA/630/R-03/003F, 2005. Washington, DC

EPA (1996) US Environmental Protection Agency Proposed guidelines for carcinogen risk assessment. Fed. Reg. 61: 17960–18011

Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, et al. (2001) Metabolic gene polymorphism frequencies in control populations. Cancer Epi Biomarker, Preven 10: 1239-1248.

Goldbohm RA, Tielemans ELJP, Heederik D, Rubingh CM, Dekkers S, Willem M, Kroese ED (2006): Risk estimation for carcinogens based on eidemiological data: A structured approach, illustrated by an example of chromium. Regul Toxicol Pharmacol 44, 294-310

Gonzales FJ (2007) The 2006 Bernard B. Brodie Award Lecture CYP2E1. Drug Metab Disposition 35, 1-8.

Gu, J., Bondy, M. L., Sigurdson, A., Spitz M. R., Hsu, T. C., Wu, X. (1999) Three Measures of Mutagen Sensitivity in a Cancer-Free Population. Cancer Genet. Cytogenet. 110, 65-69.

Hagmar l, Bonassi S, Strömberg U, Brøgger A, Knudsen lE, Norppa H, Reuterwall C (1998) Chromosomal aberrations in lymphocytes predict human cancer : A report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). Cancer Res.; 58(18): 4117-21

Haseman JK, Zeiger E, Shelby M, Margolin B, Tennant R. (1990) Predicting rodent carcinogenicity from four *in vitro* genetic toxicity assays: an evaluation of 114 chemicals studied by the National Toxicology Program. J Am Stat Assoc. 85:964–971.

Hirvonen A (1999) Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. Environ. Health Perspect. 107 (suppl. 1), 37-47.

Huang C-Y, Huang K-L, Cheng T-J, Wang J-D, Hsieh L-L (1997) The GSTT1 and CYP2E1 genotypes are possible factors causing vinyl chloride induced abnormal liver function. Arch Toxicol 71: 482-488.

Humfrey C. D. N. (2007) Recent Developments in the Risk Assessment of Potentially Genotoxic Impurities in Pharmaceutical Drug Substances Toxicological Sciences 100: 24–28

Kirkland D, Aardema M, Henderson L, Müller L. (2005) Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. Mutat Res. 584(1-2):1-256

Kirsch-Volders M, Elhajouji A, Cundari E, Van Hummelen P. (1997) The *in vitro* micronucleus test: a multi-endpoint assay to detect simultaneously mitotic delay, apoptosis, chromosome breakage, chromosome loss and non-disjunction.Mutat Res. 392:19-30

Krewski D, Cardis E, Zeise L, Feron F. (1999) Empirical approaches to risk estimation and prediction. In: *Quantitative estimation and prediction of human cancer risks.* Eds.: Moolgavkar S, Krewski D, Zeise L, Cardis E, Møller H. IARC Scientific Publications No. 131: 131–178, International Agency for Research on Cancer, Lyon

Kroes R, Galli C, Munro I, Schilter B, Tran L-A, Walker R, Würtzen G (2000) Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food and Chemical Toxicology 38, 255-312

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, et al. (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the dies. Fod and Chemical Toxicology 42, 65-83

Kroes R., A.G. Renwick, V. Feron, C.L. Galli, M. Gibney, H. Greim, R.H. Guy, J.C. Lhuguenot, J.J.M. van de Sandt (2007) Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients Food and Chemical Toxicology 45 2533–2562

Lorge E, Gervais V, Becourt-Lhote N, Maisonneuve C, Delongeas JL, Claude N. (2007) Genetic toxicity assessment: employing the best science for human safety evaluation part IV: a strategy in genotoxicity testing in drug development: some examples. Toxicological Sciences, 98: 39-42.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. (2006a) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul Toxicol Pharmacol. 44, 83-96

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. (2006b) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regul Toxicol Pharmacol. 44(2):97-110

Matthews EJ, Kruhlak NL, comino MC, Benz RD, Contrera JF (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II: Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regul Toxicol Pharcol 44, 97-110

Mayer J, Cheeseman MA, Twaroski ML (2008) Structure-activity relationship analysis tools: Validation and applicability in predicting carcinogens. Regulatory Toxicology and Pharmacology 50 (2008) 50-58

Mohrenweiser HW, Wilson DM 3rd, Jones IM. (2003) Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. Mutat Res. May 15;526(1-2):93-125

Munro IC, Renwick AG, Danielewska-Nikiel B. (2008) The Threshold of Toxicological Concern (TTC) in risk assessment. Toxicol LettAug 15;180(2):151-6

Müller L, Mauthe RJ, Riley CM, Andino MM, De Antonis D, Beels, J. et al. (2006) A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity Regul Toxicol Pharmacol 44, 198-211

Needham LL, Calafat AM, Barr DB (2007). Uses and issues of biomonitoring. Int J Hyg Environ Health. May;210(3-4):229-38

Neumann H-G, Thielmann HW, Filser JG, Gelbke HP, Greim H, Kappus H, et al. (1998) Changes in the classification of carcinogenic chemicals in the work area. J Cancer Res Clin Oncol 124, 661–669

Norppa H., Bonassi S., Hansteen I.-L., Hagmar L., Strömberg U (2006) Chromosomal aberrations and SCEs as biomarkers of cancer risk. Mutat Res 600, 37-45

O'Brien J, Renwick AG, Constable A, Dybing E, Müller DJG, Schlatter J, Slob W, Tueting W, van Benthem J, Williams GM, Wolfreys A (2006) :Approaches to the risk assessment of genotoxic carcinogens in food: A critical appraisal. Food and Chemical Toxicology 44, 1613-1635

OECD 2007: Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens. OECD Environment, Health and Safety Publications, Series on Testing and Assessment No. 31

Pavanello S. and Clonfero, E. (2000) Biological indicators of genotoxic risk and metabolic polymorphisms. Mutat. Res. 463(3):285-308.

