Application of in silico modelling to estimate toxicity of migrating substances from food packaging

Nicholas Price, Qasim Chaudhry*

The Food and Environment Research Agency, Sand Hutton, York YO41 1LZ, UK

ARTICLE INFO

Article history:
Received 13 April 2014
Accepted 26 May 2014
Available online 9 June 2014

Keywords:
QSAR
Read across
Weight of evidence
In silico methods
Computational toxicology
Packaging migrants

ABSTRACT

This study derived toxicity estimates for a set of 136 chemical migrants from food packaging materials using in silico (computational) modelling and read across approaches. Where available, the predicted results for mutagenicity and carcinogenicity were compared with published experimental data. As the packaging compounds are subject to safety assessment, the migrating substances were more likely to be negative for both the endpoints. A set of structural analogues with positive experimental data for carcinogenicity and/or mutagenicity was therefore used as a positive comparator. The results showed that a weight of evidence assembled from different in silico models and read-across from already-tested structurally similar compounds can provide a rapid and reliable means for rapid screening of new yet-untested intentional or unintentional chemical compounds that may migrate to packaged foodstuffs.

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

1. Introduction

1.1. The FACET project

The work reported here was designed to test the in silico predictive strategy used in the FACET project (Flavourings, Additives and food Contact materials Exposure Task – www.ucd.ie/facet), which was funded by the European Commission under the framework programme to produce a risk management tool consisting of a database of information on the levels of food additives, flavourings, food packaging migrants, and corresponding food consumption data (Hearty et al., 2011). The objectives within the food packaging migrant element of the project (Oldring et al., 2014a,b) also aimed at establishing the physicochemical properties that are related to migration of chemicals from food packaging, classification of foods in relation to their migration behaviour, mathematical modelling to estimate migration from packaging to various food types, and the possible use of in silico (computational) predictive models to evaluate specific toxicological endpoints of migrating compounds.

1.2. Predictive toxicology of food migrating substances

The migration of chemicals into food is a potentially serious issue relating not only to the quality of food but also to the safety of the consumer (Barnes et al., 2007). Some of the chemicals used to make food packaging have been shown to migrate from packaging into packaged foodstuffs. Generally, such compounds have been assessed by expert panels for safety, and have published toxicological data available. The amount and quality of the available data, however, differs widely among the compounds. In addition, for some compounds, the migrating substances have been found to be the breakdown products or impurities rather than the parent compound itself. Pragmatic approaches such as Threshold Level of Concern (TTC), are often based on rather crude indices of toxicity, such as the Cramer index, (Pinalli et al., 2011), and could be improved with a more data-intensive approach. The aim of this study was therefore to attempt to establish a set of “in silico” procedures that would add to the body of information on the existing migrating compounds, and serve as a starting point for assessment of new yet-untested compounds. The techniques used included a combination of expert systems and predictive computational models based on Structure Activity Relationships (SAR), and Quantitative Structure Activity Relationships, (QSAR).

Since the range of chemical types that may migrate from packaging into food can be very wide, no single model was deemed likely to suffice for all types of compounds or for all toxicological endpoints. A “weight of evidence” (WoE) approach was therefore

* Corresponding author. Tel./fax: +44 1904462584.
E-mail address: qasim.chaudhry@fera.gsi.gov.uk (Q. Chaudhry).
taken as currently recommended for regulatory predictive toxicology (Balls et al., 2006). In addition, since the methodology was intended to be used widely in the industry and by regulatory bodies, all software platforms used were those that are publicly available.

By way of validating the approach taken in the FACET project, the current work assessed the quality of mutagenicity and carcinogenicity endpoints on the packaging migrant compounds that had measured toxicological data available for the two endpoints. Since all compounds used in plastics food packaging go through a rigorous process of assessment by expert panels, it was expected that the migrant substances would not be carcinogens or mutagens. Similarly, substances used in other packaging materials such as paper/board, inks and adhesives, will have undergone some degree of toxicity pre-selection by users. Although a predictive system that would have returned a result of “negative” for these compounds could be deemed to have performed well, it would have not allowed validation of the predictive strategy used. In view of this, a series of structural analogues of the migrant compounds that were tested positive for mutagenicity and carcinogenicity was also included in the study as a positive control. Structural and compositional similarity of this “test set” to the FACET set was considered a good test of the WoE approach in identifying those structural characteristics which confer carcinogenicity or mutagenicity.

