Accepted Manuscript

The 'BlueScreen HC' assay as a decision making test in the genotoxicity assessment of flavour and fragrance materials

Sylvain Etter, Louise Birrell, Paul Cahill, Heather Scott, Nick Billinton, Richard M. Walmsley, Benjamin Smith

PII: S0887-2333(15)00098-3

DOI: <http://dx.doi.org/10.1016/j.tiv.2015.05.005>

Reference: TIV 3530

To appear in: *Toxicology in Vitro*

Received Date: 14 October 2014

Accepted Date: 3 May 2015

Please cite this article as: Etter, S., Birrell, L., Cahill, P., Scott, H., Billinton, N., Walmsley, R.M., Smith, B., The 'BlueScreen HC' assay as a decision making test in the genotoxicity assessment of flavour and fragrance materials, *Toxicology in Vitro* (2015), doi: <http://dx.doi.org/10.1016/j.tiv.2015.05.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Title:

The 'BlueScreen HC' assay as a decision making test in the genotoxicity assessment of flavour and fragrance materials

Authors:

Sylvain Etter¹; Louise Birrell²; Paul Cahill²; Heather Scott²; Nick Billinton²; Richard M Walmsley^{2,3}; Benjamin Smith¹

Affiliations:

¹Firmenich SA, Meyrin, Switzerland

²Gentronix Ltd, Alderley Park, United Kingdom

³University of Manchester, Manchester, United Kingdom

Corresponding author:

Professor Richard M Walmsley

Address:

University of Manchester

Michael Smith Building

Oxford Road

Manchester M13 9PL, UK

Telephone number: +44 (0) 161 3062746

Email address: richard.walmsley@manchester.ac.uk

1. Title:

The BlueScreen HC assay as a decision making test in the genotoxicity assessment of flavour & fragrance materials

2. Abstract

The genotoxicity of a library of 70 flavour and fragrance substances having a high proportion of *in vivo* and/or carcinogenicity test data has been assessed using the GADD45a-GLuc 'BlueScreen HC' genotoxicity assay, with and without exogenous metabolic activation. There are only limited genotoxicity and carcinogenicity study data for compounds in this applicability domain, but this study allowed the following conclusions: **i.** the BlueScreen HC results are highly predictive of positive results from regulator-required *in vitro* genotoxicity assays for the test set of materials; the moderate negative predictivity of BlueScreen HC from the *in vitro* test set of material is mainly due to the high rate of false positive in regulatory *in vitro* mammalian tests **ii.** BlueScreen HC negative results are predictive of negative *in vivo* results and provide a specific prediction of *in vivo* genotoxicity assay results. **iii.** in this applicability domain, which comprises a large proportion of relatively low molecular weight molecules, a 1 mM testing limit maintains the sensitivity of the assay, and increases specificity. **iv.** the predictive capacity and specificity to *in vivo* genotoxins and carcinogens, coupled to a microplate format with low compound requirement supports further investigation of the BlueScreen HC assay as a useful tool in prioritizing the assessment of new F&F materials and in filling data gaps on materials with no or limited regulatory test data for genotoxicity .

Abbreviations: AIT: Allyl Isothiocyanate; CN: Citronellyl Nitrile; BSHC: BlueScreen HC; CA: chromosome aberration; CPDB: Carcinogenic Potency Database; ECHA: European Chemicals Agency; ECVAM: European Centre for the Validation of Alternative Methods; EFSA: European Food Safety Authority; EURL: European Union Reference Laboratory for alternatives to animal testing; FEMA: Flavor and Extract Manufacturers Association; F&F: Flavor and Fragrance; GN: Geranyl Nitrile, GRAS: Generally Recognised As Safe; GSHC: GreenScreen HC; Gluc: *Gaussia* luciferase; HPRT: Hypoxanthine Phosphoribosyl Transferase Assay; IARC: International Agency for Research on Cancer; FAO: Food and Agriculture Organization of the United Nations; WHO: World Health Organisation; JECFA: Joint (FAO/WHO) Expert Committee on Food Additives; MLA: Mouse Lymphoma Assay; MN: Micronucleus Test ; MW: Molecular Weight; REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals; RIFM: Research Institute for Fragrance Materials; SCCS: Scientific Committee on Consumer Safety; SCCP: Scientific Committee on Consumer Product; UK COM: United Kingdom Committee on Mutagenicity of Chemicals in Food.

Corresponding author: Prof. Richard M Walmsley.
 Telephone numbers: 0161 3062746 or 01625 238743
 Email address: richard.walmsley@manchester.ac.uk

3. Introduction.

Most of the international authorities dealing with fragrance materials have enforced regulations demanding a tonnage-dependant, stepwise testing approach for genotoxicity. This begins with a battery

of *in vitro* tests that may be followed by *in vivo* tests under strict conditions and authorization by the relevant authorities. For instance, for chemicals in Europe (EC 1907/2006), no toxicological tests are required for a tonnage below 1 ton per year (tpy). An Ames test is mandatory for a tonnage between 1 and 10 tpy. However, above 10 tpy both an additional *in vitro* mammalian mutagenicity test (MLA or HPRT) and an additional *in vitro* mammalian clastogenicity/aneugenicity test (CA or MN) are required. *In vivo* genotoxicity tests are only allowed as a follow-up of positive *in vitro* genotoxicity test(s). If a substance is negative in the *in vitro* battery, no further testing is required and the substance is considered as non genotoxic. Where positive *in vitro* test results are produced, confirmatory *in vivo* tests may be performed, unless the substance is a cosmetic ingredient, in which case animal testing is prohibited (EC 1223/2009).

Since many fragrance ingredients are produced in low tonnages, the Ames test is the main test used to assess genotoxic potential. However, the Ames test uses bacterial cells, and only assesses the possibility of gene mutations. It does not determine the potential for eukaryote-specific effects such as aneuploidy and clastogenesis, and therefore has a lower sensitivity than the *in vitro* mammalian cell tests. Hence to strengthen the assessment of fragrance materials, there is a need to cover the additional endpoints with test data in a practical and efficient way. Since none of the current regulatory approved *in vitro* assays assess all of the three main genotoxicity endpoints, test batteries are required. However, a recent review of historical data revealed that the *in vitro* mammalian cell tests in the battery are particularly prone to the generation of positive results for compounds found to produce negative results in animal genotoxicity and carcinogenicity studies ("misleading" or "false" positives: Kirkland *et al.*, 2005, 2007). These misleading positives can trigger needless *in vivo* studies, in a context where there is political ambition to reduce, refine or replace animal testing. In the absence of animal tests the development of promising substances might be abandoned, or substances already present on the market might be withdrawn. Part of the solution to this problem is the use of *in silico* predictive tools including knowledge based expert system such as Derek Nexus (<http://www.lhasalimited.org/products/derek-nexus.htm>), together with systems assessing quantitative structure activity relationships (QSAR: <http://www.qsartoolbox.org/>). The Fragrance Industry has relied on "read across", which the European Chemicals Agency (ECHA) describes as "a technique for predicting endpoint information for one substance (target substance), by using data from the same endpoint from (an)other substance(s), (source substance(s))". However, there remains the requirement for more accurate *in vitro* test data on which to anchor such tools and provide a more robust addition to the weight of evidence considerations.

The development and use of reliable early screening tests that can accurately predict true negatives is a significant part of the solution to the 'false positive' problem. Indeed a recent retrospective review of data from 249 materials from the cosmetics field asked very directly "whether trying to improve assays, which were never designed to have a high specificity but to be very sensitive, is the way forward", concluding that "high priority should be given to the development of strategies that incorporate new *in vitro* assays in a weight-of-evidence approach to safety assessment" (Ates *et al.*, 2014). Definitive negative results produced in an early screen can be used to prioritise compounds for development, and where those selected produce conflicting results in subsequent regulatory testing, the screening result can contribute to weight of evidence considerations. It is in this context the GADD45a-galactosyl transferase 'BlueScreen HC' ('BSHC') assay has been evaluated as a decision making test in the genotoxicity assessment of fragrance (and also related flavour) materials. The BSHC detects gene mutagens, aneuploids and clastogens (Hughes *et al.*, 2012), and has protocols for use with and without metabolic activation by liver S9 and, therefore, may detect a wider set of genotoxicants. It was anticipated for

example that amongst the BSHC positive substances, there would be misleading Ames negative results generated by eukaryote specific genotoxins. The performance of BSHC has been validated using the lists of chemicals recommended by ECVAM for the assessment of new or improved genotoxicity tests (Kirkland *et al.*, 2008) and was found to have similar sensitivity (ECVAM Group 1, expected positives, 80%) and, importantly, specificity (ECVAM Group 2, expected negatives) to the widely used GADD45a-Green Fluorescence Protein (GFP) reporter assay “GreenScreen HC” (GSHC) assay (Hastwell *et al.*, 2006, 2009; Billinton *et al.*, 2008, 2010, Birrell *et al.*, 2010., Jagger *et al.*, 2009). Whilst the assay produces an indicator endpoint, induction of GADD45a expression, rather than an apical endpoint (mutation/chromosome damage), it is none-the-less very accurate. For example, whilst many compounds can induce apoptosis, non-genotoxic apoptogens do not induce GADD45a (Topham *et al.*, 2012). The genotoxicity data from GSHC assay validation studies, and inter-laboratory assessments of transferability and reproducibility were submitted for approval by EURL ECVAM in April 2013, however the test data set is not specifically focussed on the chemical space that comprises F&F ingredients (i.e. small, low MW, volatile materials).

