



The safety evaluation of food flavoring substances: the role of genotoxicity studies

Nigel J. Gooderham, Samuel M. Cohen, Gerhard Eisenbrand, Shoji Fukushima, F. Peter Guengerich, Stephen S. Hecht, Ivonne M. C. M. Rietjens, Thomas J. Rosol, Maria Bastaki, Matthew J. Linman & Sean V. Taylor

To cite this article: Nigel J. Gooderham, Samuel M. Cohen, Gerhard Eisenbrand, Shoji Fukushima, F. Peter Guengerich, Stephen S. Hecht, Ivonne M. C. M. Rietjens, Thomas J. Rosol, Maria Bastaki, Matthew J. Linman & Sean V. Taylor (2020) The safety evaluation of food flavoring substances: the role of genotoxicity studies, *Critical Reviews in Toxicology*, 50:1, 1-27, DOI: [10.1080/10408444.2020.1712589](https://doi.org/10.1080/10408444.2020.1712589)

To link to this article: <https://doi.org/10.1080/10408444.2020.1712589>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 12 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 395



View related articles [↗](#)



View Crossmark data [↗](#)

The safety evaluation of food flavoring substances: the role of genotoxicity studies

Nigel J. Gooderham^a , Samuel M. Cohen^b , Gerhard Eisenbrand^c , Shoji Fukushima^d ,
F. Peter Guengerich^e , Stephen S. Hecht^f , Ivonne M. C. M. Rietjens^g , Thomas J. Rosol^h ,
Maria Bastakiⁱ , Matthew J. Linmanⁱ  and Sean V. Taylorⁱ 

^aDepartment of Surgery and Cancer, Imperial College, London, UK; ^bHavlik-Wall Professor of Oncology, Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA; ^cFood Chemistry & Toxicology, University of Kaiserslautern (retired), Heidelberg, Germany; ^dJapan Bioassay Research Center, Kanagawa, Japan; ^eDepartment of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, USA; ^fMasonic Cancer Center and Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA; ^gDivision of Toxicology, Wageningen University, Wageningen, The Netherlands; ^hDepartment of Biomedical Sciences, Heritage College of Osteopathic Medicine, Ohio University, Athens, OH, USA; ⁱFlavor and Extract Manufacturers Association, Washington, DC, USA

ABSTRACT

The Flavor and Extract Manufacturers Association (FEMA) Expert Panel relies on the weight of evidence from all available data in the safety evaluation of flavoring substances. This process includes data from genotoxicity studies designed to assess the potential of a chemical agent to react with DNA or otherwise cause changes to DNA, either *in vitro* or *in vivo*. The Panel has reviewed a large number of *in vitro* and *in vivo* genotoxicity studies during the course of its ongoing safety evaluations of flavorings. The adherence of genotoxicity studies to standardized protocols and guidelines, the biological relevance of the results from those studies, and the human relevance of these studies are all important considerations in assessing whether the results raise specific concerns for genotoxic potential. The Panel evaluates genotoxicity studies not only for evidence of genotoxicity hazard, but also for the probability of risk to the consumer in the context of exposure from their use as flavoring substances. The majority of flavoring substances have given no indication of genotoxic potential in studies evaluated by the FEMA Expert Panel. Examples illustrating the assessment of genotoxicity data for flavoring substances and the consideration of the factors noted above are provided. The weight of evidence approach adopted by the FEMA Expert Panel leads to a rational assessment of risk associated with consumer intake of flavoring substances under the conditions of use.

ARTICLE HISTORY



Received 13 August 2019
Revised 29 December 2019
Accepted 3 January 2020

KEYWORDS

FEMA GRAS; genotoxicity; flavors; flavoring substance; risk assessment; weight of evidence; DNA adducts; mutagenicity; chromosomal damage; safety assessment

Table of contents

1. Introduction	2	4.1. Study quality	9
2. Regulatory approaches in the evaluation of genotoxicity information	3	4.2. Biological relevance	10
2.1. The FEMA Expert Panel approach to the genotoxicity evaluation of flavoring substances	3	4.2.1. Cytotoxicity	10
2.2. The JECFA approach to genotoxicity evaluation of flavoring substances	4	4.2.2. Dose or concentration–response	11
2.3. The EFSA approach to genotoxicity evaluation of flavoring substances	4	4.2.3. Variability and reproducibility	11
3. Genotoxicity datasets reviewed by the FEMA Expert Panel	5	4.2.4. Functional DNA repair systems	11
3.1. Genotoxicity data packages for flavoring substances	5	4.2.5. Assay performance in predicting possible human genotoxic hazard	11
4. Interpretation of genotoxicity data in FEMA GRAS evaluations	9	4.3. Human relevance	12
		4.3.1. Metabolic activation	12
		4.3.2. Mode of action	13
		4.3.2.1. DNA adducts	14
		4.3.2.1.1. DNA adducts, genotoxicity, and carcinogenicity	14
		4.3.2.1.2. Endogenous DNA adducts	15

CONTACT Sean V. Taylor  staylor@vertosolutions.net  Scientific Secretary to the FEMA Expert Panel, Flavor and Extract Manufacturers Association, 1101 17th Street, NW, Suite 700, Washington, DC 20036, USA

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

4.3.2.2. Methyl eugenol and FEMA expert panel decision to remove it from the FEMA

GRAS list	15
5. The weight of evidence and human relevance of genotoxicity testing findings	16
5.1. Sufficient evidence of genotoxicity	17
5.2. Insufficient evidence of genotoxicity	17
5.3. Consideration of carcinogenicity studies	19
6. Conclusions	20
Notes	21
Acknowledgments	21
Declaration of interest	21
ORCID	22
References	23

1. Introduction

Humans have been producing nutritious and appealing foods for thousands of years by taking basic ingredients – meats, fish, and vegetables – and curing, drying, boiling, frying, or roasting to make them edible and safe to store. One critical factor in making these foods appealing, and in some cases improving safety, has been the use of culinary enhancers, including herbs, spices, and other ingredients to impart flavor. More recently, technology has expanded the application and range of flavors far beyond basic cooking processes. Chemically defined flavorings can be isolated from natural sources or created *de novo*. Flavors can be formulated from these and also from naturally-derived essential oils, extracts, and other complex materials.

It was recognized more than 60 years ago that the safety evaluation of all flavorings, regardless of their source, was an essential element of ensuring the safety of flavored foods. In the USA, the safety evaluation of flavoring substances is based on the concept of “Generally Recognized As Safe under intended conditions of use” (GRAS) as implemented in the Food Additive Amendment of 1958 to the Federal Food, Drug and Cosmetic (FD&C) Act. Within the GRAS regulatory framework, the Flavor and Extract Manufacturers Association (FEMA), a US-based trade association, brought together the FEMA Expert Panel, a group of scientists qualified by training and experience to conduct scientifically independent evaluations of the safety of food flavoring substances (Hallagan and Hall 1995, 2009). Substances that hold FEMA GRAS status are listed in regular publications that are authored by the FEMA Expert Panel (GRAS 3-GRAS 28), with the most recent update published in 2017 (Cohen et al. 2017b) and their conditions of intended use are described therein. Additionally, the FEMA Expert Panel has published 16 safety evaluation updates on specific groups of flavoring substances (re-evaluations of FEMA GRAS flavoring substances) and 17 reviews on flavorings and issues relevant to their safety assessment including several recent additions (Cohen et al. 2018; Smith et al. 2018).

The FEMA Expert Panel applies safety standards required by the United States Food and Drug Administration (US FDA) (which are also utilized by other national, regional, and

international expert bodies) in evaluating the risk that a potential flavoring substance may pose to consumers under the conditions of use. The criteria used by the FEMA Expert Panel to assess the safety of flavoring substances for the consumers have been previously described in detail (Smith et al. 2005). In essence, the Panel follows the three elements of the well-established risk assessment paradigm: hazard identification and characterization, exposure assessment, and risk characterization.

First, hazard identification and characterization consider the identity of the substance and its physicochemical and biological properties, its metabolic fate, and its toxicity profile. The hazard characterization provides dose–response data for standard hazard metrics to enable definition of points of departure (PODs), such as no-observed-adverse-effect-levels (NOAELs), or benchmark dose (BMD) values, and any appropriate uncertainty factors or other similar adjustments based on a review of the entire database.

Second, an exposure assessment incorporates the conditions of use such as consumer food intake, levels/patterns, and range of use levels of flavoring substance in foods for the populations of interest.

The third and final step integrates the information arising from the hazard identification and characterization and the exposure assessment to conclude upon the safety of the flavoring substance under conditions of use by determining the relationship between the level of consumer intake and the applicable thresholds of toxicological concern (TTC), or other PODs. Additionally, the relevance of possible hazards identified *in vitro* or *in vivo* studies to human safety is assessed by considering the validity of such studies, the mode of action (MOA) for any effects observed, and relevant species differences between humans and the animals utilized in the studies.

Since new data and methods continue to become available and the possible consumer exposure may change, the FEMA Expert Panel also performs periodic reevaluations of the safety of flavoring substances. Within these reevaluations, all additional relevant information is reviewed and assessed. The FEMA Expert Panel considers any new data along with the previously available data and updates its safety conclusions accordingly.

While it might seem ideal for all substances to be exhaustively tested for any potential adverse outcomes, this is neither practical for a variety of reasons (e.g. material availability, time, costs) nor is it scientifically necessary, and hence not justified under the imperative to replace, reduce, and/or refine (3Rs) animal testing. Like other expert bodies that conduct safety evaluations, the FEMA Expert Panel has adopted a pragmatic approach for toxicity assessment that relies on clustering flavoring substances into congeneric groups based on chemical structural similarity (i.e. similar structural frame and shared functional groups) and similar anticipated metabolic outcomes. Therefore, the GRAS assessment performed by the FEMA Expert Panel includes a thorough evaluation of all the available data for the candidate flavoring substances as well as for structurally related substances that can be considered as part of the same chemical group.¹ Available information relevant to the absorption, distribution, metabolism and excretion of the flavoring and structurally related substances provide the basis for understanding the biochemical fate of the substance. Particular

attention is given to the generation of potentially toxic metabolites as opposed to innocuous products. Data from short-term and long-term oral administration studies of the flavoring or structurally related substances provide a fundamental basis to understand the toxic potential of the substance and the potential tissue or cellular targets, including DNA. Where available or considered necessary, specific toxicities are also evaluated by considering pathological, behavioral, neurotoxicity, immunotoxicity, developmental, and reproductive toxicity data.

In this paper, the FEMA Expert Panel describes its approach to the consideration of one aspect of toxicity – the potential for a substance to react with DNA and/or otherwise alter its function – which is commonly referred to as genotoxic potential. Herein the Panel describes its consideration of genotoxicity data within the evaluation of safety for a flavoring substance. Of note, the consideration of genotoxic potential is but one factor that is incorporated along with others into a comprehensive safety evaluation of a flavoring substance, including those flavorings that have not yet attained FEMA GRAS status as well as those that are already in the market and undergoing reevaluation for continued GRAS status.

2. Regulatory approaches in the evaluation of genotoxicity information

For the safety assessment of foods and food ingredients, the relevant national or regional agencies include the US FDA, the European Food Safety Authority (EFSA), Health Canada, the Food Standards Australia New Zealand (FSANZ) agency, the Japanese Ministry of Health, Labor and Welfare (JMHLW), and the Chinese National Center for Food Safety Risk Assessment (CFSA), among others. Additionally, the Joint FAO/WHO Expert Committee on Food Additives (JECFA) is a widely recognized expert body that provides scientific advice to the Codex Alimentarius Commission. JECFA's safety evaluations are broadly recognized by numerous regulatory bodies.

Although the outcome of JECFA's evaluations does not have any direct bearing on the regulatory approval of use of a food additive in any specific country, its evaluations are widely recognized and may affect an application for approval for a new food additive in a particular country.

The above and other regulatory agencies and the evaluation bodies within them utilize genotoxicity testing batteries that include complementary *in vitro* and *in vivo* assays to assess different modes of genotoxic potential. In general, scientific expert bodies agree that the purpose of genotoxicity testing of substances in food is

- To identify substances that have the potential to cause genetic damage in humans,
- To predict potential genotoxic carcinogens in cases where carcinogenicity data are not available, and
- To contribute to an understanding of the mode of action of chemical carcinogens.

The default position for some regulatory bodies, in those cases where genotoxicity testing has provided strong evidence of confirmed positive genotoxic potential, is that there is no acceptable level of exposure. For other regulatory bodies, a consideration of the genotoxicity data, as well as the potential exposure and possible mode of action, are used to make an assessment of genotoxic risk. In either case, some understanding of genotoxic potential is generally considered essential for completing the safety evaluation of a putative food ingredient and ultimately developing a conclusion as to whether it should be allowed for use in foods. Since the 1980s, and regardless of the level of precaution that the regulatory agency applies, a tiered approach to genotoxicity testing has been favored. In this tiered approach, prior to testing, a consideration of the chemical structure and resulting possible alerts for genotoxicity are considered. Appropriate *in vitro* genotoxicity studies are conducted as considered necessary and *in vivo* studies are conducted as follow-up testing in the case of positive results in the *in vitro* testing. One notable exception to this approach is the EU directive for cosmetics testing, which mandates that to comply with EU legislation no animal testing of cosmetic products can be performed (EU 2009).

Regulatory and other expert bodies around the world have been using read-across and weight-of-evidence approaches that incorporate considerations of all relevant data on the substance and/or structurally related substances being evaluated. Such data can provide important context when drawing conclusions about the relevance of the results of genotoxicity studies. This context can include the known or anticipated chemical reactivity (related to site-of-contact impacts, such as local inflammation), bioavailability, metabolism, toxicokinetics, target tissue(s) exposure, and target organ specificity.

2.1. The FEMA Expert Panel approach to the genotoxicity evaluation of flavoring substances

The FEMA Expert Panel's philosophy and general approach to the safety evaluation of flavoring substances have been described in the context of its criteria for the safety evaluation of chemically-defined substances (Smith et al. 2005) and of natural flavor complexes (Smith et al. 2005; Cohen et al. 2018). Although these criteria do not prescribe a specific battery of genotoxicity tests, the FEMA Expert Panel considers genotoxic potential to be a critical element that must be adequately addressed before a safety conclusion can be reached.

Genotoxicity testing, as with other toxicity testing, can provide information relevant to the hazard potential of the tested substance. For the FEMA Expert Panel, a genotoxic risk to the consumer is determined not purely by an inherent ability of a substance to interact with DNA under testing conditions (i.e. identification of a potential hazard) but also by evaluating the likelihood that such an event is manifested in an *in vivo* functional phenotype and whether that is likely to be a human-relevant risk. Theoretically, genetic damage poses a safety concern only if (a) interaction with genetic material is likely to occur *in vivo*; (b) the genetic interaction,

which is a stochastic event, occurs at a relevant genetic locus in a coding or otherwise functional DNA sequence (rather than as a silent DNA modification); (c) repair is insufficient (DNA repair capacity is exceeded); and (d) the phenotype of the genetic damage has biological consequences (i.e. leads to cancer, germ cell damage, or other cell/tissue disruption) (Vogelstein et al. 2013; Klapacz et al. 2016; Liu et al. 2016; Basu 2018).

In this light, a genotoxic risk is defined as the combination of the hazard inherently associated with a substance and the conditions necessary for the functional expression of that hazard, which is dose-dependent. Therefore, the FEMA Expert Panel conducts a complete assessment of probable risk rather than merely a hazard assessment limited to the intrinsic genotoxic potential (hazard) of flavoring substances. This can involve the use of an appropriate TTC value for genotoxic potential, currently considered as 0.15 µg/person/day (Kroes et al. 2004; Boobis et al. 2017).

In a recent publication, describing its updated procedure for the safety evaluation of natural complex mixtures used as flavoring substances, the FEMA Expert Panel incorporated the TTC concept for compounds that are potentially genotoxic (Cohen et al. 2018). Within that publication, the Panel's approach to the consideration of the genotoxic potential of known and unidentified compounds is described. The updated procedure acknowledges that some constituents of natural complex mixtures, whether identified or unidentified, may possess genotoxic potential and determines whether that potential poses appreciable genotoxicity risk to the consumer, when test data are not available. The TTC for evaluation of genotoxicity risk (TTC_{genotox}) of 0.15 µg/person/day was proposed by Kroes and colleagues (Kroes et al. 2004) as the dose below which cancer risk does not exceed 1 in 10^6 , specifically for compounds that have structural alerts for genotoxicity other than those of highly potent carcinogens, such as aflatoxin, certain azo- and azoxy-compounds or *N*-nitroso-compounds, for which no threshold can be determined. The TTC_{genotox} is 10-fold lower (more stringent) than the threshold of regulation (TOR) for cancer risk previously established for substances with no indication of DNA reactivity [for details, see (Kroes et al. 2004; Boobis et al. 2017; Patlewicz et al. 2018)]. The application of the TTC_{genotox} is consistent with a risk assessment approach rather than a strict hazard evaluation (EFSA 2016; Nohmi 2018). In the absence of test data, the safety evaluation procedure for flavoring substances proposes that intake below the TTC_{genotox} presents negligible concern for genotoxicity.

2.2. The JECFA approach to genotoxicity evaluation of flavoring substances

To date, JECFA has evaluated over 2200 flavoring substances that are used globally. JECFA reviews data for flavorings in groups of structurally similar substances ("JECFA group"). The JECFA flavoring groups undergo evaluation using a Procedure for the Safety Evaluation of Flavoring Agents, and the data and resulting conclusions are published in flavor monographs (Food Additive Series No. 40-73).² While within

the procedure applied from 1997-2016 there were no systematic approaches to the consideration of genotoxic potential nor any explicit requirements for genotoxicity data, JECFA has applied a weight-of-evidence approach to incorporate all available information. This process includes data from genotoxicity and toxicity studies, as well as established or expert knowledge on metabolism and chemical reactivity. JECFA requested additional toxicity data in some cases, including genotoxicity data, in order to have a sufficient data set upon which it could base its weight-of-evidence conclusions.

In 2016, JECFA revised its procedure for the safety evaluation of flavoring substances by incorporating consideration of genotoxicity alerts and available data as the first step before consideration of other available information (JECFA 2016). This gives priority to an assessment of genotoxic potential prior to completing the full safety evaluation through the JECFA procedure. Notably, the updated JECFA procedure does not simply assess whether a flavoring substance has given positive results in *in vitro* or *in vivo* genotoxicity studies; rather, it works to reach a conclusion as to whether the substance is anticipated or demonstrated to be a DNA-reactive carcinogen. The first JECFA evaluations of flavorings that utilize this new procedure were conducted in June 2018, and the detailed reports that describe how JECFA has applied this approach were recently published (JECFA 2019).

2.3. The EFSA approach to genotoxicity evaluation of flavoring substances

To date, EFSA has evaluated the safety of flavorings in the European market by subdividing them into 34 groups according to their chemical structure, with a chemical group designation (EC CG 1-34) as defined by the European Commission (Regulation (EC) No. 1565/2000; Annex I).³ Out of those initial 34 main groups, 28 subgroups of flavorings with α,β -unsaturated carbonyl moieties were formed and evaluated separately for genotoxicity prior to further safety evaluation. Testing, if required, was performed on specified representative substances of each subgroup (EFSA 2008a) according to EFSA's published test strategy (EFSA 2008b).