Pfuhler S, Albertini S, Fautz R, Herbold B, Madle S, Utesch D, Poth A. (2007) Genetic toxicity assessment: employing the best science for human safety evaluation part IV: Recommendation of a working group of the Gesellschaft fuer Umwelt-Mutationsforschung (GUM) for a simple and straightforward approach to genotoxicity testing. Toxicol Sci. 97:237-40

Pirkle JL, Osterloh J, Needham LL, Sampson EJ. (2005) National exposure measurements for decisions to protect public health from environmental exposures. Int J Hyg Environ Health.;208(1-2):1-5

Renwick AG, Barlow SM, Hertz-Piccotto L, Boobis AR, Dybing E, (2003) Risk characterization of chemicals in food and diet. Food Chem Toxicol 41, 1211-1271

RIP (2007) Technical Guidance Document For Preparing The Chemical Safety Assessment Part B: Hazard Assessment, Chapter R.8, 3.2-2: Human Health Hazard Assessment

Rossner P., Boffetta P., Ceppi K., Bonassi S., Smerhovsky Z. (2005) Chromosomal aberrations in lymphocytes of healthy subjects and risk of cancer. Environ Health Perspect; 113(5): 517-20

Sanner T, Dybing E, Willems MI, Kroese D. (2001) A simple method for quantitative risk assessment of non-threshold carcinogens based on the dose descriptor T25. Pharmacol. Toxicol. 88: 331–341

Sanner T, Dybing E. (2005a) Comparison of carcinogen hazard characterisation based on animal studies and epidemiology. Basic Clinical Pharmacol Toxicol. 96: 66–70

Sanner T, Dybing E. (2005b) Comparison of carcinogenic and *in vivo* genotoxic potency estimates. Basic Clinical Pharmacol Toxicol. 96: 131–139,

SCCP/SCHER/SCENIHR (2008) Use of the Threshold of Toxicological Concern (TTC) approach for the health risk assessment of chemicals - in preparation

SCHER (2006) opinion on the revised chapter for "Human Health Risk Characterisation of the Technical Guidance Document on Risk Assessment" 19 September 2006

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J et al (2001) International Programme on Chemical Safety IPCS conceptual framework for a evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol 34, 146-152

Steenland K and Greenland S (2004) Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer Am J Epidemiol, 160, 384-392

Swaen GMH. (2006) A framework for using epidemiological data for risk assessment Hum Exp Toxicol.; 25: 147-155

Swenberg JA, Fryar-Tita E, Jeong YC, Boysen G, Starr T, (2008) Biomarkers in toxicology and risk assessment: informing critical dose-response relationships. Chem Res Toxicol. 21(1):253-65

Taningher, M. Malacarne, D. Izzotti, A. Ugolini, D Parodi, S. (1999) Drug metabolism polymorphisms as modulators of cancer susceptibility. Mutat. Res. 436(3); 27-61.

Tennant RW, Margolin BH, Shelby MD, Zeiger E, Haseman JK (1987) Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicity assays Science 236:933–941

TGD (2007) Technical Guidance Document for Preparing the Chemical Safety Assessment. Chapter R.8: Human Health Hazard Assessment, R.8.8.14 Appendix 14

TGD: Technical Guidance Document on Risk Assessment in support of Directive 93/67/EEC on risk assessment for new notified substances and Regulation 1488/94 on risk assessment for existing substances and Directive 98/8/EC concerning the placing of biocidal products on the market. Human Health Risk Characterization revised chapter, final draft 10.11.2005

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T (2007a) Report of the IWGT working group on strategies and interpretation of regulatory in vivo tests. I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. Mutation Research, 627: 78–91.

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T (2007b) Report of the IWGT working group on strategy/interpretation for regulatory in vivo tests. II. Identification of in vivo-only positive compounds in the bone marrow micronucleus test. Mutation Research, 627: 92–105.

USFDA (1995) Food additives: Threshold of Regulation for substances used in Food-Contact articles: Final Rule. Fed Reg 60: 38582-38596

Veglia F, Loft S, Matullo G et al (2008) DNA adducts and cancer risk in prospective studies: a pooled analysis and a meta-analysis. Carcinogenesis 29: 932-936.

WHO (2004) Principles for modelling dose-response for the risk assessment of chemicals. World Health Organization, Geneva, 2004. http://www.who.int/ipcs/methods/harmonization/draft_document_for_comment.pdf

WHO (2005). Harmonization Project Document No.2: Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration – response assessment. World Health Organization, Geneva, 2005. Available from: http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf

WHO (2006). Assessing human health risks of chemicals: Derivation of guidance values for health-based exposure limits. Environmental Health Criteria. 170: 1-40.

WHO (2008). WHO/IPCS Harmonization Project Draft Guidance on "Mutagenicity for Chemical Risk Assessment". http://www.who.int/ipcs/methods/harmonization/areas/mutagenicity_testing/en/index.html

WHO (2008) Principles of characterizing and applying physiologically-based pharmaco-kinetic and toxicokinetic models in risk assessment. World Health Organization, Geneva, 2008 http://www.who.int/ipcs/methods/harmonization/areas/pbpkguidance_draft.pdf

Zeiger E (1998) Identification of rodent carcinogens and non-carcinogens using genetic toxicity tests: premises, promises and performance. Regul Toxicol Pharm. 28:85–95.

