Boobis, AR, Brown, P, Cronin, MTD, Edwards, J, Galli, CL, Goodman, J, Jacobs, A, Kirkland, D, Luijten, M, Marsaux, CFM, Martin, M, Yang, C and Hollnagel, HM

Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their revaluation

Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their revaluation

Alan R Boobis[1], Paul Brown[2], Mark T.D. Cronin[3], James Edwards[4], Corrado L Galli[5], Jay Goodman[6], Abby Jacobs[2], David Kirkland[7], Mirjam Luijten[8], Cyril F M Marsaux[9], Matt Martin[10], Chihae Yang[11,12,13], Heli M Hollnagel[14]

[1]Imperial College London, London, United Kingdom

[2]US Food and Drug Administration, Silver Spring, United States of America

[3]Liverpool John Moores University, Liverpool, United Kingdom

[4]DSM Nutritional Products Ltd, Kaiseraugst, Switzerland

[5]University of Milan, Milan, Italy

[6]Michigan State University, East Lansing, United States of America

[7]Kirkland Consulting, Tadcaster, United Kingdom

[8]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

[9]ILSI Europe a.i.s.b.l., Brussels, Belgium

[10]Environmental Protection Agency, Washington DC, United States of America

[11]The Ohio State University, Columbus, United States of America

[12]Molecular Networks GmbH, Erlangen, Germany

[13]Altamira LLC, Columbus, United States of America

[14]Dow Europe GmbH, Horgen, Switzerland

Corresponding author at: ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6, BE-1200 Brussels, Belgium. Tel: +3227710014; fax: +3227620044. E-mail address: publications@ilsieurope.be (Cyril F M Marsaux).

Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their revaluation

The threshold of toxicological concern (TTC) approach is a resource-effective *de minimis*method for the safety assessment of chemicals, based on distributional analysis of the results of a large number of toxicological studies. It is being increasingly used to screen and prioritise substances with low exposure for which there is little or no toxicological information. The first step in the approach is the identification of substances that may be DNA-reactive mutagens, to which the lowest TTC value is applied. This TTC value was based on analysis of the cancer potency database and involved a number of assumptions that no longer reflect the state-of-the-science and some of which were not as transparent as they could have been. Hence, review and updating of the database is proposed, using inclusion and exclusion criteria reflecting current knowledge. A strategy for the selection of appropriate substances for TTC determination, based on consideration of weight of evidence for genotoxicity and carcinogenicity is outlined. Identification of substances that are carcinogenic by a DNA-reactive mutagenicmode of action and those that clearly act by a non-genotoxic mode of action will enable the protectiveness to be determined of both the TTC for DNA-reactive mutagenicityand that applied by default to substances that may be carcinogenic but are unlikely to be DNA-reactive mutagens (i.e. for Cramer class I-III compounds). Critical to the application of the TTC approach to substances that are likely to be DNA-reactive mutagens is the reliability of the software tools used to identify such compounds. Current methods for this task are reviewed and recommendations made for their application.

Keywords: Threshold of Toxicological Concern (TTC); Threshold of Regulation (TOR); Carcinogenicity; Genotoxicity; Mutagenicity; DNA-reactivity; Cramer classification; Structural alerts; Cancer Potency Database (CPDB)

Word count: 17,846

Table of Contents

3

4

List of abbreviations

3Rs, Reduction, Refinement and Replacement of the use of animals in toxicity testing;

AOP, Adverse outcome pathways;

AUC, Area under the curve;

BMD, Benchmark dose;

BMDL, Estimate of the 95% lower confidence limit on the BMD;

BMDL10, BMDL for a 10% response;

BMDU, Estimate of the 95% upper confidence limit on the BMD;

bw, Bodyweight;

CAR, Constitutive androstane receptor;

COC, Cohort of concern;

CPDB, Cancer potency database;

ECHA, European Chemical Agency;

EFSA, European Food Safety Authority;

EMA, European Medicines Agency;

EMS, Ethyl methanesulfonate;

ENU, Ethylnitrosourea;

EPA, Environmental Protection Agency;

FAO, Food and Agriculture Organization of the United Nations;

FDA, Food and Drug Administration;

GLP, Good laboratory practice;

HBGV, Health-based guidance value;

HIV, Human immunodeficiency virus;

HPV, High Production Volume (OECD program);

ICH, International Conference on Harmonisation;

IPCS, International Programme on Chemical Safety;

JRC, Joint Research Centre;

LTD10, Lower 95% confidence limit on TD10 (the dose associated with an extra lifetime tumour risk of 10%);

MIE, Molecular initiating event;

MNU, Methylnitrosourea;

MOA, Mode of action;

MOE, Margin of exposure;

MTD, Maximum Tolerated Dose;

MXA studies, TD50 values have been derived from more than one site, combined by NCI/NTP;

MXB studies, TD50 values have been derived from more than one site, combined by Berkeley;

NCI, National Cancer Institute;

NfG, Note for Guidance;

NIEHS, National Institute of Environmental Health Sciences

NOAEL, No Observed Adverse Effect Level;

NOEL, No Observed Effect Level;

NTP, National Toxicology Program;

OECD, Organisation for Economic Co-operation and Development;

PAH, Polycyclic aromatic hydrocarbon;

PCB, Polychlorinated biphenyl;

POD, Point of departure;

PPAR, Peroxisome proliferator-activated receptor;

PPR Panel, EFSA Panel on Plant Protection Products and their Residues;

(Q)SAR, (Quantitative) structure–activity relationship;

SCCS, Scientific Committee on Consumer Safety;

SCHER, Scientific Committee on Health and Environmental Risks;

TBA studies, TD50 values have been derived from findings observed in all tumour bearing animals;

TD50, dose that induces tumours in 50% of dosed animals over the normal lifespan of the species;

TGR, Transgenic rodent;

TOR, Threshold of Regulation;

TTC, Threshold of toxicological concern;

UDS, Unscheduled DNA synthesis;

UGT, Glucuronosyltransferase;

VSD, Virtually Safe Doses;

WHO, World Health Organisation;

WOE, Weight of evidence;

Introduction

The Threshold of Toxicological Concern (TTC) approach is being increasingly used to screen and prioritise substances with low exposure for higher tier risk assessments (e.g. see Embry et al. 2014), and is being accepted more widely by regulatory bodies. The TTC values in use have been derived from analysis of extensive databases of toxicology studies in animals from which extracted information has been used to derive quantitative thresholds of low toxicological concern. The concept is pragmatic and has evolved over a number of years. It represents a major development in the Reduction, Refinement and Replacement of the use of animals in toxicity testing (3Rs). Beyond accepting the concept in principle, it is appropriate to regularly review the quantitative thresholds that have been derived, ensure they are based on the most up to date information, reflect state-of-the-science principles, and meet quality standards, so that the quantitative values can be confirmed or, if need be, refined. A thorough review of the non-cancer TTC values has been published recently (EFSA 2016). However, other than concluding that the TTC value for genotoxic carcinogens is sufficiently protective, no such review of this value has been undertaken. Hence, this paper considers the historical development of the TTC values for substances that might be carcinogenic, and provides recommendations on reviewing and updating these values, as necessary.

Role of de minimis approaches in risk assessment

There are a large number of chemicals in use for which there is little or no toxicological information. This is exacerbated by the ever increasing sensitivity and versatility of analytical chemistry to detect the presence of natural and manmade substances in our environment. There is therefore a need for methods that enable the rapid and cost-effective screening of such chemicals to enable their prioritization for further

assessment. Equally, when regulating chemical products such as food contact materials, drug impurities, food additives, industrial chemicals and consumer products, it is important that authorities and industry make the most efficient use of time and resources to focus on those exposures that are of potential concern while giving lower priority to those considered to be of negligible concern. This can be achieved by defining *de minimis* exposure levels representing negligible risk of adverse effects to human health.

One science-based approach for setting *de minimis* thresholds is to collect toxicological data on a large group of substances and to apply this knowledge to untested substances. Frawley (1967) was one of the first to describe such an approach. He analysed a large number of chronic oral toxicity studies to derive quantitative *de minimis* thresholds for dietary exposures to substances migrating from food contact materials.

Brief history of existing TTC scheme

Following up the work of Fawley, Rulis (1987) concluded that carcinogenicity was the most sensitive endpoint from exposure to chemicals (this was on the assumption of a linear non-threshold dose-response relationship for such substances) and proposed the use of analysis of the carcinogenic potencies of 343 (subsequently updated to 709 by (Cheeseman et al. 1999) substances from over 3500 experiments in the Carcinogenic Potency Database (CPDB) (Gold et al. 1984; Gold et al. 1989; Gold et al. 1995; European Chemicals Agency (ECHA) 2015) to derive a Threshold of Regulation (TOR) for new food contact substances migrating into food at only minute concentrations, and with no structural alerts for DNA-reactive mutagenicity. Following further evaluation (Munro 1990) the US Federal and Drug Administration (FDA) introduced a TOR of 0.5 ppb for such food contact materials (Federal Register 1995). Using the US FDA default values for combined food and drink consumption of 3 kg per day per person, and 60 kg

body weight (bw), the level of 0.5 ppb corresponds to a daily exposure of 0.025 µg/kg bw/day or 1.5 µg/person (see section on “FDA TOR and additional thresholds developed from it”further details). Based on the analyses by Rulis, Munro and others, this exposure was concluded to present negligible concern to public health, even if the substance in question were later identified to be a carcinogen. For new substances migrating below 0.5 ppb into packaged food and for which there are no concerns for DNA-reactive mutagenicity, US FDA requires no specific toxicity testing and performs an abbreviated safety assessment focussed mainly on intake assessment. More detail on the derivation of the carcinogen thresholds is presented in later sections of this document. In view of the fact that only a fraction of all natural and manmade substances possess structural alerts for DNA-reactive mutagenicityand are carcinogenic, and that these could be identified reasonably well based on structural features, attention turned to whether suitable*de minimis* thresholds for repeat dose toxicity other than carcinogenesis could be developed.

Munro and co-workers evaluated the use of TTC for toxicity other than carcinogenicity (613 substances) (Munro et al. 1996). Information from the most sensitive species, sex, and toxicological endpoints was recorded to identify the most conservative No Observed Effect Level (NOEL)[1] value for each substance. Structural information based on an algorithm developed by Cramer and colleagues was used to broadly group the chemicals (Cramer et al. 1978). This algorithm grouped chemicals into three structural classes based on a ‘decision tree’ approach that consists of 33 questions each of which is answered by ‘yes’ or ‘no’. Each answer leads to another question or to a final classification into one of the threeclasses of expected low, medium

1. Whilst Munro et al referred to the points of departure used as NOELs, in fact these were mostly also No Observed Adverse Effect Levels (NOAELs).

or high toxicity. Further details on the Cramer classification scheme can be found in European Food Safety Authority (EFSA,2016). This part of the TTC approach is for substances that are non-genotoxic, based on the 5th percentile of NO(A)EL frequency distributions,and assumes a lifetime exposure.

Outline of current TTC scheme

Kroes et al. (2004) undertook further review of the TTC concept with regard to potentially sensitive endpoints (i.e. for which potency would not be covered by point of departures (PODs) for systemic toxicity) such as developmental toxicity and neurotoxicity, and combined the cancer and non-cancer aspects into a tiered approach. They noted that the threshold of 0.025 µg/kg bw/day (1.5 µg/day) resulting from the FDA analysis for its TOR was applicable only to substances for which there was an absence of any indication of DNA-reactive mutagenic potential. In their analysis, Kroes et al. (2004) concluded that the appropriate value for compounds that were potentially DNA-reactive mutagens would need to be 10-fold lower, and derived a TTC for such compounds of 0.0025 µg/kg bw/day (0.15 µg/day) (see section on "Background to cancer TTC values"for details). The introduction of the 0.0025 µg/kg bw/day threshold for potentially DNA-reactive mutagenicsubstances rendered the 0.025 µg/kg bw/day tier originating from the FDA TOR development work redundant within a tiered approach to TTC. This is depicted in the tiered decision scheme as, for example, laid out by World Health Organisation (WHO)/EFSA (EFSA 2016) (**Figure 1**). However, it is perhaps worth noting that a higher threshold (1.5 µg/person per day) is used for pharmaceutical impurities that are mutagenic or that are suspected of being mutagenic (see section on "Thresholds for impurities in pharmaceuticals" below).

The TTC values derived are not applicable to chemicals belonging to a number of substance classes (referred to "exclusionary categories" in **Figure 1**) . Inorganic

chemicals, metals and organometallics, organo-silicon compounds, nanomaterials, radioactive substances, proteins and high molecular weight or poorly characterised chemicals such as polymers were excluded since they were not represented in the toxicity databases. Aflatoxin-like, azoxy and N-nitroso-compounds were excluded as highly potent genotoxic carcinogens, along with polyhalogenateddibenzo-p-dioxins, -dibenzofurans and dioxin like polychlorinated biphenyls (PCBs, potent carcinogens, bioaccumulative, with very large kinetic differences between experimental animals and humans); and steroids (potent carcinogens and lack of agreement on dose thresholds), as explained in more detail in section on “FDA TOR and additional thresholds developed from it”. Chemicals with these structural characteristics can be easily identified by, for example, ChemoTyper using the ToxPrints and the TTC category chemotypes (https://toxprint.org/).

The TTC approach is currently in use to evaluate migrant substances from packaging materials (Federal Register 1995), flavourings substances in food (Federal Register 1993; Joint Food and Agriculture Organization of the United Nations (FAO)/WHO Expert Committee on Food Additives 1996; Munro et al. 1999; WHO 1999; European Commission 2002) non-relevant plant protection product metabolites in ground water (European Commission 2003), mutagenic impurities in pharmaceutical preparations (International Conference on Harmonisation (ICH) 2014), and genotoxic constituents in herbal preparations (European Medicines Agency (EMA) 2008), albeit with slight variations in its application. The TTC concept has been acknowledged to be a science-based prioritisation and risk assessment tool by different organisations such as WHO International Programme on Chemical Safety (IPCS), EFSA, the Scientific Committees of the EU Commission and Health Canada (Joint FAO/WHO Expert Committee on Food Additives 1996; EFSA 2012; Scientific Committee on Consumer

Safety (SCCS), Scientific Committee on Health and Environmental Risks (SCHER) 2012; EFSA 2016; SCCS Note for Guidance (NfG) 2016; Health Canada, 2016).