EFSA has prescribed a systematic and step-wise approach for the generation and evaluation of data on genotoxic potential (EFSA 2011). This approach relies upon:

- a battery of *in vitro* tests that cover mutagenicity and chromosomal damage endpoints; this battery includes the bacterial reverse mutation test (OECD 1997) and an *in vitro* mammalian cell micronucleus test (OECD 2016b);
- consideration of whether specific structural features of the test substance or test conditions might require additional testing beyond the recommended *in vitro* tests (i.e. by other *in vitro* or *in vivo* tests in the basic battery);
- additional considerations in the event of positive results from the basic *in vitro* battery, including a careful review of the data and the test substance;
- where necessary, an appropriate follow-up *in vivo* study (or studies) to assess whether the genotoxic potential

observed *in vitro* is expressed *in vivo*. For instance, the *in vivo* comet assay (OECD 2014) is an indicator assay that is considered as an appropriate follow-up test to resolve equivocal or positive *in vitro* mutagenicity or chromosomal damage tests, along with the transgenic rodent mutation assay (OECD 2013).

The EFSA Scientific Committee has recently published updated guidance on the interpretation of genotoxicity testing data (Hardy et al. 2017). Part of the scope of that publication was to provide clarity and transparency on the rationale and application of the weight-of-evidence approach in the interpretation of genotoxicity data. Drawing on the previously published EFSA Scientific Opinion on genotoxicity testing strategies (EFSA 2011) and Guidance on the use of the Weight of Evidence approach in scientific assessments (EFSA 2017), EFSA's weight-of-evidence approach to genotoxicity assessment includes assembling, weighing, and assessing data quality and availability on genotoxicity itself and any other relevant data within the overall hazard assessment. EFSA emphasized consideration of uncertainties in the scientific assessments, including clear and unambiguous identification of the sources of uncertainty and their impact on the assessment outcome. EFSA considers uncertainty assessment directly relevant to cases where, based on the available *in vitro* and *in vivo* results from the standard battery of genotoxicity assays, it is not possible to conclude on the absence of genotoxicity with confidence (standard or preferred battery of tests is not available or results *in vitro* and *in vivo* are inconsistent). In these cases, EFSA considers all available data that may reduce the uncertainty, such as mode of action, results of carcinogenicity studies, reproductive toxicity, toxicokinetic studies, read-across from structurally related substances and predictions from quantitative structure-activity relationship (QSAR) models, and reliable data from non-standard tests/endpoints (e.g. presence of DNA adducts). If despite all lines of available evidence, it is still not possible to conclude on the genotoxicity, EFSA would require additional data to reduce the uncertainty before concluding on the genotoxic potential of a flavoring substance.

3. Genotoxicity datasets reviewed by the FEMA Expert Panel

Screening genotoxicity tests originally emerged as surrogates for the expensive and resource-intensive rodent bioassay based on the premise that indication of DNA damage can be a predictor of carcinogenicity, while they had the additional advantage of requiring less time to conduct and fewer resources than cancer bioassays. Currently however, these screening tests are often employed for the evaluation of genotoxicity as an endpoint in itself. Starting with the *Salmonella typhimurium* reverse mutation assay, known as the Ames assay or bacterial reverse mutation assay, other variant mutation assays were developed in mammalian cells that incorporated the complexity of chromosomal organization and assess mutations at specific gene loci (usually *tk* and *hprt*) and chromosomal damage. Additionally, *in vivo* genotoxicity

assays in rodents were soon developed. In all cases, the output is an indication of the potential of substances or their metabolites to react or interact directly with DNA. Although the results of such tests do not directly address the carcinogenic potential of a substance, they provide indicative information to determine whether further assessment may be necessary to address such a concern.

It is generally agreed upon that genotoxic activity can be due to multiple possible mechanisms and a battery of complementary tests is often used in combination with expert judgment, structural alert systems, or other relevant data to derive conclusions about genotoxic potential. To address both the possibility of mutagenicity (i.e. DNA damage resulting in irreversible and/or heritable changes to the genetic sequence of an organism) and other genotoxic effects (such as single or double-strand DNA breaks, DNA cross-linking, or structural or numerical chromosomal damage), several genotoxicity assays have been developed. Some of these assays have undergone validation and test guidelines for their proper conduct have been published by the Organization for Economic Cooperation and Development (OECD) (OECD 2017). Although the OECD testing guidelines (TG) for some of the older assays have been deleted when their utility and validity were determined to be insufficient,⁴ currently published OECD testing guidelines still include some older tests that are no longer considered reliable, including for example the mouse heritable translocation assay (TG 485) due to the number of animals required and the unscheduled DNA synthesis (UDS) test with mammalian liver cells *in vivo* which does not respond to all types of DNA damage (OECD 2017). Current genotoxicity OECD guidelines in effect today include: the bacterial reverse mutation test (Ames test, *vide supra*) (TG 471), *in vitro* mammalian chromosomal aberration test (TG 473), mammalian erythrocyte micronucleus test (TG 474); mammalian bone marrow chromosomal aberration test (TG 475); *in vitro* mammalian cell gene mutation test using the *hprt* or *xprt* locus (TG 476); rodent dominant lethal assay (TG 478); mammalian spermatogonial chromosome aberration test (TG 483); mouse heritable translocation assay (TG 485); unscheduled DNA synthesis test with mammalian liver cells *in vivo* (TG 486); *in vitro* mammalian cell micronucleus test (TG 487); transgenic rodent somatic and germ cell gene mutation assays (TG 488); *in vivo* alkaline comet assay (TG 489) and *in vitro* mammalian cell gene mutation tests using the thymidine kinase gene (TG 490).

3.1. Genotoxicity data packages for flavoring substances

The FEMA Expert Panel and JECFA have traditionally had access to the same data for the evaluation of safety of flavoring substances. The FEMA Expert Panel reviews a new application for consideration of FEMA GRAS status for each new flavoring substance individually, in what is essentially a pre-market approach in the United States and subsequently JECFA reviews the same data in groups of structurally related substances. Periodically the FEMA Expert Panel also conducts reevaluations of structurally similar substances when new

data become available or when changes in the use of the flavoring substance are likely to change the estimated consumer intake. The same data packages provided to the FEMA Expert Panel in support of new flavoring substances are also provided within the chemical group dossiers submitted for JECFA review. Further, these same data packages along with the appropriate JECFA evaluation (if previously available) and any updated literature are also reviewed by EFSA for their independent safety evaluations.

The following section examines the genotoxicity data and determinations available for flavorings as published by the FEMA Expert Panel (Adams et al. 1996, 1997, 1998; Newberne et al. 1999; Smith et al. 2002; Adams et al. 2004, 2005a, 2005b, 2005c, 2007, 2008, 2011; Marnett et al. 2014; Cohen et al. 2016; Cohen et al. 2017a; Cohen et al. 2019) or by JECFA in a series of published monographs as referenced previously (Food Additive Series No. 40-73).² Although most flavorings considered FEMA GRAS also have completed JECFA safety evaluations, there are some FEMA GRAS flavorings for which the evaluations at JECFA are pending due to the 2-year cycles of flavor evaluations at JECFA. There are also some for which additional tests have been requested to complete the evaluations at JECFA.

A summary of genotoxicity testing and frequency of negative and positive outcomes is shown in Table 1 (*in vitro*) and Table 2 (*in vivo*) for a sampling of flavoring substances within eight JECFA chemical groups. Each substance may have been tested in more than one assay, sometimes more than once in the same assay (e.g. multiple Ames assays), so both the number of tests conducted and the number of substances that have been tested in at least one genotoxicity assay are shown for each JECFA group, along with the total number of substances in each group. The summary of *in vitro* genotoxicity testing is subdivided into Ames tests (mutagenicity) and non-Ames tests, as the Ames test is the most commonly performed assay and is usually the first screening assay performed to explore possible genotoxic potential. The number of Ames assays is greater than the number of any other *in vitro* genotoxicity assay available to the FEMA Expert Panel. The non-Ames tests are further divided into the most commonly conducted assays. Several substances have been tested in less common, older, and/or non-standard assays; those are grouped as “other,” and the test names are listed in table footnotes. Typically, substances selected for testing are structurally representative of the chemical group and many are widely used (>10 kg/year).

The majority of flavoring substances have given negative results in all genotoxicity/mutagenicity tests conducted on them *in vitro* or *in vivo*. There are cases where a flavoring substance was reported to show a positive result in one of several *in vitro* or *in vivo* tests, while being negative in the rest, e.g. only one substance (isobutyraldehyde) was recorded as positive in the Ames assay, and then only with a modification of the assay (gradient plate technique) among all substances tested in chemical group 5 (Table 1). However, isobutyraldehyde was negative in all other *in vitro* tests, including standard Ames tests. Most flavoring substances were negative in the Ames assay (Table 1), while a number of flavoring substances gave positive results in non-Ames

tests. For substances with positive results in the Ames assay, e.g. in chemical groups 34 and 47, *in vivo* testing is typically available (Table 2). Generally, the majority of the positive responses for flavoring substances has typically been obtained from older, often obsolete assays that either fall short of current testing guidelines or are no longer in use due to inherent limitations. While the Panel does not disregard any available genotoxicity studies without careful review, it places particular value on those studies for which there are current OECD guidelines and for which modern methods have been used. This point is illustrated in Figure 1 for two JECFA groups (groups 4 and 30), where the proportions of negative and positive genotoxicity tests are shown for each test. Additionally, as valuable new testing approaches become available, with OECD guidelines developed,⁵ the Panel incorporates data from these assays with the same weight as other, established assays.

In cases of equivocal or positive results in the *in vitro* tests, elements of study quality and inherent limitations of each assay are considered when interpreting the data and additional *in vitro* or *in vivo* testing assists in the interpretation by providing additional information. *In vivo* tests have been primarily conducted to follow up on equivocal or positive *in vitro* findings, and thus there are fewer *in vivo* than *in vitro* assays. When tested *in vivo*, flavoring substances are typically negative for genotoxicity in the three preferred *in vivo* tests (transgenic rodent mutagenicity, bone marrow micronucleus, and comet assays) (Table 2). The results of carcinogenicity studies, when available, are used to further inform expert judgment in the weight-of-evidence assessment. The last column on the right in Table 2 reflects the overall conclusion from the weight-of-evidence assessment with the number of substances for which there is currently remaining concern of genotoxicity.

As a general observation, the larger the number of *in vitro* genotoxicity tests that have been conducted on a substance, the higher the probability that positive responses may be observed, based merely on the statistical probability of 5% false positive outcomes at the 95% confidence level typically used in statistical analysis of test results (Kirkland et al. 2005; Kirkland et al. 2007). Given that there are several factors that may contribute to a non-specific (false) positive result *in vitro* as discussed in Section 4, positive results for flavoring substances in *in vitro* assays are typically not confirmed in *in vivo* studies, with notable exceptions, e.g. 4-hydroxy-2,5-dimethyl-3(2H)-furanone (EFSA 2015a) and 3-acetyl-2,5-dimethylthiophene (Cohen et al. 2017a). These two substances also illustrate the critical role of weight-of-evidence in reaching final conclusions with regards to genotoxicity risk. For 3-acetyl-2,5-dimethylthiophene, the genotoxicity concern could not be eliminated (Table 2, JECFA group 34), partly because the biological relevance of the results could not be dismissed and relevant rodent carcinogenicity studies that could provide additional information were lacking. As a result, the FEMA Expert Panel revoked its GRAS status (see discussion in Section 5.1) (Cohen et al. 2017a). EFSA also determined that 3-acetyl-2,5-dimethylthiophene was mutagenic *in vitro* and *in vivo* and concluded that its use as flavoring substance raises a safety concern (EFSA 2013a). In

Table 1. *In vitro* genotoxicity/mutagenicity evaluation of flavoring substances by JECFA.

JECFA group no.	JECFA group name	Ames tests										Substances tested in non-Ames tests									
		Total no. of substances in group	No. of substances tested	Total no. of tests	No. of positive tests	Total no. of Ames tests	No. of positive Ames tests	No. of substances w/ Ames test	No. of substances positive Ames test	Total no. of non-Ames tests	Mammalian cell mutagenicity	MN	CA	SCE	UDS	Other	7 ^a	7 ^b	Total no. of positive non-Ames tests	No. of substances w/ positive non-Ames test	
4	Saturated aliphatic acyclic linear primary alcohols, aldehydes and acids	38	28	85	22	33	0	25	0	52	10	1	8	7	8	7 ^a		22	14		
5	Saturated aliphatic acyclic branched-chain primary alcohols, aldehydes, and acids	26	8	39	3	17	1*	8	1*	22	3	0	0	2	2	2 ^b		2	2		
6	Aliphatic Lactones	29	8	58	10	24	0	6	0	34	1	1	2	1	3	5 ^c		12	3		
7	Esters of aliphatic acyclic primary alcohols and aliphatic linear saturated carboxylic acids	69	4	10	0	6	0	4	0	4	0	0	2	0	0	1 ^d		0	0		
8	Esters of aliphatic acyclic primary alcohols and branched-chain aliphatic acyclic acids	32	1	6	1	2	0	1	0	4	0	0	1	0	0	1 ^d		1	1		
9	Esters of terpenoid alcohols and aliphatic carboxylic acids	25	2	13	2	2	0	1	0	11	1	0	1	1	1	2 ^e		2	1		
34	Sulfur-containing heterocyclic compounds	54	8	25	5	19	3	9	2	6	0	2	1	0	0	0		2	1		
47	Aliphatic linear α,β -unsaturated aldehydes, acids and related alcohols, acetals and esters	64	11	76	38	31	6	11	5	45	4	2	3	4	0	5 ^f		32	8		

*Isobutylaldehyde, see Discussion in text.

^aDNA repair, SOS chromotest, *S. cerevisiae* nonactivation suspension test, rec assay.^bRec assay, 8-azaguanine resistance assay, paper disk mutagenicity.^cFluctuation test, microtiter fluctuation, differential killing test/assay, ADP-ribosyltransferase activity, mitotic crossing over, rec assay, mitotic gene conversion, chromosomal damage, mammalian cell transformation, degranulation assay, cell growth inhibition, haploid yeast reversion, DNA pol I inhibition, sperm head abnormality, sex-linked recessive lethal (SLRL) test.^dRec assay.^ePAGE/PAGE, alkaline elution, rec assay.^fSOS chromotest, mitotic chromosomal loss, mitotic recombination, DNA repair, DNA strand break.

MN: Micronucleus; CA: Chromosomal Aberrations; SCE: Sister Chromatid Exchange; UDS: Unscheduled DNA Synthesis

Table 2. *In vivo* genotoxicity/mutagenicity evaluation of flavoring substances by JECFA.

JECFA group no.	JECFA group name	No. of substances	Substances tested										No. of substances w/ current concern			
			MN Comet	TGRM	Total no. of substances tested in these 3 tests		No. of positive results in these 3 tests		No. of positive results in other tests		Total no. of positive tests			No. of substances with carcinogenicity study		
4	Saturated aliphatic acyclic linear primary alcohols, aldehydes and acids	38	2	0	0	2	1 ^a	0	0	0	1	1 ^a	0	0	0	0
5	Saturated aliphatic acyclic branched-chain primary alcohols, aldehydes, and acids	26	0	0	0	0	0	2 ^b	0	0	0	0	2	1	0	0
6	Aliphatic Lactones	29	2	0	0	2	0	0	0	0	0	0	0	0	0	0
7	Esters of aliphatic acyclic primary alcohols and aliphatic linear saturated carboxylic acids	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Esters of aliphatic acyclic primary alcohols and branched-chain aliphatic acyclic acids	32	0	0	0	0	0	0	0	0	0	0	2	2	0	0
9	Esters of terpenoid alcohols and aliphatic carboxylic acids	25	1	0	0	0	0	1 ^c	0	0	0	0	2	1	0	0
34	Sulfur-containing heterocyclic compounds	54	3	1	1	3	1	0	0	1	1	1	0	0	0	1 ^e
47	Aliphatic linear α,β -unsaturated aldehydes, acetals and esters	64	6	0	0	6	2	3 ^d	0	2	2	2 ^f	4	1	1	0

^aAcetaldehyde; see discussion in text.^bSLRL, dominant lethal assay, CA.^cSLRL, UDS.^dCA, SCE, SLRL, DNA alkaline elution.^e3-Acetyl-2,5-dimethylthiophene; see Discussion in text.^fEthyl acrylate and 2-Hexenal; see Discussion in text.

MN: Micronucleus; TGRM: Transgenic rodent mutation.

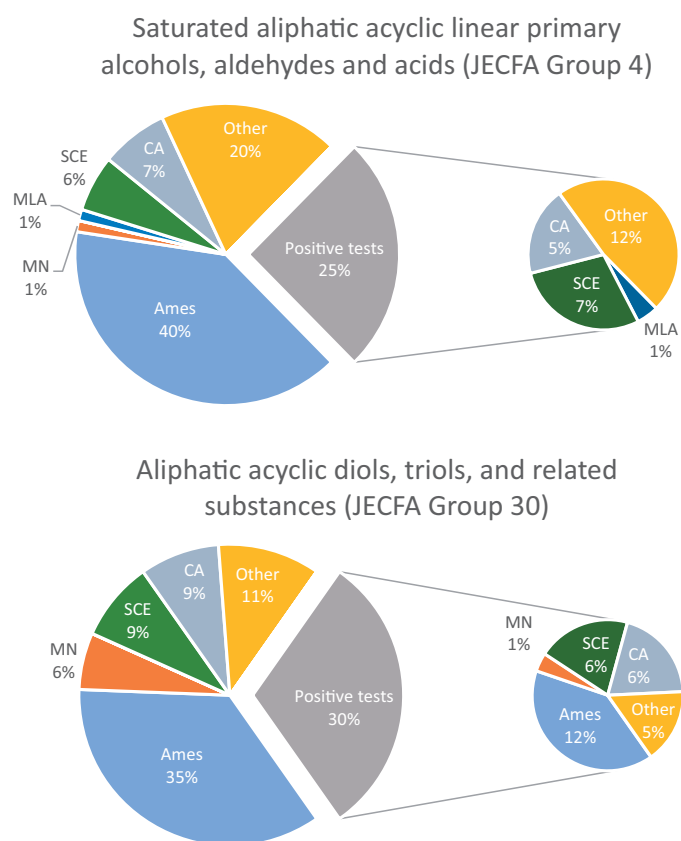


Figure 1. Negative and positive tests as percentages of all *in vitro* tests conducted for JECFA chemical groups 4: Saturated aliphatic acyclic linear primary alcohols, aldehydes and acids (85 *in vitro* tests), and 30: Aliphatic acyclic diols, triols, and related substances (82 *in vitro* tests). The pie chart on the left shows the percentages of negative and positive tests relative to all tests conducted for the chemical group: Ames assay; micronucleus (MN); mouse lymphoma assay (MLA); sister chromatid exchange; chromosomal aberrations (CA). The pie chart on the right shows the contribution of specific assays among the positive tests. The fraction “other” includes less frequently encountered tests (see Table 1 footnotes).

contrast, there was no remaining concern for the use of 4-hydroxy-2,5-dimethyl-3(2H)-furanone as a flavoring, despite the positive *in vivo* genotoxicity data. Any concern raised by the test results was eliminated based on metabolism and mode of action data consistent with reactive oxidative species formation, as well as the availability of a negative carcinogenicity study (JECFA 2005; Smith et al. 2009) and absence of gonadal effects in a male rat fertility study (EFSA 2015b). In other cases, genotoxicity concern raised by positive *in vivo* results was assessed based on specific considerations: (a) acetaldehyde was reported positive in one *in vivo* bone marrow micronucleus assay in mice at very high levels of intraperitoneal dosing, which is not considered relevant for oral exposure (JECFA 1998a, 1998b) (see Table 2, JECFA group 4); (b) ethyl acrylate and 2-hexenal were reported to increase micronuclei frequencies in the bone marrow or buccal cells, respectively, in two older studies (Table 2, JECFA group 47); however, ethyl acrylate was administered intraperitoneally in that study and the findings of both studies were superseded by negative results in later studies (JECFA 2005, 2006; Adams et al. 2008). The overall interpretation with regards to the genotoxicity of these substances was not solely based on any single study but on the quality criteria

detailed in Section 4 and the overall weight of evidence as discussed in Section 5.

