JRC Scientific and Technical Reports

The Applicability of Software Tools for Genotoxicity and Carcinogenicity Prediction: Case Studies relevant to the Assessment of Pesticides

Andrew Worth, Silvia Lapenna, Elena Lo Piparo, Aleksandra Mostrag-Szlichtyng and Rositsa Serafimova

EUR 24640 EN - 2010

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC62109

EUR 24640 EN
ISBN 978-92-79-18762-9
ISSN 1018-5593
doi:10.2788/61485

Luxembourg: Publications Office of the European Union

Printed in Italy

ABSTRACT

This report presents research results obtained in the framework of a project on the Applicability of Quantitative Structure-Activity Relationship (QSAR) analysis in the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances. During this project, which was funded by the European Food Safety Authority (EFSA), the Joint Research Centre (JRC) performed several investigations to evaluate the comparative performance of selected software tools for genotoxicity and carcinogenicity prediction, and to develop a number of case studies to illustrate the opportunities and difficulties arising in the computational assessment of pesticides. This exercise also included an investigation of the chemical space of several pesticides datasets. The results indicate that different software tools have different advantages and disadvantages, depending on the specific requirements of the user / risk assessor. It is concluded that further work is needed to develop acceptance criteria for specific regulatory applications (e.g. evaluation of pesticide metabolites) and to develop batteries of models fulfilling such criteria.

LIST OF ABBREVIATIONS

AGES	Austrian Agency for Health and Food Safety
ANN	Artificial Neural Network
CAS	Chemical Abstract Service
CPDB	Carcinogenic Potency Database
CRD	UK Chemicals Regulations Directorate
EFSA	European Food Safety Authority
EU	European Union
FDA	United States Food and Drug Administration
GHS	Globally Harmonised Classification System (United Nations)
ISS	Istituto di Sanità (Italy)
JRC	Joint Research Centre
k-NN	K Nearest Nerighbours
MRL	Maximum Residue Level
OECD	Organisation for Economic Cooperation and Development
PC	Principal Component
PCA	Principal Components Analysis
PPP	Plant Protection Product
QSAR	Quantitative Structure-Activity Relationship
SAR	Structure-Activity Relationship
SVM	Support Vector Machine
TTC	Threshold of Toxicological Concern

CONTENTS

1. INTRODUCTION

1.1 Summary

This report presents research results obtained by the Joint Research Centre (JRC) in the framework of a project on the applicability of Quantitative Structure-Activity Relationship (QSAR) analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment (JRC, 2010). During this project, which was funded by the European Food Safety Authority (EFSA), the JRC performed several investigations to evaluate the comparative performance of selected software tools for genotoxicity and carcinogenicity prediction, and to develop a number of case studies to illustrate the opportunities and difficulties arising in the computational assessment of pesticides. This exercise also included an investigation of the chemical space of several pesticides datasets. The results indicate that different software tools have different advantages and disadvantages, depending on the specific requirements of the user / risk assessor. It is concluded that further work is needed to develop acceptance criteria for specific regulatory applications (e.g. evaluation of pesticide metabolites) and to develop batteries of models fulfilling such criteria.

1.2 Background

The general objective of food safety policy is to protect consumer health. In order to achieve this objective, regulatory bodies ensure that control standards are established and enforced.

In the European Union (EU), Regulation (EC) 178/2002 (EC, 2002) lays down the general principles and requirements of food law and procedures in matters of food safety, aiming at harmonising existing national requirements in order to ensure the free movement of food and feed in the EU. The Regulation ensures a high level of protection of human life and health, taking also into account the protection of animal health and welfare, plant health and the environment. One of the most important ways of protecting plants and plant products and of increasing agricultural production yields is the use of plant protection products (PPPs). Pesticides are used to protect crops before and after harvest against harmful organisms such as funghi, insects or weeds. A possible consequence of their use may be the presence of pesticide residues in the treated products. It is necessary to ensure that such residues are not found in food or feed at levels presenting an unacceptable risk to humans. Maximum residue levels (MRLs) are therefore set by the European Commission at the lowest achievable level consistent with good agricultural practices to protect consumers from exposure to unacceptable levels of pesticide residues in food and feed. As from 1 September 2008, a new legislative framework on pesticide residues, Regulation (EC) No 396/2005 (EC, 2005), has been applicable. This Regulation achieves the harmonisation and simplification of pesticide MRLs, while ensuring better consumer protection throughout the EU. With the new rules, MRLs undergo a common EU assessment to make sure that all classes of consumers, including the most vulnerable, such as children, are protected. All decision-making in this area is science-based and a consumer intake assessment is carried out by the European Food Safety Authority before concluding on the safety of an MRL.

A dietary risk assessment is therefore a prerequisite for any MRL-setting. A major difficulty stems from the fact that only the toxicological properties of the active substance are normally directly investigated through the range of toxicological studies required according to Directive 91/414/EEC (EC, 1991). The Organisation for Economic Cooperation and Development (OECD) guidance document on the definition of residue (OECD, 2009), however, requires the consideration of human relevance for risk assessment of all metabolites the consumer is exposed to both in plant and animal commodities, raw or processed. Since any requests for further toxicological studies are restricted as far as possible to minimise the use of animals in toxicological testing, alternative methods that refine, reduce and replace animal tests, need to be developed further in order to evaluate the toxicological profile of metabolites and degradates. For this reason, a project was initiated to evaluate the possible

contribution of computational methods, and in particular QSAR analysis, to the evaluation of the toxicological relevance of metabolites and degradates of active substances of pesticides for dietary risk assessment (JRC, 2010). This project was one of three pesticide metabolism related projects sponsored by EFSA during 2009-2010. The other two addressed the possible use of Threshold of Toxicological Concern (TTC) considerations in assessing metabolite/degradate toxicity carried out by the UK Pesticides Safety Directorate (CRD, 2010) and the impact of metabolism and degradation on pesticide toxicity, performed by the Austrian Agency for Health and Food Safety (AGES, 2010). Upon the completion of these projects, EFSA intends to pool and use the results to inform both the development of an opinion on the evaluation principles of the toxicological burden of metabolites, degradation and reaction products of active substances in food commodities, and the development of a guidance document on the establishment of the residue definition for risk assessment.

Computational (*in silico*) approaches can be used to study the mechanisms and modes of action underlying chemical toxicity and to guide, focus, design experiments and limit laboratory experimentation. These approaches differ from laboratory experiments, *in vivo* and *in vitro*, in that they do not involve the use of any biological system, but are generally built utilising biological data. They are based on scientific knowledge gained from different scientific fields and on the premise that the properties (including biological activities) of a chemical depend on its intrinsic nature, can be directly predicted from its molecular structure, and inferred from the properties of similar compounds whose activities are known. More specifically, QSARs are quantitative relationship models between the chemical structures of compounds and a given property, such as a biological mechanism or endpoint, while a Structure-Activity Relationship (SAR) is a qualitative relationship between a molecular (sub)structure and the presence or absence of a given biological activity. The term substructure refers to an atom, or group of adjacently connected atoms, in a molecule. A substructure associated with the presence of a biological activity is also called a structural alert. These models can be used to predict the property values of chemicals and to support the design of new chemicals with given property values. In toxicology, SARs and QSARs, collectively referred to as (Q)SARs, are used to predict the impacts of chemicals on human health, wildlife and the environment. These models are most often used when implemented in the form of software models.

In addition to the formalised approach of QSAR analysis, it is possible to estimate chemical properties and endpoints by using a less formalised approach, based on the grouping and comparison of chemicals. The grouping approach can be used, for example, to support the results of QSAR analysis or to generate estimated data (and fill data gaps) assuming that, in general, similar compounds will exhibit similar biological activity (ECHA, 2008; OECD, 2007).

2. CHARACTERISATION OF CHEMICAL SPACE

2.1 Introduction

The chemical space of a dataset (or inventory of chemicals) can be regarded as the ranges of physicochemical properties and structural features covered by chemicals in the dataset. It is an important piece of information in the evaluation and application of computational models for several reasons: a) a model should be applied to chemicals within the applicability domain of the model, since outside of its domain, it is unlikely to give reliable predictions; b) when the predictive performance of model is assessed by challenging it with an independent (external) test set, it is useful to compare the chemical space of the test set with that of the training set; c) when the predictive performance of a model is assessed against a limited test set, and the conclusions are generalised to a wider dataset (or chemical inventory), it is important to compare the chemical space of the test set with that of the wider inventory.

The main aim of the research investigation reported in this section was to compare the chemical space of pesticides compounds studied in this project (Section 3), including those already studied in the CRD TTC project (CRD, 2010) and some of those being studied in the AGES metabolism project (AGES, 2010), with a broader inventory of heterogenous substances including pesticides and industrial chemicals - the DSSTox Carcinogenic Potency Database (CPDB).

2.2 Compilation of datasets

2.2.1 Compilation of an "internal" pesticides dataset

The chemical space of pesticides, including compounds studied in the three EFSA-funded projects, was represented by three datasets for which chemical structures were available:

a) **CRD pesticides dataset**: initially consisting of 135 parent compounds from the CRD TTC study (CRD, 2010), including 100 parent compounds used by the CRD to develop the TTC scheme, 15 to validate it and 20 metabolites. This was reduced to 128 after removal of structures that cannot be handled by computational tools (e.g. salts, organometallics).

b) **AGES pesticides dataset**: initially consisting of 67 parent compounds from the AGES study (AGES, 2010). This was reduced to 56 compounds after removal of 11 compounds common to the above-mentioned TTC dataset.

c) **PPP pesticides database**: initially containing 821 compounds on the EU list of Plant Protection Products for which the structures were available. This was reduced to 658 compounds after removal of duplicates with the CRD and AGES datasets as well as 27 structures that cannot be handled by computational tools (e.g. salts, inorganics, organometallics).

2.2.2 Compilation of an "external" heterogeneous dataset

In addition to the above-mentioned datasets, we investigated the chemical space of an "external" dataset, representing the wider universe of pesticides and industrial chemicals. The DSSTox CPDB contains 6540 chronic, long-term animal (rats, mice, hamsters, dogs, and nonhuman primates) cancer tests on 1547 diverse chemicals (pharmaceuticals, natural chemicals in the average diet, air pollutants, food additives and pesticide residues). From the initial database, the following compounds were excluded: inorganics (60), organometallics (44), compounds for which structures were not available, macromolecules (polymers, proteins, DNA, or other large biomolecular species; and mixtures (75). As a result, 1326 single organic compounds remained, of which 735 were classified as "active", 588 as "inactive" and 3 as "unspecified". Structures that cannot be handled by computational tools were

excluded, resulting in the removal of 36 of the 1326 CPDB structures, thereby leaving 1290 chemicals in the CPDB database.

2.2.3 Compilation of an "external" dataset of classified mutagens

In addition, we investigated the chemical space of an external dataset of 104 substances that had been classified as mutagens (Muta. Category 2 R46 and Muta. Category 3 R68) during the EU harmonised classification process (the corresponding GHS classifications are Muta. 1B and Muta 2., respectively). These were derived from a list of 601 R46 and R68 substances, including 594 substances extracted from the ex-ECB Classlab database (http://ecb.jrc.ec.europa.eu/classification-labelling/) and 7 substances added by an external expert (André Muller, RIVM). From the total list, 497 substances were removed because they could not be handled by computational tools.

2.3 Investigation of chemical space by Principal Components Analysis

Principal Components Analysis (PCA) is a multivariate statistical method that is used to reduce complex multi-dimensional datasets to simpler lower dimensional datasets, while minimising the loss of information (variance in the data). Trends and patterns can be more easily identified by using the Principal Components (PCs), which are linear combinations of the original descriptors. The "meaning" of each PC can be derived from the loadings of the original descriptors on the PCs.

PCA was carried out by using the MATLAB v. R2007a software (MathWorks, http://www.mathworks.com/). The final PCA model was developed on the basis of training set including 842 pesticide compounds (CRD, AGES and PPP datasets) and 35 DRAGON descriptors (Appendix 1). DRAGON can calculate as many 3224 various molecular descriptors, many of which are difficult to interpret. For the purpose of this study, a range of easily interpretable descriptors (constitutional descriptors, functional group counts and molecular properties) were used. As a result, the chemical space was built from a combination of physicochemical properties and substructural features. The input data were pre-processed by autoscaling (the data were mean-centered and scaled to unit variance).

PCA reduced the dimensions of 35 molecular descriptors to 5 representative PCs. The five PCs described a total of 70.46% of the variance in the defined chemical space, broken down as follows: 36.10% (PC1), 12.82% (PC2), 9.95% (PC3), 6.65% (PC4) and 4.94% (PC5).

2.3.1 Chemical space of individual datasets

The chemical space covered by each dataset is illustrated in the following 3D PCA plots:

- AGES pesticides (Figure 2.1)
- CRD pesticides, including chemicals with worst mutagenicity and carcinogenicity predictions (Figure 2.2)
- AGES and CRD pesticides on the same scale (Figure 2.3)
- PPP pesticides, including some extreme chemicals (Figure 2.4)
- CRD, AGES and PPP pesticides on the same scale (Figure 2.5)
- CPDB compounds, on the same scale (Figure 2.6)

Figure 2.2 shows that pesticides which are most often incorrectly predicted for mutagenicity and carcinogenicity (see Section 3) do not fall in any particular region of the chemical space. Furthermore, they do not have extreme PCA scores. This means that correctly and incorrectly predicted pesticides cannot be distinguished on the basis of their PCA scores in these plots. This is not surprising because the chemical space was constructed to represent general molecular features and physicochemical properties of the studied chemicals, and not specifically those that may be useful in the prediction of mutagenicity and carcinogenicity.

2.3.2 Chemical space comparison - pesticides *vs* broader inventory

To compare the chemical space covered by the studied pesticides with a broader inventory of compounds (including pesticides and industrial chemicals), the PCA model with 5 PCs was applied to the CPDB dataset (1290 compounds). The following PCA plots, based on the three most influential PCs (PC1, PC2 and PC3) illustrate the overlap between the pesticides datasets and the CPDB:

- 3D PCA plot of all pesticides (CRD/AGES/PPP), classified mutagens and CPDB compounds (Figure 2.7)
- 2D biplots of all pesticides (CRD/AGES/PPP), classified mutagens and CPDB compounds (Figures 2.8-2.10)

These figures indicate that the chemical space of pesticides is overlapping with the chemical space of CPDB (which is also more diffuse). This means that CPDB is a suitable dataset for assessing the applicability of QSARs to pesticides.

These figures also show that the classified mutagens (R46 and R68 compounds) are widely scattered across the chemical space, which means they cannot be distinguished by their PCA scores alone. This is not surprising because the chemical space was constructed to represent general molecular features and physicochemical properties of the studied chemicals, and not specifically those that may be useful in the identification of classified mutagens.

2.4 Investigation of chemical space by Leadscope Substructural Analysis

In the CRD TTC study (CRD, 2010), conclusions regarding the predictivity of QSARs based on a set of 115 pesticides were generalised to all pesticides. In order to make a detailed comparison of the structural space of the CRD and AGES datasets with the broader "universe"of pesticides, as represented by the PPP dataset, the structural space of the two datasets was characterised by applying the systematic substructure analysis in the Leadscope Enterprise (v.2.4.15-6) software (http://www.leadscope.com/). Subsequently, the Leadscope substructural analysis was applied to the CPDB dataset and the set of 104 classified mutagens.

Leadscope includes a pre-defined structural fragment library of approximately 27,000 features (classes) typical of small-molecule drug candidates. The compounds in each investigated dataset were automatically broken down and categorised (indexed) according to the pre-defined structural fragment. This analysis provided information on fragment frequencies (i.e. the numbers of chemical structures indexed by a particular Leadscope structural class). Major structural classes identifed by Leadscope included: amino acids, bases and nucleosides, benzenes, carbocycles, carbohydrates, elements, functional groups, heterocycles, naphthalenes, natural products, peptidomimetics, pharmacophores, protective groups and spacer groups. More detailed information is provided in Appendix 2.

In order to characterise and compare the two datasets, the values of percentage frequency (% frequency) for each identified structural class were calculated. In each case the maximal possible frequency (100%) was equal to the amount of compounds included in the dataset. The detailed results of Leadscope systematic substructure analysis, including the frequency and % frequency values for structural classes identified among the studied datasets, are given in Table 2.1. This shows that the PPP database dataset includes a range of structural classes not represented in the CRD-AGES, but which were included in the CPDB dataset (marked in bold red in Table 2.1). Specifically, the following observations can be made:

- structural fragments not present in CRD-AGES dataset but captured in PPP: bases, nucleosides, carbocycles, aldehydes, azides, mercaptans, sulfoxides, thiocarboxamides, natural products;
- structural fragments not present in CRD-AGES but captured in CPDB: bases, nucleosides, carbocycles, acid anhydrides, acid halides, aldehydes, azides, isocyanates, mercaptan, nitroso groups, sulfonic acids, sulfoxides, thiocarboxamides, natural products;

- structural fragments not present in CRD-AGES-PPP but captured in CPDB (further indication that we extended the chemical space): acid anhydrides, acid halides, isocyanates, nitroso groups, sulfonic acids;
- an organometal was captured in the dataset of 104 classified mutagens.

2.5 Conclusions

The fact that the chemical spaces of the pesticides dataset and the CPDB dataset are overlapping supports the usefulness of CPDB when assessing the applicability of QSARs to pesticides as well as other chemicals. The results obtained for a range of QSAR models, using both the pesticides dataset and CPDB are presented in Section 3.

In the CRD TTC project (CRD, 2010), a number of conclusions were made concerning the predictive performance of Derek and Toxtree based on the application of these software tools to a dataset of 115 (mostly randomly selected) pesticides. In this evaluation, we found that the CRD dataset was broadly representative of the chemical space of the PPP inventory, but lacking in a number of structural classes. In this study, the use of a broader dataset increased the coverage of structural space, thereby providing a more extensive and robust analysis.

3. APPLICABILTY OF GENOTOXICITY & CARCINOGENICITY SOFTWARE TOOLS

3.1 Introduction

To investigate the potential applicability of QSAR analysis, the predictive performances of a range of software tools were explored for mutagenicity and rodent carcinogenicity prediction. The investigation was based on three datasets: a pesticides dataset taken from the CRD and AGES studies, a larger, structurally diverse dataset, and a dataset of mutagens classified according to EU classification criteria. The analysis revealed that, in general, software tools have a greater predictive capacity for mutagenicity than for carcinogenicity. Furthermore, several software tools could be used in the context of a TTC scheme to identify genotoxic chemicals that should either be excluded from the TTC analysis and assessed on a case-by-case basis, or subjected to a lower threshold of toxicological concern.

In addition to characterising the predictive performance of a range of software tools in statistical terms, some chemical-specific case studies were examined. In one case study, analysis of the best and worst predicted pesticides led to some suggestions for the further development of predictive tools. A second case study describes a data gap filling scenario for metabolites.

3.2 Compilation of datasets

The ability to predict genotoxicity and carcinogenicity was based on the application of the various software tools to three datasets consisting of 185 pesticides, 1290 heterogeneous chemicals, and 113 heterogeneous classified mutagens. The SMILES structures of all compounds were checked and SD files were generated using the freely available Accelrys Discovery Studio Visualizer v 2.5 software (http://accelrys.com) and subjected to further processing by the prediction tools. The compilation of each dataset is described in the following paragraphs.

3.2.1 Compilation of an "internal" pesticides dataset

To assess the carcinogenicity and mutagenicity predictions, the following datasets of pesticides were compiled:

a) **CRD pesticides dataset.** From the initial TTC list containing 100 parent compounds, two were removed - a polysaccharide (heptamaloxyglucan) and a salt (magnesium phosphide). For all of the remaining 98 compounds, genotoxicity and carcinogenicity data were available. An additional set of 15 parent compounds and 20 metabolites associated with them was also presented in the CRD TTC report. After removal of duplicates present among 133 TTC compounds (i.e. fludioxonyl, metribuzin, pirimicarb, lambda-cyhalothrin) and one organometal (metiram), the total number of 128 compounds was finally taken into account. For all of these compounds, toxicity data for at least one endpoint (genotoxicity in vitro, genotoxicity in vivo and carcinogenicity) was available.
b) **AGES pesticides dataset**. Originally, the AGES dataset included 67 parent compounds. However, since 10 were common to the TTC list, these were removed, resulting in 57 additional structures. All of them were non-genotoxic.

The total number of case study structures, including CRD (128) and AGES compounds (57), was 185. Experimental data for carcinogenicity were available for 104 compounds (45 active and 59 inactive). Information on mutagenic activity was available for 181 molecules, but only 11 of them were recognised as active (Ames positive).

3.2.2 Compilation of an "external" heterogeneous dataset

The DSSTox CPDB contains the results of cancer and Ames mutagenicity tests on 1547 diverse chemicals (pharmaceuticals, natural chemicals in the average diet, air pollutants, food additives and pesticide residues). From the initial database, the following compounds were excluded: inorganics (60), organometallics (44), compounds for which structures were not available, macromolecules (polymers, proteins, DNA, or other large biomolecular species; 3) and formulations/mixtures (75). Since computational tools cannot handle certain structures (e.g. salts), these were also excluded, resulting in the removal of 36 of the 1326 CPDB structures, thereby leaving 1290 chemicals in the CPDB database.

Carcinogenicity data were available for 1288 molecules: 717 compounds were active (i.e. carcinogenic) and 571 were inactive (i.e. non-carcinogenic). Mutagenicity data were available for 748 of the 1290 DSSTox molecules: 368 compounds were active (i.e. mutagenic) and 380 inactive (i.e. non-mutagenic).