Use of the TTC concept for chemicals with specific data requirements for their regulatory approval is currently not acceptable as an alternative to a chemical-specific evaluation by regulatory authorities (EFSA 2016), although future application in the prioritisation or first tier assessment of such chemicals is possible, as regulatory authorities explore means to streamline the assessment process.

Background to cancer TTC values

Definitions

Mutagenicity is defined as the induction of permanent transmissible changes in the amount or structure of the genetic material of cells or organisms (ECHA 2015). This includes gene mutations, and structural and numerical changes in chromosomes. Genotoxicity is a broader term and refers to the capability of substances to damage DNA, for example covalent binding to DNA nucleobases, and/or cellular components regulating the fidelity of the genome – such as the spindle apparatus, topoisomerases, DNA repair systems and DNA polymerases – and includes all adverse effects, direct or indirect, on genetic information. Genotoxicity is not necessarily associated with mutagenicity. Alterations to the genetic material of cells may occur spontaneously endogenously or be induced as a result of exposure to ionising or ultraviolet radiation, or genotoxic substances. In principle, human exposure to substances that are mutagens may result in increased frequencies of mutations above background. DNA-reactive mutagens are substances that can react with and modify DNA directly, either as the parent chemical or in the form of a metabolite. Chemicals are defined as carcinogenic if they induce tumours, increase tumour incidence and/or malignancy or shorten the time

to tumour occurrence, relative to controls. Genotoxic chemicals that can cause canceras a result of their genotoxicity are typically referred to as genotoxic carcinogens (see section on “Assessment of dataset entries” below for further discussion). More specifically, where this is a consequence of DNA-reactive mutagenicity, such compounds have been referred to as DNA-reactive mutagenic carcinogens in the present paper. Non-genotoxic carcinogens exert their carcinogenic effects through other mechanisms that do not involve direct alterations in DNA, i.e. direct mutation(Adler et al. 2011). Chemicals can induce cancer by any route of exposure (e.g., when inhaled, ingested, applied to the skin or injected), but carcinogenic potential and potency may depend on the conditions of exposure (e.g., route, level, pattern and duration of exposure) (ECHA 2015).

FDA TOR and additional thresholds developed from it

On the basis that carcinogenicity by genotoxic substances represents the endpoint of greatest sensitivity and relevance at low dietary exposures of chemicals (assuming a linear non-threshold dose-response relationship),Rulis (1987) subjected a suitable subset of bioassay data from the CPDB (Gold et al. 1984; Gold et al. 1989; Gold et al. 1995) to probabilistic analysis. The objective was to determine a threshold level of dietary exposure that provides adequate protection from presumptive carcinogenic risk in excess of a specified target risk, even in the event that a substance was later discovered to be a carcinogen. The aim was to prioritise new indirect food additives, i.e. migrating food contact material components, based on exposure, for further toxicological testing. Rulis (1987) used animal dose-response data from 343 carcinogens tested by the oral route to derive TD50 values for the substances (the dose that induces tumours in 50% of dosed animals over the normal lifespan of the species). Corresponding ‘Virtually Safe Doses’ (VSDs) for 1 in one million lifetime tumour risk level in exposed individuals,

which served as the target risk, were calculated from these values. The resulting distribution of VSDs describes the relative probability that a randomly selected carcinogen would present an excess tumour risk greater than $1x10^{-6}$ over a lifetime. Rulis (1987) did not present specific numbers from the distribution, only that at a human dietary intake of ca. 0.025 µg/kg bw/day (1.5 µg/day based on 60 kg default body weight), about half of all carcinogens would exceed the 1×10^{-6}excess risk level. Below a dietary intake of 0.0025 µg/kg bw/day (corresponding to combined food and drink consumption of 3 kg per day containing 0.050 µg/kg food or drink), about 85% of the carcinogens resulted in less than 1×10^{-6}excess lifetime risk, i.e. 0.0025 µg/kg bw/day roughly represents the 15^{th} percentile of the VSD distribution.

Rulis (1987) also introduced the consideration that these risks were calculated on the unrealistic assumption that all substances were carcinogens and discusses that under the still conservative assumption that 20% of all new food contact substances were carcinogens, the 0.0025 µg/kg bw/day exposure level would result in less than 1×10^{-6}excess lifetime risk in 97% of cases. Rulis(1987) further concluded that it would be reasonable to assume that a new substance used in packaging materials would contact, at most, 5% of the diet. Hence, a dietary concentration of 50 ppt would correspond to a migration of 1 ppb from packaging material into food.

The issue of *de minimis* thresholds for carcinogens was further elaborated at a workshop organised by the Canadian Centre for Toxicology (Munro 1990), which addressed potency data for four different subsets of carcinogens from the CPDB analysed both parametrically and non-parametrically. The probability that an excess lifetime cancer target risk of 1×10^{-6} or 1×10^{-5} is not exceeded at various exposure levels was presented, as a function of the percentage of substances presumed to be carcinogens. Three of the four data subsets, one of which was that reported by Rulis

(1987), resulted in very similar VSD 15th percentiles, whereas one subset of National Toxicology Program (NTP)-conducted studies resulted in approximately four-fold greater values.

Overall, the type of data subset and the statistical approach to low dose extrapolation were reported to be significant factors in determining the risk probability. However, the choice of percentage of new substances assumed to be carcinogens had the most impact on a given risk level associated with different dietary exposures. For example, at a threshold of 0.5 ppb dietary concentration (0.025 µg/kg bw/day intake), the probability of not exceeding a 1 x 10^{-6}excess risk is 63% if one assumes that all new substances are carcinogens, and increases to 96% if one assumes that only 10% of new substances are carcinogens. This issue was discussed extensively at the workshop, without resolution. It was noted that of 260 substances tested for carcinogenicity by the US National Cancer Institute (NCI)/NTP program (Ashby and Tennant 1991; Fung et al. 1995), many of which are also listed in the CPDB, roughly 50% acted as carcinogens: it was pointed out, however, that these databases are impacted by a large selection bias because those substances tested were predominantly already suspected of being carcinogens, based on structural features or toxicological effects. Indeed, Fung et al. (1995) reported in their analysis that whilst of 400 substances tested in cancer bioassays by NCI/NTP, 210 were positive, of these, 181 had been selected for the testing program based on suspected carcinogenicity, while only 29 positive substances (7% of 400) had been chosen due to high exposure potential. Another confounding factor, which was not discussed by the above authors, is that many of the studies used doses greatly exceeding the Maximum Tolerated Dose (MTD) which probably meant that some of the tumour findings were secondary to other toxicity – see further discussion in sectionon“General criteria for the Cancer Potency database”.

Munro (1990) also reported that the workshop participants discussed the application of structural alerts and genotoxicity data to distinguish between DNA-reactive carcinogens with potential low-dose effects and non-genotoxic carcinogens demonstrating thresholds. It was suggested that it might be reasonable to set thresholds for non-genotoxic carcinogens based on classical toxicological criteria rather than on low-dose extrapolation of tumour data. This was however not addressed in the analyses, which treated all carcinogens with the same approach of linear extrapolation from the TD50 to below the VSD.

The FDA based the TOR of 0.5 ppb (corresponding to a default daily intake of 0.025 µg/kg bw/day based on 60 kg body weight and combined food and drink consumption of 3 kg per day) finally adopted on the need to provide a reasonable balance between necessary conservatism and practical utility, including analytical sensitivity. The TOR applies only to substances used in food-contact articles (Federal Register 1995) and defines a migration level below which new food contact substances qualify for an abbreviated approval process, provided that no concern exists for low dose toxicity,from DNA reactive mutagenicity, based on chemical structure or other information.

The possibility of refining the TOR approach by taking into account information on mode of action (MOA) for carcinogenicity was then investigated by Cheeseman et al. (1999), a possibility that had already been touched upon by Munro (1990). Cheeseman et al. (1999) analysed an expanded database of 709 carcinogens tested by the oral route. This analysis showed that Ames test negative carcinogens (193 out of 709) were 8-fold less potent than Ames test positive carcinogens (249 out of 709), even if a linear non-threshold dose-response was assumed, regardless of MOA. Based on this analysis it was concluded that substances that were negative in the Ames test and for

structural alerts for DNA-reactive genotoxicity (i.e. mutagenicity)might qualify for a higher dietary threshold, of around 0.25 µg/kg bw/day. Cheeseman et al. (1999) also identified structural classes associated with high potency for carcinogenicity, and on this basis recommended exclusion, i.e. N-nitroso-compounds, benzidine-structures and three other structural classes, from the TOR approach, even at the conservative value of 0.025 µg/kg bw/day.

Kroes et al. (2004) further analysed highly potent genotoxic carcinogens exceeding a 1 x 10^{-6}excess risk below 0.0025 µg/kg bw/day exposure, using a slightly expanded set of 730 carcinogens. This enabled identification of structural classes which should be excluded from the application of the TTC approach and established that otherwise an exposure threshold of 0.0025 µg/kg bw/day was adequately protective even for those substances with structures raising alerts for DNA-reactive genotoxicity (i.e. mutagenicity). Kroes et al. (Kroes et al. 2004) classified carcinogens based on the structural alerts described by Ashby and Tennant (1991) and Cheeseman et al. (1999), and some of the latter structural groupings were modified. Those structural groups identified to comprise a significant number of carcinogens exceeding 1 x 10^{-6}excess risk were aflatoxin-like compounds, N-nitroso-compounds and azoxy-compounds, as well as dioxins and steroids. The latter were however judged to represent non-genotoxic carcinogens with dose thresholds, so that only aflatoxin-like compounds, N-nitroso-compounds and azoxy-compounds were proposed to be excluded from the TTC concept on the basis of their high potency as genotoxic carcinogens, the so-called cohort of concern (COC).

EFSA (2012) discussed the fact that despite exclusion of the COC structural classes of Kroes et al. (2004), the carcinogen database of 730 substances still contained up to 30 substances exceeding the target excess risk of 1 x 10^{-6}, representing about 4%

of the database. However, this includes 2 dioxins and 5 steroids, inappropriately assessed by linear extrapolation, as they causecancer by non-genotoxic modes of action, for which there is evidence of biological thresholds, and the other substances represent only small fractions of their respective structural groups (3-14%). EFSA (2012) also noted that there is only a low probability (possibly as low as 10%) that any new substance would be carcinogenic and an even lower probability that this would be of high potency. It was further mentioned that EFSA typically assesses the level of concern from unavoidable exposure to genotoxic carcinogens using the margin of exposure (MOE), i.e. the ratio of the benchmark dose (BMD) lower confidence limit for a 10% response ($BMDL_{10}$) for a tumour response in a bioassay to measured or estimated human exposure. Assuming linear low dose extrapolation, this MOE is 'equivalent' to a $1x10^{-5}$excess risk and hence, a TTC of 0.0025 µg/kg bw per day would be more conservative than the accepted approach for substances specifically identified as genotoxic carcinogens. EFSA (2012) concluded that, after exclusion on the COC, for substances with structural alerts for DNA-reactive genotoxicity (i.e. mutagenicity), there is a very low probability (between 0-4%) of any appreciable cancer risk to humans from exposures below a TTC value of 0.0025 µg/kg bw per day. This value is equivalent to the TOR initially proposed by Rulis (1987).

To date, although the anticipated difference in potency and dose response between genotoxic and non-genotoxic carcinogens has been discussed by most authors addressing the TOR and TTC concepts, this aspect has not been included in their analyses such that it would enable derivation of quantitative thresholds for separate distributions of the respective groups of carcinogens.

Thresholds for impurities in pharmaceuticals

While the necessity to assess the multitude of chemicals used in food contact

applications or as food flavourings led to the development of the FDA TOR and the WHO/EFSA TTC concepts, respectively, it can equally be applied to other situations where there is a similar need. A pragmatic application of the TTC concept can be found in the control of mutagenic impurities in human pharmaceuticals. A recent harmonised guideline, ICH M7 (ICH 2014), provides a practical framework to limit potential carcinogenic risk from such impurities.

The focus of the ICH M7 guideline is on DNA-reactive mutagenic substances. For those impurities which are, or are suspected of being, DNA-reactive mutagens, a TTC approach is used. This includes the results of structure-based approaches to predict the outcome of bacterial mutagenicity assays. Compounds in the highly potent structural classes comprising the cohort of concern are excluded from the TTC approach. Otherwise, for compounds with structural alerts for DNA-reactive mutagenicity, or with other evidence of DNA-reactive mutagenicity , exposure limits are based on the respective TTC values.

The guideline notes that acceptable risk during the early drug development phase is set at a theoretically calculated level of approximately one additional cancer per million lifetime risk. However, risk is considered to be a function of both magnitude and duration of exposure and acceptable levels are adjusted accordingly, with shorter exposure durations during clinical trials allowing higher acceptable intake (i.e. TTC values adjusted upwards). For later stages in development and for marketed products, an increased cancer risk is considered acceptable due to the therapeutic benefits of the drug, and is set at a theoretically calculated level of approximately one additional cancer in one hundred thousand lifetime risk, corresponding to a TTC value of 1.5 µg/person per day (note that ICH explicitly expresses TTC values on a per person basis). This exemplifies that policy decisions on the risk acceptance levels can lead to different

thresholds being applied. Regardless of such choices, a large, robust database, analysed in a consistent manner, should form the basis of threshold derivation.

Underlying approach used to date to develop a TTC for substances that might be genotoxic carcinogens

The TTC analyses to date have largely applied the methodology used by Gold et al. (Gold et al. 1989) for interpretation of cancer bioassay data. The TD50 was used to characterise cancer potency, and all substances were considered positive for carcinogenicity where at least one sex of one species showed a response with statistical significance of $p \leq 0.01$. This approach is simple and pragmatic and allows derivation of conservative *de minimis* levels of exposure, but takes no account of a number of important aspects such as the shape of the dose-response curve and the human relevance or otherwise of rodent tumour findings. In many substance-specific risk assessments for carcinogenicity, increased understanding of modes of action and dose-response have led to criteria and methodologies different from those used to analyse the CPDB being developed and applied for the derivation of acceptable exposure levels. The question therefore arises of whether any of this knowledge or the approaches used can be applied to the derivation of TTC thresholds for such compounds.