4. Interpretation of genotoxicity data in FEMA GRAS evaluations

As described above, the FEMA Expert Panel endeavors to conduct a comprehensive safety evaluation when considering the GRAS status of flavoring substances, rather than a hazard assessment alone. The FEMA Expert Panel’s evaluation process leads to a conclusion of the probable risk to consumers. The FEMA Expert Panel assesses probable risk when evidence of genotoxic potential meets two conditions. First, there are either structural alerts and/or positive results in genotoxicity assays where findings are biologically relevant to humans (further discussed in Section 4.3); second, the findings indicate a concern under the conditions of use of flavoring substances. Whether assay results are clearly positive, clearly negative, or equivocal, the FEMA Expert Panel interprets individual assays within the context of all relevant data. The value of the results of each genotoxicity assay within the overall evaluation (relative to all of the available data) is determined by three critical elements: (1) the study quality, (2) biological relevance of assay results, and (3) human relevance, as discussed in detail below. Where available, negative results from well-conducted *in vivo* genotoxicity studies would typically outweigh positive *in vitro* results, provided they reflect the same genotoxic mode of action (i.e. mutagenicity or chromosomal damage). Flavoring substances that contain structural alerts for genotoxicity, such as α,β -unsaturated carbonyl moieties, reactive aldehyde moieties, α -ketone functionality, epoxide groups, or aromatic heterocyclic groups, are subject to particular scrutiny and require a comprehensive dataset of high quality genotoxicity studies as well as specific data on metabolic fate and kinetics to unequivocally eliminate any genotoxicity concern. The FEMA Expert Panel considers the totality of the scientific information to resolve conflicting data. The critical factors affecting both the outcome of the genotoxicity tests and the interpretation of the results are discussed below.

4.1. Study quality

Before the FEMA Expert Panel reviews the results of a study in detail, the quality of the study is evaluated based on broadly recognized criteria for study acceptance. Adherence to internationally accepted testing guidelines, such as those of the OECD provides strong confidence that the study is likely well-conducted, reproducible, and reliable. The OECD publishes guidelines only after extensive inter-laboratory validation of each assay has been conducted and the acceptance criteria for the proper performance and evaluation of assays are detailed within each guideline. Additionally, adherence to Good Laboratory Practices (GLP) provides confidence in the quality of experimental conditions and has been established by OECD (OECD GLP), or regulatory authorities (US FDA GLP). Non-OECD guideline studies, either predating the publication of the guidelines or not fully adhering to the

guidelines, are reviewed carefully with an eye to documentation of indicators of good study quality (similar to individual quality criteria described in the guidelines). These may include the justification of selected concentration or dose ranges tested, the adequacy of treatment time and sampling timing, detailed documentation of the conditions and procedures of tissue collection and processing, adequate data presentation, data variability, and statistical analysis. However, the FEMA Expert Panel is of the opinion that current studies are more stringent, based on knowledge of study performance, pitfalls, and limitations that has accumulated over the decades of genotoxicity testing. Particular attention is given to study conditions that are now recognized as sources of artifacts, giving rise to misleading results, or that would limit the biological relevance of the findings, even if these were not yet recognized at the time of the study publication. For example, it is critical that a study adequately documents how concentration or dose selection is justified, either by preliminary testing or previously available information on the cytotoxicity or systemic toxicity of the substance. Sufficient confidence that the test substance is properly identified and of appropriate purity is also important to understand whether the test substance was appropriate for testing, and not, for instance, degraded or oxidized. A study of good quality is also one that has adequately addressed sources of artifacts that may compromise the validity of the results (see Section 4.2 on biological relevance, below). Such artifacts include interactions between the test substance and components of the culture medium, which can lead to production of reactive oxygen species (Kirkland 2011); high osmolality, high ionic strength, and extremes of pH, which can lead to artifactual positive responses in mammalian cell genotoxicity tests (Brusick 1987; Scott et al. 1991).

As mentioned, a portion of genotoxicity data for flavoring substances reviewed by the FEMA Expert Panel is from studies that predate OECD guideline publications. Therefore, at the time of periodic updates of the GRAS status of flavoring substances, the FEMA Expert Panel reevaluates previously reviewed data using current criteria of assay validity. Updated testing may be considered necessary to confirm the safety of flavoring substances and reaffirm their GRAS status if older data are determined to be insufficient according to current criteria. It is recognized that adherence to current criteria is not a strict requirement but rather a first factor to determine whether the study could be useful in a safety assessment.

4.2. Biological relevance

For an increase in the frequency of mutants or other parameters indicating DNA damage (or absence of it) to be biologically meaningful, several factors must be scrutinized. Evidence of DNA damage may not be biologically relevant if it is the result of certain experimental conditions such as (1) excessive cytotoxicity, (2) an inadequate dose-response relationship, (3) high data variability, (4) absence of a DNA repair system within the assay system, or (5) other inherent

assay performance limitations. These conditions resulting in assay artifacts and false positives are discussed in more detail.

4.2.1. Cytotoxicity

Dose-dependent cytotoxicity is known to induce artifacts in *in vitro* assays and must be carefully considered (Kirkland et al. 2007). Cytotoxicity observed *in vitro* is used to establish the maximum concentrations up to where meaningful data are collected. For all *in vitro* assays, the treatment period is relatively short (often 3–24 h), but long enough to allow the genetic damage to occur and become heritable. Longer duration of exposure is not appropriate, since the frequency of cytogenetic damage may decrease with time either via apoptosis or by differential growth of non-damaged cells. Therefore, *in vitro* tests are typically conducted at concentrations high enough to induce a detectable level of genetic damage in short treatment periods (OECD 2017). However, since such high concentrations can lead to significant cellular perturbations and cytotoxicity, limits of cytotoxicity (e.g. ~50–60% in the *in vitro* micronucleus assay) are considered, above which genotoxicity scoring is not meaningful (Galloway 2000; Honma 2011). The Panel notes that cytotoxicity at any level can result in DNA damage and should be taken into consideration when interpreting assay results.

Among the different methods used to determine cytotoxicity, those that account for dividing cells rather than simply cell counts are preferred (Fellows and O'Donovan 2007; O'Donovan 2012). Specifically, evidence suggests that non-physiological conditions (i.e. unusual pH or osmolality) that significantly inhibit cell division often lead to irrelevant genotoxicity that results in false-positives (Brusick 1986; Brusick 1987). Measurements of cytotoxicity are used for two objectives: (1) to better define the concentrations to be used in the main experiment and (2) to demonstrate sufficient exposure of the cells. Cytotoxicity measures based on cell proliferation are preferred for genetic toxicology tests and, consequently, have been incorporated into the revised OECD TGs. As a result of recognizing the significance of cytotoxicity indicators, the OECD recently updated its published guidelines for two cytogenetic assays including the *in vitro* micronucleus (OECD 2016b) and *in vitro* chromosomal aberration assays (OECD 2016a) to include recommendations for the use of cytotoxicity indices such as relative population doubling or relative increase in cell count (RICC). These updates take cell cycles/growth into consideration, instead of relative cell counts (RCC) (Fellows et al. 2008; Fowler et al. 2012b). These changes were anticipated to reduce false positive outcomes due to cytotoxicity. Reanalysis of previously reported results of *in vitro* cytogenetic assays based on updated cytotoxicity evaluation can lead to more accurate assessment of flavoring substances that were previously determined to be genotoxic. For example, an algorithm has been developed to predict the likelihood that test results (positive or negative) would change when updated cytotoxicity indices are employed for previously published studies for the *in vitro* chromosomal aberration test (Honda et al. 2018). This algorithm was used

to examine >100 substances retrospectively in a database with *in vitro* chromosomal aberration test data. By utilizing these updated cytotoxicity indices, several false positives were reclassified as negative results (Honda et al. 2018). Thus, the method of cytotoxicity assessment employed is an essential factor that the FEMA Expert Panel takes into consideration when determining the reliability of *in vitro* cytogenetic assay results.

Evaluation of mutagenicity *in vitro* requires dividing cells through the gene expression phase of the assay and during the cloning for mutant selection. Therefore, test results are meaningful in these assays within the range of test concentrations that allow not only for cell survival but also for cell proliferation. For an *in vitro* mammalian cell gene mutation test, like the mouse lymphoma assay, the relative total growth (RTG) is the recommended measure of cytotoxicity (OECD 2016d). For other gene mutation assays, the relative survival (RS) is recommended. RS is the relative cloning efficiency of cells plated immediately after treatment and accounts for cell loss during treatment. When evaluating new substances for GRAS consideration or reevaluating those already with GRAS status, the FEMA Expert Panel examines the cytotoxicity methods used as part of the assessment of the results of any individual genotoxicity assay.

4.2.2. Dose or concentration–response

Evidence that the genotoxic effect is dose- or concentration-dependent is one of the three key criteria (along with statistically significant difference from concurrent control and exceeding of the historical control range) put forth in OECD guidelines for results of genotoxicity assays to be interpreted as positive, whether *in vitro* or *in vivo*. The requirement for a consistent increase in response with increasing concentration/dose prevents erroneous interpretation of genotoxicity on the basis of a spurious increase at a single concentration/dose level (or within a single animal).

4.2.3. Variability and reproducibility

Some genotoxicity assays, such as the *in vivo* comet assay, have significant parameter and results variability (Speit et al. 2015). Variability may be due to poor study quality, the nature of the endpoint, small effect size (e.g. amplified differences between low-frequency events such as micronuclei frequency), influence of technical artifacts (e.g. physical damage to DNA during sample preparation for the comet assay), the cell type (for *in vitro* assays), or other factors that impact the performance of the assay. Generally, significant variability within the data collected in a study requires a repeat of the experiments, although the sources of variability cannot necessarily be controlled simply by repeating the test. Large data variability and the resulting lack of reproducibility reduce the reliability of genotoxicity testing results, and this increases the uncertainty in determining whether a substance might actually possess genotoxic potential. Nonetheless, repeated experiments might provide additional data to assist interpretation. In the case of high variability, the FEMA Expert

Panel turns to other sources of evidence, including other genotoxicity data and other information to confirm the data.

4.2.4. Functional DNA repair systems

Cell-based *in vitro* test systems have been established as valid assays to identify possible genotoxic potential and are often preferred as the first screening tools. The acceptance of cell-based assays is based upon the universality of genotoxic modes of action that lead to DNA mutations or chromosomal damage. This approach is applied despite the recognition that cells in culture, particularly immortalized cell lines, behave differently than the same cells present in *in vivo* test systems. In essence, removal of the cells from their biological context (e.g. the multicellular 3-dimensional tissue environment) has a significant impact on how the cells grow, survive, and respond to xenobiotics such as those encountered in genotoxicity testing. Some cell lines commonly used in genotoxicity testing have the potential to undergo genetic drift and changes in karyotype, changes in gene expression patterns, loss of key genes, or loss of other cell functions critical for the maintenance of genetic stability, such as functional DNA repair systems (Kirkland et al. 2007; Fowler et al. 2012b; Whitwell et al. 2015). Reduced or absent DNA repair increases the probability that DNA damage is “fixed” and is associated with both higher background of DNA damage and higher responses to test substances. The increased frequency of DNA damage that escapes repair increases the probability of random damage being detected and therefore increases the frequency of false-positive results. A systematic comparison of false-positive results among commonly used cell lines has revealed that V79, CHL, or CHO cells do not have a functional p53 gene and are prone to higher frequencies of misleading positive genotoxic outcomes (60%, 66%, and 53%, respectively) compared to p53-competent cells such as human lymphocytes, TK6 cells and HepG2 cells (17%, 40%, and 23%, respectively) (Pfuhler et al. 2011; Fowler et al. 2012b; Whitwell et al. 2015). Therefore, positive results obtained using p53-deficient cells or cells lacking other DNA repair systems are interpreted with caution, preferably in a context of additional relevant data, while data from human lymphocytes are regarded as more reliable.

4.2.5. Assay performance in predicting possible human genotoxic hazard

All genotoxicity assays are experimental models that attempt to identify the possibility of genotoxic effects (or lack thereof) in humans. As models, genotoxicity assays are subject to inherent limitations for correctly detecting true genotoxic activity (assay sensitivity) and correctly eliminating concern for a non-genotoxic substance (assay specificity). Failure of an assay to identify a true positive result (known as false negative outcomes, or low sensitivity) is primarily a concern for regulators as it reflects unidentified and therefore unmitigated hazard, while failure to eliminate non-genotoxic substances (known as false positives outcomes, or low specificity) leads to unnecessary follow-up testing and unnecessary animal use. As human genotoxicity data are very limited, the

performance of genotoxicity assays has been measured against results from animal carcinogenicity studies. Typically, these are rodent bioassays, but it should be noted that some rodent tumors do not adequately reflect the human situation (discussed in Section 4.3). If not properly evaluated, rodent tumor data may lead to the incorrect conclusion that negative genotoxicity assay results are falsely negative when they are not, i.e. leading to the incorrect conclusion that the assay lacks sensitivity. In contrast, negative rodent carcinogenicity results challenge assay specificity, suggesting that positive genotoxicity assay results may be falsely positive. Considering current understanding of the complexity of carcinogenesis and its relationship to genotoxicity, and the limitations of the traditional rodent bioassay, the FEMA Expert Panel maintains reservations about the published rates of assay sensitivity (false negative rates) when measured against the rodent bioassay.

The FEMA Expert Panel interprets results cautiously when obtained from a single assay. However, the benefit of assay combinations in strengthening the reliability of results comes with increased probability of false positive results, particularly for *in vitro* assays, simply by addition of the statistical probability of 5% false positive rate per test. Indeed, false positive outcomes are considered to be a more substantial challenge to interpretation than false negative ones (Kirkland et al. 2007). Systematic analysis of assay predictivity has revealed that up to 80% of 177 non-carcinogens (i.e. negative in the rodent bioassay) were positive in at least one *in vitro* genotoxicity assay when multiple assays were conducted (Kirkland et al. 2005). The Ames assay has a significantly lower rate of false positives than other genotoxicity assays (Kirkland et al. 2005, 2007), and positive results from the Ames assay are relatively sparse in the chemical space of flavoring substances. Thus, positive results in Ames assays warrant careful consideration and often require further data to appropriately assess the possible mutagenic activity in humans. A recent analysis suggests that positive results in Ames assays are not indicative of *in vivo* mutagenic or carcinogenic activity if accompanied by negative results in two mammalian cell genotoxicity assays, regardless of whether they query mutagenicity or other types of chromosomal damage (Kirkland et al. 2005, 2007, 2014).

Based on relative measures of assay reliability, such as the relative success-to-failure ratio, or relative predictivity (correct-to-incorrect prediction rates of either genotoxic or non-genotoxic substances), the Ames assay is reported to have the highest positive predictivity, and the mouse lymphoma assay (MLA) the highest negative predictivity. Recent analysis suggests that combinations of assays can provide the highest sensitivity in predicting for carcinogenicity (Bhagat 2018).

The FEMA Expert Panel regards the results of *in vitro* assays for genotoxicity testing, if/when available, as either further confirmation of *in vitro* findings or a more conclusive means to resolve equivocal findings *in vitro*. Within the limited current database of *in vivo* results, the *in vivo* MN is reported to have a high false negative rate (Morita et al. 2016), and typically a second *in vivo* assay, e.g., comet assay or transgenic rodent mutagenicity assay is also used. The

comet assay shows higher sensitivity (89%) and specificity (78%), relative to the transgenic rodent mutagenicity assay (50% sensitivity and 69% specificity) in detecting genotoxic substances that were missed in the *in vivo* MN assay (Kirkland and Speit 2008). This may be expected as the comet assay identifies substances that induce both chromosomal aberrations and mutations, while the transgenic rodent mutagenicity assay is specifically designed to be highly sensitive for the detection of mutations only.

False positive results *in vitro* have been attributed primarily to the use of cells without a functional p53 gene or other DNA repair mechanism (e.g. TK6 cells lack repair mechanisms for double-strand breaks), and improper measures of cytotoxicity, among other factors (Kirkland and Speit 2008; Kirkland et al. 2016). Conversely, false negative results in *in vitro* genotoxicity assays have been associated with the commonly used exogenous metabolic activation system (S9, discussed below) (Kirkland et al. 2007). The sensitivity and specificity of an assay also depend on whether it measures DNA damage directly or indirectly by measuring a surrogate of DNA damage, such as compensatory (unscheduled) DNA synthesis (UDS) an indicator of DNA repair. Indirect measures of genotoxicity are less sensitive [e.g. the *in vivo* liver UDS assay has a sensitivity of <20% (Kirkland and Speit 2008)] and may display higher variability that compromises statistical power. The FEMA Expert Panel includes the above considerations in the interpretation of results from *in vitro* and *in vivo* genotoxicity assays.

4.3. Human relevance

In addition to the above considerations of assay-specific artifacts to assess biological relevance of the results and because the FEMA Expert Panel evaluates flavoring substances specifically for human safety, genotoxicity assay data are assessed for their human relevance. This entails primarily two key elements: (a) whether the metabolic activation system was appropriate for the flavoring substance, and (b) whether the mode of action applies to humans. The FEMA Expert Panel applies these considerations to both *in vitro* and *in vivo* genotoxicity assay data in evaluating human relevance as detailed in the sections below.

4.3.1. Metabolic activation

The FEMA Expert Panel recognizes the role of metabolism as a critical contributing factor to toxicity outcomes including genotoxicity. The types of metabolic pathways encountered in the safety evaluation of flavoring substances have recently been reviewed by the FEMA Expert Panel (Smith et al. 2018). Mammalian enzyme systems generally eliminate or reduce the levels of a wide variety of exogenous chemicals (xenobiotics) and facilitate their elimination from the organism. However, metabolic activation and detoxication processes determine the net balance of reactive intermediates to inactive metabolites and therefore, subsequent manifestations of toxicity. Metabolism is an essential factor in the interpretation of genotoxicity assay results, particularly when generated *in vitro*. Mammalian metabolic activation systems

are necessary for bacterial *in vitro* genotoxicity models. However, mammalian cell lines also lack or have limited ability to metabolize chemicals without an exogenous metabolic activation system (Kirkland et al. 2007; Pfuhler et al. 2011). Liver homogenate post-mitochondrial fraction (S9), available from different species, is the most commonly used exogenous metabolic activation system for simulating the metabolism of compounds in humans and other animals because it contains major metabolic enzymes (Jia and Liu 2007; Richardson et al. 2016). Typically, S9 is prepared from the livers of rats that have been treated with chemicals known to induce hepatic drug metabolism [typically Aroclor-1254 (a mixture of polychlorinated biphenyls), or a combination of phenobarbital and β -naphthoflavone]. Liver homogenates from other species (e.g. hamster or guinea pig) can be used to investigate effects of alternate metabolic pathways when the dominant pathway differs among species. Understanding the metabolic pathways in humans is necessary to interpret data that depend on particular biotransformation pathways, and human S9 fractions are available for this purpose (Cox et al. 2016).