3.2.3 Compilation of an "external" dataset of classified mutagens

To supplement the assessment of mutagenicity prediction based on the pesticides (CRD-AGES) dataset and the heterogeneous (DSSTox) dataset, we also investigated the abilities of the different software tools to correctly identify classified mutagens. This was important because the pesticides dataset was heavily biased towards non-mutagens - there were only 11 compounds having some evidence of genotoxicity in the Ames test, of which only 5 compounds (etridiazole, carbendazim, dichlorvos, thiobencarb, methyl parathion) were associated with *in vivo* test data that might result in regulatory classification. Furthermore, the larger and better balanced DSSTox dataset, while providing a good basis for assessing the ability to predict the presence and absence of Ames mutagenicity, was not suitable for assessing the ability to identify classified mutagens, since the mutagens in the DSSTox dataset had been defined on the basis of positive results in the Ames test, which is not sufficient for regulatory classification (for which evidence of *in vivo* genotoxic potential is also required).

We therefore applied the mutagenicity software tools to a dataset of 113 substances that had been classified as mutagens (Muta. Category 2 R46 and Muta. Category 3 R68) during the EU harmonised classification process (the corresponding GHS classifications are Muta. 1B and Muta 2., respectively). These were derived from a list of 601 R46 and R68 substances, including 594 substances extracted from the ex-ECB Classlab database (http://ecb.jrc.ec.europa.eu/classification-labelling/) and 7 substances added by an external expert (André Muller, RIVM). From the total list, 495 substances were removed because they could not be handled by computational tools (e.g. salts, mixtures, inorganic chemicals). The 113 classified mutagens comprised 27 substances classified as R46, and 86 substances classified as R64. The reason for using this dataset was to assess the ability of software tools to identify classified mutagens, both when used alone and in combination.

3.3 Software tools applied for genotoxicity and carcinogenicity prediction

Predictions were generated using a range of software tools, including a tool based on expert rules (Derek v.12), tools based on statistical methodologies (CAESAR, Lazar, TOPKAT v. 6.2, HazardExpert [Pallas v 3.3.2.4] and the formerly named ToxBoxes [now called ACDToxSuite]), and a hybrid tool (Toxtree v.1.60). In each case, a scheme for interpreting the model results in terms of categorical activities was adopted or devised. The predictive performances of the individual software tools were assessed and compared using the internal and external datasets (185 pesticides, 1290 DSSTox compounds, 113 classified mutagens).

The above-mentioned software tools were selected on practical grounds, taking into account the in-house availability of software as well as budgetary and procurement constraints for the acquisition of new licenses. The OECD QSAR Toolbox is becoming an increasingly used and freely available resource. However, we did not use this because it is primarily a tool for grouping chemicals and

facilitating read-across, rather than a tool which implements pre-defined QSAR algorithms. The exclusion of a given software tool from our study does not imply that it is not promising. Nevertheless, within the constraints of the project, a diverse range of methodologies (statistical and expert-based) were applied.

3.3.1 Toxtree

Toxtree is a flexible and user-friendly open-source application that places chemicals into categories and predicts various kinds of toxic effect by applying decision tree approaches. Toxtree can be downloaded from the JRC (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE) and from Sourceforge (https://sourceforge.net/projects/toxtree/)

Toxtree has been developed by the JRC in collaboration with various consultants, in particular Ideaconsult Ltd (Sofia, Bulgaria). A key feature of Toxtree is the transparent reporting of the reasoning underlying each prediction. Toxtree v 1.60 (July 2009) includes classification schemes for systemic toxicity (Cramer scheme and extended Cramer scheme), as well as mutagenicity and carcinogenicity (Benigni-Bossa rulebase and the ToxMic rulebase on the in vivo micronucleus assay). The Cramer scheme is probably the most widely used approach for structuring chemicals in order to make an estimation of the Threshold of Toxicological Concern (TTC).

The current version of Toxtree (v2.1.0, June 2010) also applies the TTC scheme of Kroes et al. (2004), alerts for skin sensitisation alerts (Enoch et al, 2008), and SMARTCyp, a two-dimensional method for the prediction of cytochrome P450-mediated metabolism (Rydberg et al, 2010). SMARTCyp predicts which sites in a molecule are labile for metabolism by Cytochromes P450.

In this study, Toxtree v 1.60 was used. The carcinogenicity/mutagenicity predictions generated by Toxtree v. 1.60 are based on a decision tree implementing the Benigni/Bossa rules (Benigni et al., 2008) and rules for the *in vivo* micronucleus assay (Benigni et al, 2010). In addition, Toxtree applies the following QSAR models to query chemicals belonging to the classes of aromatic amines or alpha,beta-unsaturated aldehydes: (i) QSAR6 - mutagenic activity of aromatic amines in the *Salmonella typhimurium* TA100 strain (Ames test); (ii) QSAR8 - carcinogenic activity of the aromatic amines in rodents (summary activity from rats and mice); (iii) QSAR 13 - mutagenic activity of alpha,beta-unsaturated aldehydes in the *Salmonella typhimurium* TA100 strain (Ames test). There are certain exceptions in the application of these QSARs, namely QSAR6 and QSAR8 in Toxtree v 1.60 apply to aromatic amines with the exclusion of aromatic amines having a sulphonic group on the same ring, and QSAR13 applies to alpha,beta-unsaturated aldehydes excluding cyclic alpha,beta-unsaturated aldehydes.

For the final assignment of genotoxicity and carcinogenicity predictions, the weight-of-evidence scheme summarised in Figure 3.1 was applied. In general, QSAR analyses provide a more refined assessment than structural alerts. The outputs of the QSARs for carcinogenicity were given more importance than the presence of structural alerts for (non)genotoxic carcinogenicity. Thus, when these QSARs gave a negative result, in spite of the presence of structural alerts, the final prediction was treated as a negative (lack of toxicity). However, in the case of genotoxicity, the *S. typhimurium* TA100 QSAR output and the structural alerts for genotoxic carcinogenicity were assigned equal weight (if either an alert or the QSAR gives a positive prediction, the overall prediction is a positive), since the outcome of a QSAR for Ames mutagenicity was regarded as an incomplete prediction of mammalian genotoxicity.

When applying Toxtree for the prediction of Ames mutagenicity (DSSTox and CRD-AGES datasets), we used only the alerts for genotoxic carcinogenicity. However, when applying it for the prediction of classified mutagens, we used both the genotoxic carcinogenicity alerts (Benigni-Bossa rulebase) and the *in vivo* micronucleus alerts (ToxMic rulebase).

The structural rules in Toxtree are based largely on expert knowledge rather than statistically derived from training sets. However, the Benigni-Bossa rulebase includes some QSARs in addition to the

structure-based rules: QSAR6 (Ames mutagenicity of aromatic amines) has 111 chemicals in its training set, QSAR8 (rodent carcinogenicity of aromatic amines) has 64 training set chemicals, and QSAR13 (Ames mutagenicity of alpha,beta-unsaturated aldehydes) has 20. Of these chemicals, very few are among the predicted chemicals in the three test sets, as shown in the following table:

Overlap between the Toxtree training sets and genotoxicity test sets

Toxtree QSAR	DSSTox overlap	CRD-AGES overlap	Classified mutagens overlap
QSAR6	0	0	0
QSAR8	1 (2,4,6-trimethylaniline)	0	0
QSAR13	1 (acrolein)	0	1 (crotonaldehyde)
Total No	2	0	1

3.3.2 CAESAR

CAESAR comprises a series of statistically-based models developed within EU-funded CAESAR project (http://www.caesar-project.eu). The models have been implemented into open-source software and made available for online use via the web. Predictions can be made for five endpoints: mutagenicity (Ames), carcinogenicity, developmental toxicity, skin sensitisation, and the bioconcentration factor.

The CAESAR prediction of mutagenicity is based on the Support Vector Machine (SVM) approach and the Kazius/Bursi database (http://www.cheminformatics.org/datasets/bursi). The SVM modelling is followed by an “expert facility” filter based on Benigni/Bossa rules, applied to the compounds presumed safe by SVM. The filter combines two sets of structural alerts with different distinguishing features: the former (the "sharp" one) has the aim to enhance the prediction accuracy attempting a precise identification of misclassified False Negatives (FN), the latter (the "suspicious" one) continues with the FN removal in such a way that this does not noticeably reduce the original prediction accuracy by generating too many False Positives (FP) as well. Compounds picked out by the first checkpoint are classified as "mutagenic" (i.e. **active**), and those picked out by the second are classified as "suspicious" (i.e. **equivocal**). Unaffected ones are finally classified as "non-mutagenic" (i.e. **inactive**).

The CAESAR mutagenicity model training set contains 3367 chemicals: 16 of these are included in the dataset of 181 predicted CRD-AGES compounds (9%); 400 are included in dataset of 748 predicted DSSTox compounds (53%); and 48 are included in the dataset of 113 classified mutagens (42%).

The prediction of carcinogenicity by CAESAR is performed on the basis of a Counter-Propagation Artificial Neural Network (CP-ANN) classification model and 805 compounds from the CPDB (http://potency.berkeley.edu/cpdb.html). The software output classifies molecules as "positive" (i.e. **active**) or "non-positive" (i.e. **inactive**), using the threshold of probability of the compounds to express activity/inactivity equal 0.5.

3.3.3 Lazar

Lazar is an open-source software programme that makes predictions of toxicological endpoints (currently, mutagenicity, human liver toxicity, rodent and hamster carcinogenicity, and Maximum Recommended Daily Dose [MRDD]) by analysing structural fragments in a training set (Helma, 2006; Maunz & Helma, 2008). It is based on the use of statistical algorithms for classification (k-nearest neighbours and kernel models) and regression (multi-linear regression and kernel models). In contrast to traditional k-Nearest Neighbour (k-NN) techniques, Lazar treats chemical similarities not in absolute values, but as toxicity dependent values, thereby capturing only those fragments that are relevant for the toxic endpoint under investigation. Lazar performs automatic applicability domain

estimation and provides a confidence index for each prediction, and is usable without expert knowledge. Lazar runs under Linux and a web-based prototype is also freely accessible (http://lazar.in-silico.de/).

The mutagenicity predictions by Lazar are based on a k-NN algorithm and two datasets: Kazius/Bursi (http://www.cheminformatics.org/datasets/bursi/) and the so-called Benchmark Data Set for In Silico Prediction of Ames Mutagenicity (http://ml.cs.tu-berlin.de/toxbenchmark/). Each prediction is associated with a prediction confidence (between 0 and 1), which gives information about the presence/absence of studied compounds within the applicability domain (AD) of the model. The developer proposed a confidence value higher than 0.025 as a reasonable hard cut-off for compounds within the AD. The accuracy of prediction decreases with the confidence value.

The training sets used to build the Lazar models comprised 4337 chemicals in the Kazius/Bursi dataset, and 6512 chemicals in the benchmark dataset. Of the 4337 chemicals in the Kazius/Bursi training set: 21 are included among the 181 predicted CRD-AGES compounds (12%); 467 are included among the 748 predicted DSSTox compounds (62%); and 58 are included among the 113 classified mutagens (51%). Of the 6512 chemicals in the benchmark training set: 29 are included among the 181 predicted CRD-AGES compounds (16%); 590 are included among the 748 predicted DSSTox compounds (79%); and 60 are in the dataset of 113 classified mutagens (53%). Despite the significant overlap between the Lazar training sets on the one hand and the DSSTox and classified mutagens datasets on the other hand, this should not affect the confidence in the Lazar mutagenicity predictions. This is because the Lazar algorithm works by building an instance-based local model that excludes the chemical being predicted from its local training set. Thus the predicted compound is never in the model training set.

3.3.4 TOPKAT

TOPKAT is a QSAR-based system, developed by Accelrys Inc. (http://accelrys.com/), makes predictions of a range of toxicological endpoints, including mutagenicity, developmental toxicity, rodent carcinogenicity, rat chronic LOAEL, rat Maximum Tolerated Dose (MTD) and rat oral LD_{50}. The QSARs are developed by regression analysis for continuous endpoints and by discriminant analysis for categorical endpoints. TOPKAT models are derived by using a range of two-dimensional molecular, electronic and spatial descriptors. TOPKAT estimates the confidence in the prediction by applying the patented Optimal Predictive Space (OPS) validation method. The OPS is TOPKAT's formulation of the model applicability domain - a unique multivariate descriptor space in which a given model is considered to applicable. Any prediction generated for a query structure outside of the OPS space is considered unreliable.

The predictions by the Weight-of-Evidence Rodent Carcinogenicity Module of the TOPKAT package are based on the scoring each query chemical as a carcinogen if: a) it is a multiple-site carcinogen in at least one sex/species combination (male or female/rat or mouse), or b) it is a single-site carcinogen in at least two sex/species combinations. TOPKAT v 6.2 provided the set of probability values, indicating if a query chemical expresses a carcinogenic activity. A computed probability below 0.3 indicates a non-carcinogen (i.e. **inactive** compound), and probability above 0.7 signifies a carcinogen (i.e. **active** compound). The probability range between 0.3 and 0.7 is the "indeterminate" zone (IND) indicating the **equivocals**.

The TOPKAT mutagenicity model was developed from compounds assayed according to the US EPA GeneTox protocol (i.e. tested against five strains of Salmonella typhimurium using the Histidine Reversion Assay). A chemical is labelled a mutagen if a positive response is observed against one or more strains. A chemical is considered a non-mutagen if a negative response is observed in all of these five bacterial strains. Therefore, when a query structure is assessed by TOPKAT to be a non-mutagen (computed probability of mutagenicity between 0.0 and 0.3), it indicates that there is a high probability of the query chemical producing a negative response in the Histidine Reversion Assay against all of the five bacterial strains. It is important to note that a non-mutagen assessment by TOPKAT does not

mean that the query chemical will be a non-mutagen in other mutagenicity tests, such as the micronucleus and Chinese Hamster Ovary tests. As suggested by the vendor, probability values can be converted into binomial ones (actives or inactives) according to the following rules:

(i) if computed probability of mutagenicity greater than 0.7, then the compound is considered to be a mutagen (i.e. **active**);

(ii) if computed probability of mutagenicity smaller than 0.3, then the compound is considered to be a non-mutagen (i.e. **inactive**);

(iii) if computed probability of mutagenicity between 0.3 and 0.7, then the prediction is **equivocal**.

Although the total number of compounds in the TOPKAT training set is not known, it is possible to check whether the query compunds are in the TOPKAT database: 14 are included in the dataset of 181 predicted CRD-AGES compounds (8%); 226 are included in dataset of 748 predicted DSSTox compounds (30%); and 43 are included in the dataset of 113 classified mutagens (38%).

3.3.5 HazardExpert

HazardExpert is a module of the Pallas software developed by CompuDrug (http://compudrug.com/). It predicts the toxicity of organic compounds based on toxic fragments, and it also calculates bioavailability parameters (logP and pKa). It is a rule-based system with an open knowledge base, allowing the user to expand or modify the data on which the toxicity estimation relies. It covers the following endpoints relevant to dietary toxicity assessment: carcinogenicity, mutagenicity, teratogenicity, membrane irritation, immunotoxicity and neurotoxicity

The results of oncogenicity and mutagenicity predictions by HazardExpert (Pallas v 3.3.2.4) are provided as relative percentage toxicity values. On the basis of the ranges of the results the authors proposed the classification of chemicals as “highly probable”, “probable”, “uncertain” and “not probable” to express oncogenic/mutagenic activity. In order to compare the HazardExpert predictions with the results of other software tools we treated “highly probable” and “probable” chemicals as **active,** “uncertain” chemicals as **equivocal,** and “not probable” ones as not active, as in the following table.

Interpretation of HazardExpert oncogenicity/mutagenicity predictions

The range of relative percentage toxicity [%]	Toxic Class	Classification	Interpretation of the results
100-60	1	Highly probable	active
59-48	2A	Probable	active
47-36	2B	Probable	active
35-3	3	Uncertain	equivocal
2-0	4	Not probable	not active

The HazardExpert training sets are not available, so it was not possible to check their overlap (in terms of common chemicals) with the three test sets (pesticides, DSSTox, classified mutagens).

3.3.6 Derek

Derek for Windows (DfW) is a SAR-based system is developed by Lhasa Ltd, a non-profit company and educational charity (https://www.lhasalimited.org/). DfW contains over 50 alerts covering a wide range of toxicological endpoints in humans, other mammals and bacteria. An alert consists of a toxicophore (a substructure known or thought to be responsible for the toxicity) and is associated with

literature references, comments and examples. A key feature of DfW is the transparent reporting of the reasoning underlying each prediction.

All the rules in DfW are based either on hypotheses relating to mechanisms of action of a chemical class or on observed empirical relationships (Sanderson & Earnshaw, 1991). Information used in the development of rules includes published data and suggestions from toxicological experts in industry, regulatory bodies and academia. The toxicity predictions are the result of two processes. The program first checks whether any alerts in the knowledge base match toxicophores in the query structure. The reasoning engine then assesses the likelihood of a structure being toxic. There are nine levels of confidence: certain, probable, plausible, equivocal, doubted, improbably, impossible, open, and contradicted. DfW can be integrated with Lhasa's Meteor software, which makes predictions of fate, thereby providing predictions of toxicity for both parent compounds and their metabolites.

DfW predictions are knowledge-based, based on the application of alerts and reasoning rules. The final toxicity assessment is a result of a two-part process: (i) the program checks whether any alerts from the knowledge base appear in the query compounds, and (ii) the reasoning model is applied in order to determine the likelihood of the compound's toxicity (expressed as the level of likelihood). If no alerts from the knowledge base can be matched against query structure, the program displays a message "Nothing to report".

Genotoxicity alerts in Derek include alerts for mutagenicity (in bacteria and mammals) and alerts for chromosome damage based on the in vitro chromosomal aberration assay and including effects that do not involve direct DNA damage (inhibition of DNA synthesis/repair, spindle function disruption, reactive oxygen species generation, energy depletion, thiol reactivity, intercalation).

When applying DfW for the prediction of Ames mutagenicity (DSSTox and CRD-AGES datasets), we used only the alerts for bacterial mutagenicity. However, when applying it for the prediction of classified mutagens, we used all of the genotoxicity alerts (mutagenicity and chromosomal aberration) to better reflect the basis for regulatory classification. In order to make the results from DfW comparable with other results, we converted the output into three categories: **active**, **equivocal** and **not active**, as in the following table.

Interpretation of Derek toxicity predictions

Level of likelihood	Interpretation of the results
Certain	active
Probable	active
Plausible	active
Equivocal	equivocal
Doubted	not active
Improbable	not active
Impossible	not active
Open	not active
Contradicted	not active
Nothing to report	not active

3.3.7 ToxBoxes

ToxBoxes (now called ACD/Tox Suite), marketed by ACD/Labs and Pharma Algorithms, provides predictions of various toxicity endpoints including hERG inhibition, genotoxicity, CYP3A4 inhibition, ER binding affinity, irritation, rodent LD50, aquatic toxicity, and organ-specific health effects (http://www.acdlabs.com/products/admet/tox/). The predictions are associated with confidence intervals and probabilities, thereby providing a numerical expression of prediction reliability. The software incorporates the ability to identify and visualize specific structural toxicophores, giving insight as to which parts of the molecule are responsible for the toxic effect. It also identifies analogues from its training set, which can also increase confidence in the prediction. Predictions are based on data from over 100,000 compounds. The algorithms and datasets are not disclosed.

The predictions of genotoxicity by ToxBoxes are based on the probability of query compounds to be genotoxic in Ames test. The training data used in the software originate from Chemical Carcinogenesis Research Information (CCRIS) and Genetic Toxicology Data Bank (GENE-TOX), containing the results of Ames genotoxicity assays for several strains of *S. typhimurium* (TA97, TA98, TA100, TA102, TA104, TA1535, TA1537, TA1538 and also E. coli strain WP2 uvrA), with or without metabolic activation. In establishing this training set, a compound was considered genotoxic if at least one of Ames results was positive; otherwise, the compound was considered non-genotoxic. In case of inconsistent results from different assays, the data were evaluated by experts and in some cases had been labelled as inconclusive. The final training set exceeded 8000 compounds with standardised Ames genotoxicity values. The neural network model was built using structural fragments as descriptors. Molecules were decomposed into atomic and chain-based fragments (chains of interconnected atoms). Atomic fragments and chains, containing 2 to 5 atoms, present in at least 10 training set molecules were utilized to develop the model. The model makes a prediction if the chemical structure is more than 75% covered by fragments in the training set. For each compound, the "probability of positive Ames test" and the "Ames test reliability index" are provided.

The method suggested by the vendor was adopted to convert the probability values into binomial ones (actives or inactives) according to the following rules:

(i) if the "Probability of positive Ames test" is bigger than 0.7, then the compound is a predicted mutagen (i.e. **active**);

(ii) if the "Probability of positive Ames test" is smaller than 0.3, then the compound is a predicted non-mutagen (i.e. **inactive**);

(iii) if the "Probability of positive Ames test" is between 0.7 and 0.3, then the result is predicted as **equivocal**.

While the ToxBoxes training set is accessible via its database, it is not practical to verify the overlap with the test sets, since the database can only be checked chemical by chemical, and cannot be extracted or searched in an automated manner.

When performing the analysis with ToxBoxes, we noticed that to avoid some errors in toxicity prediction, it was better to encode aromaticity as alternate single and double bounds in the input file, rather than representing aromatic rings with delocalized electrons.