Factors potentially influencing the TTC values for DNA-reactive mutagenic and/or carcinogenic compounds

Implicit assumptions in current approach

The current, lowest TTC value of 0.0025 µg/kg bw per day was developed to protect against substances for which there is concern that they might be carcinogenic via a genotoxic MOA (i.e. DNA-reactive mutagenicity), assuming a linear, non-threshold dose-response relationship and an acceptable excess risk of 1×10^{-6}. However, unlike

other TTC values, this value was not obtained directly from distributional analysis, but included a number of other considerations (see previous sections). Amongst assumptions that have been made in the development of this TTC value are 1) the compounds in the CPDB are representative of the world of carcinogens, 2) there are sufficient compounds to obtain a reliable estimate of the distribution of their PODs, 3) carcinogens have a linear dose-response curve, 4) the carcinogenic response observed in laboratory animals is relevant to humans, 5) the acceptable excess risk for such compounds is 1×10^{-6}. In the application of this TTC value it is assumed that 1) it is possible to identify reliably carcinogenic compounds acting via a genotoxic MOA (i.e. DNA-reactive mutagenicity, 2) those compounds that are not so identified are adequately covered by existing, higher TTC values (which would usually be 1.5 µg/kg bw per day, i.e. the TTC for Cramer class III compounds, or higher).

Assumption that non-genotoxic carcinogens are covered by higher, existing TTC values

Implicit in the current approaches to application of TTC values for genotoxic carcinogens is that if a compound is not predicted (or shown to be) a DNA-reactive mutagen, it will be adequately covered by higher TTC values, even if it is carcinogenic by some other MOA. In practice, this means that the acceptable exposure to such a compound would be up to 1.5 µg/kg bw per day, or higher depending on the Cramer class. There is certainly evidence that non-DNA reactive carcinogens are less potent that DNA-reactive carcinogens (Cheeseman et al. 1999). In addition, as such compounds are carcinogenic secondary to an adverse effect on the morphology or function of a target tissue (which is not necessarily the tissue in which carcinogenesis is observed), the application of default threshold-dependent considerations would apply. In practice, this means that if the human population is protected against the primary toxicity (which will

occur at or below the critical POD on which the health-based guidance value is based (HBGV)), they should be protected against the carcinogenic effect. Typically, the HBGV would be obtained by applying an uncertainty factor of 100 to the POD for the critical effect, such as the $BMDL_{10}$or NOAEL.

The TTC value of 0.0025 µg/kg bw per day was developed to provide a nominal level of protection of 1×10^{-6}overbackground. This was calculated assuming a linear, non-threshold dose-response. This is equivalent to dividing the $BMDL_{10}$ by 10^5. Given that for a thresholded response the $BMDL_{10}$would normally be divided by 10^2, the equivalent protection for non-DNA reactive carcinogens would be provided by a value of 2.5 µg/kg bw per day. Hence, the TTC value of 1.5 µg/kg bw per day for Cramer class III compounds (which is the likely default for most such compounds) would be adequately protective of such carcinogens, even if they were of similar potency to DNA-reactive genotoxic carcinogens. However, this analysis is predicated on the above assumptions, the implications of which would need to be determined before having confidence in this approach.

Database used to establish the TTC of 0.15 µg/day is many years old and has not been updated for several years

The database on which the TTC value for genotoxic (DNA-reactive mutagenic) carcinogens was based is several decades old, and whilst it has been updated since the original analysis, it has been 'frozen' for almost a decade. This is certainly a potential criticism. However, it should be noted that when the database was updated from 477 carcinogens by approx. 50%, to 709 carcinogens, this did not significantly alter either the range of potencies or the peak position for the distribution (Cheeseman et al, 1999). In terms of assessing DNA-reactive mutagenic compounds, the static nature of the database may be of less concern than it first appears, for at least two key reasons. The

chemicals first evaluated for carcinogenicity by the NCI and then the NTP were those that were of most concern in terms of exposure and potential potency (Munro 1990). As understanding of the structural characteristics leading to (DNA-reactive) mutagenicity increased, those companies developing novel chemicals in all sectors (e.g. agrochemicals, human medicines, food additives, industrial chemicals) introduced structural analysis and in vitro testing for mutagenicity at a very early stage of design and development. Hence, very few DNA-reactive genotoxic (DNA-reactive mutagenic) carcinogens are identified among new compounds requiring approval prior to use. In practice, this means that whilst it would be valuable to expand the CPDB, for the present exercise it is unlikely that very many suitable new compounds would be available.

Advances in understanding both the implications of genotoxicity for carcinogenicity and in mechanisms of carcinogenesis

Since the original publications on the TOR and TTC some two to three decades ago, there have been considerable advances not only in understanding modes of action for chemical carcinogenesis but also in the acceptability of such information in informing regulatory risk assessment. Hence, where a carcinogenic response has been shown to occur secondary to some other toxicological effect, it is accepted that this has a threshold and that often the application of conventional uncertainty factors to the critical effect in establishing HBGVs would be sufficiently protective against such a carcinogenic response. This also extends to consideration of the MOA for genotoxicity/mutagenicity. Where there is good toxicokinetic or, more often, toxicodynamic evidence for the existence of a threshold, conventional approaches to establishing HBGVs may be appropriately protective. Examples of the former are hydroquinone and phenol, which are rapidly detoxified in vivo. Examples of the latter

are topoisomerase inhibitors, tubulin inhibitors, and inducers of reactive oxygen species (Parry et al. 1994; Committee on Mutagency of Chemicals in Food, Consumer Products and the Environment 2010). Indeed, recently evidence has emerged that even some DNA-reactive genotoxins such as ethyl methanesulfonate (EMS) have a clear threshold for mutagenicity (Doak et al. 2007; Gocke and Müller 2009).

In 2007, it was discovered that a manufacturing batch of the drug Viracept (nelfinavir), an human immunodeficiency virus (HIV) protease inhibitor, was contaminated with EMS at levels that exceeded permissible levels by more than 1000-fold (Muller and Singer, 2009). EMS is an established DNA-reactive mutagen, rodent carcinogen and teratogen (reviewed inGocke et al. 2009b). In the absence of human data, EMS exposure limits were based on the generic TTC-derived limit of 1.5 µg/person per day as recommended in ICH M7 (discussed above). However, Doak et al. (2007) had provided reliable evidence for a threshold for chromosomal damage and gene mutations induced by EMS in vitro. Higher concentrations of EMS were clearly mutagenic, and other alkylating agents used (ethylnitrosourea (ENU) and methylnitrosourea (MNU)) exhibited apparently linear dose-effect relationshipsfor mutagenicity in these studies. The DNA alkylations induced by EMS, mainly at N7 and O6 positions of guanine, were considered to be repairable error-free by Base Excision Repair and Methyl Guanine Methyl Transferase. Additional testing in vitro and in vivo in a suitable range of assays confirmed that the mutagenicity of EMS exhibited a threshold (Gocke et al. 2009a). An EMS-specific permissible daily exposure of 100 µg/day, based on the NOEL for induction of mutations in vivo and highly conservative safety factors (amounting to >10,000), for lifetime exposure to EMS was determined. This is clearly above the ICH TTC for DNA-reactive mutagenic carcinogens of 1.5 µg/person per day (and the EFSA/WHO TTC value of 0.15 µg/person per day). Indeed,

it is above the TTC for Cramer class III chemicals, of 60 µg/person per day.It should perhaps be noted that it is not yet clear how general is the acceptability of such evidence for thresholded mutagenicity and further work on this is ongoing.

The CPDB of necessity comprises information obtained primarily on rodent cancer bioassays. Whilst, in general it is assumed, in the absence of evidence to the contrary, that the carcinogenicity of compounds acting by a DNA-reactive MOA will be relevant to humans, for those acting by other modes of action this is often not the case. A number of carcinogenic effects in rodents are now considered not to be relevant to humans, such as thyroid follicular cell tumours in rats resulting as a consequence of induced clearance of thyroid hormones by glucuronosyltransferase (UGT) (Cohen et al. 2004), renal tumours in male rats resulting from binding to alpha$_{2U}$-globulin (Meek et al. 2003), rodent liver tumours arising from activation of constitutive androstane receptor (CAR) or peroxisome proliferator-activated receptor (PPAR)-alpha(Klaunig et al. 2003; Corton et al. 2014; Elcombe et al. 2014). Those compounds that are carcinogenic by a MOA not relevant to humans should be excluded from the specific dataset used to determine an appropriate TTC value for genotoxic (i.e. DNA-reactive mutagenic) carcinogens, although in practice as the number of such compounds is small, this is unlikelyto impact on the distribution to any appreciable extent.

Even for those compounds acting by a DNA-reactive mutagenic MOA, there are often quantitative differences, as the ultimate carcinogen is usually a reactive metabolite and the enzymes involved are frequently less active in humans than in rodents (DeKeyser et al. 2011). However, the same may be true for detoxifying reactions, so the net species difference would need to be determined on a case-by-case basis.

Advances in dose-response modelling and recommendation for use of BMD approach

The analyses on which the TTC value for genotoxic (DNA-reactive mutagenic) carcinogens is based used the TD50 as the POD for the carcinogenic response. The reliability of this as an estimate of carcinogenic potency has been questioned by a number of authors and organisations (e.g. EFSA 2009). As an alternative, the $BMDL_{10}$ is recommended (e.g. EFSA 2017). This takes better account of experimental variability in the bioassay and unnecessary extrapolation outside the experimental range of observation. The CPDB was updated to include estimates of the LTD_{10}, the definition of which is analogous to that of the $BMDL_{10}$. However, the estimates provided in the CPDB were derived mathematically from the existing TD50 values. The extent to which these vary from those obtained by *de novo* benchmark dose modelling is not known. It may be that they are acceptable surrogates, but this should be determined for an appropriate sub-set of compounds before accepting the values for further analyses. A judgement would then have to be made as to whether the improvement in accuracy would be justified by the effort necessary to achieve this. EFSA (2016) concluded that such effort would not be warranted in the case of non-cancer endpoints.

Need to re-assess TTC values for genotoxic and carcinogenic compounds

From the foregoing, it is apparent that the current TTC values for substances assumed to be DNA-reactive mutagens and carcinogens could be questioned, for a number of reasons. These values are likely to be conservative, as recognised by EFSA and WHO, who concluded that their re-assessment was not considered a priority (EFSA 2016). Nevertheless, the value of such a re-assessment was noted, in that it could enhance the power and range of chemical structures covered and that a number of the assumptions used in their derivation may be unnecessarily conservative (EFSA 2016).

From a scientific perspective there is a clear case to be made for the re-assessment of these values. Whilst existing values appear to be adequately protective of human health, a number of assumptions and approaches used in their original derivation have been superseded by advances in knowledge. Hence, to ensure a robust, transparent basis for all aspects of the TTC approach, re-assessment of these values is timely. In addition, restrictions on the use of chemicals should be commensurate with their risk, which may not be the case with the application of the current values.

Outline of proposed re-evaluation of TTC values – proposed hierarchical analyses

Genotoxic carcinogens

Genotoxic chemicals are those that are capable of interacting with DNA (adduct formation, for example) directly or following formation of a DNA-reactive metabolite (Weisburger and Williams 1981) or of affecting the number or structure of chromosomes. Chemicals that are not capable of such effects are described as being non-genotoxic. However, the most potent genotoxic carcinogens are those that are DNA-reactive, leading to gene mutation. As these are the most readily identified by computational determination of structural alerts, and hence would be amenable to the application of the TTC approach, the emphasis in any re-evaluation should be on clearly identifying such compounds in the CPDB. There are a number of *in vitro* (e.g., Ames test, mammalian micronucleus assay, mammalian chromosome aberration assay, mouse lymphoma *tk* mutation assay and *hprt* mutation assay, described in OECD Test Guidelines 471, 487, 473, 490 and 476 respectively) and *invivo* (e.g., mammalian chromosome aberration, mammalian micronucleus, transgenic rodent (TGR) mutation, comet assay, and *Pig-a* mutation assay described in OECD Test Guidelines 475, 474,

488, 498 and in Dertinger & Heflich, 2011 respectively) methods that can be employed to evaluate the genotoxic potential of a chemical. Most of these (e.g., micronucleus, *tk*, *hprt*, comet and *Pig-a* assays) provide an indirect measure of DNA damage, which is presumed, and interpreted as, being indicative of a capacity to cause mutations. When interpreting assays to evaluate genotoxicity, it is imperative that the impact of cytotoxicity, which can lead to false positive results, be considered, particularly for assays in vivo and with mammalian cellsin vitro. The objective of the evaluation of chemicals for their genotoxic potential is to gain knowledge regarding their potential ability to cause mutations *invivo*, under realistic conditions of exposure. Furthermore, *invivo* data are often more relevant for extrapolation to the consequences of human exposures than *in vitro* data.