The relevance of the source of the exogenous metabolic system and the range of metabolites generated are criteria used in the interpretation of *in vitro* genotoxicity assay results. The representation of metabolic enzymes in the exogenous mix may differ quantitatively and qualitatively between the source species, as well as from the *in vivo* context depending on the choice of the chemical used to induce metabolic enzymes. Furthermore, even when a key enzyme is present, it may not be active in the S9 mix if the required co-factors are absent. Treatment of rats with Aroclor-1254 preferentially induces oxidative liver enzymes, particularly cytochrome P450 families 1-3 (Dubois et al. 1996) and favors oxidative activation. However, the conjugating activity of such S9 mix is limited, especially in the absence of added cofactors (Glatt et al. 2012; Honda et al. 2016), and thus this model does not totally represent the mammalian biotransformation capabilities. Furthermore, the cytochrome P450 (P450) enzymes induced in rat liver are not fully representative of the P450 activity profile of human liver (Dubois et al. 1996). Among the key human enzymes poorly represented in rat liver S9 are sulfotransferase (SULT), *N*-acetyl transferase (NAT), and some extrahepatic P450 enzymes such as CYP1B1 (Jin et al. 2018). Therefore, the metabolite profile can be substantially different in genotoxicity models compared to the human metabolite profile and may lead to either false positive or false negative results depending on whether the biotransformation is skewed toward the generation of reactive intermediates or detoxication products of the primary compounds (Glatt et al. 2012; Honda et al. 2016).

The absence of conjugation enzymes may be associated with false positive or false negative results *in vitro*, since conjugation reactions that generally facilitate detoxication and urinary elimination of xenobiotics may also convert several compounds to reactive products. Many pro-mutagens are activated to mutagens *in vivo* by the SULT enzyme family (Glatt 2000; Glatt and Meinel 2005). Given that conjugation pathways are generally underrepresented in exogenously added metabolic systems, *in vitro* systems yield false negative

results compared to *in vivo* assays for substances that are activated following SULT conjugation.

Importantly, the *in vivo* glutathione (GSH) transferase conjugation pathway is typically limited if not lacking in *in vitro* assays unless the metabolic activation systems are explicitly modified to include added GSH. This is particularly relevant in the interpretation of data for flavoring substances dependent on the GSH conjugation pathway for detoxication and elimination, as recognized for high concentrations of α,β -unsaturated aldehydes, where depletion of GSH levels promotes oxidative responses such as the release of nucleolytic enzymes that induce DNA fragmentation, cellular damage, and apoptosis (Eisenbrand et al. 1995; Kiwamoto et al. 2012).

Additionally, the efficacy of an exogenously added metabolic system is compromised because it functions extracellularly and metabolites may not be membrane permeable. This limitation of exogenous biotransformation systems leads to false negative results if the reactive metabolite cannot reach the intracellular target.

Mitigation strategies for the above limitations of metabolic activation options are adopted to fit the particular metabolic context. The use of HepaRG cells mitigates the limitation of extracellular enzyme systems to some extent because they express intracellularly metabolic pathways similar to those operating in human hepatocytes, albeit with quantitative differences (Ramaiahgari et al. 2017). Alternatively, genetic engineering of *in vitro* systems to express key human recombinant enzymes, including P450s, alone or in combination with conjugation enzymes SULT or NAT2, has been successfully applied to circumvent the limitations of insufficient metabolic representation of human enzymes *in vitro* (Crespi et al. 1991; Glatt and Meinel 2005; Glatt et al. 2012; Glatt et al. 2016). These cell systems allow for selection of the most appropriate metabolic transformation option based on prior knowledge, if available, on the metabolic fate of the substance in mammalian organisms and particularly humans. HepaRG cells offer a more complete representation of the complexity of human metabolic enzymes compared to the engineered cell lines available to date, that generally co-express one to two enzymes and not an extensive range of enzymes to sufficiently represent the complexity *in vivo*. In addition, the genetic engineering approach is conditional upon availability of data on the metabolic fate of a substance in humans.

The FEMA Expert Panel scrutinizes the details of the metabolic activation system in the context of the above considerations when interpreting positive results of *in vitro* genotoxicity assays that were obtained only in the presence of metabolic activation but not in its absence. As this is indicative of metabolic activation processes, additional data or testing may be sought to understand the relevance of the active metabolites to humans.

4.3.2. Mode of action

When results obtained in *in vivo* genotoxicity assays are positive or equivocal, further investigation into the mode of action is warranted. Genotoxic modes of action are typically

considered to be either direct interactions between the test chemical and DNA, or DNA damage that occurs indirectly (i.e. resulting from the perturbation of other cell mechanisms by the test chemical). Typically, indirect modes of action for genotoxicity, such as the generation of oxidative species, are assessed using a non-linear or threshold-based dose-response model (Pratt and Barron 2003; EPA 2004; Tritscher 2004; EPA 2005; Foth et al. 2005; EU 2009; Barlow and Schlatter 2010; EFSA 2011). Indirect modes of action are, therefore, evaluated within the framework of a human risk assessment process that includes identification of a point of departure (POD), consideration of exposure, and determination of a margin of safety (MOS). Indirect modes of action in genotoxicity assays usually present no concern for consumer safety due to the low levels of exposures typically seen with flavoring substances, which often provide a sufficient MOS. On the contrary, when evidence is consistent with a direct genotoxic mode of action that is considered to be both biologically relevant and of potential human impact, the Panel requires additional information such that the genotoxicity hazard can be appropriately addressed within a comprehensive risk assessment. Examples of this are shown in Section 5.1 and 5.2 below.

Data from both *in vitro* and *in vivo* assays can be useful in demonstrating a direct mode of action of genotoxicity for the flavoring substances or their (relevant) metabolites. For example, even though the Ames assay is based on a bacterial test system, Ames assay data are considered relevant to human safety assessment because this test can assess a direct mode of action resulting from an interaction between a substance (or a metabolite) and DNA. Confidence in the support that data from Ames and other genotoxicity assays provide for a direct mode of action is conditional upon sufficient scrutiny of possible sources of artifacts resulting in false positive results (see Biological relevance section) and of evidence indicating an indirect mode of action. Most *in vitro* genotoxicity assays are designed to identify direct DNA damage, with the exception of the UDS, which indicates DNA repair. The micronucleus assay is informative with regards to the mode of action when it includes centromere or kinetochore staining to distinguish between direct DNA interaction (clastogenicity) and spindle-mediated chromosomal separation (aneugenicity) (OECD 2016b). Therefore, a substance is unlikely to have a direct genotoxic effect if *in vitro* genotoxicity assay results are negative.

The demonstration of whether genotoxicity modes of action are indirect, e.g. mediated by reactive oxygen species, either *in vitro* or *in vivo*, requires collection of additional evidence and careful consideration of their source. The FEMA Expert Panel considers that oxidative species may be generated in the test system from two sources: artifacts of the test system or cell/tissue injury. The generation of DNA-reactive agents as artifacts in the test system, such as hydrogen peroxide from phenolic substances under aerobic conditions in the presence of trace metals, can be mitigated by addition of catalase (eliminating hydrogen peroxide) or antioxidant supplementation (Kirkland et al. 2007; Kurutas 2015). Oxidative species that are generated as a result of cell or tissue injury occur only at cytotoxic concentrations and therefore the displayed genotoxic effects are not directly due to the test

substance but occur secondary to toxicity. DNA damage from these sources is evaluated within a human risk assessment framework that considers cytotoxicity (*in vitro*) or clinical toxicity (*in vivo*), as well as the responses observed at concentrations and/or doses giving lower levels of toxicity.

4.3.2.1. DNA adducts. DNA adduct studies have been conducted for only a small number of flavoring substances and potential metabolites, including formaldehyde, acetaldehyde, the α,β -unsaturated aldehydes acrolein, crotonaldehyde, and related compounds (Wang et al. 2000; Hecht et al. 2001; Hecht et al. 2011; Yang et al. 2019), 2-hexenal and *trans*, *trans*-2,4-hexadienal (Frankel et al. 1987; Eder et al. 1993; Eisenbrand et al. 1995; NTP 2003); estragole (Ishii et al. 2011; Paini et al. 2012; Ding et al. 2015), and methyl eugenol (Phillips et al. 1984; Herrmann et al. 2013; Williams et al. 2013; Herrmann et al. 2014; Monien et al. 2015; Tremmel et al. 2017). Most flavoring substances have chemical structures that make them unlikely candidates for DNA adduct formation (e.g. they are not inherently reactive electrophiles or form electrophilic metabolites).

DNA adduct formation is one direct mode of action for substances with genotoxic potential. DNA adducts are sometimes interpreted as biomarkers of biological effect but can be considered as biomarkers of exposure to a substance not necessarily resulting in biological consequences. The reasons for different interpretations result from a complex set of considerations involving the adduct structure, measured levels, repair capacity, endogenous formation, and other factors (De Bont and van Larebeke 2004; Paini et al. 2011; Swenberg et al. 2011; Basu 2018). When presented with data on DNA adducts, the FEMA Expert Panel has concluded that several factors derived from a large body of literature must be considered. These factors include (a) the structure of DNA adducts detected, (b) the repair/persistence of DNA adducts, (c) levels of adducts detected relative to those occurring endogenously or as the result of non-flavor related background exposure, (d) the methods for detecting and measuring adducts, (e) the dose-response relationship for adduct formation, (f) the metabolic profile of the flavoring substance, (g) evidence to determine a direct or indirect mode of action, (h) consistency with data from *in vivo* mutagenicity assay, if available, and (i) other target tissue pathology. The FEMA Expert Panel considers these factors in the context of evolving science on the association of DNA adducts with mutagenicity as discussed in detail below.

4.3.2.1.1. DNA adducts, genotoxicity, and carcinogenicity. While DNA adduct formation can be a critical component of mutagenesis and carcinogenesis due to the miscoding properties of some DNA adducts, the role of any particular DNA adduct in these processes is highly dependent on multiple factors. These include the extent of adduct formation under physiological conditions, the structure and stability of the adduct formed, the shape of the dose-response curve for adduct formation, persistence of the DNA adduct, DNA adduct repair mechanisms, DNA polymerases involved in error prone bypass and resulting mutagenesis, the location of the adduct in the genome, and other factors (Peterson 2017;

Barnes et al. 2018; Fukushima et al. 2018; McCullough and Lloyd 2019; Pottenger et al. 2019). Thus, it is difficult to generalize the potential mutagenic and carcinogenic effects of DNA adduct formation besides their utility as exposure biomarkers, but rather it is necessary to consider a particular DNA adduct in the context of the biological systems under investigation. Context may determine whether the formation of DNA adducts is a good predictor of mutagenicity and of the potential for direct biological consequences, e.g. carcinogenicity (Hecht et al. 2011; Paini et al. 2011; Swenberg et al. 2011). Common environmental and lifestyle exposures lead to DNA adduct formation and have been associated with mutations but not necessarily with carcinogenesis (Wang et al. 2000; Hecht 2003; Wang et al. 2006; Zhang et al. 2006; Lao et al. 2007; Balbo et al. 2008; Wang et al. 2009; Balbo et al. 2012). Other reports indicate that DNA adducts do not always lead to permanent mutations but are subject to repair at rates and efficiencies dependent on the adduct levels and structure (Povey 2000; De Bont and van Larebeke 2004; Gocke and Muller 2009; Swenberg et al. 2011; Broustas and Lieberman 2014; Kobets and Williams 2016; Geacintov and Broyde 2017). As one example, the major acetaldehyde DNA adduct (*N*²-ethylidene-dG) from consumption of a single alcoholic beverage increases rapidly but transiently in the oral cavity and blood (Balbo et al. 2008; Hecht et al. 2011; Balbo et al. 2012). A dose–response relationship has been shown between alcohol consumption and DNA adduct formation, although the relationship of adduct presence to genotoxic or carcinogenic effects has not been characterized in detail.

4.3.2.1.2. Endogenous DNA adducts. DNA adducts can be detected at extremely low levels, e.g. the reported limit of detection (LOD) is as low as 1 adduct per 10¹¹ nucleotides and the limit of quantitation (LOQ) is as low as 5 adducts per 10¹¹ nucleotides by LC/MS (Zhang et al. 2006; Monien et al. 2015; Villalta et al. 2017; Yang et al. 2019) and is even lower by accelerator MS (Hummel et al. 2018; Madeen et al. 2019). Therefore, it is critical to evaluate whether the level of DNA adducts detected is biologically significant. In some cases, interpretation of DNA adduct data requires comparison against reported background adduct levels arising from non-flavor related exposures, which can vary widely depending on the structures of adducts (Povey 2000; De Bont and van Larebeke 2004; Swenberg et al. 2011). One source of these background DNA adducts are DNA interactions with electrophilic molecules and reactive oxygen species that are produced endogenously from normal physiological energy metabolism and oxidative processes, e.g. inflammation, mitochondrial respiration, lipid peroxidation, estrogen oxidation, endogenous alkylating agents, e.g. *S*-adenosylmethionine, *N*-nitroso compounds, and others (Tornqvist et al. 1989; Bartsch et al. 1992; De Bont and van Larebeke 2004; Yager and Davidson 2006). In other cases, adducts are derived from the natural occurrence of some chemicals in foods and the environment (e.g. methyl eugenol in basil) (De Bont and van Larebeke 2004; Herrmann et al. 2013; Tremmel et al. 2017). The FEMA Expert Panel considers the background level of DNA damage to be a crucial element in the interpretation of data for either direct or indirect-acting genotoxic substances,

with particular consideration of the background frequency of the same type of adducts in the same target tissue (Povey 2000; Swenberg et al. 2011).

4.3.2.2. Methyl eugenol and FEMA expert panel decision to remove it from the FEMA GRAS list. Methyl eugenol, a naturally occurring allylalkoxybenzene substance found in sweet basil and other herbs (Miele et al. 2001) forms DNA adducts (Phillips et al. 1984; Randerath et al. 1984; Williams et al. 2013; Alhusainy et al. 2014; Tremmel et al. 2017) as well as protein adducts (Gardner et al. 1996) in rodents. An older ³²P-postlabelling study suggested dose related increases in the levels of DNA adducts in the liver and at the top dose only in the glandular stomach of rats administered methyl eugenol for 28 days (Ellis et al. 2007). Similarly, in a later study designed to investigate its tumor-initiating potential, gavage administration of methyl eugenol to rats three times a week for 8 weeks or 16 weeks resulted in dose-dependent increases in liver DNA adducts as measured by ³²P-postlabelling (Williams et al. 2013). Adduct levels were reduced (by 70–80%) during a 24-week post-treatment recovery period with or without the promoter phenobarbital (Williams et al. 2013). In the same study, DNA adducts correlated with a dose-dependent increase in hepatocyte proliferation (based on PCNA staining) at all dose levels. Hepatic preneoplastic lesions (based on GST-P immunohistochemistry staining) and hepatocellular adenomas increased significantly in the middle and high dose groups during the 24-week post-treatment period (Williams et al. 2013).

The FEMA Expert Panel examined all available data and assessed the genotoxic potential of methyl eugenol. The metabolism of methyl eugenol plays a critical role in its mode of action and the interpretation of the findings. Earlier studies described a dose-dependent metabolic shift for allylalkoxybenzenes (Zangouras et al. 1981; Caldwell 1987; Smith et al. 2010). The predominant pathway at doses >10 mg/kg bw/day was supposed to lead to a reactive product, and at doses <10 mg/kg bw/day the primary pathway was presumed to effectively facilitate excretion with only limited formation of reactive intermediates (Smith et al. 2002; Punt et al. 2009). However, more recent evidence points to sulfation with human and murine SULTs of hydroxylated methyl eugenol metabolites as key to their metabolic activation, and resulting DNA adducts and mutagenicity in the Ames assay (Herrmann et al. 2012, 2014). Because DNA adducts have been detected at dose levels as low as 5 mg/kg bw/day in the liver (Ellis et al. 2007), the hepatic bioactivation of methyl eugenol at lower doses could not be definitively excluded and could be considered indicative of a direct genotoxicity risk upon metabolic activation. However, others (Williams et al. 2013) have proposed that only cumulative DNA damage that exceeds repair capacity leads to preneoplastic and neoplastic lesions, based on the observations that a cumulative (long-term) intake of up to 3000 mg/kg bw (spread over 16 weeks) in rodents (~26 mg/kg bw daily average; up to 62 mg/kg bw three times a week, intermittent intake) resulted in measurable DNA adducts

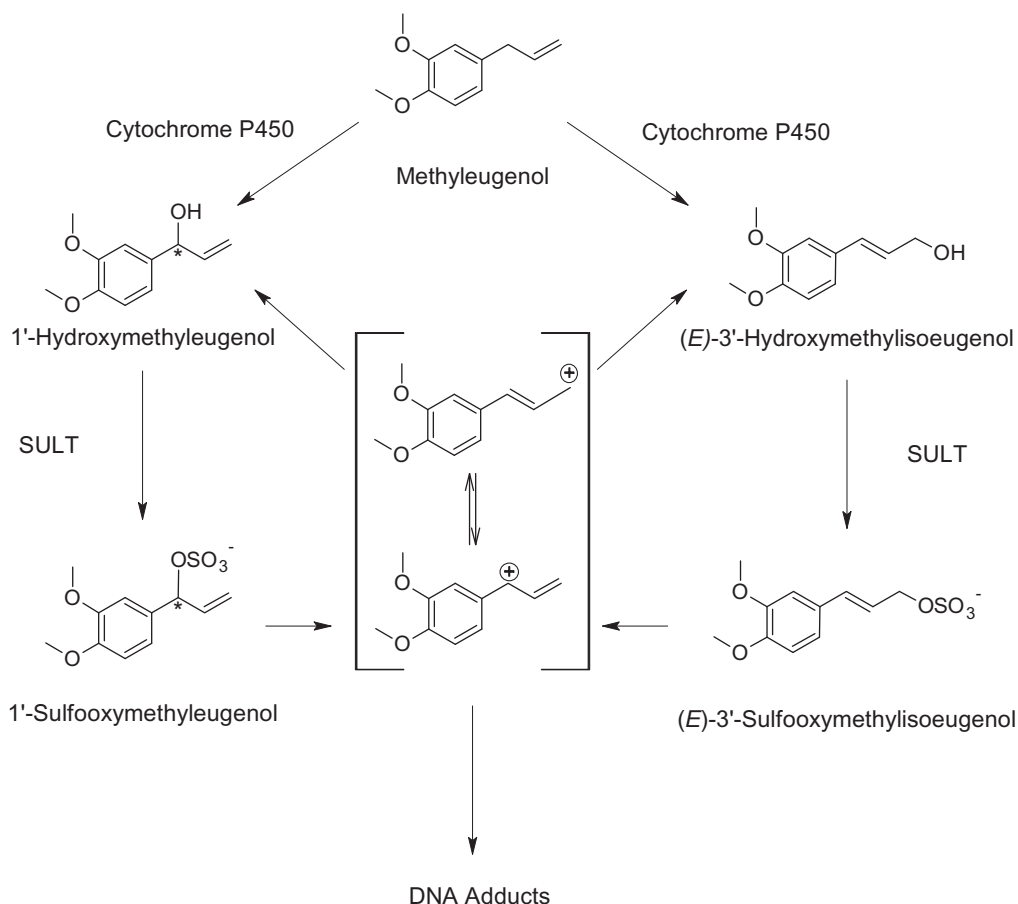


Figure 2. Metabolic pathways for methyleugenol leading to electrophilic metabolites. Chiral centers are marked with an asterisk. Reproduced from Herrmann et al. (2012).

but did not result in neoplasia even in the presence of a promoter (i.e. 500 ppm phenobarbital in the diet). Methyl eugenol-induced DNA adducts (up to 37 per 10^8 nucleosides or 4700 adducts per diploid genome) were detected in liver of 29 of 30 subjects (median of 13 per 10^8 nucleosides or 1700 adducts per diploid genome) (Herrmann et al. 2013). Considering that methyl eugenol occurs naturally in common herbs such as basil and fennel, this background of DNA adducts in humans may be related to chronic dietary exposure. The FEMA Expert Panel considered the overall evidence in its reevaluation of methyl eugenol in early 2015 and concluded that DNA adduct formation in humans was directly related to the formation of bioactivated metabolites. In humans, methyl eugenol undergoes bioactivation via 1'-hydroxylation and subsequent sulfation, forming reactive metabolites (Al-Subeihi et al. 2012; Herrmann et al. 2012; Herrmann et al. 2014) (Figure 2).

