

Environmental Project No. 636 2001
Miljøprojekt

Report on the Advisory list for selfclassification of dangerous substances

The Danish Environmental Protection Agency will, when opportunity offers, publish reports and contributions relating to environmental research and development projects financed via the Danish EPA.

Please note that publication does not signify that the contents of the reports necessarily reflect the views of the Danish EPA.

The reports are, however, published because the Danish EPA finds that the studies represent a valuable contribution to the debate on environmental policy in Denmark.

Contents

SUMMARY	5
1 BACKGROUND, CONTENTS, AND USE OF THE LIST	7
1.1 BACKGROUND	7
1.2 QSAR MODELS - AN ALTERNATIVE METHOD FOR ASSESSMENT OF DANGER	8
1.3 THE ADVISORY LIST FOR SELFCLASSIFICATION OF DANGEROUS SUBSTANCES	9
1.4 THE DUTY OF MANUFACTURERS AND IMPORTERS TO CARRY OUT SELFCLASSIFICATION	11
2 TECHNICAL DESCRIPTION OF THE CREATION OF THE LIST AND THE QSAR MODELS USED	13
2.1 INTRODUCTION	13
2.1.1 SAR / QSAR	13
2.1.2 The domain of the models	14
2.1.3 Accuracy of the model predictions	14
2.1.4 Software	15
2.2 METHODOLOGY IN MAKING THE LIST	17
2.2.1 The selected dangerous properties	17
2.2.2 The evaluated chemical substances	17
2.2.3 Test data	17
2.2.4 Use of QSAR models	18
2.2.5 The result	18
2.3 ACUTE ORAL TOXICITY	19
2.4 SENSITIZATION BY SKIN CONTACT	21
2.5 MUTAGENICITY	24
2.6 CARCINOGENICITY	27
2.7 DANGER TO THE AQUATIC ENVIRONMENT	31
3 REFERENCES	35

Summary

This report features a description of the Danish Environmental Protection Agency's (EPA) Advisory list for selfclassification of dangerous substances. The substances have been identified by means of computer models, so-called QSAR models (Quantitative Structure-Activity Relationship). The list is intended as an aid to producers / importers in their selfclassification.

Part I of this report features a description of the background of the list, its contents, and its application. Part II comprises a technical description of the QSAR models used, the creation of the list, and its relationship to the criteria for classification of selected dangerous properties. The list can be found on the Danish EPA's homepage (www.mst.dk) under the heading "chemicals".

With the aid of QSAR models, the Danish EPA has examined approximately 47,000 chemical substances, identifying 20,624 substances which are deemed to require classification for one or more of the following dangerous properties: Acute oral toxicity, sensitization by skin contact, mutagenicity, carcinogenicity, and danger to the aquatic environment.

According to classification criteria, classification should be carried out on the basis of the knowledge available, which is most often from the results of laboratory tests on animals. However, in the experience of the Danish EPA, manufacturers / importers find it difficult to comply with their duty to assess whether a substance they wish to introduce to the market should be classified because of lack of available data. The fact is that only very little information is available on the dangerous properties of chemical substances. The Danish EPA estimates that for approximately 90 per cent of all substances, only few or no test results from animal testing etc. are available on any dangerous properties to humans or the environment.

In addition to results from animal testing, the criteria for classification also provide opportunities for using alternative methods. This could for instance be studies which do not require the use of laboratory animals, but are based on comparisons with other similar chemicals by so-called structure-activity relationships.

QSAR modelling is such an alternative method to assess the potential danger of chemical substances. For several years now, the Danish EPA has carried out work to develop and apply QSAR models in order to predict the properties of chemical substances. The models used here are now so reliable that they are able to predict whether a given substance has one or more of the properties selected with an accuracy of approximately 70-85 per cent.

In spite of the general lack of data, reliable information on the dangerous properties of substances from suitable animal testing, etc. might be available for some substances found in the Advisory list for selfclassification of dangerous substances. To the extent that this is the case, such information should be employed for selfclassification in preference to the recommendations of this list.

It should be emphasized that the list is not binding. The responsibility for carrying out correct classification still rests with the manufacturer / importer. The Danish EPA calls upon importers/manufacturers to use the Advisory list for selfclassification of dangerous substances as a tool in their assessment of the dangerous properties of chemicals in cases of insufficient or no data for the selected dangerous properties.

1 Background, contents, and use of the list

1.1 BACKGROUND

When chemical substances are to be classified in terms of the danger they represent, their inherent properties are assessed on the basis of the knowledge and information available /1,57/. Such assessment is often carried out on the basis of results from animal testing. Assessment must be carried out individually for each property, which means that extensive animal testing may be required for a single substance. Thus, complete identification of all the properties that are classified at present can entail up to 30 animal studies on animals for just one substance.

Studies have shown that very little information is available on the danger posed to human beings and the environment by chemical substances in the European market. In 1999, the European Commission assessed the scope of available test data for substances which are available on the market in large quantities (more than 1,000 tonnes per manufacturer/importer per year in the EU). The Commission found that the minimum information on dangerous properties of substances required under EU regulations in order to carry out risk assessment of industrial chemicals was only available for 14 per cent of all the substances studied. For 21 per cent of all substances, no test data at all was available as regards their toxicity towards human beings or the environment /2/.

In 2000, the Danish EPA carried out a study to determine the extent of the data available on the danger, presented by the approximately 100,000 substances in the EU Inventory of Existing Substances* /3/, in two of the world's largest sources of publicly available test data (RTECS, 2000**; AQUIRE, 1994***). This study showed that test data on selected types of effects were available for the following percentages of all Eines substances /4/:

TABLE 1

Acute toxicity	13.4 per cent
Toxic to reproduction	2.5 per cent
Mutagenicity	3.9 per cent
Carcinogenicity	1.8 per cent
Danger to the aquatic environment	3.5 per cent

* Eines: European Inventory of Existing Chemical Substances: Inventory of substances which were reported by industry as being present within the European market during the period 1971 to 1981.

** RTECS: Registry of Toxic Effects of Chemical Substances, The National Institute of Occupational Safety and Health, Washington, D.C.

*** AQUIRE: AQUatic toxicity Information Retrieval, United States Environmental Protection Agency, Mid-Continent Ecology Division, Duluth, MN.

Thus, in the assessment of the Danish EPA, information on the dangerous properties of chemical substances is at present incomplete or absent for approximately 90 per cent of all substances listed in the EINECS. This means that many chemical substances within the European market can have unknown dangerous properties even though they have been used for many years. Issues regarding animal ethics and financial considerations mean that it is unlikely that test data on the dangerous properties of these substances will be available within the foreseeable future.

The criteria on classification describe how available experimental test data (from animal testing, etc.) should be used in assessment and classification of the toxicity of substances for human beings and the environment. These criteria also describe how the danger presented by substances can be assessed by means of comparison to other, similar substances with known toxic properties (SARs, Structure-Activity Relationships). Finally, these criteria include the use of expert judgements, e.g. from practical experience of a given substance, as the basis for classification /1,57/.

In Denmark as well as internationally in the EU and the OECD the importance of developing alternative methods, which are not based on animal testing, are emphasized. Lower organisms such as algae and bacteria are already being used in tests for certain properties, and today good results have been achieved by means of alternative tests rather than tests on animals. A test method for skin irritation, which does not require the use of living animals, was recently added to the rules on classification /1,57/. As regards many dangerous properties, however, efforts made to discover suitable methods for testing which do not require use of laboratory animals have until now not succeeded.

1.2 QSAR MODELS - AN ALTERNATIVE METHOD FOR ASSESSMENT OF DANGER

Quantitative structure-activity relationships (QSAR models) can be used for assessment of dangerous properties as an alternative to animal testing. A QSAR model relates an effect with molecular descriptors found to be tied to this effect. Using information on the relevant molecular descriptors the models can predict effect for substances without test data. By using the ability of computers to go through large quantities of information, QSAR models have in this project been used to assess a big number of substances.

The principle behind structure-activity relationships is that substances with comparable structures possess similar properties. SARs and QSARs are well-known tools for assessment of chemical substances. These tools are used by authorities in the USA and the EU, as well as by industry, to assess physico-chemical, toxicological, and eco-toxicological properties and to predict the fate of substances in the environment.

The criteria for EU classification include the possibility of using expert judgements as well as conclusions based on structural analogies /1,57/. SARs and QSARs have been used for classification of effects on the aquatic environment in cases where no test data on toxicity or degradation in the aquatic environment were available. As regards classification for impacts on human health, SARs have been applied in specific cases, and this tool was recently used in a discussion of two special properties: Narcotic effect and

defatting properties.

1.3 THE ADVISORY LIST FOR SELFCLASSIFICATION OF DANGEROUS SUBSTANCES

The Danish EPA has carried out work on QSAR models for several years, an area that continues to develop. At present, the Danish EPA has access to reliable models which are capable of predicting whether a substance possesses one or more of the dangerous properties selected in this context. The substances on this list have been assessed for the following dangerous properties: Acute oral toxicity, sensitization by skin contact, mutagenicity, carcinogenicity, and danger to the aquatic environment. According to validation results the models available to the Danish EPA identify the substances which possess these properties with a degree of accuracy of approximately 70 - 85 per cent, depending on the model used.

The basis for the list was the European Inventory of Existing Substances, Einecs. For technical reasons, the QSAR models can only assess chemical substances with unambiguous chemical structure, so-called discrete substances. The Danish EPA has used validated QSAR models to carry out a systematic assessment of the approximately 47,000 discrete organic substances in Einecs. Also, the approximately 7,000 chemical substances which have already been classified by EU authorities, have not been included in the assessment*.

The criteria for computer-model selection of substances for a given property have been defined to match the criteria for classification of chemical substances as closely as possible. /1/. For properties, where the criteria are open to interpretation, such definitions have been specified in accordance with the Danish EPA's best judgement with a view to providing the public with an operative list. The preparation of this list is described in more detail in Part II.

The result of the computer-based assessment is this Advisory list for selfclassification of dangerous substances, which comprises 20,624 chemical substances with suggested classifications for one or more of the dangerous properties selected.

By making this Advisory list for selfclassification of dangerous substances available to the public, the Danish EPA wishes to offer manufacturers / importers a tool which can be used when carrying out selfclassification of chemical substances for those dangerous properties which are included in the list. Enterprises are encouraged to include the advisory classifications provided in this list in their assessment of chemical substances where no

* These substances can be found in the List of dangerous substances /5/. As far as possible, such substances have been removed from consideration prior to the assessment of the substances for the Advisory list for selfclassification of dangerous substances. However, substances may appear in both lists. This is due to the fact that no official overview of the substances covered by the group entries in the List of dangerous substances is available. Another reason is that a single chemical may be found under the heading of several CAS numbers. Where substances listed in the Advisory list for selfclassification of dangerous substances also appear in the List of dangerous substances, the recommended classification provided in this list must be discounted in favour of classification as indicated in the List of dangerous substances.

results from animal testing or other reliable data on the relevant dangerous properties are available.