It has often been assumed that if a carcinogenic chemical is positive in a bacterial mutation (Ames) test then it can be classified as a genotoxic (DNA-reactive mutagenic) carcinogen. On a mechanistic basis, mutagenicity is not carcinogenicity, and the induction of a mutation is only one of a number of obligate steps in the progression of a normal cell and tissue to malignancy. Kirkland et al. (2006; 2014)have stated that 20-30% of chemicals that are positive in the Ames test are not carcinogens. Bacteria and mammalian tissues have different structures, metabolism and defence mechanisms. ‘S9’ (postmitochondrial supernatant) from the liver of Aroclor 1254-treated rats,used as an exogenous metabolic system, might increase the likelihood of detecting mutagens requiringmetabolic activation, it is not very representative of the constitutive enzymes of xenobiotic metabolism present in mammalian cells. Hence, although a positive result in the Ames test is considered highly important in identifying a potential genotoxic (DNA-reactive mutagenic) carcinogen, a carcinogenic chemical that is also mutagenic/genotoxic in a number of different test systems, particularly in vivo, and

across different endpoints (chromosome or DNA damage as well as gene mutation) is considered more likely to exert its carcinogenic activity via a mutagenicMOA. In vivo genotoxicity in the target species and target tissue for carcinogenicity provides considerable weight in determining whether a compound acts by a DNA-reactive mutagenic MOA. It is proposed that the most likely candidates to be considered genotoxic (DNA-reactive mutagenic) carcinogens, and therefore priority chemicals for determining TTCs, are those described in **Table 1**. The strength of the genotoxicity profile is highest in Group 1, where positive results are obtained for more than one endpoint (gene mutation, chromosomal or DNA damage) both in vitro and in vivo. Chemicals with the weakest profile show positive results in only a single test in addition to being positive in the Ames test. If chemicals show mixed positive and negative results, or only in vitro results are available, they may still be genotoxic carcinogens but the evidence linking the positive genotoxicity results to tumour induction is less strong. Given that the *in vivo* liver unscheduled DNA synthesis (UDS) test has been shown to be insensitive to a number of carcinogens suspected of operating via a DNA-reactive mutagenic MOA (see Kirkland and Speit, 2008), if a carcinogen is negative in the UDS test but gives positive results in other in vivo tests it could be included for consideration. Thus, positive results in the UDS assay are considered a relevant indicator of carcinogenic potential. However,since the comet assay is considered an 'indicator test', in that DNA strand breaks may be effectively repaired or lead to lethality, positive results only in the comet assay may not be indicative of the ability to induce permanent DNA changes. Therefore, for any carcinogens that are positive in the Ames test, but for which the only in vivo result is a positive comet assay, the appropriateness of considering such a compound as a DNA-reactive mutagenic carcinogen is uncertain and if possible additional information should be sought.

It is hoped that a sufficient number of carcinogens with a strong genotoxicity/mutagenicity profile could be identified by such an analysis to allow an adequate re-assessment of the TTC values. However, if too few carcinogens are found using this approach, the next step would be to add carcinogens for which there are mixed (both positive and negative) *in vivo* genotoxicity results. A careful evaluation of the robustness of the respective positive and negative results would be needed, such that a conclusion of most likely genotoxic MOA could be made. Given the propensity of mammalian cell genotoxicity tests to give 'misleading' positive results (Kirkland et al. 2006; Kirkland et al. 2007) particularly from older published studies where extreme conditions may not have been controlled, or where p53-deficient rodent cell lines of unauthenticated origin were used, it is recommended that a conclusion of genotoxic MOA should not rely solely on *in vitro* results.

Non-genotoxic carcinogens

Traditionally, it has been assumed that a TTC value set for genotoxic (DNA-reactive mutagenic) carcinogens would be sufficiently protective for non-genotoxic carcinogens i.e. thresholds for non-genotoxic carcinogens would be expected to be higher than those for DNA-reactive mutagenic carcinogens and would be covered by existing, higher TTC values. However, this has never been independently tested (but see Cheeseman et al. (1999), discussed above). In order to do that, a priority list of non-genotoxic carcinogens relevant and potentially relevant to humans should be identified.

It is theoretically more demanding to identify a non-genotoxic MOA for a carcinogen than a genotoxic MOA. Since mutagenic, clastogenic and/or aneugenic activities may be involved in one or more key steps of the carcinogenic process, negative results would need to be obtained across all of these endpoints in order to conclude absence of genotoxic activity unequivocally. The tests in which such negative

results were obtained would also need to be rigorous, for example, the following when conducted according to the Organisation for Economic Co-operation and Development (OECD) or equivalentguidelines (see also Eastmond et al. 2009):

- Bacterial mutation tests including *S. typhimurium* or one of the *E. Coli* WP2 strains
- Chromosomal aberration or micronucleus tests performed to acceptable levels of cytotoxicity, scoring sufficient cells, and including short and prolonged treatment times
- Mammalian cell gene mutation tests, in which sufficient cells have been treated, subcultured and plated for mutant selection.

In addition, absence of structural alerts, and absence of adduct formation with DNA would add weight to a conclusion of absence of genotoxic activity. At this time we are not aware of any curated databases of clearly non-genotoxic chemicals, and therefore identifying a priority list of human-relevant (and potentially relevant) non-genotoxic carcinogens for re-assessment of TTC values will probably need to be carried out on a case-by-case basis. Despite the challenges in forming a priority list of human-relevant (and potentially relevant) non-genotoxic carcinogens, there are several well documented processes that can lead to tumour formation via a non-genotoxic MOA (see Hernandez et al. 2009; BAuA, 2014; Luijten et al. 2016; Jacobs et al. 2016), for example:

- (Peroxisome proliferation)
- (CAR/PXR activation)
- AhR activation
- Cytotoxicity

- Growth stimulation (mitogenesis)
- Inflammation
- Immunosuppression
- Endocrine modification.

The first two modes of action are considered of doubtful, if any, significance to humans (Corton et al. 2014; Elcombe et al. 2014). Substances acting via these agreed modes of action, but excluding any possible DNA-reactivity/genotoxicity, could be used to identify a priority list of human-relevant (and potentially relevant) non-genotoxic carcinogens.

Proposed approach to re-assess TTCs for carcinogens

Revise and expand CPDB

Inclusion/exclusion criteria

The current TTC value of 0.0025 µg/kg bodyweight per day for genotoxic (DNA-reactive mutagenic) carcinogens was based on analyses of information from the CPDB. Re-evaluation of this database to determine TTC values for DNA-reactive mutagenic and non-genotoxic carcinogens relevant and potentially relevant to humans could be achieved using the hierarchical approach outlined above. A set of criteria are proposed for selecting suitable data to ensure the dataset underlying these TTC values is scientifically robust and transparent. These criteria include general study design parameters as well as more specific inclusion and exclusion criteria. Since this exercise will be resource intensive, it is proposed that it would be more efficient to collect all potentially useful studies in a new database. From this database, the data qualifying for determining the TTC values can be selected. These data are further referred to as the

‘dataset’. A database constructed in this way would be suitable for a number of applications, additional to the selection of a subset of data for re-assessment of the TTC values.

General criteria for the CPDB

In total, five general criteria for the database were defined and evaluated. These are as follows:

Box 1: General criteria

(1) Study type: carcinogenicity studies;

(2) Relevance and reliability: studies with Klimisch score 1, 2 or 4;

(3) Route of exposure: studies that used either oral dosage regimen or inhalation;

(4) Species: include any;

(5) Studies listed in CPDB as ‘TBA’, ‘MXA’, or ‘MXB’: exclude.

(1) Study type: carcinogenicity studies. Preferably, the cancer potency database should contain studies performed according to, or consistent with, OECD Test Guideline 451 on Carcinogenicity Studies (OECD 2009a) or Test Guideline 453 on Combined Chronic Toxicity/Carcinogenicity Studies (OECD 2009b). In any case, only studies examining tumour responses will be relevant for the database.

(2) Relevance and reliability: studies with Klimisch score 1, 2 or 4. The reliability categories as described by Klimisch et al. (1997) have proven to be useful in regulatory risk assessment, for example in the OECD High Production Volume (HPV) program and under REACH. They take into account the reliability, relevance and adequacy of

data for use in reaching hazard/dose-response conclusions on an endpoint for a specific substance. In addition, consideration is given as to whether the administered material has been appropriately characterized and that information on this has beenprovided.Studies with a Klimisch score of 1 areperformed according to generally valid and/or internationally accepted testing guidelines or in which the test parameters documented are based on a specific (national) testing guideline (preferably performed according to good laboratory practice (GLP)) or in which all parameters described are closely related/comparable to a guideline method. Studies with a Klimisch score of 2 are those in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable. Studies with a Klimisch score of 3 are those in which there are interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g., unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for an assessment and which is not convincing for an expert judgment, Studies with a Klimisch score of 4 are those which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.).It is proposed that studies with a Klimisch score of 1 (acceptable without restrictions), 2 (acceptable with restrictions) or 4 (not assignable) should be considered for inclusion, in terms of ‘potentially useful’, in the database. Inclusion of studies with a Klimisch score of 4 in the dataset used to re-evaluate the TTC values would have to be decided on a case-by-case basis. Discriminating between studies with a Klimisch score of 1 versus a score of 2 may be difficult due to a lack of

information provided on guideline compliance. This, however, should not pose a problem when re-evaluating data. Studies with a Klimisch score of 3 are not considered suitable for inclusion in the database. The Klimisch score should be noted in the database.

*(3) Route of exposure: studies that used either oral dosage regimen or inhalation.*For the database, data for chemicals studied using oral (diet, gavage, drinking water) or inhalation as the route of exposure should be collected, as they are physiologically relevant exposure routes in the safety assessments of chemicals. For volatile chemicals, in particular, a significant number of systemic carcinogens might be missed if inhalation studies are excluded. Information on the extent of systemic exposure should be recorded, when available and the absence of such information noted. Inclusion of data from studies using inhalation exposure should be included in the TTC dataset only when relevance to oral exposure is plausible and when it is possible to calculate the equivalent oral exposure. When calculating the POD for the TTC dataset, such exposures should be recalculated to give mg/kg bw per day. Studies using dermal application or intraperitoneal, intravenous or subcutaneous injection as the route of exposure are generally not considered suitable for derivation of potency values that would be physiologically relevant for normal routes of exposure and should therefore be omitted.

*(4) Include any species.*Studies in any species are considered potentially relevant and should be included in the database. Whether or not a given study should be included into the TTC dataset will depend on various factors, such as, for example, availability of historical control data and animal numbers used. It is noted that most studies with unusual species will probably not meet the quality criteria.

*(5) Studies listed as 'TBA', 'MXA', or 'MXB' in CPDB: exclude.*The CPDB reports TD50 values that have been derived from either statistically significant findings in a single tissue, which should be included in the dataset for relevant studies, or from findings observed in all tumour bearing animals (TBA), from more than one site, combined by NCI/NTP (MXA), or from more than one site, combined by Berkeley (MXB). Data from studies listed in the CPDB as 'TBA', 'MXA', or 'MXB' should be excluded from the dataset, as the biological relevance of such grouping, comprising a range of pathologies and potential modes of action, is difficult to interpret. The FDA CFSAN 2012 study (J. Aungst, K. Arvidson, D. Rua, D. Hristozov, B. Mugabe and A. McCarthy, C. Yang, M. A. Cheeseman. CFSAN/OFAS 2012) also excluded these mixed combined tumours.

Criteria for inclusion in the cancer potency dataset

In addition to these general criteria, specific inclusion criteria for acceptable positive studies and acceptable negative studies are proposed. Negative studies will be useful to be able to conduct a Weight of Evidence (WOE) analysis for substances with mixed results, *i.e.,* with different results from multiple studies (see below). The inclusion criteria listed below should be applied as guidelines with some flexibility to ensure that no studies relevant for potency setting are lost. In general, all criteria should be met.

Box 2: Inclusion criteria

Acceptable positive studies

(1) Tumour findings relevant or assumed to be relevant to humans;

(2) Tumour incidences based on one species and one sex per one tissue/organ;

(3) Studies with exposures shorter than usual (≤ 18 months for mice and ≤ 24 months for rats) included if statistically significant (for example, one-tailed test for pair-wise comparisonfor a common tumour type, P≤0.01);

(4) Sufficient number of animals per sex and dose group evaluated for carcinogenicity by the end of the study to enable a reliable estimate of the POD.Acceptable negative studies

(1) Exposure duration ≥ 18 months for mice and ≥ 24 months for rats;

(2) MTD achieved,limit dose reached for practical reasons, or acceptable multiple of human exposure (internal dose) achieved;

(3) Group size minimum (per sex) is 40, with survival being at least 50% at end of study;

(4) Notumour shows an incidence statistically significantly different from that in the concurrent control group.

Acceptable positive studies

Studies in which the substance is considered positive or negative for carcinogenicity by an authoritative body, such as the NTP, should be so noted in the database. However, it is recommended that the findings should also be evaluated independently, using the criteria discussed below, and a conclusion on the acceptability of positive or negative findings based on this assessment should also be noted.

(1) Tumour findings relevant to humans. Unless there is strong and accepted (by authoritative bodies) evidence to the contrary, a carcinogenic response induced in experimental animals by a DNA-reactive mutagenic MOA (likely or possible) will be considered relevant to humans. This applies even to tissues where there is no human counterpart, such as the Harderian gland.

There are a few site-of-contact carcinogens which exhibit some genotoxicity in vitro but where the evidence for their carcinogenicity is for a MOA that does not involve DNA reactivity. An example would be ethyl acrylate in the rat and mouse forestomach, where the concentration is uniquely high due to the particular anatomy and physiology of this rodent-specific organ and the MOA appears to involve local irritation (Proctor et al. 2007). Where there is general acceptance by the scientific community (and by one or more authoritative bodies) that such a response is not relevant to humans because there is no human counterpart to the forestomach, a similar position should be adopted.

Where a carcinogenic response is through a non-genotoxic MOA, human relevance assessed using the WHO IPCS Human Relevance Framework (Boobis et al. 2006; Boobis et al. 2008; Meek et al. 2014) should be determined. Where there is general acceptance by the scientific community (and by one or more authoritative bodies) that a carcinogenic response in experimental animals is not relevant to humans this will be deemed sufficient evidence to exclude these data from TTC analysis. In the case of non-genotoxic carcinogens, site of effect is an important factor in determining human relevance. Where such a non-genotoxic response is observed only in a tissue with no human counterpart, this will be strong evidence for exclusion of the data from TTC analysis (see Edler et al. 2014).

(2) Tumour incidences based on one species and one sex. Tumour incidences and PODs in the database should be based only on specific site/sex, species and tissue/organ. Expert judgement and transparent weight-of-evidence assessment will be needed to determine if isolated increases in one sex of one species in one tissue or whether conflicting results in different studies in the same tissue, sex, strain and species are sufficient evidence of a relevant response for inclusion in the dataset. Data where tumour findings from both males and females or different sites have been added together would be deemed not informative. There is no firm recommendation in terms of statistical test to be performed, maximal P-value, or minimal sample size for inclusion into the database. However, only substances with a relevant positive (i.e. statistically significant) tumour response should be included into the TTC dataset.

(3) Studies with exposures shorter than usual (≤ 18 months for mice and ≤ 24 months for rats) included if statistically significant (e.g. one-tailed test for a common tumour type, P≤0.01). Potent carcinogens and/or high doses may cause tumour incidences to increase early, relative to controls, so that studies of shorter duration than normal may still be sufficient to enable characterisation of the dose-response relationship. Hence, these studies may still be useful and should therefore be included into the database. Responses in this type of study would have to be extrapolated for lifetime exposure when estimating potency. Inclusion into the TTC dataset should be decided on a case-by-case basis (expert judgement).