Given that the human SULTs are more effective than the murine counterparts but formation of the 1'-hydroxy intermediate is less efficient in humans than rodents (Al-Subeihi et al. 2012), updated information related to the dose-dependent metabolic fate and relative bioactivation and detoxication rates of methyl eugenol in humans and the potential for DNA repair is needed to better understand the importance of the relative increase of DNA adducts from its use as a flavoring substance and to support continuation of

its GRAS status. Until such information becomes available the FEMA Expert Panel concluded that methyl eugenol no longer met the criteria for GRAS status.

5. The weight of evidence and human relevance of genotoxicity testing findings

The interpretation of genotoxicity data by the FEMA Expert Panel, JECFA and EFSA is similar when considering the emphasis that is placed on identification of a genotoxic hazard and the adoption of a weight-of-evidence approach. Similar to the FEMA Expert Panel's approach, EFSA's weight-of-evidence approach integrates the evidence for all endpoints, assesses data quality and availability on genotoxicity itself and any other relevant data within the overall hazard assessment. Furthermore, EFSA emphasizes consideration of sources of uncertainty and their impact on the assessment outcome. This approach is particularly important in cases where, based on the standard battery of *in vitro* and *in vivo* genotoxicity assays, it is not possible to conclude on the absence of genotoxicity with confidence (i.e. standard or preferred battery of tests is not available or results are inconsistent) (Hardy et al. 2017). The FEMA Expert Panel and EFSA consider all data that may reduce uncertainty, such as mode of action, results of carcinogenicity studies, reproductive toxicity (indicative of germ cell DNA damage), toxicokinetic

studies, read-across from structurally related substances and predictions from QSAR models, and reliable data from non-standard tests/endpoints (e.g. DNA adducts). When evidence for genotoxicity of a flavoring substance is inconclusive, the FEMA Expert Panel does not proceed to finalize a safety evaluation (places its evaluation on hold), pending additional data submission. Similarly, EFSA requests additional data to reduce the uncertainty before concluding on genotoxicity.

The FEMA Expert Panel concludes that evidence of *in vivo* genotoxic hazard should be reviewed in the context of all relevant information, including carcinogenicity and developmental data if available. Negative carcinogenicity bioassay results can be used to interpret the human relevance of positive *in vivo* genotoxicity findings as discussed below. Other data can also be used to come to an overall conclusion regarding the *in vivo* genotoxic potential, as illustrated (*vide infra*).

5.1. Sufficient evidence of genotoxicity

In the reevaluation of 12 related thiophenes, the results of genotoxicity tests were interpreted within the context of specific structural features relevant to the reactivity of metabolic intermediates of these substances (Cohen et al. 2017a). In OECD guideline-compliant genotoxicity studies, 3-acetyl-2,5-dimethylthiophene was positive in Ames assays in the presence of metabolic activation, and an *in vivo* transgenic rodent mutation assay (MutaTMMouse) showed a dose-dependent increase in the mutant frequency of the *lacZ* transgene in the liver, which was statistically significant in the middle and high dose groups and exceeded both concurrent and historical control means. No mutagenicity was seen in the duodenum and no increases in micronucleated cells of the bone marrow. These results indicated that mutagenicity was limited to test conditions in the presence of metabolic activation and were consistent with the formation of genotoxic biotransformation products. Therefore, consideration of the metabolic fate of this flavoring substance was directly relevant to the interpretation of these results. Substituted thiophenes are subject to biotransformation by oxidative reactions, including S-oxidation and/or ring epoxidation/hydroxylation.

Detailed data on the metabolic fate of substituted thiophenes demonstrate that the path of oxidative transformation depends significantly on structural features such as (a) the presence and number of substitution groups, (b) the type of substitution groups, e.g. alkyl or acyl side chains, and (c) the location of substitution groups on the ring. Relative to the other thiophenes, 3-acetyl-2,5-dimethylthiophene contains unique features and, unlike the other members of the group, could not be expected to be metabolized to non-reactive intermediates and/or efficiently conjugated and readily excreted. Instead, it was the only member of the group that was predicted to produce reactive metabolic intermediates. For 3-acylthiophenes such as 3-acetyl-2,5-dimethylthiophene, S-oxidation (Figure 3, Panel B) is favored 5-fold over ring oxidation (Figure 3, Panel A). For 3-acetyl-2,5-dimethylthiophene, the methyl substituents in positions 2 and 5,

together with the position of the acyl group (3- versus 2-) result in a more reactive intermediate (S-oxide intermediates are thought to be more reactive compared to ring epoxide intermediates). Furthermore, the two additional substitution groups at positions 2- and 5- (where GSH conjugation typically occurs) prevent migration of the oxygen atom (Figure 3, Panel B, to product 5), dimer formation (Figure 3, product 6), and GSH conjugation (Figure 3, product 7) that would detoxify the reactive sulfoxide intermediate. The weight of evidence, in this case, led the FEMA Expert Panel to the conclusion that 3-acetyl-2,5-dimethylthiophene is a direct acting genotoxic substance and in the absence of additional information its GRAS status should be revoked (Cohen et al. 2017a).

5.2. Insufficient evidence of genotoxicity

The pitfalls often encountered in interpretation of *in vivo* genotoxicity testing results are illustrated in the case of perillaldehyde (*p*-mentha-1,8-dien-7-al), a naturally occurring cyclic α,β -unsaturated aldehyde.⁶ Due to its structure and mixed results in older, mostly non-guideline compliant genotoxicity studies, an *in vivo* comet/micronucleus combination assay was performed in rats at the request of a regulatory agency (EFSA). Rats were treated by oral gavage at dose levels of 175, 350, and 700 mg/kg bw/day. The pattern of DNA damage was described as a statistically significant increase in mean tail intensity only at the high dose (compared to concurrent control) and a statistically significant linear trend. This encompassed a small increase of mean tail intensity at the highest dose relative to the concurrent control. Mean tail intensity in the low and middle dose groups were similar to the average of the historical control range and none of the data including the high dose level exceeded the historical control range. The increased mean tail intensity was found driven by two animals in the high dose group. A direct correlation with biochemical and histopathological evidence of liver hepatocellular toxicity was observed. The study directors interpreted this pattern as consistent with a mode of action of DNA damage secondary to cytotoxicity. The results were interpreted differently by EFSA. Upon independent review, the FEMA Expert Panel concurred with the study directors that the results had no biological relevance if interpreted by the full criteria in the relevant Guideline (Cohen et al. 2016; Hobbs et al. 2016). The divergence of opinion by EFSA was based on the interpretation of a statistically significant increase as a sufficient criterion alone. Statistical analysis is instrumental in distinguishing random natural variation from changes large enough in magnitude to be considered non-random but attributable to the presence of the test substance. In recognition of this, the OECD Guidelines prescribe appropriate statistical tests and list specific criteria for consistent interpretation. According to OECD TG 489 guidelines, a positive result in the comet assay requires all acceptability criteria to be met. If these are not met, paragraph 62 offers some guidance (OECD 2016c). The final determination by EFSA that perillaldehyde is genotoxic was based on two of the three criteria for a positive test, namely the statistically

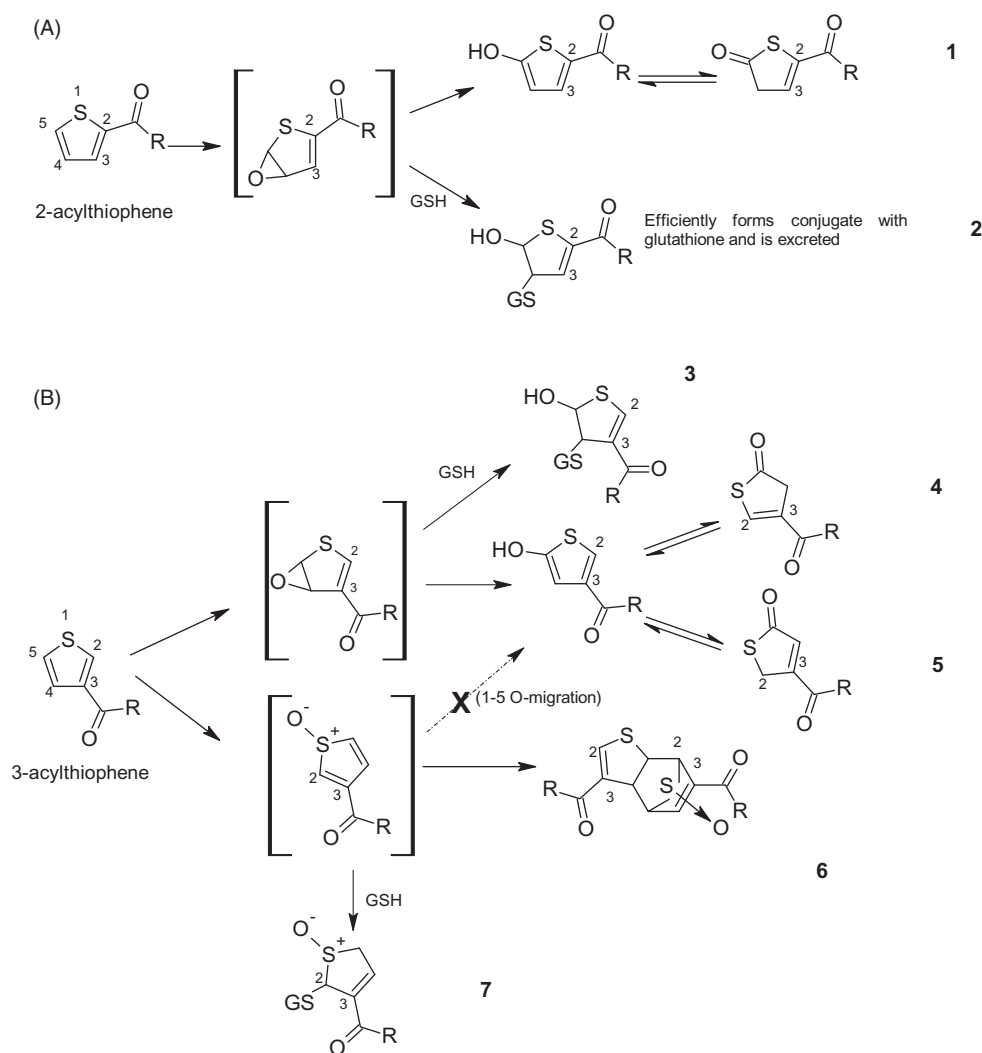


Figure 3. Possible metabolic pathways of substituted thiophenes. Substituted thiophene oxidation via sulfoxide and epoxide intermediates for 2-acylthiophene (A) compared to 3-acylthiophene (B), based on documented empirical evidence and *in silico* modeling. In the case of 3-acetyl-2,5-dimethylthiophene, S-oxidation is favored 5-fold over ring oxidation due to the position of the acyl group (3- versus 2-), and results in a more reactive intermediate (S-oxide intermediate) (as in B). Meanwhile, dimer formation (6) and GSH conjugation (7) are unlikely due to steric hindrance leaving the reactive intermediate (sulfoxide) to interact with cellular components including DNA. Reproduced with modifications from (Cohen et al. 2017).

significant difference between one treatment group from its concurrent control group, and statistical evidence of dose-response. The third criterion for a positive comet assay, that an increase in %Tail DNA must be outside of the historical control range, was not met and therefore expert judgment is essential. Expert judgment cannot dismiss the impact of toxicity at the high dose on the outcome of the comet assay and the absence of biological relevance. This is explicit in the Guideline which also proposes that in cases of confounded results “clinical chemistry measures can provide useful information on tissue damage and additional indicators such as caspase activation, TUNEL stain, Annexin V stain, etc. may also be considered” (para 55, OECD TG 489) (OECD 2016c). In the perillaldehyde study, the two animals with significant increases in tail intensity were also those with the most substantial evidence of hepatotoxicity, based on histopathology and elevated liver enzyme levels in the serum. In the weight of evidence, the other genotoxicity assays (reverse bacterial mutation assay, *in vitro* micronucleus, *in vitro* *hprt* mutation,

in vivo micronucleus) did not corroborate genotoxicity for perillaldehyde (Hobbs et al. 2016). The FEMA Expert Panel concluded that the findings of the comet assay were driven by the hepatotoxicity and thus not biologically relevant, and stated that disregard of the laboratory historical controls and interpretation of the data outside the OECD guidelines was neither appropriate nor justified (Cohen et al. 2016). Meanwhile, the recent JECFA evaluation concluded that the genotoxicity data for *p*-mentha-1,8-dien-7-al raise concerns for potential genotoxicity and, thus, it was not further considered by the Procedure for Safety Evaluation of Flavoring Agents (JECFA 2019).

Another recent example of the necessity of using the OECD guideline within the interpretation of the outcome of genotoxicity studies is furan-2(5*H*)-one.⁷ Due to the presence of the α , β -unsaturated carbonyl moiety within the structure, a battery of *in vitro* genotoxicity assays was requested for furan-2(5*H*)-one by EFSA (EFSA 2013b). In a standard OECD TG 471-compliant Ames assay the substance gave no

indication of mutagenic potential in *S. typhimurium* strains TA98, TA100, TA1535, TA1537, and TA102 (Bowen, 2011). Three independent OECD TG 487-compliant *in vitro* micronucleus assays were performed. In two separate micronucleus studies conducted in human peripheral blood lymphocytes, treatment of cells for a 3 h exposure period in the presence of S-9 metabolic activation followed by a 21-h recovery period resulted in statistically significant increases in the frequency of micronucleated binuclear cells. In the same studies, treatment of cells with furan-2(5H)-one for 3 h with a 21-h recovery period in the absence of S-9, and for 24 h with no recovery period in the absence of S-9, resulted in no increases in the frequency of micronucleated binuclear cells at concentrations at or below the OECD guideline-recommended cytotoxicity levels ($55 \pm 5\%$). The Panel concluded that treatment of cells with furan-2(5H)-one did result in increases in micronuclei when assayed in cultured human peripheral lymphocytes for 3 + 21 h in the presence of S-9 (Whitwell 2012; Watters 2013). In a third *in vitro* micronucleus assay conducted in TK6 cells, no increases in micronucleus frequencies were encountered at any concentration in any of the treatment arms of the study (4 h exposure in the presence of S-9 with a 20 h recovery period, 4 h exposure in the absence of S-9 with a 20 h recovery period, and 24 h exposure in the absence of S-9 without a recovery period) (Dutta 2018). There is not a clear explanation as to why there were differing outcomes within these studies. The Panel notes that *in vitro* micronucleus studies using TK6 cells are generally considered to be appropriately sensitive, if not more sensitive, than those conducted in human peripheral blood lymphocytes (Fowler et al. 2012a; Pfuhler et al. 2011; Fowler 2014; Whitwell et al. 2015; OECD 2016b).

The results from these three *in vitro* micronucleus studies are clearly inconsistent. In its review the Panel has noted that the two *in vitro* micronucleus studies conducted in human peripheral blood lymphocytes displayed exceedingly steep cytotoxicity curves and shifts from trial to trial in the cytotoxicity measured at the same or very similar concentrations. The cytotoxicity curves for the *in vitro* micronucleus assay in TK6 cells were less steep and it was correspondingly easier, it appears, to choose concentrations for scoring of micronuclei. There is not a clear explanation as to why there were differing outcomes within these studies.

To probe whether the *in vitro* effects could also be identified in an *in vivo* system, a comet/micronucleus combination assay was conducted in Han Wistar rats at 62.5, 125, and 250 mg/kg bw/day (with the doses set based on the results from a preliminary dose-range finder assay). There were no changes in clinical chemistry parameters. Decreases in glycogen vacuolation in the liver were reported in animals in the top dose group, while in the duodenum of the top dose group villous tip necrosis was observed. There were no increases in micronuclei induction, or increases in % tail DNA or % tail moment in the duodenum, observed at any tested doses. At the top dose of 250 mg/kg bw/day, small, less than two-fold increases in % tail DNA and tail moment were observed in the liver. The Panel notes that the increase in % tail DNA and tail moment at the top dose were within both the historical control range and the 95% reference range of

the historical controls. Additionally, overlap between tail DNA values were reported for concurrent vehicle control animals and those in the top dose group.

By applying the criteria to assess comet assay results as is described in OECD TG 489, the Panel concluded that the criteria for a clear positive outcome were not met, and when considered in combination with the negative bacterial reverse mutation outcome, the negative *in vivo* micronucleus results, and the inconsistent results in the *in vitro* micronucleus studies, concluded that based upon weight of evidence that furan-2(5H)-one did not display genotoxic potential. This conclusion is different than that reached by EFSA (EFSA 2019). In their opinion, EFSA did not apply one of the three criteria as outlined in OECD TG 489, stating that, "The Panel considered that the third criterion (*any of the results are outside the distribution of the historical negative control data for a given species, vehicle, route, tissue, and number of administrations*) mentioned in the OECD TG 489 was not applicable in this case because of the very wide range for historical negative controls reported (95% reference range for the vehicle control ranging from 0.02 to 11.39; 95% reference range for the positive control ranging from 7.15 to 65.07)."