3.4 Predictive performances of the models for genotoxicity

For each model, categorical predictions were compared with the assessments of genotoxic potential, as taken from the CRD and AGES reports, or the assessment of Ames mutagenicity taken from DSSTox. Thus, the original interpretations of the experimental data were retained. This led to the calculation of the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), from which the following statistics were calculated:

Statistics for classification models

Statistic	Definition	Meaning (proportion/percentage) of …
Sensitivity	= TP / (TP+FN)	known positives that are correctly predicted
Specificity	= TN / (TN+FP)	known negatives that are correctly predicted
False positive rate	= 1-specificity = FP/(TN+FP)	known negatives that are incorrectly predicted as positive
False negative rate	= 1-sensitivity = FN/ (TP+FN)	known positives that are incorrectly predicted as negative
Positive predictivity	= TP / (TP+FP)	positive predictions that are true positives (probability of a positive prediction being correct)
Negative predictivity	= TN / (TN+FN)	negative predictions that are true negatives (probability of a negative prediction being correct)

The statistics for genotoxicity prediction are given in Tables 3.1 (pesticides) and Table 3.2 (DSSTox), and for carcinogenicity prediction in Table 3.3 (pesticides) and Table 3.4 (DSSTox).

To visualise the relationship between sensitivity and false positive rate, Receiver Operating Characteristic (ROC) curves were plotted. In a ROC curve, a model on the diagonal is a poor model, having predictions no better than chance, whereas a model located in the top left corner is the ideal model, having a perfect (100%) prediction of positives and a perfect (0%) false positive rate. Typically, the ability to predict positives is made at the expense of the false positive rate.

When interpreting these statistics in relation to a test set of chemicals, it should be remembered that most statistical methods contain a definite but variable proportion of the test chemicals in their training sets. In extreme cases (majority of test chemicals are in the training set), this could mean that the statistics are not so much measures of predictivity as measures of goodness-of-fit. In less extreme cases (minority of test chemicals are in the training set), this may affect the weight given to the statistics when comparing models. Therefore, wherever practically possible, the extent of overlap between the test and training sets was checked.

3.4.1 Genotoxicity prediction results: CRD-AGES pesticides dataset

Based on the pesticides dataset, the majority of software tools had sensitivities between 45% and 64% (Table 3.1 and Figure 3.2). Surprisingly, a perfectly high sensitivity (100%) was obtained for ToxBoxes. In contrast, a surprisingly low sensitivity of 45% was observed for the Toxbenchmark-based Lazar model. However, these results are misleading and are due to the fact that there were only 11 active pesticides in the dataset. In the case of ToxBoxes, only 4 active compounds were correctly predicted, and the remaining 7 molecules were considered equivocal. Thus, the results based on the ability to identify known positives (sensitivity and false negative rate) should not be used to draw general conclusions.

In contrast, the results based on the ability to identify known negatives (specificity and false positive rate) should be more robust, since there were 170 inactive compounds. The highest specificity was obtained for Derek (87%) followed by ToxBoxes (84%), Lazar (76%, for both training sets) and CAESAR (76%). The high specificity of Derek is due to the fact that Derek has “nothing to report” for substances which are not recognized by its alerts. In the case of Toxboxes, the high specificity is at least partly due to the high percentage of substances that are considered equivocal (43/181=24%).

In addition to these statistics, we examined the ability of each software tool to correctly identify genotoxic chemicals in the CRD-AGES pesticides dataset. These true positives are listed in Table 3.5. The same information, including chemical structures, is given in Appendix 3.

3.4.2 Genotoxicity prediction results: DSSTox heterogeneous chemicals dataset

On the basis of the larger and better balanced DSSTox dataset, the assessment should be more robust (Table 3.2 and Figure 3.4). The statistical software tools (CAESAR, TOPKAT,) as well as knowledge-based ones (Derek, Toxtree) showed similar sensitivities between 83-86%, which are acceptable values in the range of experimental error. ToxBoxes showed the highest sensitivity of 93%, whereas Lazar showed the lowest sensitivity (66% or 67%, depending on the training set, Toxbenchmark or Kazius/Bursi). The apparent superior ability to predict positives (in percentage terms) of ToxBoxes is partly due to the higher number of compounds classified as equivocal compared with the other software tools (thus fewer chemicals were included in the percentage calculation) and partly to its reliance on a more extensive training set. This is not to say that the result should be disregarded – in fact, it can be argued that the generation of fewer but more reliable results is an advantage, especially if the model is being used to identify positives in a stepwise assessment strategy. The highest specificity was identified for ToxBoxes (93%), corresponding to the lowest false positive rate of 7%, and the lowest specificity was found for HazardExpert (61%), corresponding to the highest false positive rate of 39%. The other software tools showed specificities from 70% to 86%.

The above-mentioned statistics are sometimes referred to as “producer statistics” – they focus on the characteristics of the model. It is also informative to look at so-called “user statistics” (Tables 3.6 and 3.7) which focus on the characteristics of the predictions. The positive predictivity gives an indication of the probability that a positive prediction is correct. For genotoxicity prediction, this ranges from 69% (HazardExpert) to 93% (ToxBoxes). The negative predictivity, which gives an indication of the probability that a negative prediction is correct, ranges from 71% (Lazar) to 93% (ToxBoxes). Thus the positive and negative predictivities fall in a similar range.

In general, ToxBoxes seems to have the highest predictive performance, in terms of its ability to identify positives and negatives. However, it does not necessarily follow that ToxBoxes should be the preferred tool for genotoxicity prediction. It is not known what percentage of these accurately predicted chemicals are in the ToxBoxes training set, so the value of 93% may be misleadingly high in comparison with the other tools evaluated. The tool with the next highest sensitivity is CAESAR (sensitivity of 86%), although it should be noted that 53% of the compounds predicted are in the CAESAR model training set. Finally, some users may place more confidence in knowledge-based predictions (from Derek or Toxtree), or in a combination of statistically-based and knowledge-based predictions.

3.4.3 Genotoxicity prediction results: classified mutagens dataset

The abilities of the different software tools to identify classified mutagens is summarised in Table 3.8. In the case of this dataset of 113 mutagenic chemicals, it was possible to calculate the number of true positives, the sensitivity, the number of false negatives, and the false negative rate. It was not possible to calculate the number of false positives, the false positive rate and the specificity since there were no negatives in the data set.

The results show that the highest sensitivity (87%) and thus the lowest false negative rate (13%) was obtained for Toxtree (using the in vivo micronucleus rulebase), followed by HazardExpert (sensitivity 77%; false negative rate 23%). In the case of Toxtree, a detailed examination of the predictions shows that the *in vivo* micronucleus assay generates a positive prediction for all positive predictions generated by the Benigni-Bossa rulebase, as well as a few additional ones. This is expected since the *in vivo* micronucleus rulebase consists of the Benigni-Bossa alerts, with the exclusion of the alerts specific for non-genotoxic carcinogenicity, and with the inclusion of five additional alerts associated with the generation of micronuclei *in vivo*. In other words, it is sufficient to use the *in vivo* micronucleus

rulebase when predicting *in vivo* genotoxicity in Toxtree. These good results are in line with the fact that most classifications for mutagenicity are based on positive micronucleus tests (A. Muller, personal communication).

The lowest sensitivity (50%) and thus the highest false negative rate (50%) was observed for Lazar Toxbenchmark, although this is reduced to 39% if the two Lazar models are used in combination (if either model gives a positive prediction, the overall prediction is considered positive). This is in agreement to the results previously obtained to predict Ames mutagenicity (Table 3.2). The next highest false negative rate is associated with TOPKAT (42%).

To explore the possibility of further reducing the false negative rate, various two-software combinations were considered (Table 3.8). In these combinations, if either tool gives a positive prediction, then the overall prediction is considered positive. Not surprisingly, by combining the use of two software tools his way it is possible to reduce the false negative rate (the lowest false negative rate of 8% being for the combined use of Toxtree and Derek).

By examining the individual predictions for the 113 substances, it was found that 30 classified mutagens were correctly identified by all seven software tools (Toxtree, TOPKAT, CAESAR, HazardExpert, Lazar, Derek, ToxBoxes), and that 7 classified mutagens were incorrectly predicted as false negatives. These "best" and "worst" predicted substances are listed in Appendix 4, which indicates, for each mutagen, the number of software tools that correctly detect the mutagenic potential.

3.4.4 Carcinogenicity prediction results

Based on the pesticides dataset, the points representing sensitivity/specificity of six software tools used for calculations are all situated near the diagonal of the ROC plot (Table 3.3 and Figure 3.3). TOPKAT had the highest sensitivity (58%) and thus the lowest false negative rate (42%), but this was associated with the lowest specificity (53%) and thus the highest false positive rate (47%). In contrast, the lowest sensitivity (31%) was observed for CAESAR, which turned out to have a specificity of 68%. The highest specificity was observed for Lazar (84%). Overall, the relatively poor ability of the software tools to identify carcinogenic pesticides could be attributed to the fact that carcinogenic pesticides are less well represented in the model training sets than other types of carcinogenic chemicals. Furthermore, since genotoxic carcinogens are probably never used as pesticides, it is likely that most carcinogenic pesticides are non-genotoxic. These mechanisms of carcinogenicity are less well accounted for by current QSAR models.

On the basis of the larger DSSTox dataset, the assessment should be more robust (Table 3.4 and Figure 3.5). The sensitivities of TOPKAT, CAESAR, Derek, and Toxtree are comparable and vary from 67-71%, corresponding to false negative rates of 29-33%. The lowest sensitivity (38%) was observed for Lazar, corresponding to a false negative rate of 62%, although this is associated with the highest specificity (88%) and thus the lowest false positive rate (12%). For the other software tools, the specificity was between 53% (HazardExpert) and 70% (CAESAR), corresponding to false positive rates from 30-47%.

The positive predictivity (Table 3.7) ranged from 66% (HazardExpert) to 80% (Lazar), whereas the negative predictivity ranged from 53% (Lazar) to 64% (CAESAR and Toxtree). Thus, the overall ability to identify carcinogens is better than the ability to identify non-carcinogens.

These statistics should be considered collectively, but some may be more important than others, depending on the regulatory context. The statistics imply that Lazar is the best tool for identifying non-carcinogens (highest specificity). However this is at the expense of identifying carcinogens (lowest sensitivity). As a screening tool, Lazar would therefore be the least effective tool for identifying and filtering carcinogens. However, when Lazar predicts a chemical to be carcinogenic, it is more likely to be correct compared with other tools (highest positive predictivity). Thus, it is crucial to consider whether a tool is being used as a screen, or as a means of directly filling a data gap.

3.5 Case study 1: best and worst predicted chemicals

The dataset of pesticides (CRD/AGES) was used to identify those chemicals which are most often correctly and falsely predicted, both for mutagenicity and carcinogencity (Appendix 5). In the regulatory assessment of pesticides, particular interest should focus on chemicals that are incorrectly predicted, and in particular the false negative rate should be minimised, since false negatives may not receive adequate health protection measures.

3.5.1 Best and worst predictions for genotoxicity

Best predicted chemicals: only one chemical (parathion-methyl) is correctly predicted as mutagenic by all software (except ToxBoxes which gave an equivocal prediction). In addition, 34 chemicals are correctly predicted as non-mutagenic by all software tools (Appendix 5).

Worst predicted chemicals: the worst predicted chemical is sodium nitroguiacolate, a non-mutagen which is falsely predicted to be mutagenic by all software tools. This false positive may be related to the aromatic nitro moiety (the alert triggered in Derek is "aromatic nitro compound"). The prediction of another chemical (metconazole) is difficult to evaluate, since interpretation of the underlying experimental data is variable. In the CRD report, this chemical is treated as an *in vitro* genotoxicant (clastogen); in the AGES report it is treated as a non-mutagen.

3.5.2 Best and worst predictions for carcinogenicity

Best predicted chemicals: analysis of the internal (CRD-AGES) pesticides dataset shows that no single carcinogenic chemical is correctly identified by all software tools. Three carcinogens (chlorothalonil, amitrole, diuron) belonging to different chemical classes, are correctly predicted by all 6 software models. In addition, 5 non-carcinogenic chemicals carbosulfan, methiocarb, prohexadione calcium, boscalid, thiodicarb methomyl) are correctly predicted by all 6 models.

Worst predicted chemicals: three carcinogenic chemicals (forchlorfenuron, tebufenpyrad, thiodicarb) are wrongly predicted as non-carcinogenic by all 6 models. Such false negatives would represent the most concern in an assessment strategy based on the application of QSARs in a TTC scheme. It is recommended that in-depth investigations are performed for these chemicals to understand the reasons for false prediction. It would be desirable to analyse data concerning their ADME characteristics (e.g. metabolism and solubility), and examine how the carcinogenicity conclusions were drawn from the underlying experimental data. In the case of models based on expert knowledge, these false negatives are probably related to the absence if an appropriate structural alert. In the case of models based on statistical methodology, it is more difficult to rationalise these false negatives – it could be due to the absence of close analogues in the training set, incompleteness in the descriptor set, or the choice of statistical methodology. Finally, one known non-carcinogen (bifenox) is wrongly predicted by the software models as active (false positive).

3.6 Case study 2: filling data gaps for pesticide metabolites

Appendix 8 summarises the mutagenicity and carcinogenicity predictions for a series of parent compounds (fenamidone, fludioxonil, bitertanol, dimethoate, metconazole, pirimicarb, proquinazid, spirotetramat, thiodicarb) and their metabolites contained in the CRD-AGES pesticide dataset together with the experimental outcomes extracted from the CRD and AGES reports. In cases where experimental data are available for both parent compounds and their metabolites, the metabolites have the same toxicity (or absence of toxicity) as the parent, except for thiodicarb which is a mutagenic/carcinogenic parent with a non-mutagenic/non-carcinogenic metabolite. However, in most cases, data are missing for the metabolites. The table therefore represents a realistic scenario where risk assessment decisions based on QSAR predictions might be needed in the absence of experimental data. In such cases, it is recommended to check whether the toxicity of the parent compound is correctly predicted by one or more software tools, since the correct prediction of parent compound

toxicity will increase the confidence that the toxicities of metabolites are correctly predicted as well. In addition, other analogues for which data are available could be sought for additional confirmation. In the case of statistical models, it is useful to know whether such analogues (including the parent compound) are in the training set. However, their presence in the training set does not undermine the argumentation. On the contrary, if close analogues are present in the training set, this increases the confidence that the chemical of interest is in the applicability domain of the model. In the case of the nine parent compounds given in Appendix 6, only dimethoate was found in any of the model training sets (for the Caesar, Topkat, and Lazar mutagenicity and carcinogenicity models, as well as the ToxBoxes mutagenicity model).

Concerning the mutagenicity prediction of parents and metabolites, it can be seen that ToxBoxes either gives the correct answer for compounds in Appendix 6 (which is consistent with the statistics reported above) or it gives an equivocal outcome, indicating that not all the fragments present in the submitted compound are covered by the ToxBoxes training set. In addition, Derek gave consistently accurate results for mutagenicity with the only exception of the non-mutagen dimethoate, which is a false positive.

For carcinogenicity, most of the metabolites have missing experimental values. For instance, in the case of fludioxonil, all software tools predict metabolite 1 (CGA192155) and metabolite 2 (CGA308103) as non-carcinogenic and just one software (Topkat) gives a carcinogenicity alert for metabolite 3 (CGA339833). These metabolites could be predicted as non-carcinogens. The same conclusion could be applied to the metabolite of bitertanol (triazolylalanine), where only two software tools (Topkat and Lazar) falsely predict the parent to be carcinogenic. In the case of spirotetramat, for which five metabolites were considered, the carcinogenicity prediction of the parent compound was predicted correctly by all the software tools except Toxtree. All the metabolites were predicted as non-carcinogens by almost all software tools. Only metabolite1 had a positive carcinogenicity prediction by two software tools (Topkat and Caesar). Thus, in this case, it is recommended not to rely on the results of QSAR analysis alone – a more detailed experimental analysis for this metabolite might be needed.

3.7 Conclusions and recommendations

The results described in this Section provide a statistical characterisation of the overall ability of a range of software tools to predict genotoxicity (including mutagencity) and carcinogenicity. This general statistical analysis is supplemented with a number of chemical-specific case studies (best and worst predicted chemicals, and a data gap filling scenario). The statistical analyses show that the overall concordance in predicting Ames mutagenicity (Table 3.2) ranges from reasonable (71%) to high (93%), whereas the overall concordance in predicting rodent carcinogenicity (Table 3.4) ranges from poor (49%) to modest (69%), with a greater ability to identify carcinogens than non-carcinogens (Table 3.7).

The usefulness of a model, and in particular the adequacy of a model prediction, can only be considered in the context of the specific application, including the regulatory purpose, in which the prediction is being used (e.g. in a weight-of-evidence assessment with experimental data) and the consequence of being wrong. During the course of this project, based on the outcome of the CRD TTC project, an EFSA working group concluded that in pesticide risk assessment, QSAR analysis should be applied in the context of a TTC scheme, in order to enhance the high degree of protectiveness already present. The TTC is a generic human exposure level for chemicals below which there is low probability of risk to human health, assuming lifetime exposure. The principle of TTC is built on the premise that a safe level of exposure can be identified for chemicals present at low concentrations in the diet, even for those with unknown toxicity, on the basis of their chemical structure (Barlow, 2005).

In other words, QSARs are not considered here as standalone methods to directly fill data gaps in hazard assessment. Instead, they are being used to identify particular health concerns that may warrant specific thresholds of toxicological concern. In the TTC scheme by Kroes *et al*. (2004), and in the subsequent modifications by Munro *et al*. (2008) and Felter *et al*. (2009), there are three Cramer

classes (I - low, II- moderate, and III-high) for different levels of non-cancer life-time risk, corresponding to threshold doses of 1800, 540 and 90 μg/day/person, respectively. In the case of chemicals containing a structural alert for potential genotoxicity, a lower TTC of 0.15 μg/day is applied. These schemes refer to structural alerts, although the prediction of potential genotoxicity by a QSAR is presumably equivalent. However, it is unclear what is meant by an alert or QSAR for potential genotoxicity – should this be any genotoxic effect (e.g. Ames mutagenicity) or should it be limited to effects that are strong enough to warrant regulatory classification? In this study we analysed the predictive abilities of various software tools, both in terms of their ability to predict Ames mutagenicity as well as their ability to identify classified mutagens. Another open question is where structural alerts and QSARs for carcinogenicity fit in such TTC schemes. Presumably, such models would also be used to trigger a TTC of 0.15 μg/day, especially since most models for potential carcinogenicity are effectively modelling DNA reactivity (like models for potential genotoxicity). However, some models (e.g. Toxtree Benigni-Bossa) make predictions of non-genotoxic carcinogenicity. Non-genotoxic carcinogens which also have the potential to bioaccumulate are typically excluded from TTC schemes. Furthermore, high potency carcinogens (e.g. aflatoxin-like, azoxy and N-nitroso compounds) are also excluded, not as a matter of principle, but because there has been insufficient analysis of their potency distributions on the basis of existing TTC databases.

Another largely unexplored question is how to combine the use of QSAR analysis with available *in vitro* and *in vivo* data. For example, Felter *et al*. (2009) have proposed that the presence of negative Ames data should overrule concern based on structural alerts for genotoxicity. This is reasonable to the extent that alerts for genotoxicity are modelling bacterial mutagenicity. However, it does not necessarily follow that a negative Ames test result should also overrule a positive prediction of *in vivo* mutagenicity. In this case, the two pieces of information (predicted *in vivo* mutagenicity and experimentally determined *in vitro* mutagenicity) are not strictly comparable, because the underlying mechanism could be different in each case (for example, point mutations versus micronuclei formation). However, an open question is whether a positive prediction of *in vivo* mutagenicity for a given endpoint (*in vivo* micronucleus) should overrule a negative *in vitro* test result for the same endpoint (*in vitro* micronucleus).

The crucial question in the application of QSAR is whether any model, or combination of models, is "good enough" for the regulatory purpose (in this case the identification of potential genotoxins). This cannot be answered in the absence of clearly defined performance criteria, and these should be set by the risk assessor and risk manager. In conducting this study, it was considered premature to define clear acceptance criteria. However, EFSA indicated that for the purpose of pesticide risk assessment, the most important criterion is minimisation of the false negative rate (even though an acceptable threshold value was not defined). The generation of false negative predictions could not be adequately assessed from the CRD-AGES pesticides dataset, due to the low number of known positives. However, more robust false negative rates could be established by using the DSSTox/CPDB dataset and the classified mutagens dataset. In the case of carcinogenicity prediction, the false negative rate ranges from 29-62%, depending on the software tool (Table 3.4). This could probably be reduced by combining the use of two or more carcinogenicity models. In the case of genotoxicity prediction, the generation of false negatives ranges from 7-34% when predicting Ames mutagenicity (Table 3.2) and from 13-56% when predicting mutagenic effects (Table 3.8) that are sufficient to result in a regulatory classification (risk phrases R46 or R68 in the EU classification system).

When assessing the predictive performance of a (software) model based on global statistics, it is important to assess, if possible, the extent of overlap between the model training set and the test set to which the model is being applied. At one extreme, all of the test chemicals are in the training set, in which case the statistics reflect goodness-of-fit rather than predictivity. At the other extreme, none of the test chemicals are in the training set, which means that the statistics are a more realistic reflection of the ability of the software to predict "unknown" chemicals. It is therefore recommended, wherever possible, to check the extent of overlap between a test set and the training sets of models assessed against the test set. However, this consideration is not applicable to knowledge-based models (such as

Derek and Toxtree), in which the rules are based on expert knowledge and it is not applicable to certain types of statistical model, such as Lazar, which makes predictions using instance-based models (in which nearest neighbours are used for the prediction, but the chemical of interest is excluded). The extent to which the predicted datasets were in the model training sets is described above (Section 3.3 and Table 3.8). In particular, 30% (and 38%) of the chemicals in the DSSTox (and classified mutagens) test sets are present in the TOPKAT database; whereas 53% (and 42%) of the chemicals in the DSSTox (and classified mutagens) test sets are present in the CAESAR training set. This does not mean that these statistics should be disregarded or downplayed – it simply means that caution should be applied when comparing statistics between models. Such comparisons are further complicated by the fact the some models make indeterminate predictions (e.g ToxBoxes), and in some cases, the extent of overlap between training and test sets cannot be verified (e.g. ToxBoxes, HazardExpert). In short, global statistics can be helpful in assessing and comparing models, but they should not be the only consideration.

