Development of BlueScreen and GreenScreen Genotoxicity Assays and Applicability to Cosmetic Ingredients

What are these assays? Why or when are they used?

Supporting data from published validation studies:
154 compounds including “ECVAM lists”, and marketed pharmaceuticals,
70 flavour and fragrance materials:
(largest single genotoxicity F&F study reported)

Professor Richard Walmsley
University of Manchester & Gentronix Ltd
The sector wake-up call

Retrospective analysis of the mutagenicity/genotoxicity data of cosmetic ingredients, Annexes of European cosmetic legislation Ates et al. (2014).

Considered 249 cosmetics ingredients, including 27 fragrance ingredients

125 materials (50 %) produced positive in vitro genetox results. But, 106 of these 125 positives (85%!) were negative in vivo...

That is 106 substances, currently used in cosmetics, which would have been lost from the market without animal tests...

...and couldn’t have been developed in Europe and some others!

It would be useful to have an assay that more accurately predicts in vivo hazard.
Outline of talk

• Brief introduction to genetic toxicology – 4 slides
• Genotoxicity assessment in the F&F sector – 2 slides
• The GADD45a assays – 6 slides
• Data from 70 compounds supplied by Firmenich – 10 slides
• Conclusions – 2 slides
First, the briefest intro to genetic tox!

1771: John Hill noted nasal cancer amongst snuff users
1776: Sir Percival Potts noted scrotal cancer amongst sweeps
1915: Yamagiwa & Ichikawa coal tar causes rabbit skin carcinoma
1932: Sir Ernest Kennaway coal tar B[a]P is the carcinogen!

The field of chemical carcinogenesis was born

1973: Bruce Ames B[a]P is Ames positive

The field of genetic toxicology was born
20/21st Century: Regulatory genotoxicity tests

• The **Ames** bacterial **reverse mutation** assay
  
  More colonies = more mutation
  *compound hungry, not generally for screening*
  *(also microplate methods)*

• **Mammalian cell** chromosome aberration assays
  mis-segregation (**aneugens**),
  breakage (**clastogens**)
  **aneugens and clastogens** produce **micronuclei**
  some automated flow methods available (LitronLabs)

• **Mouse Lymphoma Assay** forward cell mutation (MLA).
  *why another mutation assay? we are not bacteria!*
  innacurate time/labour consuming

• **Animal testing:**
  Generally: assess micronuclei and chromosome aberration, or breaks (comet assay)
  Rarely: unscheduled DNA synthesis
  Expensive: Mutation using “Big Blue” or “MutaMouse”
  NEW! “Pig-A” trials looking good (cheaper and easier, generally blood)
Historical issues with the *in vitro* tests


Most carcinogens are positive in the battery of tests – good!

Most non-Carcinogens have a positive result too - bad!

Many non-Carcinogens are ‘uniquely positive’ in MLA/MNT/CA

Response:

New guidance on toxicity; top dose reduced from 10 mM to 1 mM (pharma) consideration of other approaches, and weight of evidence
Why are there so many misleading *in vitro* positive results in the published data?

- Testing at excessive toxicity
- Testing at irrelevantly high doses
- Particular properties of the test cells
  - bacteria are not eukaryotes
  - cell lines are not animals
  - cell lines are often p53 mutant (repair deficient)
- Growth media/conditions
- The now abandoned “LD 50” benchmark tests in animals overestimated cancer risk –
  - which meant that high *in vitro* test doses were needed, to get positive results for the “high dose carcinogens”.
Genotoxicity assessment for F&F is relevant to personal care sector

- It is tonnage based and stepwise:
  - EC: *in vitro* tests may only be followed by *in vivo* under strict conditions
    - < 1 ton per year: none required;
    - 1-10 tons per year: an Ames test;
    - >10 tons per year: Ames+MLA/HPRT/MNT/Cytogenetics
  - Animal testing is only allowed for *in vitro* positives: & NOT for cosmetics

- Most fragrance ingredients are low tonnage, so Ames is the only data...
  (and Ames doesn’t detect eukaryotic effects: aneugenesis/clastogenesis)

- No single regulatory test covers all events: but remember the graph!
  Combinations of low specificity tests lead to misleading/conflicting data.