2. Materials and methods

2.1. The OECD QSAR toolbox

The OECD Toolbox is a software application intended to be used to fill gaps in toxicity and ecotoxicity data needed for assessing the hazards of chemical substances. The Toolbox incorporates databases on chemical data (e.g. properties), experimental toxicological and ecotoxicological data, and estimated values from a large range of QSAR tools, together with incorporated QSAR modelling and Expert Systems, built within a regulatory application chassis. This package therefore allows the user to perform a number of functions (OECD, 2008), e.g. to:

- identify analogues for a chemical, retrieve experimental results available for those analogues and fill data gaps by read-across or trend analysis;
- categorise large inventories of chemicals according to mechanisms or modes of action;
- fill data gaps for any chemical by using the library of QSAR models;
- evaluate the robustness of a potential analogue for read-across;
- evaluate the appropriateness of a Q SAR model for filling a data gap for a particular target chemical; and
- build QSAR models.

For this study version 2.3 of the Toolbox was used that had been augmented with a number of extra publicly available databases with carcinogenicity and mutagenicity data.

2.2. The VEGA platform

The VEGA platform (www.vega-qsar.eu), has been developed by the Istituto di Ricerche Farmacologiche Mario Negri in Milan with a number of collaborating organisations and through a series of EU-funded projects. The models used in VEGA for carcinogenicity and mutagenicity originated in the EU project CAESAR, (www.caesar-project.eu), with subsequent improvements and additions from contributing organisations, and incorporating some of the models in the US-EPA Toxicity Estimation Software Tool; T.E.S.T (www.epa.gov/nrmrl/std/qsar/qsar.html).

VEGA models generate transparent, reproducible, and verifiable results. The system comprises a series of tools that have been optimised so that the results obtained for a target compound can also be related to those for other structurally related compounds. VEGA also has a comprehensive 5-point validation system that allows the user to assess the reliability of predictions.

2.3. Study substances

Initially 78 “substances” were selected for the study that are used in inks and plastic food packaging. Using the summary information in the expert assessments of the EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids, published in the EFSA Journal, a set of discrete chemical compounds representing the actual, or most likely migrating moieties of each of the plastic substances was identified. In most cases but not all, the migrating moiety was the same as the parent compound. However, particularly in the case of oligomers, polymers, or mixtures, it was necessary to identify the specific migrant compound(s) and in some cases there were two or more. In some cases the same compound was found to be the most appropriate migrating substance for more than one parent. This resulted in a total of 136 discrete chemical compounds that were used in this study, being derived from the initial 78. The study set of 136 consisted of a wide variety of organic compounds including alcohols, aldehydes, alkanes, alkenes, amides, carboxylic acids, esters, ketones, organophosphates, phenols, phenolic acids, piperidines and sulphonamides. Names and SMILES notation for all of the compounds, (including where a the same compound was a migrant from more than one parent substance) can be found in Supplementary Table 1.


Of the 136 study compounds, 70 had published results of Ames mutagenicity tests, whilst only 37 had data for carcinogenicity. For the purposes of this study, it was assumed that the remaining compounds were negative for these endpoints based on the expert scrutiny to which they had been subjected prior to approval as food packaging substances.

All 136 compounds were input into software using SMILES strings or .mol files. This structural information was obtained by searching the same databases as described above. Where structural data were not available, mollies were constructed from known fragments using Accelrys Draw, http://accelrys.com.

2.4. Carcinogenic and mutagenic analogues

Further structural analogues of the migrant compounds were identified from toxicity databases for which measured mutagenicity or carcinogenicity data were available to regard them carcinogens or mutagens. This was done to assess the ability of the SAR and QSAR systems to predict the endpoints of structurally similar, but toxicologically dissimilar compounds. The selection criteria used were that the analogue had to be more than 70% similar to the parent compound, with positive published results for either mutagenicity and/or carcinogenicity, and be within the prediction domain for the QSARs in the OECD QSAR Toolbox. Structural similarity of analogues was assessed by using the similarity module in the QSAR Toolbox. Each parent compound was submitted to a search of the Toolbox databases for >70% similarity based on the Dice algorithm (Dice, 1945), using atom pairs, atom type and cycles as the basis for comparison. The resulting analogues were screened for those with positive carcinogenicity and/or mutagenicity data, and also for their domain suitability for the Toolbox QSARs for those endpoints. From these, compounds that were most similar to the target packaging migrant compounds were selected for further study. Not all of the migrant compounds yielded suitable analogues with >70% structural similarity, and only 49 compounds were found that met the criteria and are shown in Supplementary material.