(4) Sufficient number (usually at least 10) of animals per sex and dose group evaluated for carcinogenicity by the end of the studyto enable a reliable estimate of the POD. If a positive study has only a few animals per group, the uncertainty in the TD50 estimate becomes very large. Applying a BMD approach, i.e. using a $BMDL_{10}$ instead of a TD50

as POD, will not overcome this issue because the number of dose groups in a carcinogenicity study is usually fairly low (see Slob, 2014). A judgement would have to be made on the acceptability of the POD estimate, based on the difference between the BMD, BMDL and BMDU (estimates of the 95% lower and upper confidence limit on the BMD, respectively) (e.g. EFSA 2009).

Acceptable negative studies

(1) Exposure duration ≥ 18 months for mice and ≥ 24 months for rats. This criterion is proposed because in cases where studies are negative, it study duration is short it cannot be excluded that this was the primary reason for the absence of an observable increase in tumour incidence. For other species (hamster, monkeys, dogs, etc), ideally life-time exposure should be considered; however, for species other than rat or mouse, adequacy of study duration should be by expert judgement on a case-by-case basis.

(2) MTD achieved,limit dose reached for practical reasons, or acceptable multiple of human exposure (internal dose) achieved. In a carcinogenicity study, it is usually required that the highest dose achieved is the MTD to ensure the adequacy of neoplastic hazard identification. In combination with lower doses the use of the MTD also enables a dose response evaluation. It is noted that for pharmaceuticals and to some extent also agrochemicals, the top dose can be set on considerations of human exposure and/or saturation of exposure. Classical considerations to define whether an MTD or adequate dosage has been achieved include the following (see also Rhomberg et al. 2007).

The highest dose level elicited evidence of toxicity (2.1, 2.2, or 2.3 below), is a limit dose (2.4), or is an acceptable multiple of human exposure (particularly for human pharmaceuticals (2.6):

(2.1) Depression of body weight gain (approximately 10%) or any other adverse effect that may limit the high dose level that could be humanely achieved (OECD 2009a)

(2.2) Data from dose-setting studies that indicate the dose level is approaching a toxic level, even if not observed in the carcinogenicity study itself. For example, an incrementally higher dosage in the dose setting study was overtly toxic;

(2.3) Although an MTD in terms of body weight gain was not achieved, overt organ toxicity was observed;

(2.4) Alternatively, the high dose may be defined as a maximum practical dosage, on the nutritional basis that the concentration of the chemical should not normally exceed 5% of the total diet (OECD 2009a) or is a limit dose (e.g. 1000 mg/kg/body weight in OECD 408 or dose setting studies);

(2.5) There is robust evidence that systemic exposure does not increase beyond the top dose used, even if below a limit dose, due to saturable absorption.

(2.6) The high dose results in a systemic exposure that is a large multiple of the human area under the exposure curve (AUC), typically 25-fold or greater, for substances that have similar metabolic profiles in humans and rodents and low organ toxicity in rodents (i.e., high doses are well tolerated in rodents). This approach is most often used for human pharmaceuticals, for which suitable pharmacokinetic data should be available.

Fulfilment of classic MTD acceptance criterion may be of limited importance for substances of high neoplasticpotency. Use of sufficiently high dosage is likely to be

a more important criterion for inclusion of substances of low potency, for which no tumours were observed.

The main need is to be able to define a POD for any neoplastic change in a study. It is quite possible that an individual carcinogenicity study fails to reach a MTD for non-neoplastic change but is highly relevant due to high neoplastic potency. Conversely, it is quite possible that an individual carcinogenicity study fails to reach a MTD for non-neoplastic change or neoplastic change, and fails to show hazard potential, but nevertheless is sufficient to indicate that the substance has little or no carcinogenic potency and thus a high POD (if at all).

It is also possible that a study is considered of uncertain relevance for POD determination if excessive toxicity occurred at all dosages, and a carcinogenic response seen in the study is considered likely secondary to the toxicity observed. This would need to be judged on a case-by-case basis.

In summary, whether MTD was achieved or not is of limited relevance to the use of study data in the database and dataset, and generally would not be a basis to consider a study as scientifically inadequate for use in defining carcinogenic threshold(s), if potency information is available. Therefore the MTD requirement should in general be considered using expert judgement on a case-by-case basis.

(3) Group size minimum (per sex) is 40, with survival being at least 50% at end of study. In any cancer bioassay, as indeed in any quantitative scientific study, it is not possible to prove a negative. Statistical analyses are undertaken to determine whether the null hypothesis can be rejected, in this case that substance administration has no effect on tumour incidence. Lack of statistical significance does not prove no effect, only that there is no evidence for an effect within the power of the study. Hence, to ensure adequate power in studies with negative findings, and provide reasonable

confidence in the upper limit of any possible response in the absence of statistical significance, the minimum group size per dose per sex should be 40 animals (for larger species, such as dogs and monkeys expert judgement will be needed to assess acceptable group size). In addition, because of the latency of some tumour types, at least 50% of the animals should be evaluable histologically at study termination, to ensure that there is sufficient power for assessment at the end of the study. Whilst large group sizes could result in survival of at least 20 animals at the end of the study, even with extensive precedent mortality, this may obscure a particularlyaggressivecarcinogenic response and hence would not be acceptable for the purposes of the present exercise.

(4) No tumour shows an incidence statistically significantlydifferent from that in the concurrent control group. In studies that meet all of the inclusion criteria, including adequacy of dosing regimen, dose group size and study duration, where there is no statistically significant increase in tumour incidence in any tissue (e.g. one-tailed test, $P \leq 0.01$ for pairwise comparison of a common tumour type), they should be considered negative for carcinogenicity.

Consistent with the inclusion criteria for acceptable positive and acceptable negative studies a total of four exclusion criteria were proposed.

Box 3: Exclusion criteria

Unacceptable positive studies

(1) Tumour responses observed only at doses that exceed the MTD;

(2) Studies performed in short-term animal models of carcinogenicity and other study types;

(3) Single dose group studies;

(4) Tumour findings irrelevant for humans.

Unacceptable positive studies

(1) Tumour responses observed only at doses that exceed the MTD. The MTD may be deemed to be exceeded based on:

(1.1) Excessive depression of body weight gain;

(1.2) Excessive mortality (from non-cancer causes);

(1.3) Signs of excessive toxicity such as marked non-neoplastic histopathological changes, clinical signs, haematology – even if there is not an excessive depression of body weight gain.

It is recommended that the option to evaluate any bioassay that reports a positive tumour response, on a case-by-case basis, with regard to whether or not the MTD has been exceeded be kept open. Also, it should be borne in mind that target organ toxicity may be the precursory basis for tumour development.

(2) Studies performed in short-term animal models of carcinogenicity and other study types. Models such as transgenic mice and partially hepatectomised rats might be of value in assessing carcinogenic hazard but they are not helpful in assessing potency towards normal organisms. Hence, it is recommended that data obtained from such models should not be used for re-evaluating the TTC values. Similarly, any other study type for which there is insufficient experience on potency correlation to normal exposure situations or organisms, e.g. genetically highly susceptible mouse strains or newborn mice, should be considered irrelevant for this purpose.

(3) Single dose group studies. Data from single dose studies should be collected for the database, because they are potentially useful in a WOE approach. However, single dose

studies do not permit meaningful determination of potency and hence should not be included in the TTC dataset.

*(4) Tumour findings irrelevant for humans.*See inclusion criteria.

Assessment of dataset entries

Application of the hierarchical analysis outlined in section on "Outline of proposed re-evaluation of TTC values – proposed hierarchical analyses", combined with the criteria listed above, would hopefully result in a dataset comprising a sufficient number of carcinogens that are considered most likely to have a genotoxic MOA. In addition, a dataset of non-genotoxic carcinogens relevant or potentially relevant to humans should be identified.

Evidence of a causal role of genotoxicity in the carcinogenic MOA

Each carcinogenic response included in the database should be assessed using rigorous and transparent weight-of-evidence for the contribution of genotoxicity to the MOA for the carcinogenic response observed. It is proposed that the IPCS MOA Human Relevance framework be used for this purpose (Boobis et al. 2006; Boobis et al, 2008; Meek et al. 2014). Preston & Williams (2005) provide a specific example of application of the framework in this way. Multiple sources of information, including data on genotoxicity, precursor effects, mechanistic studies and structural analogues will be of value in such assessment. It is suggested that, ideally, compounds should be identified as DNA-reactive mutagenic carcinogens (clear threshold, no clear evidence of a threshold for carcinogenicity), carcinogens involving a different genotoxic MOA (i.e. not via DNA-reactive mutagenicity),non-genotoxic carcinogens (relevant or potentially relevant to humans), genotoxic and carcinogenic but insufficient evidence to determine

causality (clear threshold, no clear evidence of a threshold for carcinogenicity). Such identification will enable sensitivity analyses to be performed when deriving TTC values.

POD

Ideally, the same method would be used to determine the POD for all substances in the TTC dataset. As indicated above, there is a clear preference that this should be the $BMDL_{10}$, based on the results on animal bioassays (see section on "Advances in dose-response modelling and recommendation for use of BMD approach"). This would likely require benchmark dose modelling of a large number of datasets. As discussed in section on "Need to re-assess TTC values for genotoxic and carcinogenic compounds", it is unknown whether the estimates of the LTD_{10}, currently available in the CPDB, are sufficiently reliable and similar to $BMDL_{10}$ values calculated by *de novo* benchmark dose modelling. This should be evaluated for appropriate subsets of chemicals, both for DNA-reactive mutagenic and human-relevant (and potentially relevant) non-genotoxic carcinogens, before using LTD_{10} values for further analyses. Given the number of decisions that have to be made when deriving BMDs and BMDLs, it will be essential that a consistent, transparent approach be used for any such modelling (e.g. EFSA 2009).

There may be multiple BMDLs for a given tumour response, either in the same study because more than one model is acceptable, or in more than one study of comparable design. Consideration will need to be given as to which BMDL value should be used in the analysis (lowest, mean, model average, etc). Guidance on this can be found in EFSA (2017).

Where administration of the test article was less than daily (for oral exposures) or not for 24 h per day (for inhalation), appropriate adjustment of the POD will be

necessary (seeEnvironmental Protection Agency (EPA), 2005).In situations where duration of exposure was for less than lifetime, see above for inclusionand exclusion criteria.Where necessary, appropriate defaults may have to be used for physiological variables, such as breathing rate, food consumption and water consumption, such as in the EPA (2011) Exposure Factors Handbook.

Extrapolation to human equivalent dose

In all previous analyses to derive TTC values for genotoxic (DNA-reactive mutagenic) carcinogens, linear extrapolation from the PODs in animal cancer bioassays to a nominal minimal human risk value has been used to determine VSDs for distributional assessment. The minimal excess human risk value used has been 1×10^{-6}, other than for impurities in human pharmaceuticals in late stage development or clinical use, when a value of 1×10^{-5} has been used (see sections on "FDA TOR and additional thresholds developed from it" and on "Database used to establish the TTC of 0.15 µg/day is many years old and has not been updated for several years"). In the first analyses to reassess the TTC values for genotoxic carcinogens, it is suggested that VSDs based on a minimal excess human risk value of 1×10^{-6} be used, to enable comparison with previous TTC values. Additional analyses could then be performed using different approaches, as appropriate, for example different minimal human risk values, relaxation of the assumption of linear, no threshold depending on MOA, WOE for genotoxicity, and nature of genotoxicity.

Importance of rigorous quality control of data and open access

It is strongly recommended that all data be subjected to rigorous quality assurance and quality control, in order to prevent data entry errors and create a reliable, sound and trustworthy database. This database and all associated information should be freely

available to the public. Permitting open access to the database will allow other researchers to conduct similar or related analyses; moreover, it will enhance adoption and implementation of any new TTC values for risk assessment purposes.

Identification of 'relevant' genotoxins using structural alerts

The cancer TTC value is applicable only to those compounds that are likely to be carcinogenic by a DNA-reactive mutagenic MOA. Hence, in applying the TTC approach it is necessary to identify those compounds that are likely to be DNA-reactive mutagens. Genotoxicity comprises a number of different modes of action, amongst which direct interaction with DNA of the compound itself or a metabolite leading to covalent modification of DNA is of most concern. Indeed, there is good evidence that other genotoxic modes of action either do not lead to carcinogenesis or exhibit a clear threshold, and would be covered by higher TTC values. Indeed, this is the basis for the various current TTC decision trees; it is necessary only to identify putative DNA-reactive mutagens for application of the lowest TTC value of 0.0025 µg/kg bw per day and compounds that do not fall into this category will be covered by other, higher TTC values. It is thus critical in the application of the TTC approach that those compounds likely to be mutagenic via a DNA-reactive mode action can be reliably identified on the basis of structure alone.

EFSA and other work on identification of DNA-reactive mutagenic compounds using structural alerts

EFSA, through their Panel on Plant Protection Products and their Residues (PPR Panel), reported the findings of an EFSA-commissioned study undertaken by the European Commission’s Joint Research Centre (JRC) to evaluate the toxicological relevance of metabolites of pesticide active substances (EFSA 2012). The JRC report (2010)

conceived a conceptual framework, based on the OECD Principles for the Validation of QSARs (quantitative structure–activity relationships), to review (Q)SAR approaches – for the purposes of this manuscript the term (Q)SAR is inclusive of both structural alert and related SAR approaches as well as statistical models such as QSARs; where necessary in this manuscript SARs are considered separately from QSARs. In addition, for DNA-reactive mutagenicity, a number of expert systems, representing a knowledge-driven rulebase of structural alerts (DEREK), statistical models (CAESAR, LAZAR, TOPKAT, HazardExpert and ToxBoxes) and a hybrid rule-based / statistical system (Toxtree) were evaluated with regard to predictions of the results of over 1,500 substances. The chemical space of the substances, as compared to e.g. pesticides, was evaluated, providing some – if not complete - insight into the relative applicability domains.