In the absence of clear acceptance criteria, it is not possible to draw firm conclusions about which software tool would be most useful. However, if it is assumed for the purposes of illustration that the false negative rate in the identification of classified mutagens should not exceed 20%, then only one tool would qualify for standalone use (the Toxtree *in vivo* micronucleus rulebase). In addition, various two-software combinations for genotoxicity prediction would also qualify (Table 3.8). The user could decide which is the most appropriate or convenient combination; for example, the combined use of Toxtree and CAESAR gives a false negative rate of 11%. This combination might be the preferred one since it combines two freely available tools, of which one (CAESAR) is statistically based and the other knowledge-based (Toxtree). The combined use of Lazar and Toxtree would be equivalent (false negative rate of 10%). The lowest false positive rate was given by the combined use of Toxtree (*in vivo* micronucleus rulebase) and Derek (false negative rate of 8%).

Although not explored in this study, it is noted that the way in which the estimated data are interpreted in terms of positives and negatives can also be varied for each software tool (in this study we followed the recommendation of the developer/supplier). For example, the probability cut-off values for positive and negative predictions can be adapted when using TOPKAT, CAESAR and ToxBoxes. To minimise the false negative rate of a given model, it is necessary to increase its sensitivity (e.g. by lowering the cut-off value between negatives and positives, so that the model "captures" more positives). However, this most likely will be accompanied by a decrease in the specificity and thus an increase in the false positive rate. For individual statistically-based models, there is a trade-off between sensitivity and specificity. The trade-off can in principle be avoided by combining the use of multiple models (including a high-sensitivity model and a high-specificity model). However, finding the optimal combination of models (against a set of pre-defined criteria) would require further research.

To further develop a transparent and scientifically robust basis for applying QSAR analysis in the context of the TTC approach, the following actions are recommended:

1) the role of carcinogenicity prediction models in TTC schemes should be clarified, including models for non-genotoxic carcinogenicity

2) the definition of "potential genotoxicity", which trigger a lower TTC threshold, should be clarified (any genotoxic effect, or a weight-of-evidence genotoxicity meeting more stringent criteria, such as regulatory classification criteria)

3) EFSA should establish, in discussion with its risk assessors and managers, a set of statistical acceptance criteria for the use of QSAR models or combinations of QSAR models in defined regulatory contexts.

4) models for genotoxicity and carcinogenicity, including combinations of models, should be further developed and assessed in the light of the established acceptance criteria.

5) studies should be carried out to develop proposals for the combined use of QSARs, structural alerts and available *in vitro* genotoxicity data.

Table 2.1. Leadscope systematic substructure analysis

Structure Set	Frequency						% Frequency					
	AGES	CRD	CRD-AGES	PPP	CPDB	Classified mutagens	AGES	CRD	CRD-AGES	PPP	CPDB	Classified mutagens
Amino acids	0	8	8	23	51	1	0.00	6.25	4.35	3.50	3.95	0.96
Bases, nucleosides	**0**	**0**	**0**	**2**	**16**	**2**	**0.00**	**0.00**	**0.00**	**0.30**	**1.24**	**1.92**
Benzenes	43	93	136	377	642	62	76.79	72.66	73.91	57.29	49.77	59.62
Carbocycles	**0**	**0**	**0**	**14**	**26**	**0**	**0.00**	**0.00**	**0.00**	**2.13**	**2.02**	**0.00**
Carbohydrates	0	1	1	5	21	0	0.00	0.78	0.54	0.76	1.63	0.00
Elements	1	10	11	83	33	5	1.79	7.81	5.98	12.61	2.56	4.81
Functional groups	56	128	184	633	1263	97	100.00	100.00	100.00	96.20	97.91	93.27
acid anhydride	**0**	**0**	**0**	**0**	**2**	**0**	**0.00**	**0.00**	**0.00**	**0.00**	**0.16**	**0.00**
acid halide	**0**	**0**	**0**	**0**	**1**	**0**	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.00**
alcohol	9	20	29	93	293	16	16.07	15.63	15.76	14.13	22.71	15.38
aldehyde	**0**	**0**	**0**	**10**	**18**	**2**	**0.00**	**0.00**	**0.00**	**1.52**	**1.40**	**1.92**
alkene	4	23	27	145	232	17	7.14	17.97	14.67	22.04	17.98	16.35
alkyne	0	2	2	9	7	0	0.00	1.56	1.09	1.37	0.54	0.00
amidine	2	2	4	11	5	0	3.57	1.56	2.17	1.67	0.39	0.00
amines	22	56	78	153	443	35	39.29	43.75	42.39	23.25	34.34	33.65
azide	**0**	**0**	**0**	**1**	**1**	**0**	**0.00**	**0.00**	**0.00**	**0.15**	**0.08**	**0.00**
carbamate	0	13	13	26	26	3	0.00	10.16	7.07	3.95	2.02	2.88
carbonyl	28	99	127	322	548	25	50.00	77.34	69.02	48.94	42.48	24.04
carboxamide	10	36	46	59	134	12	17.86	28.13	25.00	8.97	10.39	11.54
carboxylate	8	17	25	109	111	6	14.29	13.28	13.59	16.57	8.60	5.77
carboxylic acid	2	13	15	90	119	0	3.57	10.16	8.15	13.68	9.22	0.00
ether	31	55	86	168	229	31	55.36	42.97	46.74	25.53	17.75	29.81
guanidine	3	0	3	11	13	0	5.36	0.00	1.63	1.67	1.01	0.00
halide	42	64	106	264	290	16	75.00	50.00	57.61	40.12	22.48	15.38
hydrazine	0	7	7	26	83	2	0.00	5.47	3.80	3.95	6.43	1.92
hydroxylamine	0	3	3	22	8	0	0.00	2.34	1.63	3.34	0.62	0.00
imine	0	1	1	2	21	1	0.00	0.78	0.54	0.30	1.63	0.96
iminomethyl	6	7	13	136	63	0	10.71	5.47	7.07	20.67	4.88	0.00
isocyanate	**0**	**0**	**0**	**0**	**1**	**1**	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.96**
ketone	0	9	9	35	98	5	0.00	7.03	4.89	5.32	7.60	4.81
mercaptan	**0**	**0**	**0**	**2**	**4**	**0**	**0.00**	**0.00**	**0.00**	**0.30**	**0.31**	**0.00**

Structure Set	Frequency						% Frequency					
	AGES	CRD	CRD-AGES	PPP	CPDB	Classified mutagens	AGES	CRD	CRD-AGES	PPP	CPDB	Classified mutagens
misc nitrogen groups	5	3	8	7	200	5	8.93	2.34	4.35	1.06	15.50	4.81
misc oxygen groups	0	1	1	2	4	1	0.00	0.78	0.54	0.30	0.31	0.96
misc sulfur groups	1	0	1	10	15	2	1.79	0.00	0.54	1.52	1.16	1.92
nitrile	2	7	9	27	10	0	3.57	5.47	4.89	4.10	0.78	0.00
nitro	9	9	18	28	119	11	16.07	7.03	9.78	4.26	9.22	10.58
nitroso	**0**	**0**	**0**	**0**	**139**	**2**	**0.00**	**0.00**	**0.00**	**0.00**	**10.78**	**1.92**
organometal	**0**	**0**	**0**	**0**	**0**	**1**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.96**
phosphorous groups	0	2	2	15	17	3	0.00	1.56	1.09	2.28	1.32	2.88
quinones	0	1	1	4	16	0	0.00	0.78	0.54	0.61	1.24	0.00
silicon groups	0	1	1	0	0	0	0.00	0.78	0.54	0.00	0.00	0.00
sulfide	0	3	3	31	26	1	0.00	2.34	1.63	4.71	2.02	0.96
sulfonamide	19	9	28	21	23	2	33.93	7.03	15.22	3.19	1.78	1.92
sulfonate	0	1	1	3	38	0	0.00	0.78	0.54	0.46	2.95	0.00
sulfone	0	3	3	14	8	0	0.00	2.34	1.63	2.13	0.62	0.00
sulfonic acid	**0**	**0**	**0**	**0**	**32**	**0**	**0.00**	**0.00**	**0.00**	**0.00**	**2.48**	**0.00**
sulfonyl group	21	10	31	42	72	4	37.50	7.81	16.85	6.38	5.58	3.85
sulfoxide	**0**	**0**	**0**	**2**	**3**	**0**	**0.00**	**0.00**	**0.00**	**0.30**	**0.23**	**0.00**
thiocarboxamide	**0**	**0**	**0**	**2**	**4**	**0**	**0.00**	**0.00**	**0.00**	**0.30**	**0.31**	**0.00**
thioxomethyl	1	1	2	11	36	1	1.79	0.78	1.09	1.67	2.79	0.96
urea	1	18	19	36	57	5	1.79	14.06	10.33	5.47	4.42	4.81
Heterocycles	44	76	120	258	444	32	78.57	59.38	65.22	39.21	34.42	30.77
Naphthalenes	0	1	1	9	43	2	0.00	0.78	0.54	1.37	3.33	1.92
Natural products	**0**	**0**	**0**	**1**	**18**	**0**	**0.00**	**0.00**	**0.00**	**0.15**	**1.40**	**0.00**
Pharmacophores	56	127	183	633	1213	89	100.00	99.22	99.46	96.20	94.03	85.58
Protective groups	1	1	2	42	140	3	1.79	0.78	1.09	6.38	10.85	2.88
Spacer groups	24	37	61	190	329	13	42.86	28.91	33.15	28.88	25.50	12.50

In bold red, structural classes in the CRD and AGES datasets that are not represented in the PPP dataset

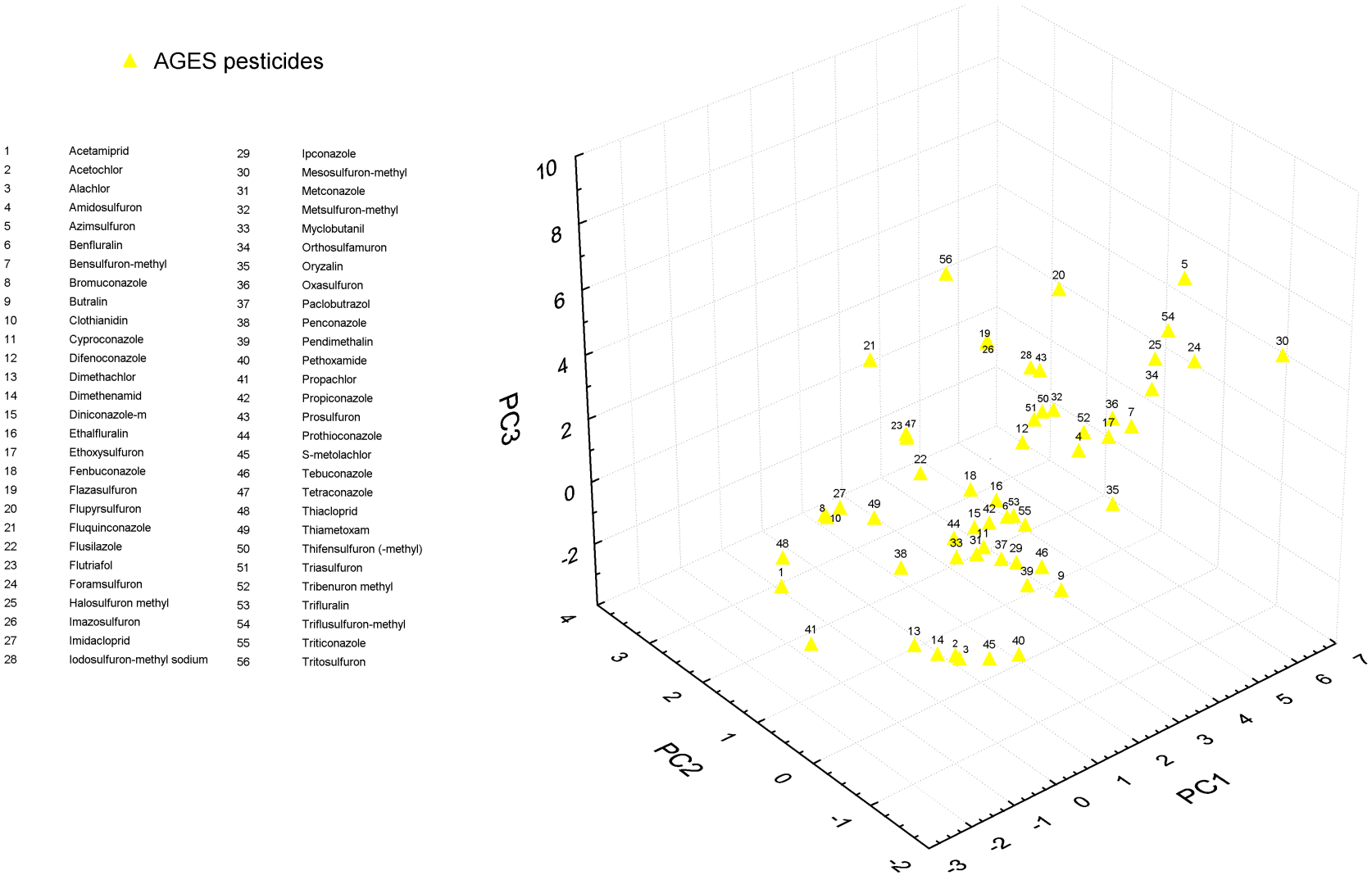


Figure 2.1 Chemical space of pesticides in the AGES dataset

1	1-MCP	65	Nicosulfuron
2	2,4-DB	66	Novaluron
3	Abamectin	67	Parathion-methyl
4	Aldicarb	68	Pencycuron
5	Amisulbrom	69	Phosmet
6	Amitrole	70	Picoxystrobin
7	Aviglycine	71	Pirimicarb
8	Azinphos-methyl	72	Prohexadione calcium
9	Azoxystrobin	73	Propyzamide
10	Benalaxyl-m	74	Prosulfocarb
11	Benthiavalicarb	75	Pymetrozine
12	**Bifenox**	76	Pyraclostrobin
13	Bifenthrin	77	Quinoclamine
14	Carbendazim	78	Rimsulfuron
15	Carbosulfan	79	Silthiofam
16	Carboxin	80	Sintofen
17	Chlorothalonil	**81**	**Sodium nitroguiacolate**
18	Chlorotoluron	82	Sodium p_nitrophenolate
19	Chlorsulfuron	83	Spinosad
20	Cinidon ethyl	84	Spiromesifen
21	Cyanamide	85	Sulfosulfuron
22	Cypermethrin	**86**	**Tebufenpyrad**
23	Desmedipham	87	Teflubenzuron
24	Dicamba	88	Tefluthrin
25	Dichlorvos	89	Thifensulfuron
26	Dicloran	90	Thiobencarb
27	Dicofol	91	Tralkoxydim
28	Dimethenamid-p	92	Triadimenol
29	Dimethomorph	93	Tri-allate
30	**Dinocap**	94	Tribenuron
31	Dinoterb	95	Triflumizole
32	Diuron	96	Triflumuron
33	Epoxiconazole	97	Valiphenal
34	Ethephon	98	Lambdcyhalothrin
35	Ethofumesate	99	Bitertanol
36	Etridiazole	100	Boscalid
37	Famoxadone	101	Carbaryl
38	Fenhexamid	102	Dimethoate
39	Fenitrothion	103	Fenamidone
40	Fenoxaprop-p	104	Formetanate
41	Fluazifop-p-butyl	**105**	**Metconazole**
42	Fluazinam	106	Proquinazid
43	Fludioxonil	107	Spirotetramat
44	Flufenacet	**108**	**Thiodicarb**
45	Flumioxazine	109	Bitertanol_triazolylalanine
46	Fluometuron	110	Dimethoate_omethoate
47	Fluopicolide.	111	Fenamidone_rpa405862
48	Flupyrsulfuron-methyl	112	Fenamidone_rpa717879
49	Flurprimidol	113	Fenamidone_rpa408056
50	Flurtamone	114	Fludioxonil_cga192155
51	**Forchlorfenuron**	115	Fludioxonil_cga308103
52	Fosthiazate	116	Fludioxonil_cga339833
53	Haloxyfop r	117	Metconazole_triazolylalanine
54	Laminarin	118	Metiram_etu
55	Linuron	119	Metiram_eu
56	Lufenuron	120	Pirimicarb_r31805
57	Maleic hydrazide	121	Pirimicarb_r34865
58	Mcpb	122	Proquinazid_in-mm671
59	Metaflumizone	123	Spirotetramabyi 08330-desmethyl-ketohydroxy
60	Metazachlor	124	Spirotetramabyi 08330-dihydroxy
61	Methiocarb	125	Spirotetramabyi 08330-enol
62	Metosulam	126	Spirotetramabyi 08330-ketohydroxy
63	**Metrafenone**	127	Spirotetramabyi 08330-monohydroxy
64	Metribuzin	128	Thiodicarb_methomyl

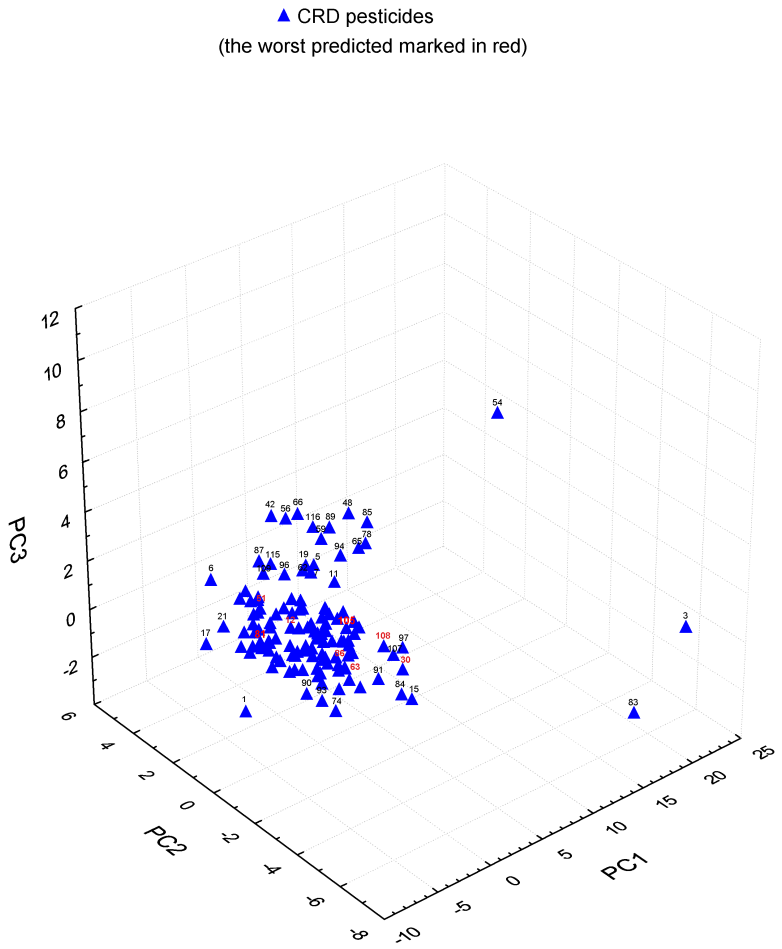


Figure 2.2 Chemical space of pesticides in the CRD dataset, including the worst predicted chemicals for mutagenicity and carcinogenicity