- Go to animals if you can... but if not, then what?
  Screen for safety! How???
Genotoxicity assessments for F&F

_in silico_ tools: knowledge-based expert systems (DEREK Nexus), QSAR
“Read across”, needs enough ‘similars’, with enough data.

So, there is a need for additional test data:

to anchor _in silico_ tools, and add “weight of evidence”.

What additional data? The “BlueScreen HC” assay was chosen:

**Screening assay with accurate negative prediction**, to address the “false positive” problem presented by low specificity _in vitro_ mammalian tests

Definitive negative results produced in an early screen can:

- prioritise compounds for development
- **contribute to WOE** for compounds with conflicting regulatory results
BlueScreen HC? GreenScreen HC?

Similarities and Differences

BlueScreen HC

• GADD45a-**Gluc Reporter** in TK6 cells. With/without S9
  
  – Flash (not glow): needs injection luminometer
  – Cell toxicity estimated by DNA content using thiazole orange
  – 96/384 well microplate
  – 4 compounds per plate

No restriction on use

GreenScreen HC

• GADD45a-**GFP Reporter** in TK6 cells. With/without S9

  – Fluorescence based, plate reader assay (S9 needs flow)
  – Cell toxicity by absorbance (or PI in flow cytometer)
  – 96-well microplate
  – 4 compounds per plate

Restricted use in F&F sector
The GreenScreen GADD45a reporter assay

A 96 well, genotoxicity and cytotoxicity assay, designed to test 4 compounds at 9 dilutions, with two dose positive controls

Uses TK6 cells - human, p53 wt
Low compound requirement
- 5 mg for 0.5 mg/ml test
20 minutes to set up a plate, results in 48 hours
Test >60/day, >180/week
Simple transparent Excel data processing

RSG - toxicity

GFP - genotoxicity

Safe

Genotoxic
BlueScreen HC

• 4 compounds, 8 dilutions, duplicate strains

Duplicate tests in just one strain

Why the difference?

GSHC needs control strain without GFP to compensate for cellular fluorescence

BSHC needs duplicate test because of variability in signal output
### Performance of in vitro tests?

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity &amp; Specificity to genotoxic carcinogens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(% carcs positive) &amp; (% non carcs negative)</td>
</tr>
<tr>
<td>Ames</td>
<td>60.3% 77.3%</td>
</tr>
<tr>
<td>MNT</td>
<td>80.9% 53.8%</td>
</tr>
<tr>
<td>GADD45a</td>
<td>87.5% 95%</td>
</tr>
<tr>
<td>In silico</td>
<td>15-68% 48-95%</td>
</tr>
</tbody>
</table>

GADD45a assays have high specificity – this property is important for screening.

In this presentation I also discuss figures for the **Predictivity** of the tests:

i.e. proportion of GADD45a **negatives**/positives that are **negatives**/positive in another test

e.g. “how well does a negative GADD45a result predict ‘safe’ Ames negative results?”
Hypothesis for high specificity?

- p53 RE is **not at the promoter**
- p53 only affects the promoter *indirectly*, via WT1 binding
- p53 RE and WT1 mutations equally repress reporter

MYC blocks p53 interaction with promoter WT1
Copied from Tolstonog and Deppert
Nat. Str. Mol. Biol. 16 900-901 (2009)

Removal of MYC from WT1 brings specificity to p53 regulation:
*There are non-genotoxic p53 inducers!*
Assessment of the genotoxicity of S9-generated metabolites using the GreenScreen HC GADD45a–GFP assay. Jagger et al., 2009

Development and validation of a higher throughput screening approach to genotoxicity testing using the GreenScreen HC assay. Knight et al., 2009

Analysis of 75 marketed pharmaceuticals using the GADD45a-GFP ‘GreenScreen HC’ genotoxicity assay. Hastwell et al., 2009

Evaluation of High-throughput Genotoxicity Assays Used in Profiling the US EPA ToxCast Chemicals. Knight et al, 2009

GADD45a-GFP GreenScreen HC assay results for the ECVAM recommended lists. Birrell et al., 2010

A pre-validation transferability study of the GreenScreen HC assay with a metabolic activation system S9. Billinton et al. 2010 Ring Trial


Review: How accurate is in vitro prediction of carcinogenicity Walmsley&Billinton, 2011

More recent...