Zeiger E, Haseman JK, Shelby MD, Margolin BH, Tennant RW. (1990) Evaluation of four *in vitro* genetic toxicity tests for predicting rodent carcinogenicity: confirmation of earlier results with 41 additional chemicals. Environ Mol Mutagen. 16 (suppl. 18):1–14

Znaor A., Fučic´A., Strnad M., Barkovic´ D., Skara M., Hozo I. (2003) Micronuclei in peripheral blood lymphocytes as a possible cancer risk biomarker: a cohort study of occupationally exposed workers in Croatia. Croat Med J. 44(4): 441-6

7. LIST OF ABBREVIATIONS

ALARA	As Low As Reasonably Achievable
BMD	Benchmark Dose
BMDL	Benchmark Dose Lower Limit
CI	Correlation Indicator
DFG	German Research Foundation
DMEL	Derived Maximum Exposure Level
EFSA	European Food Safety Authority
EPA	Environmental Protection Agency
GHS	Globally Harmonized System
GSTT1	Glutathione S-transferase theta 1
HPRT	Hypoxanthine Guanine Phosphoribosyl Transferase
IARC	International Agency for Research of Cancer
ICSAS	Informatics and Computational Safety Analysis Staff
IPCS	International Programme on Chemical Safety
IRIS	Integrated Risk Information System
LED	Lowest Effective Dose
MOE	Margin of Exposure
MTD	Maximum Tolerated Dose
NO(A)EL	No Observed (Adverse) Effect Level
OECD	Organisation for Economic Co-operation and Development
(Q)SAR	(Quantitative) Structure-Activity Relationship
SCs	Scientific Committees
TGD	Technical Guidance Document
TTC	Threshold of Toxicological Concern
UDS	Unscheduled DNA Synthesis

8. GLOSSARY

Aneuploidy	Abnormal number of chromosomes
Carcinogen	Substance or agent that increases the incidence of neoplasms in experimental animals and/or humans
Clastogens	Substance or agent that causes chromosome breakage and (or) consequent gain, loss or rearrangement of pieces of chromosomes.
False positive / negative	A result that is erroneously positive/negative
Genotoxic agent	Substance or agent that affects the genetic material of cells.
Genotoxic carcinogen	Substance capable of producing cancer by directly altering the genetic material of target cells
HT25	$HT25=T25/(\text{weight human/weight animal})^{0.25}$
In vitro	Refers to the technique of performing a given experiment in a controlled environment outside of a living organism
In vivo	Refers to experimentation done in or on the living tissue of a whole, living organism
LED10	The 95% per cent lower confidence interval for a 10 per cent increase in tumour incidence and is determined by fitting dose-response data to various mathematical models (US EPA, 1996)
Micronuclei	Small extra nuclear chromatin-containing bodies in cells
Mutagenic agent	Substance or agent capable of causing permanent alterations of the genetic material (mutations)
Neoplasm	Structure resulting from the abnormal proliferation of cells.
Non-genotoxic carcinogen	Substance capable of producing cancer by mechanism not related to initial damage to the genetic material.
S9 mix	Artificial system for the extra cellular metabolic activation of the test agent in *in vitro* systems
T25	Chronic daily dose which will give tumours in 25% of the animals above background at a specific tissue site (Dybing et al., 1997)
Transgenic	Refers to an organism whose genome has been altered by the transfer of one or more genes.

ANNEX I

TEST SYSTEMS FOR GENOTOXICITY

Genotoxic effects, i.e. toxic effects on the genome of cells, comprise both mutations, which are alterations of the information content or the segregation of DNA that can be passed on to the next cell generation or the offspring, and damage to the genetic material which cannot be inherited but has the potential to cause the formation of mutations after further processing.

The terms "genotoxic" and "mutagenic" or "genotoxicity test" and "mutagenicity test" must, therefore, not be considered synonymous. Whereas the detection of mutations caused by a test agent in mutagenicity test systems may enable direct conclusions with regard to its impact on the genetic material, the detection of mere damage is usually less meaningful.

The demonstration of DNA damage or cellular responses specific to the presence of DNA damage by so-called "indicator tests" may, however, yield important information on the potential mutagenicity of an agent in cases where mutagenicity is difficult to determine directly, such as in primary cells *in vitro* which do not proliferate to any significant extent or in certain tissues *in vivo* in which the induction of mutations is difficult or impossible to determine.

For most of the test described below guidelines are available from OECD, EU ICH (Cimino 2006; Lorge et al., 2007; Tweats et al., 2007a, b; WHO, 2008)

I. In vitro test systems for genotoxicity

In vitro systems for the detection of genotoxicity are generally comprised of test cells in which genotoxic effects (the so-called "endpoints") can be detected, and an added artificial system for the extracellular metabolic activation of the test agent. This artificial system is generally required as the test organisms employed, e.g. bacteria and mammalian cell lines have a very limited capacity for the metabolic activation of chemicals. The standard system used in nearly every *in vitro* test for genotoxicity is the so-called "S9 mix", a mixture of a cell fraction obtained as supernatant after centrifugation of liver homogenate of rodents, usually rats, at 9000 x g ("S9") and an enzymatic system for the generation of NADPH, a cofactor required for the metabolism of chemicals by cytochromes P450 (CYPs). The enzymatic activity of the S9 is generally increased by pre-treatment of the animals with chemical inducers of the CYP system, and its activating capacity is further augmented by the virtual absence of the activity of inactivating conjugating enzymes due to the lack of their enzymatic cofactors. However, this lack can also prevent the detection of genotoxic effects of compounds which require the activity of conjugating enzymes for their activation. Moreover, other potentially activating enzymes are not present in the S9 to any significant extent, and in specific cases their absence may prevent the detection of genotoxic effects of test agents. Available information on the potential route(s) of metabolic activation of the test agent can provide the possibility to supplement the presumably missing enzymes or cofactors. Other problems associated with the use of S9 mix may result when metabolites are formed that are charged and unlikely to penetrate the bacterial cell wall or whose half-life is so short that they cannot reach the genetic material of the test cell. In this case, the use of metabolically competent cells, e.g. primary hepatocytes, in which indicator effects can be detected, or transgenic cell lines engineered to express the relevant enzymes may provide an alternative test strategy.

Mutagenicity tests

Mutations which can be induced by test agents in *in vitro* test systems comprise gene mutations, chromosome mutations and genome mutations. Whereas the induction of gene mutations can be assessed in bacteria and in mammalian cells, the analysis of the induction of chromosome and genome mutations is restricted to mammalian cells and yeasts.

