GENERAL DISTRIBUTION

OCDE/GD(94)28

ENVIRONMENT MONOGRAPH NO. 88

US EPA/EC JOINT PROJECT ON THE EVALUATION OF (QUANTITATIVE) STRUCTURE ACTIVITY RELATIONSHIPS

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 1994

012625

FOR TECHNICAL REASONS, GRAPHS, TABLES & FACSIMILES ARE NOT AVAILABLE ON OLIS

ENVIRONMENT MONOGRAPHS

The OECD Environment Monograph series is designed to make selected technical reports available to a wide readership. The Joint Meeting of the Chemicals Group and Management Committee recommended that this report be derestricted. It has been made public under the authority of the Secretary-General. Copies of Environment Monographs on a limited basis can be forwarded on request. Please contact the OECD Environment Directorate, Environmental Health and Safety Division, 2 rue André-Pascal, 75775 Paris Cedex 16.

FOREWORD

In October 1989, an OECD workshop was held on notification schemes for new chemicals (see also Environment Monograph No. 35, A Survey of New Chemicals Notification Procedures in OECD Member Countries, OECD, 1990). One of the most important recommendations of this workshop was that an attempt should be made to evaluate the predictive power of the QSARs (Quantitative Structure Activity Relationships) used by the United States Environmental Protection Agency (EPA). It was also recommended that this evaluation be achieved by applying QSAR methods to chemicals for which extensive test data were already available, and then comparing the properties predicted by SAR with the properties observed from experimental testing.

The recommendations of the OECD workshop were the starting point for the collaborative project between the European Community and the United States described in this report. US and EC experts met on 14-16 October 1992, at the Umweltbundesamt in Berlin, to discuss the results of the joint project. Following that meeting, this final report was prepared for onward transmission to the OECD. On 25 May 1993, previous to the 20th Joint Meeting of the OECD Chemicals Group and the Management Committee of the Special Programme on the Control of Chemicals, a session took place on the "US/CEC Special Comparison of Estimated and Measured MPD [Minimum Pre-Marketing set of] Data".

RESUME

Un atelier de l'OCDE sur les systèmes de notification de nouvelles substances chimiques a été tenu en octobre 1989 (voir la Monographie sur l'Environnement No 35, Aperçu général des procédures de notification des produits nouveaux dans les pays Membres de l'OCDE, OCDE 1990). Une recommandation importante de cet atelier était qu'il fallait procéder à une évaluation de la valeur des relations quantitatives structure-activité mises au point par l'EPA (United States Environmental Protection Agency) en tant que moyen d'estimation. Il était également recommandé de faire l'évaluation de ces relations en les appliquant à des substances pour lesquelles beaucoup de données expérimentales étaient disponibles et de comparer les valeurs estimées avec celles obtenues par expérimentation.

Suite à ces recommandations, les Communautés Européennes et les Etats-Unis ont collaboré à un projet qui fait l'objet du présent rapport. Une première réunion, au cours de laquelle des experts américains et européens ont examiné les résultats, a eu lieu au Umweltbundesamt, à Berlin du 14 au 16 octobre 1992. A la suite de cette réunion un rapport final a été soumis à l'OCDE. Une seconde réunion intitulée "US/CEC Special Comparison of Estimated and Measured MPD Data" a été tenue à Paris, le 25 mai 1993, juste avant la 20e Réunion conjointe du Groupe des produits chimiques et du Comité de gestion du Programme spécial sur le contrôle des produits chimiques.

Sur recommandation de la Réunion conjointe, le rapport est mis en diffusion générale par le Secrétaire général sous sa propre responsabilité.

TABLE OF CONTENTS

1. BACKGROUND

In October 1989 the OECD organised, in the context of that organisation's Chemicals Programme, a workshop on notification schemes for new chemicals (see also Environment Monograph No. 35, A Survey of New Chemicals Notification Procedures in OECD Member Countries, OECD, 1990). The major objective of this meeting was to review, in the light of the 1981 OECD Council Act on the Mutual Acceptance of Data [C(81)30(Final)] and the 1982 Council Act on the Minimum Pre-Marketing Set of Data in the Assessment of Chemicals [C(82)196(Final)], the notification schemes applied by the Member Countries of the OECD. The 1982 Council Act recommended that countries require manufacturers/importers to supply a certain minimum pre-marketing data set (MPD) before placing a new chemical substance on the market: the test data to be generated experimentally using standard OECD testing guidelines.

From the information presented at the workshop, it was apparent that the majority of Member Countries had introduced notification schemes based on the principle of an MPD although the content of the testing package often diverged from that recommended in the Council Act. One notable exception to this general tendency was, however, the United States where the notification scheme for new chemicals established under the 1976 Toxic Substances Control Act (TSCA) did not, a priori, oblige manufacturers/importers to carry out testing before placing a new substance on the market. Essentially, the scheme established under TSCA required the submission of available data, often extremely limited, to the regulatory authority, in this case the Environmental Protection Agency (EPA). Faced with this paucity of experimental data, the EPA was obliged to place increasing reliance on techniques known collectively as (Quantitative) Structure Activity Relationships, (Q)SAR, in order to carry out a preliminary hazard/risk assessment of notified substances: (Q)SARs are predictive methods which estimate the properties (activity) of a chemical, e.g. melting point, vapour pressure, toxicity and ecotoxicity, on the basis of its structure.

One of the most important recommendations from the OECD workshop was that an attempt be made to evaluate the predictive power of the (Q)SAR used by the EPA (see Room Document No. 6 from the 13th Joint Meeting of the Chemicals Group and the Management Committee of the Special Programme on the Control of Chemicals). It was in addition recommended that this evaluation be achieved by applying the (Q)SAR methods to chemicals for which extensive test data were already available, and then comparing the properties predicted by SAR with the properties observed from experimental testing.

In the European Community, a new chemicals notification scheme came into force in 1981 in accordance with the rules laid down in Directive 79/831/EEC, being the sixth amendment to Directive 67/548/EEC on the classification, packaging and labelling of dangerous substances. The notification procedure required manufacturers/importers to submit a standardized data set (roughly similar to the OECD MPD) with experimental data being generated according to prescribed test methods (essentially equivalent to OECD test guidelines). By 1989, the EC notification scheme had been in force for over eight years and several hundred notifications had been received. The OECD workshop therefore

recommended that the predictive power of the (Q)SAR methods used by the EPA should be evaluated against the data submitted on chemicals in the context of the notification scheme established in the European Community.

The recommendations from the OECD workshop were the starting point for the collaborative project between the European Community and the United States, which is described in this report. It must be emphasized that the scope of this project was limited to that defined by the OECD workshop, namely an evaluation of the predictive power of the (Q)SAR techniques used by the EPA in the context of the new chemicals notification scheme established under the Toxic Substances Control Act. The project is not, and was not designed to be, an evaluation of QSAR techniques in general.

2. PROJECT DESIGN

2.1. Competent bodies

In the United States, the Agency responsible for processing the new chemicals notifications, and the body responsible for the realization of this collaborative project, is the Environmental Protection Agency.

In the European Community, each of the twelve Member Countries has designated national Competent Authorities responsible for the implementation of the notification scheme established under Directive 67/548/EEC as amended. The Commission of the European Communities is also involved in the implementation of the notification scheme, as well as being responsible for ensuring co-ordination between the Member States. For the purposes of this project, the Commission of the European Communities was mandated by the national Competent Authorities to act as the contact point with the EPA. For the detailed realization of the project, the input from the EC was co-ordinated by the Commission with advice and support from the national Competent Authorities.

Lists of the EPA and EC experts who were responsible for carrying out the detailed analyses upon which this report is based are included as Annex 1.

N.B.: New chemicals notification schemes in the United States and the European Community. In order to understand fully the design of the collaborative project, its implementation and the conclusions which can be drawn from it, it is essential to understand the details of the notification schemes as they are applied in the United States under the Toxic Substances Control Act and in the European Community under Directive 67/548/EEC as amended. Descriptions of the schemes are to be found in chapter 3 of this report.

2.2. Confidentiality

Directive 67/548/EEC, as amended, makes clear that the confidential data included in a notification dossier can only be made available to the national Competent Authorities designated as being responsible for implementing the Directive, and the European Commission. Within the national Competent Authorities and the Commission only a restricted number of staff are allowed access to this confidential information, and extensive measures are taken to ensure the physical security of this information.

Given the obligations imposed under the Directive, the confidential data submitted to the European Authorities could not be made available to the EPA without the specific permission of the manufacturers/importers who had submitted the notifications in Europe. Therefore, prior to the start of the project, the national Competent Authorities in the EC Member States wrote to all notifiers asking for permission to release confidential data to the EPA for the purpose of this collaborative project. It was made clear to the notifiers that the EPA had undertaken to accord the same degree of protection to confidential data submitted under this project as they would to confidential business information submitted as part of a new chemical notification under TSCA.

A total of 107 companies responded positively to the request made by the national Competent Authorities. A list of these companies is attached as Annex 2 to this report. The EPA, the national Competent Authorities and the European Commission would like to thank these companies for their assistance, without which this project could not have been carried out.

Confidential information exchanged between the EPA and the European authorities was taken by hand from the notification unit located in Direction General XI of the European Commission in Brussels to the mission of the United States to the European Commission. From there the information was transferred by diplomatic bag to the EPA in Washington. While in the EPA, the data were held in secure areas dedicated to the storage and processing of confidential business information. At the end of the project, confidential documents supplied to the EPA were destroyed.

2.3. How the project was organised

Discussions with EC notifiers regarding the release of confidential data to the US authorities were completed by December 1990. Altogether, companies gave permission for information on a total of 175 substances to be included in the project. Chemicals were removed from the study if, for example, they were on the original TSCA inventory or had been submitted under the US notification scheme and had been accompanied by the equivalent of the MPD. This reduced the test set of chemicals to a total of 144. The various use categories of substances notified under the EC scheme were reasonably well represented in this set of 144. The dates of notification ranged from 1983 to 1990. For the US, however, the scarcity of polymers and the inclusion of pesticides and pharmaceutical intermediates represents a somewhat atypical data set of chemicals and, as such, may not have been as good a match with the US experience as could be desired.

In autumn 1991, DG XI of the European Commission communicated to the EPA the following information in relation to each of the substances selected for the study:

- IUPAC name
- CAS number (where available)
- physical form
- melting point
- use (where this was adequately described in the original dossier).

Prior to the dispatch of information, the Commission and the national Competent Authorities were provided by the EPA with details of the (Q)SAR methods that the EPA would use during the collaborative project.

The EPA treated this input data in exactly the same way that they would have treated data submitted under the TSCA new chemicals notification scheme, applying (Q)SARs to predict the properties of the chemical and carrying out a preliminary hazard assessment. For each substance the EPA drew up a one to two page summary of their analysis. These summaries were delivered to DG XI of the EC Commission in March 1992, and thereafter to the national Competent Authorities.

In April 1992, DG XI communicated the full test dossiers on each of the 144 substances to the EPA.

Between April 1992 and September 1992, the US EPA on the one hand and the EC Member States/Commission (DG XI) on the other reviewed and analysed the result of the study. Between 14-16 October 1992, a joint meeting of US and EC experts took place at the Umweltbundesamt in Berlin to discuss the results of the project. Following that meeting, this final report was prepared for onward transmission to the OECD.

3. NOTIFICATION SCHEMES IN THE EUROPEAN COMMUNITY AND IN THE UNITED STATES

3.1. Essential features of the notification scheme for new chemical substances in the European Community

- **Overview/legal basis**

The new chemicals notification scheme is established within the framework of Directive 67/548/EEC on the classification, packaging and labelling of dangerous substances. The notification scheme was in fact introduced in the sixth amendment to the basic Directive (Directive 79/831/EEC), which came into force in the EC Member States in 1981. [A copy of the sixth amendment is attached as Appendix 1].