-The 'BlueScreen HC' assay as a decision making test in the genotoxicity assessment of flavour and fragrance materials. Etter et al. (2015). Toxicol In Vitro, 29(7), 1425-1435


-Cytotoxicity and genotoxicity of urban particulate matter in mammalian cells. Dumax-Vorze et al., (2015). Mutagenesis


Histone deacetylase inhibitors produce positive results in the GADD45a-GFP GreenScreen HC assay. Johnston and Walmsley Mutation Research, 751, 96-100

What is an appropriate top dose?

- Pharma sector reduced top *in vitro* tests dose from 10 mM to 1 mM

- Misleading *in vitro* positive results at high concentrations, are prevalent in low molecular weight compounds - many compounds in this study fall into the low molecular weight category!

- So, we reviewed GADD45a validation data from 161 compounds in order to discover how a reduction in top testing dose from 10 mM to 1 mM would affect its performance:

  - 5 substances were positive at 10 mM top dose that would be GADD45 negative at 1 mM top dose:
    - valacyclovir, *not carcinogenic*, and azidothymidine (high dose female-specific rodent carcinogen) adversely affect DNA replication and repair (Antiviral nucleoside analogues)
    - EDTA and p-nitrophenol, produce **artifactual positive** results: only positive in the 3h S9 exposure protocol at concentrations unreachable in the standard 48h exposures
    - Ofloxacin, a bacterial gyrase inhibitor: **negative in vivo and carc**. Positive results in other *in vitro* genotoxicity assays: *a photo-mutagen, equivocal GSHC data when light is excluded*.  

- Lowered top dose does would not lead to a loss of sensitivity in the GADD45a assays.  
  There is a strong case for following the Pharma guidance:  
  Classify all *in vitro* mammalian positive results >1 mM as “misleading” or “not relevant”  
  ONLY 10mM top dose studies are considered further (ask about 1 mM)
The F&F substances in this study

- 70: selected with as full data sets as possible
  i.e. with animal data where possible (not easy!)

- broad range of chemical structures commonly encountered in the F&F industry: esters, ketones, aldehydes, nitriles, alcohols and thiols

- All tested to 10 mM – to be consistent with reference test data
<table>
<thead>
<tr>
<th>Test Data Available</th>
<th># of Compounds With Test Data Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlueScreen HC</td>
<td>70</td>
</tr>
<tr>
<td>Ames</td>
<td>61</td>
</tr>
<tr>
<td><em>In vitro</em> mammalian gene mutation (MLA/HPRT)</td>
<td>31</td>
</tr>
<tr>
<td><em>In vivo</em> MN</td>
<td>15</td>
</tr>
<tr>
<td><em>In vitro</em> CA</td>
<td>38</td>
</tr>
<tr>
<td><em>In vivo</em> MN</td>
<td>33</td>
</tr>
<tr>
<td><em>In vivo</em> CA</td>
<td>7</td>
</tr>
<tr>
<td><em>In vivo</em> Comet</td>
<td>8</td>
</tr>
<tr>
<td><em>In vivo</em> UDS</td>
<td>10</td>
</tr>
<tr>
<td><em>In vivo</em> Transgenic Studies</td>
<td>7</td>
</tr>
<tr>
<td><strong>Global <em>in vivo</em> Data</strong></td>
<td><strong>36</strong></td>
</tr>
<tr>
<td><strong>Rodent Carcinogenicity</strong></td>
<td><strong>21</strong></td>
</tr>
</tbody>
</table>

**Observations:**
There are many “data gaps” in this applicability domain...
...reliability of analyses is greatest for substances with most data!
Overview of the BlueScreen Positive Results


- Compounds tested up to 10 mM (or limit of solubility)
- **20 of the 70 (28%)** produced BlueScreen positive results
- **9 of the 70 (13%)** only positive above 1 mM

**Positive only with S9**
- Quinole
- Methyl Eugenol
- Glyoxal
- Delta damascone
- Alpha damascone
- Oxacyclohexadec-12(+13)-en-2-one

**Positive only without S9**
- Isophorone
- Octahydro-5-methoxy-4,7-methano-1(H)-indene-2-carboxaldehyde
- 5,8-Methano-2H-1-benzopyran-2-one, 6-ethylideneoctahydro-