The selected dangerous properties and classifications are listed in Table 2.

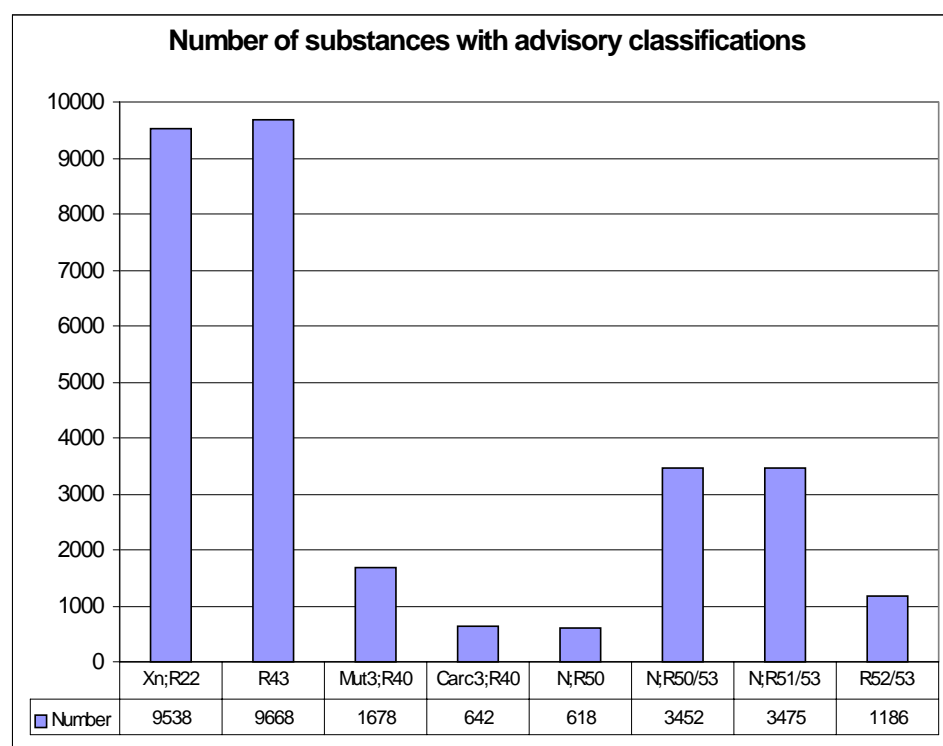
TABLE 2

Dangerous property	Classification	Wording of classification
Acute oral toxicity	Xn;R22	Harmful; harmful if swallowed
Sensitization by skin contact	R43	May cause sensitization by skin contact
Mutagenicity	Mut3;R40	Mutagen, category 3; possible risk of irreversible effects
Carcinogenicity	Carc3;R40	Carcinogen, category 3; possible risk of irreversible effects
Danger to the aquatic environment	N;R50	Dangerous for the environment; very toxic to aquatic organisms
	N;R50/53	Dangerous for the environment; very toxic to aquatic organisms, may cause long-term adverse effects in the aquatic environment
	N;R51/53	Dangerous for the environment; toxic to aquatic organisms, may cause long-term adverse effects in the aquatic environment
	R52/53	Harmful to aquatic organisms, may cause long-term adverse effects in the aquatic environment

For each substance on the list, the following information is included in addition to the advisory classifications: Eines name in Danish and English, Eines number, and CAS number.

Figure 1 shows how many of the 20,624 substances in this list have been included with advisory classifications for each dangerous property.

FIGURE 1:



It should be noted that the Advisory list for selfclassification of dangerous substances is not an exhaustive list of all the dangerous substances in the EU Inventory of Existing Substances (Einecs). As was mentioned above, the substances assessed here comprise only approximately half of all substances in Einecs.

At the same time, the list cover only those selected dangerous properties which feature the most reliable computer-generated predictions. Therefore these substances may well possess other dangerous properties.

Finally, for each of the selected dangerous properties, only the substances for which the model predictions are most reliable, have been included in the list. As a result, substances that were assessed but not included in this list may well possess one or more of the dangerous properties selected.

Similarly, if a substance is included in this list and does not have an advisory classification for e.g. carcinogenicity, the substance can nevertheless have this property. The reason for this could be that the models for carcinogenicity applied do not have good coverage for this specific chemical substance.

If a substance is not included in the list, or it is on the list but without one or more advisory classifications, this can then be due to the models predicting that the substance does not possess these dangerous properties, or it can be because the models are not able to give a good prediction in these cases.

Finally, out of the substances that a model cover, it can sometimes erroneously estimate substances as not having a property which they in fact do have (false negatives). With other substances, the models will attribute a specific property to a substance, which actually does not possess that property (false positive). The Advisory list for selfclassification of dangerous substances can be used to identify substances that do possess dangerous properties, well knowing that that some predictions will be false positives. If the list had contained negative predictions, a part of these would also be incorrect (false negatives). By this, substances which in reality possess dangerous properties would be advised not to be classified for this.

This list, only containing positive predictions, can not be used to “acquit” substances of dangerous properties.

1.4 THE DUTY OF MANUFACTURERS AND IMPORTERS TO CARRY OUT SELFCLASSIFICATION

Manufacturers or importers are responsible for investigating the properties of chemical substances and for classifying them in accordance with their inherent dangerous properties before marketing them. Such selfclassification must be carried out on the basis of available information on substances in accordance with the criteria of the Statutory Order on Classification /1,57/.

As regards the approximately 7,000 substances for which harmonised classification has been adopted, the classification of the List of dangerous substances shall be applied /5/. For the remaining approximately 93,000 of the 100,000 substances in the EU Inventory of Existing Substances /2/, importers/manufacturers are obliged to assess whether such a substance should be classified as dangerous (selfclassification). Selfclassification must be

carried out in accordance with the criteria in Appendix 1 of the Statutory Order on Classification.

The Advisory list for selfclassification of dangerous substances is intended as a tool to help manufacturers / importers fulfil their duty to carry out correct classification in those cases where no other information is available on a given substance. When preparing this list, the Danish EPA has not examined whether data on individual substances is available in literature. The duty to map available information on substances for selfclassified lies with manufacturers / importers. Reliable test results or relevant specialist knowledge on specific substances should always be used in preference to computer predictions. This is to say that where such information - which runs contrary to the recommendations of this list - is available, it should be used instead of the classifications featured in this advisory list. At the same time, it should be emphasised that this advisory list includes only some of the dangerous properties which must be considered by manufacturers / importers in their assessment of substances. Manufacturers / importers should also carry out assessment of other properties regarding flammability, explosivity, and danger to human health and the environment.

Use of the list

It is recommended that the list be used for selfclassification in the following way:

1. Examine if the substance is on the List of dangerous substances /5/. If so it should be classified accordingly.
2. If the substance is not on the List of dangerous substances, it should be classified according to the criteria in the Statutory Order on Classification /1/.
3. In cases of an otherwise insufficient data basis for selfclassification, it is recommended to use the Advisory list for selfclassification of dangerous substances.

2 Technical description of the creation of the list and the QSAR models used

2.1 INTRODUCTION

In a field developing as rapidly as QSAR's are today, there will always be better models, better validations and new endpoints becoming available - and consequently never a "right" time to release advisory classification based on them. It is however, felt that considerable information has been accumulated which can now be of help in the otherwise difficult task of assessing the toxicology of many thousands of otherwise untested chemicals. This knowledge may also be of assistance in helping to direct future testing plans to the areas for which it is most urgently needed.

2.1.1 SAR / QSAR

The concept that similar structures will have similar properties is not new. Already in the 1890's it was discovered, for example, that the anaesthetic potency of substances to aquatic organisms was related to their oil/water solubility ratios, a relationship which led to the use of LogP (octanol/water) as a prediction of this effect. Today it is known that all chemicals will exhibit a minimum or "basal" narcotic effect, which is related to their absorption to cell membranes, and which is well predicted by their lipophilic profile.

SARs and QSARs ((Quantitative) Structure Activity Relationships) are based on a comparison of the structure and physico-chemical properties (descriptors) with measured parameters or endpoints for a range of chemicals called a training set.

The endpoint may for instance be another physical-chemical property or it may be a biological effect. The descriptors may include LogP, molecular indices, quantum mechanical properties, shape, size, charge, distributions, etc. The comparison is often made with statistical tools. The goal is to determine which descriptor(s) are in an essential way connected with the endpoint in question, and to set up a relationship between these descriptors and the endpoint.

When the result is expressed qualitatively the relationship is a SAR, and when the result is expressed quantitatively the relationship is a QSAR. A QSAR is a relation between the quantitative descriptors for chemical substances and a more or less graduated scale of property or effect.

Once a correlation between structure / properties is established it can be used for predictions of the endpoints for other chemicals, for which the descriptors are known or can be estimated. In general, development and use of the correlations are done by computers.

2.1.2 The domain of the models

The domain limits the QSARs use to the endpoint being modelled and the group of substances for which it is valid. The domain of the QSAR is defined in the selection of the training set; the coverage of the descriptors of the training set define the “area” of “the chemical universe” for which the model is valid.

2.1.3 Accuracy of the model predictions

In order to check a models predictive ability it should be validated. Validation is a trial of the model performance for a set of substances independent of the training set, but within the domain of the model. The model predictions for these substances are compared with measured endpoints for the substances in order to establish the accuracy of the models.

Ideally all models should be assessed by seeing how well they predict the activity of chemicals, which were not used to make them. This is not, however, always simple. In part valuable information may be left out by setting aside chemicals to be used in such an evaluation, and in part it can be extremely difficult to assess how “external” chemicals relate to the model’s domain; that is, if they represent a random distribution within this thereby giving a fair picture of the performance of the model.

This problem is often addressed by using one or another form of cross-validation. Statistical evaluation is an extremely important method of determining the performance of these models, and in some cases (where there is little or no test data to be found which was not used to develop the model) it is the only method available.

The validation techniques most commonly mentioned in this report include the “drop one” “Q²” procedure, where one substance at a time is removed, and then predicted by a model made on the remainder of the training set. This is done once for every substance. While widely used, this form of cross-validation can have a tendency to over-predict goodness of fit.

A more robust technique for these data sets is for example the “3x10% out”, which consists of removing a random sample of 10% three times, and each time making a new model which is then used to predict the excluded chemicals. Instead of running this process three times it can be run until all of the chemicals have been estimated. However, three runs will generally be sufficient to establish the correlation /50/.