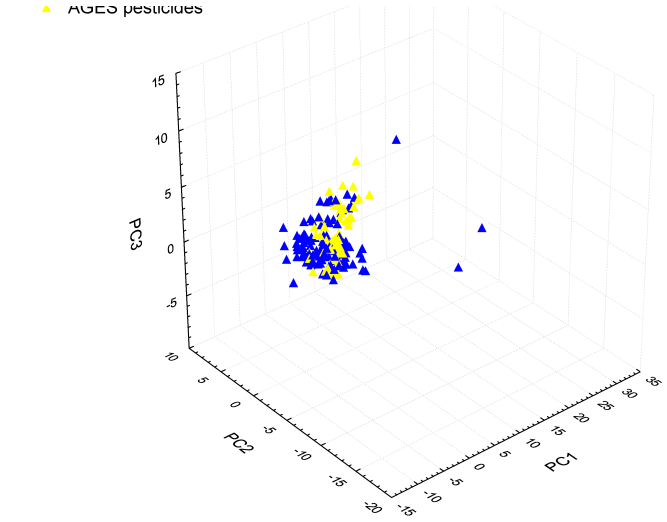


Figure 2.3 Chemical space of pesticides in the AGES and CRD datasets

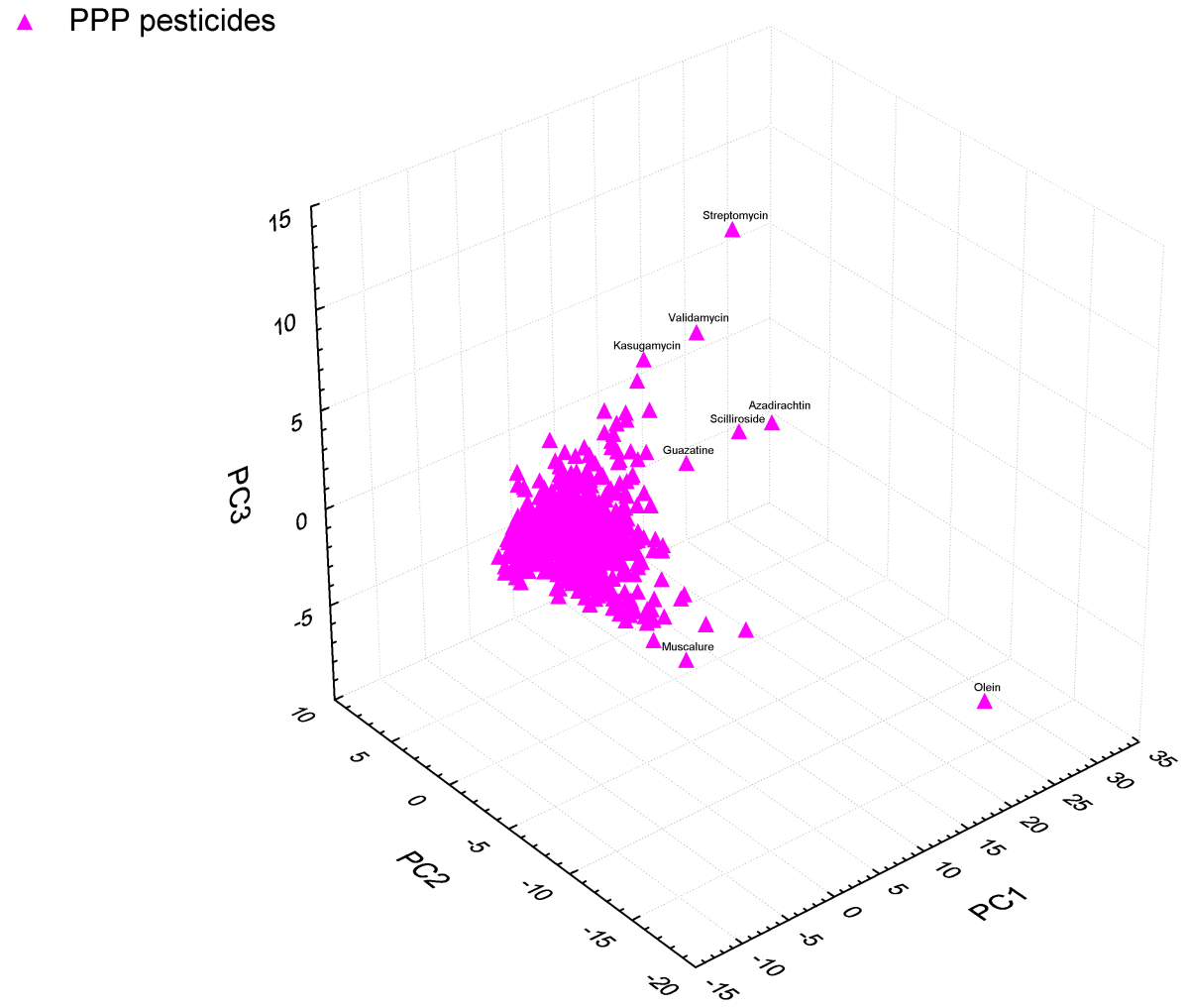


Figure 2.4 Chemical space of pesticides in the PPP database

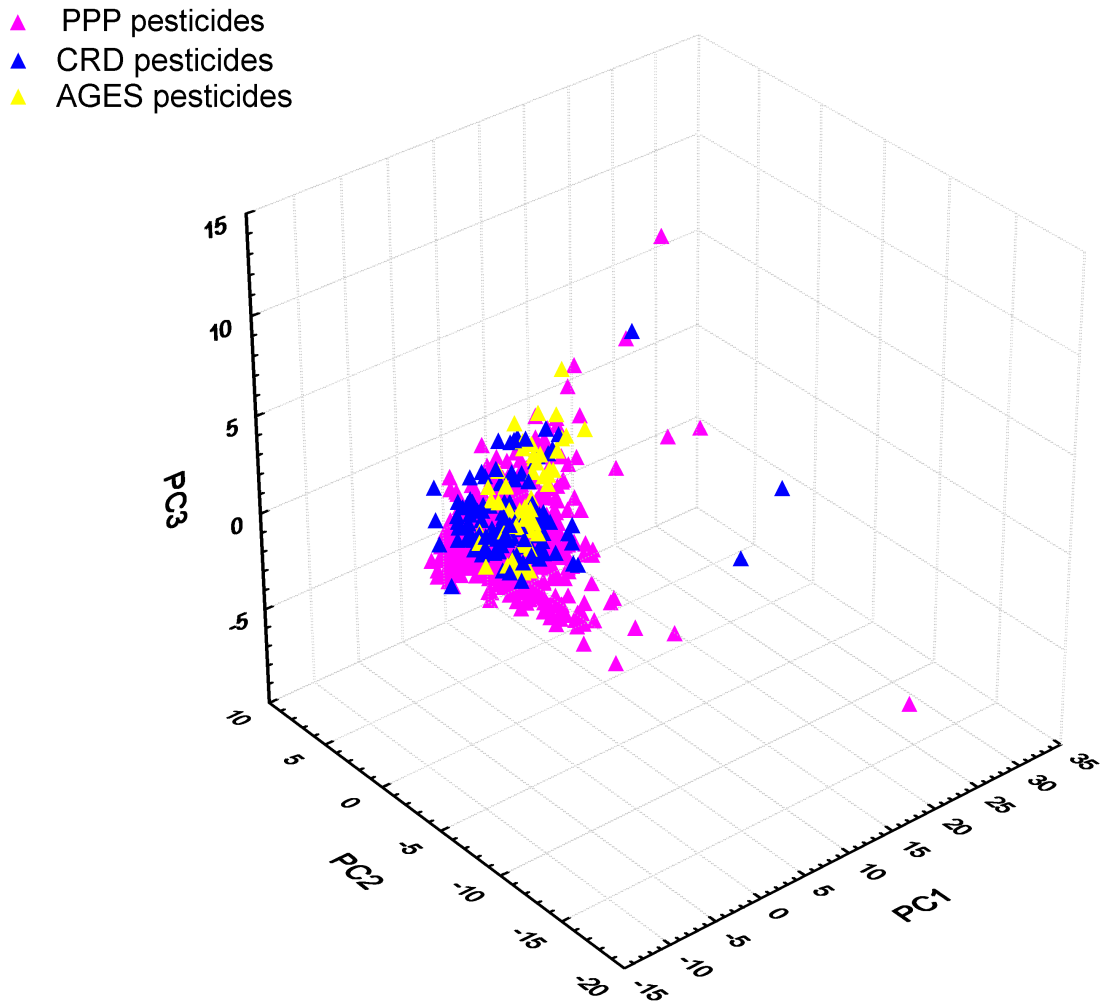


Figure 2.5 Chemical space of pesticides in the CRD, AGES and PPP datasets

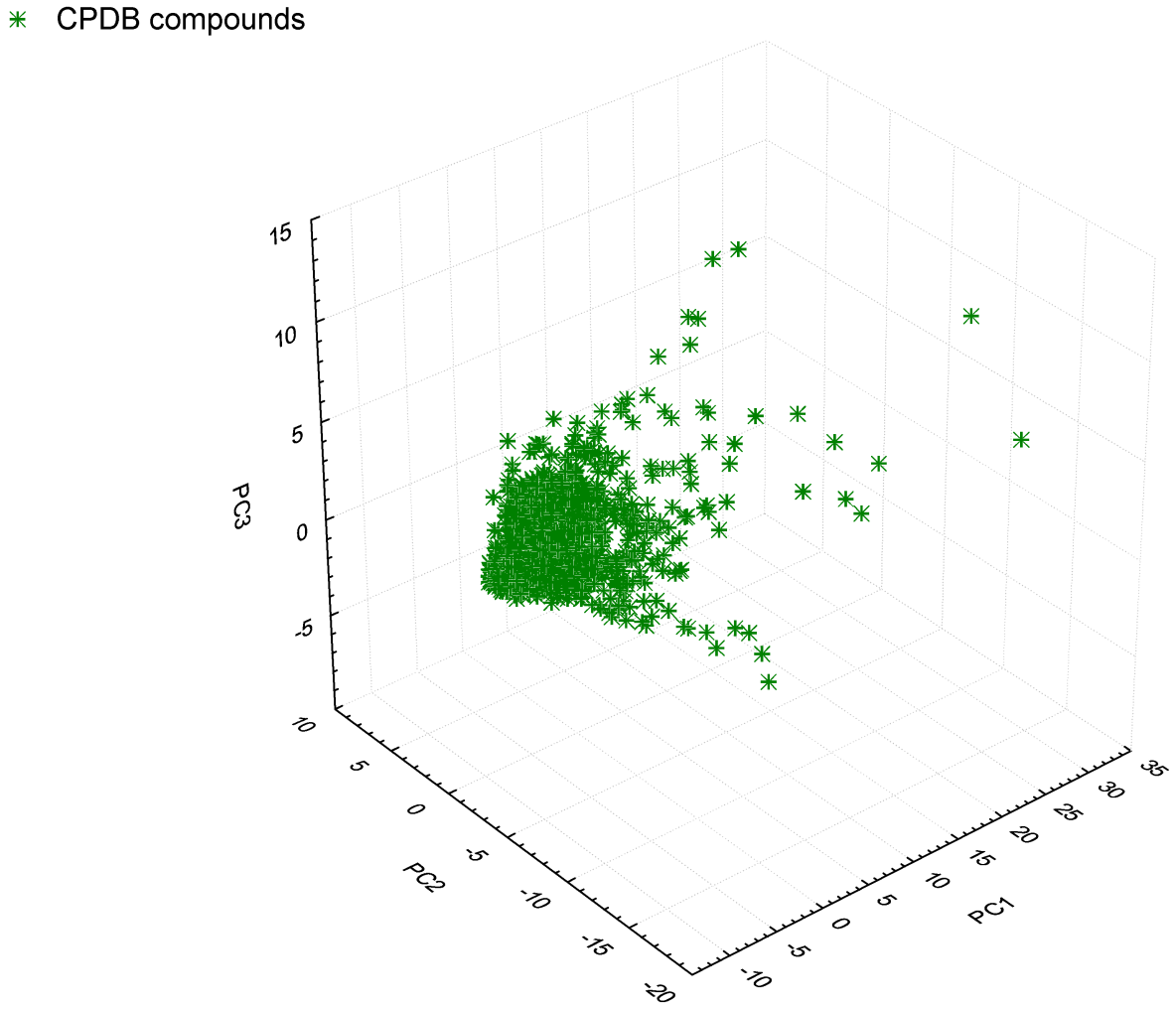


Figure 2.6 Chemical space of a broad chemical inventory (CPDB)

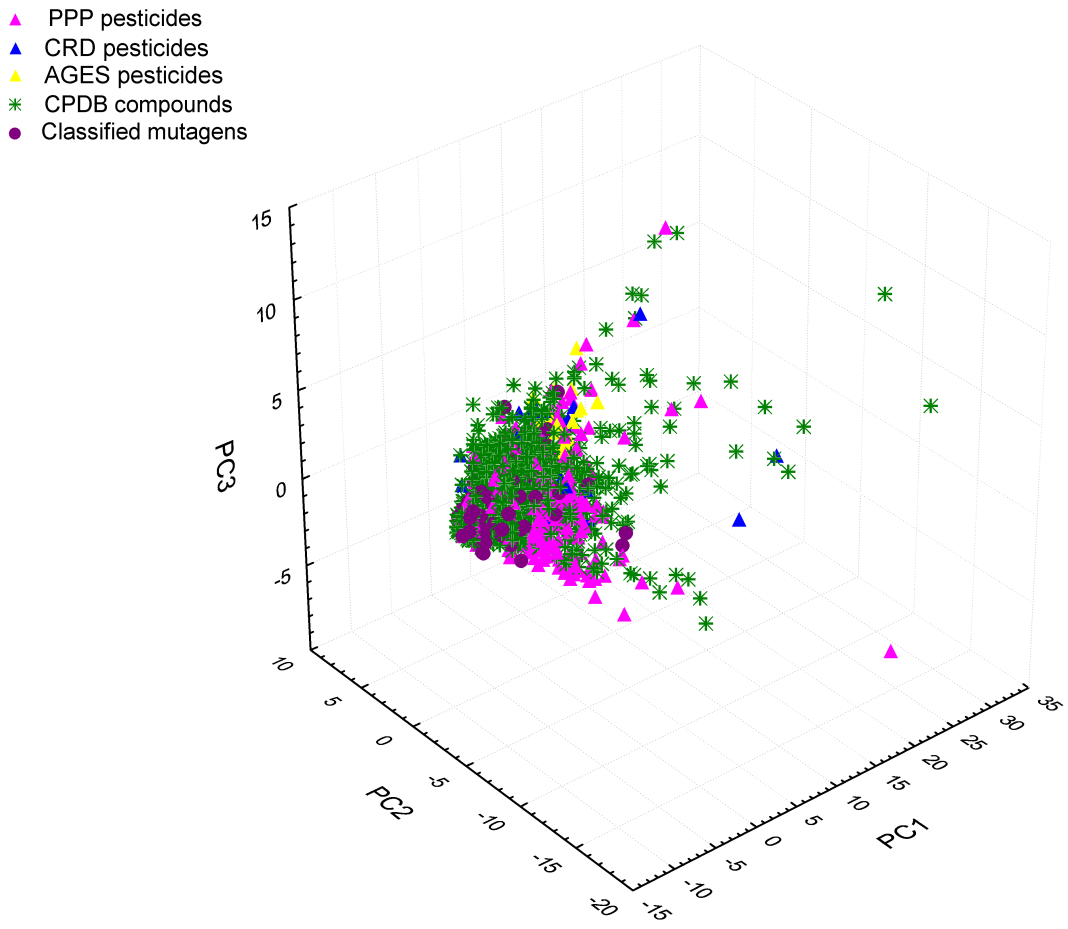


Figure 2.7 Comparison of chemical space of pesticides with the CPDB database

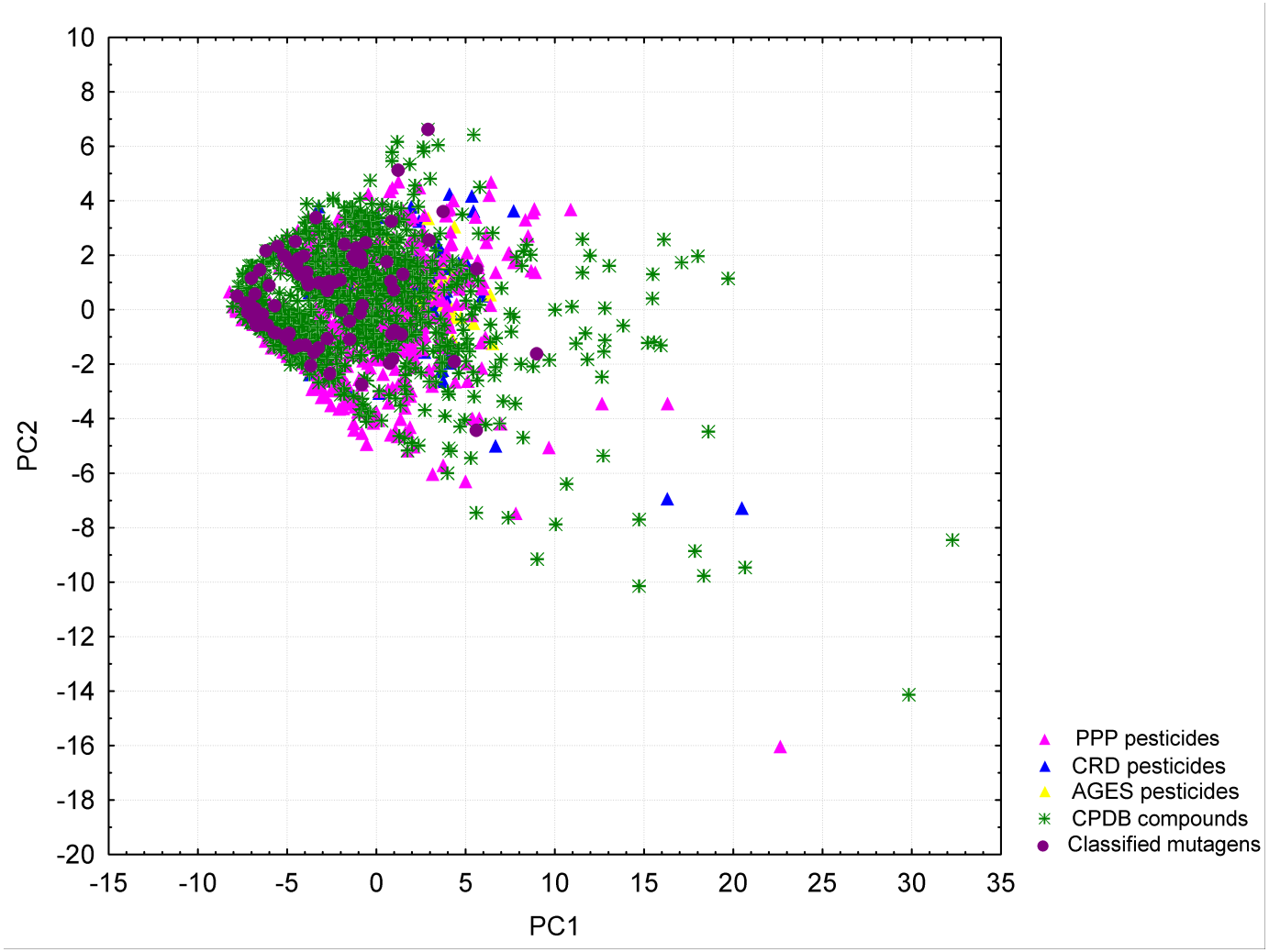


Figure 2.8 Comparison of chemical space of pesticides with the CPDB database (PC1 vs PC2)

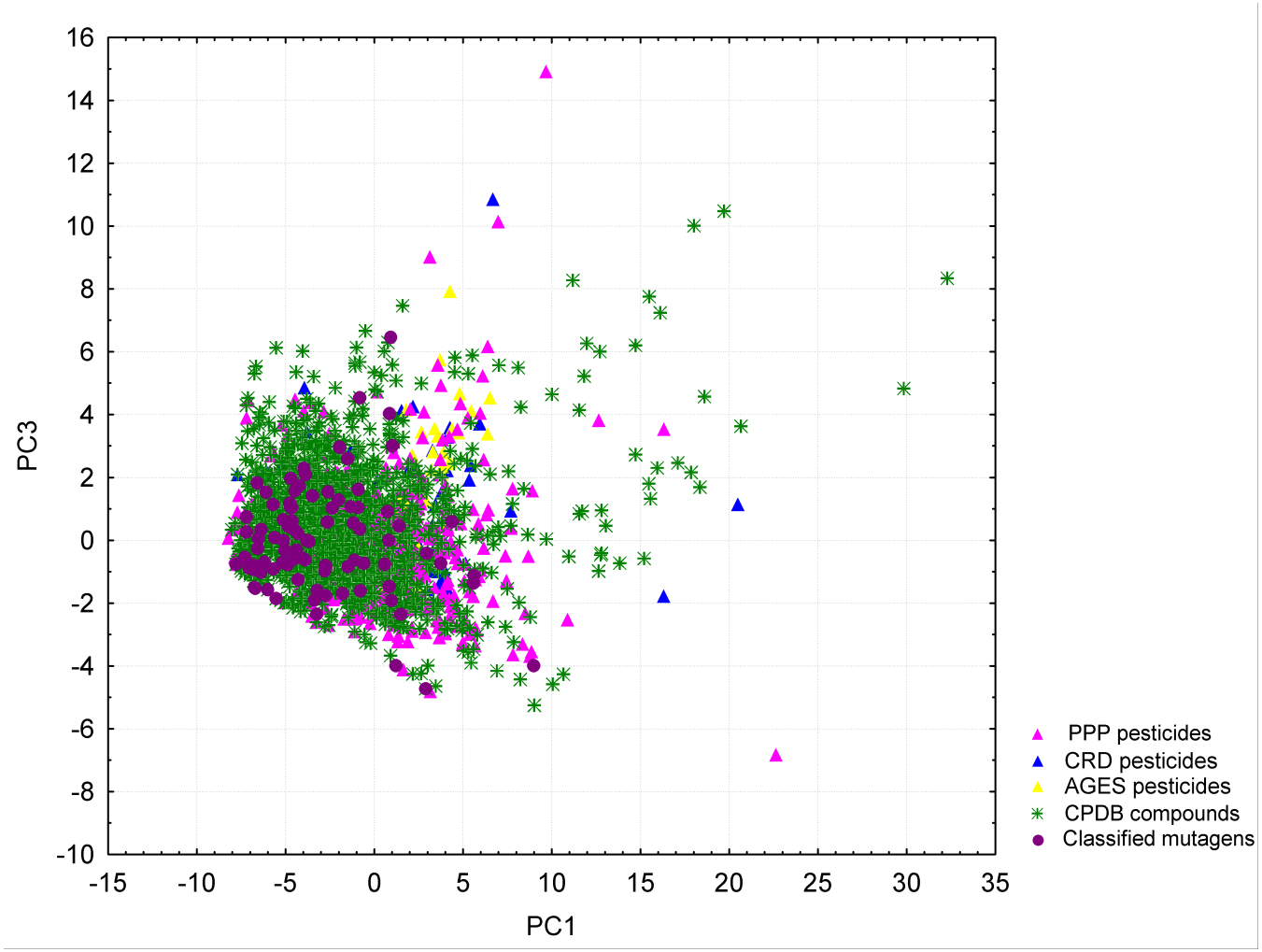


Figure 2.9 Comparison of chemical space of pesticides with the CPDB database (PC1 vs PC3)