5.3. Consideration of carcinogenicity studies

A well-conducted rodent carcinogenicity bioassay (or an equivalent modern alternative assay, e.g. a shorter duration study in transgenic animals) is sometimes proposed by some investigators as an option to confirm whether the observed genotoxicity has measurable biological consequences. However, a bioassay is rarely conducted as a follow up to investigate evidence of genotoxicity for flavoring substances. Although lifetime carcinogenicity studies have been conducted for several flavoring substances, most have been performed by the NTP, following nominations unrelated to the context of their use as flavors. Typically, the FEMA Expert Panel considers the combined evidence of available genotoxicity tests to be sufficient to determine whether a substance presents a genotoxic hazard, particularly when results are reproducible.

The FEMA Expert Panel reviews data of rodent carcinogenicity studies, when available, guided by criteria that determine relevance of the findings to humans and distinguish genotoxic from non-genotoxic modes of carcinogenicity (Hernandez et al. 2009). The FEMA Expert Panel recognizes that in some instances the pathogenesis of the observed tumors (and other endpoints) is not always relevant to humans based on three extensively documented limitations of the bioassay (Sonich-Mullin et al. 2001; Cohen 2004; Holsapple et al. 2006; Doi et al. 2007; Proctor et al. 2007; Boobis et al. 2016; Cohen and Arnold 2016):

- a. Tumors are often observed at the highest dose, typically the maximum tolerated dose (MTD) which may be associated with significant toxicity, qualitatively different kinetics, including different metabolism and biological activity, cell death with compensatory proliferation, conditions where DNA replication errors may be propagated.

- b. Species-specific differences have been described for the mode of action at the molecular level, such as the expression of $\alpha_2\mu$ -globulin in renal tubular epithelium in male rats, or the hormone feedback loops in thyroid function in rats (Doi et al. 2007; Hard 2018) explaining why many substances that induced rodent tumors are not carcinogens in humans (Boobis et al. 2016).
- c. Laboratory rodent species have shown increased background incidences of spontaneous lesions over time (Maronpot et al. 2016), such as liver tumors in mice or kidney chronic progressive nephropathy in rats. Future research on genetic or physiological differences between rodents and humans may reveal additional rodent-specific responses.

The findings of carcinogenicity studies may be investigated with follow up genotoxicity testing to determine whether a substance is a genotoxicant or produces tumors via a non-genotoxic mode of action. In addition, specific *in vitro* mechanistic studies contribute significantly to interpretation of human relevance of rodent lesions (e.g., providing evidence for irritation, oxidative stress, species-specific pathophysiology). As an example, forestomach tumors reported in rodents following gavage administration of *trans*, *trans*-2,4-hexadienal (NTP 2003) were concluded to be the result of a non-genotoxic mode of action (supported by the absence of mutagenicity in the Big Blue assay and to have resulted from tissue regeneration secondary to local irritation (Adams et al. 2008).

New models of *in vivo* carcinogenicity promise to further improve the validity of the outcomes for human safety, while also generating parallel genotoxicity data. These new models include: transgenic rodent models with increased sensitivity and significantly shorter duration (Cohen et al. 2001) and extended subchronic toxicity studies (e.g., 90-day studies) (Cohen 2010). These are designed to produce data on specific endpoints considered to be critical events in the pathogenesis of neoplastic lesions, such as pre-neoplastic lesions, evidence of increased cell proliferation, immunosuppression, interference with hormonal homeostasis, gene expression profiles associated with adverse outcome pathways, etc. (Cohen 2004; Holsapple et al. 2006; Boobis et al. 2009; Cohen 2010, 2017). Among transgenic mice models for carcinogenicity assessment developed in the 1990s (Cohen et al. 2001; ILSI/HESI 2001; Nambiar et al. 2012; Urano et al. 2012), the Tg.*rasH2* transgenic mouse model carrying the human prototype virus c-Ha-ras oncogene has the potential to reduce the length of in-life testing to 6 months (Shah et al. 2012). This model is responsive to genotoxic and non-genotoxic carcinogens and the animals have fewer and well-characterized spontaneous tumors, while the clinically relevant tumors are similar to those of the two-year bioassay (Paranjpe, et al. 2013; Paranjpe, et al. 2013). This model has gained regulatory acceptance (Sistare et al. 2011; Morton et al. 2014) and is widely used for testing of pharmaceuticals (Robinson and MacDonald 2001).

While the new models of carcinogenicity may be significant improvements in predicting human carcinogenicity, data from such models are not yet available for flavoring substances.

6. Conclusions

Many *in vitro* and *in vivo* mutagenicity and genotoxicity assays are available and can also be applied for testing flavoring substances. The FEMA Expert Panel does not require a standard battery for genotoxicity testing but will request relevant data as necessary to evaluate the genotoxic potential of a flavoring substance. When positive *in vitro* results are seen that are possibly biologically relevant, the FEMA Expert Panel generally considers the results from an *in vivo* micronucleus or comet assay or an appropriate *in vivo* mutagenicity assay to be helpful in addressing the question of genotoxic potential. Transgenic rodent mutation assays are highly sensitive in detecting *in vivo* mutagens. Due to the costs of the transgenic rodent mutation assays and the number of animals required, they are not routinely employed for flavoring substances. Instead, they have been used to confirm positive (or equivocal) *in vitro* genotoxicity tests or to probe the mechanisms of toxicity. Other assays are also included in the weight of evidence when flavoring substances are evaluated.

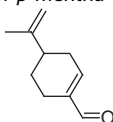
The role of genotoxicity assays in safety assessment is now well-established and can provide useful information on whether test substances are genotoxic hazards. Data from *in vitro* and *in vivo* genotoxicity assays are evaluated for study quality, biological relevance, and the relevance of the findings to humans. Biological relevance includes evaluation of cytotoxicity, dose-response relationship, variability and reproducibility of the assays, presence of functional DNA repair, and assay performance in detecting genotoxic substances. Human relevance includes the relevance of the metabolic system and the mode of action, including data from DNA adduct studies, carcinogenicity bioassays, and *in vitro* mechanistic tests.

The FEMA Expert Panel adopts a risk assessment approach in the evaluation of flavoring substances. Specifically, emphasis is placed on the weight of evidence of data from all assays, biological and human relevance, and exposure context to assess probable risk of genotoxicity to humans, including the potential for efficient metabolic detoxication and elimination, plausibility of genotoxic intermediate formation *in vivo*, and the context of background DNA lesions. Therefore, the weight-of-evidence approach adopted by the FEMA Expert Panel, as well as other regulatory and scientific expert bodies, is not limited to hazard assessment but aims for a realistic assessment of probable risk from consumer intake of flavoring substances under the conditions of use.

In view of the long history of safe use of naturally-derived and synthesized flavors in foodstuffs, it is important to note that only a very small percentage (2%) of flavoring substances evaluated were positive for mutagenicity or genotoxicity. Furthermore, substances are removed from the FEMA GRAS list when the weight of evidence no longer supports the definition of FEMA GRAS for these substances under the law (Cohen et al. 2016, 2017a). In conclusion, the FEMA Expert Panel uses genotoxicity data to aid their assessment of toxicity in a weight-of-evidence approach that leads to a realistic determination of probable risk for the consumer from intake of flavoring substances under the conditions of intended use.

Notes

1. Chemical group refers to the systematic organization of flavoring substances in groups according to specific structural features as adopted by the European Union and JECFA. See Sections 2.3 and 3.1 for explanation and the rest of the document.
2. JECFA list of monographs is available here: <https://www.who.int/foodsafety/publications/monographs/en/>
3. Commission Regulation No 1565/2000 of 18 July 2000 laying down the measures necessary for the adoption of an evaluation programme in application of Regulation (EC) No 2232/96. OJ L 180, 19.7.2000, p. 8–16. In addition to the chemical group designations, QSAR modeling was used to subdivide further substances with structural alerts for genotoxicity. Reference: EFSA (2008a). Genotoxicity test strategy for substances belonging to subgroups of FGE.19 [1] – Statement of the panel on food contact materials, enzymes, flavourings and processing aids (CEF). EFSA Journal. 854:1-5.
4. Deleted OECD guidelines on genetic toxicology include: Genetic Toxicology: *Escherichia coli*, Reverse Assay (OECD TG 472); Sex-linked recessive lethal test in *Drosophila melanogaster* (OECD TG 477); *In vitro* sister chromatid exchange assay in mammalian cells (OECD TG 479); *Saccharomyces cerevisiae*, gene mutation assay (OECD TG 480); *Saccharomyces cerevisiae*, mitotic recombination assay (OECD TG 481); Unscheduled DNA synthesis in mammalian cells *in vitro* (OECD TG 482) and mouse spot test (OECD TG 484).
5. For example, the Pig-a assay is a promising mutation assay for which there is both significant interest and completed work towards developing an OECD guideline: see https://www.oecd.org/chemicalsafety/testing/TGP%20work%20plan_September%202018.pdf and <http://www.oecd.org/chemicalsafety/testing/pig-a-gene-mutation-assay-detailed-review-paper.pdf>.
6. Structure of perillaldehyde or *p*-mentha-1,8-dien-7-al:



7. Structure of furan-2(5H)-one:



Acknowledgments

The authors gratefully acknowledge the assistance provided by Mr. Michael Armesto in preparing Tables 1 and 2 while he was an employee of Verto Solutions. The authors also acknowledge Ms. Dana Ramanan, an employee of Verto Solutions, for her assistance in identifying relevant literature for review and in the formatting of citations. These individuals were not considered for authorship as they were only involved in the preparation of the manuscript as described above, and they did not participate in the discussions of the contents of the manuscript, nor in developing the conclusions reached by the authors regarding those contents.

Declaration of interest

This work was supported by the Flavor and Extract Manufacturers Association (FEMA), a US-based trade association comprised of member companies that manufacture and/or use food flavorings. The FEMA Expert Panel is a group of scientists qualified by training and experience to conduct scientifically independent evaluations of the safety of food flavoring substances. The FEMA Expert Panel is scientifically and procedurally independent but is financially supported by FEMA. The employment affiliation of each co-author is declared above, and each of these authors participated in the review process and preparation of this paper as independent professionals and not as a representative of his or her employer. The manuscript reflects the knowledge and judgment of the

FEMA Expert Panel members as applied for the assessment of flavoring substances for GRAS status. The opinions expressed, and final conclusions set out in this overview paper were those of the listed authors and no one else.

Dr. Cohen is a member of the FEMA Expert Panel. Dr. Cohen has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Cohen to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Cohen does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Cohen reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Cohen receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS. Dr. Cohen was involved with the research on pulegone mode of action for bladder lesions in female rats that was supported by FEMA (Toxicol. Sci., 128: 1-8, 2012).

Dr. Eisenbrand is a member of the FEMA Expert Panel. Dr. Eisenbrand has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Eisenbrand to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Eisenbrand does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Eisenbrand reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Eisenbrand receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS. Dr. Eisenbrand is also a member of the Genotoxicity Adjunct Group of the Research Institute for Fragrance Materials, Inc. (Woodcliff Lake, NJ, USA), a group of scientific experts in toxicology who review and evaluate genotoxicity data of fragrance ingredients.

Dr. Fukushima is a member of the FEMA Expert Panel. Dr. Fukushima has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Fukushima to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Fukushima does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Fukushima reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Fukushima receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS. Dr. Fukushima has provided consulting services for pharmaceutical companies, chemical companies, and a contract research organization through his work with the Association for Promotion of Research on Risk Assessment (APRRA) and the Japan Bioassay Research Center (JBRC), but the consulting fees were paid to APRRA and JBRC, and Dr. Fukushima did not receive fees from the companies, APRRA, or JBRC for these services.

Dr. Gooderham is a member of the FEMA Expert Panel. Dr. Gooderham has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Gooderham to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Gooderham does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Gooderham reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Gooderham receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS. Dr. Gooderham did provide scientific contributions to a FEMA-sponsored PhD scholarship that was awarded to Dr. P. Carmichael of Imperial College London; Dr. Gooderham received no personal financial gain from his contributions to this award.

Dr. Guengerich is a member of the FEMA Expert Panel. Dr. Guengerich has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Guengerich to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Guengerich does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Guengerich reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Guengerich receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS.

Dr. Hecht is a member of the FEMA Expert Panel. Dr. Hecht has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Hecht to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Hecht does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Hecht reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Hecht receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS.

Dr. Rietjens is a member of the FEMA Expert Panel. Dr. Rietjens has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Rietjens to her position on the FEMA Expert Panel; rather, she was appointed by the other Expert Panel members following review of her qualifications and experience. Dr. Rietjens does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Rietjens

reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Rietjens receives an honorarium and reimbursement of expenses by FEMA for her service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS.

Dr. Rosol is a member of the FEMA Expert Panel. Dr. Rosol has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. FEMA did not appoint Dr. Rosol to his position on the FEMA Expert Panel; rather, he was appointed by the other Expert Panel members following review of his qualifications and experience. Dr. Rosol does not have a consulting relationship with any FEMA member company regarding anything to do with flavors in the context of the FEMA GRAS program. Within the context of safety evaluations by the FEMA Expert Panel for FEMA GRAS status for flavorings, Dr. Rosol reviews safety dossiers, including genotoxicity data, that are prepared and submitted by the applicant, without knowledge of the identity of the submitting company. As a member of the FEMA Expert Panel, Dr. Rosol receives an honorarium and reimbursement of expenses by FEMA for his service on the Panel. This financial remuneration is provided regardless of whether or not the FEMA Expert Panel concludes that any substances are GRAS.

Dr. Bastaki is employed by Verto Solutions which provides scientific and management support services to industry trade associations, including FEMA. Dr. Bastaki provides scientific support to FEMA, the FEMA Expert Panel and the International Organization of the Flavor Industry (IOFI). She serves as the Scientific Director of the International Association of Color Manufacturers (IACM). Dr. Bastaki has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. Dr. Bastaki has co-authored publications on the genotoxicity of food flavoring substances and food color additives. These publications were sponsored by IOFI, FEMA, and IACM.

Dr. Linman is employed by Verto Solutions which provides scientific and management support services to industry trade associations, including FEMA. Dr. Linman provides scientific support to FEMA and the FEMA Expert Panel. He serves as the Assistant Scientific Director of FEMA. Dr. Linman has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments.

Dr. Taylor is a Managing Director of Verto Solutions which provides scientific and management support services to industry trade associations, including FEMA. Dr. Taylor is the Scientific Secretary of the FEMA Expert Panel and the Scientific Director of the International Organization of the Flavor Industry. Dr. Taylor has no financial conflicts of interest related to the manuscript, and within the last 5 years has not appeared in legal proceedings, regulatory proceedings or advocacy roles related to the use of genotoxicity in flavor safety assessments. Dr. Taylor has co-authored publications on the genotoxicity of food flavoring substances.