For the validation of a parametric model the result can be expressed as the sensitivity, the specificity and the concordance of the model. The sensitivity is a measure of how well the model “catches” the substances with the effect being modeled. A sensitivity of 80% means that 80% of the “true positives” in the validation set were correctly predicted as positives, and that the remaining 20% were falsely predicted as negatives (false negatives). The specificity is a measure of how many false positives the model predicts. A specificity of 80% means that 80% of the “true negatives” in the validation set were correctly predicted as negatives, and the remaining 20% of the negatives were falsely predicted as positives (false positives). The concordance is an overall measure of the correctness of the predictions. A concordance of 80% means that 80% of the substances in the validation set were correctly predicted as positives or

negatives, and the remaining 20% are the false predictions (false negatives and false positives).

Predictive ability will vary depending on both the method used, and the endpoint in question. In general, predictive ability of contemporary QSAR systems can often correctly predict the activity of about 70 – 85% of the chemicals examined, provided that the query structures are within the domains of the models /53,54/. This also applies to the models described in this paper. Of course, a model can never be more accurate than the test data on which it was based. Therefore it is extremely important to be aware of the accuracy and reproducibility of the test data used for making a model. If a biological test gives the wrong results 17% of the time, the “perfect” model based on these tests would also be wrong in 17% of the time.

In addition to assessing the predictive ability of a model, it is also necessary to consider in which context it will be used. In some cases a large number of “false positives” or “false negatives” may be acceptable, while in others they will not be. In this exercise there was no deliberate attempt to adjust the weight of these factors in either direction. The specific “context” in which these models have been used is simply that where there are no tests or other information available, the alternative is that the substance is not assessed at all for the endpoints covered.

2.1.4 Software

Today numerous computerized systems exist for predicting a large range of effects reaching from biodegradability to cancer. These include fragment based[†] statistical systems such as TOPKAT and M-CASE, as well as three-dimensional Modelling of ligand docking^{**} such as Comparative Molecular Field Analysis (COMFA). Mention should also be made of OASIS /46,47/, a sophisticated program package able to estimate a wide variety of effects using 3-D and Quantum Mechanical parameters, and which is currently being used to estimate binding of chemicals to Estrogen receptors /48/.

In essence, these programs don't really do anything “new.” They are simply grouping substances with similar structures and similar effects, including use of global or local parameters such as LogP and electrophilicity in much the same way as an expert might do. However, they do this at very high speed and take account of a large number of factors simultaneously (such as critical inter-atomic distances) which can assist an expert in finding hitherto unobserved relationships. In addition, the programs TOPKAT and M-CASE described below, emulate another human characteristic, and reject estimates for chemicals where there is simply not enough information to provide a sound prediction. They accomplish this by iterative statistical methods rather than by human intelligence or intuition.

M-CASE

M-CASE is a knowledge-based artificial intelligence system capable of learning directly from data. Models made in this program can predict various toxic endpoints on the basis of discrete structural fragments found to be

[†] In fragment based programs the prediction is based on the occurrence of molecular substructures.

^{**} Ligand docking models give predictions of how well a chemical substance fits in a certain 3D structure of a macromolecule with a biological function like a receptor (for example a hormone receptor).

statistically relevant to a specific biological activity, either increasing or decreasing it. The program can thus provide a “chemical” explanation to observed biological properties. It assumes that the presence of fragments previously found in a number of active compounds is indicative of potential activity. This fragment-based method is assumed to be a reasonable basis to assess the activity of new molecules. On the basis of the presence of the fragments in a query molecule the program will estimate a value for its potency by using “local QSARs” for the various fragments. If so found, “global QSARs” like the relation between LogP and toxicity to aquatic organisms may also be included in the model. The program gives a warning if there are fragments in the query molecule, that are not found in the training set of the model, indicating that the query molecule is outside the domain of the model /38,43/. Estimates for substances found to be within the domain of the model and for which sound predictions could be made are referred to as AOKs (“All OK chemicals”) in this paper.

TOPKAT

TOPKAT assesses toxicity of chemicals from their molecular structure utilizing QSTR (Quantitative Structure Toxicity Relationship) models for assessing specific adverse health effects /56/. When querying the program by entering a code for chemical structure, the program determines the compound class of the structure for those models which have class-specific sub-models. Next, the system computes the descriptors needed for the specific toxicity model. These consist of for example electrotopological state, kappa indices, molecular weight and symmetry indices. Then the program checks whether all the fragments present in the query molecule were present with adequate frequency in the training set for the specific equation. If there are no missing fragments, the program next checks whether the query is within the optimum prediction space of the equation. If this is the case, the training set of the model is searched for the compounds most similar to the query molecule, and the concordance between the actual and predicted values for those compounds is determined /45/. If there is reasonable agreement between observed / predicted values for the four most similar substances the estimate is accepted and referred to as AOKs in this paper.

Epiwin

This suite of programs developed by Syracuse Research Corporation was used to estimate three ecotoxicological parameters – Biodegradation, LogP and Bioconcentration. Unlike TOPKAT and M-CASE, Epiwin does not attempt to define a predictive space, and all estimates were used “as is”.

Chem-X

This program has features for making estimates for a large number of physical-chemical properties of chemicals, making 2D- and 3D-QSARs and storing large amounts of data and chemical structures in databases.

The Danish EPA has built up a database in Chem-X which contains QSAR predictions for about 166.000 substances /55/, including almost all of the discrete organic chemicals in Einecs, a total of approximately 47,000 substances. Estimates are available for a number of endpoints covering both health- and environmental concerns. The QSAR estimates for these chemicals create the background for the recommended selfclassifications. Detailed facilities for searching, displaying and manipulating chemical structures are also available in this data package. This tool was used extensively to compare

test data, predictions and selected sub-substructures while performing “expert” assessment of the QSAR’s.

Possibilities for dissemination of this database and the detailed QSAR predictions are currently unclear due to issues of copyright.

2.2 METHODOLOGY IN MAKING THE LIST

2.2.1 The selected dangerous properties

The following endpoints were addressed:

- Acute oral toxicity
- Sensitization by skin contact
- Mutagenicity
- Carcinogenicity
- Danger to the aquatic environment

2.2.2 The evaluated chemical substances

The overall purpose of the project was to evaluate as many as possible of the substances in Einecs (European Inventory of Existing Commercial Chemical Substances) /2/. The list consists of 100.116 entries, covering organic and inorganic substances in both single substance entries and mixtures.

The screening was limited to cover “discrete organics,” meaning that UVCBs (Unknown, Variable Composition and Biologicals) and other ill-defined structures or mixtures were excluded for practical reasons – if you don’t know what it is, you can’t really make a model. Exceptions were made where this seemed logical (C12 – C16 n-alcohols has been entered as C14 n-alcohol – hydrochloride salts have been entered as the parent compound, etc.).

Inorganic substances have likewise not been evaluated. These are usually better approached by simpler methods of evaluating the availability of the respective an- and cations with well known hazard profiles. “Organo-metallics” have also been excluded as being poor candidates for modeling. Finally, as a matter of resources, only such chemicals as were available with 3-D structural information were used /7/.

In so far as this was possible using a CAS number comparison, all substances already classified on Annex I of the formal EU list (List of dangerous substances) were also removed as they should never be the subject of provisional classification.

This resulted in a total of 46,707 or about half of all Einecs chemicals, which could be subjected to screening.

2.2.3 Test data

For the vast majority of the chemicals no measured data was available. However, if measured data were available as part of the model, this was generally used in preference to the estimates.

It is important to stress that no attempt was made to search the world's published or unpublished databases for toxicological information to determine whether a QSAR was even necessary for each endpoint. This task is the responsibility of the manufacturer / importer of the individual chemicals.

2.2.4 Use of QSAR models

The technical specifications for the models and a description of the criteria for assignment of advisory classifications for each effect are given in the technical sections for the individual endpoints.

It should also be stressed that the models available do not predict a "classification" – they predict biological activity that may lead to a classification. Further criteria have therefore been applied to each endpoint to try and link the biological prediction with a risk phrase. Because of the large number of chemicals involved, "rules" were used to achieve this purpose. Such rules are also imperfect, but in essence the process is no different than that imposed upon a human expert forced to use common sense to provide a provisional classification for any given substance for which the desired test data does not exist.

Only model predictions that satisfied a formal criteria were used: For TOPKAT the predictions had to fall within the optimum prediction space of the model, and the four most statistically relevant observed/predicted chemicals referenced by the model should be within acceptable agreement. The predictions fulfilling these criteria are referred to as AOKs.

For M-CASE the predictions had to fall within the optimum prediction space of the model, meaning that there were no unknown fragments, and that there was sufficient knowledge about the known fragments to give an unequivocal prediction.

As described in the technical sections, expert inspection has been undertaken where time allowed to confirm the probable activities given by the QSARs. This has included evaluation of the QSAR estimates in comparison with known biological activities and chemical properties. No in depth toxicological assessment of the individual chemical substances has been undertaken. Questionable QSAR predictions for each endpoint were excluded.

The effort used on expert inspection varied with the endpoint in question. In general most time was used in assessing the predictions for Mutagenicity and Carcinogenicity, and least was used on Allergy and Aquatic Effects.

2.2.5 The result

It is important to understand that the results as given in the Advisory list only represent POSITIVE predictions. No distinction has been made between a negative prediction for an endpoint, and an unreliable prediction (a non-AOK prediction) which was simply discarded.

Evaluated substances not on the list, or substances which are on the list but without advisory classifications for one or more of the selected dangerous properties, may have been predicted as not having this / these dangerous properties, or the models may not have been valid for this substance.

Therefore the advisory list can not be used to conclude that these substances do not possess dangerous properties. Depending on the endpoint in question, unreliable predictions were obtained for between 5 and 65% of the chemicals examined.

2.3 ACUTE ORAL TOXICITY

EU criteria for classification

The formalized criteria for classification for acute oral toxicity includes a number of options of tests including fixed-dose procedure and interpretation of the various sources of information about acute oral toxicity, but is often based on acute LD₅₀ tests in the rat for which the following classification criteria are used:

TABLE 3 CLASSIFICATION CRITERIA

Classification criteria	Classification
LD ₅₀ oral, rat ≤ 25 mg/kg	T+;R28 (very toxic; very toxic if swallowed)
25 mg/kg < LD ₅₀ oral, rat ≤ 200 mg/kg	T;R25 (toxic; toxic if swallowed)
200 mg/kg < LD ₅₀ oral, rat ≤ 2,000 mg/kg	Xn;R22 (harmful; harmful if swallowed)

Evaluation based on QSAR models

An advisory classification of Xn;R22 is recommended in all cases where a rat oral LD₅₀ of ≤ 2000 mg/kg is predicted or based on measured data. For reasons indicated below, no attempt was made to differentiate between the different levels of acute toxicity, and it is important to recognize that this classification will often be less stringent than classification based on measured data.

If test results measured in the rat were readily available (had been used to make the model) these took precedence over any predictions.