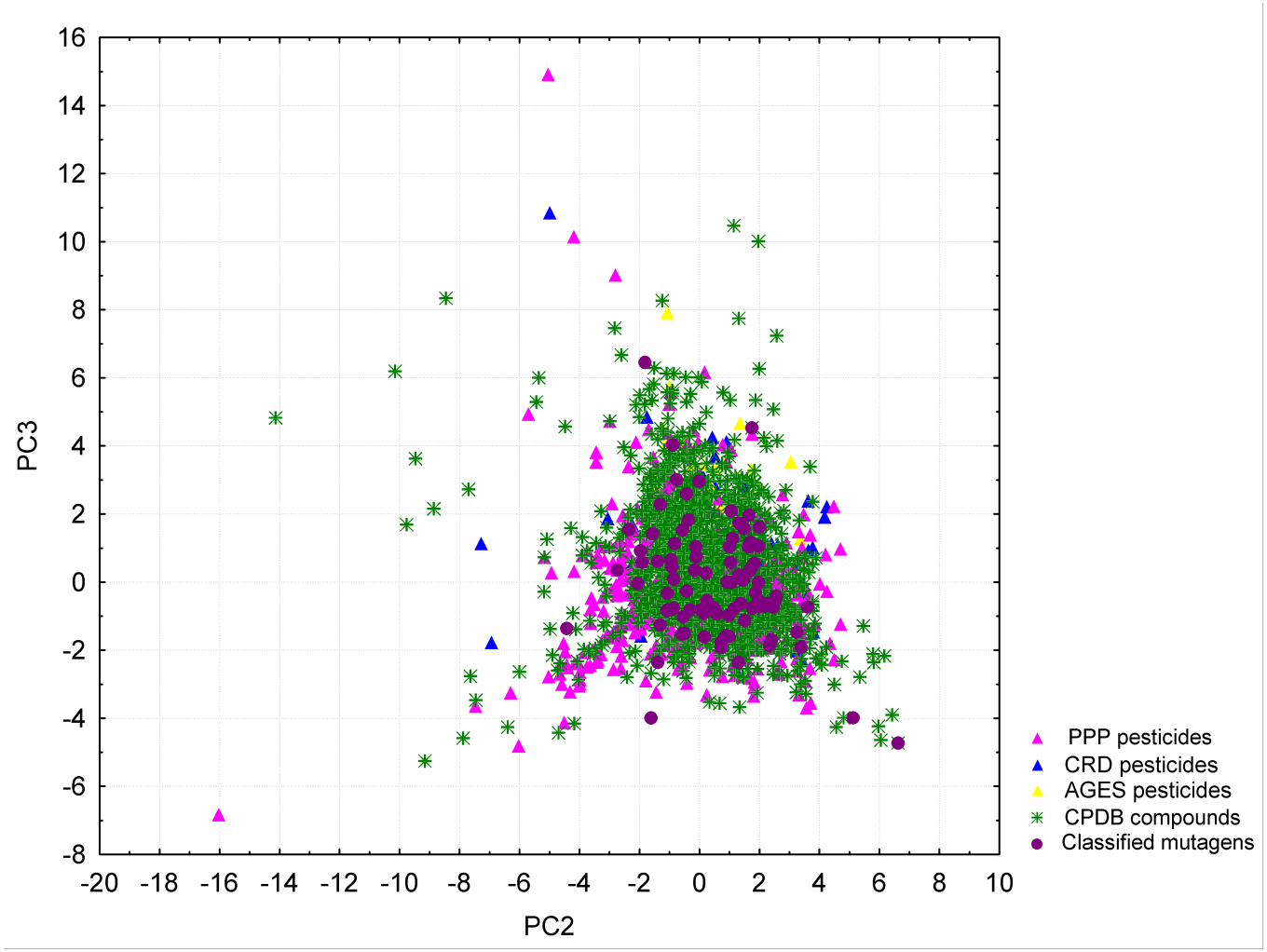


Figure 2.10 Comparison of chemical space of pesticides with the CPDB database (PC2 vs PC3)

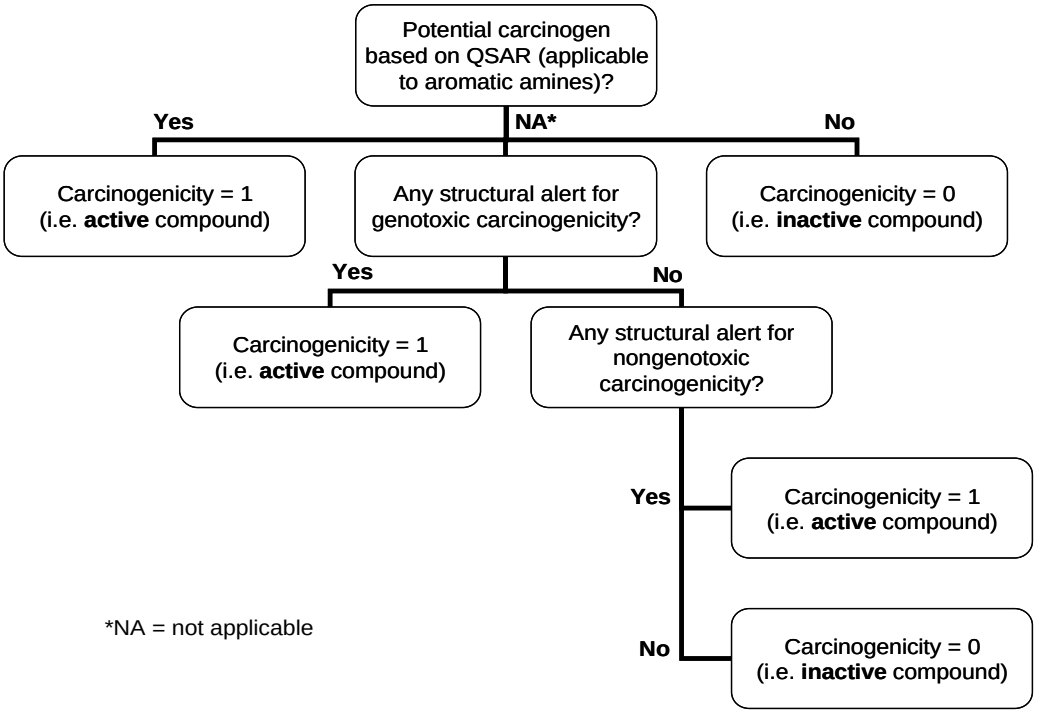


Figure 3.1a Interpretation of carcinogenicity predictions generated by Toxtree

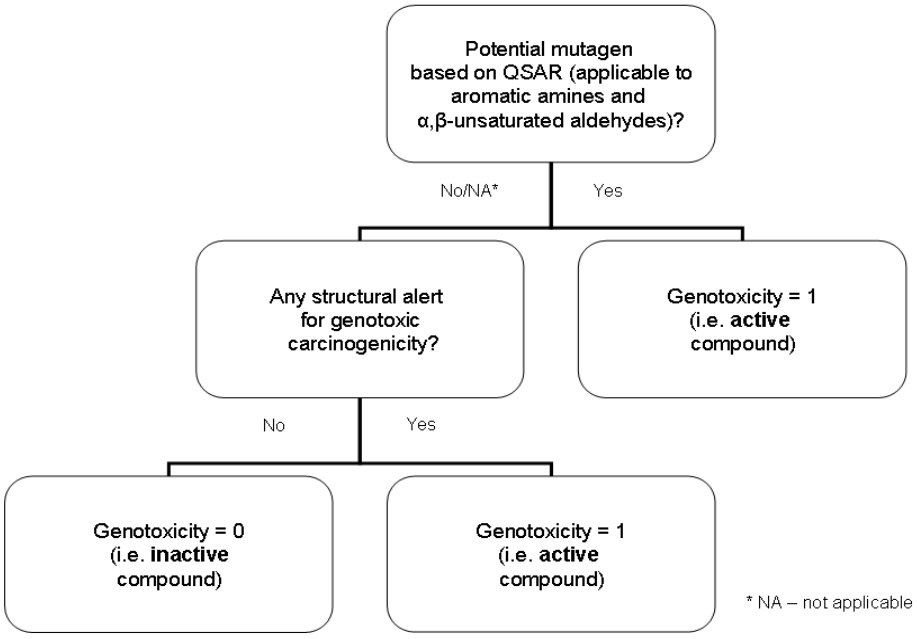


Figure 3.1b Interpretation of gentoxicity predictions generated by Toxtree

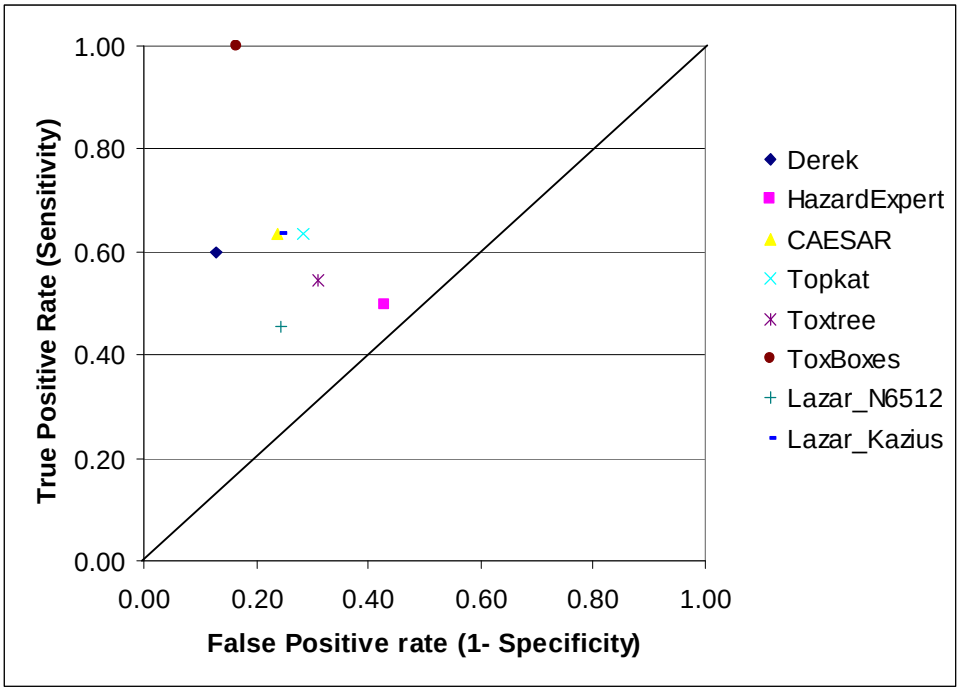


Figure 3.2 ROC curve for pesticide genotoxicity predictions

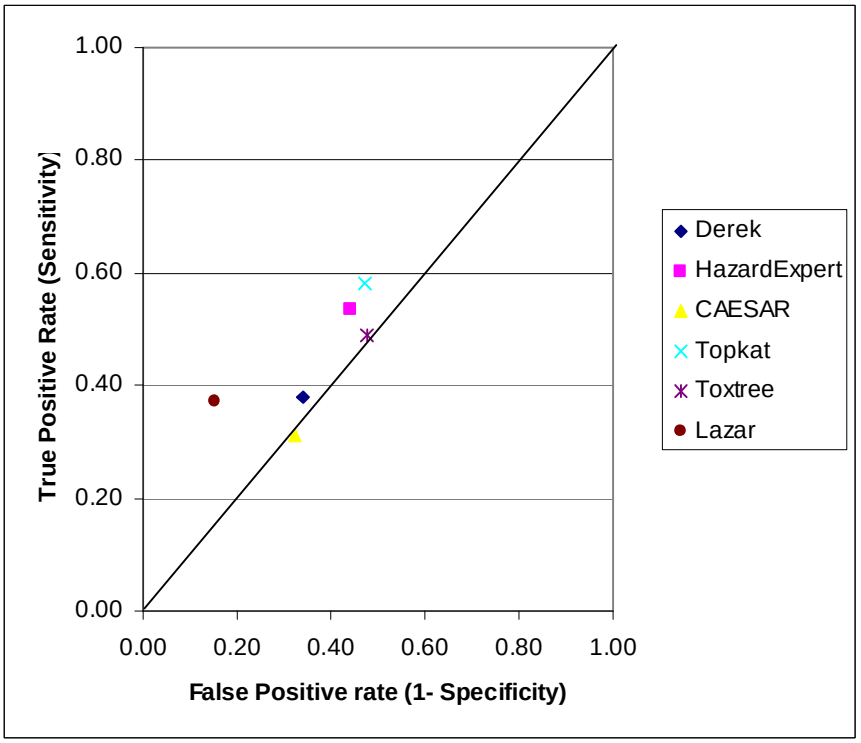


Figure 3.3 ROC curve for pesticide carcinogenicity predictions

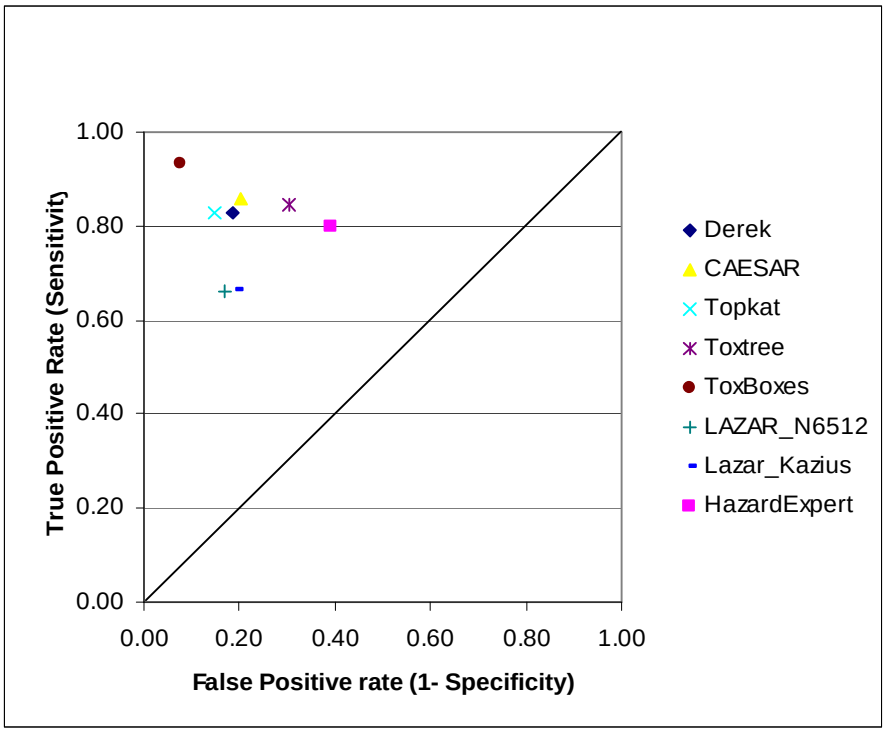


Figure 3.4 ROC curve for DSSTox genotoxicity predictions

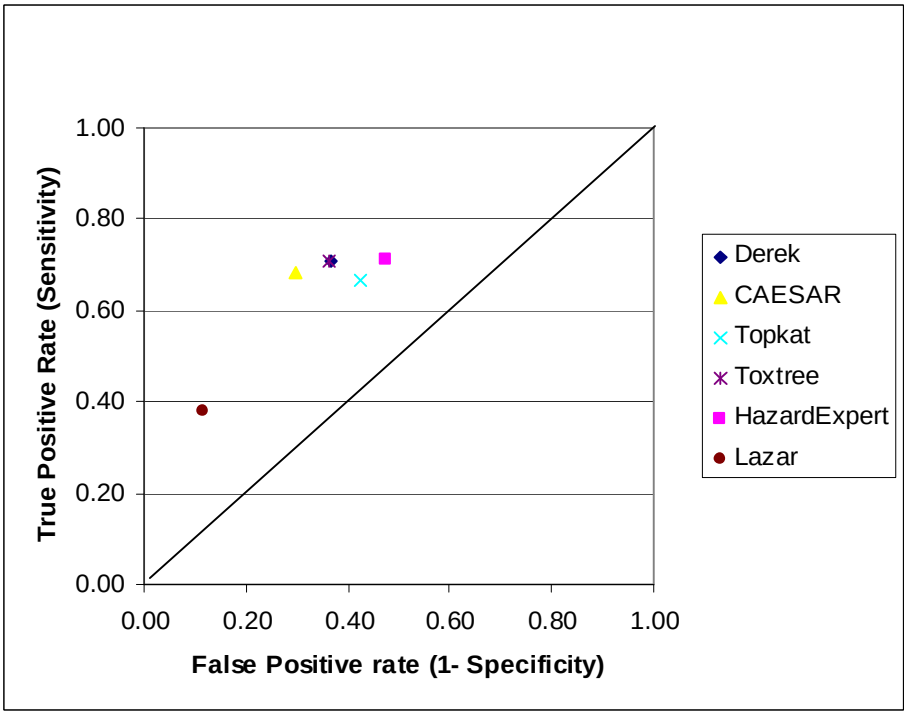


Figure 3.5 ROC curve for DSSTox carcinogenicity predictions

Table 3.1. Genotoxicity prediction results for the pesticides dataset

Number of compounds: 185
Experimental values available: 181
Exp. active compounds: 11
Exp. inactive compounds: 170

SOFTWARE	STATISTICS										
	TP	TN	FP	FN	EQ	ND	SP	SE	CONC	1-SE	1-SP
CAESAR	7	129	40	4	1	0	0.76	0.64	0.76	0.36	0.24
Derek	6	148	22	4	1	0	0.87	0.60	0.86	0.40	0.13
HazardExpert	5	95	71	5	5	0	0.57	0.50	0.57	0.50	0.43
Lazar (Kazius/Bursi)	7	127	41	4	0	2	0.76	0.64	0.75	0.36	0.24
Lazar (Toxbenchmark)	5	127	41	6	0	2	0.76	0.45	0.74	0.55	0.24
TOPKAT	7	121	48	4	0	1	0.72	0.64	0.71	0.36	0.28
ToxBoxes	4	112	22	0	43	0	0.84	1.00	0.84	0.00	0.16
Toxtree (Benigni-Bossa)	6	117	53	5	0	0	0.69	0.55	0.68	0.45	0.31

TP – true positives; **TN** – true negatives; **FP** – false positives; **FN** – false negatives; **EQ** – compounds predicted as equivocal; **ND** – the number of compounds that were not handled by the software; **SP** – specificity; **SE** – sensitivity; **CONC** – overall concordance; **1-SE** – false negative rate; **1-SP** – false positive rate

Table 3.2. Genotoxicity prediction results for the DSSTox dataset

Number of compounds: 1290
Experimental values available: 748
Exp. active compounds: 368
Exp. inactive compounds: 380

SOFTWARE	STATISTICS										
	TP	TN	FP	FN	EQ	ND	SP	SE	CONC	1-SE	1-SP
CAESAR	315	298	76	52	7	0	0.80	0.86	0.83	0.14	0.20
Derek	299	298	68	62	20	1	0.81	0.83	0.82	0.17	0.19
HazardExpert	285	199	128	72	64	0	0.61	0.80	0.71	0.20	0.39
Lazar (Kazius/Bursi)	245	305	74	123	0	1	0.80	0.67	0.74	0.33	0.20
Lazar (Toxbenchmark)	243	316	64	124	0	1	0.83	0.66	0.75	0.34	0.17
TOPKAT	286	315	55	59	26	7	0.85	0.83	0.84	0.17	0.15
ToxBoxes	300	301	22	24	101	0	0.93	0.93	0.93	0.07	0.07
Toxtree	311	265	115	57	0	0	0.70	0.85	0.77	0.15	0.30

TP – true positives; **TN** – true negatives; **FP** – false positives; **FN** – false negatives; **EQ** – compounds predicted as equivocal; **ND** – the number of compounds that were not handled by the software; **SP** – specificity; **SE** – sensitivity; **CONC** – overall concordance; **1-SE** – false negative rate; **1-SP** – false positive rate

Table 3.3. Carcinogenicity prediction results for the pesticides dataset

Number of compounds: 185
Experimental values available: 104
Exp. active compounds: 45
Exp. inactive compounds: 59

SOFTWARE	STATISTICS										
	TP	TN	FP	FN	EQ	ND	SP	SE	CONC	1-SE	1-SP
CAESAR	14	40	19	31	0	0	0.68	0.31	0.52	0.69	0.32
Derek	17	39	20	28	0	0	0.66	0.38	0.54	0.62	0.34
HazardExpert	23	30	24	20	7	0	0.56	0.53	0.55	0.47	0.44
Lazar	16	49	9	27	0	3	0.84	0.37	0.64	0.63	0.16
TOPKAT	25	30	27	18	0	4	0.53	0.58	0.55	0.42	0.47
Toxtree	22	31	28	23	0	0	0.53	0.49	0.51	0.51	0.47

TP – true positives; **TN** – true negatives; **FP** – false positives; **FN** – false negatives; **EQ** – compounds predicted as equivocal; **ND** – the number of compounds that were not handled by the software; **SP** – specificity; **SE** – sensitivity; **CONC** – overall concordance; **1-SE** – false negative rate; **1-SP** – false positive rate

Table 3.4. Carcinogenicity prediction results for the DSSTox dataset

Number of compounds: 1290
Experimental values available: 1288
Exp. active compounds: 717
Exp. inactive compounds: 571

SOFTWARE	STATISTICS										
	TP	TN	FP	FN	EQ	ND	SP	SE	CONC	1-SE	1-SP
CAESAR	490	401	170	227	0	0	0.70	0.68	0.69	0.32	0.30
Derek	505	358	209	209	6	1	0.63	0.71	0.67	0.29	0.37
HazardExpert	495	283	255	198	56	1	0.53	0.71	0.63	0.29	0.47
Lazar	220	412	54	361	0	241	0.88	0.38	0.49	0.62	0.12
TOPKAT	453	306	227	228	62	12	0.57	0.67	0.63	0.33	0.43
Toxtree	508	364	207	209	0	0	0.64	0.71	0.68	0.29	0.36