2.5. Carcinogenicity and mutagenicity predictions

2.5.1. Predictive QSAR models

A number of validated QSARs are built into the Toolbox and the ones used for this study were all developed or adopted by the Danish Environmental Protection Agency for use in their Danish (QSAR Database (http://qsar.food.dtu.dk/). These were male mouse carcinogenicity (MultiCASE commercial model AG4); male rat carcinogenicity (MultiCASE commercial model AG1); female rat carcinogenicity (MultiCASE commercial model AG2); and Ames mutagenicity (MultiCASE commercial model A2H).

Other models used included male/female rat carcinogenicity and in vitro mutagenicity based on the Ames test, using Salmonella typhimurium, (http://130.226.165.14/User_Manual_Danish_Database.pdf); and VEGA QSARs as described above. The results were used only for those compounds that had a good or moderate reliability based on the VEGA domain assessment scores.

2.5.2. Read across

Read across is a technique for predicting toxicity endpoints for a given compound from the measured toxicity values of other compounds that have a close structural or mechanistic similarity to the query compound (Gallegos Saliner et al., 2005). Thus, for each target compound requiring predictions of carcinogenicity and mutagenicity, a category of analogues was constructed using one criterion for structural similarity, and another for mechanistic similarity. As before,
structural similarity was assessed using the Dice algorithm, and compounds assessed as >70% similar were further screened for mechanistic similarity using one or more of the Toolbox profilers for carcinogenicity or mutagenicity depending on the endpoint required for prediction. Read across was performed on the resulting categories, and acceptable predictions recorded as ‘positive’ or ‘negative’ were based on a 66% or better probability. Any results below 66% probability were recorded as “equivocal”.

2.5.3. Weight of evidence

There is no definitive method for devising a weight of evidence from given sets of data. The European Chemicals Agency recommends the combination of reliability, relevance and adequacy, when constructing a weight of evidence argument, (ECHA, 2010). Within this overall scheme, each criterion is to be assessed on a complex set of value judgements. This approach encompasses all types of evidence including the data obtained through in-silico methods. Specifically for (QSAR data, the US FDA offers a pragmatic approach to WoE using the results from several (Q)SAR methods (Matthews, 2009). As this approach was more directly related to the situation in this study, a modified US FDA approach was adopted as follows:

- Data with low reliability from a (Q)SAR were excluded.
- Data from “out of domain” compounds were excluded.
- “Positive” predictions were assigned if 2 or more methods returned a positive prediction and less than 2 gave negative or equivocal predictions.
- “Negative” predictions were assigned if 2 or more methods returned a negative prediction and less than 2 gave positive or equivocal predictions.
- “Equivocal” predictions were assigned if one of the above conditions was not fulfilled.

The final prediction of toxicity for each endpoint was based on a combination of results from read across, and QSAR models within VEGA and the OECD Toolbox. The assessment showed that 65 out of the 136 migrant compounds were outside the domain of any of the relevant QSARs in the Toolbox. For these compounds SAR alerts were used for carcinogenicity or mutagenicity, together with read across and VEGA results where validated. These SAR alerts are part of the profiling module of the Toolbox and comprise carcinogenicity and mutagenicity (Ames tests) alerts from the expert system Toxtree (http://toxtree.sourceforge.net/). An unequivocal prediction was recorded where all methods gave the same prediction, or if the result from one of the methods was equivocal but at least two others were unequivocal. In the case of the analogues of the migrant compounds QSARs were always available because this was one of the criteria for their selection.

3. Results

3.1. Packaging migrant compounds

Table 1 shows the results of predictions for carcinogenicity and mutagenicity for the packaging migrant compounds studied. Of the 136 compounds, 70 had published results for Ames mutagenicity; of which 65 were negative, 5 equivocal, and 3 positive.