In order to test the ability of BSHC to predict higher tier *in vivo* genotoxicity and carcinogenicity assays in the F&F domain, it was evaluated using a selection of substances with as full data sets as possible. Whilst *in vivo* tests are seldom used in the F&F industry, this study was able to identify a collection of 70 substances reflecting a broad range of chemical structures commonly encountered in the F&F industry (Table 1). This included several materials found in the ECVAM recommended lists for the validation of new/improved genotoxicity test. The majority have either carcinogenicity (21 materials) and/or *in vivo* genotoxicity data (36 materials). BSHC is now reported for all of them. The aims of this initial study were to 1.) assess the utility of BSHC in the F&F material applicability domain and to produce the first full data set for a collection of F&F materials from a single genotoxicity assay and 2.) provide a rational basis by which to judge whether or not BSHC could be an effective prioritization tool prior to regulatory genotoxicity assessment.

4. Materials and Methods:

a. Selection of test materials

The 70 F&F material along with consolidated *in vivo*/carcinogenicity test data considered in this study are listed in Table 1. Data from 8 F&F materials belonging to the ECVAM list were taken from Hughes *et al.*, 2012. The majority of the remaining materials were supplied by Firmenich. Allyl isothiocyanate (AIT), Geranyl Nitrile (GN) and Glyoxal were purchased Sigma-Aldrich. Glyoxal is not an F&F ingredient *per se* but is often found in finished cosmetic products as residue from polymerization reactions, in fermented food and beverages due to microbial activity, and as a non-enzymatic autoxidation of oil or browning reactions of saccharides. Furthermore, it shares structural similarities with a number of fragrance materials and was thus included as an additional control material. In addition to the purchased materials, all of the other materials were of commercial quality and met industry and regulatory specifications. Purity of “market quality” materials can only be given as a range. CAS numbers are provided except for proprietary substances. This was considered a large enough collection to allow an initial assessment of assay performance, as the library contains materials with significant *in vitro* and *in vivo* existing genotoxicity data available. The substances were chosen to represent a broad range of chemical structures common in the F&F industry. They included for instance esters, ketones, aldehydes, nitriles, alcohols and thiols (Table 1). Sixty one (61) of the materials had some existing test data for the

mutagenicity, clastogenicity or aneugenicity endpoints (*in vitro* and/or *in vivo*), and these were selected to challenge BSHC with substances already evaluated with regulatory tests. Within this subset, 25 materials have at least one positive genotoxicity test result, and the remaining 36 have negative genotoxicity test data. Some materials indicated to be of putative concern (with genotoxicity still under investigation) were intentionally selected in order to challenge BSHC. The collection includes 18 proprietary compounds for which only the chemical class is described (Table 1). All tested negative. Their inclusion provides a broader view of the incidence of genotoxicity in this applicability domain which has relatively little published data. Nine of these materials, mainly low or zero tonnage had no genotoxicity data at all.

From the 10 F&F materials also present in the ECVAM recommended lists (Kirkland *et al.*, 2008), 4 materials (limonene, cyclohexanone, tert-butyl alcohol, pyridine) belong to the ECVAM Group 2 of chemicals that should give negative results in *in vitro* mammalian test and 6 belong to the ECVAM group 3, 'false positive' chemicals (menthol, benzyl alcohol, isobuteraldehyde, ethyl acrylate, resorcinol, eugenol) which have been reported to induce chromosomal aberrations or TK mutations in mouse lymphoma cells. All 10 are non-DNA-reactive chemicals. Other genotoxicity data for comparative purposes were collated from the published literature and proprietary tests reports from industry testing programs that were made available to the authors of this study.

b. The GADD45a-Gluc BlueScreen HC assay.

The BSHC assay provides a quantitative genotoxicity hazard assessment, and has been described in detail by Hughes *et al* (2012), with examples of dose-response data showing positive results for genotoxins and pro-genotoxins and negative results for non-genotoxins. It exploits human-derived TK6 lymphoblastoid cells, which in contrast to many cell lines used historically in genetic toxicology (V79, CHO etc.) are p53 competent and hence maintain an effective response to DNA damage. TK6 was recently more generally recommended for genotoxicity testing (Fowler *et al.*, 2014).

The decision to investigate the utility of the BSHC assay in the F&F sector was drawn from the published validation studies on the closely related GADD45a-Green Fluorescence Protein (GFP) reporter assay GSHC. GSHC could not be used for this study, because of current patent restrictions on the use of GFP. Both assays use TK6 cells containing a reporter plasmid which includes the promoter and expressed sequences of the GADD45a gene, except that a reporter gene (Gaussia luciferase in BSHC, GFP in GSHC) replaces sequences between translational start and the p53 response element in intron 3. The latter has an essential role in the proper regulation of GADD45a (Johnson *et al.*, 2013). The published validation exercises of both GSHC and BSHC have included the ECVAM lists of chemicals which include diverse mechanistic classes of genotoxins.

96 well microplate set-up. 4 compounds were each tested in duplicate across 8, two-fold dilutions, along with in-plate controls to determine data acceptability. Controls included the following: different concentrations of a control genotoxin to confirm that the assay is producing an appropriate dose response; growth medium only to identify contamination; test compound alone to determine interference from optical properties of test compounds. Samples were prepared in 2 % DMSO at a concentration of 20 mM unless limited by solubility, which is halved to 10 mM (1% DMSO) by the addition of double strength growth media during plate set up. Incubation was for 48 hours at 37 °C in a CO₂ incubator. In the S9 protocol, samples were incubated for 3 hours with 1 % S9 fraction (Aroclor-1254

induced male Sprague-Dawley rat extracts Moltox Inc., Boone, NC, USA), then washed and suspended in fresh growth medium for 45 hours before data collection.

Data collection. GADD45a-Gluc reporter luminescence measurement: coelenterazine was prepared as a 5-mM stock solution in acidified methanol and stored at -80°C . Approximately 30 min before use, it was diluted to $2.5\ \mu\text{M}$ in a buffer consisting of 400 mM Tris-HCl and 5 mM β -cyclodextrin, adjusted to pH 7.4. Luminescence data were collected using a Tecan Infinite F500 plate reader (Tecan UK Ltd., Reading, UK) fitted with an injector syringe system. $50\ \mu\text{L}$ of the diluted coelenterazine solution was injected into each well, shaken within the microplate reader for 2 s, before data collection (OD1 attenuation; integration time of 5 s).

Cell density measurement: thiazole orange (TO) binds to RNA and DNA and fluoresces brightly, providing a sensitive measure of cell density in low culture volumes (Hughes *et al.*, 2012). TO reagent was prepared by combining $50\ \mu\text{L}$ of TO (12.5 mM in 100 % DMSO) with 10 mL of a solution containing 4 % v/v Triton X-100 solution in D-PBS. A $50\ \mu\text{L}$ aliquot of this was added to each well of the microplate, and fluorescence data were collected using standard fluorescein (FITC) filters with excitation at 485 nm and emission at 535 nm.

Data handling. The TO relative cell density (RCD) data was used to normalise the reporter output to produce a “brightness” value, so that a small number of strongly light emitting cells (strongly genotoxic) can be distinguished from a high number of weakly light emitting cells (not genotoxic).