ORCID

Nigel J. Gooderham  <http://orcid.org/0000-0001-7645-8219>

Samuel M. Cohen  <http://orcid.org/0000-0002-5047-0962>

Gerhard Eisenbrand  <http://orcid.org/0000-0003-4079-2463>

Shoji Fukushima  <http://orcid.org/0000-0001-6422-3316>

F. Peter Guengerich  <http://orcid.org/0000-0002-7458-3048>

Stephen S. Hecht  <http://orcid.org/0000-0001-7228-1356>

Ivonne M. C. M. Rietjens  <http://orcid.org/0000-0003-1894-3544>

Thomas J. Rosol  <http://orcid.org/0000-0003-3737-1190>

Maria Bastaki  <http://orcid.org/0000-0003-4487-1150>

Matthew J. Linman  <http://orcid.org/0000-0003-4243-2568>

Sean V. Taylor  <http://orcid.org/0000-0002-7806-4173>

References

- Adams TB, Cohen SM, Doull J, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Smith RL, Waddell WJ, et al. 2004. The FEMA GRAS assessment of cinnamyl derivatives used as flavor ingredients. *Food Chem Toxicol.* 42(2):157–185.
- Adams TB, Cohen SM, Doull J, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Smith RL, Waddell WJ, et al. 2005a. The FEMA GRAS assessment of benzyl derivatives used as flavor ingredients. *Food Chem Toxicol.* 43(8):1207–1240.
- Adams TB, Cohen SM, Doull J, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Smith RL, Waddell WJ, et al. 2005b. The FEMA GRAS assessment of hydroxy- and alkoxy-substituted benzyl derivatives used as flavor ingredients. *Food Chem Toxicol.* 43(8):1241–1271.
- Adams TB, Cohen SM, Doull J, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Smith RL, Waddell WJ, et al. 2005c. The FEMA GRAS assessment of phenethyl alcohol, aldehyde, acid, and related acetals and esters used as flavor ingredients. *Food Chem Toxicol.* 43(8): 1179–1206.
- Adams TB, Doull J, Goodman JI, Munro IC, Newberne P, Portoghese PS, Smith RL, Wagner BM, Weil CS, Woods LA, et al. 1997. The FEMA GRAS assessment of furfural used as a flavour ingredient. *Food Chem Toxicol.* 35(8):739–751.
- Adams TB, Gavin CL, McGowen MM, Waddell WJ, Cohen SM, Feron VJ, Marnett LJ, Munro IC, Portoghese PS, Rietjens IMCM, et al. 2011. The FEMA GRAS assessment of aliphatic and aromatic terpene hydrocarbons used as flavor ingredients. *Food Chem Toxicol.* 49(10): 2471–2494.
- Adams TB, Gavin CL, Taylor SV, Waddell WJ, Cohen SM, Feron VJ, Goodman JI, Rietjens IMCM, Marnett LJ, Portoghese PS, et al. 2008. The FEMA GRAS assessment of alpha, beta-unsaturated aldehydes and related substances used as flavor ingredients. *Food Chem Toxicol.* 46(9):2935–2967.
- Adams TB, Greer DB, Doull J, Munro IC, Newberne P, Portoghese PS, Smith RL, Wagner BM, Weil CS, Woods LA, et al. 1998. The FEMA GRAS assessment of lactones used as flavour ingredients. *Food Chem Toxicol.* 36(4):249–278.
- Adams TB, Hallagan JB, Putnam JM, Gierke TL, Doull J, Munro IC, Newberne P, Portoghese PS, Smith RL, Wagner BM, et al. 1996. The FEMA GRAS assessment of alicyclic substances used as flavour ingredients. *Food Chem Toxicol.* 34(9):763–828.
- Adams TB, McGowen MM, Williams MC, Cohen SM, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Smith RL, et al. 2007. The FEMA GRAS assessment of aromatic substituted secondary alcohols, ketones, and related esters used as flavor ingredients. *Food Chem Toxicol.* 45(2):171–201.
- Alhusainy W, Williams GM, Jeffrey AM, Iatropoulos MJ, Taylor S, Adams TB, Rietjens IMCM. 2014. The natural basil flavonoid neovadensin protects against a methyleugenol-induced marker of hepatocarcinogenicity in male F344 rat. *Food Chem Toxicol.* 74:28–34.
- Al-Subeihi AAA, Spengelink B, Punt A, Boersma MG, van Bladeren PJ, Rietjens IMCM. 2012. Physiologically based kinetic modeling of bioactivation and detoxification of the alkenylbenzene methyleugenol in human as compared with rat. *Toxicol Appl Pharmacol.* 260(3):271–284.
- Balbo S, Hashibe M, Gundy S, Brennan P, Canova C, Simonato L, Merletti F, Richiardi L, Agudo A, Castellsague X, et al. 2008. N2-ethyldeoxyguanosine as a potential biomarker for assessing effects of alcohol consumption on DNA. *Cancer Epidemiol Biomark Prev.* 17(11):3026–3032.
- Balbo S, Meng L, Bliss RL, Jensen JA, Hatsukami DK, Hecht SS. 2012. Kinetics of DNA adduct formation in the oral cavity after drinking alcohol. *Cancer Epidemiol Biomark Prev.* 21(4):601–608.
- Barlow S, Schlatter J. 2010. Risk assessment of carcinogens in food. *Toxicology and Applied Pharmacology.* 243(2):180–190.
- Barnes JL, Zubair M, John K, Poirier MC, Martin FL. 2018. Carcinogens and DNA damage. *Biochem Soc Trans.* 46(5):1213–1224.
- Bartsch H, Ohshima H, Pignatelli B, Calmels S. 1992. Endogenously formed N-nitroso compounds and nitrosating agents in human cancer etiology. *Pharmacogenetics.* 2(6):272–277.
- Basu AK. 2018. DNA damage, mutagenesis and cancer. *IJMS.* 19(4):970.
- Bhagat J. 2018. Combinations of genotoxic tests for the evaluation of group 1 IARC carcinogens. *J Appl Toxicol.* 38:81–99.
- Boobis AR, Brown P, Cronin MTD, Edwards J, Galli CL, Goodman JI, Jacobs A, Kirkland D, Luijten M, Marsaux C, et al. 2017. Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation. *Crit Rev Toxicol.* 47(8): 710–732.
- Boobis AR, Cohen SM, Dellarco VL, Doe JE, Fenner-Crisp PA, Moretto A, Pastoor TP, Schoeny RS, Seed JG, Wolf DC. 2016. Classification schemes for carcinogenicity based on hazard-identification have become outmoded and serve neither science nor society. *Regul Toxicol Pharmacol.* 82:158–166.
- Boobis AR, Cohen SM, Doerr NG, Galloway SM, Haley PJ, Hard GC, Hess FG, MacDonald JS, Thibault S, Wolf DC, et al. 2009. A data-based assessment of alternative strategies for identification of potential human cancer hazards. *Toxicol Pathol.* 37(6):714–732.
- Bowen R. 2011. Reverse mutation in five histidine-requiring strains of *Salmonella typhimurium*. Furan-2(5H)-one. Covance Laboratories Ltd. Study no. 8233099. September 2011. Unpublished report submitted by EFFA to FLAVIS Secretariat.
- Broustas CG, Lieberman HB. 2014. DNA damage response genes and the development of cancer metastasis. *Radiat Res.* 181(2):111–130.
- Brusick D. 1986. Genotoxic effects in cultured mammalian cells produced by low pH treatment conditions and increased ion concentrations. *Environ Mutagen.* 8(6):879–886.
- Brusick DJ. 1987. Implications of treatment-condition-induced genotoxicity for chemical screening and data interpretation. *Mutat Res.* 189(1): 1–6.
- Caldwell J. 1987. Human disposition of 14C-ORP/178. Unpublished report provided to the Expert Panel of the Flavor and Extract Manufacturers Association, Washington, DC, USA.
- Cohen SM. 2004. Human carcinogenic risk evaluation: An alternative approach to the two-year rodent bioassay. *Toxicol Sci.* 80(2):225–229.
- Cohen SM. 2010. An enhanced thirteen-week bioassay as an alternative for screening for carcinogenesis factors. *Asian Pac J Cancer Prev.* 11(1):15–17.
- Cohen SM. 2017. The relevance of experimental carcinogenicity studies to human safety. *Curr Opin Toxicol.* 3:6–11.
- Cohen SM, Arnold LL. 2016. Critical role of toxicologic pathology in a short-term screen for carcinogenicity. *J Toxicol Pathol.* 29(4):215–227.
- Cohen SM, Eisenbrand G, Fukushima S, Gooderham NJ, Guengerich FP, Hecht SS, Rietjens IMCM, Bastaki M, Davidsen JM, Harman CL, et al. 2019. FEMA GRAS assessment of natural flavor complexes: citrus-derived flavoring ingredients. *Food Chem Toxicol.* 124:192–218.
- Cohen SM, Eisenbrand G, Fukushima S, Gooderham NJ, Guengerich FP, Hecht SS, Rietjens IMCM, Davidsen JM, Harman CL, Taylor SV. 2018. Updated procedure for the safety evaluation of natural flavor complexes used as ingredients in food. *Food Chem Toxicol.* 113:171–178.
- Cohen SM, Fukushima S, Gooderham NJ, Guengerich FP, Hecht FM, Rietjens IMCM, Smith LM, Bastaki M, Harman CG, McGowen MM, et al. 2017a. Safety evaluation of substituted thiophenes used as flavoring ingredients. *Food Chem Toxicol.* 99(99):40–59.
- Cohen SM, Fukushima S, Gooderham NJ, Guengerich FP, Hecht SS, Rietjens IMCM, Smith RL, Bastaki M, Harman CL, McGowen MM, et al. 2016. FEMA expert panel review of p-mentha-1,8-dien-7-al genotoxicity testing results. *Food Chem Toxicol.* 98(Part B):201–209.
- Cohen SM, Fukushima S, Guengerich FP, Gooderham NJ, Hecht SS, Rietjens IMCM, Smith RL. 2017b. GRAS flavoring substances 28. Washington, DC: Flavor & Extract Manufacturers Association.
- Cohen SM, Robinson D, MacDonald J. 2001. Forum: alternative models for carcinogenicity testing. *Toxicol Sci.* 64(1):14–19.
- Cox JA, Fellows MD, Hashizume T, White PA. 2016. The utility of metabolic activation mixtures containing human hepatic post-mitochondrial supernatant (S9) for in vitro genetic toxicity assessment. *MUTAGE.* 31(2):117–130.
- Crespi CL, Gonzalez FJ, Steimel DT, Turner TR, Gelboin HV, Penman BW, Langenbach R. 1991. A metabolically competent human cell line expressing five cDNAs encoding procarcinogen-activating enzymes: application to mutagenicity testing. *Chem Res Toxicol.* 4(5):566–572.

- De Bont R, van Larebeke N. 2004. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*. 19(3):169–185.
- Ding W, Levy DD, Bishop ME, Pearce MG, Davis KJ, Jeffrey AM, Duan JD, Williams GM, White GA, Lyn-Cook LE, et al. 2015. In vivo genotoxicity of estragole in male F344 rats. *Environ Mol Mutagen*. 56(4):356–365.
- Doi AM, Hill G, Seely J, Hailey JR, Kissling G, Bucher JR. 2007. alpha2u-Globulin nephropathy and renal tumors in National Toxicology Program studies. *Toxicol Pathol*. 35(4):533–540.
- Dubois M, De Waziers I, Thome JP, Kremers P. 1996. P450 induction by Aroclor 1254 and 3,3',4,4'-tetrachlorobiphenyl in cultured hepatocytes from rat, quail and man: interspecies comparison. *Compar Biochem Physiol Part C, Pharmacol Toxicol Endocrinol*. 113(1):51–59.
- Dutta A. 2018. Furan-2(5H)-one: In vitro mammalian cell micronucleus assay in TK6 cells. Unpublished report provided by the International Organization of the Flavor Industry to the Expert Panel of the Flavor and Extract Manufacturers Association, Washington, DC, USA.
- Eder E, Scheckenbach S, Deininger C, Huffman C. 1993. The possible role of alpha, beta-unsaturated carbonyl compounds in mutagenesis and carcinogenesis. *Toxicol Lett*. 67(1–3):87–103.
- EFSA. 2008a. Genotoxicity test strategy for substances belonging to subgroups of FGE.19 [1] - Statement of the panel on food contact materials, enzymes, flavourings and processing aids (CEF). *EFSA J*. 854:1–5.
- EFSA. 2008b. List of alpha, beta-unsaturated aldehydes and ketones representative of FGE.19 substances for genotoxicity testing – statement of the panel on food contact materials, enzymes, flavourings and processing aids (CEF). *EFSA J*. 910:1–7.
- EFSA. 2011. Scientific Opinion on genotoxicity testing strategies applicable to food and feed safety assessment. *EFSA J*. 9(9):2379.
- EFSA. 2013a. Scientific Opinion on re-evaluation of one flavouring substance 3-acetyl-2, 5-dimethylthiophene [FL-no 15.024] from FGE.19 subgroup 52. *EFSA J*. 11(5):3227.
- EFSA. 2013b. Scientific Opinion on Flavouring Group Evaluation 217, Revision 1 (FGE.217Rev1). Consideration of genotoxic potential for α,β -unsaturated ketones and precursors from chemical subgroup 4.1 of FGE.19: lactones. *EFSA J*. 11(7):3304.
- EFSA. 2015a. Scientific Opinion on Flavouring Group Evaluation 99 Revision 1 (FGE.99Rev1): Consideration of furanone derivatives evaluated by the JECFA (63rd, 65th and 69th meetings). *EFSA J*. 13(11):4286.
- EFSA. 2015b. Scientific Opinion on Flavouring Group Evaluation 220 Revision 3 (FGE.220Rev3): consideration of genotoxic potential for α,β -unsaturated 3(2H)-Furanones from subgroup 4.4 of FGE.19. *EFSA J*. 13(5):4117.
- EFSA. 2016. Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. *EFSA Support Publ*. 13(3):1–50.
- EFSA. 2017. Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA J*. 15(8):e04971.
- EFSA. 2019. Scientific Opinion on Flavouring Group Evaluation 217 Revision 2 (FGE.217Rev2), consideration of genotoxic potential for α,β -unsaturated ketones and precursors from chemical subgroup 4.1 of FGE.19: lactones. *EFSA J*. 17(1):5568.
- Eisenbrand G, Schuhmacher J, Goelzer P. 1995. The influence of glutathione and detoxifying enzymes on DNA damage induced by 2-alkenals in primary rat hepatocytes and human lymphoblastoid cells. *Chem Res Toxicol*. 8(1):40–46.
- Ellis JK, Carmichael PL, Gooderham NJ. 2007. Toxicological assessment of low dose exposure to the flour methyl eugenol. Unpublished report provided to the Expert Panel of the Flavor and Extract Manufacturers Association, Washington, DC, USA.
- EPA. 2004. An examination of EPA risk assessment principles and practices. Staff Paper Prepared for the U.S. Environmental Protection Agency by members of the Risk Assessment Task Force, Office of the Science Advisor, Washington DC. EPA/100/B-04/001, March 2004.
- EPA. 2005. Guidelines for carcinogen risk assessment. Risk Assessment Forum. U.S. Environmental Protection Agency Washington, DC. EPA/630/P-03/001B, March 2005.
- EU. 2009. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on Cosmetic Products.
- Fellows MD, O'Donovan MR. 2007. Cytotoxicity in cultured mammalian cells is a function of the method used to estimate it. *Mutagenesis*. 22(4):275–280.
- Fellows M D, O'Donovan M R, Lorge E, Kirkland D. 2008. Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test. II: Practical aspects with toxic agents. *Mutation Res*. 655(1-2):4–21.
- Foth H, Degen GH, Bolt HM. 2005. New aspects in the classification of carcinogens. *Arh Hig Rada Toksikol*. 56(2):167–175.
- Fowler P. 2014. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. III: Sensitivity of human cell types to known genotoxic agents. *Mutation Res*. 767:28–36.
- Fowler P, Smith K, Young J, Jeffrey L, Kirkland D, Pfuhrer S, Carmichael P. 2012a. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. I. Choice of cell type. *Mutation Res*. 742:11–25.
- Fowler P, Smith R, Smith K, Young J, Jeffrey L, Kirkland D, Pfuhrer S, Carmichael P. 2012b. Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. II. Importance of accurate toxicity measurement. *Mutat Res*. 747(1):104–117.
- Frankel EN, Neff WE, Brooks DD, Fujimoto K. 1987. Fluorescence formation from the interaction of DNA with lipid oxidation degradation products. *Biochim Biophys Acta (BBA) – Lipids Lipid Metab*. 919(3):239–244.
- Fukushima S, Gi M, Fujioka M, Kakehashi A, Wanibuchi H, Matsumoto M. 2018. Quantitative Approaches to assess key carcinogenic events of genotoxic carcinogens. *Toxicol Res*. 34(4):291–296.
- Galloway SM. 2000. Cytotoxicity and chromosome aberrations in vitro: experience in industry and the case for an upper limit on toxicity in the aberration assay. *Environ Mol Mutagen*. 35(3):191–201.
- Gardner I, Bergin P, Stening P, Kenna JG, Caldwell J. 1996. Immunochemical detection of covalently modified protein adducts in livers of rats treated with methyleugenol. *Chem Res Toxicol*. 9(4):713–721.
- Geacintov NE, Broyde S. 2017. Repair-resistant DNA lesions. *Chem Res Toxicol*. 30(8):1517–1548.
- Glatt H. 2000. Sulfotransferases in the bioactivation of xenobiotics. *Chem Biol Interact*. 129(1-2):141–170.
- Glatt H, Meinel W. 2005. Sulfotransferases and acetyltransferases in mutagenicity testing: technical aspects. *Methods Enzymol*. London: Academic Press; p. 230–249.
- Glatt H, Sabbioni G, Monien BH, Meinel W. 2016. Use of genetically manipulated *Salmonella typhimurium* strains to evaluate the role of human sulfotransferases in the bioactivation of nitro- and aminotoluenes. *Environ Mol Mutagen*. 57(4):299–311.
- Glatt H, Schneider H, Murkovic M, Monien BH, Meinel W. 2012. Hydroxymethyl-substituted furans: mutagenicity in *Salmonella typhimurium* strains engineered for expression of various human and rodent sulphotransferases. *Mutagenesis*. 27(1):41–48.
- Gocke E, Muller L. 2009. In vivo studies in the mouse to define a threshold for the genotoxicity of EMS and ENU. *Mutation Res*. 678(2):101–107.
- Hallagan J, Hall R. 1995. FEMA GRAS – a GRAS assessment program for flavor ingredients. *Regul Toxicol Pharmacol*. 21(3):422–430.
- Hallagan J, Hall R. 2009. Under the conditions of intended use – new developments in the FEMA GRAS program and the safety assessment of flavor ingredients. *Food Chem Toxicol*. 47(2):267–278.
- Hard GC. 2018. Mechanisms of rodent renal carcinogenesis revisited. *Toxicol Pathol*. 46(8):956–969. 192623318797071.
- Hardy A, Benford D, Halldorsson T, Jeger M, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A. 2017. Clarification of some aspects related to genotoxicity assessment. *EFSA J*. 15(12):e05113.
- Hecht SS. 2003. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer*. 3(10):733–744.
- Hecht SS, McIntee EJ, Wang M. 2001. New DNA adducts of crotonaldehyde and acetaldehyde. *Toxicology*. 166(1-2):31–36.
- Hecht SS, Upadhyaya P, Wang M. 2011. Evolution of research on the DNA adduct chemistry of N-nitrosopyrrolidine and related aldehydes. *Chem Res Toxicol*. 24(6):781–790.