Tests for the induction of gene mutations

The bacterial reverse mutation assay, also colloquially referred to as "Ames test", is the most frequently employed *in vitro* mutagenicity test. The test is performed with various strains of *Salmonella typhimurium* and *Escherichia coli* which, as a consequence of specific, well-defined mutations in genes coding for enzymes for the biosynthesis of a certain amino acid, have lost the ability to grow in the absence of this amino acid. Mutagenic chemicals which induce reverse mutations at selected loci in the genome of these cells convey back the capability to synthesize the respective amino acid and to grow and form visible colonies. The number of these colonies is a direct measure of the mutagenicity of the test agent in the individual strains, which also enables conclusions with respect to the underlying mechanism of mutation induction. The system is very sensitive for the detection of mutations because of the tailor-made properties of the cells and the use of very large numbers of cells (about 10^8 per treatment), which allows the detection of small increases in mutant frequency. The high sensitivity of the assays puts high demands on the purity of test agents, as very minor mutagenic impurities can produce false positive results. As the structure of the bacterial chromosome is different from the much more complex structure of the chromatin in mammalian cells, some types of mutagens, e.g., compounds eliciting chromosomal mutations in mammalian cells as a consequence of an interaction with DNA-associated proteins cannot be detected in bacterial test systems.

In contrast to the bacterial reverse mutation assay, gene mutation assays with mammalian cells detect forward mutations, i.e. the loss of a specific function, usually an enzyme activity, as a consequence of a mutation. The functionally altered cells are identified by selection procedures which allow the detection and quantification of few mutants within a large number of non-mutated cells. In general, selection is based on the mutation-induced resistance to toxic base analogs which will kill the non-mutated cells. The mutants survive and form colonies because the resistance is passed on to the daughter cells.

For routine testing, two assays referred to as "HPRT gene mutation test" and "$TK^{+/-}$ gene mutation test" are usually employed. The HPRT gene mutation test, which is generally performed with Chinese hamster cell lines, e.g., V79 or CHO cells, detects the heritable loss of the activity of the enzyme hypoxanthine guanine phosphoribosyl transferase (HPRT). Loss of this enzyme activity results in the resistance to guanine analogs like 6-thioguanine. It primarily identifies mutations within the *hprt* gene, e.g., base substitutions or small deletions/insertions. Large deletions, which extend into neighboring essential genes and which show up microscopically as chromosome aberrations, are frequently lethal and cannot be detected by the HPRT test. The $TK^{+/-}$ gene mutation test is routinely performed with the mouse lymphoma cell line L5178Y and then also referred to as "mouse lymphoma assay". It detects the heritable loss of the activity of the enzyme thymidine kinase (TK) in cell lines heterozygous for the thymidine kinase gene ($tk^{+/-}$). Loss of the TK activity leads to the insensitivity of the cells to thymidine analogs like trifluorothymidine. The test detects mutations limited to the *tk* gene. Mutants produced by a mutation within the tk^+ allele, which is not essential for growth, form rapidly growing, large colonies. This type of mutation apparently corresponds to the mutations detectable with the HPRT system. However, the test also detects large mutations which can be identified microscopically as chromosome aberrations and which become apparent as slowly growing, small colonies of trifluorothymidine-resistant cells. It appears that compounds or treatment conditions, such as those associated with excessive toxicity, which result in the formation of chromosomal aberrations (see below) also lead to the appearance of small colonies. Electrophilic properties or a reactivity of the compound or its metabolites with DNA are not essential for the induction of this type of mutation.

Thus, the responsiveness of the L5178Y $TK^{+/-}$ system to different types of mutations is broader than that of the HPRT system which does not detect chromosomal mutations.

Tests for the induction of chromosome aberrations

The term "chromosome aberrations" includes both alterations of chromosome structure that are microscopically visible in mitotic cells (structural aberrations) and alterations of chromosome number (numerical aberrations).

The chromosome aberration test detects structural aberrations only, and it can provide an indication of polyploidy. It cannot detect numerical aberrations elicited by aneugens. The first mitoses (metaphases) occurring after treatment of the cells with the test substance are scored, i.e., the analysed cells have not yet divided. Thus, the aberration assay does not detect mutations in the narrower sense, but potential precursors of heritable chromosome aberrations, i.e., chromosome mutations. Nevertheless, the terms "chromosome aberration test" and "chromosome mutation test" are frequently used as synonyms. Chemicals which induce structural chromosome aberrations are also called "clastogens".

Chromosomal aberrations induced by chemicals are usually assessed in human lymphocytes or, more frequently, certain karyotypically stable cell lines, such as various Chinese hamster cell lines or L5178Y mouse lymphoma cells. Scoring for aberrations is performed on metaphase cells, as only in this phase of the cell cycle the chromosomes are amenable to an analysis of their structure by light microscopy. Metaphase analysis requires considerable effort to produce statistically robust data sets. The quality of the performance of the assay and the scoring of the aberrations is of vital importance for the significance of results obtained with this assay.

A prerequisite for the induction of chromosomal aberrations by a compound is its ability to cause DNA double strand breaks. These breaks can be induced directly by an attack of the compound on DNA, or they can be formed during the processing or repair of DNA damage. However, even though the formation of DNA double strand breaks is an essential step in the formation of chromosome aberrations, the clastogenicity of a compound is not necessarily due to the reactivity of the compound or its metabolites with DNA. Chromosomes are complex, dynamic structures made up of DNA and proteins, and their structural integrity depends on a large number of factors including the maintenance of a physiological intracellular milieu, such as a proper osmolality and pH, and the activity of several enzymes. Changes of chromosome structure can be induced by, e.g., disturbances of DNA replication and repair, effects on topoisomerases, depletion of cellular energy, interference with cell membrane function or triggering of apoptosis. Therefore, this assay system is highly sensitive to interferences caused by the toxicity of the test compound and therefore prone to the generation of "false positive" results.

In the *in vitro* micronucleus test, cells are treated with the test agent, and, after they have completed cell division or, in a variant of the assay, only nuclear division, scored for the presence of micronuclei. Micronuclei are small extra-nuclear chromatin-containing bodies. They contain either fragments of chromosomes (produced by chromosome breakage) or complete chromosomes (as a consequence chromosome malsegregation) that have not been integrated into a nucleus of one of the daughter cells during nuclear division. Micronuclei can be easily identified by light microscopy after staining. Whereas the formation of chromosomal fragments (also called acentric fragments) is due to the clastogenic activity of the test agent, the formation of micronuclei containing whole chromosomes is caused by disturbances of the mitotic apparatus of the cells.