Mutagenicity predictions for all 136 study compounds were similar whether the WoE was gathered by the FDA approach or by a simple majority approach, in which compounds were assigned positive if more models gave a positive prediction than negative, and vice-versa for negative predictions. The two respective approaches predicted 91% and 94% of the compounds as negative in the Ames test. Using a simple majority approach, the number of equivocal results was reduced from 10 to 2, but the positives increased from 2 to 6. Carcinogenicity predictions were less clear due to a large number of compounds that came out as equivocal. Using the simple majority approach, the WoE regarded 29 compounds as equivocal for carcinogenicity as compared to 51 when WoE was gathered by the FDA method. The percentage of the compounds predicted negative was 60% for the FDA method and 73% for a simple majority approach. From these data, there is no way of knowing how good these predictions were, as we had assumed all the compounds to be negative for both endpoints. Published data were available for 71 of these compounds for mutagenicity and 37 for carcinogenicity.

Table 2 shows the result from a FDA weight of evidence analysis of these compounds.

The results showed a good rate of accurate prediction of the measured values for mutagenicity. However, as explained previously, these are mostly negative as would be expected. The results for carcinogenicity were comparatively less consistent as a high proportion of results from the individual models showed a conflict, resulting in equivocal predictions. Interestingly, the measured data for carcinogenicity appears similarly conflicting with 7 of the compounds having both positive and negative results in the various databases.

3.2. Positive analogues

The FDA weight of evidence prediction for the positive analogues is shown in Table 3.

In the case of the positive analogues, albeit with smaller numbers of compounds, the results for carcinogenicity are similar to the non-carcinogenic migrant compounds, whilst those for mutagenicity are not so consistent for the positive analogues. The obvious question therefore was whether any of the individual prediction methods used was significantly better or worse than the others that could have resulted in the overall lowering of predictivity when used in a weight of evidence context.

3.3. Performance of the individual models

In order to compare the individual models that were used on both sets of compounds to derive the weight of evidence predictions, the standard parameters, (so called Cooper statistics, Cooper et al., 1979) were used to assess the quality of the predictions, as below:

\[
\text{Sensitivity (True positive rate)} = \frac{TP}{TP + FN} \\
\text{Specificity (True negative rate)} = \frac{TN}{TN + FP} \\
\text{Accuracy} = \frac{(TN + TP)}{(TN + FP + FN + TP)} \\
\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}
\]

where TP = True positive, TN = True negative, FP = False positive, FN = False negative.

Sensitivity is defined as the percentage of correctly classified positive predictions among the total number of positive instances. Specificity is the percentage of correct negative predictions compared to the total number of negatives.

Accuracy (concordance or “Q”) is defined as the total number both positive and negatives correctly predicted among the total number of compounds.

MCC (Matthews correlation coefficient) is a weighted value that overcomes any imbalance in the data classes which might lead to over optimistic values of Q. (Matthews, 1975) A MCC value of 1 indicates that the model can predict the data classes of unknown

<table>
<thead>
<tr>
<th>Weight of evidence</th>
<th>Mutagenicity</th>
<th>Carcinogenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
<td>Equivocal</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Equivocal</td>
</tr>
<tr>
<td>FDA method</td>
<td>124</td>
<td>10</td>
</tr>
<tr>
<td>Simple majority</td>
<td>128</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 2
Weight of evidence predictions from the FACET compounds with available measured endpoint data. M = measured values, P = predicted values, NA = not available.

<table>
<thead>
<tr>
<th>Result</th>
<th>Mutagenicity</th>
<th>Carcinogenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>P</td>
</tr>
<tr>
<td>Negative</td>
<td>65</td>
<td>64</td>
</tr>
<tr>
<td>Positive</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Equivocal</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>False negative</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>False positive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No result</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>71</td>
</tr>
</tbody>
</table>

Table 3
Weight of evidence predictions from the analogues compounds with positive endpoint values. M = measured values, P = predicted values, NA = not available.

<table>
<thead>
<tr>
<th>Result</th>
<th>Mutagenicity</th>
<th>Carcinogenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>P</td>
</tr>
<tr>
<td>Negative</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Positive</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>Equivocal</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>False negative</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>False positive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No result</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>26</td>
</tr>
</tbody>
</table>

compounds perfectly, whilst a MCC value of 0 indicates that the predictions are no better than random guessing, and a MCC value of −1 indicates total disagreement between the predicted data and the actual data.