- Hernandez LG, van Steeg H, Luijten M, van Benthem J. 2009. Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach. *Mutation Res.* 682(2-3):94–109.
- Herrmann K, Engst W, Appel KE, Monien BH, Glatt H. 2012. Identification of human and murine sulfotransferases able to activate hydroxylated metabolites of methyleugenol to mutagens in *Salmonella typhimurium* and detection of associated DNA adducts using UPLC-MS/MS methods. *Mutagenesis.* 27(4):453–462.
- Herrmann K, Engst W, Meinel W, Florian S, Cartus AT, Schrenk D, Appel KE, Nolden T, Himmelbauer H, Glatt H. 2014. Formation of hepatic DNA adducts by methyleugenol in mouse models: drastic decrease by Sult1a1 knockout and strong increase by transgenic human SULT1A1/2. *Carcinogenesis.* 35(4):935–941.
- Herrmann K, Schumacher F, Engst W, Appel KE, Klein K, Zanger UM, Glatt H. 2013. Abundance of DNA adducts of methyleugenol, a rodent hepatocarcinogen, in human liver samples. *Carcinogenesis.* 34(5):1025–1030.
- Hobbs CA, Taylor SV, Beevers C, Lloyd M, Bowen R, Lillford L, Maronpot R, Hayashi S-m. 2016. Genotoxicity assessment of the flavouring agent, perillaldehyde. *Food Chem Toxicol.* 97:232–242.
- Holsapple MP, Pitot HC, Cohen SH, Boobis AR, Klaunig JE, Pastoor T, Dellarco VL, Dragan YP. 2006. Mode of action in relevance of rodent liver tumors to human cancer risk. *Toxicol Sci.* 89(1):51–56.
- Honda H, Fujita Y, Kasamatsu T, Fuchs A, Fautz R, Morita O. 2018. Necessity for retrospective evaluation of past-positive chemicals in in vitro chromosomal aberration tests using recommended cytotoxicity indices. *Genes Environ.* 40:2.
- Honda H, Minegawa K, Fujita Y, Yamaguchi N, Oguma Y, Glatt H, Nishiyama N, Kasamatsu T. 2016. Modified Ames test using a strain expressing human sulfotransferase 1C2 to assess the mutagenicity of methyleugenol. *Genes Environ.* 38(1):1–5.
- Honma M. 2011. Cytotoxicity measurement in in vitro chromosome aberration test and micronucleus test. *Mutation Res.* 724(1-2):86–87.
- Hummel JM, Madeen EP, Siddens LK, Uesugi SL, McQuistan T, Anderson KA, Turteltaub KW, Ognibene TJ, Bench G, Krueger SK, et al. 2018. Pharmacokinetics of [¹⁴C]-Benzo[a]pyrene (BaP) in humans: impact of co-administration of smoked salmon and BaP dietary restriction. *Food Chem Toxicol.* 115:136–147.
- ILSI/HESI. 2001. ILSI/HESI alternatives to carcinogenicity testing project. *Toxicol Pathol.* 29(Supplement):1–351.
- Ishii Y, Suzuki Y, Hibi D, Jin M, Fukuhara K, Umemura T, Nishikawa A. 2011. Detection and quantification of specific DNA adducts by liquid chromatography-tandem mass spectrometry in the livers of rats given estragole at the carcinogenic dose. *Chem Res Toxicol.* 24(4):532–541.
- JECFA. 1998a. 51st Joint FAO/WHO Expert Committee on Food Additives (JECFA) meeting – Food additives. WHO Technical Report Series. Geneva, Switzerland: World Health Organization.
- JECFA. 1998b. Safety evaluations of certain food additives and contaminants WHO Food Additives Series. Geneva, Switzerland: World Health Organization.
- JECFA. 2005. 63rd Joint FAO/WHO Expert Committee on Food Additives (JECFA) meeting – Food additives. WHO Technical Report Series. Geneva, Switzerland: World Health Organization.
- JECFA. 2006. Safety evaluations of certain food additives and contaminants WHO Food Additives Series. Geneva, Switzerland: World Health Organization.
- JECFA. 2016. 82nd Joint FAO/WHO Expert Committee on Food Additives (JECFA) meeting – Food additives. WHO Technical Report Series. Geneva, Switzerland: World Health Organization.
- JECFA. 2019. 86th Joint FAO/WHO Expert Committee on Food Additives (JECFA) meeting – Food additives. WHO Technical Report Series. Geneva, Switzerland: World Health Organization.
- Jia L, Liu X. 2007. The conduct of drug metabolism studies considered good practice (II): in vitro experiments. *CDM.* 8(8):822–829.
- Jin G, Cai L, Hu K, Luo Y, Chen Y, Glatt H, Liu Y. 2018. Mutagenic activity of N-nitrosodiethylamine in cell lines expressing human CYP2E1—adequacy of dimethylsulfoxide as solvent. *Environ Mol Mutagen.*
- Kirkland D. 2011. Improvements in the reliability of in vitro genotoxicity testing. *Expert Opin Drug Metab Toxicol.* 7(12):1513–1520.
- Kirkland D, Aardema M, Henderson L, Müller L. 2005. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. *Mutation Res.* 584(1-2):1–256.
- Kirkland D, Kasper P, Martus H-J, Müller L, van Benthem J, Madia F, Corvi R. 2016. Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. *Mutation Res.* 795:7–30.
- Kirkland D, Pfuhrer 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. *Mutation Res.* 628(1):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. *Mutation Res.* 654(2):114–132.
- Kirkland D, Zeiger E, Madia F, Gooderham N, Kasper P, Lynch A, Morita T, Ouedraogo G, Parra Morte JM, Pfuhrer S, et al. 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? I. Reports of individual databases presented at an EURL ECVAM Workshop. *Mutation Res.* 775–776:55–68.
- Kiwamoto R, Rietjens IMCM, Punt A. 2012. A physiologically based in silico model for trans-2-hexenal detoxification and DNA adduct formation in rat. *Chem Res Toxicol.* 25(12):2630–2641.
- Klapacz J, Pottenger LH, Engelward BP, Heinen CD, Johnson GE, Clewell RA, Carmichael PL, Adeleye Y, Andersen ME. 2016. Contributions of DNA repair and damage response pathways to the non-linear genotoxic responses of alkylating agents. *Mutation Res Rev Mutation Res.* 767:77–91.
- Kobets T, Williams GM. 2016. Chapter 2 – Thresholds for hepatocarcinogenicity of DNA-reactive compounds. In: Nohmi T, Fukushima S, editors. Thresholds of genotoxic carcinogens. Boston: Academic Press; p. 19–36.
- Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG, et al. 2004. Structure-based Thresholds of Toxicological Concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem Toxicol.* 42(1):65–83.
- Kurutas EB. 2015. The importance of antioxidants which play the role in cellular response against oxidative/nitrosative stress: current state. *Nutr J.* 15(1):71–71.
- Lao Y, Yu N, Kassie F, Villalta PW, Hecht SS. 2007. Analysis of pyridyloxobutyl DNA adducts in F344 rats chronically treated with (R)- and (S)-N'-nitrososornicotine. *Chem Res Toxicol.* 20(2):246–256.
- Liu B, Xue Q, Tang Y, Cao J, Guengerich FP, Zhang H. 2016. Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat Res Rev Mutat Res.* 768:53–67.
- Madeen E, Siddens LK, Uesugi S, McQuistan T, Corley RA, Smith J, Waters KM, Tilton SC, Anderson KA, Ognibene T, et al. 2019. Toxicokinetics of benzo[a]pyrene in humans: Extensive metabolism as determined by UPLC-accelerator mass spectrometry following oral micro-dosing. *Toxicol Appl Pharmacol.* 364:97–105.
- Marnett LJ, Cohen SM, Fukushima S, Gooderham NJ, Hecht SS, Rietjens IMCM, S R.L, Adams TB, H J.B, Harman C, et al. 2013. GRAS Flavoring Substances 26: The 26th publication by the Expert Panel of the Flavor and Extract Manufacturers Association provides an update on recent progress in the consideration of flavoring ingredients generally recognized as safe under the Food Additive Amendment. *Food Technol.* 67(8):38–56.
- Marnett LJ, Cohen SM, Fukushima S, Gooderham NJ, Hecht SS, Rietjens IMCM, Smith RL, Adams TB, Bastaki M, Harman CL, et al. 2014. GRASr2 evaluation of aliphatic acyclic and alicyclic terpenoid tertiary alcohols and structurally related substances used as flavoring ingredients. *J Food Sci.* 79(4):R428–R441.
- Maronpot RR, Nyska A, Foreman JE, Ramot Y. 2016. The legacy of the F344 rat as a cancer bioassay model (a retrospective summary of three common F344 rat neoplasms). *Crit Rev Toxicol.* 46(8):641–675.

- McCullough AK, Lloyd RS. 2019. Mechanisms underlying aflatoxin-associated mutagenesis – Implications in carcinogenesis. *DNA Repair*. 77: 76–86.
- Miele M, Dondero R, Ciarallo G, Mazzei M. 2001. Methyleugenol in *Ocimum basilicum* L. Cv. Genovese Gigante. *J Agric Food Chem*. 49(1): 517–521.
- Monien BH, Schumacher F, Herrmann K, Glatt H, Turesky RJ, Chesné C. 2015. Simultaneous detection of multiple DNA adducts in human lung samples by isotope-dilution UPLC-MS/MS. *Anal Chem*. 87(1):641–648.
- Morita T, Hamada S, Masumura K, Wakata A, Maniwa J, Takasawa H, Yasunaga K, Hashizume T, Honma M. 2016. Evaluation of the sensitivity and specificity of in vivo erythrocyte micronucleus and transgenic rodent gene mutation tests to detect rodent carcinogens. *Mutation Res/Genetic Toxicol Environ Mutagenesis*. 802:1–29.
- Morton D, Sistare FD, Nambiar PR, Turner OC, Radi Z, Bower N. 2014. Regulatory forum commentary: alternative mouse models for future cancer risk assessment. *Toxicol Pathol*. 42(5):799–806.
- Nambiar PR, Turnquist SE, Morton D. 2012. Spontaneous tumor incidence in rasH2 Mice: review of internal data and published literature. *Toxicol Pathol*. 40(4):614–623.
- Newberne P, Smith RL, Doull J, Goodman JI, Munro IC, Portoghese PS, Wagner BM, Weil CS, Woods LA, Adams TB, et al. 1999. The FEMA GRAS assessment of trans-anethole used as a flavouring substance. *Food Chem Toxicol*. 37(7):789–811.
- Nohmi T. 2018. Thresholds of genotoxic and non-genotoxic carcinogens. *Toxicol Res*. 34(4):281–290.
- NTP. 2003. NTP Technical Report on the Carcinogenesis Studies of 2,4-hexadienal (CAS No 142-83-6) in F344/N Rats and B6C3F1. NTP TR 509 ed. Research Triangle Park, NC: NIH Publication No. 04-4443. National Toxicology Program.
- O'Donovan M. 2012. A critique of methods to measure cytotoxicity in mammalian cell genotoxicity assays. *Mutagenesis*. 27(6):615–621.
- OECD. 1997. Test No. 471: Bacterial Reverse Mutation Test. OECD Publishing.
- OECD. 2013. Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays. OECD Publishing.
- OECD. 2014. Test No. 489: In Vivo Mammalian Alkaline Comet Assay. OECD Publishing.
- OECD. 2016a. Test No. 473: In Vitro Mammalian Chromosomal Aberration Test.
- OECD. 2016b. Test No. 487: In Vitro Mammalian Cell Micronucleus Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, <https://doi.org/10.1787/9789264264861-en>
- OECD. 2016c. Test No. 489: In Vivo Mammalian Alkaline Comet Assay.
- OECD. 2016d. Test No. 490: In Vitro Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Gene.
- OECD. 2017. Overview on genetic toxicology TGs, OECD Series on Testing and Assessment, No. 238, Paris: OECD Publishing.
- Paini A, Punt A, Scholz G, Gremaud E, Spenkelink B, Alink G, Schilter B, van Bladeren PJ, Rietjens IMCM. 2012. In vivo validation of DNA adduct formation by estragole in rats predicted by physiologically based biodynamic modelling. *Mutagenesis*. 27(6):653–663.
- Paini A, Scholz G, Marin-Kuan M, Schilter B, O'Brien J, van Bladeren PJ, Rietjens IMCM. 2011. Quantitative comparison between in vivo DNA adduct formation from exposure to selected DNA-reactive carcinogens, natural background levels of DNA adduct formation and tumour incidence in rodent bioassays. *Mutagenesis*. 26(5):605–618.
- Paranjpe MG, Elbekaei RH, Shah SA, Hickman M, Wenk ML, Zahalka EA. 2013. Historical control data of spontaneous tumors in transgenic CBYB6F1-Tg(HRAS)2Jic (Tg.rasH2) mice. *Int J Toxicol*. 32(1):48–57.
- Paranjpe MG, Shah SA, Denton MD, Elbekai RH. 2013. Incidence of spontaneous non-neoplastic lesions in transgenic CBYB6F1-Tg(HRAS)2Jic mice. *Toxicol Pathol*. 41(8):1137–1145.
- Patlewicz G, Wambaugh JF, Felter SP, Simon TW, Becker RA. 2018. Utilizing Threshold of Toxicological Concern (TTC) with high throughput exposure predictions (HTE) as a risk-based prioritization approach for thousands of chemicals. *Computat Toxicol*. 7:58–67.
- Peterson LA. 2017. Context matters: contribution of specific DNA adducts to the genotoxic properties of the tobacco-specific nitrosamine NNK. *Chem Res Toxicol*. 30(1):420–433.
- Pfuhler S, Fellows M, van Benthem J, Corvi R, Curren R, Dearfield K, Fowler P, Frotschl R, Elhajouji A, Le H L, et al. 2011. In vitro genotoxicity test approaches with better predictivity: summary of an IWGT workshop. *Mutation Res*. 723(2):101–107.
- Phillips DH, Reddy MV, Randerath K. 1984. 32 P-Post-labelling analysis of DNA adducts formed in the livers of animals treated with safrole, estragole and other naturally-occurring alkenylbenzenes. II. Newborn male B6C3F1 mice. *Carcinogenesis*. 5(12):1623–1628.
- Pottenger LH, Boysen G, Brown K, Cadet J, Fuchs RP, Johnson GE, Swenberg JA. 2019. Understanding the importance of low-molecular weight (ethylene oxide- and propylene oxide-induced) DNA adducts and mutations in risk assessment: Insights from 15 years of research and collaborative discussions. *Environ Mol Mutagen*. 60(2):100–121.
- Povey AC. 2000. DNA adducts: endogenous and induced. *Toxicol Pathol*. 28(3):405–414.
- Pratt IS, Barron T. 2003. Regulatory recognition of indirect genotoxicity mechanisms in the European Union. *Toxicol Lett*. 140-141:53–62.
- 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(2):313–326.
- Punt A, Paini A, Boersma MG, Freidig AP, Delatour T, Scholz G, Schilter B, Bladeren PJ, Rietjens IMCM. 2009. Use of physiologically based biokinetic (PBPK) modeling to study estragole bioactivation and detoxification in humans as compared with male rats. *Toxicol Sci*. 110(2): 255–269.
- Ramaiahgari SC, Waidyanatha S, Dixon D, DeVito MJ, Paules RS, Ferguson SS. 2017. From the cover: three-dimensional (3D) HepaRG Spheroid model with physiologically relevant xenobiotic metabolism competence and hepatocyte functionality for liver toxicity screening. *Toxicol Sci*. 159(1):124–136.
- Randerath K, Haglund RE, Phillips DH, Reddy MV. 1984. 32 P-Post-labelling analysis of DNA adducts formed in the livers of animals treated with safrole, estragole and other naturally-occurring alkenylbenzenes. I. Adult female CD-1 mice. *Carcinogenesis*. 5(12):1613–1622.
- Richardson SJ, Bai A, Kulkarni AA, Moghaddam MF. 2016. Efficiency in drug discovery: liver S9 fraction assay as a screen for metabolic stability. *DML*. 10(2):83–90.
- Robinson DE, MacDonald JS. 2001. Background and framework for ILSI's collaborative evaluation program on alternative models for carcinogenicity assessment. *Toxicologic Pathol*. 29 (Suppl):13–19.
- Scott D, Galloway SM, Marshall RR, Ishidate M, Brusick D, Ashby J, Myhr BC. 1991. Genotoxicity under extreme culture conditions. *Mutat Res/Rev Gen Toxicol*. 257(2):147–205.
- Shah SA, Paranjpe MG, Atkins PI, Zahalka EA. 2012. Reduction in the number of animals and the evaluation period for the positive control group in Tg.rasH2 short-term carcinogenicity studies. *Int J Toxicol*. 31(5): 423–429.
- Sistare F D, Morton D, Alden C, Christensen J, Keller D, Jonghe S D, Storer R D, Reddy M. V, Kraynak A, Trela B, et al. 2011. An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. *Toxicol Pathol*. 39(4):716–744.
- Smith RL, Adams TB, Doull J, Feron VJ, Goodman JI, Marnett LJ, Portoghese PS, Waddell WJ, Wagner BM, Rogers AE, et al. 2002. Safety assessment of allylalkoxybenzene derivatives used as flavouring substances — methyl eugenol and estragole. *Food Chem Toxicol*. 40(7):851–870.
- Smith B, Cadby P, Leblanc JC, Setzer RW. 2010. Application of the margin of exposure (MoE) approach to substances in food that are genotoxic and carcinogenic: example: methyleugenol, CASRN: 93-15-2. *Food Chem Toxicol*. 48 (Suppl 1):S89–S97.
- Smith RL, Cohen SM, Doull J, Feron VJ, Goodman JI, Marnett LJ, Munro IC, Portoghese PS, Waddell WJ, Wagner BM, et al. 2005. Criteria for the safety evaluation of flavoring substances. *Food Chem Toxicol*. 43(8):1141–1177.
- Smith RL, Cohen SM, Fukushima S, Gooderham NJ, Hecht SS, Guengerich FP, Rietjens IMCM, Bastaki M, Harman CL, McGowen MM, et al. 2018. The safety evaluation of food flavouring substances: the role of metabolic studies. *Toxicol Res*. 7(4):618–646.
- Smith RL, Waddell WJ, Cohen SM, Feron VJ, Marnett LJ, Portoghese PS, Rietjens IMCM, Adams TB, Lucas Gavin C, McGowen MM, et al. 2009.

- GRAS 24: the 24th publication by the FEMA Expert Panel presents safety and usage data on 236 new generally recognized as safe flavoring ingredients. *Food Technol.* 63(6):46–105.
- Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, et al. 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. *Regul Toxicol Pharmacol.* 34(2):146–152.
- Speit G, Kojima H, Burlinson B, Collins AR, Kasper P, Plappert-Helbig U, Uno Y, Vasquez M, Beevers C, De Boeck M, et al. 2015. Critical issues with the in vivo comet assay: a report of the comet assay working group in the 6th International Workshop on Genotoxicity Testing (IWGT). *Mutation Res Gen Toxicol Environ Mutagenesis.* 783:6–12.
- Swenberg JA, Lu K, Moeller BC, Gao L, Upton PB, Nakamura J, Starr TB. 2011. Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci.* 120(Supplement 1):S130–S145.
- Tornqvist M, Gustafsson B, Kautiainen A, Harms-Ringdahl M, Granath F, Ehrenberg L. 1989. Unsaturated lipids and intestinal bacteria as sources of endogenous production of ethene and ethylene oxide. *Carcinogenesis.* 10(1):39–41.
- Tremmel R, Herrmann K, Engst W, Meinel W, Klein K, Glatt H, Zanger UM. 2017. Methyleugenol DNA adducts in human liver are associated with SUL1A1 copy number variations and expression levels. *Arch Toxicol.* 91(10):3329–3339.
- Tritscher AM. 2004. Human health risk assessment of processing-related compounds in food. *Toxicol Lett.* 149(1–3):177–186.
- Urano K, Tamaoki N, Nomura T. 2012. Establishing a laboratory animal model from a transgenic animal: *Rash2* mice as a model for carcinogenicity studies in regulatory science. *Vet Pathol.* 49(1):16–23.
- Villalta PW, Hochalter JB, Hecht SS. 2017. Ultrasensitive high-resolution mass spectrometric analysis of a DNA adduct of the carcinogen Benzo[a]pyrene in human lung. *Anal Chem.* 89(23):12735–12742.
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. 2013. Cancer genome landscapes. *Science.* 339(6127):1546–1558.
- Wang M, Cheng G, Balbo S, Carmella SG, Villalta PW, Hecht SS. 2009. Clear differences in levels of a formaldehyde-DNA adduct in leukocytes of smokers and nonsmokers. *Cancer Res.* 69(18):7170–7174.
- Wang M, McIntee EJ, Cheng G, Shi Y, Villalta PW, Hecht SS. 2000. Identification of DNA adducts of acetaldehyde. *Chem Res Toxicol.* 13(11):1149–1157.
- Wang M, Yu N, Chen L, Villalta PW, Hochalter JB, Hecht SS. 2006. Identification of an acetaldehyde adduct in human liver DNA and quantitation as N2-ethyldeoxyguanosine. *Chem Res Toxicol.* 19(2):319–324.
- Watters G. 2013. Furan-2(5H)-one: Induction of micronuclei in cultured human peripheral blood lymphocytes. Unpublished report provided by the International Organization of the Flavor Industry to the Expert Panel of the Flavor and Extract Manufacturers Association, Washington, DC, USA.
- Whitwell J. 2012. Furan-2(5H)-one: Induction of micronuclei in cultured human peripheral blood lymphocytes of furan-2(5H)-one. Unpublished report provided by the International Organization of the Flavor Industry to the Expert Panel of the Flavor and Extract Manufacturers Association, Washington, DC, USA.
- Whitwell J, Smith R, Jenner K, Lyon H, Wood D, Clements J, Aschcroft-Hawley K, Gollapudi B, Kirkland D, Lorge E, et al. 2015. Relationships between p53 status, apoptosis and induction of micronuclei in different human and mouse cell lines in vitro: implications for improving existing assays. *Mutation Res Gen Toxicol Environ Mutagenesis.* 789–790:7–27.
- Williams GM, Iatropoulos MJ, Jeffrey AM, Duan JD. 2013. Methyleugenol hepatocellular cancer initiating effects in rat liver. *Food Chem Toxicol.* 53:187–196.
- Yager JD, Davidson NE. 2006. Estrogen carcinogenesis in breast cancer. *N Engl J Med.* 354(3):270–282.
- Yang J, Balbo S, Villalta PW, Hecht SS. 2019. Analysis of acrolein-derived 1, N(2)-propanodeoxyguanosine adducts in human lung DNA from smokers and nonsmokers. *Chem Res Toxicol.* 32(2):318–325.
- Zangouras A, Caldwell J, Hutt AJ, Smith RL. 1981. Dose dependent conversion of estragole in the rat and mouse to the carcinogenic metabolite, 1-hydroxyestragole. *Biochem Pharmacol.* 30(11):1383–1386.
- Zhang S, Villalta PW, Wang M, Hecht SS. 2006. Analysis of crotonaldehyde- and acetaldehyde-derived 1, n(2)-propanodeoxyguanosine adducts in DNA from human tissues using liquid chromatography electrospray ionization tandem mass spectrometry. *Chem Res Toxicol.* 19(10):1386–1392.