The obligation to submit a standard notification dossier harmonized at the level of the EC falls upon any manufacturer or importer wishing to place a <u>new</u> substance <u>on the market</u> in quantities greater than one tonne per annum per manufacturer. [Notice that the EC scheme is a pre-marketing scheme and not pre-manufacture as is the case in the United States.]

A "new substance" is defined as one that is not to be found on the European Inventory of Existing Commercial Chemical Substances (EINECS). EINECS contains over 100,000 chemicals on the EC market before 18th September 1981.

Even if a chemical is new, it may not need to be notified if it falls into one of the exempted product sectors, e.g. pharmaceuticals, or substance classes, e.g. polymers containing "old" monomers, which are specified in Articles 1 and 8 of the Directive respectively.

Notifiers are required to submit a notification dossier relating to the substance as marketed, including any impurities and additives necessary for keeping the substance stable but without separable solvent. This means that the substance or entity assessed is very rarely a pure substance and indeed some of the properties observed may be due to the impurities or additives present in the "substance". This means that the assessment is made on the entity to which man or the environment will actually be exposed rather than on the pure substance.

- **Information to be provided by the notifiers**

Notifiers must submit a notification dossier, including an extensive technical dossier containing the results of the experimental testing carried out on the substance. The contents of the technical dossier are laid down in Annex VII to the Directive. This standard testing package is known as the "base set" test dossier. When the marketing levels for a substance reach ten tonnes per annum per notification, the authorities may require further testing. When marketing levels reach 100 tonnes and 1000 tonnes per annum, the notifier is required to carry out further testing. These obligatory supplementary testing packages are known as the level 1 and level 2 testing packages respectively, and are laid down in Annex VIII to the Directive.

The testing methods to be used in carrying out testing of chemicals for the purpose of notification are laid down in Annex V to the Directive.

The "base set" test package is approximately equivalent to the OECD Minimum Pre-Marketing Data Set (MPD) and the testing methods in Annex V are, for the majority of tests, equivalent to the corresponding OECD Test Guidelines. Requiring testing according to agreed standard test methods has the distinct advantage of facilitating comparison of substances.

- **How does the notification scheme work?**

The notifier submits a notification dossier to the competent authority in the Member State where the substance is manufactured or imported. Forty-five days after the authority is in receipt of a dossier which conforms to the Directive, the notifier can place the substance on the market anywhere in the European Community.

The authority receiving the notification prepares a summary dossier which is circulated through the Commission in Brussels to the other eleven Member States (a copy of the summary dossier is attached as Appendix 2).

The other Member States and the Commission can request the lead authority to make changes to the dossier or ask the notifier for further information.

The essential feature to note about the notification scheme is that it is a decentralized one: the lead authority effectively takes the decision as to the acceptability of the notification dossier on behalf of the rest of the Community. In order for this decentralized approach to work effectively, the degree of flexibility/subjectivity the system can tolerate is rather small: it is not one single group of people which take the decisions, but twelve different national authorities each acting alone, with the Commission playing the role of co-ordinator. This is one of the main reasons for the perceived rigidity in the EC notification scheme, which is based upon a fixed set of information which must be supplied for each substance. This loss of flexibility is one of the costs to be paid for the benefit of having a notification scheme which has worked effectively across twelve different countries for over ten years.

- **Classification and labelling**

Directive 67/548/EEC as amended contains detailed and extensive rules for the classification and labelling of dangerous substances. Substances are classified on the basis of objective, often very precise, criteria which are laid down in Annex VI to the Directive (the version of Annex VI in force at the time of this study is included as Appendix 3). The classification criteria are in turn based upon the results of the tests carried out on the substance. The rules laid down in Annex VI also determine whether the labelling of a substance should carry a pictogram/symbol indicating certain types of danger, and also whether the label should indicate certain standard phrases describing the risk of the substance, so-called R-phrases, as well as certain standard phrases describing how the substance can be used safely, so-called S-phrases.

In addition to determining the labelling of a substance, the classification is the starting point for the risk assessment in the European Community and also drives downstream legislation concerned with aspects of risk management, e.g. worker protection.

As can be understood from the short description given above, classification and labelling, and in particular classification, are central elements in the EC chemicals legislation. However, the criteria for classification are often extremely precise: for example, substances are classified as "very toxic" if the acute oral LD50 is less than or equal to 25 mg per kilogram, but as "toxic" if the value is above 25 mg but less than or equal to 200 mg per kilogram. Classification schemes which demand such a high degree of precision to discriminate between substances allocated to one category or another obviously demand a high degree of precision in the estimates made of the chemical’s properties. Experimental testing does generate precise values and even though this precision may be more apparent than real, it does provide an effective basis for building an objective classification scheme. (Q)SAR methods, on the other hand, usually generate less objective/precise estimates of chemical properties, and therefore do not immediately lend themselves as input data constructing classification schemes.

3.2. Essential features of the notification scheme for new chemical substances in the United States

- **Overview/legal basis**

Persons who plan to manufacture or import a new chemical substance for a commercial purpose are required to provide the Environmental Protection Agency (EPA) with a pre-manufacture notification (PMN) at least 90 days prior to the activity. Section 5 of the Toxic Substances Control Act (TSCA) was designed to enable the Agency to review activities associated with manufacture, processing, use and disposal of any new chemical substance before it enters the market place. If necessary, the EPA is empowered to take action to prevent unreasonable risks before they occur (pollution prevention at its basic level). This is accomplished by requiring pre-manufacture reporting. [A copy of the relevant part of the TSCA is attached as Appendix 4.]

TSCA defines "new chemical substances" as chemical substances not listed on the TSCA Chemical Substance Inventory and not otherwise excluded by the regulations. The Inventory includes chemicals in commercial production between 1975 and 1979, and any chemicals reviewed in the PMN programme which have subsequently been commercially produced. The Inventory currently contains over 70,000 chemical substances, of which over 7500 substances have been added to the Inventory through the submission of notifications of commencement to manufacture (NOCs) after those substances had completed the PMN review process and were manufactured for commercial purposes.

The PMN programme has been in place since 1979 and, through fiscal year 1992, has reviewed over 21,500 notices. The Agency took action to protect health and the environment from potential risks posed for over 1800 of these new substances.

- **The PMN review process**

EPA developed the PMN review process to meet the statutory mandate of TSCA §5. Under the US programme, any person who intends to manufacture or import a new chemical substance is required to provide to EPA available data on the chemical structure, production, use, release, exposure, and health and environmental effects. However, section 5 does not require chemical companies to test their new chemical substances for potential toxic effects. Therefore, EPA’s review (and 5(e) regulatory actions) are often conducted in the absence of data. The Agency relies on Structure Activity Relationships (SAR) to make predictions concerning the environmental fate and effects (health and environmental) of PMN chemicals. Each PMN proceeds through a screening process to determine whether more detailed review is required and to identify candidates for regulatory action. The Structure Activity Team (SAT), made up of a multidisciplinary group of experts, is responsible for the initial assessment of fate and effects. EPA focuses on the relatively few new chemicals of greatest concern - those which are structurally related to known toxic chemicals, and those about which little is known.

a. Initial screen. PMN notices go through a multidisciplined initial review designed to ascertain whether regulatory action on a more detailed analysis is warranted. Preliminary chemistry, Structure Activity Relationship (SAR) analysis, exposure, and environmental fate analyses are conducted.

b. Use of SAR in hazard assessment. Given the qualitative and quantitative limitations of the test data provided with PMNs (over half of all PMNs contain no test data), the EPA has developed innovative approaches to characterize the potential hazards associated with new chemical substances. The major components of the EPA's SAR-based approach to hazard analysis are the following:

- critical review of submitted test data, if any, on the PMN chemical;

- identification and selection of potential analogues and/or prediction of key PMN metabolites, followed by critical review of test data available on these chemicals;

- use of QSAR (Quantitative Structural Activity Relationships) methods when available and applicable; and

- the experience and judgement of scientific assessors in interpreting, weighing, and integrating the often limited information yielded by the above hazard analysis components.

The TSCA PMN reporting requirements can be compared with the European Community's "pre-marketing" notification requirements. As the terms indicate, pre-manufacture notification under TSCA is required at an earlier point in the development of a chemical than is the case for the EC's pre-market notification procedure. Many of the information reporting requirements under the EC Directive are similar to those in TSCA, with the major difference that the EC Directive requires, as a mandatory part of the notification, a specified "base set" of health, environmental, and physico-chemical test data. Therefore, a minimum set of test data is available on pre-market notification EC chemicals, whereas the hazard assessment of TSCA PMN chemicals often starts out with fewer or no data.

c. Cases completing their initial review are brought to the first regulatory decision meeting, called "Focus". At this meeting, the results of the Initial Screen analyses are presented and considered and a decision rendered on each PMN case. The possible outcomes include: drop the case from review; hold it over for more investigation (standard review); or move directly toward a regulatory outcome for certain standard categories of chemicals. To date, the Agency has developed over 35 chemical "categories of concern" to facilitate the new chemicals review process.

d. For chemicals which are not screened out early, the standard review includes:

- conducting a chemistry analysis;

- identifying structurally analogous substances;

- searching the literature for toxicity data,

- analysing test data on the substance or analogous substances;

- analysing potential releases to the environment;

- estimating exposures to workers and the general population;

- estimating potential concentrations in surface waters;

- investigating additional uses which could significantly alter exposure.

e. Cases completing standard review are taken to the PMN Disposition Meeting for a final decision. The meeting can result in a decision to drop a case from further review, to regulate (and require controls) under section 5(e) or 5(f) (see below), or to "ban" the substance pending the receipt and evaluation of "upfront testing".

f. If a regulatory decision to impose certain controls on the manufacture, process, use, distribution, or disposal of a new substance is reached, the EPA staff communicate and negotiate with the submitter. Similarly, if "up-front" testing is recommended in face of banning the new substance, this decision is also communicated to the submitter by the EPA staff.

g. Notice of Commencement (NOC) of Manufacture or Import. An NOC must be submitted within 30 days of commencement of commercial production of a chemical substance which has completed the 90-day review period. The substance is then added to the TSCA Inventory.

- **Regulating new chemical substances under TSCA**

Section 5(e) and 5(f) of TSCA authorize the EPA to prohibit or limit the manufacture, processing, distribution in commerce, use, and disposal of a new chemical substance if the EPA makes the following determinations:

a. Section 5(e) findings:

- Available information on the substance is insufficient to permit a reasoned evaluation of its health or environmental effects; and

- (1) The manufacture, processing, distribution in commerce, use, or disposal of the substance <u>may present</u> an unreasonable risk of injury to health or the environment (referred to as a "may present" or risk-based determination); or

- (2) the substance will be produced in substantial quantities and (A) may reasonably be anticipated to enter the environment in substantial quantities, or (B) there may be significant or substantial human exposure (referred to as an "exposure-based" finding). An exposure-based review is triggered by an estimated threshold production volume of 100,000 kilograms per year. For those substances meeting significant or

substantial human exposure criteria, chemical manufacturers may be asked to perform some or all of the following tests on their PMN substance: an Ames assay, an in vivo mouse micronucleus test, a 28-day (oral) repeat dose toxicity test, and an acute oral toxicity test. PMN substances meeting the environmental release criterion may be tested for algal acute toxicity, daphnid acute toxicity, and fish acute toxicity. Additional elements of the exposure-based testing policy may include environmental fate testing and, for PMN substances having higher production volumes, developmental toxicity testing requirements.

b. Section 5(f) findings:

- There is a reasonable basis to conclude that the manufacture, processing, distribution in commerce, use, or disposal of the substance will present an unreasonable risk of injury to human health or the environment before a TSCA §6 rule can be issued to prevent the risk (referred to as a "will present" determination):

- A section 5(f) rule, which limits activities involving a new substance, is a section 6(b) proposed rule which is immediately effective upon proposal. A section 5(f) order prohibits all activities involving the substance. (To date, EPA has issued three section 5(f) rules and no section 5(f) orders, although a number of PMNs have been withdrawn from review after EPA notified the submitters that the Agency intended to ban the substances.)

c. Practices under section 5(e):

To date, there have been five outcomes, depending upon the facts of the case, when EPA has made a determination under section 5(e):

- The company may withdraw the PMN.