Tables 4 and 5 show the Cooper statistics for all individual models used and for the overall FDA weight of evidence.

4. Discussion

4.1. Weight of evidence predictions

On the assumption that all 136 original packaging migrant compounds studied were negative for both mutagenicity and carcinogenicity, the FDA weight of evidence approach correctly predicted around 90% of them as negative for mutagenicity, but only 60% for carcinogenicity. One reason for this was the large number of equivocal results recorded under the FDA WoE method. If a simple majority method was used instead, the number of equivocal results dropped significantly, and the correct prediction rate for negative carcinogenicity rose to 73%. When considering only those compounds with measured data, the success rate for the FDA WoE was 92% for mutagenicity and 70% for carcinogenicity. This perhaps demonstrates the more complex nature of the carcinogenicity endpoint compared to mutagenicity. Indeed examination of some of the carcinogenicity databases shows a variation in the approaches used to derive classification data (positive/negative) from continuous data such as a TD50. For example, in selecting the CAESAR database (used for the model of the same name in the VEGA platform), any compound for which a finite value for TD50 was recorded, was allocated to the carcinogen class. The Benigni Bossa dataset were derived using different criteria as some compounds allocated as carcinogens in the CAESAR set are allocated as non-carcinogens in the Benigni–Bossa set. This is likely to be the case in other databases, as a number of compounds found in the various databases of the Toolbox have been found to have conflicting measured data. In some cases this may simply be due to experimental variation at the margins but in others there is clearly a conflict across databases, one reason for which may be differences in the classification criteria.

For the positive mutagens and carcinogens, the results were less promising and only 70% and 60% of the compounds respectively could be correctly predicted by the FDA WoE approach. Clearly an investigation of the contribution of the individual models to the overall WoE was necessary.

4.2. Contribution of Individual Models to WoE

The relative abilities of the individual models to predict mutagens and carcinogens are shown in Tables 4 and 5. For mutagenicity, the true negative rate (or Specificity) is high in all cases, though the true positive rate (or Sensitivity) is variable, with the Toolbox FDA Ames QSAR model being the best overall predictor with an overall accuracy (Q) of 95.4% and a Matthews correlation coefficient of 0.9, indicating a nearly perfect predictivity of this model on this very mixed array of compounds. The SarPy model from the Vega suite performed well, but read across was poor at picking up the mutagens, possibly reflecting a lack of robust structural and mechanistic analogues to form a category. It may also be that the detailed procedure used to assemble the category was flawed in some way, and a systematic examination of the factors which make up good read across analogues may be beneficial.

Clearly in this case, although WoE ended up giving a good set of predictions (Q = 94%, MCC = 0.85), the use of the Toolbox FDA Ames QSAR in this instance would have sufficed.

The picture for the carcinogens is rather different. The Benigni–Bossa model from Vega, which did not perform very well on the

Table 4
Cooper statistics for mutagenicity predictions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>VEGA (CAESAR)</th>
<th>VEGA (SarPy)</th>
<th>VEGA (Benigni Bossa)</th>
<th>Toolbox (Danish EPA Ames)</th>
<th>Toolbox (Read across)</th>
<th>WoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>68.70</td>
<td>78.3</td>
<td>65.0</td>
<td>87.5</td>
<td>64.7</td>
<td>78.2</td>
</tr>
<tr>
<td>Specificity</td>
<td>94.10</td>
<td>100</td>
<td>100</td>
<td>100.0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Accuracy (Q)</td>
<td>84.60</td>
<td>93.6</td>
<td>87.9</td>
<td>95.4</td>
<td>91.5</td>
<td>94.0</td>
</tr>
<tr>
<td>MCC</td>
<td>0.65</td>
<td>0.85</td>
<td>0.74</td>
<td>0.90</td>
<td>0.76</td>
<td>0.85</td>
</tr>
</tbody>
</table>
mutagen data appeared the best of the group on the carcinogen data with an accuracy (Q) of 97.1% and an MCC of 0.93, indicating almost perfect prediction. Again sensitivity was poorer than specificity across the piece, with the exception of the Benigni–Bossa model. Read across performed fairly well, and the resulting WoE predictions were very good (Q = 97.4%, MCC = 0.95). These results show the benefit of using both positive and negative “controls” in the same way that would be done in any scientific bench research. The controls allow the user to pinpoint any anomalies and to indicate for a given endpoint and set of compounds the most likely models to give accurate predictions. The results also indicate that WoE is not always necessarily the best, but may be the safest approach in the absence of a detailed examination of each model as presented here.