TP – true positives; **TN** – true negatives; **FP** – false positives; **FN** – false negatives; **EQ** – compounds predicted as equivocal; **ND** – the number of compounds that were not handled by the software; **SP** – specificity; **SE** – sensitivity; **CONC** – overall concordance; **1-SE** – false negative rate; **1-SP** – false positive rate

Table 3.5. True positives (genotoxic substances) in the CRD-AGES pesticides dataset

Substance	CAESAR	Derek	HazardExpert	Lazar (Kazius/Bursi)	Lazar (Toxbenchmark)	TOPKAT	ToxBoxes	Toxtree
Carbendazim	X		X	X	X	X		
Dichlorvos	X	X				X	X	X
Dinocap	X	X	X	X	X		X	X
Ethephon		X		X		X		X
Fenitrothion		X	X	X	X	X		X
Metiram ETU						X		
Parathion-methyl	X	X	X	X	X	X		X
Phosmet	X	X						
Thiodicarb	X			X			X	
Tri-allate	X		X	X	X	X	X	X

Table 3.6. Positive and negative predictivities for genotoxicity (DSSTox dataset)

Software	Positive predictivity	Negative predictivity
CAESAR	0.79	0.84
Derek	0.81	0.83
HazardExpert	0.69	0.73
Lazar (Kazius/Bursi)	0.77	0.70
Lazar (Toxbenchmark)	0.79	0.71
TOPKAT	0.84	0.84
ToxBoxes	0.93	0.93
Toxtree	0.73	0.82

Table 3.7. Positive and negative predictivities for carcinogenicity (DSSTox dataset)

Software	Positive predictivity	Negative predictivity
CAESAR	0.74	0.64
Derek	0.71	0.63
HazardExpert	0.66	0.59
Lazar	0.80	0.53
TOPKAT	0.67	0.57
Toxtree	0.71	0.64

Table 3.8. Ability of software tools to identify classified mutagens

Software (used alone)	ND	EQ	TP	SE	FN	1-SE	No TS
Toxtree (genotoxic carcinogenicity)	0	0	86	0.76	27	0.24	NA
Toxtree (in vivo micronucleus)	0	0	98	0.87	15	0.13	NA
Toxtree (genotoxic carcinogenicity or in vivo micronucleus)	0	0	98	0.87	15	0.13	NA
TOPKAT	1	0	65	0.58	47	0.42	43
CAESAR	1	0	82	0.73	30	0.27	48
HazardExpert	0	5	82	0.77	25	0.23	Not known
Lazar (Kazius/Bursi)	0	0	65	0.58	48	0.42	58*
Lazar (Toxbenchmark)	0	0	56	0.50	57	0.50	60*
Lazar (Kazius/Bursi or Toxbenchmark)	0	0	69	0.61	44	0.39	74*
Derek (mutagenicity or chromosome damage)	0	2	81	0.73	30	0.27	NA
ToxBoxes	0	22	68	0.75	23	0.25	Not known
Software (used in combination)							
Toxtree or CAESAR	0	0	101	0.89	12	0.11	48
Derek or CAESAR	0	0	96	0.85	17	0.15	48
Derek or Lazar	0	0	92	0.81	21	0.19	74*
Derek or TOPKAT	0	0	89	0.79	24	0.21	43
Toxtree or Lazar	0	0	102	0.90	11	0.10	74*
Toxtree or Derek	0	0	104	0.92	9	0.08	NA
HazardExpert or CAESAR	0	0	94	0.83	19	0.17	• 48

Test set of 113 classified mutagens; **ND** – not determined**; EQ** – compounds predicted as equivocal; **TP** – true positives; **SE** – sensitivity; **FN** – false negatives; **1-SE** – false negative rate; **No TS** – number of chemicals already in the training set of the model (where applicable); **NA** – not applicable

***** For Lazar it is not important whether a substance is in the dataset used to build the model, since an instance-based prediction is generated by a local model built from data that exclude the query chemical

4. Acknowledgements and Disclaimer

This work was carried out under the terms of a grant awarded by the European Food Safety Authority (EFSA) to the European Commission's Joint Research Centre (JRC), Institute for Health & Consumer Protection, Ispra, Italy (Contract number: SLA/EFSA-JRC/2009-01).

Any conclusions and opinions expressed in this document are those of the authors as individual scientists and do not constitute an official position by the JRC or the European Commission.

The authors appreciate the advice provided by EFSA staff (Hans Steinkellner, Luc Mohimont, Miriam Jacobs) and members of the EFSA Steering Group (André Muller, Emilio Benfenati, Claudia Bolognesi). In addition, the authors are grateful to Matteo Stocchero (Soluzioni Informatiche, Vicenza) for support in the generation and interpretation of ToxBoxes data, as well as to Christoph Helma (*in silico* toxicology, Basel) for providing data from the Lazar software. The authors also acknowledge the useful comments and editorial revisions provided by JRC staff (David Asturiol and Jutta Triebe).

5. References

AGES (2010). Impact of metabolic and degradation processes on the toxicological properties of residues of pesticides in food commodities. Report from the Austrian Austrian Agency for Health and Food Safety (AGES) to the European Food Safety Authority (EFSA). Available from http://www.efsa.europa.eu/en/scdocs/scdoc/49e.htm

Barlow S (2005). Threshold of toxicological concern: a tool for assessing substances of unknown toxicity present at low levels in the diet. ILSI Concise Monograph Series, ISBN 1-57881-188-0, ILSI Press, Washington DC and Brussels.

Benigni R, Bossa C, Jeliazkova N, Netzeva T & Worth A (2008). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. EUR 23241 EN. http://ecb.jrc.ec.europa.eu/qsar/publications/

Benigni R, Bossa C, Tcheremenskaia O & Worth A (2009). Development of structural alerts for the in vivo micronucleus assay in rodents. EUR 23844 EN. http://ecb.jrc.ec.europa.eu/qsar/publications/

Benigni R, Bossa C & Worth A (2010). Structural analysis and predictive value of the rodent in vivo micronucleus assay results Mutagenesis, in press. Available online: doi:10.1093/mutage/geq010

CRD (2010) Applicability of thresholds of toxicological concern in the dietary risk assessment of metabolites, degradation and reaction products of pesticides. Report from the UK Chemicals Regulation Directorate (CRD) to the European Food Safety Authority (EFSA). Available from http://www.efsa.europa.eu/en/scdocs/scdoc/44e.htm

EC (1991). Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market. Official Journal of the European Union, L 230/1 of 19.08.1991. Office for Official Publications of the European Communities (OPOCE), Luxembourg.

EC (2002). Regulation (EC) No 178/2008 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal L 031, 1-24. 1 February 2002.

EC (2005). Regulation (EC) No 396/2005 of the European Parliament and of the Council of 23 February 2005 on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EEC. Official Journal L070, 1-16. 16 March 2005.

ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R6. European Chemicals Agency, Helsinki, Finland. Available at: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1252064523#r6

Enoch SJ, Madden JC & Cronin MTD (2008). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. SAR and QSAR in Environmental Research 19, 555-578.

Felter S, Lane RW, Latulippe ME, Llewellyn GC, Olin SS, Scimeca JA & Trautman TD (2009). Refining the threshold of toxicological concern (TTC) for risk prioritization of trace chemicals in food. Food & Chemical Toxicology 47(9), 2236-2245.

Helma C (2006). Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity 10, 147-158.

JRC (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Report from the European Commission's Joint Research Centre (JRC) to the European Food Safety Authority (EFSA). Available from http://www.efsa.europa.eu/en/scdocs/scdoc/50e.htm

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG & Würtzen G (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food & Chemical Toxicology 42(1), 65-83.

Munro IC, Renwick AG & Danielewska-Nikiel B (2008). The Threshold of Toxicological Concern (TTC) in risk assessment. Toxicology Letters 180(2), 151-156.

OECD (2007). Guidance on Grouping of Chemicals. Series on Testing and Assessment Number 80. ENV/JM/MONO(2007)28. Organisation for Economic Cooperation and Development, Paris, France. Available at: http://www.oecd.org/

OECD (2009). Guidance Document on the Definition of Residue. Series on Testing and Assessment No. 63 and Series on Pesticides no.31. 28 July 2009. Available at: http://www.oecd.org/

Rydberg P, Gloriam DE, Zaretzki J, Breneman C & Olsen L (2010). SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. ACS Medicinal Chemistry Letters **1**, 96-100.

Sanderson DM & Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; the DEREK system. Human and Experimental Toxicology 10, 261-273.

6. Appendices

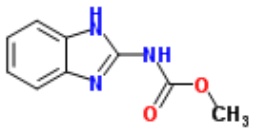
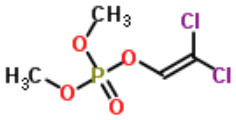
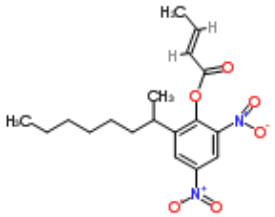
Appendix 1. DRAGON descriptors used for the investigation of chemical space

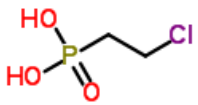
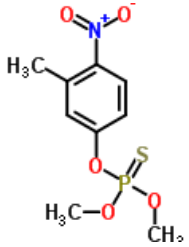
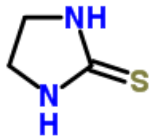
ID	Explanation	Type of descriptor
MW	Molecular Weight	Constitutional
Sv	Sum of atomic van der Waals volumes	Constitutional
Se	Sum of atomic Sanderson electronegativities	Constitutional
Sp	Sum of atomic polarizabilities	Constitutional
Ss	Sum of Kier-Hall electrotopological states	Constitutional
nAT	Number of atoms	Constitutional
nSK	Non-H atoms	Constitutional
nBT	Number of bonds	Constitutional
nBO	Non-H bonds	Constitutional
nBM	Multiple bonds	Constitutional
ARR	Aromatic ratio	Constitutional
nCIC	Number of rings	Constitutional
RBN	Rotatable bonds	Constitutional
nDB	Number of double bonds	Constitutional
nTB	Number of triple bonds	Constitutional
nAB	Number of aromatic bonds	Constitutional
nH	H atoms	Constitutional
nC	C atoms	Constitutional
nN	N atoms	Constitutional
nO	O atoms	Constitutional
nP	P atoms	Constitutional
nS	S atoms	Constitutional
nF	F atoms	Constitutional
nCL	Cl atoms	Constitutional
nBR	Br atoms	Constitutional
nX	Halogen atoms	Constitutional
nR03	3-membered rings	Constitutional
nR05	5-membered rings	Constitutional
nR06	6-membered rings	Constitutional
nR09	9-membered rings	Constitutional
nR10	10-membered rings	Constitutional
nHDon	Hydrogen bond donors	Functional groups counts
nHAcc	Hydrogen bond acceptors	Functional groups counts
Hy	Hydrophilic factor	Molecular properties
MLOGP	Moriguchi octanol-water part. coeff.	Molecular properties

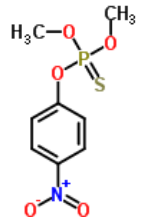
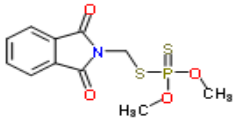
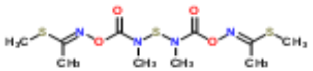
Appendix 2. Structural classes (features) identified by Leadscope

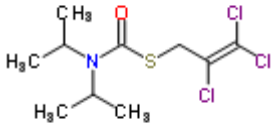
Structural classes	Definition/contents/ subclasses
Amino acids	This class contains 20 naturally occurring amino acids, homocysteine, homoserine, isovaline, ornithine in D-, L- and non-sterospecific forms.
Bases, nucleosides	This class contains the purine and pyrimidine bases adenine, cytosine, guanine, thymine, and uracil, their N-glycosyl derivatives, and the nucleosides inosine and xanthosine. For the nucleosides, both the d-ribo and 2'-deoxy-d-ribo glycosides (specific stereo and non-stereospecific (NS)) are included.
Benzenes	This class is divided into five major categories: 1,2-, 1,3-, and 1,4-substituted benzenes, substituents, and substitution patterns.
Carbocycles	This class contains common ring systems such as adamantine as well as bridged and spiro ring systems
Carbohydrates	A large variety of 4-, 5-, and 6-carbon monosaccharides are contained here within five major sub-branches. These subbranches are furanoses, pyranoses, pentoses, hexoses, and inositols. In all cases, the OH substituents have been replaced with a Z (representing {N,S,O}) in order to include the common amino and thio analogs of sugars. Each feature is represented in D-, L-, and No Stereo (NS) forms. The anomeric carbon of the furanoses and pyranoses may have either stereochemistry.
Elements	This class contains features of the form E-A, where E = {As, B, M, P, Se, Si, Te}; M = any metal atom, and A = any atom. The bond between E and A is any type.
Functional Groups	49 main functional group classes were investigated: acid anhydride, acid halide, alkohol, aldehyde, alkene, alkeny, amidine, amines, azide, boron groups, carbamate, carbonyl, carboxamide, carboxylate, carboxylic acid, ether, guanidine, halide, hydrazine, hydroxylamine, imane, iminomethyl, isocyanate, isonitrile, ketone, merkaptan, misc nitrogen groups, misc oxygen groups, misc sulfur, groups, nitryle, nitro, nitroso, organometal, phosphorous groups, quinones, silicon groups, sulfide, sulfonamide, sulfonate, sulfone, sulfonic acid, sulfonyl group, sulfonyl halide, sulfoxide, thiocarboxamide, thiocarboxylates, thiocarboxylic acids, thioxomethyl, urea.
Heterocycles	This class is the largest one, with 132 subclasses. The subclasses can be further divided into two categories: rings with substituents and rings without substituents.
Naphthalenes	This class contains 1- and 2- substituted naphthalene rings with various substituents. The naphthalene may be further substituted and embedded in a larger ring system.
Natural products	The only subclass contained here is the steroid subclass. This class contains various singly substituted and partially unsaturated steroid ring systems.
Peptidomimetics	The peptidomimetics class contains only one subclass, amide bond mimetics.
Pharmacophores	The features grouped here are pairs of generalized physiochemical atom types joined by a path of 3-8 atoms/bonds of indeterminate type, i.e. 2-D topological pharmacophores. The physiochemical types are aromatic (ARO), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), and positive (PCC) and negative (NCC) charge centers.
Protective Groups	The features contained here are the IUPAC-recommended blocking groups for heteroatoms.
Spacer Groups	This class is based on the premise that many drug-like molecules contain carbon chains separating structural features important for their activity. The subclasses are organized by chain length. Atoms in the spacer are acyclic, unfunctionalized carbons with no heteroatom attachments.

Appendix 3. True positives (genotoxic substances) in the CRD-AGES pesticides dataset

	CAESAR	Derek	HazardExpert	Lazar (Kazius/Bursi)	Lazar (Toxbenchmark)	TOPKAT	ToxBoxes	Toxtree
Carbendazim 	X		X	X	X	X		
Dichlorvos 	X	X				X	X	X
Dinocap 	X	X	X	X	X		X	X

	CAESAR	Derek	HazardExpert	Lazar (Kazius/Bursi)	Lazar (Toxbenchmark)	TOPKAT	ToxBoxes	Toxtree
Ethephon 		X		X		X		X
Fenitrothion 		X	X	X	X	X		X
Metiram ETU 						X		

	CAESAR	Derek	HazardExpert	Lazar (Kazius/Bursi)	Lazar (Toxbenchmark)	TOPKAT	ToxBoxes	Toxtree
Parathion-methyl 	X	X	X	X	X	X		X
Phosmet 	X	X						
Thiodicarb 	X			X			X	

	CAESAR	Derek	HazardExpert	Lazar (Kazius/Bursi)	Lazar (Toxbenchmark)	TOPKAT	ToxBoxes	Toxtree
Tri-allate 	X		X	X	X	X	X	X

Appendix 4. Ranking of classified mutagens from best to worst predicted

Chemical name	Toxtree	TOPKAT	CAESAR	ToxBoxes	HazardExpert	Lazar	Derek	No Correct
butane	0	0	0	0	0	0	0	0
isobutane	0	0	0	0	0	0	0	0
benzene	0	0	0	0	0	0	0	0
2-chloro-6-fluoro-phenol	0	0	0	0	IND	0	0	0
ibutyltin dichloride	0	ND	ND	0	0	0	0	0
2-methyl-1,3-butadiene (isoprene)	0	0	0	0	0	0	0	0
phenol	0	0	0	0	IND	0	0	0
hexamethylphosphoramide,	0	0	0	IND	0	1	0	1
trifluoroiodomethane	0	0	0	IND	0	0	1	1
1,4-dihydroxybenzene	0	0	0	0	IND	0	1	1
5-allyl-1,3-benzodioxole (safrole)	1	0	0	0	0	0	IND	1
cycloheximide	1	0	0	0	0	0	0	1
fenthion	0	1	0	0	0	0	1	2
furan	1	0	0	0	0	0	1	2
pyrogallol	1	0	0	IND	IND	0	1	2
phenolphthalein	1	0	0	0	IND	1	0	2
(±) tetrahydrofurfuryl (R)-2-[4-(6-chloroquinoxalin-2-yloxy)phenyloxy]propionate	1	0	0	0	1	0	0	2
2-(isocyanatosulfonylmethyl)benzoic acid methyl ester	1	0	1	IND	0	0	0	2
5-(2,4-dioxo-1,2,3,4-tetrahydropyrimidine)-3-fluoro-2-hydroxymethyltetrahydrofuran	1	0	0	IND	1	0	0	2
O-hexyl-N-ethoxycarbonylthiocarbamate	1	0	0	0	1	0	0	2
4'-ethoxy-2-benzimidazoleanilide	1	0	1	0	0	0	0	2
N-(2,3-dihydroxypropoxymethyl)-2-methylacrylamide	1	0	0	IND	1	0	0	2

Chemical name	Toxtree	TOPKAT	CAESAR	ToxBoxes	HazardExpert	Lazar	Derek	No Correct
methyl acrylamidoglycolate	1	0	0	IND	1	0	1	3
colchicine	1	1	0	0	0	0	1	3
N-[6,9-dihydro-9-[[2-hydroxy-1-(hydroxymethyl)ethoxy]methyl]-6-oxo-1H-purin-2-yl]acetamide	1	0	1	IND	0	0	1	3
monocrotophos	1	0	1	IND	0	1	0	3
chloro-1-ethylcyclohexyl carbonate	1	0	1	0	1	0	IND	3
aniline	1	0	0	0	1	1	0	3
trichloroethylene	0	0	1	IND	0	1	1	3
1,3-Bis(vinylsulfonylacetamido)propane	1	1	0	0	IND	0	1	3
2-methyl-N-(2-methylacryloylaminomethoxymethyl)-acrylamide	1	1	0	IND	1	0	0	3
1,3-butadiene	0	1	0	1	0	0	1	3
methyl acrylamidomethoxyacetate	1	0	0	IND	1	1	1	4
benomyl	1	0	0	0	1	1	1	4
thiophanate-methyl	1	0	1	1	1	0	0	4
glyoxal	1	1	1	IND	0	0	1	4
4,4'-bis(dimethylamino)benzophenone	1	0	1	IND	1	0	1	4
4-chloro-o-toluidine	1	0	1	1	1	0	0	4
methacrylamide	1	1	1	0	1	0	0	4
O-isobutyl-N-ethoxycarbonylthiocarbamate	1	0	1	IND	1	1	0	4
N-[2,3-bis-(2-methylacryloylaminomethoxy)propoxymethyl]-2-methylacrylamide	1	1	1	IND	1	0	0	4

Chemical name	Toxtree	TOPKAT	CAESAR	ToxBoxes	HazardExpert	Lazar	Derek	No Correct
2-nitrotoluene	1	0	1	IND	1	1	1	5
4,6-dinitro-o-cresol	1	0	1	1	1	1	0	5
1-phenylazo-2-naphthol	1	0	1	IND	1	1	1	5
4-aminophenol	1	1	0	IND	1	1	1	5
4-ethoxyaniline	1	1	1	IND	1	1	0	5
3-chloro-4-(3-fluorobenzyloxy)aniline	1	1	1	IND	1	0	1	5
9-vinylcarbazole	0	1	1	1	1	1	0	5
phosphamidon	1	1	1	1	0	1	0	5
carbendazim	1	1	1	IND	1	1	1	6
acrylamide	1	1	1	0	1	1	1	6
(4-hydrazinophenyl)-N-methylmethanesulfonamide hydrochloride	1	1	0	1	1	1	1	6
2,4-dinitrotoluene	1	0	1	1	1	1	1	6
2,6-dinitrotoluene	1	0	1	1	1	1	1	6
2,3-dinitrotoluene	1	0	1	1	1	1	1	6
3,4-dinitrotoluene	1	0	1	1	1	1	1	6
3,5-dinitrotoluene	1	0	1	1	1	1	1	6
2,5-dinitrotoluene	1	0	1	1	1	1	1	6
2-aminophenol	1	1	0	1	1	1	1	6
2-methoxyaniline	1	1	1	1	1	1	0	6