The overall consensus from JRC and the EFSA panel is that there is utility in predicting DNA-reactive mutagenicity from (Q)SAR models. Improvements in accuracy of prediction are seen when forming some type of consensus between a rule–based and a statistical system. Overall the conclusions appear sound – (Q)SAR methods may be used with some confidence, within defined areas of chemical space. This may be assisted by the use of read-across (assuming sufficient data are available).

Advances in software for identifying structural alerts

Structural alerts are fragments of molecular structures or functional groups that convey some type of biological activity. They vary widely in their form, use and application. In the context of this application they will assist in the identification of DNA-reactive mutagens(and hence putative carcinogens) since they are ideally suited for identifying fragments of molecules associated with mutagenicity. They are based on a rich and long history of toxicological understanding, with the original fragments predating

computational technologies to make them useable e.g. the bay region of polycyclic aromatic hydrocarbons(PAHs), which is the area bounded by three contiguous adjacent bonds, one from each of three fused aromatic rings of a PAH molecule (Lehr et al. 1985). The use of structural alerts crystallised with the seminal publication of Ashby and Tennant (1988) that compiled, *in parchmento*(in their paper), over 20 fragments associated with electrophilic chemistry and known DNA-reactive genotoxic carcinogenicity. It took several years to codify Ashby and Tennant's structure-activity relationships computationally, but since then the 'alerts' have been expanded and defined.

Structural alerts for DNA-reactive mutagenicity are effective for mechanisms that are based around electrophilic reactivity, i.e. compounds that will be positive in the Ames test. Such mechanisms are well defined from an organic chemistry point of view and can be readily captured (Enoch and Cronin 2010; Enoch and Cronin 2012). There are then a number of technologies to identify these structural alerts, the premise being that the presence of an alert in a molecule would indicate the potential for activity (e.g. Votano et al. 2004; Tropsha and Tropsha 2010). Currently technologies include commercial systems from the major software houses, to the use of open access systems. Whilst the process of defining an electrophilic fragment itself may be straightforward, a number of decisions need to be borne in mind when applying a fragment. The level of definition of an alert is important and they may be designed to be broad e.g. to identify any molecule with a specific functional group. Alternatively, the level of definition of an alert may be designed to be narrower, to take into account the 'molecular environment' e.g. other parts of the molecule that may decrease or increase activity. Broad alerts are useful for grouping and category formation – the OECD and OASIS DNA binding profilers are examples of these. Implicitly they are likely to be over-

predictive. Narrower, or more defined alerts, e.g. in DEREK Nexus are likely to be more applicable in hazard assessment.

This *in silico* process of identification of structural alerts is an important consideration in the hazard identification of a compound with few data. However, it does not provide a quantitative threshold for toxicological concern and does not diminish the importance of reliable TTC values (Figure 1). It is always difficult to know which of the various types of software to utilise and there are no hard and fast rules. It must also be borne in mind that all *in silico* approaches, for making important decisions on individual chemicals, must be considered in context and on a chemical-by-chemical basis. One approach, e.g. ICH M7 for DNA-reactive mutagenic impurities (ICH 2014), has been to consider a rule-based (SAR) and statistical (QSAR) approach. For DNA-reactive substances, it may be possible to refine this process. For instance, combination of a chemistry based profiler (e.g. the OECD profiler for DNA reactivity in the OECD QSAR Toolbox) with appropriate use of the rules for DNA reactivity from a knowledge based profiler (e.g. DEREK Nexus) may give complementary and increased coverage. QSAR models may be less applicable for the identification of directly DNA-reactive substances than structural alert based approaches as they are derived from statistical models utilising global, or general, molecular descriptors as opposed to predictions directly from known DNA-reactive fragments. However, they may again provide complementary information. Thus, there may be merit in considering the use of two types of structural alert approach, possibly in addition to a QSAR model. The disadvantage is the increased complexity of using the information from multiple models, e.g. how and when a consensus would be reached. This will be, in part, down to the user to decide how conservative they require the assessment to be.

A further, critical, aspect to the application of structural alerts is that of metabolism and how this is included, or not, into the model. Alerts can be written to implicitly take account of metabolism, for instance, alerts for aromatic amines assume the metabolic steps that lead to the reactive nitrenium ion metabolite; thus it is not the parent compound itself that is genotoxic/mutagenic but the assumed metabolite. Other alerts do not assume the metabolic step. Thus, the metabolites themselves must be predicted and subsequently screened for structural alerts. Likewise, with SAR or statistical approaches, the metabolic step may be implicit. Whilst there are computational methods to predict metabolites (Kirchmair et al. 2015), they can predict a large number of metabolites, many of which are irrelevant(i.e. whilst they are theoretically possible they are not formed in biological systems),from which it can thus be difficult to identify those important for toxicity. Thus, metabolism must be considered in the hazard assessment of compounds that do not have a structural alert identified. The consideration of metabolism will inevitably affect the performance of the models and better consideration of how to include these effects, and those metabolic conversions most relevant for genotoxicity, is required. Thus, with the new, and refinements of existing, methods, there is a need to revisit this process to determine if the overall scheme and approach can be improved even further.

A further issue to be noted with the use of structural alerts for DNA-reactive mutagenicity is the meaning of a ‘negative’ prediction. In other words, if a compound does not contain a structural alert, how much confidence is there that it is not a DNA-reactive mutagen? This topic has stimulated much debate (Ellison et al. 2011). The reality is that it is widely acknowledged that predictions of DNA-reactive mutagenicity from structural alerts are likely to be more acceptable (from a precautionary perspective) than negative predictions. Nevertheless, in their recent review of the TTC

approach, EFSA and WHO concluded that negative predictions using appropriate software were acceptable in the application of the TTC approach (EFSA 2016) and the reliability of negative predictions is supported by a recent analysis by Williams et al. (2016).

With regard to the practical application of structural alerts to identify DNA-reactive mutagenic carcinogens, software can be freely available such as SMARTS (2016) and the chemoTyper(Altamira LLC; Yang et al. 2015), or associated with a cost e.g. requiring payment or on a commercial basis (e.g. DEREK Nexus). In addition, there are different philosophies in the derivation of rules, varying from direct mechanistic interpretation and toxicological expert knowledge (e.g. DEREK Nexus), to derivations from organic reaction mechanisms and chemistry (e.g. OECD DNA binding profiler), to machine learning approaches (e.g. MultiCASE). The compilations of alerts that are freely available are summarised in **Table 2**. The most comprehensive of these is available in the OECD QSAR Toolbox (freely downloadable from http://www.qsartoolbox.org/). This comprises a number of profilers that can be applied to identify DNA-reactive mutagenic carcinogens, including those from the European Commission's Joint Research Centre's ToxTree software and United States Environmental Protection Agency's Oncologic software. At this time, however, there is no coherent strategy on how to apply all these profilers and it should be borne in mind that the purpose of the OECD QSAR Toolbox is primarily to facilitate grouping and not to identify hazard. The commercial software to predict mutagenicity from structural knowledge is summarised in **Table 3**. In addition, there are a number of techniques which can be thought of as being 'hybrid' (a mixture of structural alerts and models), or statistically derived QSARs, and these are summarised in **Table 4**.

Discussion

The TTC approach was recently reviewed at an Expert Workshop organised by EFSA and the WHO (EFSA 2016). At that meeting, whilst evidence supporting the non-cancer TTC values was reviewed, and proposals for improvement were made, with respect to the cancer TTC value it was concluded that:

> 'Expanding the TTC cancer dataset (e.g. with the ToxRef database) would enhance the power and range of chemical structures covered. However, this is not considered a priority as it would be resource demanding and is not expected to significantly affect the approach.'
> 'If a revision of the carcinogenicity/genotoxicity based TTC were to be envisaged, it is recommended considering approaches other than TD50-based linear extrapolation from the most sensitive species and most sensitive site, which may be overly conservative.'

Hence, reanalysis of the appropriateness of the TTC value for DNA-reactive mutagenic carcinogens as outlined in this document would add to the confidence in the overall TTC approach.

Rigorous, transparent basis for TTC value for compounds that are DNA-reactive mutagenic carcinogens

The derivation of a TTC value for compounds that are DNA-reactive mutagenic carcinogens is currently not as transparent as for the other TTC values in current use in the TTC decision trees of organisations such as the WHO and EFSA. Whereas the latter are based on distributional analyses of publicly accessible databases, the former is based on a hybrid approach, difficult to follow and document (e.g. Kroes et al. 2004). The database on which this TTC value is based comprises carcinogens with a range of modes of action, human relevance and quantitative outcomes in rodent bioassays. Re-analysis of the database, with expansion if possible, according to the criteria outlined in

the current manuscript would provide a much more transparent and rigorous basis for derivation of this TTC value.

Appropriateness of non-genotoxic TTC values for compounds that might be carcinogenic by another MOA

It is implicit in the TTC approach that any compound that is not captured by the lowest TTC value would be covered by the higher TTC values in the decision tree. Most often this would be 1.5 µg/kg bw per day or higher (unless the compound has a structure suggesting inhibition of acetylcholinesterase activity). Yet, carcinogenicity, *per se*, is not reliably predictable on the basis of structure alone. Hence, it has to be assumed that any compound that is carcinogenic by a MOA other than DNA-reactive mutagenicity, will be adequately covered by these higher TTC values. As discussed above, there is evidence that this would be the case, at least for the current TTC value for DNA-reactive mutagenic carcinogens. However, detailed analysis as described above would enable robust conclusions to be reached on a) what is the level of protection provided by the higher TTC values for carcinogens acting by a non-genotoxic MOA and b) how does this compare with any revised TTC value for DNA-reactive mutagenic carcinogens.

Reliability/protectiveness of use of structural alerts for DNA-reactive mutagens

Axiomatic in this approach is the ability to identify those compounds likely to be DNA-reactive mutagens. As discussed above, rigorous evaluation of the original data on genotoxicity, and application of a robust weight-of-evidence approach, will provide a more reliable dataset on not only the genotoxicity of the compounds in the database but also on the MOA for their genotoxicity. In addition, assessment of likelihood of *in vivo* genotoxicity will be invaluable in such analyses. This information will enable a detailed

evaluation of the reliability and suitability of different software packages in identifying those compounds of most concern with respect to the TTC value for DNA-reactive mutagenic carcinogens. Also, as described in section on “Identification of 'relevant' genotoxins using structural alerts”, the *in silico* techniques used may utilise different approaches e.g. chemistry and knowledge derived structural alerts and QSAR models. A well curated database is essential to determine the performance of such models. The accuracy of the structural alerts and QSAR models, either individually or through weight-of-evidence or consensus, can be compared with both the genotoxicity and the carcinogenicity data, enabling conclusions to be reached on which software packages are most suitable for application in the TTC approach. This may also highlight well performing (in terms of protectiveness) alerts and provide insight into the role of metabolism. Their level of protection can be calculated based on the false positive and false negative rates.

More detailed analysis could reveal structural alerts that do not contribute meaningfully to the overall conclusions and hence could lead to revision of the software for use in TTC applications.

Possibility of including structural alerts for other cancer MOAs

Initiatives such as the WHO Mode of Action for Chemical Carcinogens and Human Relevance, the OECD Adverse Outcome Pathways (AOP) database and the National Institute of Environmental Health Sciences (NIEHS) Hallmarks of Cancer are leading to the identification of early key events, including molecular initiating events (MIEs) that are causally related to the carcinogenicity of chemicals, by a variety of both genotoxic and non-genotoxic modes of action. When the primary molecular interaction is known (the MIE), it is likely that structural or other physiochemical predictors will be established and that these can be used as alerts for such modes of action. A database

constructed as described above should enable the utility of some such alerts to be evaluated and perhaps ultimately to be included in a revised TTC decision tree. However, the benefit of this should be judged against the relative level of protection provided compared to that obtained using the existing TTC values.

Reflections on COC

In applying the TTC approach, a number of compounds are excluded, *a priori*, based on their membership of specific structural groups, e.g. aflatoxins, nitrosamines. These have been termed the cohort of concern and they are excluded because some members of the respective groups are such potent carcinogens that even the TTC for DNA-reactive mutagenic carcinogens would not be sufficiently protective and a value low enough to be protective would be of no practical value. However, detailed evaluation of the carcinogenicity, dose-response data and species differences of such compounds as envisaged above would enable their exclusion to be reassessed. For those classes where continuing exclusion was considered appropriate, the possibility of a group-specific TTC could be assessed.

Potential applications of such a database beyond re-evaluation of TTC values

The availability of an up-to-date, high quality, fully populated, curated, publicly available database of the nature envisaged above would have potential application well beyond re-evaluation of the TTC value for DNA-reactive mutagenic carcinogens. Obvious areas of application are the development of predictive algorithms for various types of genotoxicity and for carcinogenicity by a number of non-genotoxic modes of action. The information in the database should prove of value in developing AOPs for a variety of chemicals, for a number of endpoints in addition to cancer, given that information on non-cancer precursor effects will be included as appropriate. The

database may also be of value in addressing the appropriateness or revision of adjustments of the TTC value for less-then-lifetime exposures, e.g. as is current practice for DNA-reactive mutagenicimpurities in human pharmaceuticals (ICH 2014), although this might require expanding the database with information from studies of relevant duration for such an analysis.

Recommendations for depositing and maintaining database

Clearly, for the database to fulfil its full potential, it would have to be maintained well beyond the lifetime of the TTC re-evaluation. Ideally, it would be deposited on a reliable, publicly accessible site, hosted by an organisation that would ensure at least technical maintenance, i.e. continuing accessibility. There are a number of such sites potentially suitable, such as the OECD, the eChem portal of the EU and various initiatives arising out of the US EPA CompTox program. Ideally, once deposited, the database would continue to be extended by addition of new information on chemicals already listed and by the addition of new chemicals. Some mechanism for quality control and data curation would need to be established. Given the public interest in the TTC approach and non-animal methods in general, it is to be hoped that support of the maintenance of such a database would be forthcoming from government and/or supra-national bodies involved in chemical risk assessment.