**Positive with & without S9**
- Eugenol
- Beta asarone
- Furfural
- Allyl isothiocyanate
- 4-Hydroxy-2,5-dimethyl-3(2H)furanone
- Resorcinol
- 4-Phenylbut-3-en-2-one
- p-Mentha-1,8-dien-7-al
- 5-Methylthiophene-2-carbaldehyde
- Maltol
- 7-methyloct-3-en-2-one

How did BlueScreen HC compare to and predict Ames and *in vitro* test results?
# Prevalence of positive results from various tests

<table>
<thead>
<tr>
<th>Genotoxicity Test</th>
<th>Proportion of F&amp;F positive results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames</td>
<td>10 / 61 (16.3%)</td>
</tr>
<tr>
<td>\textit{In vitro} MLA / HPRT</td>
<td>10 / 31 (23.3%)</td>
</tr>
<tr>
<td>\textit{In vitro} MN</td>
<td>11 / 15 (73.3%)</td>
</tr>
<tr>
<td>\textit{In vitro} CA</td>
<td>19 / 38 (50.0%)</td>
</tr>
<tr>
<td>Global \textit{in vitro} positives (e.g. positive in at least 1 assay)</td>
<td>27 / 61 (44.0%)</td>
</tr>
</tbody>
</table>

- Ames Retains 84% of compounds 😊
- but how many Ames negatives are safe?
Negative Prediction by BlueScreen HC developable compounds

• Important part of screening – need confidence in negatives!

• 95% (38 / 40) BlueScreen HC negatives were Ames negative
  – Exceptions: Trans-2-hexenal and Pent-1-en-3-one, both Ames positive but *in vivo* negative
  – WOE: An Ames positive with a BSHC negative is absolutely worth taking forward!

• 75.6% (31 / 41) of BlueScreen HC negatives were *in vitro* negative
  – Does this mean there were 10 BlueScreen ‘false negatives’?
  – No! The 10 *in vitro positives* were largely *in vivo* negative / non-carcinogens (next slides)
  – WOE: An *in vitro* positive with a BSHC negative is worth taking forward

A BlueScreen HC negative compound is highly likely to be Ames negative,
(and very likely to be *in vitro* negative)

The F&F Strategy is based on negative Ames data, so good prediction is useful!
**BSHC v in vitro Conflicts**

9 BlueScreen negatives with positive *in vitro* mammalian test data: are they safe?

<table>
<thead>
<tr>
<th></th>
<th>In vivo Test Results</th>
<th>Rodent Carc Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>D,L-Menthol</td>
<td>NEGATIVE</td>
<td>NEGATIVE</td>
</tr>
<tr>
<td>Benzyl Alcohol</td>
<td><em>In vivo MN</em> NEGATIVE</td>
<td>NEGATIVE</td>
</tr>
<tr>
<td></td>
<td><em>In vivo Comet</em> POSITIVE</td>
<td></td>
</tr>
<tr>
<td>Isobutyraldehyde</td>
<td><em>In vivo MN</em> NEGATIVE</td>
<td>NEGATIVE</td>
</tr>
<tr>
<td></td>
<td><em>In vivo CA</em> POSITIVE</td>
<td></td>
</tr>
<tr>
<td>Ethyl acrylate</td>
<td>NEGATIVE</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>Styrene</td>
<td>EQUIVOCAL</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>Benzaldehyde</td>
<td>NEGATIVE</td>
<td>POSITIVE</td>
</tr>
<tr>
<td>Geranyl nitrate</td>
<td>POSITIVE</td>
<td>N/A</td>
</tr>
<tr>
<td>Trans-2-hexenal</td>
<td>NEGATIVE</td>
<td>N/A</td>
</tr>
<tr>
<td>Pent-1-en-3-one</td>
<td>NEGATIVE</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- 5 / 9 are clearly *in vivo* Negative (8 / 9 if Equivocal counted Negative)

- 3 have only Negative Carc data

- 3 with Positive Carc data are *in vivo* negative/equivocal (mechanism?):
  - Ethyl Acrylate is *non-geno Carc*
  - Styrene requires *special S9*, (not used in this study)
  - Benzaldehyde is *mouse specific* carcinogen

- Geranyl nitrate positive most likely to be synthesis impurity related

- Trans-2-hexenal and Pent-1-en-3-one are also Ames misleading positive! Negative *in vivo*

These results suggest that BlueScreen Negative F&F Materials should be progressed.
Let’s look again at 3 of those 9 compounds

- All Ames negative, all \textit{in vitro} positive, all at risk – \textbf{with carcinogenicity data}

If this was an anonymous collection,  
and the compounds had been BSHC screened

Their negative BSHC results would have allowed them to be carried forward for Ames.