- The company may develop toxicity information sufficient to permit a reasoned evaluation of the health or environmental effects of the substance prior to the conclusion of the review period ("upfront" or "voluntary" testing). Where exposures or releases cannot be controlled pending testing to address EPA’s concerns, or the requested testing is relatively cheap and not very time-consuming, this may be the only option available to the PMN submitter short of withdrawing the PMN.

- The company may develop and provide to EPA other information on the potential effects of the substance or its analogues, the potential exposures, or both, which if accepted by the Agency, would negate the potential unreasonable risk determination.

- The company may, together with EPA, suspend the notice review period, and negotiate and enter into a section 5(e) Consent Order. The Consent Order would permit limited manufacture, processing, distribution in commerce, use, and disposal of the substance pending the development of information. A Consent Order may contain a requirement that toxicity data be submitted to EPA when a specified volume of the chemical has

been produced. This production volume level is set where EPA estimates that profits from the chemical will support the cost of testing.

- The company may refuse to withdraw the PMN, negotiate a Consent Order with EPA, and/or conduct up-front testing or develop other information. EPA would then unilaterally develop a Proposed Order, under the procedures in section 5(e), to ban manufacture or import.

4. RESULTS

4.1. Introduction

For this project, the test set of chemicals comprised a maximum of 144 substances (sometimes fewer, depending upon the end-point and the results available). Each substance was assigned a number and is referred to in the report by means of that number. A short generic description of each substance included in the project is given in Annex 3.

In the sections which follow, the results are generally presented in a summary form, not substance by substance. However, detailed annexes presenting the results by end-point and by substance are appended to the report.

4.1.1. Evaluation criteria

For each end-point, specific criteria were agreed between the US and EC experts for assessing the "success", "failure", "hit-rate" of the (Q)SAR methods, e.g. for most physico-chemical and the ecotoxicity data, agreement was defined as being reached if the difference between measured and predicted value did not exceed a factor of 10. In addition to these end-point specific criteria, the following, more general, considerations were also taken into account in relation to each end-point.

- Can the predicted data be used on a one-to-one basis in the place of the test results foreseen in the OECD Minimum Pre-Marketing Data Set (MPD) or other similar test-based notification schemes?

- Can the results of the predictive approach be used in the context of schemes for the classification and labelling of chemicals which employ pre-defined cut off values?

- If estimated values based on predictive methods are used instead of test data for the purposes of preliminary hazard assessment, are the predictive methods sufficiently reliable in relation to each end-point and what is the likelihood of false negatives in relation to each end-point?

- The OECD MPD and other test-based systems for screening of new chemicals frequently do not include important end-points. To what extent do predictive methods allow one to go beyond the scope of fixed data sets and to assess additional end-points?

4.1.2. Complicating factors

Issues addressed with regard to each end-point are discussed in connection with that end-point. Nevertheless, a number of common problems can be identified which complicated the comparison of predicted and observed results in relation to all end-points.

- **Pure substances vs. notified substances**

In the EC notification scheme substances are notified essentially as they are marketed, including impurities but minus any separable solvent. This means that impurities or non-separable solvents may contribute significantly to the observed properties. In contrast, the (Q)SAR methods are based on pure substances and impurities are only taken into account in the US system if the manufacturer is aware of their existence/identity and reports this information to EPA.

For the above reason, the (Q)SAR methods will often fail to predict properties which are due to the presence of impurities.

- **Effect quantification**

Experimental data reported from the EC notification dossiers may display considerable variability (extremely wide confidence limits). Furthermore, both predicted and experimental data were often expressed as >n, or as <n or as ranges. In these cases, agreements had to be reached end-point by end-point as to how to make effective comparisons.

- **End-point selection**

When considering properties such as acute aquatic toxicity or biodegradation, the precise end-points addressed by the experimental testing and the (Q)SAR predictive methods were sometimes different, e.g. 24-hour toxicity as opposed to 48-hour; "ready biodegradability" as opposed to an estimate of the time required for complete biodegradation. Again in such cases, agreement had to be reached on a realistic basis for comparison.

- **Descriptive narrative assessment vs. numerical data**

(Q)SAR methods frequently generate predictions placing substances in concern categories such as low, medium or high. Again, agreement had to be reached as to how such predictions should be compared with an objective value such as a numerical (e.g. 35 mg/kg bodyweight/day) Lowest Observed Adverse Effect Level (LOAEL) in a 28-day repeated dose toxicity study.

- **Nominal vs. measured concentrations**

Test results for aquatic toxicity test, in the EC notification dossiers, particularly dossiers received early in the life of the notification scheme, were frequently based upon nominal rather than measured substance concentrations. In such cases it is entirely possible that the predicted value for aquatic toxicity generated by (Q)SAR is nearer to the "real value" than the result reported from the experimental determination.

4.2. Detailed analysis of results

A detailed description of the end-point by end-point comparison of the values predicted by (Q)SAR and the values generated by experimental determination in the EC notification dossiers is given below. For ease of presentation, the abbreviations "EC" or "EPA" have been used as a convenient shorthand to identify the approaches used in the European Community and the United States Environmental Protection Agency respectively.

4.2.1. Physico-chemical and environmental fate parameters

4.2.1.1. Boiling point

For predicting the boiling point, EPA experts use estimation methods, e.g. PCGEMS (Meissner's method), data on analogues, and experimentally determined data obtained from the published literature investigations. Impurities are in general neglected in the predictions. The application of the estimation techniques was not possible for all the chemicals within this study.

Even though the boiling point is required for notified chemicals at "base set" level in the EC, for many substances in this study experimentally determined boiling points were not available as it was technically not possible to conduct the tests.

The boiling point is used to characterize the material, it is not directly used for risk or safety evaluations. The boiling point may serve as an input parameter for estimating vapour pressure, if the latter is unavailable from experiment.

Only for 30 chemicals out of the 144 were measured/estimated boiling point values available for comparison. The following criteria were applied for the analysis:

- for all values assigned with <n or >n, the signs are deleted and the values are directly compared;

- the values are considered to be in agreement if the difference between calculated and measured data does not exceed ± 50 degrees C.

The comparison of the SAR and MPD data is given in Table 1; for detailed analysis of the boiling point data, see Annex 4.

TABLE 1: Comparison of boiling point data

	N° of chemicals	%
Total	30	100
Agreement	15	50
Disagreement	15	50

If the literature data were included in the analysis, an additional eleven chemicals would be added, for which the US boiling points were all in agreement with the EC data. The agreement was below 50% for solid substances.

- **Conclusions**

The data set for analysis was very small, so only limited conclusions are possible. The boiling point is not used directly in the hazard/risk assessment nor is it used in the classification schemes. On the other hand, the boiling point is a basic piece of information about a chemical which manufacturers should normally be aware of; furthermore, boiling point determination by testing is relatively inexpensive. Thus it is concluded that it is preferable, in the EC scheme, to continue to measure the boiling point when it is technically possible to do so.

4.2.1.2. Vapour pressure

The vapour pressure of the chemicals under consideration is predicted by the EPA using methods based on the Antoine equation or the Watson equation, or by applying the PCNOMO technique. The vapour pressure contributes indirectly to the EPA's risk assessment, as it is used as an input parameter to the exposure and fate analysis.

Also within the EC risk assessment, the vapour pressure serves as a basic parameter for human health and environmental exposure evaluation. Measured vapour pressure data are required at "base set" level in the EC; however, calculation methods can be used according to Annex V for range finding purposes, for justifying the non-performance of the test, or for providing an estimate or limit value in cases where the experimental method cannot be applied due to technical reasons (including where the vapour pressure is very low).

For 113 chemicals out of the 144 test chemicals measured, data on vapour pressure were available, and predictions were available for all chemicals. The predictions are given in the majority of the cases as upper/lower bounds. In order to compare the SAR values with the measured data, all values were converted to like units (torr). The following criteria for comparison analysis were applied:

- for all values assigned with <n or >n, the signs are deleted and the values are directly compared;

- the lower limit is set at 10^{-6} torr. All SAR and MPD values that are less than this value are arbitrarily set to 10^{-6} torr;

- the values are considered to be in agreement if they are within ± 1 log unit.

The results of the comparison of the SAR and MPD data are given in Table 2; the detailed analysis of the vapour pressure data is to be found in Annex 5.

TABLE 2: Comparison of vapour pressure data

	N° of chemicals	**%**
Total	113	100
Agreement (± 1 log unit)	71	62.8
Disagreement	42	37.2
- of these, predictions which were not at all in agreement (>3 log units difference)	[23]	[20]

The data pairs which show big deviations were more rigorously investigated: in some cases the disagreement can be put down to the fact that the material used for the experimental determination contained volatile impurities, whereas the predictions are carried out for the pure substance.

- **Conclusions**

The best agreement was observed between the PCNOMO estimates and the measured values. In general, the predictions tend to underestimate the vapour pressure. Assessing the deviations with respect to chemical classes is not possible with the small data set available. Imprecise predictions of very high or very low vapour pressure do not affect the overall assessment, but more precise values are needed in the decision-relevant range. Vapour pressure contributes to the exposure portion of the risk assessment in the EC and the US; however, it is not normally used for the purpose of classifying chemicals within the EC classification scheme. Under/overestimation of vapour pressure can result in an under/overestimation of the exposure associated with a chemical and thus contribute to an under/overestimation of the risks. The majority of methods for the experimental determination of vapour pressure are relatively inexpensive, and therefore notification schemes based upon testing will probably continue to require experimental determination. Schemes based upon predictive methods may need to be adjusted to foresee a more systematic

approach to the experimental determination of this parameter for some of the chemicals which are identified as being of concern on the basis of a preliminary hazard/risk assessment.

4.2.1.3. Water solubility

The methods used by the EPA experts for predicting water solubility are based on log P_{OW} values (PCGEMS). However, most new chemicals do not match the application criteria of the available QSARs, e.g. applicability recommended for liquid substances or only for certain log P_{OW} ranges. Within EPA hazard/risk assessment scheme, water solubility serves as an input parameter for the environmental fate analysis and ecotoxicity assessment. The lower prediction limit for fate and ecotoxicity assessment is $\leq$1 ug/l; for some other purposes it may be around 1 mg/l. In cases of concern, e.g. for chemicals with higher production volumes, measured water solubility is required.

In the EC, experimentally determined water solubility data, which are required at "base set" level, are also used in environmental exposure assessment; they may also contribute to the classification "dangerous for the environment".

Measured numerical values were not available for 13 of the 144 chemicals, as their determination was technically not possible, but in six cases out of the 13, qualitative test data were available which could be used for comparison. In four further cases the SAR data cannot be used for the comparative analysis. This means there were 133 data pairs for comparison. An additional problem affecting meaningful comparison is the lack of precision in the data (both predicted and measured): many data, in particular the majority of the predicted data, are given as ranges or upper/lower bounds; in case of measured data the values given as bounds are mostly without an indication of detection limit.

The following criteria were applied for the comparison analysis:

- for all values assigned with <n or >n, the signs are deleted and the values are directly compared;

- for data given as ranges, the average is taken for comparison;

- the lower limit is set at 0.01 mg/l and the upper limit at 10,000 mg/l. All SAR and MPD values that are less than the lower limit value, or above the upper limit value, are arbitrarily set to 0.01 mg/l or 10,000 mg/l, respectively;

- the values are considered to be in agreement if they are within ± 1 log unit.