Decision thresholds: The threshold for determining a positive genotoxicity result for BSHC was derived by calculating the mean and standard deviation of the brightness values from assays of toxic and non-toxic, non-genotoxic compounds. This mirrored the approach used for the GADD45a-GFP assay. A positive result is concluded where any dose produces a brightness value greater than three standard deviations above the value from the solvent control. This produces a threshold of 1.8-fold increase in assays without S9 incubation and a 1.5 fold increase for the S9 protocol. RCD (see iii) data qualify genotoxicity data: data were rejected at $\text{RCD} < 30\%$. Compounds that produced a single positive BSHC dose, and compounds that showed an upward, sub-threshold trend before toxicity led to data rejection, were retested over a shallower dilution. The lowest toxic effect concentration (‘LEC’) is defined as $\text{RCD} = 80\%$.

c. Genotoxicity Data Analysis

A major reason for carrying out genotoxicity testing is to detect potential genotoxic carcinogens. For the limited number of substances with animal carcinogenicity and genotoxicity data, sensitivity and specificity values have been calculated i.e. the percentage of animal genotoxic carcinogens/non-carcinogens producing positive/negative BSHC data. Genotoxins can produce other, heritable genetic defects, though the data base for these is very small, and they are not considered further here. Positive and negative prediction values have also been generated i.e. the percentage of BSHC positive (or negative) substances producing positive (or negative) data in a different *in vitro* or animal test. Prediction values are significant for an early screening programme because data are often used to select substances appropriate for further development. It is important to know how accurately a new test predicts the first tier of the regulatory *in vitro* regulatory genotoxicity test battery as this will determine how substances are prioritised at a later stage.

d. Selection of top testing dose.

In the pharmaceutical sector, there has recently been a reduction from 10 mM to 1 mM in the top testing dose required for the *in vitro* mammalian tests (ICH S2R). This arose in part from the observation that many non-carcinogens produced misleading *in vitro* positive results at high concentrations, and it was especially true for low molecular weight compounds. A comprehensive data review concluded that the lower dose would maintain sensitivity but increase specificity (Parry *et al.*, 2010). Since many of the compounds in this study fall into the low molecular weight category, published GADD45a assay study results from 161 compounds were reviewed in order to discover how a reduction in top testing dose from 10 mM to 1 mM would affect its performance. There were 5 substances which produced positive results using a 10 mM top dose that would be re-classified as negative using a 1 mM top dose: two antiviral nucleoside analogues that affect nucleotide pool sizes *in vitro* and *in vivo*, and adversely affect DNA replication and repair (valacyclovir, a non- carcinogen, and azidothymidine, a high dose female-specific rodent carcinogen 11,600 mg/kg/day (i.e. misleading/irrelevant positives); EDTA and p-nitrophenol, which produce artifactual positive results in the 3 hour S9 exposure protocol at concentrations that are unacceptably toxic (< 30 % RCD) in the standard 48 hour exposure (i.e. misleading/ irrelevant positives); Ofloxacin, a fluoroquinolone bacterial gyrase inhibitor that produces negative results in *in vivo* and carcinogenicity studies, but positive results in other *in vitro* genotoxicity assays. It is a photo-mutagen that produces equivocal GSHC data when light is excluded (Hastwell *et al.*, 2009). Thus like the regulatory tests, a lowered top dose does not lead to a loss of sensitivity. Following the revision of the top testing dose for pharmaceuticals from 10 mM to 1 mM (ICH S2R) there is a strong case for classifying all the positive results produced above 1 mM as misleading or irrelevant as the reduced has no effect on sensitivity, but increases specificity. For this reason the data analysis in this relatively low molecular weight domain includes an assessment of genotoxicity at both 10 mM and 1 mM.

5. Results.

a. Prevalence of published genotoxicity and carcinogenicity data from F&F compounds selected for testing with BSHC.

In this library of F&F materials, a significant proportion of the materials have higher tier *in vivo* data (genotoxicity or carcinogenicity) that were intentionally chosen in order to challenge BSHC sensitivity and specificity. As *in vivo* tests are generally performed as a follow up to *in vitro* positive test results, this relatively small library (70 F&F materials) can be considered as typifying F&F materials which face issues during *in vitro* regulatory genotoxicity tests. Amongst the 70 substances tested in BSHC, there were Ames data from 61 (87 %), *in vitro* chromosome aberration (CA) data from 38 (54 %), mouse lymphoma (MLA) or HPRT assay data from 31 (44 %), *in vitro* micronucleus test (MN) from 15 (21 %), *in vivo* genotoxicity data from 36 (51 %) and rodent carcinogenicity data from 21 substances (30 %): see Table 2. This reveals the scale of “data gaps” in this applicability domain, and means that the reliability of the analyses that follows is greatest for the test substances with most data, and least for those with sparse data. The proportion of positive results produced by a genotoxicity test is a function of the prevalence of genotoxic carcinogens in the collection, and the sensitivity (generation of true positives) and specificity (generation of true negatives) of the tests applied.

b. Positive and negative predictivity of the BlueScreen assay for regulatory *in vitro* test battery results.

Of the 70 materials, 50 (71%) were BSHC negative and 20 (29 %) were BSHC positive using a cut-off of ≤ 10 mM . If a 1 mM cut-off is applied, 61 are BSHC negative (87) and 9 (13 %) are BSHC positive. The

assessment of the utility for a new screening assay is often more reliant on the accuracy of predictions of regulator-required *in vitro* and *in vivo* genotoxicity assays than of the results of animal carcinogenicity studies. It is those regulatory data on which safety assessment will rely and therefore the prediction of their outcomes that will guide the ranking of substances, and the establishment of their priority for development and follow up testing. In order to assess the utility of the BSHC as a screening test, the probability that substances producing positive (or negative) BSHC data would produce positive (or negative) data for regulatory genotoxicity assays was calculated (summarized in Table 3).

Positive predictivity: Of the 20 materials producing positive BSHC results ≤ 10 mM, all had Ames and *in vitro* genotoxicity data except 4-Hydroxy-2,5-dimethyl-3(2H)furanone (Ames only).

Ames test: 8 of 20 (40 %) materials positive in BSHC at 10 mM were positive in Ames. If a 1 mM cut-off is applied, 5 of 9 (56 %) BSHC positives are Ames positive.

Regulatory *in vitro* mammalian tests: 17 of the 19 (89 %) materials positive in BSHC were positive in one or more *in vitro* mammalian tests. If a 1 mM cut-off is applied, 9 of 9 (100 %) BSHC positives are *in vitro* positive. These high values were anticipated from the generally high prevalence of positive results from the *in vitro* mammalian tests in other chemical domains. The two materials positive in BSHC ≤ 10 mM, negative at ≤ 1 mM and negative *in vitro* are methyl eugenol and Oxacyclohexadec-12(+13)-en-2-one.

Negative predictivity: Of the 50 materials producing a negative BSHC result, 41 (82 %) had *in vitro* genotoxicity data. All had Ames data and 29 had data from at least one regulatory *in vitro* mammalian test.

Ames: BSHC negative results are highly predictive of Ames negative results: 39 of 41 (95 %) BSHC negative substances were negative in Ames ≤ 10 mM. The two materials with positive Ames data were pent-1-en-3-one and trans-2-hexenal.

If a 1 mM cut-off is applied, an additional 11 BSHC positives with Ames data are scored negative, of which 8 are Ames negative. This produces a negative prediction figure of 90 % (47/52). An Ames positive result is probably the least easy to manage, so the high prediction from a BSHC negative would be very a reliable indication that a substance could be carried forward for development. The 3 additional BSHC negative with positive Ames data using a 1 mM threshold are: 4-Hydroxy-2,5-dimethyl-3(2H)furanone, glyoxal and p-mentha-1,8-dien-7-al.

Regulatory *in vitro* mammalian tests (MLA/HPRT, MN, CA): 21 of 29 (72 %) of the BSHC negative substances were negative in one or more of *in vitro* mammalian cell tests ≤ 10 mM. A higher proportion of *in vitro* mammalian test positives (c.f. Ames positives) amongst the BSHC negatives was expected because of their lower specificity. The 8 BSHC negative materials that were positive in the mammalian tests were trans-2-hexenal, ethyl acrylate, benzaldehyde, menthol, benzyl alcohol, isobutyraldehyde, geranyl nitrile (GN) and styrene. All of these except GN and styrene are either not genotoxic, non-genotoxic carcinogens or non-carcinogens.