Next steps

Basedon the foregoing, there is a strong argument for updating the CPDB based on the current state of knowledge and to use this as the basis for re-assessment of the TTC value for substances that are likely to be DNA-reactive mutagens, based on their chemical structure. Such an analysis would provide similar transparency and confidence to this TTC value as now exists for the other TTC values used in the decision tree

developed by EFSA and the WHO (EFSA 2016). The analyses proposed would also establish the minimum level of protection provided by the existing TTC values for substances that are not likely to be DNA-reactive mutagens, and hence pass the first step of the decision tree, but may prove to be carcinogenic by some other MOA, i.e. that progress to step 4 of Fig 1. Finally, retrospective application of software tools for the identification of likely DNA-reactive mutagens to the substances in the dataset used for the reassessment of these TTC values would enable the robustness of the strategy used for step 2 of the decision tree (Fig. 1) to be determined, and enable specific recommendations to be made with respect to software tools and approaches for this purpose.

Declaration of interest

This work was conducted by an expert group of the European branch of the International Life Sciences Institute, ILSI Europe. This publication was coordinated by the Threshold of Toxicological Concern Task Force. Industry members of this task force are listed on the ILSI Europe website at http://ilsi.eu/task-forces/food-safety/threshold-of-toxicological-concern/. Experts are not paid for the time spent on this work; however, the non-industry members within the expert group were offered support for travel and accommodation costs from the Threshold of Toxicological Concern Task Force to attend meetings to discuss the manuscript and a small compensatory sum (honorarium) with the option to decline. The expert group carried out the work, i.e. collecting/analysing data/information and writing the scientific paper separate to other activities of the task force. The research reported is the result of a scientific evaluation in line with ILSI Europe's framework to provide a precompetitive setting for public-private partnership (PPP).ILSI Europe facilitated scientific meetings and coordinated the overall project management and administrative tasks relating to the completion of this work. For further information about ILSI Europe, please email info@ilsieurope.be or call +3227710014. The opinions expressed herein and the conclusions of this publication are those of the authors and do not necessarily represent the views of ILSI Europe nor those of its member companies.

References

Adler S, Basketter D, Creton S, Pelkonen O, Van Benthem J, Zuang V, Andersen KE, Angers-Loustau A, Aptula A, Bal-Price A. 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85:367–485.

Altamira LLC. The ChemoTyper application. Available from: https://chemotyper.org/

Ashby J. 1994. Two million rodent carcinogens? The role of SAR and QSAR in their detection. Mutat. Res. Mol. Mech. Mutagen. 305:3–12.

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

Ashby J, Tennant RW. 1991. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat Res. 257:229–306.

Aungst J, Arvidson K, Rua D, Hristozov D, Mugabe BEM, McCarthy A, Yang C, Cheeseman MA, CFSAN/OFAS UF. 2012. Revisiting the TTC approach for cancer assessment: Part of a road map of computational methods at FDA CFSAN OFAS. In: 51st Annual Meeting and ToxExpo 2012, March 11–15, 2012, San Francisco, California. p. 61.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek MEB, Vickers C, Willcocks D, Farland W. 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. 36:781–792.

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit Rev Toxicol. 38:87-96.

BUaU. 2014. Risk-related concept of measures for activities involving carcinogenic hazardous substances (TRGS 910). http://www.baua.de/en/Topics-from-A-to-Z/Hazardous-Substances/TRGS/TRGS-910.htm (accessed 20/02/17)

Cheeseman MA, Machuga EJ, Bailey AB. 1999. A tiered approach to threshold of regulation. Food Chem Toxicol 37:387–412.

Cohen SM, Klaunig J, Meek E, Hill RN, Pastoor T, Lehman-Mckeeman L, Bucher J, Longfellow DG, Seed J, Dellarco V, et al. 2004. Evaluating the human relevance of chemically induced animal tumors.Toxicol Sci 78:181–186.

Committee on Mutagency of Chemicals in Food, Consumer Products and the Environment. 2010. Guidance Statement: Thresholds for *In Vivo* Mutagens. London, UK

Corton JC, Cunningham ML, Hummer BT, Lau C, Meek B, Peters JM, Popp JA, Rhomberg L, Seed J, Klaunig JE. 2014. Mode of action framework analysis for receptor-mediated toxicity: The peroxisome proliferator-activated receptor alpha (PPARα) as a case study.Crit Rev Toxicol. 44:1-49.

Cramer GM, Ford RA, Hall RL. 1978. Estimation of toxic hazard--a decision tree approach. Food Cosmet Toxicol. 16:255–276.

Daylight Theory: SMARTS - A Language for Describing Molecular Patterns. 2016. Available from: http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html

DeKeyser JG, Shou M, DeKeyser JG, Shou M. 2011. Encyclopedia of Drug Metabolism and Interactions. Hoboken: John Wiley & Sons, Inc. Chapter 5, Species Differences of Drug-Metabolizing Enzymes.

Dertinger SD, Heflich RH. 2011. In vivo assessment of Pig-a gene mutation-recent developments and assay validation. Environ Mol Mutagen. 52:681-4.

Doak SH, Jenkins GJS, Johnson GE, Quick E, Parry EM, Parry JM. 2007. Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. Cancer Res. 67:3904–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17440105

Eastmond DA, Hartwig A, Anderson D, Anwar WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C. 2009. Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS Harmonized Scheme.Mutagenesis 24:341-349.Edler L, Hart A, Greaves P, Carthew P, Coulet M, Boobis A, Williams GM, Smith B. 2014. Selection of appropriate tumour data sets for Benchmark Dose Modelling (BMD) and derivation of a Margin of Exposure (MoE) for substances that are genotoxic and carcinogenic: considerations of biological relevance of tumour type, data quality and uncertai. Food Chem Toxicol. 70:264–289.

EFSA. 2009. Guidance of the Scientific Committee on Use of the benchmark dose approach in risk assessment. EFSA J. 7:1150. Available from: http://doi.wiley.com/10.2903/j.efsa.2009.1150

EFSA. 2012. Scientific Opinion on Exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). EFSA J. 10:2750. Available from: http://doi.wiley.com/10.2903/j.efsa.2012.2750

EFSA. 2016. Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. EFSA Support. Publ. 13. Available from:http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2016.EN-1006/pdf (accessed 21/03/17).

EFSA. 2017. Update: use of the benchmark dose approach in risk assessment. EFSA J. 15:4658. Available from: http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2017.4658/epdf (accessed 22/3/2017).

Elcombe CR, Peffer RC, Wolf DC, Bailey J, Bars R, Bell D, Cattley RC, Ferguson SS, Geter D, Goetz A, et al. 2014. Mode of action and human relevance analysis for nuclear receptor-mediated liver toxicity: A case study with phenobarbital as a model constitutive androstane receptor (CAR) activator. Crit Rev Toxicol. 44:64–82.

Ellison CM, Sherhod R, Cronin MTD, Enoch SJ, Madden JC, Judson PN. 2011. Assessment of methods to define the applicability domain of structural alert models. J Chem Inf Model. 51:975–985.

Embry MR, Bachman AN, Bell DR, Boobis AR, Cohen SM, Dellarco M, Dewhurst IC, Doerrer NG, Hines RN, Moretto A, Pastoor TP, Phillips RD, Rowlands JC, Tanir JY, Wolf DC, Doe JE. 2014. Risk assessment in the 21st century: roadmap and matrix. Crit Rev Toxicol. 44 Suppl 3:6-16.

Enoch SJ, Cronin MTD. 2010. A review of the electrophilic reaction chemistry involved in covalent DNA binding. Crit Rev Toxicol. 40:728–748. Available from: http://www.tandfonline.com/doi/full/10.3109/10408444.2010.494175

Enoch SJ, Cronin MTD. 2012. Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding. Mutat Res Toxicol Environ Mutagen. 743:10–19. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1383571812000083

EPA. 2005. Guidelines for Carcinogen Risk Assessment. Available from: https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf (accessed 27/02/17)

EPA. 2011. Exposure Factors Handbook 2011 Edition (Final). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-09/052F. Available from: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252 (accessed on 27/02/17).

European Chemicals Agency. 2015. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. Helsinki (Finland). Available from: https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf (accessed on 27/02/17).

European Commission. 2002. Commission Decision of 23 January 2002 amending Commission Decision 1999/217/EC as regards the register of flavouring substances used in or on foodstuffs. Off J Eur Comm. 49:1–160.

European Commission. 2003. Guidance document on the assessment of the relevance of metabolites in groundwater of substances regulated under council directive 91/414/eec. Brussels. Available from: https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_ppp_app-proc_guide_fate_metabolites-groundwtr.pdf

European Medicines Agency. 2008. Guideline on the Assessment of Genotoxicity of Herbal Substances/Preparations Adoption by HMPC for Release for Consultation. London. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003569.pdf

Federal Register. 1993. Toxicological principles for the safety assessment of direct food additives and color additives used in food. Fed Regist. 58:16536.

Federal Register. 1995. Food Additives; Threshold of Regulation for Substances Used in Food-Contact Articles. Fed Regist. 60:36582-36596

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

Gocke E, Ballantyne M, Whitwell J, Muller L. 2009a. MNT and MutaMouse studies to define the in vivo dose response relations of the genotoxicity of EMS and ENU. Toxicol Lett. 190:286–297.

Gocke E, Bürgin H, Müller L, Pfister T. 2009b. Literature review on the genotoxicity, reproductive toxicity, and carcinogenicity of ethyl methanesulfonate. Toxicol Lett. 190: 254–265.

Gocke E, Müller L. 2009. *In vivo* studies in the mouse to define a threshold for the genotoxicity of EMS and ENU. Mutat Res Toxicol Environ Mutagen. 678:101–107.

Gold LS, Manley NB, Slone TH, Garfinkel GB, Ames BN, Rohrbach L, Stern BR, Chow K. 1995. Sixth plot of the carcinogenic potency database: results of animal bioassays published in the General Literature 1989 to 1990 and by the National Toxicology Program 1990 to 1993. Environ Health Perspect. 103 Suppl:3–122.

Gold LS, Sawyer CB, Magaw R, Backman GM, de Veciana M, Levinson R, Hooper NK, Havender WR, Bernstein L, Peto R. 1984. A carcinogenic potency database of the standardized results of animal bioassays. Environ. Health Perspect. 58:9–319.

Gold LS, Slone TH, Bernstein L. 1989. Summary of carcinogenic potency and positivity for 492 rodent carcinogens in the carcinogenic potency database. Environ Health Perspect. 79:259–272.

Hanser T, Barber C, Rosser E, Vessey JD, Webb SJ, Werner S. 2014. Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge. J. Cheminform. 6:1.

Health Canada. 2016. Threshold of Toxicological Concern (TTC)-based approach for certain substances (https://www.ec.gc.ca/ese-ees/326E3E17-730A-4878-BC25-D07303A4DC13/HC%20TTC%20SciAD%20EN.pdf – accessed 28/11/2016)

Hernandez LG, van Steeg H, Luijten M, van Benthem J. 2009. Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach. Mutat. Res. 682:94–109.

ICH. 2014. Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk. In: International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Geneva.

Jacobs MN, Colacci A, Louekari K, Luijten M, Hakkert BC, Paparella M, Vasseur P. 2016. International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. ALTEX. 33:359–392.

Joint FAO/WHO Expert Committee on Food Additives. 1996. Toxicological evaluation of certain food additives and contaminants in food. Geneva.

JRC. 2010. Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. EFSA Support. Publ. 7.

Kirchmair J, Göller AH, Lang D, Kunze J, Testa B, Wilson ID, Glen RC, Schneider G. 2015. Predicting drug metabolism: experiment and/or computation? Nat. Rev Drug Discov. 14:387–404.

Kirkland D, Aardema M, Muller L, Makoto H. 2006. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. Mutat Res. 608:29–42.

Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, et al. 2007. How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. Mutat Res. 628:31–55.

Kirkland D, Speit G. 2008. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: III. Appropriate follow-up testing in vivo. Mutat. Res Toxicol Environ Mutagen. 654:114–132.

Kirkland D, Zeiger E, Madia F, Corvi R. 2014. Can *in vitro* mammalian cell genotoxicity test results be used to complement positive results in the Ames test and help predict carcinogenic or *in vivo* genotoxic activity? II. Construction and analysis of a consolidated database. Mutat Res Genet Toxicol Environ Mutagen. 775–776:69–80.

Klaunig JE, Babich MA, Baetcke KP, Cook JC, Corton JC, David RM, Deluca JG, Lai DY, Mckee RH, Peters JM, et al. 2003. PPARα Agonist-Induced Rodent Tumors: Modes of Action and Human Relevance. Crit Rev Toxicol. 33:655–780.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol. 25:1–5.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, Van Schothorst F, Vos JG. 2004. Structure-based Thresholds of Toxicological Concern (TTC): Guidance for Application to Substances Present at Low Levels in the Diet. Food Chem Toxicol. 42:65–83.

Lehr RE, Kumar S, Levin W, Wood AW, Chang RL, Conney AH, Yagi H, Sayer JM, Jerina DM. 1985. The Bay Region Theory of Polycyclic Aromatic Hydrocarbon Carcinogenesis. In: Polycyclic Hydrocarbons and Carcinogenesis. p. 63–84. Available from: http://pubs.acs.org/doi/abs/10.1021/bk-1985-0283.ch004

Luijten M, Olthof ED, Hakkert BC, Rorije E, van der Laan J-W, Woutersen RA, van Benthem J. 2016. An integrative test strategy for cancer hazard identification. Crit Rev Toxicol. 46:615–639.

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. 2003. A Framework for Human Relevance Analysis of Information on Carcinogenic Modes of Action. Crit Rev Toxicol. 33:591–653. Available from: http://www.tandfonline.com/doi/full/10.1080/713608373

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J, Vickers C. 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol. 34:1–18.

Mekenyan O, Dimitrov S, Serafimova R, Thompson ED, Kotov S, Dimitrova N, Walker JD. 2004. Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. Chem Res Toxicol. 17:753–766.

Mekenyan O, Todorov M, Serafimova R, Stoeva S, Aptula A, Finking R, Jacob E. 2007. Identifying the structural requirements for chromosomal aberration by incorporating molecular flexibility and metabolic activation of chemicals. Chem Res Toxicol. 20:1927–1941.

Munro IC. 1990. Safety assessment procedures for indirect food additives: an overview. Report of a workshop. Regul Toxicol Pharmacol. 12:2–12.