Results of the comparison between SAR and MPD data are given in Table 3; the detailed analysis of water solubility data in Annex 6.

TABLE 3: Comparison of water solubility data

	N° of chemicals	%
Total	133	100
Agreement (± 1 log unit)	90	67.7
Disagreement	43	32.3

A rigorous scientific analysis of the estimated and measured data for water solubility was not possible due to the imprecise nature of both data sets. Tendencies of over- or underestimation of water solubility are not observed. A relatively high rate of disagreement is detected for low solubility values (<1 mg/l).

- **Conclusions**

Water solubility is a significant parameter in risk assessment and might have a decisive impact on the classification "dangerous for the environment". Under/overestimation of water solubility can result in a under/overestimation of exposure and thus contribute to an under/overestimation of the risks. SAR-based predictions may not always be of sufficient reliability, especially in the range of low solubility, i.e. <1 mg/l, due to the complexity of factors influencing a chemical's water solubility. The experimental determination of water solubility is relatively inexpensive, therefore notification schemes based upon testing will probably continue to require experimental determination. Schemes based upon predictive methods may need to be adjusted to foresee a more systematic approach to the experimental determination of this parameter for chemicals at higher production levels or which are identified as being of concern for the aquatic environment on the basis of a preliminary hazard/risk assessment.

4.2.1.4. Partition coefficient

The partition coefficient is a key parameter to evaluate a chemical's impact on the environment.

Furthermore, its particular importance is underlined as, in the SAR methodologies, several other predictions, e.g. ecotoxicity/toxicity, are based upon it. The SAR prediction methods applied by EPA use the MedChem ClogP Software package; the respective estimations are based on a fragment method. In cases of missing fragments, their values are estimated from expert knowledge. The upper prediction limit applied by the EPA for fate assessment is log $P_{OW} \geq 6$. For ecotoxicity assessment, no upper limit is considered for some chemical classes.

In the test-driven, stepwise assessment scheme of the EC, the partition coefficient is also used in the decision taking process on further testing (e.g. for bioaccumulation potential); in addition, the log P_{OW} contributes to the criteria for classification as "dangerous for the environment" within the EC classification scheme: the log P_{OW} value 3 represents the cut-off value for decisions on further testing and for classification. The EC notification scheme requires experimentally determined partition coefficient data at "base set" level. Nevertheless, Annex V recommends to estimate log P_{OW} for deciding which of the experimental methods is appropriate, for selecting appropriate test conditions and for providing a calculated log P_{OW} in cases where the experimental methods cannot be applied for technical reasons. Therefore, in a number of cases, only estimated values were available in the EC dossiers. Those values were not taken into consideration for the comparative analysis of the SAR/MPD data.

Eighty-two chemicals with both measured and predicted log P_{OW} values are available for the comparative study. The analysis included the application of the following criteria for comparison:

- for all values assigned with <n or >n, the values are directly compared;
- for values given as ranges, the arithmetic average is used;
- the lower limit is set at log P_{OW} = 0; all values that are below 0 are arbitrarily set to 0;
- the upper limit is set to log P_{OW} = 6; all values above 6 are arbitrarily set to 6;
- the values are considered to be in agreement if they are within ± 1 log unit.

The results of the comparison of the SAR and MPD data are given in Table 4; the detailed analysis of log P_{OW} is attached (see Annex 7).

TABLE 4: Comparison of partition coefficient data

	N° of chemicals	**%**
Total	82	100
Agreement (± 1 log unit)	50	61
Disagreement	32	39
- Overestimation	25	30.5
- Underestimation	7	8.5

- **Conclusions**

The log P_{OW} estimates are in general reasonably accurate. However, estimations are in poor agreement for certain classes of compounds (e.g. dissociated compounds, charged compounds, surfactants, chelating compounds, organometallics, organophosphorous compounds, compounds with unknown fragment values, UVCB compounds) and are not applicable for them. Calculated log P_{OW} values above 4 tend to overestimate. Calculations in the range of 0-2 possibly underestimate log P; however, the data set available is too small for exhaustive analysis. EPA calculation methods are in general successful at calculating log P values <0.

The results of this exercise indicate that the predictive methods for log P_{OW} may be of further importance in the EC in future, i.e. submission of predicted log P_{OW} values by the notifiers instead of measured data might be regarded as a possible option. However, the log P_{OW} range around the value 3, which is of particular importance for the EC classification and stepwise risk assessment scheme, will anyhow have to be taken into special consideration and may continue to require experimental determination as well as in the case of suspected underestimation.

4.2.1.5. Biodegradation

The data on this end-point were difficult to compare because different scales/definitions are used. The biodegradation estimates are given in semi-quantitative terms, indicating the appropriate time for complete degradation ("days", "days to weeks", "weeks", "weeks to months", "months" or "months or longer"), whereas the OECD-based standard 28-day tests, which are available in the EC at "base set" level, result either in the decision "readily biodegradable" or "not readily biodegradable".

The EPA predictions concern biodegradability in terms of primary and ultimate biodegradability, using structural analogies with previously studied chemicals. The applied estimation methods are based on expert judgement. The biodegradation predictions are used within EPA risk assessment scheme as an important factor of the environmental fate analysis.

Biodegradation data are required in the EC for risk assessment and also for the classification "dangerous for the environment".

One hundred fifteen substances were available for comparison of predicted with experimental data. By relating estimates of "days" and "days-weeks" to the definition "readily biodegradable", five of the nine substances experimentally determined as being readily biodegradable have been identified as such by the predicting methods (= 55.5%). The other four readily biodegradable substances are predicted to degrade in "weeks", "weeks-months" or "months or longer". At the same time, for four substances which did not pass the experimental criteria for ready biodegradability, a rapid degradation was predicted ("days-weeks"). In general, as the predictive methods indicated increasing time required for complete degradation, the better they correlated with test results indicative of a lack of ready biodegradability. The overall results of the comparative study are summarised in Table 5; the detailed analyses of the data is to be found in Annex 8.

TABLE 5: Comparison of biodegradation results

Test result	**Prediction**	
	correct	incorrect
Total	107 (93%)	8 (7%)
Readily biodegradable	5	4
Not readily biodegradable	102	4

- **Conclusions**

The EPA methods are likely to identify those substances which are not "readily biodegradable", i.e. slowly degrading chemicals. However, they do not appear to work as well in identifying chemicals which readily degrade. The use of biodegradation predictions as a tool for establishing suitable testing strategies within a stepwise assessment scheme is considered a possible option for the future in the EC. On the basis of EPA results, it appears that if the predicted biodegradability is "weeks" or longer, testing for "readily biodegradability" would not be indicated. Instead, a test for inherent degradability or another suitable test that provides further information on the biodegradation process should be carried out. If the predicted biodegradability is "days" or "days-weeks" corresponding to "readily biodegradability", then a "ready biodegradability test" would be needed for confirmation.

4.2.1.6. Hydrolysis

EPA dossiers include data hydrolysis only if it is likely to occur. The applied estimation methods evaluate the rate of hydrolysis if relevant (hydrolysable) functional groups are present in the molecule. For a few compound classes, the HYDRO programme is applied. Hydrolysis tests are not mandatory in the EC at "base set" level; for 41 of the chemicals included in this study, hydrolysis data were given. Only for six chemicals were both measured and predicted hydrolysis data available. A comparative analysis of this end-point was therefore not carried out.

4.2.1.7. Soil sorption

The environmental fate analysis carried out by EPA includes in general the prediction of log K_{OC}. For the majority of the chemicals within this study log K_{OC} predictions were available. The applied estimation methods are mostly based on log P_{OW}, but they are of limited applicability. The fragment method can be applied more widely, but it also does not satisfy all requirements.

Under the sixth amendment, no tests on soil sorption are required in the EC; for notifications according to the seventh amendment, a screening test on adsorption/desorption will be mandatory. For this study, no test results were available for comparison.

4.2.1.8. Photodegradation

The environmental fate analysis of the EPA experts includes estimates of the photolysis of the substance (direct and indirect) in water. Measured photolysis data are not required at "base set" level and are therefore in general not available. A comparative study is not possible on the data available.

4.2.2. Ecotoxicity parameters

4.2.2.1. Toxicity to aquatic organisms

For predicting aquatic toxicity, approximately 300 SAR models are available to the EPA experts for various (about 100) chemical classes. The estimation methods are mostly based on log P_{OW}; only calculated values of this latter parameter are used. Expert knowledge is required for the selection of the appropriate SAR model. The selection is based on the chemical class, not on the mode of action. The EPA's SAR predictions cover both acute and chronic toxicity for aquatic organisms. Fish, daphnia, algae and, for some pesticide structures, also vascular plants are considered. For some chemical classes, if log P_{OW} is above 5 it is assumed that there are no acute toxic effects. Nevertheless, for those substances, and similarly for chemicals for which no toxic effect is predicted at the water solubility limit, chronic effects may still be substantial. The data on aquatic toxicity are used for risk assessment and assignment of "level of concern".

In the EC, according to the requirements of Directive 79/831/EEC (sixth amendment) at "base set" level, normally only acute fish and daphnia studies are conducted. Chronic effects and effects on species other than fish and daphnia, e.g. algae, are in general not addressed at this stage. The aquatic toxicity data are used for risk assessment and for the classification "dangerous for the environment".

In several cases, the data were given as >n, <n or as NTS (Non Toxic at Saturation). LC/EC50 data given as <n are difficult to interpret because, in those cases, the actual LC/EC50 value can be much lower than the given limit. For this reason, those data were excluded from analysis. Values given as >n, however, can be used because, usually, the given limit will be regarded as a worst case estimate of the toxicity. The analysis includes therefore those chemicals for which exact and "higher than" (>n) effect concentrations are supplied; data presented as NTS are also included.

The comparative analysis is carried out applying the following criteria:

- for all values given as >n, the numbers are directly compared without considering the signs;

- for data pairs with both values above 100 mg/l, no differentiation is made between the numerical values: the ratio of estimated/measured value therefore is 1;

- the values are considered to be in agreement if they are within ± 1 log unit;

- for data pairs in which one value is given as NTS and the other as a numerical value, the results are assessed considering the water solubility: for a numerical value much higher than the water solubility (>100 mg/l) the SAR and experimental value are deemed to be in agreement; for effect concentrations closer to the water solubility (<100 mg/l) the two values are deemed to be inconsistent with one another (disagree).

The results of the comparative analyses are given in Table 6 (toxicity to fish) and Table 7 (toxicity to daphnia); the detailed analyses are given in Annexes 9 and 10.

TABLE 6: Comparison of data on toxicity to fish

	N° of chemicals	**%**
Total	130	100
Agreement	107	82.3
Disagreemen	23	17.7
- Overestimation	14	10.8
- Underestimation	9	6.9

TABLE 7: Comparison of data on toxicity to daphnia

	N° of chemicals	**%**
Total	127	100
Agreement	90	70.9
Disagreement	37	29.1
- Overestimation	20	15.7
- Underestimation	17	13.4

Some of the differences in predicted and experimental toxicity can be attributed to nominal instead of measured concentrations, the use of solvents to enhance water solubility, and to different test durations (24/48 hr for daphnia). For only five chemicals were measured and predicted data on algae toxicity available. In four cases, agreement between SAR/MPD data is observed (data: see Annex 11).

- **Conclusions**

Information on aquatic toxicity is used both for risk assessment and for classification purposes. Overall, SAR predictions of aquatic toxicity are quite good. For fish toxicity, the predictions tend to overestimate the toxicity. For daphnia, over- and underestimations occurred at about the same rate. Further effort is desirable to explain the cases where the reason for the underestimation (false negative predictions) is not evident. Nevertheless, if used with the required caution, SAR predictions can be very effective in the context of the US notification scheme.