If a 1 mM cut-off is applied, an additional 10 BSHC substances with *in vitro* mammalian test data are scored negative, of which 3 are *in vitro* mammalian test negative. This produces a negative prediction figure of 23 /39 (59 %). The lower negative BSHC prediction values for the *in vitro* mammalian tests (c.f.

Ames) is consistent with the much higher specificity of BSHC c.f. regulatory *in vitro* tests, previously reported in the validation of BSHC and the GSHC sister assay (Hughes *et al.*, 2012).

Summary of *in vitro* predictivity data.

These results from the F&F materials show that BSHC has a very high negative predictivity for Ames test results (95 %) and a high positive predictivity for *in vitro* mammalian tests (89 %). The negative predictivity of BSHC for the *in vitro* mammalian tests is relatively low (72 %), as expected from the high prevalence of misleading positive results reported in broader historical reviews of data from those assays. The positive predictivity for the Ames test is relatively low, reflecting the mechanistic limitation of the Ames test to mutagens.

Animal genotoxicity and carcinogenicity data

It is more appropriate to calculate sensitivity and specificity values for the prediction of the outcome of animal studies, but for completeness, positive and negative prediction values are also calculated for *in vivo* genotoxicity data (Table 3).

c. Positive and negative predictivity of the BlueScreen assay for regulatory *in vivo* genotoxicity tests

Of the 70 materials, 36 had *in vivo* genotoxicity data of which 19 were negative in BSHC and 17 were positive in BSHC. Some of these substances have both positive and negative *in vivo* genotoxicity data from historical studies (see discussion).

Positive predictivity: Of the 17 materials positive in BSHC (≤ 10 mM) having *in vivo* data, 4 were *in vivo* positive (quinoline, methyl eugenol, 4-Hydroxy-2,5-dimethyl-3(2H)furanone, glyoxal), 1 was *in vivo* equivocal (eugenol) and 12 were negative *in vivo*. The positive predictivity of BSHC excluding eugenol is 25% (4/16: or 31% counting Eugenol). The results shown in Table 1 present the current state of the studies on these materials at the time of submission. It is anticipated that these may change as more genotoxicity study data become available.

If a 1 mM cut-off is applied, there are 7 BSHC positives with *in vivo* data, of which one is *in vivo* positive (14 %).

Negative predictivity: Of the 19 materials negative (≤ 10 mM) in BSHC having *in vivo* data, 15 (79 %) were negative *in vivo*, 3 have both positive and negative *in vivo* study data (styrene, benzyl alcohol and isobuteraldehyde) and 1 is positive (see discussion). Thus the *in vivo* negative predictivity of BSHC excluding equivocal data is 15/16 (94 %). If equivocal results are counted as negative this becomes 95% (18/19), if equivocal results are counted as positive it becomes 79 % (15/19).

If a 1 mM cut-off is applied, there are 29 materials negative in BSHC with *in vivo* data, of which 21 (72 %) are *in vivo* negative and 2 are equivocal. The negative predictivity including equivocal as negative is 79 % (23/29).

d. Sensitivity and Specificity of BSHC results amongst 36 substances with data from *in vivo* genotoxicity studies.

Sensitivity: 7 materials had positive *in vivo* test data and 2, eugenol and styrene, had equivocal *in vivo* test data. 57 % (4/7) of these were positive in BSHC. Counting equivocal results as positive the sensitivity decreases slightly to 56 % (5/9). With the exception of geranyl nitrile, all these substances with *in vivo* genotoxicity data also have carcinogenicity data and they are discussed in more detail in section e below.

If a 1 mM cut-off for BSHC data is applied, 1 of 7 (14 %) of the *in vivo* positives is BSHC positive.

Specificity: 27 materials produced only negative results in *in vivo* tests, of which 15 (56 %) were BSHC negative. Counting the 2 equivocal results (eugenol and styrene) as negative, increases specificity to 59 % (17/29).

If a 1 mM cut-off for BSHC data is applied, there are 6 further substances classified as negative, increasing specificity to 21/27 (78 %). Counting the 2 *in vivo* equivocal results as negative increases the specificity to 23/29 (79 %).

e. Sensitivity and Specificity of BSHC results amongst substances with data from carcinogenicity studies.

Specificity and sensitivity figures are the gold standard measures of *in vitro* test performance. There were 21 materials with carcinogenicity data.

Sensitivity: 3 substances were genotoxic carcinogens (quinoline, methyl eugenol, beta asarone) and all were BSHC positive. The sample number is low, but confirms the high sensitivity of BSHC in other applicability domains (Hastwell et al, 2006). If a 1 mM cut-off is applied, methyl eugenol produces a negative result.

Specificity: 10 materials were non-genotoxic carcinogens of which 7 (70 %) were BSHC negative (d-limonene, tert-butyl alcohol, pyridine, benzophenone, ethyl acrylate, styrene, benzaldehyde) and 3 were BSHC positive (furfural, isophorone, allyl isothiocyanate). A further 8 were non-carcinogens of which 4 were BSHC negative (cyclohexanone, menthol, benzyl alcohol, isobuteraldehyde) and 4 were BSHC positive (eugenol, 4-Hydroxy-2,5-dimethyl-3(2H)furanone, glyoxal, resorcinol). The specificity of BSHC for non-carcinogens and non-genotoxic carcinogens is therefore 61 % (11/18). Furanone derivatives like ascorbic acid, are able to reduce metals ions and produce superoxide anions to generate hydroxy radicals that cleave DNA (EFSA, 2012a). This provides a plausible explanation for positive results in *in vitro* tests and some higher tier *in vivo* tests (MN). 4-Hydroxy-2,5-dimethyl-3(2H)furanone also tested negative in a 2y carcinogenicity study and in a dominant lethal test in mouse. Of the 7 non-carcinogens and non-genotoxic carcinogens which produced positive BSHC results, 6 also produced positive results in the *in vitro* mammalian mutagenicity tests (MLA/HPRT) and in the *in vitro* clastogenicity or aneugenicity tests (CA and/or MN). No *in vitro* mammalian data were found for the 7th (4-Hydroxy-2,5-dimethyl-3(2H)furanone): it was Ames positive.

All 21 of the substances with carcinogenicity data have relatively low molecular weights (average 112). Seven of the 10 substances producing positive BSHC data in this group are only positive at concentrations exceeding 1 mM (eugenol, methyl eugenol, furfural, isophorone, 4-Hydroxy-2,5-dimethyl-3(2H)furanone, glyoxal, resorcinol). Two of the 3 genotoxic carcinogens are positive at 1 mM. The exception is methyl eugenol (BSHC LEC 5 mM). Seventeen (17) of the 18 (94 %) non-carcinogens or non-genotoxic carcinogens are negative below 1 mM in BSHC. The exception is AIT, discussed below. This

suggests that in this applicability domain, with a preponderance of small molecules, the application of a 1 mM threshold maintains the high specificity of BSHC (~ 95 %) observed in wider validation studies. The impact of this lowered testing threshold on other published BSHC and GSHC data was considered earlier in this paper.

6. Discussion.

The new data collected here, together with information drawn from various other cited sources, allows several conclusions to be drawn from this study. Amongst these flavour and fragrance compounds (a) BSHC results show good predictivity of positive and negative results from the standard regulator-required *in vitro* genotoxicity assays (b) BSHC negative data (≤ 1 mM) are highly predictive of negative *in vivo* genotoxicity data (c) BSHC negative data (≤ 1 mM) are highly specific in the identification of non-carcinogens and non-genotoxic carcinogens (d) a 1 mM testing limit maintains the sensitivity of the assay, and increases specificity.