Munro IC, Ford RA, Kennepohl E, Sprenger JG. 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem Toxicol. 34:829–867.

Munro IC, Kennepohl E, Kroes R. 1999. A procedure for the safety evaluation of flavouring substances. Joint FAO/WHO Expert Committee on Food Additives. Food Chem. Toxicol. 37:207–232.

OECD. 2009a. Test No. 451: Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4. Paris, France: Organization for Economic Cooperation and Development.

OECD. 2009b. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies. OECD Publishing (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: Organization for Economic Cooperation and Development. Available from: http://www.oecd-ilibrary.org/environment/test-no-453-combined-chronic-toxicity-carcinogenicity-studies_9789264071223-en

Parry JM, Fielder RJ, McDonald A. 1994. Thresholds for aneuploidy-inducing chemicals. Advisory Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment of the UK Department of Health. Mutagenesis 9:503–504.

Preston RJ, Williams GM. 2005. DNA-reactive carcinogens: mode of action and human cancer hazard.Crit Rev Toxicol. 35:673–683.Rhomberg LR, Baetcke K, Blancato J, Bus J, Cohen S, Conolly R, Dixit R, Doe J, Ekelman K, Fenner-Crisp P, et al. 2007. Issues in the design and interpretation of chronic toxicity and carcinogenicity studies in rodents: approaches to dose selection. Crit Rev Toxicol. 37:729–837.

Proctor DM, Gatto NM, Hong SJ, Allamneni KP. 2007. Mode-of-action framework for evaluating the relevance of rodent forestomach tumors in cancer risk assessment Toxicol Sci. 98:313-326.

Rulis AM. 1987. De minimis and the threshold of regulation. In: Felix CW, editor. Food Protection Technology. Chelsea: Lewis Publishing. p. 29–37.

SCCS, SCHER S. 2012. Joint Opinion on the Use of the Threshold of Toxicological Concern (TTC) Approach for Human Safety Assessment of Chemical Substances with focus on Cosmetics and Consumer Products. Brussels. Available from:

http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_092.pdf

SCCS NfG. 2016. The SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation. Paper presented at: 11th plenary meeting of 29 September 2015; Brussels, Belgium.

Serafimova R, Todorov M, Pavlov T, Kotov S, Jacob E, Aptula A, Mekenyan O. 2007. Identification of the structural requirements for mutagencitiy, by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model. Chem Res Toxicol. 20:662–676.

Slob W. 2014. Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. Crit Rev Toxicol. 44:557-567.

Tropsha A, Tropsha A. 2010. Recent Advances in Development, Validation, and Exploitation of QSAR Models. In: Burger's Medicinal Chemistry and Drug Discovery. Hoboken, NJ, USA: John Wiley & Sons, Inc. Available from: http://doi.wiley.com/10.1002/0471266949.bmc002.pub2

Votano JR, Parham M, Hall LH, Kier LB, Oloff S, Tropsha A, Xie Q, Tong W. 2004. Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis 19:365–377.

Weisburger JH, Williams GM. 1981. Carcinogen testing: current problems and new approaches. Science. 214:401–407.

Williams RV, Amberg A, Brigo A, Coquin L, Giddings A, Glowienke S, Greene N, Jolly R, Kemper R, O'Leary-Steele C, Parenty A, Spirkl HP, Stalford SA, Weiner SK, Wichard J. 2016. It's difficult, but important, to make negative predictions.Regul Toxicol Pharmacol. 6:79-86.

WHO. 1999. Safety Evaluation of Certain Food Additives. WHO Food Additives Series, No, 42. Geneva, Switzerland: World Health Organization.

Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, et al. 2015. New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling. J. Chem. Inf. Model. 55:510–528. Available from: http://pubs.acs.org/doi/abs/10.1021/ci500667v

Table 1. Genotoxicity profiles of Ames-positive carcinogens (taken from the approaches used in compiling the database of Kirkland et al. (2014) that are considered the strongest candidates to exert a genotoxic mode of action for carcinogenicity in descending order of confidence (Group1 strongest, Group 6 weakest evidence).In vivo effects in the target tissue (and target species) for carcinogenicity would carry particular weight.

Group	*In vivo* mammalian mutation	*In vivo* mammalian cytogenicity	*In vivo* mammalian UDS[a] or comet		*In vitro* mammalian cell mutation	*In vitro* mammalian cell cytogenicity
1	+	+	+ or NA[a]		+	+
	+	+ or NA	+		+	+
	+ or NA	+	+		+	+
2	+	+	+ or NA		+	NA
	+	+ or NA	+		+	NA
	+ or NA	+	+		+	NA
	+	+	+ or NA		NA	+
	+	+ or NA	+		NA	+
	+ or NA	+	+		NA	+
3	+	+	+ or NA		NA	NA
	+	+ or NA	+		NA	NA
	+ or NA	+	+		NA	NA
4	+	NA	NA		+	+

	NA	+	NA		+	+
	NA	NA	+		+	+
5	+	NA	NA		+	NA
	+	NA	NA		NA	+
	NA	+	NA		+	NA
	NA	+	NA		NA	+
	NA	NA	+[c]		+	NA
	NA	NA	+[c]		NA	+
6	+	NA	NA		NA	NA
	NA	+	NA		NA	NA
	NA	NA	+[c]		NA	NA

+ Evidence of genotoxicity using accepted criteria for the test

[a] Some chemicals that gave negative results in liver UDS tests but were positive in the *in vivo* comet assay were included since the liver UDS test has been shown to be insensitive to a number of carcinogens (Kirkland and Speit 2008)

[b] NA: Result not available, i.e. test was not performed

[c] Care should be taken to evaluate the quality and source of the data if the only *in vivo* positive result is in a comet assay

Table 2.A summary of freely available sets of rules, or profilers, which may be used to identify genotoxic carcinogens from the OECD QSAR Toolbox.

Profiler / Rulebase	**Details**	**Reference, if applicable**
Protein binding alerts for chromosomal aberration by OASIS v1.1	28 structural alerts accounting for interactions of chemicals with specific proteins, such as topoisomerases, cellular protein adducts, etc. The scope of this profiler is to investigate the ability of target molecules to elicit clastogenicity and aneugenicity	Mekenyan et al. 2007
DNA binding by OASIS v.1.3	Ames mutagenicity model (part of the OASIS TIMES system – see Table 3) with 78 structural alerts for interaction with DNA	Mekenyan et al. 2004; Serafimova et al. 2007
DNA binding by OECD	60 mechanistic organic chemistry fragments (in the form of structural alerts) for the binding of organic compounds to DNA.	Enoch and Cronin 2010; Enoch and Cronin 2012
Carcinogenicity (genotox and nongenotox) alerts by ISS	A decision tree for estimating carcinogenicity, based on a list of 55 structural alerts (SAs). Of the alerts, 35 derive from the Toxtree module and 20 were derived separately. Most of the new SAs are relative to non- genotoxic	

	carcinogenicity, whereas the SAs in the initial list mainly coded genotoxic carcinogenicity	
in vitro mutagenicity (Ames test) alerts by ISS	Mutagenicity/Carcinogenicity module from Toxtree comprising a list of 30 structural alerts (SAs).	
in vivo mutagenicity (micronucleus) alerts by ISS	ToxMicrulebase from the Toxtree software comprising 35 structural alerts (SAs) for a preliminary screening of potentially in vivo mutagens.	
OncoLogic Primary Classifier	Molecular definitions developed to mimic the structural criteria of chemical classes of potential carcinogens covered by the U.S. Environmental Protection Agency's OncoLogic™ Cancer Expert System for Predicting the Carcinogenicity Potential	
ChemoTyper (Altamira/ Molecular Networks)	Ashby Tennant genotoxic carcinogenic rules implemented in ChemTyper. This tool was implemented by a contract from US FDA CFSAN.	Ashby and Tennant 1991; Ashby 1994; Yang et al. 2015

1

Table 3.A summary of commercially available expert systems based on structural alerts for the identification of genotoxic carcinogens.

System and Supplier	**Endpoint**	**Reference / Link**
DEREK Nexus from Lhasa Ltd	Various rulebases for mutagenicity including bacterial ….. in vitro cytogenicity in mammalian cells, in vitro micronucleus studies	https://www.lhasalimited.org/derek_nexus/
ToxAlert, HazardExpert (CompuDrug)	Mutagenicity	http://www.compudrug.com/
Case Ultra MultiCASE	Mutagenicity	http://www.multicase.com/
Genetox Expert Alerts Suite (Leadscope)	Genetic Toxicity	http://www.leadscope.com/genetox_expert_alerts/
ChemTunesToxGPS (Molecular Networks/Altamira)	Rule-base for mutagenicity, in vitro chromosome aberration, and micronucleus (WOE prediction)	https://www.mn-am.com/products/chemtunes_studio

1 Table 4.A summary of computational systems to predict mutagenicity and related

2 endpoints based on SAR, QSAR or hybrid systems.

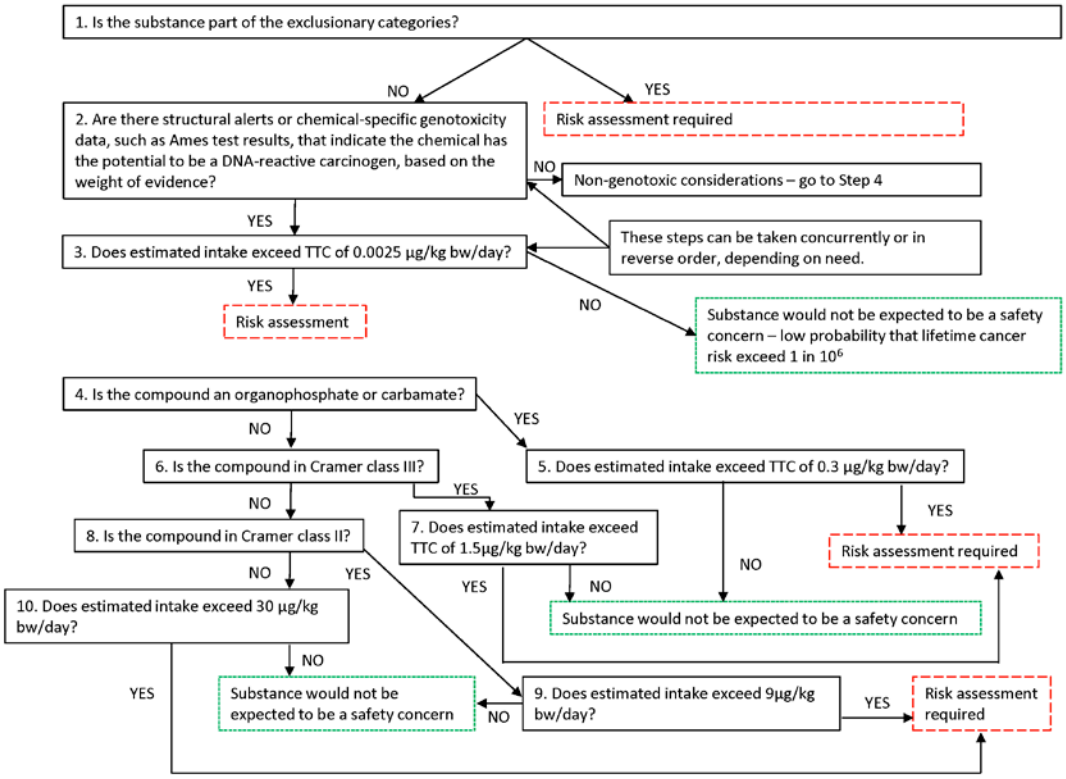
System and Supplier	**Endpoint(s)**	**Type of systems**	**Reference / Link**
ToxRead (Mario Negri Institute, Milan, Italy)	Ames mutagenicity	QSAR (chemical similarity) – freely available	http://www.toxgate.eu/
VEGA (Mario NegriInstitute, Milan, Italy)	Ames mutagenicity	Hybrid (ToxTree) rules and QSAR – freely available	http://www.vega-qsar.eu/
Lazar	Ames mutagenicity	QSAR – freely available	http://lazar.in-silico.de
Toxicity Estimation Software Tool (TEST) (US EPA)	Ames mutagenicity	QSAR – freely available	https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
Tox Suite (ACD Labs)	Ames mutagenicity	QSAR - commercial	http://www.acdlabs.com/products/pc_admet/tox/tox/
Non-human Genetic Toxicity Suite (Leadscope)	*Salmonella* and*E. coli* mutagenicity, mouse lymphoma, *in vitro* chromosome	QSAR - commercial	http://www.leadscope.com/genetic_toxicity_suite/

	aberrations, *in vivo* micronucleus		
Multicase MCASE/MC4PC (MultiCASEInc)	Many endpoints relating to mutagenicity and genotoxicity	QSAR - commercial	http://www.multicase.com/
OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University)	mutagenicity, chromosomal aberrations and micronucleus formation including metabolism prediction	Hybrid rules and QSAR - commercial	http://oasis-lmc.org/?section=software&swid=4
PASS (Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences)	Various mutagenicity	Hybrid rules and QSAR - commercial	http://www.pharmaexpert.ru/PASSOnline/
MolCode Toolbox	Mutagenicity	QSAR - commercial	http://molcode.com/
OpenTox	Micronucleus	QSAR – freely available	http://apps.ideaconsult.net:8080/ToxPredict

TOPKAT (Accelrys)	Ames mutagenicity	QSAR - commercial	http://accelrys.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html
ChemTunesToxGPS (Molecular Networks/Altamira)	Ames Mutagenicity; *in vitro* chromosome aberration; *in vivo* micronucleus	QSAR (WOE prediction) – commercial	https://www.mn-am.com/products/chemtunes_studio
Sarah Nexus (Lhasa Ltd)	Mutagenicity	QSAR – commercial	Hanser et al. 2014

1

2



1

2 Figure 1.Tiers of the TTC concept as described in the WHO EFSA Report 2016 (EFSA,
3 2016). For explanation of the “exclusionary categories” see section on “Outline of
4 current TTC scheme”. The green broken-line boxes indicate where a conclusion on the
5 acceptability of estimated human exposure can be reached. The red broken-line boxes
6 indicate where safety at estimated human exposure cannot be assured and further
7 information would be necessary to enable completion of the assessment.