As acute toxicity data from the mouse following a variety of different routes of administration was also available in some cases, this was used to predict rat oral LD₅₀'s using the QSARs preferentially as follows /8,9/:

TABLE 4

1. $\text{Log LD}_{50} \text{ oral, rat} = 0.731 + 0.841 * (\text{Log LD}_{50} \text{ oral, mouse})$ RTECS data 1989, n=3919, R ² = 0.750, Q ² = 0.749
2. $\text{Log LD}_{50} \text{ oral, mouse} = 0.682 + 0.373 * (\text{Log LD}_{50} \text{ iv, mouse}) + 0.518 * (\text{Log LD}_{50} \text{ ip, mouse})$ RTECS data 1994, n = 286, R ² = 0.766, Q ² = 0.764
3. $\text{Log LD}_{50} \text{ oral, mouse} = 0.731 * (\text{Log LD}_{50} \text{ iv, mouse})$ RTECS data 1994, n=286, R ² = 0.724, Q ² = 0.724
4. $\text{Log LD}_{50} \text{ oral, mouse} = 0.945 + 0.802 * (\text{Log LD}_{50} \text{ iv, mouse})$ RTECS data 1994, n=286, R ² = 0.689, Q ² = 0.688

iv: Intravenous

ip: Intraperitoneal

Biological data consisting of LD₅₀'s in mice or rats was available for just over 10% of the chemicals processed. If no biological data were available, rat oral LD₅₀ was estimated according to the QSTR model TOPKAT (v 5.01). According to TOPKAT, the model contains about 4,000 substances and their own cross-validation for this endpoint indicates 86-100% of estimations falling within a factor of five from test results /10/.

Danish EPA's external evaluation of this model using 1,840 chemicals not contained in the TOPKAT data set gave somewhat poorer results; R² = 0.31. According to this evaluation 86% of estimations fall within a factor of ten from test results /11/. The distribution can be seen in table 5.

TABLE 5

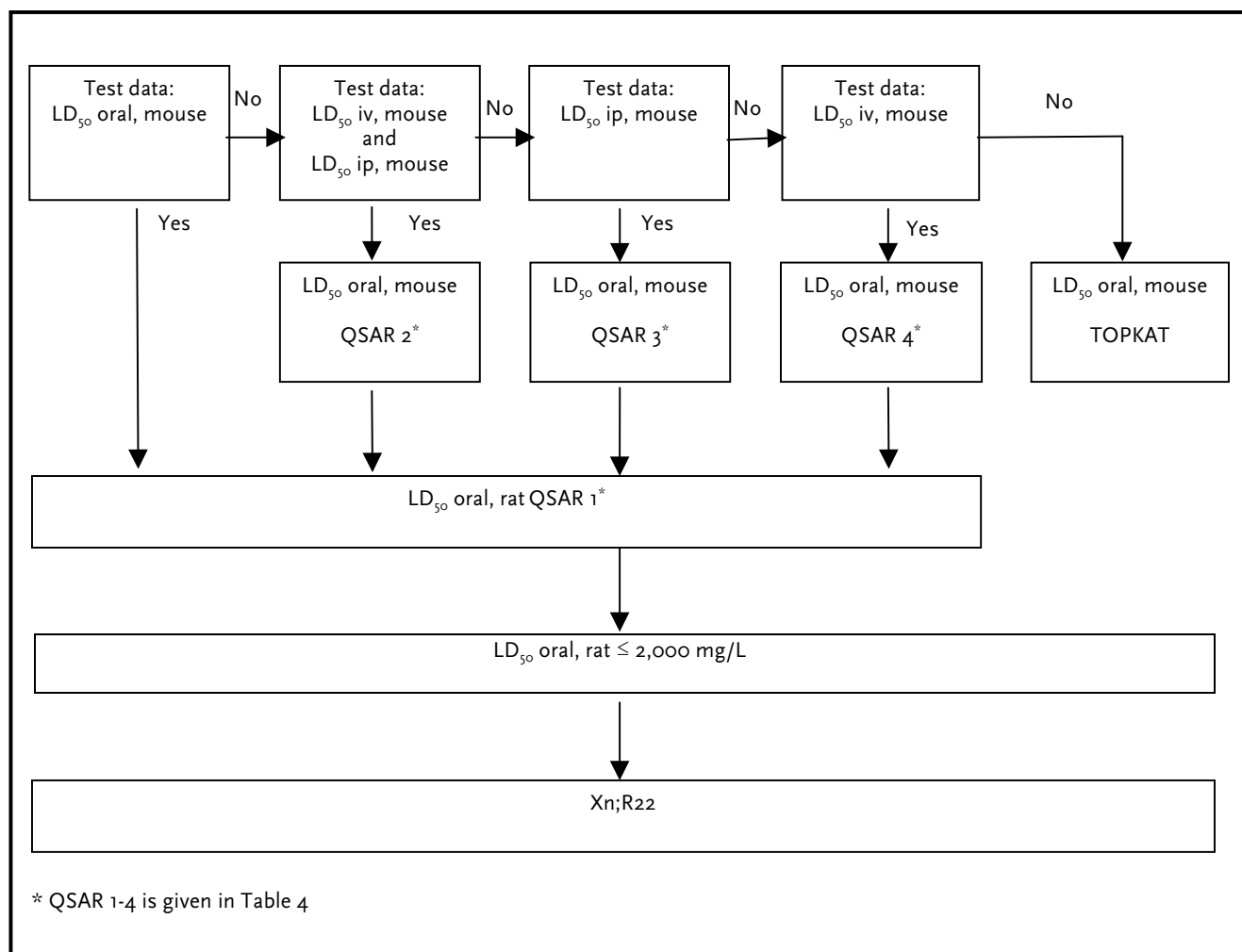
Result predicted within a factor of:	%	N (cumulative)
2	42	671
4	67	1,069
6	78	1,235
8	83	1,323
10	86	1,368

In modern LD₅₀ tests using small numbers of animals, statistical variation is often within a factor of 2-4, and inter-laboratory variations of up to 10 are not uncommon /12/. While the TOPKAT model is clearly not perfect, it is still considered sufficient to give an approximation for the suggested least strict classification for acute toxicity, Xn;R22. However, the accuracy of the model is not considered to be sufficient to differentiate between the three different levels of acute toxicity ("hamful", "toxic" and "very toxic"). It is therefore important to recognize, that there will be substances, though given with the advisory classification Xn;R22 on the list, which on the basis of for instance animal test should be classified as T;R25 or Tx;R28.

Where TOPKAT was able to make a robust prediction (AOK) it found 57% of all chemicals to have an acute oral LD₅₀ in rat of ≤ 2,000 mg/kg. The percentage of chemicals with acute toxicity's of ≤ 2,000 mg/kg for 12,632 chemicals tested for acute toxicity in rat found in the Registry of Toxic Effects of Chemical Substances (RTECS 1998) /52/ was 61%. That these two percentages are so similar is not surprising, since RTECS data was also the chief source of biological information used to construct the TOPKAT model.

A schematic diagram of the systematic evaluation is given in figure 2.

FIGURE 2 THE SYSTEMATIC EVALUATION



Approximately 10,200 compounds were estimated as having an acute LD₅₀ in rat of 2,000 mg/kg or less*. About 700 were removed by expert judgement in an attempt to exclude amino-acid and protein-type compounds which were considered likely to break down due to the effects of gastric acidity, or substances for which gastric absorption was expected to be poor. This resulted in 9,538 substances with an advisory classification of Xn;R22.

2.4 SENSITIZATION BY SKIN CONTACT

EU criteria for classification

Classification as sensitizing by skin contact, R43 (“May cause sensitization by skin contact”), is based either on animal studies or practical experience or combinations thereof. The animal criterion is based on either an adjuvant or non-adjuvant test.

* TOPKAT calculations were also performed for Rat Chronic Lowest Observed Adverse Effect Level. Cross validated accuracy of this model was similar to that used for acute toxicity, with 95% of predictions being within a factor of 3-5 of the measured values /13,44/. However, as there is no EU classification criteria related specifically to this endpoint (but rather to “serious morphological or toxicological effects after repeated dosing, R48”) no classification suggestions were applied for this endpoint.

Different adjuvant tests exist, but the Magnusson-Kligmann's method (GPMT: Guinea Pig Maximization Test) is preferred. Response in 30% of the animals results in classification. For a non-adjuvant test (for example the Büehler test) 15% responding animals is regarded as positive. The human data can be results from patch testing, case studies or epidemiological studies.

Evaluation based on QSAR models

Two approaches were used to estimate contact sensitisation /14,15/.

The first approach uses two TOPKAT QSTR models. The first model was used to predict "Allergy versus non-allergy", and, in cases where this was positive, the second model was used to predict "Strong versus weak/moderate allergy". The models used were primarily related to the GPMT. Only predictions of "Strong allergy" were considered as being likely to fulfill the EU criteria for R43.

In a second approach, predictions were also made using M-CASE. The data set used to produce the M-CASE models differed somewhat from the TOPKAT set, in that both data from the GPMT and human data were represented. Only positive predictions with M-CASE scores of ≥ 40 (corresponding to "very active") were selected.

TABLE 6 THE MODELS USED

Model	Technical specifications
TOPKAT (v. 5.01 1998) No Sensitization vs Any	n=389 GPMT Cross validation result (Q ²) /14/: Sensitivity 84-94% Specificity 87-96%
TOPKAT (v. 5.01 1998) Strong vs Weak/Moderate	n=266 GPMT Cross validation result (Q ²) /14/: Sensitivity 88-96% Specificity 88-98%
M-CASE (v. 3.320 1999) Model A33: Allergic Contact dermatitis	n=1,034 GPMT or data from human experience Cross validation result (3*10% out) /15/: Sensitivity 69 – 89% Specificity 89– 94% Chi ² > 50, p < 0.0001

External validation of both TOPKAT and M-CASE models was also attempted using confidential results from the EU New Chemicals program. Using the two-stage TOPKAT model (n= 64 AOK predictions) 67% of positives were correctly identified, and 77% of negatives. For M-CASE, (n= 75 AOK predictions) 45% of positives were correctly predicted, and 81% of negatives /16/.

It is difficult to know how representative New Chemicals are with regard to the universe of Existing Chemicals. Generally New Chemicals are more complex structures with higher molecular weights. Perhaps the most surprising aspect of this exercise was to find that for over three thousand chemicals that should have been assessed for this endpoint, such a tiny percentage of useful test data could be found.