Since the formation of micronuclei consisting of acentric fragments and whole chromosomes is based on completely different mechanisms, differentiation of the two types of micronuclei can yield valuable information on the mechanism of action of a test agent. The differentiation is achieved by examining the micronuclei for the presence of centromere-specific chromosome regions by either anti-kinetochore antibodies or fluorescence *in situ* hybridization (FISH) with centromere-specific DNA probes. Thus, the micronucleus test allows the identification of clastogens**,** which cause chromosome breaks, and of spindle poisons, which disturb chromosome segregation. The micronucleus test is particularly valuable because no other assays for the reliable detection of aneuploidies are currently available.

The *in vitro* micronucleus test can be conducted with any mammalian cell line and certain primary cells, such as mitogen-stimulated freshly isolated hepatocytes. Due to the ease of evaluation, large numbers of cells can be scored which increases the reliability of the results obtained. Like the chromosome aberration test, however, it appears to be prone to false-positive results although a sound quantitative comparison of the susceptibility of the two assays to in vitro artefacts is not yet possible due to the relatively limited database for the micronucleus test.

Indicator tests

The term "indicator tests" signifies a group of genotoxicity assays which do not detect the induction of mutations in the progeny of the treated cells. Rather, various effects mechanistically associated with the formation of mutations are identified directly in the treated cells.

Tests for covalent binding of chemicals to DNA

The capability to react with DNA and to form covalent DNA adducts is a characteristic property of many mutagens and genotoxic carcinogens or their metabolites. Covalent binding of a test chemical to DNA can be assessed by treating cells with the radioactively labelled compound and subsequently analysing them for radioactivity covalently bound to the DNA. The practical utility of this approach is, however, limited by the problem that the chemical in question has to be available in radioactively labelled form. Moreover, laborious control experiments are generally necessary to exclude experimental artefacts, such as the exchange of tritium atoms between the test chemical and the DNA when tritium-labelled compounds are used, the enzymatic incorporation of radioactive decomposition products of the test chemical into DNA and the apparent radioactive labelling of the DNA as a consequence of its contamination by radioactively labelled proteins. For these reasons, studies on the covalent binding are unsuited for routine testing. Immunological methods and the ^{32}P post-labelling technique allow very sensitive detection of DNA adducts without being dependent on radioactively labelled test compounds, but these approaches are also associated with substantial experimental effort. They are, however, particularly useful in their *in vivo* version for the detection of organ-specific genotoxicity.

Tests for differential killing

In the past, various tests were utilized in genotoxicity testing which compare the growth inhibitory effect of a test agent in two bacterial strains one of which, as a consequence of a mutation in a gene relevant for DNA repair, has a limited capacity to repair DNA damage. Different growth rates of the two strains in the presence of a test agent were interpreted as an indication that the test compound had induced DNA damage. These assays, which were performed with various species of bacteria and commonly designated "repair assay" or given specific names (e.g. "Rec assay"), frequently yielded positive results with test chemicals which did not exhibit genotoxic activity in other test systems. This was probably due to the fact that the product of the mutated gene, e.g. the recA protein, frequently has pleiotropic functions and may render cells less sensitive to certain cytotoxic effects of test agents. For this reason, among others, tests for differential killing are no longer used in genotoxicity testing. Positive results from these assays can, at the best, only support findings from other tests that yielded positive results.

Tests for the induction of the SOS response

In this type of test, the increased expression of specific stress genes ("SOS genes", specifically *sfiA*, *umuC*, *recA*) as a consequence of DNA damage or an inhibition of DNA synthesis is determined in bacteria. The increase is detected indirectly *via* the demonstration of an increased expression of a reporter gene (*lacZ*, the structural gene of the enzyme γ-galactosidase) which is quantitated colorimetrically. The tests can be performed with different bacterial species, e.g., *E. coli* („SOS chromotest") or *Salmonella typhimurium* ("Umu test", "RecA test"). The responsiveness of the tests to inhibitors of DNA synthesis can yield positive results with compounds which would not give positive results in, e.g., the bacterial gene mutation assay. Positive results from this type of tests may provide supportive evidence for positive results obtained in other test systems.

Tests for the induction of DNA strand breaks

DNA single and double strand breaks, if not repaired correctly, may result in mutation induction. DNA strand breaks can be formed directly, e.g., by attack of reactive species at the sugar phosphate backbone of the DNA. In addition, strand breaks can result from the repair of DNA damage, and certain DNA lesions called ”alkali labile sites” are converted into breaks under the alkaline conditions generally used for strand break analysis.

Alkaline conditions are employed in order to separate the DNA strands and to uncover single strand breaks. Subsequently, characteristic changes of the physicochemical properties of the DNA strands, which occur as a consequence of strand breaks, are measured. Primarily two techniques, the alkaline elution method and, more recently, the alkaline single cell gel electrophoresis method have been employed in genotoxicity testing approaches.

In the alkaline elution method, the DNA of the cells to be analysed is deposited on membrane filters and then eluted with an alkaline buffer. The rate of elution through the membrane pores depends on the length of the DNA molecules, as shortened strands of DNA resulting from strand breaks will be eluted prior to intact DNA. By comparing the elution rates of the DNA from untreated and from treated cells, it can be determined whether the test compound induced DNA strand breaks. Modifications of the experimental protocol are available which preferentially detect DNA-DNA- and DNA-protein crosslinks as well as DNA double strand breaks.

In the alkaline single cell gel electrophoresis method, also known as “Comet assay”, the cells are exposed to a test compound, embedded in agarose on a microscope slide and lysed. Subsequently, they are subjected to electrophoresis under alkaline conditions. The DNA is then stained and individual cells are analysed under the microscope. Whereas undamaged DNA does not migrate towards the anode under the test conditions applied, strand breaks and alkali labile sites cause a partial relaxation of the compact DNA structure and enable the movement of DNA fragments and loops out of the nucleus towards the anode. This movement results in the appearance of a “comet tail“ which can be quantitated and which represents a measure for the strand-break inducing capacity of the test agent. Advantages of the single cell gel electrophoresis technique, which can be applied on every cell type containing DNA, are the high sensitivity of strand break detection and the possibility to analyse individual cells. The *in vivo* version of this test has gained practical importance in genotoxicity testing with regard to the detection of organ-specific genotoxic events (see below).