The predictions are considered to represent a very useful future option to support the decision taking process within a stepwise risk assessment scheme for carrying out toxicity tests.

4.2.2.2. Classification "dangerous for the environment"

The EC scheme for classification "dangerous for the environment" is driven by toxicity, biodegradability and/or bioaccumulation potential. For certain types of substances (those which show low solubility in water), the water solubility may also be taken into account when determining the final classification.

The EC classification criteria and the resulting risk phrases (R-phrases) for the aquatic environment are as follows:

R 50: Very toxic to aquatic organisms

Acute toxicity:	96 hr LC 50 (for fish)	≤1 mg/l
or	48 hr EC 50 (for daphnia)	≤1 mg/l
or	72 hr IC 50 (for algae)	≤1 mg/l

R 50: Very toxic to aquatic organisms
and
R 53: May cause long-term adverse effects in the aquatic environment

Acute toxicity:	96 hr LC 50 (for fish)	≤1 mg/l
or	48 hr EC 50 (for daphnia)	≤1 mg/l
or	72 hr IC 50 (for algae)	≤1 mg/l

and the substance is not readily degradable
or the log P_{OW} ≥3.0.

R 51: Toxic to aquatic organisms
and
R 53: May cause long-term adverse effects in the aquatic environment

Acute toxicity:	96 hr LC 50 (for fish)	1 mg/l< LC 50 ≤10 mg/l
	or 48 hr EC 50 (for daphnia)	1 mg/l< EC 50 ≤10 mg/l
	or 72 hr IC 50 (for algae)	1 mg/l< IC 50 ≤10 mg/l

and the substance is not readily degradable
or the log P_{OW} ≥3.0.

R 52: Harmful to aquatic organisms
and
R 53: May cause long-term adverse effects in the aquatic environment

Acute toxicity:	96 hr LC 50 (for fish)	10 mg/l< LC 50 ≤100 mg/l
	or 48 hr EC 50 (for daphnia)	10 mg/l< EC 50 ≤100 mg/l
	or 72 hr IC 50 (for algae)	10 mg/l< IC 50 ≤100 mg/l

and the substance is not readily degradable.

R 53: May cause long-term adverse effects in the aquatic environment

Substances not falling under the criteria above, but which, on the basis of the available evidence concerning their persistence, potential to accumulate, and predicted or observed environmental fate and behaviour may nevertheless present a long-term and/or delayed danger to the structure and/or functioning to the aquatic ecosystems.

For example, poorly water soluble substances, i.e. substances with water solubility <1 mg/l, will be covered by this criteria if:

a) they are not readily degradable
b) and the log P_{OW} ≥3.0.

Further details are to be found in the complete EC classification and labelling guide which is attached as Appendix 3.

In this comparative study, the EPA's quantitative predictions are used to classify the chemicals according to the EC criteria. The results are compared to those classifications based on the measured data. All 144 chemicals in the project were classified for the comparison purpose on the data available, independent of whether the data sets - both measured and predicted - were complete or not. The comparison and the results are given in Tables 8 and 9.

TABLE 8: Comparison of classification "dangerous for the environment" according to the EC scheme based on MPD vs. SAR data

Classif. based on MPD data	Total	Classification based on SAR data				
		N.c.*	R53	R52/53	R51/53	R50/53	R50
Not class.	48	28	6	6	3	3	2
R53	23	2	17	-	-	4	-
R52/53	26	8	4	4	7	3	-
R51/53	34	5	3	3	14	9	-
R50/53	13	1	2	1	2	7	-
R 50	-	-	-	-	-	-	-
Total	144	44	32	14	26	26	2

*** Not classified**

TABLE 9: Result of the comparison of classification "dangerous for the environment"

	N° of chemicals	%
Total	144	100
Agreement	70	48.6
Disagreement	74	51.4
- Overclassification	43	29.9
- Underclassification	31	21.5

- **Conclusions**

The overclassifications can be considered acceptable as being conservative. The agreement of 78% when including the overclassifications is encouraging, even though the underclassifications give cause for concern since potentially dangerous substances may not be recognized.

The concordance in classification of chemicals "dangerous for the environment" is in general reasonably good. However, for the purpose of classification within a legislative scheme, the use of measured data is clearly preferable.

4.2.3. Toxicological properties/health effects

4.2.3.1. Absorption

The likely extent of absorption of a chemical via skin, lungs and gastro-intestinal tract is predicted by the EPA experts on the basis of the physico-chemical properties of the chemical (particularly log P_{ow}, which is usually a predicted value, and the physical form of the chemical). The initial opinion on this basis may be modified in the light of any available test data on the chemical itself or on a closely related structural analogue. Good, moderate, poor or no absorption will be predicted for each route of exposure (dermal, inhalation and oral).

The prediction of the likely extent of absorption following exposure by a particular route will be used when taking decisions on whether the chemical may present an unreasonable risk to human health and/or on testing requirements in the USA.

Absorption is not investigated in the base set level testing in the EC, but whether any absorption has occurred can be inferred to an extent from evidence of systemic toxicity in the acute and repeated dose studies. It is less easy to decide that absorption has not occurred - the chemical may be well absorbed and show no systemic toxicity in the particular test(s) already conducted. However, it may cause adverse effects in other test systems not yet applied. Evidence of absorption (i.e. systemic effects) may have an influence on the timing of further testing. When there is no evidence from the currently available test data, the timing of further testing may be influenced by the likelihood of absorption based on the physico-chemical properties of the chemical and/or the extent of human exposure expected.

- **Conclusions**

There were too few studies conducted using the inhalation route for an accurate assessment of concurrence between SAR calls for absorption from the lungs and derived absorption estimates from toxicity test results.

Based on the 136 chemicals for which dermal toxicity studies were available, it is considered that acute dermal studies are inadequate to judge dermal absorption. There were too few 28-day studies to serve as a basis for definitive judgement on dermal absorption calls.

The SAR calls for gastro-intestinal absorption were essentially in agreement with estimates based on the oral toxicity test results: when they differed it was only in degree of absorption and not, with one exception, giving a completely different assessment of whether or not a chemical was absorbed at all. For some chemicals, which were classified in the EC on the basis of their oral toxicity, the relatively low extent of absorption predicted may be of some concern. However, none of these chemicals were predicted to have "no absorption" and thus would not have been dropped from EPA evaluations.

4.2.3.2. Acute toxicity

Acute toxicity data are used to predict the potential effects in humans of a single exposure to a chemical (e.g. during maintenance work or in an accident). They are also used to help in setting dose levels for other toxicity tests.

Prediction of acute toxicity is not emphasised in the EPA evaluation of a new chemical, which focuses on long-term or sub-chronic effects. For the purposes of this project, however, predictions of acute toxicity following oral administration were made. (There were too few chemicals with data from inhalation or dermal acute toxicity tests which were suitable for conducting comparisons of the two approaches to evaluation.)

The following criteria were used to rank chemicals on the basis of their oral LD50 values, and so provide a means of comparing the predicted toxicity with that observed in the tests:

Oral LD50 (mg/kg)	**Toxicity**
> 2000	Low (L)
> 1000 < 2000	Low-Medium (L-M)
> 500 < 1000	Medium (M)
> 50 < 500	Medium-High (M-H)
< 50	High (H)

These criteria give more categories of acute toxicity than are conferred by the EC classification system (below), but the same criterion (LD50 >2000 mg/kg) is used to differentiate chemicals of low concern with regard to acute oral toxicity from those of some level of concern.

Oral LD50 (mg/kg)	**EC classification**
> 2000	Not classified
> 200 < 2000	Harmful
> 25 < 200	Toxic
< 25	Very toxic

Acute oral toxicity tests had been conducted on 142 chemicals (two chemicals had not been tested: chemicals 4 and 107 are corrosive and react violently with water). A prediction of acute oral toxicity had been made for all of the 142 chemicals which had been tested, plus the two which had not.

There were 21 chemicals for which the toxicity indicated by the test data differed from that predicted (15%). Twenty of these were found to have greater acute toxicity than had been predicted, but for 14 of these there was overlap between the predicted and observed toxicity categories (see Table 10). One chemical had lower toxicity than had been predicted (number 124).

Twenty-one chemicals had been classified in the EC on the basis of their acute oral toxicity: 20 of them are included in Table 10 and were predicted to have lower toxicity than was observed, though for 14 there was an overlap between predicted and observed toxicity categories. However, 18 of the classified chemicals (12%) were predicted to be of "low" acute oral toxicity, and thus would apparently be considered of low concern with regard to this end-point (false negatives). The classified chemical which is not in Table 10 (number 281) was predicted, by analogy to data in the EPA confidential data base, to have "medium" acute toxicity and this was observed (LD50 = 850 mg/kg). Details of the oral toxicity predictions and test results are given for all chemicals in the project in Annex 12.

- **Conclusions**

Using arbitrary criteria to compare LD50 values with descriptions of predicted acute oral toxicity, there was a tendency to under-prediction of the level of toxicity for chemicals which, when tested, were shown to have significant acute oral toxicity. However, the majority of the chemicals were correctly predicted to be of low concern with regard to acute oral toxicity.

Predicted toxicity for 18 (12%) of the classified chemicals was "low", indicating that one-to-one substitution of predictive methods for testing would result in chemicals being missed which are, in fact, of some potential concern because of acute toxicity. It should be noted that two of these chemicals had been classified as "toxic if swallowed" (numbers 307 and 330).

In most cases there were overlaps between the predicted and the observed toxicity for the classified chemicals, and between the toxicity predicted for the classified chemicals and those not classified. Hence, the predictive methods could not readily be used to classify chemicals within the context of a scheme using pre-defined criteria.

Thus, this comparative study shows that the predictive methods can be used to identify correctly the >80% of a batch of 142 heterogeneous new chemicals which are of low acute toxicity. However, it is of concern that some 12% of this set of chemicals did have an appreciable level of acute oral toxicity which was not predicted (false negatives). Because of this outcome, if assessment of acute toxicity is an important consideration in a given evaluation scheme, the submission of test data will be needed to assess this end-point adequately. This is especially so in instances where a quantitative assessment of acute toxicity is needed.

TABLE 10: Differences between SAR evaluations and acute oral toxicity test data

Chemical	LD50	Label[1]	MPD tox[2]	SAR tox[2]
47	1800	R22	L-M	L
49	>200 <2000	R22	L-M	L
54	1984	R22	L-M	L
124	2300	-	L	M
156	1800M 1960F	R22	L-M	L
197	612	R22	M	L
219	1670	R22	L-M	L
241	585	R22	M	L
242	520	R22	M	L
300	1011	R22	L-M	L
307	88	R25	M-H	L
312	1774	R22	L-M	L
330	104	R25	M-H	L
340	1750	R22	L-M	L
360	>1000 <2000	R22	L-M	L
370	1400	R22	L-M	L
413	1200	R22	L-M	L
425	1650	R22	L-M	L
436	899	R22	M	L
441	450	R22	M-H	M
443	320	R22	M-H	M

[1] See Appendix 3 for list of "R phrases".
[2] See abbreviations above.

4.2.3.3. Irritation

Knowledge of the potential for skin, eye and respiratory irritation is important when evaluating safe handling practices for chemicals. Skin and eye irritation test data are used to predict the likelihood that exposure of human skin or eyes to a chemical will result in adverse effects (corrosion or irritation). An indication of the duration/reversibility of effects is also usually obtained.

There is not a test method for respiratory irritation in either the EC or the OECD set of accepted test methods for the toxicity testing of chemicals.

Prediction of irritation is not usually part of the routine evaluation of new chemicals in the US, but predictions were made for the purposes of this project, although EPA did not attempt to characterise the degree of irritation.