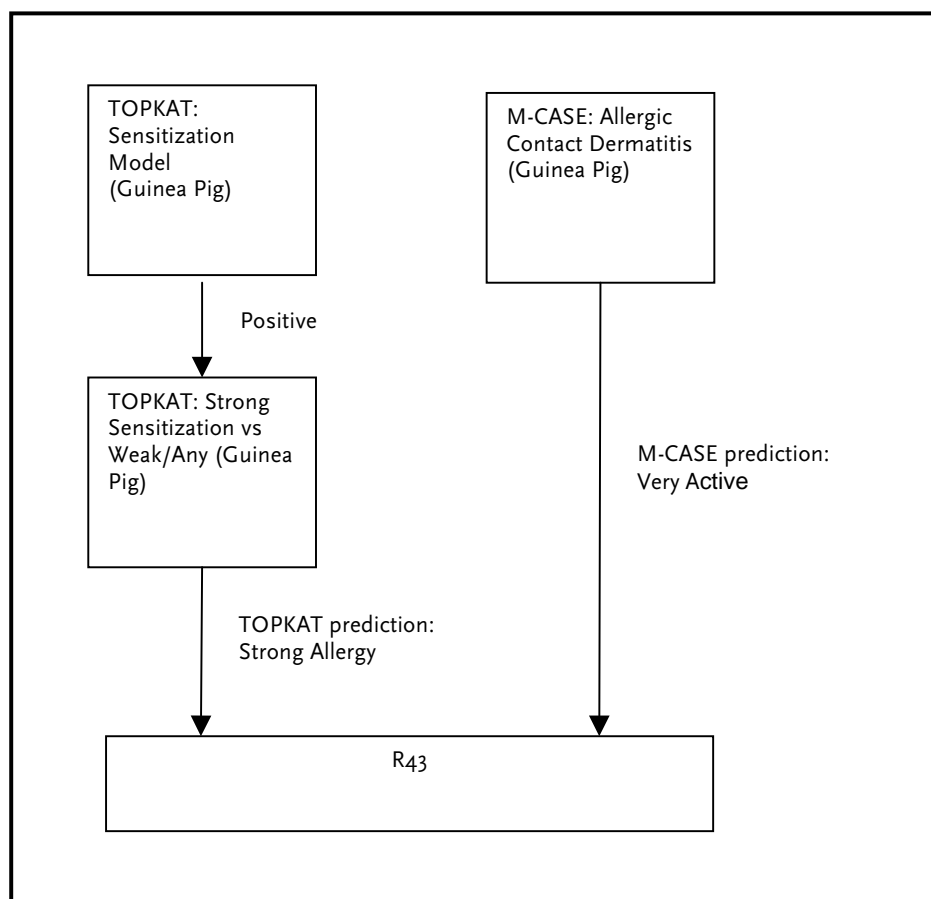
Compounds predicted as positive by either TOPKAT or M-CASE according to the above criteria were selected, provided that they were either AOK in the first, or contained no unknown fragments or equivocal results in the latter.

While it was considered to use “positive” in *both* models as a criteria, in the end this seemed inefficient, not so much due to lack of concordance between model predictions, but because the acceptance domains (AOK or all fragments known) of the two methods differed considerably.

No attempt was made to further reduce the list by systematically applying expert judgement.

A schematic diagram of the systematic evaluation is given in figure 3.

FIGURE 3 THE SYSTEMATIC EVALUATION



9,668 chemicals met the above criteria, for which an advisory classification of R43 is suggested. This strike many experts as being a rather large number of chemicals and while these models represent the current “state-of-the-art” it may indicate that they are over-sensitive. However, it was very difficult to obtain any reliable indication of how many Existing Chemicals would cause contact allergy if actually tested in animals or humans. Estimates of percentages of allergens on EINECS ranged from 5-25%, with some preference being expressed for 10%, which is the number of Annex I substances currently classified for this effect. It is not possible, however, to estimate the influence of confounders on the distribution represented in Annex I. Positive bias can have been introduced because chemicals testing positive are over-represented. Negative bias can have been caused by the fact that most of the chemicals have never been tested at all. The question of numbers remains open.

2.5 MUTAGENICITY

EU criteria for classification

The criteria for classification for mutagenicity is divided into 3 different categories:

Classification as mutagen, category 1 (mut1;R46, may cause heritable genetic damage) is based on evidence of a causal association between human exposure to the substance and heritable genetic damage.

Classification as mutagen, category 2 (mut2;R46, may cause heritable genetic damage) is based on animal studies showing mutagenicity to germ cells either in assays on germ cells or by demonstrating mutagenic effects in somatic cells *in vivo* or *in vitro* as well as metabolic proof that the substances reaches the germ cells.

The criteria for classification as mutagen, category 3 (mut3;R40, possible risks of irreversible effects) is based either on *in vivo* mutagenicity tests or on cellular interactions with *in vitro* tests acting as supportive evidence. For this classification, it is not necessary to demonstrate germ cell mutations.

Evaluation based on QSAR models

A number of models were applied for this endpoint. The different models predict a number of genotoxicity endpoints. Induction of micronuclei *in vivo*, was required, as this demonstrates chromosomal damage in somatic cells *in vivo*. The remaining endpoints reflect *in vitro* genotoxicity, where positive results would not normally lead to classification for this effect. However, positive results for these endpoints provide supporting evidence for data from *in vivo* estimates.

TABLE 7 THE MODELS USED

Model	Technical specifications
M-CASE (v. 3.320 1999) Model A2E: Structural Alerts for DNA Reactivity	n=784 Cross validation result (3*10% out) /24/: Sensitivity 85-98% Specificity 60-69% Chi ² >22, p< 0.0001
M-CASE (v. 3.320 1999) Model A62: Induction of Micronuclei	n=238 GeneTox chemicals Cross validation result (3*10% out) /30/: Sensitivity 80 –100% Specificity 50 – 70% Chi ² >4, p <0.05
TOPKAT (v. 3.01, 1998) Salmonella (Ames) Mutagenicity,	n=1,866 Cross validation result (Q ²) /25/: 10 sub-modules with sensitivities and specificity's of 75-100%. External evaluation (Danish EPA, 1998, n=118) /21/: 82% correct negative predictions, 76% correct positive predictions
M-CASE (v. 3.320 1999) Model A2H: Salmonella (Ames) Mutagenicity	n=2,034 NTP or GeneTox tests Cross validation result (3*10% out) /27/: Sensitivity 75-78.5% Specificity 78.2 – 90% Chi ² >150, p <0.0001
M-CASE (v. 3.320 1999) Model A61: Chromosomal Aberrations	n=233 NTP tests in cultured CHO cells Cross validation result (3*10% out) /28/: Sensitivity 44-80% Specificity 50-80% Chi ² < 2, p>0.15 (further validation being undertaken)
M-CASE (v. 3.320 1999) Model A2F: Mutations in Mouse Lymphoma	n=210 NTP thymidine kinase in L5178Y cells Cross validation result (3*10% out) /29/: Sensitivity 64-100% Specificity (not determined) Chi ² ≈ 2, p=0.15 (further validation ongoing)

If classification had been proposed on measured data, a positive result in the *in vivo* micronucleus test would have been sufficient evidence on which to base the classification. Since the data is predicted and not measured data, additional support for the prediction was obtained by including a number of other indicators of genotoxicity.

It is not suggested that positive *in vitro* evidence should also be necessary when classifying substances with positive *in vivo* test data. However, it was not felt that the QSAR model for the mouse micronucleus test alone was sufficient, and data estimates from additional QSAR's relevant to the endpoint were therefor used to increase the likelihood of a correct positive prediction.

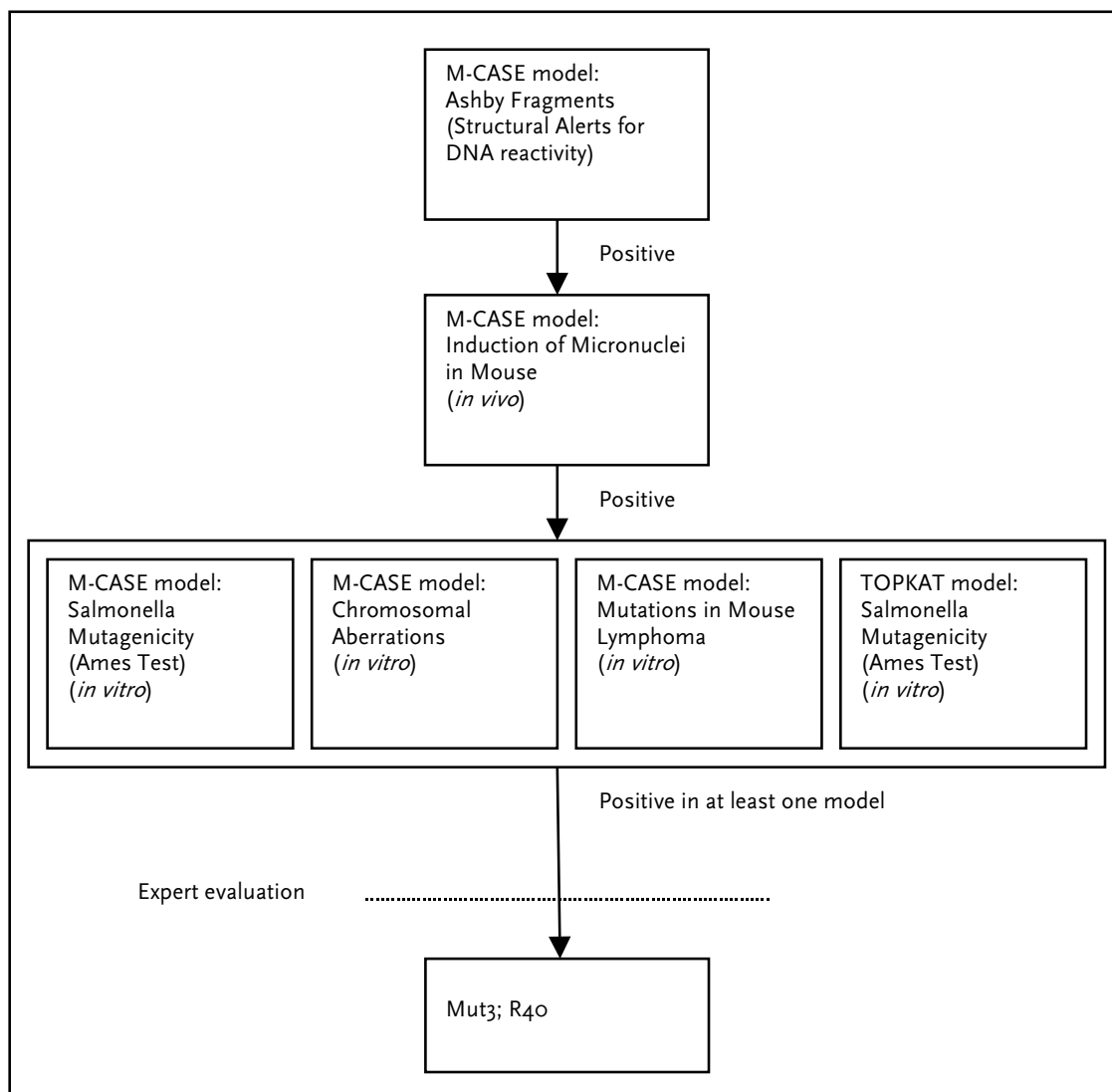
Chemicals for which model estimates were positive for mouse micronucleus and structural alerts for DNA reactivity (here an exception was made in that predictions with one unknown fragment were also accepted) and which also had two positive genotoxicity endpoints, passed the criteria for the systematic evaluation.