The interpretation of experiments measuring the induction of DNA strand breaks by test agents is complicated by the fact that strand breaks can also be induced by agents which do not directly damage DNA. Such compounds include those affecting the function of specific proteins involved in the maintenance of DNA structure, inhibitors of DNA polymerases and DNA repair enzymes or intercalating compounds which interfere with the activity of DNA topoisomerases. Strand breaks can also occur as a consequence of unspecific cytotoxic effects, such as membrane damage, shifts of the intra- and extracellular distribution of ions, lack of energy and inhibition of protein synthesis, and during apoptosis. Thus, for the interpretation of observations from strand break experiments it is mandatory to discriminate between unspecific or indirect actions of chemicals and specific effects originating from the reactivity of the test compound or its metabolites with the DNA. This differentiation is often difficult and can constitute a major problem for the interpretation of the results from strand break measurements.

Tests for the induction of sister chromatid exchanges (SCEs)

SCEs represent the exchange of corresponding stretches of DNA molecules between the two chromatids of a chromosome during DNA synthesis. The exchange reaction is based on breaks of the DNA molecules of both chromatids at genetically identical sites with subsequent rejoining of both DNA molecules at the breakage sites during replication. Thus, the exchange is “reciprocal” and does not usually lead to alterations in DNA sequence.

To make SCEs detectable by microscopic examination of the cells, chromatids are labelled with the synthetic nucleoside 5-bromo-2'-deoxyuridine (BrdUrd). After two rounds of DNA synthesis the chromosomes consist of two chromatids which are asymmetrically substituted with BrdUrd and can thus be discriminated. SCEs become apparent by color switches between the two chromatids. Quantitative evaluation of the experiments is performed by counting the number of color switches per metaphase.

SCE tests, which are easy to perform, have been preferentially performed using Chinese hamster cell lines, such as V79 or CHO. The molecular mechanisms resulting in SCE formation are largely unclear. In general, SCEs appear to be the consequence of a perturbation of DNA synthesis. Such perturbations can be caused by the presence of unrepaired DNA lesions or by an inhibition of DNA repair. However, numerous other mechanisms which result in an inhibition of DNA replication also induce SCEs. Since an exactly "reciprocal" exchange between sister chromatids is not associated with alterations of the nucleotide sequence of the DNA, a SCE *per se* is not a mutagenic event. The exchange process between the two double stranded DNA molecules is, however, a very complex, potentially error-prone biochemical process. A compound which is able to initiate this process is, therefore, generally regarded as potentially mutagenic. This notion is supported by the observation that many chemicals which induce gene mutations or chromosomal aberrations also increase the frequency of SCEs.

The SCE test is not used any more in routine testing as the micronucleus test is also easy to perform and to evaluate and because the genetic endpoints detected with the micronucleus test, i.e., clastogenicity and aneugenicity, are better defined and of higher biological relevance.

Test for the induction of unscheduled DNA synthesis (UDS)

The *in vitro* UDS test determines the capability of a chemical to induce DNA repair in cultured mammalian cells. DNA repair is quantified by measuring DNA repair synthesis. DNA repair synthesis is a step in the process called "excision repair" which enables cells to excise DNA lesions from the damaged strand. The resulting gaps are subsequently filled by repair synthesis which can be measured *via* the incorporation of radioactively labelled thymidine. Repair synthesis is discriminated from normal replicative DNA synthesis by microscopic evaluation of single cells in autoradiographs which allows the exclusion of the heavily labelled cells in the S phase of the cell cycle from the analysis. In general, primary cultures of rat hepatocytes are employed because of their outstanding metabolic competence. Thus, this assay does not rely on the use of S9 mix for metabolic activation.

The measurement of the induction of DNA repair synthesis enables the detection of the genotoxic properties of many different types of DNA damaging chemicals. Since the biological consequences of the corresponding DNA lesions can be very different, no quantitative conclusions regarding the mutagenicity of chemicals can be drawn from the results of repair tests. However, an increased repair synthesis clearly indicates a chemical reactivity of the test compound or its metabolite(s) with the DNA. Since toxic effects not associated with genotoxicity do not result in an induction of repair, false positive results do not occur.

Cell transformation assays

In certain mammalian cells, carcinogenic chemicals induce heritable alterations in cell morphology or growth behaviour which cause these cells to resemble tumour cells. The alterations occurring as a consequence of this "transformation" include the loss of growth control, which can become manifest in various steps, such as the loss of contact inhibition, the acquired ability to grow in soft agar, or the capability for unlimited cell division (immortalisation). Thus, cell transformation appears to be a complex process having many aspects in common with the process of cancer development.

The molecular mechanisms responsible for the individual steps of cell transformation are largely unknown. It appears, however, that test systems for cell transformation do not only detect mutagens, but also "non-genotoxic carcinogens" such as certain types of tumour promoters. Therefore, cell transformation assays are not counted among the

classical genotoxicity tests and have not gained broad acceptance in routine testing, primarily because of the uncertainties regarding the molecular mechanisms of cell transformation, the experimental effort required, and their sensitivity against modifications of experimental parameters. Large data base are available for systems employing mouse fibroblast cell lines, such as C3H10T1/2, Balb/c3T3, and Syrian hamster embryo cells.

II. In vivo test systems for genotoxicity

In vivo test systems for genotoxicity are used to put positive results from *in vitro* genotoxicity assays into perspective, i.e. to determine if the capability of a test compound to induce genotoxic effects *in vitro* is realised in the whole animal with fully functional absorption, distribution, metabolism and excretion. Moreover, *in vivo* test systems are used to detect genotoxic agents that are poorly detected *in vitro*. The *in vivo* tests are generally performed on rodents, i.e. mice, rats, or hamsters, but there are also well established test systems using the fruit fly *Drosophila melanogaster*.