Negative predictivity: Ames: There were two materials positive in the Ames and negative in BSHC (with and without S9): pent-1-en-3-one and trans-2-hexenal. The first of these pent-1-en-3-one is uniquely positive in the Ames TA 100 strain, but negative in the mammalian cell (HPRT) gene mutation assay and in the *in vitro* micronucleus test. Oxidative, free radical or weak electrophilic mechanisms are likely to be responsible for positive test results for pent-1-en-3-one (EFSA, 2012b). Mammalian cell systems may be better equipped to deal with oxidative stress, though culture conditions can contribute to the generation of positive results in this class (Halliwell 2003, 2012). Substances such as these support a recent ECVAM workshop conclusion that substances producing a single Ames positive result, and at least two negative *in vitro* mammalian genotoxicity test results, covering two different endpoints, can be considered a negative signature for genotoxic carcinogenicity (ECVAM, 2013). An *in vivo* COMET assay should confirm the harmlessness of this material (pent-1-en-3-one) and the unreliability of the Ames test for such materials. Trans-2-hexenal was the second material positive in Ames (TA100) and negative in BSHC. It is only marginally positive in CA in Namalva cells that are deficient in the antioxidant GSH and lack glutathione-S-transferase activity (Eisenbrand 1995). It is noteworthy that trans-2-hexenal produced negative *in vivo* test results (UDS, MN) and as an α,β -unsaturated aldehyde it is expected to be readily metabolized *in vivo*: incubation with S9 fractions of rat liver/small intestine with NAD⁺ produce 2-hexenoic acid, which does not readily form adducts and is efficiently conjugated with glucuronic acid and excreted in urine (EFSA, 2014c). However, whilst a Muta™Mouse study (negative) has supported the overall negative conclusion for this compound, the most recent EFSA opinion has not ruled out genotoxicity concerns, as clastogenicity has not yet been assessed at point of first contact (liver). This case highlights the apparently limited ability of some regulatory *in vitro* genotoxicity tests to mimic the fate of chemicals under real exposure conditions and to indicate their impact on the genetic material of organisms.

Negative predictivity: *in vitro* mammalian tests: there were eight materials negative in BSHC that were positive in at least one mammalian *in vitro* assay. This most likely reflects the documented low specificity of regulatory genotoxicity assays, particularly in historical data sets (Kirkland *et al.*, 2005). For example five of the *in vitro* positives producing BSHC negative results that have carcinogenicity studies are non-DNA-reactive chemicals. Four of the 5 (Menthol, benzyl alcohol, ethyl acrylate, isobuteraldehyde) are non-carcinogens or non-genotoxic carcinogens (ECVAM group 3 chemicals: Kirkland *et al.*, 2008), and the 5th, benzaldehyde, is a mouse specific non genotoxic carcinogen (JECFA 2002a).

The last three materials negative in BSHC and positive in *in vitro* mammalian tests are trans-2-hexenal (discussed in the Ames section above), styrene and geranyl nitrile. Styrene has been classified as a carcinogen with a probable genotoxic mode of action (Kirkland & Fowler 2010). However it is not carcinogenic in rats (by the oral or inhalation routes) and in mice, it is carcinogenic uniquely by inhalation (IARC 2012). According to the IARC, the existing data suggest at least two plausible modes of action of styrene that are not mutually exclusive. One involves a DNA-reactive mode of action initiated by the conversion of styrene to styrene 7,8-oxide and the subsequent induction of DNA damage in target tissues. The other involves cytotoxic effects in lungs of mice exposed to styrene. For either mode of action, interspecies differences in the metabolism and toxicokinetics of styrene and styrene 7,8-oxide that exist between rats and mice are likely to play a key role (IARC 2012). *In vivo*, genotoxicity tests results depend on the species and route of administration which supports that metabolism to styrene 7,8-oxide is required to get a positive genotoxicity test results. These differences are likely to explain the negative result observed in BSHC which is consistent with the negative carcinogenicity tests in rats and the negative results obtained in carcinogenicity studies in mice by gavage.

Geranyl nitrile (GN) is an α,β -unsaturated nitrile that was used as a fragrance material before to being banned due to a positive *in vitro* CA test (with metabolic activation) that was confirmed by a positive *in vivo* MN test. A very close analogues of GN such as citronellyl nitrile (3,7-dimethyl-6-octenenitrile; the α,β -saturated form of GN) and PC-06, (an α,β -unsaturated nitrile sharing the same geranyl structural core as GN) are negative in the *in vivo* micronucleus tests. However, like GN, citronellyl nitrile (CN) was a positive in the *in vitro* CA test. Metabolism studies performed on GN and CN in primary hepatocytes from mice, rats and humans did not provide an explanation for the difference in genotoxicity study results from GN and CN (Kemper *et al.*, 2006). The Mode of Action and the genotoxicity of GN still remain unclear and this uncertainty has been strengthened by additional negative *in vivo* studies performed on GN (both Comet assay and liver adduct formation were negative, unpublished research study by Firmenich). These more recent studies were carried out using analytical grade materials, produced using a different synthesis route to that used historically. It is possible or perhaps likely therefore that the genotoxicity previously reported was due to the presence of a genotoxic impurity in the past commercial batches.

By taking into account the carcinogenicity data and considering the discussion of trans-2-hexenal, GN and styrene above, the negative predictivity of BSHC for reliable *in vitro* mammalian assay data rises to between 97 % (if trans-2-hexenal would be concluded as clastogenic on the first contact site) and 100 %.

The high negative predictivity and specificity of BSHC data seen in this study indicate its usefulness as a screening assay for this sector. Substances producing negative BSHC data appear unlikely to produce consistently positive results in either the *in vitro* regulatory testing battery, or in higher tier (*in vivo*) follow-up genotoxicity testing, where there are conflicting Ames and *in vitro* mammalian regulatory data.

Positive predictivity: The materials that were flagged positive in BSHC and which are negative in the standard *in vitro* battery were either clearly genotoxic *in vivo*, or materials for which reported negative *in vitro* clastogenicity studies were found questionable following a deeper evaluation of the test study reports.

Methyl eugenol is a genotoxic carcinogen (NTP 2000); its genotoxicity is due to the *in vivo* metabolic 1' hydroxylation and sulfonation of the allylic group which forms DNA adducts *in vivo* (Randerath 1984). Its

genotoxicity is not detected by the standard *in vitro* genotoxicity test battery or by the *in vivo* micronucleus test (NTP 2000) but it is detected by BSHC, but only above the proposed testing limit, with S9 (i.e. $c = 5\text{mM}$).

Oxacyclohexadec-12(+13)-en-2-one is a macrocyclic lactone, that was positive in BSHC only in presence of S9 and only at the highest tolerated dose (2.56 mM). It was reported as negative in the Ames, CA and MLA in tests performed in 1995. A deeper evaluation of the CA study reports suggests that an equivocal conclusion would have been more appropriate as both positive and negative data were generated in presence of metabolic activation (Proprietary data). Furthermore, a very close analogue of oxacyclohexadec-12(+13)-en-2-one, Oxacyclopentadec-10-en-2-one, 13-methyl- (CAS 329925-33-9), which is a methylated C14 lactone (c.f. C15 lactone for oxacyclohexadec-12(+13)-en-2-one) was clearly positive in CA +S9: a negative *in vivo* MN study was necessary to discredit the positive *in vitro* test.

Thus depending on how the data are reviewed and compared, these results for BSHC (positive LEC 2.56 mM) could be viewed as 'false' positive BSHC results (*vs in vitro* tests). Perhaps better it should be considered as an irrelevant positive, adding to the case for considering a top testing dose of 1 mM, whose relevance on further testing is perhaps of limited or no concern. As with all screening and early tier assays the results need to be viewed in context: this study supports the view that BSHC is both a good predictive test for materials that are equivocal /positive *in vitro* and a sensitive assay for *in vivo* genotoxicants that are hard to detect in the *in vitro* regulatory test battery.

By taking into account the latter facts on Oxacyclohexadec-12(+13)-en-2-one, it would make sense to consider a positive predictivity of BSHC for *in vitro* genotoxicity of 95 % (18/19).

Positive predictivity is an important parameter for an assay that is intended to screen and prioritize new materials in development. This is particularly so in an industry where *in vivo* tests are not mandatory and even discouraged. Furthermore the high specificity of BSHC, discussed further below, provides confidence in the likely success of *in vivo* genotoxicity studies used to follow up positive *in vitro* regulatory tests. However, the poor specificity of *in vitro* mammalian regulatory tests coupled to the ban of *in vivo* tests by the EU Cosmetic Directive (EC 1223/2009) raises an issue for the new materials that are candidates for use in cosmetics. Currently there are no validated alternative non-animal test methods that allow follow up of positive results from standard *in vitro* assays [SCCP/1212/09]. Even if evidence of genotoxicity from positive *in vitro* test results can be countered using more recently developed *in vitro* tests (Kirkland and Fowler 2010, Fowler *et al.* 2012, Fowler *et al.* 2014), a number of the materials tested in this library and widely used in cosmetics (including some very commonly used materials such as for instance benzaldehyde or menthol) wouldn't be approved for cosmetic use if they had to face the current *in vitro* regulatory tests. Both substances produced negative BSHC data. This supports a recent and more significant retrospective analysis of the mutagenicity/genotoxicity data of the cosmetic ingredients present on the Annexes of the European cosmetic legislation made by Ates *et al.* (2014). The authors show that on a library of 249 cosmetics ingredients (including 27 fragrance ingredients) present on the Annexes III-VI of the current EU Cosmetic Directive (1223/2009), 125 materials (50 %) faced *in vitro* positive results. Of these, 85 % (106/125) were shown to be false positives by subsequent *in vivo* tests. These 106 substances currently used in cosmetics would have been lost for the market without *in vivo* follow-up tests.