Two models for Salmonella (Ames) mutagenicity were used, a TOPKAT and a M-CASE module respectively. This related primarily to the fact that the models differed with regard to domain, and often a robust prediction was only available for one model. If robust predictions were available for both models,

and in disagreement, this was taken into account on a case-by-case basis during the final evaluation.

A schematic diagram of the systematic evaluation is given in figure 4.

FIGURE 4 THE SYSTEMATIC EVALUATION



2,272 Eines chemicals met the criteria in the systematic evaluation.

As none of these models identifies germ cell mutagenicity, the current QSAR's do not allow discrimination between the EU categories for mutagenic effects in the three categories and the lower classification is therefore assigned as advisory classification in all cases.

Expert judgment was undertaken to confirm the robustness of the predictions of these 2,272 chemicals. This process included examination of the 2- or 3-d chemical structure, and visual comparison with test data within structural groups. If this procedure raised any doubt, substances were removed from the list for more detailed consideration in the future. This resulted in a final selection of 1,678 substances with an advisory classification mut3;R40.

2.6 CARCINOGENICITY

EU criteria for classification

This end-point can result in classification in 3 different categories:

Classification as carcinogen in category 1 (carc1;R45, Toxic; may cause cancer or carc1;R49, Toxic; may cause cancer by inhalation) is based on strong causal relationship in humans.

Classification as carcinogen in category 2 (carc2;R45, Toxic; may cause cancer or carc2;R49, Toxic; may cause cancer by inhalation) is based on conclusive animal data from 2 species or 1 species with supportive evidence such as genotoxic effects *in vitro* or *in vivo*.

Classification as carcinogen in category 3 (carc3;R40, Harmful; possible risks of irreversible effects”) is subdivided into two:

- a) Well-investigated substances with restricted tumorigenic effects. It is normally based on clear data of tumour formation in one species. Mutagenicity data *in vitro* and *in vivo* can be used as supportive evidence.
- b) Substances that are insufficiently investigated, but raising concern for man.

Evaluation based on QSAR models

While there are many non-genotoxic carcinogens acting by a wide variety of often-unknown mechanisms, it was chosen to focus here on chemicals likely to cause cancer through a genotoxic mechanism. Therefore a pre-selection criteria for genotoxicity was set up.

The criteria for the pre-selection for carcinogenicity was a positive estimate for structural alerts for DNA reactivity (AOK or one unknown fragment) and two positive AOK genotoxicity predictions out of five models for genotoxicity. The technical specifications for the models used to predict genotoxicity is given in the chapter ”Mutagenicity”.

As opposed to the selection criteria for mutagenicity, a positive mouse micronucleus test was not demanded, as not all genotoxic carcinogens are necessarily clastogenic (cause loss, addition or rearrangement of parts of chromosomes). This gave a pre-selection of 3.362 EINECS chemicals.

A total of ten cancer models were available, plus four sub-models.

TABLE 8 THE MODELS USED

Model	Technical specifications
TOPKAT (v. 3.01 1998) NTP Carcinogenicity: Male Rat	366 NTP rodent studies Cross validation result (Q ²) /32/: Sensitivities of 82-87% Specificity's of 82-88%
TOPKAT (v. 3.01 1998) NTP Carcinogenicity: Female Rat	
TOPKAT (v. 3.01 1998) NTP* Carcinogenicity: Male Mouse	
TOPKAT (v. 3.01 1998) NTP Carcinogenicity: Female Mouse	
TOPKAT (v. 5.01n 1998) FDA Carcinogenicity: Male Rat	n=384 Cross validation result (Q ²) /33/: Sensitivity 91% Specificity 90%
Sub-model: Single vs Multiple Organ Tumors	n= 131 Cross validation result (Q ²) /33/: Sensitivity 91% Specificity 96%
TOPKAT (v. 5.0 Feb. 1998) FDA** Carcinogenicity: Female Rat	n=383 Cross validation result (Q ²) /33/: Sensitivity 84% Specificity 89%
Sub-model: Single vs Multiple Organ Tumors	n= 125 Cross validation result (Q ²) /33/: Sensitivity 92% Specificity 96%
TOPKAT (v. 5.0 Feb. 1998) FDA Carcinogenicity: Male Mouse	n=316 Cross validation result (Q ²) /33/: Sensitivity 82% Specificity 90%
Sub-model: Single vs Multiple Organ Tumors	n=93 Cross validation result (Q ²) /33/: Sensitivity 93% Specificity 94%
TOPKAT (v. 5.0 Feb. 1998) FDA Carcinogenicity: Female Mouse	n=312 Cross validation result (Q ²) /33/: Sensitivity 86% Specificity 89%
Sub-model: Single vs Multiple Organ Tumors	n=100 Cross validation result (Q ²) /33/: Sensitivity 95% Specificity 95%
M-CASE (v. 3.320 1999) Carcinogenic Potency Database model: Rat (Danish EPA version of AoD, Feb. 2000)	n=870 chemicals from the CPDB Cross validation result (3*10%) /34/: Sensitivity 52-67% Specificity 63-68% Chi ² ≈ 6, p<0.014 (further validation ongoing)
M-CASE (V. 3.320 1999) Carcinogenic Potency Database model: Mouse (Danish EPA version of AoE, Jan. 2000)	n=720 chemicals from the CPDB Cross validation result (3*10%) /35/: Sensitivity 45-50% Specificity 64-72% Chi ² ≈ 2, p = 0.15 (further validation ongoing)

* NTP: National Toxicology Program, USA

** FDA: Food and Drug Administration, USA

The accuracy of these models can be difficult to determine, as there are few independent tests that have not already been used in the construction of the models themselves, which can be used for an independent assessment. This is particularly the case for TOPKAT's models, where the only real estimates consist of the producers own "1 out" Q² cross-validations. For M-CASE, other statistical methods are available.

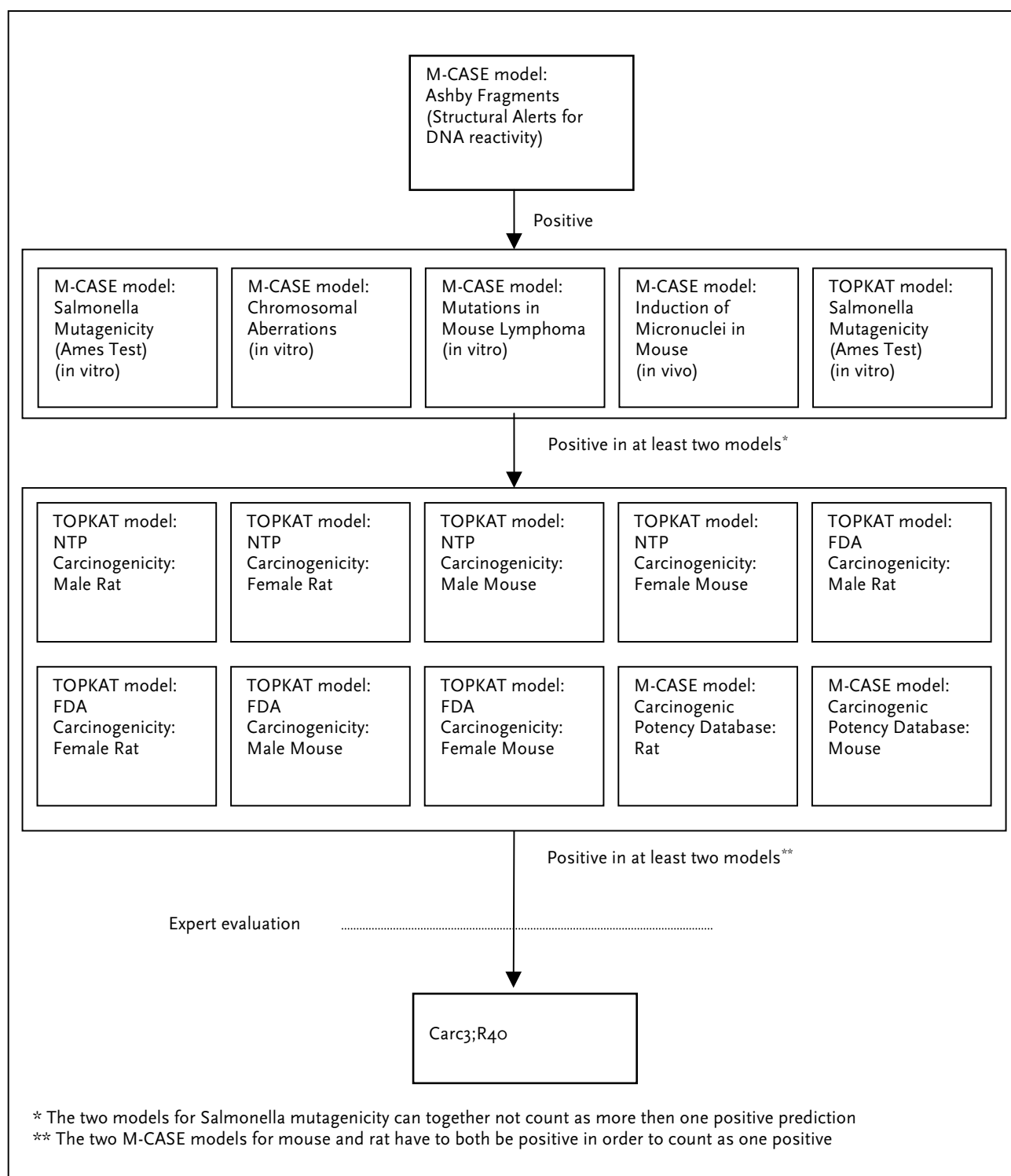
In a long-running project, where several cancer models predicted the outcome for NTP chemicals which had not yet been tested, upon completion of these tests (for 45 substances) the general conclusion was that accuracy of around 70% was achieved for clearly carcinogenic or non-carcinogenic substances /31/. Due to the small number of chemicals in this analysis it is difficult to know how much weight can be assigned to the conclusion.

3,362 substances met the pre-selection criteria for genotoxicity. For a substance to be selected as a probable carcinogen it was necessary for the following criteria to be fulfilled:

At least two positive predictions (sub-models excluded) for carcinogenicity. An exception was made for the M-CASE CPDB models. Because the data is less homogeneous, both rat and mouse predictions had to be positive to count as one prediction, and in addition to this the carcinogenic potency had to include TD_{50} 's for tumor induction of less than 1,000 mg/kg/day. These two CPDB models were developed by Danish EPA using M-CASE methodology which is described for this data set in the following references /34,35,40/.