Mutagenicity tests

Drosophila melanogaster

Somatic cell tests

These tests detect mutations and/or mitotic recombination in embryonic somatic cells of *Drosophila*. Larvae with specific genetic constitutions with respect to certain marker genes affecting the appearance of the eyes ("white-ivory eye spot test") or the wings ("wing spot test" or "wing somatic mutation and recombination test (SMART)") are treated with the test agent by injection, feeding or inhalation. A mutated cell in the embryo will grow into a region of mutated cells among normal cells in the adult fly. After hatching of the adult flies the eyes and the wings, respectively, are microscopically examined for characteristic morphological alterations. Whereas the eye spot test specifically detects gene mutations, the wing spot test detects both various types of mutational events and mitotic recombination.

Germ cell test

The "sex-linked recessive lethal (SLRL) test" is a germ cell test which detects the induction of X-chromosomal recessive lethal mutations. For routine testing males, which contain only one X chromosome per germ cell, are exposed to a potential mutagen. In two successive crosses the treated X chromosome is transferred to males of the F2 generation. Here, males carrying the treated chromosome can be detected based on easily scorable phenotypic characteristics (eye shape and colour). The X chromosome in F2 males is hemizygous, and therefore any recessive mutation is expressed: The presence of a recessive lethal kills all these males before they have developed into a fly, and this is easily scored as the absence of one mendelian class among the progeny.

Rodents

Tests for the induction of gene mutations in somatic cells

The Mouse Spot Test is an *in vivo* mutation assay with somatic cells which detects the transplacentary induction of gene mutations, but also chromosome mutations and somatic recombinations. Mouse embryos which are heterozygous for several different recessive coat colour genes are exposed to the test agent *in utero*. Mutational and recombinational events in the embryonic melanoblasts which result in the mutation or loss of the wild type allele and the expression of the recessive marker allele in the growing cell clones lead to localised coat colour changes that can be detected as a colour spot after the birth of the mouse. The frequency of the spots in treated animals is compared with that of controls. Mitotic recombination can be differentiated from gene mutations by microscopic examination of the colour granula in the hairs. .

The spot test shows certain limitations with respect to the toxicokinetics of the test agents due to the transplacentary delivery of the compound or its metabolites. Moreover, the sensitivity of the assay is low as there are only 150-200 melanoblasts in the embryo at the time of treatment. In order to attain an acceptable sensitivity, large numbers of

animals must be used. For this reason, the Mouse Spot Test is unsuited for routine *in vivo* testing for mutagenicity.

Gene mutation tests with transgenic mice or rats generally detect mutations in a prokaryotic reporter gene in genomic DNA of specific proprietary transgenic animals. The largest data bases exist for the so-called Muta™Mouse and Big Blue® mice and rats. In the Muta™ system, the reporter gene is the *lacZ* gene, in the Big Blue® system it is the *lacI* gene. These target genes are incorporated as multiple copies at specific chromosomal locations. Using shuttle vector technology, the transgene is isolated from the organ(s) of interest of control mice and treated mice and used to transfect *E. coli* cultures for the determination of mutant frequency. Cell clones carrying gene mutations can by identified by the colour changes resulting from the lack of expression (mutated *lacZ*) or the induced expression (mutated *lacI*) of the enzyme γ-galactosidase, which can cleave the chromogenic substrate X-gal to produce a blue colour. Optionally, the mutations detected may be characterised at the molecular level. Besides *lacZ* and *lacI* other reporter genes have been transferred into the mouse genome as mutational targets, but there is still little experience with these models.

A decisive advantage of the assays compared to any other mutagenicity test is the capability to isolate the target gene from any organ or tissue which enables the assessment of both systemic and local mutagenic effects. The sensitivity of the assay is limited by a relatively high frequency of spontaneous mutations which probably arise as a consequence of the suboptimal surveillance of the silent bacterial transgene by transcription-coupled DNA repair systems of the host. Sensitivity can be increased by repeated application of the test agent over several weeks. A major drawback of the shuttle vector technology is that it is not compatible with the detection of transgenes carrying large deletions, and clastogenic events resulting in chromosomal mutations can therefore not be detected. This drawback has been overcome by the recent generation of transgenic animals which carry the target gene not in a shuttle vector, but in a plasmid vector incorporated into their genome.

Tests for the induction of chromosome aberrations in somatic cells and germ cells

The rodent micronucleus assay is the most frequently employed *in vivo* mutagenicity assay. It is usually performed with mice, but rats or hamsters may also be used. Micronuclei can only be found in cells which have undergone cell division following exposure to the test agent. Since the bone marrow is one of the continuously proliferating tissues in adult animals and as it is easy to obtain, it has been used as the primary target organ for this assay. Following treatment of the animals, polychromatic (young) erythrocytes in the bone marrow which have extruded the main nucleus during the last stage of hematopoiesis are scored for the presence of micronuclei. In order to detect potential cytotoxic effects of the treatment on the bone marrow, the ratio between polychromatic and normochromatic (mature) erythrocytes is determined. It is also possible to analyse reticulocytes in peripheral blood for the presence of micronuclei. The absence of the main nucleus in these cells facilitates scoring, but the micronucleus assay can also be performed with nuclei containing cells in other proliferating tissues, such as gut epithelial cells or early spermatids. The assay can also be conducted with hepatocytes from young animals or with hepatocytes from adult animals following stimulation of cell division by partial hepatectomy or treatment with chemical mitogens such as 4-acetylaminofluorene.