Sensitivity and Specificity of BSHC: BSHC seems *prima facie* having a modest sensitivity with regards to *in vivo* positive results (57 %, 4/7 *in vivo* positive are BSHC positive) but 2 of the 3 negative materials are non-carcinogens (benzyl alcohol, benzaldehyde), and the third, geranyl nitrile, was discussed earlier. In this light,

we can conclude from a limited dataset, and with the consideration that *in vivo* genotoxins are not supported as F&F materials unless found in extracts of natural materials widely consumed (e.g. methyl eugenol in nutmeg), that all the substances known to be genotoxic carcinogens in this study, were correctly predicted by BSHC. This confirms the high sensitivity of BSHC and GSHC in earlier published studies.

The moderate specificity of BSHC for *in vivo* tests (56 %) and for non-carcinogens and non-genotoxic carcinogens (61 %) is significantly increased by considering a top testing dose of 1 mM (78 % and 94 %, respectively). The non-carcinogen allyl isothiocyanate remains positive in BSHC < 1mM. The EFSA Panel considered it reasonable to assume that reactive oxygen species might be involved in the genotoxic effects of AIT observed *in vitro* and concluded that there is no concern with respect to genotoxicity since no evidence on genotoxicity was observed in studies performed on different genetic endpoints *in vivo* (EFSA, 2010).

7. Conclusions:

A consolidated library of 70 materials including 21 F&F materials with carcinogenicity data and 36 materials with *in vivo* genotoxicity data was used to challenge BSHC in the new applicability domain of F&F materials. Materials in this library had also been tested with *in vitro* regulatory tests. Of these, 61 materials had Ames data and 48 materials had standard *in vitro* genotoxicity with data from one or more *in vitro* mammalian genotoxicity tests. This is the largest single genotoxicity study reported for F&F substances.

Unlike the pharma industry where *in vivo* tests are mandatory, the genotoxicity safety assessment of F&F materials relies primarily on an *in vitro* genotoxicity test battery. For this reason, the predictivity of the BSHC assay for the *in vitro* regulatory genotoxicity tests (Ames, CA, MLA/HPRT, MN) was evaluated first. BSHC was found to have a high positive predictivity for the *in vitro* test battery (95 %) and a high negative predictivity for materials that gave negative results in the Ames test (95 %). We believe that the lower negative predictivity for the *in vitro* mammalian test battery (72 %) reflects the previously reported high prevalence of misleading positive results from these tests for non-carcinogens and/or non-genotoxic carcinogens – substances that are more likely to produce negative BSHC data. The negative predictions of BSHC were correct for 94 % of the materials that were either non-carcinogens or carcinogens that have no genotoxic mode of action. Specificity to genotoxic carcinogens with higher molecular weight using the 1 mM cut-off was similar to that reported at 10 mM. This confirms the much higher specificity of BSHC vs the standard regulatory *in vitro* genotoxicity tests. This was already shown in other studies used to validate the BSHC and the GSHC sister assay. Specificity is a key criterion in a screening assay as it allows more confident challenge to conflicting *in vitro* test results produced in higher tier tests. In this study, the negative predictivity of BSHC for *in vivo* negative results and non-carcinogen/non-genotoxic carcinogens is 94 % and 100 % respectively although we caveat further analysis on a larger dataset. As a positive result in an *in vivo* genotoxicity test will severely threaten the commercial viability of a new substance, a negative BSHC provides a reliable indication that a substance could be carried forward for development and would also offer stronger indications that a substance that is negative in the BSHC screening assay will not face genotoxic issues in higher tier *in vivo* tests.

There were too few genotoxic carcinogens in this study to provide a robust figure for sensitivity to genotoxic carcinogens, but 3 of 3 (or 2 of 3 at 1 mM) were correctly identified by BSHC. Whilst this is a relatively small number, the broader data set for the assay from other studies includes an additional 4 volatile genotoxic carcinogens, of which 3 (75%) produced positive results: Methyl methane sulphonate, mw 110, LEC 225 uM; Dimethylnitrosamine, mw 74, LEC 625uM, Safrole, mw 162, LEC 1.25 mM). The exception, acrylamide (mw

71), requires cytochrome P450 2E1 metabolism to produce the genotoxin glycidamide *in vivo*, and produces only negative genotoxicity results in the Ames test and mammalian cell tests (Koyama et al., 2011). In this small sample, the sensitivity of the assay to volatile genotoxic carcinogens is very similar to that found in published validation studies (80%: Hastwell *et al.*, 2006, 2009). The use of a lower top testing dose increased the specificity of the assay (*in vivo* negative: 78 %, non-genotoxic carcinogen and non-carcinogens: 94 %) in this applicability domain, whilst only marginally effecting sensitivity.

In conclusion this study demonstrated that the BSHC assay appears to be a viable candidate for the *in vitro* genotoxicity screening of F&F materials and a useful tool to help avoid follow-up animal testing in line with the recently published EURL ECVAM Strategy to Avoid and Reduce Animal Use in Genotoxicity Testing (Corvi *et al.*, 2013).

Conflict of interest

RMW is founder and CSO of Gentronix Ltd which developed and sells the GADD45aGLuc, BlueScreen HC assay, which is protect by European Patent Application No 05744681.7.

8. References:

- Ates G, Doktorova TY, Pauwels M, Rogiers V (2014). Retrospective analysis of the mutagenicity/genotoxicity data of the cosmetic ingredients present on the Annexes of the Cosmetic EU legislation (2000-12). *Mutagenesis*. 29 115-121.
- Bakhtyari, N.G., Raitano, G., , Benfenati, E., Martin, T., Young, D. (2013): Comparison of in silico models for prediction of mutagenicity. *J. Env.. Sci. Health, Part C: Env. Carc. Ecotox. Rev.*, 31 45-66
- Billinton N, Hastwell PW, Beerens D, Birrell L, Ellis P, Maskell S, Webster TW, Windebankd S, Woestenborghs F, Lynch AM, Scott AD, Tweats DJ, Gompel JV, Rees RW, Walmsley RM. (2008). Inter-laboratory assessment of the GreenScreen HC GADD45a-GFP genotoxicity screening assay: an enabling study for independent validation as an alternative method. *Mutat. Res.* 653 22-33
- Billinton N, Bruce S, Hansen JR, Hastwell PW, Jagger C, McComb C, Klug ML, Pant K, Rabinowitz A, Rees R, Tate M, Vinggaard AM, Walmsley RM. (2010). A pre-validation transferability study of the GreenScreen HC GADD45a-GFP assay with a metabolic activation system (S9). *Mutat. Res.* 700, 44-50
- Birrell L, Cahill P, Hughes C, Tate M, Walmsley RM. (2010). GADD45a-GFP GreenScreen HC assay results for the ECVAM recommended lists of genotoxic and non-genotoxic chemicals for assessment of new genotoxicity tests. *Mutat. Res.* 695, 87-95
- Corvi, R., Madia, F., Worth, A., Whelan M. (2013) EURL ECVAM Strategy to avoid and reduce animal use in genotoxicity testing. doi: 10.2788/43865
- EC 1907/2006. Registration, evaluation, authorisation and restriction of chemicals (REACH)
- EC 1223/2009. Cosmetics products directive, from 2013
- ECVAM (2013) EURL ECVAM progress report on the development, validation and regulatory acceptance of alternative methods (2010-2013). doi:10.2788/90736
- 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 40-46
- EFSA (2010). Scientific opinion on the safety of allyl isothiocyanate for the proposed uses as a food additive. *EFSA J.* 8(12):1943.
- EFSA (2012a). Scientific opinion on flavouring group evaluation 99. *EFSA J.* 10 (10): 2901
- EFSA (2012b) Scientific opinion on flavouring group evaluation 205. *EFSA J* 10(10):2902
- EFSA (2013) Scientific opinion on flavouring group evaluation 224 (FGE.224):Consideration of genotoxic potential for two α,β -unsaturated thiophenes from subgroup 5.2 of FGE.19 by EFSA. *EFSA J* 11(2):3093 (5-Methyl Thiophencarbaldehyde)
- EFSA (2014a) Scientific opinion on flavouring group evaluation 212, Revision 2 (FGE.212Rev2): α,β -Unsaturated alicyclic ketones and precursors from chemical subgroup 2.6 of FGE.19. *EFSA Journal* 2014;12(2):3584 (Isophorone)