If one or more positive tests could be seen (part of the training set for the model) for any cancer endpoint, this took precedence over model results and resulted in an over-all positive classification recommendation. While in most cases this resulted in little change (the models are heavily biased towards making a correct prediction for substances used to make them), it was felt that there was no reason to artificially reduce the quality of the advisory classification by neglecting to use data, which happen to be present. A schematic diagram of the systematic evaluation is given in figure 5.

FIGURE 5 THE SYSTEMATIC EVALUATION



According to these criteria, 1,272 substances were selected for advisory classification for carcinogenicity. Expert judgment was performed on the QSARs. In this process, all data was used including predictions of TOPKAT FDA Carcinogenicity sub-models, the probability of rapid metabolism or excretion, and where appropriate, predictions of aryl hydroxylase activity /37/. Where any doubts were raised, substances were removed from this version of the list to be considered in more detail in the future.

This resulted in 652 substances selected for advisory classifications for carcinogenicity. It is not felt that the models employed allow discrimination between classification in the three categories, so the lower classification Carc3;R40 was applied in all cases.

2.7 DANGER TO THE AQUATIC ENVIRONMENT

EU criteria for classification

The classification criteria are composed of three main elements: Biodegradability, Bioconcentration potential, and Toxicity to aquatic organisms. Classifications are assigned according to the following scheme:

TABLE 9 CLASSIFICATION CRITERIA

Classification	Criteria for acute toxicity to aquatic organisms*
N;R50 (Dangerous for the environment; very toxic to aquatic organisms)	Acute toxicity ≤ 1.0 mg/l
N;R50/53 (Dangerous for the environment; very toxic to aquatic organisms; may cause long-term adverse effects in the aquatic environment)	Acute toxicity ≤ 1.0 mg/l and not readily degradable or BCF** ≥ 100
N;R51/53 (Dangerous for the environment; toxic to aquatic organisms; may cause long-term adverse effects in the aquatic environment)	Acute toxicity ≤ 10 mg/l and not readily degradable or BCF ≥ 100
R52/53 (Harmful to aquatic organisms; may cause long-term adverse effects in the aquatic environment)	Acute toxicity ≤ 100 mg/l and not readily degradable
R53 (Harmful to aquatic organisms)	Solubility in water < 1 mg/l and not readily degradable and BCF ≥ 100

* The lowest effect concentration for fish, daphnia or algae is used

** BCF: Bioconcentration factor

Evaluation based on QSAR models

Advisory classifications were assigned on basis of combinations of estimates for biodegradation, bioconcentration and acute toxicity according to the criteria in table 9. Classification with risk phrase R53 alone was not done in this exercise, as the strong co-linearity between water solubility and bioconcentration factor made it redundant.

Biodegradation

Biodegradability was estimated using the Syracuse BIOWIN program (v. 3.02) /17,41/. Only the linear equation for rapid/non-rapid biodegradation was applied. Previous validation of this parameter compared with MITI “ready/not-ready” results showed that while a number of “not-ready” chemicals were missed, 93% of “not ready” predictions were correct /18/. In other words while this model may fail to identify all “non-ready” substances, the number of false predictions for lack of degradability will be acceptably low. A total of about 14,000 EINECS chemicals were found to be “not-readily degradable” according to this criteria /51/.

* While too late for this phase of the project, Danish EPA has now developed a M-CASE model for ready biodegradation based on new MITI data, which appears to offer significantly better predictions. 81% correct “not ready” and 76% correct “ready” predictions (3x10% out). An external validation using 72 “not ready” chemicals that had not been used to produce the model gave 89% correct predictions. Analysis and fine-tuning of this model is continuing /19/.

Bioconcentration

The classification and labeling guidelines prefer measured data for Bioconcentration, but as this is seldom available, a LogP (octanol/water) of greater than three is recommended as an indication that BCF will be 100 or greater, in accordance with the linear equation of Vieth and Kosian /41/. While a good rule-of-thumb, this relation both over- and underestimates BCF for many classes of chemicals, and takes no account of the fact that bioconcentration is a bilinear function of LogP, decreasing when this is sufficiently high.

Bioconcentration was therefore predicted using Syracuse BCFWIN (v. 2.13), a method based on a combination of logP (octanol-water) relations and structural fragment categories. This method was evaluated by its authors as having a statistical accuracy of $R^2 = 0.74$ ($n = 694$, S.D. 0.65, mean error = 0.47), which is a significant improvement over the standard equation of Vieth and Kosian ($\log \text{BCF} = 0.85 * \log \text{Kow} - 0.70$) where predictions for the same 694 compounds had a statistical accuracy of $R^2 = 0.32$ (S.D. 1.62 and mean error = 1.12) /20/. About 11,000 Eines chemicals were found with BCF estimates of equal to or greater than 100.

No attempt was made to further assess bioaccumulation potential caused by possible presence in the diets of aquatic organisms, as it was not felt that an appropriate general model was available.

Acute toxicity

For aquatic toxicity classifications, values $(L(E)C_{50})$ for fish, daphnia and algae are recommended (although seldom available for most existing chemicals). In the current exercise it was decided to only use predictions for fish, due to their robustness and the availability of high quality test data for model construction.

For Acute aquatic toxicity to fish a M-CASE model developed by Danish EPA using 96h LD_{50} data on 569 chemicals from the Duluth Fathead minnow Database /22/ was applied. The model had an R^2 of 0.85. Cross-validation of this model gave a Q^2 of 0.735 (3*10% out). A description of the M-CASE methodology used for the Fathead minnow data can be found in the following references /21,42/. Only predictions within the optimum prediction space of the model (no fragment or other warnings) were used.

As there was insufficient test data on the Fathead minnow for very lipophilic substances the M-CASE model was only applied for chemical substances with LogP of six or less. Another relationship was used for chemicals with a LogP of greater than six. Here, all substances were assumed to act by non-polar narcosis, and toxicity at equilibrium was estimated according to a relation to the predicted Bioconcentration factor:

$$LC_{50} \text{ (equilibrium)} = 8.15 \text{ mmol /BCF}$$

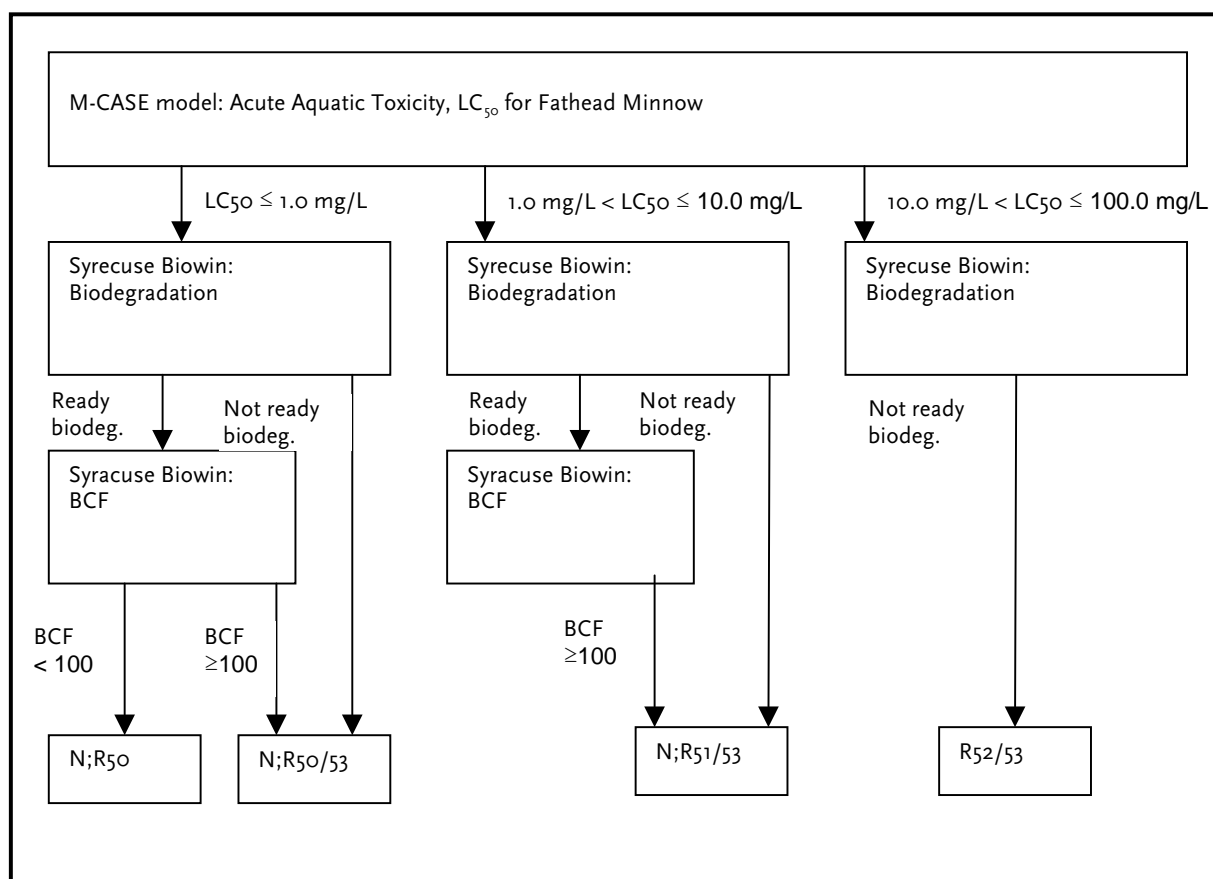
The choice of 8.15 mmol corresponds to the theoretical level inducing aquatic effects represented by the non-polar narcosis fish QSAR recommended in the EU TGD /41/. Non-polar narcosis Lethal Body Burden's for fish are generally assumed to be within the range of about 2–8 mmol /23,58/.

While simple LogP (octanol/water) relationships exist for predicting the non-polar narcotic toxicity for fish, daphnia and algae /41/, these do not distinguish specific toxicity's unique to any of the three taxa, and were not felt to offer any advantage over using the fish models alone, which also adequately predict non-polar narcosis. For all practical purposes, non-polar narcosis induces effects at the same concentration levels in all three taxa /18/.

Using both estimates, about 10,000 EINECS chemicals were found with toxicity's to fish of $LC_{50} \leq 100$ mg/l.

A schematic diagram of the systematic evaluation is given in figure 6.

FIGURE 6 THE SYSTEMATIC EVALUATION



A total of 8,731 substances were selected according to one of the four classification categories as indicated above. Considering that the number of robust (AOK) predictions for fish toxicity was just fewer than one-half of the chemicals screened, this number seems in reasonable concordance with what would be expected for Existing Chemicals.