4.3. Predictions in context

The Cooper statistics shown above give a better overall picture of the predictions than is actually the case. This is because the parameters calculated do not take account of equivocal results. The unequivocal predictions from this study show a very good concordance with measured values, and demonstrate that this approach could be useful in predicting human health endpoints of a wide range of structural types. If such an approach is used on compounds with unknown carcinogenicity or mutagenicity profiles, in the event of obtaining equivocal predictions it could be suggested that other models be applied or that in-vitro or in-vivo tests need to be performed.

5. Conclusions

An increasing emphasis is being placed on in silico predictive models for regulatory toxicology, (Serafimova et al., 2010; Worth and Mostrag-Szllichtyng, 2012), with many regulatory authorities adopting or testing (Q)SAR models for their regulatory procedures for different regulated chemical substances, such as medicines, pesticides, food and cosmetic additives. In Europe, the REACH regulations have been a driver for the improvement of in-silico predictive toxicology, since, like some other countries, animal testing is now being phased out, where possible. Such regulations require many thousands of chemical compounds to be tested for a range of health and environmental endpoints, whilst minimizing the use of animal testing. Thus in-silico testing is becoming more refined and fit for this purpose. In the past, there was little standardization in predictive toxicology and no proper guidelines as to what contributes to reliable predictions. The development and adoption of the Setubal (OECD) Principles, (OECD, 2004) in which a five-point validation process was proposed, led the way to in-silico predictions being admissible under a number of regulatory schemes across the world. For this purpose it is recommended that more than one method is used to obtain endpoint prediction and that these results should contribute to an overall weight of evidence, leading to a prediction (ECHA, 2008).

It is clear that results from all types of in-silico predictions need to be assessed with expert judgement, but in the case of read across such expert knowledge is needed to obtain robust predictions in the first place. Good read across predictions will also depend on the availability of good mechanistic and structural analogues upon which to base predictions. The results from this study

<table>
<thead>
<tr>
<th></th>
<th>VEGA (CAESAR)</th>
<th>VEGA (Benigni–Bossa)</th>
<th>Toolbox (Danish EPA female mouse)</th>
<th>Toolbox (Danish EPA male mouse)</th>
<th>Toolbox (Danish EPA female rat)</th>
<th>Toolbox (Danish EPA male rat)</th>
<th>Toolbox (Read across)</th>
<th>WoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>75.0</td>
<td>95.8</td>
<td>75.0</td>
<td>61.9</td>
<td>65.2</td>
<td>81.8</td>
<td>81.3</td>
<td>94.1</td>
</tr>
<tr>
<td>Specificity</td>
<td>83.3</td>
<td>100</td>
<td>89.5</td>
<td>85.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Accuracy (Q)</td>
<td>77.8</td>
<td>97.1</td>
<td>82.1</td>
<td>73.2</td>
<td>81.8</td>
<td>89.5</td>
<td>91.9</td>
<td>97.4</td>
</tr>
<tr>
<td>MCC</td>
<td>0.55</td>
<td>0.93</td>
<td>0.65</td>
<td>0.48</td>
<td>0.69</td>
<td>0.81</td>
<td>0.84</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**Table 5** Cooper statistics for carcinogenicity predictions.

![Fig. 1. Schematic outline for user workflows for the prediction of toxicity of chemical compounds.](image-url)
on a diverse set of chemical structures also points up the value of conducting predictions on “control” compounds for which measured endpoint values are known. On the basis of the findings of this study, a schematic outline for the use of in silico tools for the estimation of toxicity of untested chemical substances is proposed in Fig. 1.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Transparency Document

The Transparency document associated with this article can be found in the online version.

Acknowledgements

Development of the FACET exposure tool was funded by the European Union under Grant Agreement 211686 (Project FACET – Flavours, Additives and Food Contact material Exposure Task), with top-up funds (to Fera) from a consortium of 12 European packaging trade associations and from the UK Food Standards Agency. The findings and the conclusions in this paper are the responsibility of the authors alone and they should not be taken to represent the opinion of the funding bodies.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.fct.2014.05.022.

References