4.2.3.3.i Skin irritation

The criteria used for conducting the comparisons were to compute "primary irritation scores" from the test data, by taking the average of the total erythema and oedema scores for both the 24 and 72 hour readings:

Primary irritation index	**Irritant category**
2 or less	Mild/nil (low)
>2 to 5	Moderate
>6	Severe

The category "corrosive" was also used.

In addition, chemicals were also considered according to whether they had been classified as "Corrosive" or "Irritating to skin" in the EC.

Of the total of 144 chemicals in the project, there were 140 on which skin irritation tests had been conducted. All 144 chemicals had been considered when predicting the potential for skin irritation as a consequence of dermal exposure to the chemicals.

Correct predictions of low concern for skin irritation were made for 104 of the 122 chemicals (including the untested polymer, chemical number 267) for which the test results indicated little or no irritancy (83% of the 122 chemicals; 73% of the total number of chemicals in the project). There were 18 chemicals which were predicted to be irritating to skin, but were found not to be irritant in the test conducted, i.e. false positives.

The test results (or physico-chemical characteristics of three chemicals: numbers 4, 107 and 194) showed that 22 chemicals either were, or could be expected to be, at least moderate skin irritants. Twelve of these had been classified as "Corrosive" in the EC, and six as "Irritating to skin". The outcome of the comparisons for the classified chemicals is shown in Table 11.

Ten of these were identified by EPA as being skin irritants, while for the remaining eight, EPA did not identify a concern for skin irritation (false negatives). The group of false negatives included six corrosive chemicals.

TABLE 11: Comparison of predicted skin irritancy with that observed

Chemical	Label[1]	MPD result[2]	SAR result	Agreement[3]
4	R35	Corrosive[4]	Acute	Yes
49	R34	Corrosive	Irritant	Yes
53	R38	Mod - Sev	Irritant	Yes
107	R35	Corrosive[4]	Acute	Yes
118	R34	Corrosive	No comment	False -ve
182	R34	Corrosive	No comment	False -ve
192	R34	Corrosive	No comment	False -ve
194	R34	Corrosive[4]	No comment	False -ve
222	R38	Moderate	Irritant	Yes
235	R34	Corrosive	No comment	False -ve
237	R38	Low - Mod	Irritant	Yes
278	R38	Moderate	Irritant	Yes
370	R34	Corrosive	Irritant	Yes
373	R38	Moderate	No comment	False -ve
425	R34	Corrosive	Irritant	Yes
436	R34	Corrosive	No comment	False -ve
437	R38	Mod - Sev	Irritant	Yes
443	R34	Corrosive	No comment	False -ve

[1] See Appendix 3 for list of "R phrases".
[2] According to the criteria above, using primary irritation score.
[3] Predicted relative to test-derived level of skin irritancy.
[4] Chemicals not tested: EC assumed corrosivity based on physico-chemical properties.

The overall results for the comparison of SAR calls and MPD data for skin irritation are summarised in Table 12. In this Table, MPD positive includes the three chemicals considered corrosive in the EC on the basis of physico-chemical properties (chemicals 4, 107 and 194); and SAR negative includes the two chemicals for which the prediction was "uncertain". Details of the data on skin irritation for all chemicals are to be found in Annex 13.

TABLE 12: Overall results for skin irritation

	SAR Positive	**SAR Negative**
MPD Positive	14 (10%)	8 (5.5%)
MPD Negative	18 (12.5%)	104 (72%)

- **Conclusions - skin irritation**

Incorrect predictions were obtained for 18% of the chemicals: 12.5% were false positives and 5.5% were false negatives. The predictive methods used are not adequate for classification of chemicals using a system based on severity of response, and thus the test cannot be replaced on a one-to-one basis by the predictive approach when knowledge of the potential for skin irritation/ corrosion is needed.

4.2.3.3.ii Eye irritation

The criteria used to compare the test data with the SAR call for eye irritation could not be made on a severity index as the SAR evaluations did not usually include this index. From the test data summaries, a chemical was considered to produce significant eye irritation if redness, swelling or corneal opacity persisted beyond seven days or if effects were not reversible by 21 days or corrosion was reported. Eye testing was not conducted on chemicals with predictable corrosivity because of their physico-chemical characteristics or, for some chemicals (see Table 13), if corrosive effects had been recorded in a previously conducted skin test.

Classification according to the EC system (for which the criteria are a combination of scores and duration of effects), on the basis of the results of the eye irritation studies, was obviously also considered as indicating that the classified chemicals were eye irritants.

Of the total of 144 chemicals in the project, there were 140 on which eye irritation tests had been conducted, three were predicted to be corrosive and one (number 267) could not be tested for technical reasons. All 144 chemicals had been considered when predicting the potential for eye irritation as a consequence of ocular exposure to the chemicals.

On the basis of the test results, 105 chemicals were considered to be of low concern for eye irritation, as was chemical 267, which had not been tested. Correct predictions of low concern were made for 87 of these (83% of the "negative" chemicals, 60% of the total set of chemicals). The other 18 were predicted by the EPA to be irritant, i.e. they were false positives.

The 38 remaining chemicals were either corrosive (12 chemicals), or irritant according to the criteria given above. The outcome of the comparisons between the predicted and test results for the classified chemicals is given in Table 13; the detailed analysis for all chemicals in the project is given in Annex 13.

TABLE 13: Comparison of predicted eye irritancy with that observed

Chemical	Label[1]	MPD result[2]	SAR result	Agreement[3]
4	R35	Corrosive[4]	Acute	Yes
47	R41	Severe	Uncertain	False -ve
49	R34	Corrosive	Irritant	Low
87	R41	Severe	No comment	False -ve
107	R35	Corrosive[4]	Acute	Yes
118	R34	Corrosive[4]	No comment	False -ve
124	R36	Irritant	Irritant	Yes
151	R36	Irritant	Irritant	Yes
170	R41	Severe	Irritant	Yes
182	R34	Corrosive	Irritant	Low
192	R34	Corrosive	No comment	False -ve
194	R34	Corrosive[4]	No comment	False -ve
197	R41	Severe	No comment	False -ve
222	R36	Irritant	Irritant	Yes
235	R34	Corrosive[4]	No comment	False -ve
237	R36	Irritant	Irritant	Yes
256	R36	Irritant	Irritant	Yes

TABLE 13 - continued

Chemical	Label[1]	MPD result[2]	SAR result	Agreement[3]
263	R36	Irritant	Irritant	Yes
270	R36	Irritant	No comment	False -ve
281	R36	Irritant	Irritant	Yes
370	R34	Corrosive[4]	Irritant	Low
425	R34	Corrosive[4]	Irritant	Low
436	R34	Corrosive[4]	No comment	False -ve
441	R41	Severe	Irritant	Yes
442	R41	Severe	Irritant	Yes
443	R34	Corrosive[4]	No comment	False -ve

[1] See Appendix 3 for list of "R phrases".

[2] According to the criteria given in the text.

[3] Predicted relative to test-derived result.

[4] Chemicals not tested: corrosivity assumed based on physico-chemical properties or results of skin irritation study.

From the comparisons given in Table 13, it can be seen that, for the 26 classified chemicals, 16 were correctly predicted to be eye irritants and ten were incorrectly assessed (false negatives).

The overall results for the comparison of the SAR calls and the MPD test results are summarised in Table 14.

TABLE 14: Overall results for eye irritation

	SAR Positive	**SAR Negative**
MPD Positive	26 (18%)	13 (9%)
MPD Negative	18 (13%)	87 (60%)

- **Conclusions - eye irritation**

Incorrect predictions were made for 22% of the chemicals (9% were false negatives, 13% false positives). As with skin irritation, predictive methods are not adequate for classification of chemicals with regard to severity of the response and thus cannot replace test results on a one-to-one basis.

4.2.3.3.iii Respiratory irritation

New chemicals are not tested for respiratory irritation in the EC, but the potential for respiratory respiration had been considered by the EPA predictors.

Predictions of potential respiratory or mucous membrane irritation had been made for nine (6%) of the chemicals in this study.

- **General conclusions**

The majority of this group of new chemicals were of low concern for skin (85%) and eye (74%) irritancy. Thus, the extent to which an assessment can be made of the power of the predictive methods to discriminate between chemicals on the basis of their skin or eye irritation potential is limited.

The majority (>80%) of the low concern chemicals were predicted correctly and 18% were overpredicted for either or both skin and eye irritancy. The latter observation means that, for these substances, the risk assessment would err on the side of caution but would lead to "overlabelling" if the predictive methods replaced the tests.

The incidence of false negatives and the limitations in assessing severity of response are of some concern, and indicate that replacement of testing with prediction cannot yet be recommended with confidence.

Respiratory irritation is an important end-point which is not investigated in the MPD. It would be prudent to take note of chemicals predicted to be respiratory irritants.

4.2.3.4. Sensitisation

Knowledge of the sensitising potential of chemicals is important when evaluating safe handling practices.

Prediction of sensitisation is not usually part of the routine evaluation of a new chemical in the US, but it was considered for this project.

In the EC, chemicals are tested for their skin sensitising potential. There is not an internationally recognised test method for respiratory sensitisation. Classification of notified new chemicals as skin sensitisers in the EC is based

on the proportion of animals showing a positive response in a particular test. In the EC, chemicals may be classified as respiratory sensitisers if they show close structural similarity to known chemical respiratory sensitisers.

Skin sensitisation tests, mostly maximisation tests, were conducted on 137 of the chemicals in the project. Twenty-eight chemicals were classified as skin sensitisers (including one of those which had not been tested). A further 18 induced some positive responses, but the number of animals responding was below the threshold for classification in the EC.

Seventeen chemicals were predicted to be sensitisers; four of these were predicted to be respiratory sensitisers and one was predicted to be a photosensitiser. Two were predicted not to be sensitisers. For most of the chemicals there was no comment on skin sensitisation - this is equivalent to considering the chemical of low concern/negative for this end-point.

For 108 chemicals (75% of the whole set in the project), both the test results and the predictions indicated low concern for skin sensitisation.

The results of the comparisons of the test data and the predictions are given in Table 15 for the 28 chemicals classified as skin sensitisers in the EC.

TABLE 15: Comparison of results for chemicals classified as skin sensitisers

Chemical	SAR	Result and comments
47	–	False negative
76	+	Agree
96	–	False negative
118	–	False negative
133	–	False negative
173	+	Agree
194	–	False negative NB: chemical not tested
196	+	Agree
197	–	False negative
200	–	False negative
222	+	Agree Chemical also classified and predicted as a respiratory sensitiser
235	–	False negative

TABLE 15 - continued

Chemical	SAR	Result and comments
256	+	Agree
271	–	False negative
275	–	False negative
330	+	Agree
341	+	Agree Chemical also classified and predicted as a respiratory sensitiser
344	–	False negative
348	–	False negative
376	+	Agree
393	–	False negative
401	–	False negative
413	–	False negative
416	–	False negative
437	–	False negative
442	–	False negative
444	–	False negative

Five other chemicals were predicted by the US to be skin sensitisers: one did not have adequate test data (240); two did induce some positive responses in the tests conducted (253, 312); and two were apparently false positive predictions (340, 364).

Two other chemicals were predicted to be potential respiratory sensitisers (69, 101).

For the set of comparable skin sensitisation data (140 chemicals), the comparisons in Table 16 can be made.

TABLE 16: Overall results for skin sensitisation

	SAR Positive	SAR Negative
MPD Positive	9 (6.5%)	19 (13.5%)
MPD Negative	4*(3%)	108 (77%)

* includes two substances for which positive responses, below the threshold for classification, were observed in the tests

- **Conclusions**

The incidence of false negatives precludes use of the predictive methods to replace the tests on a one-to-one basis or to classify chemicals for their skin sensitisation potential. However, the concurrence of positive predictions with positive test results needs to be further assessed with a larger set of chemicals, as confidence in the ability to predict positives could perhaps replace testing of chemicals predicted to be skin sensitisers.