The micronucleus assay detects both clastogenic and aneugenic effects of agents. The endpoint is easy to score and provides good statistical power as many cells can be scored. Scoring of blood cells can be further improved and facilitated by the use of flow cytometry. It can be incorporated in ongoing toxicological studies, and several samples can be obtained from the same animal. For the evaluation of negative results obtained it is essential to assure that the test agent actually reached the respective organ, e.g. the bone marrow, in its active form. The specificity of the assay for detecting genotoxic carcinogens is apparently quite high, i.e. false positives are very rare. However, there is a growing body of evidence that compound-related disturbances in the physiology of the

rodents used in these assays can result in increases in micronucleated cells that are not related to an intrinsic genotoxicity of the compound under test. These disturbances include changes in the core body temperature (hypo- and hyperthermia), increases in erythropoiesis following prior toxicity to erythroblasts, such as haemolysis, the inhibition of protein synthesis or the direct stimulation of cell division in these cells. Potential contributions of these disturbing factors to micronucleus formation must be taken into consideration..

In the chromosomal aberration test with mice, rats or hamsters, the clastogenicity of the test agent is assessed by analysing the first metaphases of the cells after treatment of the animal with the test agent. To meet this requirement, animals are treated only once. A mitogenic activity of the target tissue is a prerequisite for the assay. It is increasingly displaced by the micronucleus assay because structural chromosomal aberrations are much more difficult to score than micronuclei, are scored with less statistical power, show less reproducibility between laboratories, and, most importantly, do not yield information on potential aneugenic effects.

Indicator tests

Tests for the induction of DNA strand breaks

The detection of strand break induction *in vivo* is nowadays almost exclusively achieved using the Comet assay. This assay can be performed with any organ from which single cell suspensions can be obtained, and it is particularly suited for the detection of local, organ-specific genotoxic events. Following treatment of the animals, cells are isolated from the organ(s) of interest and analysed *in vitro* for the presence of strand breaks.

As DNA strand breakage can be induced by a variety of mechanisms including both a direct interaction of the test agents or its metabolite(s) with the DNA and indirect mechanisms occurring as a consequence of cytotoxicity, it is indispensable to get information on potential cytotoxicity in the respective organ in case an induction of strand breaks is observed. The most straightforward approach is the histopathological examination of the tissue for signs of toxicity, e.g. the presence of necrotic or apoptotic cells. In case cytotoxicity can be excluded, the agent can be regarded as genotoxic in the organ or tissue studied. It is not possible, however, to draw any sound conclusion with respect to the potential clastogenicity of the agent, as double strand breaks, the cause of chromosomal aberrations, are generally not distinguished from single strand breaks.

Tests for the induction of sister chromatid exchanges (SCEs)

Mice, rats or Chinese hamsters are exposed to the test agent. To enable a discrimination of the sister chromatids, the animals are treated with 5-bromodeoxyuridine by repeated injection or *via* an implanted tablet or minipump. SCEs can be assessed in proliferating somatic cells, such as bone marrow cells, and germ cells (differentiated spermatogonia).

The SCE test does not measure chromosomal aberrations, but intrachromosomal recombinational events, and an induction of SCEs can be due to a direct interaction of the test agent or its metabolites with the DNA or to an interference with DNA synthesis mediated by non-genotoxic mechanisms. For this reason, SCE tests are hardly used any more in genotoxicity testing.

Test for the induction of unscheduled DNA synthesis (UDS)

In principle, the assay can be conducted with any organ or tissue from which preparations of single cells can be obtained. The most common version is the rat liver UDS test. In general, animals are treated with the test agent once, and after 12-16 hours hepatocytes are isolated by collagenase perfusion of the liver. They are then exposed to ^{3}H-thymidine for the *in vitro* labelling of the ongoing DNA repair synthesis and analysed for the induction of UDS by autoradiography and microscopic examination. In case of a negative test result, the treatment is usually repeated using shorter exposure times of only 2-4 hours.

The hepatic cytotoxicity of the exposure is monitored by assessing the morphological appearance of the cells. Interpretation of the results is facilitated by the fact that even strong cytotoxicity does not result in UDS induction, i.e. false positives. This feature is

responsible for the high specificity (i.e., ability to correctly identify non-genotoxic hepatocarcinogens as non-genotoxic) of the assay. As there is no information on the kind of DNA damage responsible for repair induction, it is not possible to draw quantitative conclusions regarding the mutagenic potential of the test agent.

Germ cell tests involving an examination of the progeny

The following *in vivo* tests for mutagenicity are assays in which mutations in germ cells of exposed animals are identified via the scoring of specific changes in the next generation (F1). These tests are generally very laborious and require large numbers of animals. They are, therefore, hardly used for the detection of potential genotoxic carcinogens but rather for the identification of germ cell mutagens or the assessment of the risk of humans to develop heritable genetic changes.

Tests for the induction of gene mutations

The specific locus test is a mutagenicity assay which detects heritable mutations of specific marker genes in germ cells of treated mice on the basis of changes in coat colour, eye colour or ear size in the progeny of the F1 generation. The discernable mutations are base substitutions or small deletions. The test is restricted to mice as only mouse strains with the required marker genes are available. The test is used primarily for the quantification of the genetic risk associated with exposure of humans to mutagenic agents.

Tests for the induction of chromosome aberrations

The dominant lethal test, which can be performed with mice and rats, is an indirect germ cell assay for the clastogenic activity of an agent. Males are exposed to the test agent and mated with untreated females. After fertilisation, chromosomal aberrations in sperm cells result in the death of the embryos, an event that can be recognized by the subsequent examination of the uteri of pregnant females. The dominant lethal test is used for the identification of germ cell mutagens and the characterisation of the mutagenic effects induced in specific stages of spermatogenesis.

The test for heritable translocations, which also can be performed with mice and rats, identifies a specific subset of chromosomal aberrations which are transmitted from the exposed animals to the progeny and which are compatible with the survival of the embryos. In general the males are treated for this kind of assay. The offspring with inherited chromosomal aberrations of this type exhibit reduced fertility, i.e. they are semi-sterile or sterile. Individuals of the F1 generation carrying translocations are identified by the assessment of their fertility and, preferably, by the cytogenetic confirmation of the translocation. Because of the large number of animals required, the test for heritable translocations is usually conducted only when the dominant lethal test has indicated an effect of an agent on a specific stage of spermatogenesis. Similar to the specific locus test, the assay for heritable translocations serves to quantify the genetic risk to humans associated with an exposure to the test agent and is not used in routine mutagenicity testing.