- EFSA (2014b) Scientific opinion on flavouring group evaluation 215 (FGE.215): Seven α,β -Unsaturated Cinnamyl Ketones from subgroup 3.2 of FGE.19. EFSA Journal 2014;12(4):3623 (4-phenylbut-3-en-2-one)
- EFSA (2014c) Scientific opinion on flavouring group evaluation 200. EFSA J 12(6):3709 (trans-2-hexenal and derivatives)
- EFSA (2014d): Scientific opinion on flavouring group evaluation 213, Revision 1 (FGE.213Rev1): Consideration of genotoxic potential for α,β -Unsaturated Alicyclic ketones and precursors from chemical subgroup 2.7 of FGE.19 EFSA Journal 2014; 12(5):3661 (Maltol, damascones)
- Fowler, P, Smith, K, Young, J, Jeffrey, L, Kirkland, D, Pfuhler, S, Carmichael, P (2012) Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. I. Choice of cell type. *Mutat. Res.* 742, 11–25
- Fowler, P, Smith, R, Smith, K, Younge, J, Jeffrey, L, Carmichael, P, Kirkland, D, Pfuhler, S. (2014) Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. III: Sensitivity of human cell types to known genotoxic agents. *Mutat. Res.* 767 28–36
- Halliwell, B (2003) Oxidative stress in cell culture: an under-appreciated problem? *FEBS Lett.* 540 3-6
- Halliwell, B (2012). The antioxidant paradox: less paradoxical now? *Br. J. Clin. Pharma.* 75 637-644.
- Hastwell PW, Chai LL, Roberts KJ, Webster TW, Harvey JS, Rees RW, Walmsley RM. (2006). High-specificity and high-sensitivity genotoxicity assessment in a human cell line: Validation of the GreenScreen HC GADD45a-GFP genotoxicity assay. *Mutat. Res.* 607, 160-175
- Hastwell PW, Webster TW, Tate M, Billinton N, Lynch AM, Harvey JS, Rees RW, Walmsley RM. (2009). Analysis of 75 marketed pharmaceuticals using the GADD45a-GFP 'GreenScreen HC' genotoxicity assay. *Mutagen.* 24, 455-463
- Hughes, C, Rabinowitz, Adam, Tate, M, Birrell, L, Alsop, J, Billinton, N, Walmsley, R.M. (2012). Development of a high-throughput Gaussia luciferase reporter assay for the activation of the GADD45a gene by mutagens, promutagens, clastogens, and aneugens. *J. Biomol. Scr.* 17, 1302-1316.
- IARC, 2012. Some traditional herbal medicines, some mycotoxins, naphthalene and styrene. IARC monographs on the evaluation of carcinogenic risks to humans. 82 437-522
- Jagger C, Tate M, Cahill PA, Hughes C, Knight AW, Billinton N, Walmsley RM. (2009). Assessment of the genotoxicity of S9-generated metabolites using the GreenScreen HC GADD45a-GFP assay. *Mutagen.* 24 35-50
- JECFA, 2002a. Safety evaluation of certain food additives and contaminants. Fifty-seventh meeting of the Joint FAO/WHO Expert Committee on Food Additives, WHO Food Additives Series: 48. IPCS, WHO, Geneva.
- Johnson, D, Hastwell, P, Walmsley, R. (2013). The involvement of WT1 in the regulation of GADD45a in response to genotoxic stress. *Mutagen.* 28, 393-399
- Kemper R.A., Nabb D.L., Gannon S.A., Snow T.A., Api A.M. (2006). Comparative metabolism of geranyl nitrile and citronellyl nitrile in mouse, rat, and human hepatocytes. *Drug. Metab. Dispos.* 34 1019-29

- Kirkland D, Aardema M, Henderson L, Müller L. (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. *Mutat. Res.* 584 1–256.
- Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F. (2007). How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of an ECVAM Workshop. *Mutat. Res.* 628 31–5
- Kirkland, D, Kasper, P, Müller, L, Corvi, R, Speit, G. (2008). Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A follow-up to an ECVAM workshop. *Mutat. Res.* 653 99-108
- Kirkland, D, Fowler, P. (2010), Further analysis of Ames-negative rodent carcinogens that are only genotoxic in mammalian cells in vitro at concentrations exceeding 1 mM, including retesting of compounds of concern. *Mutagen.* 25 539-553.
- Kirkland, D, Reeve, L, Gatehouse, D, Vanparys (2011) A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. *Mutat. Res.* 721 27-73.
- Koyama, N, Yasui, M, Oda, Y, Suzuki, S, Satoh, T, Suzuki, T, Matsuda T, Masuda, S, Kinae, N, Honma, M (2011). Genotoxicity of acrylamide in vitro: acrylamide is not metabolically activated in standard in vitro systems. *Env. Mol. Mutagen.* 52 11-19.
- Nesslany, Simar-Meintières S, Talahari I, Nkili-Mboui EN, Marzin D (2007). In vivo Comet assay on isolated kidney cells to distinguish genotoxic carcinogens from epigenetic carcinogens or cytotoxic compounds. *Mutat. Res.* 630 28-41
- NTP, 2000. Toxicology and carcinogenesis studies of methyleugenol (Cas No. 93-15-2) In F344/N rats and B6c3f1 mice (Gavage Studies). NTP TR 491 NIH Publication No. 00-3950.
- Parry, JM, Parry, E, Phrakonkham, P Corvi, R (2010). Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. *Mutagenesis.* 25 531–538
- Randerath, K, Haglund, RE, Phillips DH (1984). 32P-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. *Carcinogen.* 5 1613-1622
- Topham, CH, Billinton, N, Walmsley RM (2012). Nongenotoxic apoptosis inducers do not produce misleading positive results in the TK6 cell-based GADD45a-GFP genotoxicity assay. *Tox. Sci.* 128 79-91.
- Turner SD, Tinwell H, Piegorsch W, Schmezer P, Ashby J (2001). The male rat carcinogens limonene and sodium saccharin are not mutagenic to male Big Blue rats. *Mutagen.* 16 329-32

Table 1: BlueScreen HC assay results for 71 Flavor & Fragrance materials.