Their Ames negative results would then allow them into production.

But if tonnage had required \textit{in vitro} tests (>10 tons),  
their positive results would have required \textit{in vivo} testing

2 of 3 would be negative (as predicted): no further tests. They are not genotoxic carcinogens

The third, styrene would eventually be found to be a non-relevant, metabolism specific effect

- 3 with Positive Carc data are \textit{in vivo}  
  Negative or Equivocal (mechanism?)

- Ethyl Acrylate is \textbf{non-genO Carc}

- Styrene requires \textit{special S9}, not used in this study

- Benzaldehyde is \textbf{mouse specific carcinogen.}

BSHC data could have been used as weight of evidence before these extensive studies...
Of the 20 materials producing positive BSHC results (≤ 10 mM), all but one* had Ames and *in vitro* genotoxicity data

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

This is consistent with validation studies: the GADD45a assays are effective in identifying eukaryote-specific *genotoxins missed by the Ames test*

**Regulatory *in vitro* mammalian tests:** 17 of the 19 (89 %)** materials positive in BSHC at 10 mM 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.

* 4-Hydroxy-2,5-dimethyl-3(2H)furanone (Ames only).
**exceptions: methyl eugenol and Oxacyclohexadec-12(+13)-en-2-one.
Conclusions

70 flavour and fragrance substances (a high proportion with *in vivo* and/or carcinogenicity test data) were tested (+/- S9) to a top dose of 10 mM (solubility/toxicity permitting)).

i. BlueScreen HC negative results are highly predictive of negative *in vivo* genotoxicity, and negative Ames results: *i.e. high safety rating*

ii. BlueScreen HC positive results are highly predictive of positive results from regulator-required *in vitro* genotoxicity assays *i.e. compounds likely to cause problems*. (The moderate negative Predictivity of BlueScreen HC for the *in vitro* test set of material is due mainly to the high rate of ‘false positives’ in regulatory *in vitro* mammalian tests)

iii. In the flavour and fragrance domain, which comprises a large proportion of relatively low molecular weight molecules, (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 with simple low compound-requiring microplate format) supports further investigation of the utility of BSHC as a tool in prioritizing the assessment/development of new F&F materials in data gap filling for materials with no/limited regulatory test data.
Acknowledgements

**Firmenich** for collaborating and letting us present their findings: Ben Smith and Sylvain Etter

**Gentronix** – the team who did all the BSHC testing

– Matt Tate, Louise Birrell, Paul Cahill, Heather Scott, Nick Billinton

**PCPC** – For the opportunity to present this study to you!
Thank you!

Questions?

Ask for a pdf of the paper!
BSHC Negative Predictivity:
spotting developable compounds

1. Ames test results – most prevalent:
   and positive Ames is least easy to manage

   50 materials which produced a negative BSHC result had Ames test data

10 mM cut-off:
   95% (38/40) BSHC negative substances were negative in Ames.

1 mM cut-off:
   90%, (47/52) BSHC negative substances were negative in Ames.

The high Predictivity of Ames negative results by BSHC provides a reliable indication that a substance can be carried forward for development without Ames liability.
BSHC Negative Predictivity: spotting developable compounds

2. Regulatory *in vitro* mammalian test data - less prevant

29 materials which produced a BSHC negative result had *in vitro* data

10 mM cut-off: 21 of 29 of the BSHC negative substances (72%) were negative in one or more *in vitro* mammalian cell tests.

The 8 BSHC negative *in vitro* positives were all non-genotoxic carcinogens, non-carcinogens etc (e.g. geranyl nitrile and styrene above).

1 mM cut-off: An additional 10 BSHC substances with mammalian test data tested negative, of which 3 are *in vitro* mammalian test negative. This produces a negative *in vitro* prediction figure of 23 /39 (59%).

The lower negative BSHC prediction of *in vitro* mammalian tests (c.f. Ames) is consistent with the much higher specificity of BSHC c.f. regulatory *in vitro* tests.