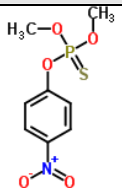
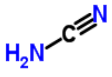
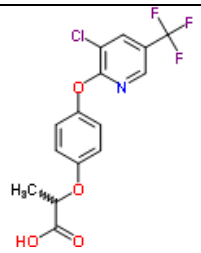
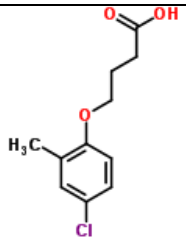
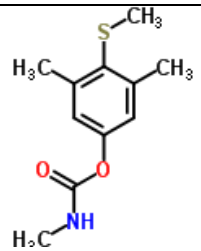
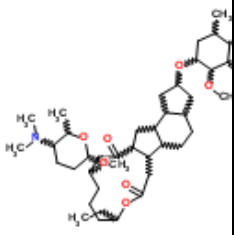
Chemical name	Toxtree	TOPKAT	CAESAR	ToxBoxes	HazardExpert	Lazar	Derek	No Correct
N,N'-diacetylbenzidine	1	0	1	1	1	1	1	6
2,4-toluenediamine	1	1	1	1	1	1	0	6
quinoline	0	1	1	1	1	1	1	6
N,N',N''-tris(2-methyl-2,3-epoxypropyl)-perhydro-2,4,6-oxo-1,3,5-triazine	1	0	1	1	1	1	1	6
diethyl sulphate	1	1	1	1	0	1	1	6
oxirane	1	1	1	1	1	0	1	6
methyloxirane (propylene oxide)	1	1	1	1	1	0	1	6
1,2:3,4-diepoxybutane (2,2'-bioxirane)	1	1	1	1	1	0	1	6
(2-chloroethyl)(3-hydroxypropyl)ammonium chloride	1	1	1	1	0	1	1	6
aziridine	1	1	1	1	1	0	1	6
1,3,5-tris(oxiranylmethyl)-1,3,5-triazine-2,4,6(1H,3H,5H)-trione	1	0	1	1	1	1	1	6
1,3,5-tris-[(2S and 2R)-2,3-epoxypropyl]-1,3,5-triazine-2,4,6-(1H,3H,5H)-trione	1	0	1	1	1	1	1	6
isobutyl nitrite	1	1	1	1	0	1	1	6
dimethyl sulphate	1	1	1	1	0	1	1	6
2,3-dichloropropene	1	1	1	1	0	1	1	6
butyl glycidyl ether	1	1	1	1	1	0	1	6
2,3-epoxypropan-1-ol (oxiranemethanol)	1	1	1	1	1	0	1	6
1,2-epoxy-3-phenoxypropane (phenyl glycidyl ether)	1	0	1	1	1	1	1	6
2,3-epoxypropyltrimethylammonium chloride	1	1	1	1	1	0	1	6
1-chloro-4-nitrobenzene	1	0	1	1	1	1	1	6
N,N,N',N'-tetraglycidyl-4,4'-diamino-3,3'-diethyldiphenylmethane	1	1	1	1	1	0	1	6
trimethylopropane tri(3-aziridinylpropanoate)	1	1	1	1	1	0	1	6
R-2,3-epoxy-1-propanol	1	1	1	1	1	0	1	6

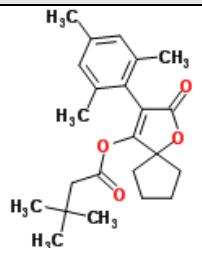
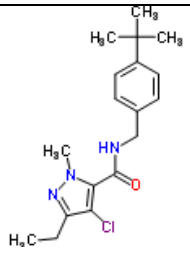
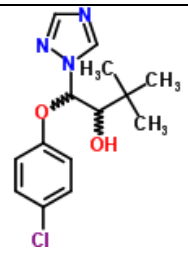
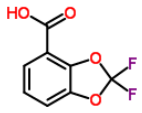
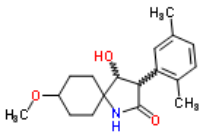
Chemical name	Toxtree	TOPKAT	CAESAR	ToxBoxes	HazardExpert	Lazar	Derek	No Correct
benzo[a]pyrene	1	1	1	1	1	1	1	7
p-aminophenyl ether	1	1	1	1	1	1	1	7
chrysene	1	1	1	1	1	1	1	7
(3-chlorophenyl)-(4-methoxy-3-nitrophenyl)methanone	1	1	1	1	1	1	1	7
azobenzene	1	1	1	1	1	1	1	7
4-(phenylazo)benzene-1,3-diamine	1	1	1	1	1	1	1	7
4-phenylazophenylene-1,3-diamine monohydrochloride	1	1	1	1	1	1	1	7
4,4'-diaminodiphenylmethane	1	1	1	1	1	1	1	7
2,6-toluenediamine	1	1	1	1	1	1	1	7
o-phenylenediamine	1	1	1	1	1	1	1	7
o-phenylenediamine dihydrochloride	1	1	1	1	1	1	1	7
m-phenylenediamine	1	1	1	1	1	1	1	7
m-phenylenediamine dihydrochloride	1	1	1	1	1	1	1	7
2,4-diaminoanisole	1	1	1	1	1	1	1	7
diaminobenzidine	1	1	1	1	1	1	1	7
ethidium bromide	1	1	1	1	1	1	1	7
1,2-dibromo-3-chloropropane	1	1	1	1	1	1	1	7
Bromomethane	1	1	1	1	1	1	1	7
3-chloropropene	1	1	1	1	1	1	1	7
allyl 2,3-epoxypropyl ether	1	1	1	1	1	1	1	7
[p-tolyloxy)methyl]oxirane	1	1	1	1	1	1	1	7
[m-tolyloxy)methyl]oxirane	1	1	1	1	1	1	1	7
2,3-epoxypropyl o-tolyl ether	1	1	1	1	1	1	1	7
[(tolyloxy)methyl]oxirane	1	1	1	1	1	1	1	7
resorcinol diglycidyl ether	1	1	1	1	1	1	1	7
6-glycidyloxynapht-1-yl oxymethyloxirane	1	1	1	1	1	1	1	7
4-nitrosophenol	1	1	1	1	1	1	1	7
2-butenal (crotonaldehyde)	1	1	1	1	1	1	1	7
Phenylhydrazine	1	1	1	1	1	1	1	7
oxiranemethanol,4-methylbenzene-sulfonate	1	1	1	1	1	1	1	7

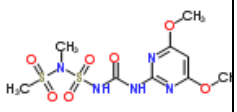
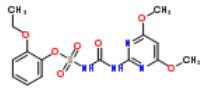
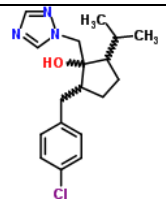
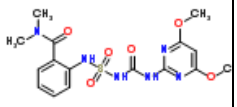
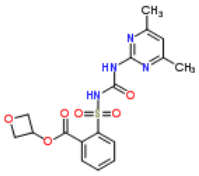
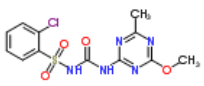
0=negative; 1=positive; IND=indeterminate

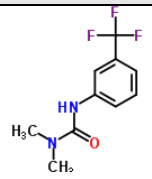
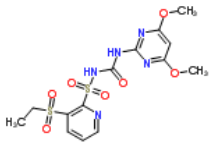
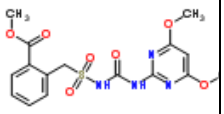
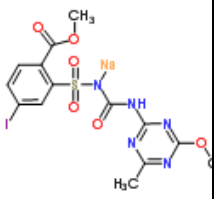
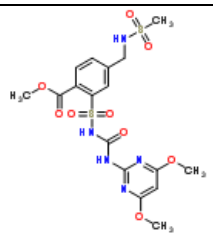
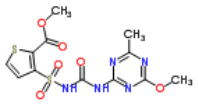
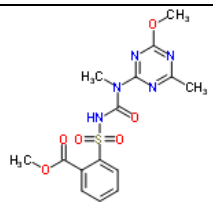
Appendix 5. Best and worst predicted pesticides in the CRD-AGES dataset

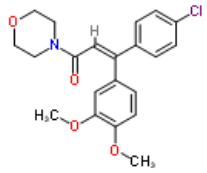
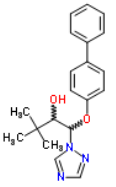
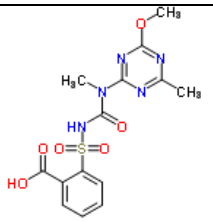
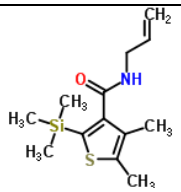
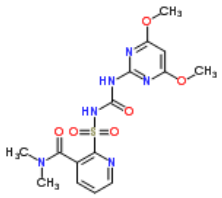
Best predicted pesticides (mutagenicity)

Name	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	ToxBox	Lazar 6512	Lazar Kazius
Parathion-methyl		1	1	1	1	1	1	Equiv	1	1
Cyanamide		0	0	0	0	0	0	0	0	0
Haloxyfop R		0	0	0	0	0	0	0	0	0
MCPB		0	0	0	0	0	0	0	0	0
Methiocarb		0	0	0	0	0	0	0	0	0
Spinosad		0	0	0	0	0	0	0	0	0

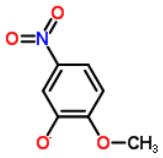
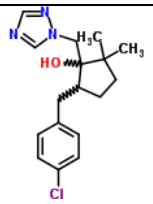
Name	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	ToxBox	Lazar 6512	Lazar Kazius
Spiromesifen		0	0	0	0	0	0	0	0	0
Tebufenpyrad		0	0	0	0	0	0	0	0	0
Triadimenol		0	0	0	0	0	0	0	0	0
CGA192155		0	0	0	0	0	0	0	0	0
CGA339833		0	0	0	0	0	0	0	0	0
BYI 08330-desmethyl-ketohydroxy		0	0	0	0	0	0	0	0	0
BYI 08330-dihydroxy		0	0	0	0	0	0	0	0	0
BYI 08330-ketohydroxy		0	0	0	0	0	0	0	0	0
BYI 08330-monohydroxy		0	0	0	0	0	0	0	0	0

Name	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	ToxBox	Lazar 6512	Lazar Kazius
Amidosulfuron		0	0	0	0	0	0	0	0	0
Ethoxysulfuron		0	0	0	0	0	0	0	0	0
Ipconazole		0	0	0	0	0	0	0	0	0
Orthosulfamuron		0	0	0	0	0	0	0	0	0
Oxasulfuron		0	0	0	0	0	0	0	0	0
Prosulfuron		0	0	0	0	0	0	0	0	0
Chlorsulfuron		0	0	0	0	0	0	0	0	0

Name	Structure	Exp data	Derek	Hazard Expert	CAESAR	Topkat	Toxtree	ToxBox	Lazar 6512	Lazar Kazius
Fluometuron		0	0	0	0	0	0	0	0	0
Rimsulfuron		0	0	0	0	0	0	0	0	0
Bensulfuron-methyl		0	0	0	0	0	0	0	0	0
Iodosulfuron-methyl sodium		0	0	0	0	0	0	0	0	0
Mesosulfuron-methyl		0	0	0	0	0	0	0	0	0
Thifensulfuron (-methyl)		0	0	0	0	0	0	0	0	0
Tribenuron methyl		0	0	0	0	0	0	0	0	0

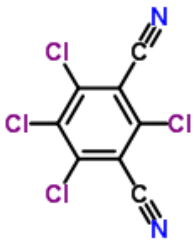
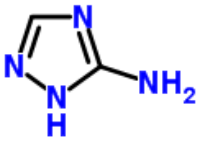
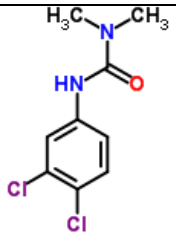
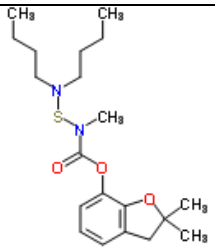
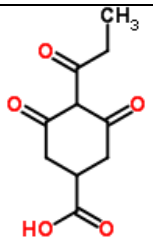
Name	Structure	Exp data	Derek	Hazard Expert	CAESAR	Topkat	Toxtree	ToxBox	Lazar 6512	Lazar Kazius
Dimethomorph		0	0	0	0	0	0	0	0	0
Thifensulfuron		0	0	0	0	0	0	0	0	0
Tribenuron		0	0	0	0	0	0	0	0	0
Silthiofam		0	0	0	0	0	0	0	0	0
Nicosulfuron		0	0	0	0	0	0	0	0	0

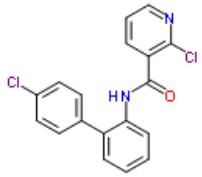
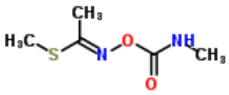
Worst predicted pesticides (mutagenicity)

Name (Outcome)	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	ToxBoxes	Lazar 6512	Lazar Kazius
Sodium nitroguiacolate (FP)		0	1	1	1	1	1	Equiv	1	1
Metconazole (ambiguous)		0 (AGES) 1 (TTC)	0	0	0	0	0	Equiv	0	0

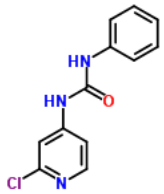
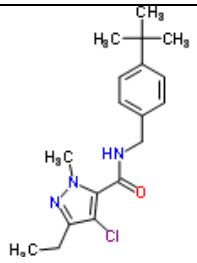
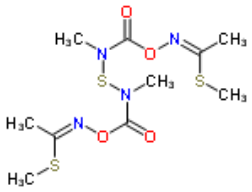
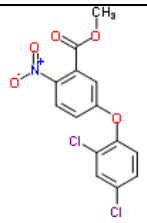
FP=false positive

Best predicted pesticides (carcinogenicity)

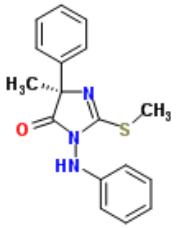
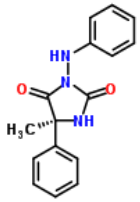
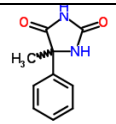
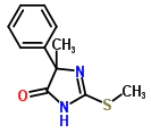
Name	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	Lazar
Chlorothalonil		1	1	1	1	1	0	1
Amitrole		1	0	1	1	1	1	1
Diuron		1	1	1	0	1	1	1
Carbosulfan		0	0	0	0	0	0	0
Methiocarb		0	0	0	0	0	0	0
Prohexadione calcium		0	0	0	0	0	0	0

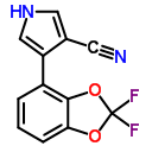
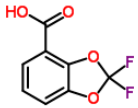
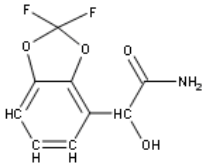
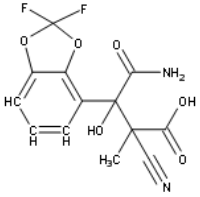
Name	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	Lazar
Boscalid		0	0	0	0	0	0	0
Thiodicarb_Methomyl		0	0	0	0	0	0	0

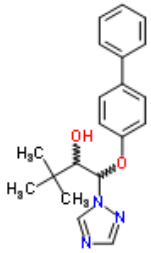
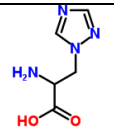
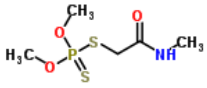
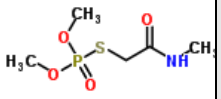
Worst predicted pesticides (carcinogenicity)

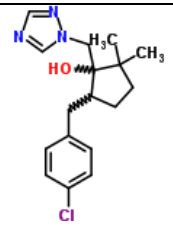
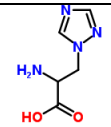
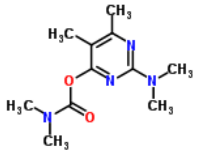
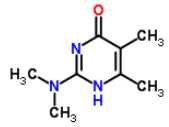
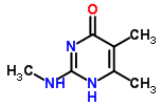
Name (Outcome)	Structure	Exp data	Derek	Hazard Expert	Caesar	Topkat	Toxtree	Lazar
Forchlorfenuron		1	0	0	0	0	0	0
Tebufenpyrad		1	0	0	0	0	0	0
Thiodicarb		1	0	0	0	0	0	0
Bifenox		0 (equivocal in rat)	1	1	1	1	1	1

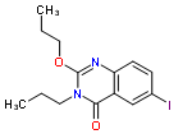
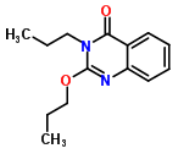
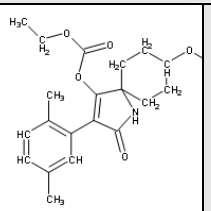
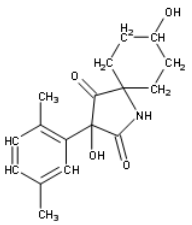
Appendix 6. Data gap filling for pesticide metabolites

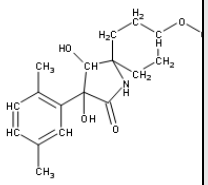
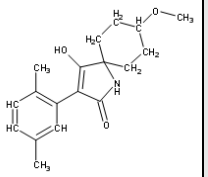
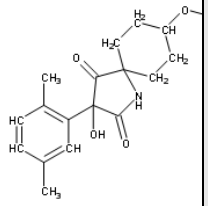
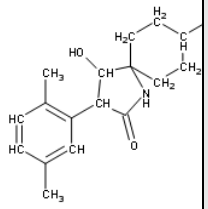
Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
fenamidone		**0**	0	1	1	0	1	0	1	1	**0**	0	1	1	0	1	0
Met1: RPA405862		**0**	0	1	1	1	1	0	1	1	**ND**	0	1	1	1	1	1
Met2: RPA717879		**0**	0	0	1	0	0	0	0	0	**ND**	1	0	0	0	0	0
Met3: RPA408056		**0**	0	0	0	1	0	0	0	0	**ND**	0	1	1	0	0	0

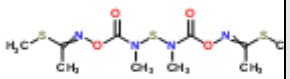
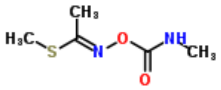
Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
fludioxonil		**0**	0	0	0	0	0	EQ	0	0	**0**	0	0	0	0	0	1
Met1: CGA192155		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	0	0	0	0
Met2: CGA308103		**0**	0	0	1	0	0	0	0	0	**ND**	0	0	0	0	0	0
Met3: CGA339833		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	0	1	0	0

Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
bitertanol		**0**	0	0	0	0	0	0	1	1	**0**	0	0	0	1	0	1
Met1: triazolylalanine		**0**	0	0	0	0	0	0	1	0	**ND**	0	0	0	0	0	0
dimethoate		**0**	1	0	1 In TS	1 In TS	0	EQ In TS	0 In TS	0 In TS	**0**	0	EQ	0 In TS	0 In TS	0	0 In TS
Met1: omethoate		**ND**	1	0	1	1	0	0	0	0	**ND**	0	EQ	0	0	0	0

Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
metconazole		**0 (CRD) 1 (AGES)**	0	0	0	0	0	EQ	0	0	**1**	0	0	1	0	1	0
Met1: triazolylalanine		**0**	0	0	0	0	0	0	1	0	**ND**	0	0	0	0	0	0
pirimicarb		**0**	0	0	1	0	1	0	0	0	**1**	1	0	0	1	1	0
Met1: R31805		**0**	0	EQ	1	0	1	EQ	0	0	**ND**	1	0	0	0	1	1
Met2: R34865		**0**	0	EQ	0	0	1	EQ	0	0	**ND**	1	0	1	1	1	1

Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
proquinazid		**0**	0	0	0	0	0	EQ	0	1	**1**	0	0	0	1	0	ND
Met1: IN-MM671		**0**	0	0	0	0	0	0	1	1	**ND**	0	0	0	1	0	ND
spirotetramat		**0**	0	0	0	0	1	0	0	ND	**0**	0	0	0	0	1	0
Met1: BYI 08330-desmethyl-ketohydroxy		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	1	1	0	0

Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
Met2: BYI 08330-dihydroxy		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	0	ND	0	0
Met3: BYI 08330-enol		**0**	0	0	0	0	1	0	0	ND	**ND**	0	0	0	0	1	0
Met4: BYI 08330-ketohydroxy		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	0	1	0	0
Met5: BYI 08330-monohydroxy		**0**	0	0	0	0	0	0	0	0	**ND**	0	0	0	0	0	0

Name	Structure	Mut exp	Derek (mut)	Hazard Expert (mut)	Caesar (mut)	Topkat (mut)	Toxtree (mut)	Tox Boxes (mut)	Lazar N6512 (mut)	Lazar Kazius (mut)	Carc exp	Derek (carc)	Hazard Expert (carc)	Caesar (carc)	Topkat (carc)	Toxtree (carc)	Lazar
thiodicarb		**1**	0	0	1	0	0	1	0	1	**1**	0	0	0	0	0	0
Met1: methomyl		**0**	0	0	0	0	0	EQ	0	1	**0**	0	0	0	0	0	0

EQ – Equivocal; In TS – compound in model training set; ND – not determined; 0 – non-toxic; 1 - toxic

European Commission

EUR 24640 EN – Joint Research Centre – Institute for Health and Consumer Protection
Title: The Applicability of Software Tools for Genotoxicity and Carcinogenicity Prediction: Case Studies relevant to the Assessment of Pesticides
Author(s): Andrew Worth, Silvia Lapenna, Elena Lo Piparo, Aleksandra Mostrag-Szlichtyng and Rositsa Serafimova
Luxembourg: Publications Office of the European Union
2010 – 75 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-18762-9
doi:10.2788/61485

Abstract

This report presents research results obtained in the framework of a project on the Applicability of Quantitative Structure-Activity Relationship (QSAR) analysis in the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances. During this project, which was funded by the European Food Safety Authority (EFSA), the Joint Research Centre (JRC) performed several investigations to evaluate the comparative performance of selected software tools for genotoxicity and carcinogenicity prediction, and to develop a number of case studies to illustrate the opportunities and difficulties arising in the computational assessment of pesticides. This exercise also included an investigation of the chemical space of several pesticides datasets. The results indicate that different software tools have different advantages and disadvantages, depending on the specific requirements of the user / risk assessor. It is concluded that further work is needed to develop acceptance criteria for specific regulatory applications (e.g. evaluation of pesticide metabolites) and to develop batteries of models fulfilling such criteria.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24640-EN-C

Publications Office

ISBN 978-92-79-18762-9

9 789279 187629