For respiratory sensitisation, reliance is currently placed on predictive methods, based on structure, to classify new chemicals in the EC, and the unclassified substances predicted, in this project, to be potential respiratory sensitisers should be re-evaluated in the EC with regard to classification.

It is not possible to comment on the single prediction of potential photosensitisation.

4.2.3.5. Repeated dose toxicity

Repeated dose toxicity covers the adverse effects which may arise in humans exposed to a chemical at frequent, regular intervals over a prolonged period of time, for example at their daily work. To facilitate evaluation of safe handling practices for chemicals, it is important to have knowledge of the potential systemic effects which may occur on repeated exposure.

In the EC, general effects on the whole animal and effects on tissues, organs and/or systems are investigated. Special effects (e.g. neurotoxicity, reproductive toxicity, carcinogenicity) are investigated in specific tests, but indications of potential reproductive toxicity, neurotoxicity or immunotoxicity may be detected in repeated dose toxicity studies.

For most of the chemicals in this project, only 28-day, and/or occasionally 90-day, study results were available. In the EC study summaries used for this project, dose levels used, a description of toxic signs, including clinical chemistry and haematology, gross and microscopic changes in a selected set of tissues/organs, and NOEL, NOAEL, LOEL and LOAEL (no/low observed effect/adverse effect level) values are usually included or can be deduced. In general, only effects of biological significance are included, and species-specific effects (e.g. peroxisome proliferation and, in the more recent summaries, male rat-

specific light hydrocarbon nephropathy) are not. Chemicals are classified for repeated dose toxicity in the EC on the basis of adverse effects (of biological/human significance) occurring at or below dose levels specified according to the route of exposure and the duration of the study.

Predictions of repeated dose toxicity are particularly important in the EPA evaluation process, with identification of potentially toxic chemicals as the goal. Efforts are also made to assess potential target tissue/organ/system.

Test data were not available for seven chemicals (three corrosive chemicals, two polymers, one organoclay, and one chemical not tested in the light of test data available for another notified chemical, of very similar structure). Two chemicals had been tested in 28-day inhalation studies and eight in dermal 28-day studies. For one of the latter group, a 90-day study had also been conducted. The remaining 127 chemicals had been tested using 28-day oral toxicity studies and three also had results available from 90-day studies.

Eight chemicals had been classified in the EC on the basis of their repeated dose toxicity.

The comparison of repeated dose toxicity test results with predicted toxicity was the most difficult to do, as interpretation of observed effects in terms of severity and significance is a matter of professional judgement. The factors considered in the evaluation were the perceived seriousness of the toxic effect, the number of organ-specific parameters affected, with microscopic pathology given the heaviest weight, multiplicity of target organs, the toxic effect(s) at the LOAEL, the numerical value of the NOAEL, dose-related effects, and the spacing of the dose levels used.

The systemic toxicity data from the test results were scored as high, moderate or low using the following general criteria (sometimes modified according to professional judgement):

Concern level	Criteria
Low (L)	No systemic toxicity (NOAEL 1 g/kg/day or more); only minor clinical signs of toxicity; liver and/or kidney weight increase or clinical chemistry changes; LOAEL >500 mg/kg/day.
Moderate (M)	Organ pathology (gross and/or microscopic) with LOAEL 500 mg/kg/day or less; clinical chemistry changes and organ weight changes at <500 mg/kg/day; NOAEL <100 mg/kg/day.
High (H)	Death, organ pathology (microscopic) at LOAEL 100 mg/kg/day or less; multiple organ toxicity; NOAEL <10 mg/kg/day.

"Split-levels" (L-M; M-H) were adjustments for specific multiple organ toxicity, borderline effect levels, and professional judgement.

The outcome of the comparisons of repeated dose toxicity on the basis of concern level is summarised in Table 17.

TABLE 17: Matrix analysis of systemic toxicity concern levels

SAR	L	L-M	M	M-H	H
MPD					
L	62	10	5	0	0
L-M	23	11	2	0	0
M	11	1	5	1	1
M-H	3	1	2	3*	0
H	1	0	0	0	1*

* One chemical in each of these groups was corrosive and predicted to have acute effects

One chemical (337) is not included in the matrix. It was M-H according to the test results, but there was no prediction of repeated toxicity.

Sixty-two chemicals (43%) were considered of low concern, both following testing and by the predictive methods.

Twenty chemicals (14%) with greater than "low" concern were predicted to have the same level of concern as was deduced from the test data using the criteria given above. This group included the two corrosive chemicals which were predicted to have "acute" effects (numbers 4 and 107) and chemical 292, for which data were available from the product literature.

The concern level was underpredicted for 42 chemicals (29%), though for 27 chemicals there were overlapping concern levels from the test and predicted results; and 23 of these predicted to be of low concern were only low-moderate from test results. For the other 15, the concern level predicted was at least one whole level lower than that deduced from the test data. Six of this sub-set of 15 were chemicals classified in the EC on the basis of repeated dose toxicity.

Toxicity concern was apparently overpredicted for 19 chemicals. However, the extent of repeated dose toxicity testing of these chemicals was limited to 28-day studies (18 oral studies, one dermal). It will be of interest, if/when

90-day, or longer, study data become available, to re-compare the predicted toxicity with that found on testing.

Overall, the correct level of concern (according to the criteria given above) was predicted for 57% of the chemicals, but was underpredicted for 29%. Toxicity was apparently overpredicted for 13% of the chemicals.

Details of the organ toxicity predictions and test results are given for all the chemicals in the project in Annex 14.

- **Conclusions**

Just over half (57%) of this group of 143 heterogeneous chemicals were correctly predicted to be either of low concern (43% of the total) or to have the same level of concern (14% of total) in relation to repeated dose systemic toxicity. The concern level was apparently overpredicted for a further 13%, but if/when longer-term studies are conducted the predicted effects may be induced.

Underprediction of the level of concern on the basis of repeated dose toxicity was noted for 42 chemicals (29% of the total), although for 23 of these the test data indicated only low-moderate concern and EPA predicted low concern. For 15 chemicals, there was at least one whole "level of concern" difference, and six of the eight classified chemicals were in this group.

On the basis of these comparisons, although for 74% of the chemicals in this study correct or near-correct predictions of concern level were made, it is not considered possible to consider the predictive methods as an adequate substitute for conducting repeated dose toxicity testing of a random/heterogeneous group of chemicals because of underprediction of toxicity. As classification of a chemical as dangerous following repeated exposure depends not only on the effects seen, but also on the doses at which they occur, the predictive methods for repeated dose toxicity would not provide a firm basis for classification.

4.2.3.6 Mutagenicity

Chemicals which increase the incidence of mutations in the cells of exposed humans may thereby increase the incidence of cancer (from mutations in somatic cells) or genetic defects in the offspring (from mutations in germ cells). It is generally thought prudent to assume that there is no threshold exposure level, below which exposure would give rise to only low concern, for chemical mutagens. Thus, chemicals identified as mutagens are subject to stringent controls so that human exposure is minimised.

Because of the serious and irreversible effects which may occur in humans exposed to chemical mutagens, testing for mutagenicity usually employs a number of tests, in vitro and in vivo, which are conducted either as a battery or (as in the EC) in series. In the EC, all notified chemicals must, if it is technically possible, be tested in a bacteriological test for gene mutation and in a test in mammalian cells for chromosomal effects at the "base set" level of

supply. The latter test may be either an *in vitro* test or a test conducted *in vivo*. Maximised conditions are used, though short of conditions likely to cause artefactual positive results; and *in vitro* tests are conducted both with and without exogenous metabolic activation. Further testing is conducted to investigate in more detail positive test results, as necessary, and/or as supply tonnages reach the trigger levels. Classification of chemicals on the basis of mutagenicity is done according to criteria defined in Annex VI to the dangerous substances Directive. Chemicals are not usually classified unless there is evidence of mutagenicity from tests conducted *in vivo*, so positive *in vitro* test data will trigger the need for testing *in vivo*.

The EPA predictions for mutagenicity, based on e.g. chemical class, analogue data, likely metabolites, alkylating potential, represent an overall for mutagenic potential. EPA also considers available data concerning mutagenicity test systems and their sensitivity towards different classes of chemicals. Thus, the criteria for comparing the predicted with the test results involved more than a simple comparison of EPA predictions with the test data. In addition, the test results for a few (six, 4%) chemicals with borderline responses were not always interpreted in the same way by the EPA and EC experts.

Tests had not been conducted on five of the 144 chemicals in the project - three for technical reasons (chemicals 4 and 107 were corrosive and chemical 267 was an insoluble polymer) and for the other two (chemicals 194 and 445) data from analogues were considered acceptable. Predictions had been made for the first three (all were "low concern" for mutagenicity) but there were no test data to compare them with. Thus, there were 141 data pairs for comparison. All of the 139 chemicals tested had Ames test data and all had at least a result from one other study. The *in vivo* micronucleus test occurred most frequently as the second study, and the *in vitro* chromosome aberration test was the next most common. Tests in *E. coli* (always alongside the Ames test when the *E. coli* test had been conducted), *in vivo* chromosome aberration, nuclear anomaly and sister chromatid exchange (SCE) tests, and *in vitro* mammalian cell gene mutation assays, unscheduled DNA synthesis, and SCE tests also occurred in this set of tests. Interestingly, for no chemical was there more than one positive test.

One hundred twenty chemicals gave negative results in both a bacteriological (Ames) test and a non-bacteriological test. Some of these chemicals also had negative results from gene mutation tests in *E. coli* and/or from other non-bacteriological tests. Two chemicals were assumed by analogy to structurally similar chemicals to be negative and were not tested. Thus, following testing, 122 chemicals (85% of the chemicals in the project) were considered negative. SAR predictions of low concern for mutagenicity were made for 107 chemicals in this group (88% of the MPD "negatives").

Depending on how the analysis is done, "false positive" predictions were made for 14 (10% of total) or two (1.4%) chemicals. A direct reading of the MPD results would lead one to conclude that there were 14 false positive predictions. However, EPA considers that positive results would be produced if tests were performed using assay systems other than those used already to test the affected chemicals. The EPA conclusions are based on the existence of data on analogues (chemical or mechanistic) indicating positive results in certain

test systems. It will be of interest (and potential importance) to see whether the predictions of positive mutagenicity are fulfilled if further test data become available.

Six chemicals (4% of the total) with positive test data were predicted "low" (false negatives) because of absence of the known positive data in analogues.

The test results (including, where appropriate, an indication of weak positive results), EPA predictions, and results of comparison are given in Annex 15 for all of the chemicals in the project.

- **Conclusions**

A high proportion of the chemicals in this project were negative for mutagenicity, and a high proportion of these were correctly identified by the EPA.

Although the number of test-positive chemicals was small, it is also of concern that six of them were called low. The observation that 123 of 142 data pairs (87%) were apparently correctly predicted thus has to be seen in the light of the above comment. For this reason it would not be prudent at this time to replace mutagenicity testing of new chemicals in the EC with the predictive methods used in the US for PMN chemicals.

As the EC classification system for mutagenicity, as applied to notified new chemicals, depends essentially on testing in vivo to investigate whether effects observed in vitro are expressed in vivo, the predictive methods used here, which do not make this distinction, could not be used for classification in the EC.

4.2.3.7. Other effects

A number of effects were considered using the predictive methods which had not yet been investigated in the EC testing programme for the chemicals in this project, i.e. reproductive and developmental toxicity, neurotoxicity and oncogenicity. For some chemicals, indications of some of these effects (e.g. clinical signs of neurotoxicity; changes affecting the reproductive organs) may be reported for the acute or repeated dose tests. Such reports were made for some chemicals in this project: five chemicals had significant indications of potential reproductive toxicity (76, 151, 186, 200 and 292), and reproductive toxicity was predicted for chemicals 200 and 292 but not for the others (developmental toxicity was predicted for chemical 76). Signs of neurotoxicity were seen with six chemicals (54, 268, 340, 342, 431 and 434) and neurotoxicity was predicted for two of these (54 and 340).