Chemical	CAS n°	BlueScreen HC Genotoxicity						In vitro Genotoxicity						In vivo Genotoxicity						Carcinogenicity	Refs for animal studies
		Highest conc ⁿ Tested*		LEC		Influence of top dose, mM		Ames	MLA/hprt	CA	MN	GLOBAL w/o Ames	GLOBAL incl. Ames	UDS study	MN	CA	Comet Assay	Transgenic mice	Global in vivo		
		mM	mg/ ml	-S9 μM	+S9 μM	10	1														
D-limonene	5989-27-5	10	1.36	-	-	-	-	-	-	-		-	-				-	-	-	+ NG ECVAM Gr.2	a, b, c, d
Cyclohexanone	108-94-1	10	0.98	-	-	-	-	-	-	-		-	-			-			-	- ECVAM Gr.2	a, e
Tert-butyl alcohol	75-65-0	10	0.74	-	-	-	-	-	-	-		-	-		-				-	+ NG ECVAM Gr.2	a, f
Pyridine	110-86-1	10	0.8	-	-	-	-	-	-	-		-	-		-			-	-	+ NG ECVAM Gr.2	a, g, h
D,L-menthol	15356-70-4	10	1.82	-	-	-	-	-	E	+		+	+		-				-	- ECVAM Gr.3	a.
Benzyl alcohol	100-51-6	10	1.08	-	-	-	-	-	+	+		+	+		-		+		+	- ECVAM Gr.3	a, i, j
Eugenol	97-53-0	2.5	0.41	1250	2500	+	-	-, TC	+	+	+	+	+	-	E				E	- ECVAM Gr.3	a, j
Isobutyraldeyde	78-84-2	10	0.72	-	-	-	-	-	+	+		+	+		-	+			+	- ECVAM Gr.3	a.
Quinoline	91-22-5	10	1.29	-	313	+S9	+S9	+S9	+S9	+		+	+	E	+	+		+	+	+ G	k.
Methyl eugenol	93-15-2	5	0.89	-	5000	+S9	-	-		-		-	-	+	-				+	+ G	h.
Beta asarone	5273-86-9	2.5	0.52	1063	156	+	+	+S9		+S9	+	+	+							+ G	l.
Benzophenone	119-61-9	1.25	0.23	-	-	-	-	-	-	-	-	-	-		-			-	-	+ NG	h.
Furfural	98-01-1	10	0.96	5000	104	+	-	-	-S9	+		+	+	-		-		-	-	+ NG	m.
Ethyl acrylate	140-88-5	10	1	-	-	-	-	-	+	+	+	+	+		-	-			-	+ NG ECVAM Gr.3	a, k
Isophorone	78-59-1	10	1.38	104	-	-S9	-	-	-S9	+		+	+		-				-	+ NG	n,o
Styrene	100-42-5	2.5	0.26	-	-	-	-	-		+	+	+	+		E	E			E	+ NG, mouse by inhalation	p.
Allyl isothiocyanate (AIT)	57-06-7	10	0.99	78.1	131	+	+	+	+	+		+	+	-	-			-	-	+ NG	q,r.
Benzaldehyde	100-52-7	10	1.06	-	-	-	-	-	+	+		+	+		-				-	+ NG, mouse specific	k.
4-Hydroxy-2,5-dimethyl-3(2H)furanone	3658-77-3	10	1.28	1250	5000	+	-	+					+	+	+			-	+	-	s.
Glyoxal	107-22-2	10	0.58	-	1280	+S9	-	+	+	+		+	+	+	-				+	-	t.
Resorcinol	108-46-3	10	1.1	2500	8000	+	-	-	+	+	+	+	+		-				-	- ECVAM Gr.3	a.
Geranyl nitrile	5146-66-7	10	1.49	-	-	-	-	-	-	+S9		+	+		+	+	-		+		u.

Perillyl alcohol (p-Mentha-1,8-dien-7-ol)	536-59-4	2.5	0.38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Anti-carcinogenic	v.
Methyl dihydrojasmonate	24851-98-7	10	2.26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		z
4-Phenylbut-3-en-2-one	122-57-6	10	1.46	313	819	+	+	+S9	-S9	-	-	+	+	-	-	-	-	-	-	- NTP skin study	z
p-Mentha-1,8-dien-7-al	2111-75-3	10	1.5	1582	5000	+	-	+S9	-	+	-	+	+	-	-	-	-	-	-		z
trans-2-hexenal	6728-26-3	5	0.49	-	-	-	-	+	+	-	+	+	+	-	-	-	-	-	-		w.
5-Methylthiophene-2-carbaldehyde	13679-70-4	10	1.26	5625	625	+	+	-	-	-	+S9	+	+	-	-	-	-	-	-		x.
Maltol (3-Hydroxy-2-methyl-4h-pyran-4-one)	118-71-8	10	1.26	625	2500	+	+	+	-	+	+	+	+	-/+	-	-	-	-/+	-	k; y	
Octahydro-5-methoxy-4,7-methano-1(H)-indene-2-carboxaldehyde	86803-90-9	10	1.94	625	-	-S9	-S9	-	-	+S9	-	+	+	-	-	-	-	-	-		z
5,8-Methano-2H-1-benzopyran-2-one, 6-ethylideneoctahydro-	69486-14-2	5	0.96	1250	-	-S9	-	-	-	-S9	-	+	+	-	-	-	-	-	-		z
Delta damascone	71048-82-3	5	0.96	-	1250	+S9	-	-	-	+S9	-	+	+	-	-	-	-	-	-		z
7-methyloct-3-en-2-one	33046-81-0	10	1.4	156	1250	+	+	-	-	-	+	+	+	-	-	-	-	-	-		z
Pent-1-en-3-one	1629-58-9	10	0.84	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-		z
Dodec-2-enal	4826-62-4	5	0.91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		z
PC01 unsaturated alicyclic nitrile		5	0.85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		z
PC06 unsaturated nitrile		10	1.63	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		z
PC07 unsaturated nitrile		10	1.35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		z
Alpha damascone	24720-09-0	10	1.92	-	78.1	+S9	+S9	-	-	-	+	+	+	-	-	-	-	-	-		
Acetic acid, 2-(1-oxopropoxy)-, 1-(3,3-imethylcyclohexyl) ethyl ester	236391-76-7	10	2.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Geranyl palmitate	3681-73-0	2.5	0.98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Propyl (2S)-2-(1,1-dimethylpropoxy)-propanoate	319002-92-1	10	2.02	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
3-Methylcyclopentadecenone	82356-51-2	10	2.36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Carvacrol	499-75-2	10	1.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
2H-Pyran, tetrahydro-4-methyl-2-phenyl-	149713-23-5	10	1.76	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Oxacyclohexadec-12(+13)-en-2-one	111879-80-2	5	1.19	-	2560	+S9	-	-	-	-	-	-	-	-	-	-	-	-	-		

[illegible]

(1). Column headers: LEC, lowest effective concentration; -S9, result without metabolic activation; +S9: result with metabolic activation.
(2). Within Table: '-', negative result with and without S9; '+', positive result with and without S9; '+S9', only positive with S9; '-S9', only positive without S9;
G, Genotoxic; NG, not Genotoxic; E, equivocal; TC, Technically Compromised;

22

Table 2

Prevalence of study data and positive results in the test collection.

Test	N	Positive (%)
BlueScreen (10 mM)	70	20 (29 %)
BlueScreen (1 mM)	70	9 (13 %)
Ames	61	10 (16 %)
Chrom. Abs	38	19 (50 %)
MLA/HPRT	31	10 (32 %)
MN <i>vitro</i>	15	11 (73 %)
MN <i>vivo</i>	36	8 (22 %)

Table 3
Predictivity, Specificity and Sensitivity of BSHC for the library of 71 F&F chemicals

	10 mM threshold	10 mM threshold considering <i>in vivo</i> or carcinogenicity data	1mM threshold (ICH criteria)
Positive Predictivity			
<i>In vitro</i> mammalian test battery ^(a)	95 % ^(18/19) ^(a)	100 % ^(a)	100 % (9/9)
<i>In vivo</i> tests ^(a,b)	29 % (5/17) ^(b)	36 % (5/14) ^(b)	14 % (1/7) ^(b)
Negative Predictivity			
Ames ^(c)	95 % (39/41)	≥ 98 % ^(c,d)	90 % (47/52)
<i>In vitro</i> mammalian test battery ^(c)	72 % (21/29)	90 % (26/29) to 100 % (29/29) ^(c)	59 % (23/39)
<i>In vivo</i> tests ^(b,c)	79 % (15/19) to 95 % (18/19) ^(b)	79-94%	72 % (21/29) to 79 % (23/29) ^(b)
Sensitivity			
<i>In vivo</i> tests	56 % (5/9)	63 % (5/8)	14 % ^(d) (1/7)
Genotoxic Carcinogenicity	100 % ^(e) (3/3)	N.A.	67 % (2/3)
Specificity			
<i>In vivo</i> tests	56 % (15/27)	78 % (21/27) ^(f)	78 % (21/27)
Carcinogens & non genotox carcinogens	61 % (11/18)	N.A.	94 % (17/18)
(a) See Discussion, paragraph on positive predictivity (b) The <i>in vivo</i> results present the current state of the studies on these materials at the time of submission but these may change as some material were still under investigation for genotoxicity. (c) See Results and Discussion, paragraphs on negative predictivity (d) 4 of the 7 materials positive <i>in vivo</i> are in fact non-carcinogens (see discussion) (e) Styrene is an animal- and exposure-specific carcinogen, the negative result observed in BSHC which is consistent with the negative carcinogenicity tests in rats and the negative results obtained in carcinogenicity in mice by gavage (see discussion) (f) 6 <i>in vivo</i> positives are non genotoxic carcinogens			

Highlights for review

- i. The genotoxicity of 70 flavour and fragrance substances was assessed using BlueScreen HC.
- ii. Positive results are highly predictive of positive regulator-required *in vitro* assays
- iii. Negative results were highly predictive of negative *in vivo* results.
- iv. A top 1 mM or 0.5 mg/ml dose maintains sensitivity and increases specificity.
- v. The assay can fill data gaps for materials with no/limited regulatory test data.