The advantages of being able to predict toxic effects specific to both fish, daphnia and algae are obvious, and this can hopefully be accomplished in the future. A M-CASE model for acute toxicity to daphnia has recently been completed by Danish EPA ($n = 574$, $R^2 = 0.826$, $3 \times 10\%$ out $Q^2 = 0.692$). It is still being refined, and predictions for all chemicals will soon be available. A M-CASE model for toxicity to algae is under development.

3 References

1. Danish Ministry of Environment and Energy Statutory Order no. 1065 of November 30, 2000, on Classification, Packaging, Sale and Storage of Chemical Substances and Products.
2. EINECS (European Inventory of Existing Commercial Chemical Substances). O.J. C146A, 15.6.1990, p.1.
3. Allanou et al., 1999: Remi Allanou, Bjørn Hansen and Yvonne van der Bilt (1999): Public Availability of Data on EU High Production Volume Chemicals (EUR 18996 EN).
4. Niemelä, 1992: EU project; Priority Setting for the Purpose of future Classification and Labelling of Dangerous Substances (Contract No. B91/B4.3044/12200), Danish Environmental Protection Agency, Copenhagen, November 1992 and Jay Niemelä, 1994, Danish EPA, working document with update of the data from the EU project on priority setting for future classification and labelling (available on request).
5. Danish Ministry of Environment and Energy, Statutory Order no. 733 of July 31, 2000, List of Dangerous Substances.
6. European Commission: Commission communication 90/C146A/01 pursuant to Article 13 of Council Directive 67/548/EEC of 27 June 1967 on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances, as amended by Directive 79/831/EEC –
7. Oxford Molecular, Chem-X version July 1998, SMILES to 3-D conversion module.
8. Niemela, J., “Non-Structural Activity Coefficients for Acute Oral Toxicity in the Mouse and Rat”, Danish EPA, working document, 1992 (available on request)
9. Niemela, J., “Acute Toxicity versus route of Administration in Mice,” Danish EPA, working document, 1995 (available on request)
10. Health Designs Inc., “TOPKAT 3.0 QSAR Module Summary, Rat Oral LD₅₀”, v3ld50.d94, Rochester, New York, U.S.A., 1997
11. Niemela, J., “Computer Predictions of LD₅₀ in the Rat using TOPKAT v. 5.01, Danish EPA, working document, 1999 (available on request).
12. Hunter, W.J., et. al., “Intercamparison Study on the Determination of Single Administration Toxicity in Rats,” J. ASSOC. AFF. ANAL. CHEM., Vol. 62, no. 4, 1979.
13. Health Designs Inc., “TOPKAT 3.0 QSAR Model Summary, Rat Chronic LOAL,” v3loal.595, Rochester, New York, U.S.A., 1997.

14. Health Designs, Inc., "New Skin Sensitization Model," Computational Toxicology News, no. 21, 1998
15. M-CASE; Model A33 Allergic contact dermatitis, July, 1998
16. Niemela, J., "QSAR's for the Estimation of Sensitization Potential," Danish EPA, working document, 1999 (available on request).
17. Howard, P.H., et. al., "Predictive Model for aerobic biodegradability developed from a file of evaluated biodegradation data," Environ. Toxicol. Chem, 11; 593-603, 1992. (see also BIOWIN Biodegradation Probability Program for Microsoft Windows 3.1, Syracuse Research Corporation, Syracuse, NY, U.S.A., 1994.
18. Pedersen, F., Tyle, H., Niemela, J., Guttman, B., Lander, L., and Wedebrand, A., 1995, "Environmental Hazard Classification – data collection and interpretation guide for substances to be evaluated for classification as dangerous to the environment. Appendix 9; Validation of the BIODEG Probability Program." TemaNord Report 589, 153-156.
19. Niemelä, J., and Wedebye, E.B., "MITIMOD, M-CASE Model for Aerobic Biodegradation, Danish EPA, internal working document, April, 2000.
20. Meylan, W.M., et. al., "Improved Method for Estimating Bioconcentration Factor (BCF) from Octanol-Water Partition Coefficient," Third Update Report, Syracuse Research Corporation, Syracuse, NY, U.S.A. August, 1997.
21. Klopman, Gilles, Saiakov, R., and Rozenkranz, S., "Multiple Computer-Automated Structure Evaluation Study of Aquatic Toxicity II. Fathead minnow," Environmental Toxicology and Chemistry, Vol. 19, No 2, pp. 441-447.
22. Brooke, L.T., et. al., "Acute Toxicities of Organic Chemicals to Fathead Minnows (*Pimephales promelas*)," Center for Lake Superior Environmental Studies, University of Wisconsin – Superior, 1988.
23. Hermens, J.L., et. al., "Aquatic Toxicity of Polar Narcotic Pollutants," University of Utrecht, Utrecht, Netherlands, 1998
24. M-CASE Structural Alerts for DNA Reactivity ("Ashby fragments", model A2E, July 1998.
25. Health Designs Inc., "TOPKAT 3.0 QSAR Module Summary: Mutagenicity (Ames), V3mut.d94, New York, U.S.A., 1997
26. Niemela, J., "Computer Prediction of Ames Test Mutagenicity (progress report, May 1998)," Danish EPA, working document, 1998 (available on request)
27. M-CASE Salmonella (Ames) Mutagenicity, model A2H, data evaluation, July 1999.

28. M-CASE Chromosomal Aberrations, model A61, data evaluation, July 1998
29. M-CASE Mutations in Mouse Lymphoma, model A2F, data evaluation, July 1998.
30. M-CASE Induction of Micronuclei, model A62, data evaluation, July 1998.
31. Zhang, Y.P., et. al., "Development of Methods to Ascertain the Predictivity and Consistency of SAR Models: Application to the U.S. National Toxicology Program Rodent Carcinogenicity Bioassays," *Quant. Struct-Act. Relat*, 16, 290-295, 1997.
32. Health Designs Inc., "TOPKAT 3.0 QSAR Module Summary: Rodent Carcinogenicity, v3carc.995, NY, U.S.A., 1997.
33. Health Designs Inc., "Computational Toxicology News, "MTA with FDA produces new Carcinogenicity QSTR's," no. 21, February 1998.
34. Cunningham, A.R., Klopman, G., Rosenkranz, H.S., "Identification of structural features and associated mechanisms of action for carcinogens in rats", *Mutation Research*, 405; 9-28, 1998.
35. Cunningham, A.R., Rosenkranz, H.S., Zhang, Y.P., Klopman, G., "Identification of "genotoxic" and "non-genotoxic" alerts for cancer in mice: The Carcinogenic potency database," *Mutation Research*, 398; 1-17, 1998.
36. Gold, L.S., et. al., "The Carcinogenic Potency Database," National Institute of Environmental Health Sciences, NIEHS Center, University of California, Berkeley, 1999
37. M-CASE, Binding to Aryl Hydrocarbon Hydroxylase, Model A68, July 1998.
38. Klopman, G., "MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program," *Quant. Struct.-Act. Relat.*, 11; 176-184, 1992.
39. Oxford Molecular Group, Inc., Health Designs, Inc., *Computational Toxicology, "TOPKAT 5.0 Reference Manual," 1997.*
40. Mathews, E.J., Contrera, J.F., "A New Highly Specific Method for Predicting the Carcinogenic Potential of Pharmaceuticals in Rodents Using Enhanced MCASE QSAR-ES Software", *Reg.Tox. and Pharm.*, 28; 242-264, 1998.
41. European Commission, "Technical Guidance Document in Support of commission directive 93/67EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No. 1488/94 on Risk Assessment for Existing Substances," Part III, ISBN 92-827-8013-9, 1996, p. 554.

42. Klopman, G., Saiakov, R, Rosenkranz, HS, Hermans, JLM, "M-CASE study of aquatic toxicity, I. Guppy," *Environ. Toxicol. Chem.* 18: 2497-2505.
43. G. Klopman and O.T. Macina, "Drug Design Based on Artificial Intelligence," *Computer Aided Innovation of New Materials II*, Elsevier Science Publishers, 1992, pp 1135-1140.
44. Mumtaz, M.M., et. Al. "Assessment of effect levels of chemicals from quantitative structure-activity (QSAR= models. I. Chronic lowest-observed-adverse-effect (LOAEL)," *Toxicology Letters*, 79 (1995), 131-143
45. Gomber, V. K., "Quantitative Structure-Activity Relationships in Toxicology; From Fundamentals to Applications," Health Designs Inc., 183 East Main Street, Rochester, NY 14604, USA.
46. O. Mekenyan, S. Karabunarliev and D. Bonchev, The OASIS Concept for Predicting the Biological Activity of Chemical Compounds, *J. Math. Chem.*, 4, 207-215 (1990).
47. O. Mekenyan, S. Karabunarliev and D. Bonchev, The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds, *Comput. & Chem.*, 14, 193-200 (1990).
48. Mekenyan, et. al. "COREPA Method Used for Evaluation of Reactivity Profile for High ER Binding Affinity," *Quant. Struct.-Act. Relat.*, 18:139-153, 1999.
49. Moudgal, C. J., Lipscomb, R. M. Bruce, "Potential health effects of drinking water disinfection by-products using quantitative structure toxicity relationships, *Toxicology*, vol 147, 2000, pp. 109-131.
50. Eriksson, L, Johansson, E., Keteth-Wold, N., and Wold, S. "Introduction to Multi- and Megavariate Analysis using Projection Methods (PCA & PLS," Umetrics, Sweden, 1999 pp. 251-260.
51. OECD Environmental Health and Safety Publications Series on Testing and Assessment, No. 24, "Revised Draft Guidance Document on the use of the Harmonized System of Classification of Chemicals which are Hazardous for the Aquatic Environment," Environment Directorate, Organization for Economic Cooperation and Development, Draft, Oct. 2000 (ENV/JM/HCL (2000) 15/REV2).
52. Registry of Toxic Effects of Chemical Substances (RTECS), The National Institute of Occupational Safety and Health, Washington, D.C, 1998 (Silver Platter electronic version).
53. M-CASE, Users Guide, Version 3.30 (Rev. 1.0), Multicase inc., 25825 Science Drive Park, suite 100, Cleveland, OH, 44122.
54. Rosenkranz, H.S., et.al, "Development, Characterization and Application of Predictive Toxicology Models," *SAR and QSAR in Environmental Research*, Vol. 10, pp. 277-298, 1999.

55. Wedebye, E.B., and Niemela, J. R., "Database with QSAR Estimates for more than 166,000 Chemicals," QSAR 2000, Ninth International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences, September 16-20 2000, Bourgas, Bulgaria. Abstracts, PV.2
56. Enslein, Kurt, Health Designs, Inc., 183 E. Main Street, Rochester, NY 14604: "QSTR applications in acute, chronic, and developmental toxicity, and carcinogenicity".
57. Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances.
58. ECETOC Technical Report No. 67, "The Role of Bioaccumulation in Environmental Risk Assessment: The Aquatic Environment and Related Food Webs", 1995.