Adverse effects on reproduction and/or development were predicted by EPA for 51 chemicals (35%); 27 chemicals were predicted to be neurotoxic (19%) and 33 (23%) to be oncogenic. This is of particular concern as these potential effects are not specifically investigated in the initial testing of new chemicals in the EC.

The health concerns for which the MPD data set does not provide data were analysed for a number of chemicals for which such concerns were expressed, and the frequency of occurrence. Of the 144 chemicals, 66 (44%) had concerns that addressed health effects outside the scope of the MPD data set. The breakdown by effect and frequency of occurrence is presented in Table 18.

TABLE 18: Health concerns not addressed by the MPD data set

Concern	Number of chemicals	% of total chemicals
Oncogenicity	33	23
Developmental toxicity	46	32
Reproductive toxicity	13	9
Neurotoxicity	2	15
Immunotoxicity	2	-
Photosensitisation	1	-
Lung	1	-
Respiratory sensitisation	1	-

This table indicates that potential adverse effects beyond those in the MPD were identified for a substantial number of the chemicals, which implies that hazards and possibly risks may be underestimated if these effects are not considered. There may be a need for early focused testing in at least some of these cases.

5. OVERALL CONCLUSIONS

5.1. Conclusions: US perspective

5.1.1. Introduction/overview

The purposes of the study were to compare the results obtained in assessing a series of European Community (EC) new chemicals using two methods - the US SAR-based (Structure Activity Relationships) approach and the EC's testing-based approach using the Minimum Pre-marketing set of Data (MPD) - and to estimate

the extent to which the US hazard[1] conclusions on new chemicals might change given a "base set" of test data. The study would also provide insights into the strengths and weaknesses of specific SAR approaches and allow EPA to judge how well SAR works in other areas of application, e.g. priority setting for existing chemicals and testing.

The results of the study, as expected, were quite useful in judging many of the strengths and weaknesses of the US approach, as well as determining the utility of MPD-type data in improving US assessment capabilities. It must be pointed out, however, that as useful as the study was, there are some limitations that must be considered in the overall evaluation of the exercise. These limitations include: the small data set available, the end-points used for comparison were limited to the tests included in the MPD data set, different approaches to ascertaining certain parameters, and indirect measurement in some MPD data sets of one or more physico-chemical properties (i.e. extrapolation) which may or may not give a "true" result. These limitations are discussed in more detail in the following sections. However, taking into account these limitations, the MPD/SAR exercise served to confirm that the SAR approach to screening new chemicals[2] is useful and effective in identifying chemicals that may be toxic and in need of further scrutiny for US regulatory purposes. However, the SAR approach appears to have limitations in predicting physico-chemical properties under some circumstances, and in predicting the exact type and level of toxicity of the chemical, especially with regard to general systemic (health) effects.

5.1.2. Results

The end-points that were assessed have been divided into four categories (physico-chemical properties, biodegradability, health effects, and ecotoxicity) for discussion purposes and appear below.

5.1.2.1. Physico-chemical properties

The physico-chemical properties routinely predicted by the SAT are: log P_{OW}, boiling point/melting point, water solubility, vapour pressure, Henry's Law constant, as well as the soil sorption coefficient and the bioconcentration

---

1 This study examined hazard (or toxicity) predictions and did not exmaine exposure or risk issues, other than to consider predictions of environmental fate.

2 In the US scheme, PMN chemicals are initially reviewed by EPA's Structure Activity Team (SAT), which "screens" the chemicals to assess their fate and effects. For cases which are determined to present potentially significant risk concerns, a more detailed assessment is prepared. The present study compared the results of SAT (screening) assessments with the results of the MPD testing.

factor. The MPD data set contains either measured or calculated values for log P_{OW}, boiling point/melting point, water solubility, vapour pressure, and Henry's Law constant. Of these properties, there were sufficient data pairs for meaningful comparison of log P_{OW}, vapour pressure, and water solubility.

For log P_{OW} comparisons of the 144 chemicals, there were 35 for which either SAR and/or MPD data were missing; additionally, a number of the MPD values were calculated or estimated, which allowed for a comparison of estimation methods but did not provide an opportunity to compare the US estimated values with actual measured values. Applying a US/EC agreed upon standard of ± 1 order of magnitude for "good agreement", the overall agreement between the US estimates and the EC measured values was around 60%. In analysing the 40% which were in disagreement, it became apparent that the estimation techniques for log P_{OW} were of limited value with certain classes of chemicals (e.g. classes where all the molecular fragment constants have not been measured, ionic compounds, organometallics, inorganics, and classes/compounds which are readily hydrolysed). For those classes where the estimation techniques are appropriate, the agreement was acceptable and predictive approaches were judged to provide a useful alternative to experimentally determining log P_{OW}. For chemicals where models are not appropriate, experimental determination of log P_{OW} is the preferred method.

Vapour pressure comparisons presented a number of analytical problems. In the US PMN programme, vapour pressures below 10^{-3} torr are routinely considered "negligible" and not of concern for either worker/consumer exposure or volatilization from the pure state. Thus estimated values of less than 10^{-3} torr are in general not determined. The EC, however, considers vapour pressures relevant to 10^{-6} torr and thus requires values to be provided. In order to adjust for the differing requirements, a set of rules was generated and agreed to by the US and EC. Additionally, the vapour pressure for the EC chemicals was measured on the substance "as marketed" in the EC (i.e. a mixture or formulation, in many cases), whereas the US estimate was made for the pure chemical. The results of the analysis showed that 63% of the US estimated values were in agreement ($\pm$1 log unit) with the measured EC values. Of the 37% (42 chemicals) of the comparisons that were in disagreement, the disagreement for 30 of the chemicals can be accounted for by the following reasons:

- the "measured" vapour pressure value was extrapolated from a value measured at a higher temperature, which tends to overestimate the true actual atmospheric vapour pressure;

- the pre-market substance tested contained a volatile solvent and/or impurities;

- the substance decomposed during the measurement procedure;

- the measured value reflected water which was being driven off by the measurement procedure;

- vapour pressure was the lowest value measured and thus represents the upper limit rather than an actual value.

The best agreement was observed between the PCNOMO estimates and the measured values. Overall, however, vapour pressure estimates were judged to have marginal acceptability since the values were both over- and underestimated by the US. As was stated previously, vapour pressure contributes to the exposure portion of the risk assessment for new chemicals and over/underestimation can result in an over/underestimation of the exposure associated with a chemical and thus contribute to an over/underestimation of the risks. Thus incorrectly estimating vapour pressure may unnecessarily put the worker/consumer at risk or burden the manufacturer with unnecessary constraints, depending upon the direction of the estimation error. Vapour pressure is a relatively inexpensive parameter to measure and, as such, it may be more cost effective and less risky/burdensome to obtain experimental data to confirm the estimated value in cases where vapour pressure is an important contributor to the risk projection.

Water solubility comparisons presented some similar problems to the vapour pressure comparisons. In the US PMN programme water solubilities below 1 mg/l are not routinely estimated, because reasonably accurate estimation of extremely low water solubilities is difficult. On the other hand, the EC data measure water solubilities of < 0.1 mg/l in many cases. In addition, the EC measured value is not necessarily done on the pure chemical but many times on the substance "as marketed", whereas the US estimated value is for the pure chemical. The results of the analysis showed that 68% of the US estimated values were in agreement (± 1 log unit) with the measured EC values. Of the 32% of the comparisons (43 chemicals) that were in disagreement, the disagreement for 26 of the chemicals can be accounted for by the following reasons:

- the "measured" value was not actually measured, but reported as a lower limit of detection or the lowest value measured;

- the pre-market substance tested contained a solvent and/or impurities which complicated interpretation of water solubility values;

- the measured value was measured spectrophotometrically;

- the substance decomposed or reacted with the water during the measurement procedure.

Overall, the water solubility estimates were judged to have marginal acceptability since the values were both over- and underestimated by the US. Water solubility contributes to the hazard and exposure portions of the risk assessment for new chemicals, and over/underestimation can result in an over/underestimation of the hazard/exposure associated with a chemical and thus contribute to an over/underestimation of the risks. Thus incorrectly estimating water solubility may put the worker/consumer unnecessarily at risk or burden the manufacturer with unnecessary constraints, depending upon the direction of the estimation error. Water solubility is a relatively inexpensive parameter to measure and, as such, it may be more cost effective and less risky/burdensome to obtain experimental data to confirm the estimated value in cases where the water solubility is an important contributor to the risk projection.

5.1.2.2. Biodegradability

Comparison of the US and EC biodegradability data was difficult due to the fundamental incompatibility of the evaluation approaches used for assessing biodegradability in the US versus the EC. The US estimates biodegradability in terms of "days, weeks, or months", which refer to the approximate amount of time (not half-life) required for complete primary and ultimate biodegradation of the chemical in aquatic environments. In contrast, the EC requires a laboratory test which evaluates the "ready" biodegradability of chemicals. Thus, while chemicals that degrade easily in the EC testing scheme would most likely be easily degraded in the environment, it is not necessarily true that chemicals not degraded in the EC tests would not be degraded under environmental conditions, which is what the US approach attempts to predict. For the purposes of this exercise, chemicals that did not pass the EC test, i.e. did not degrade under conditions of the test, were considered to correspond to the descriptors "weeks or longer" and ones that passed, i.e. degraded, were considered to correspond to the descriptors "days" and "days to weeks" in the US scheme. Using these criteria, there was a 93% agreement between the US predictions and the EC test results.

The US scheme for predicting biodegradability aims for a realistic assessment of the ultimate fate of a chemical under environmental conditions. In contrast, the EC testing scheme is designed to determine ready biodegradability under precise laboratory conditions. While the EC scheme may provide more quantitative results, it can be argued that the modelling by the US represents a more realistic estimate albeit qualitative. Biodegradability testing under conditions that duplicate actual environmental conditions may not be feasible either from a scientific or a cost perspective. Although the MPD/SAR analysis has significant uncertainty due to the basic differences between the two approaches, the present US modelling scheme appears to be reasonably effective in predicting biodegradability that is consistent with experimentally derived results. However, given the uncertainty in the analysis, in the instances for which fate is a major contributor to the overall risk projection, or for classes of chemicals where there is insufficient data for modelling, it is advisable to confirm the prediction with appropriate testing.

5.1.2.3. Health effects

Although the EC requires that a base set of toxicity data be submitted with all their new chemicals, the data are used principally to classify and label the chemicals according to a set scheme. This is in contrast to the US practice, where hazard information is evaluated and integrated with potential exposure to ascertain risk. In addition, under the EC scheme additional testing on the new chemical must be provided as production grows (known as the "step system"). In the US, on the other hand, if controls or testing requirements are not implemented before manufacture commences, the new chemical authorities under TSCA no longer apply. Thus any controls or testing must be done under TSCA’s existing chemical provisions, which carry a much heavier burden for the government. Thus the emphasis on end-points tends to differ under the two schemes, with more weight given to acute effects (i.e. lethal dose, eye and skin irritation and sensitisation) in the EC scheme and more attention paid to

long-term or sub-chronic effects in the US, with relatively little emphasis given to acute effects. Nonetheless, because the US does not routinely predict acute effects for new chemicals (end-points which are well represented in the MPD), but focuses its efforts on predicting long-term effects (many of which are not covered by the MPD), the study was somewhat limited in its ability to compare health hazard predictions with MPD results. These points will be discussed in more detail below.

For the analysis of the comparison between predicted effects and test data, each end-point was compared and analysed separately. An overall analysis was also done which attempted to compare the US and EC "bottom line" health assessments for each chemical regardless of effect.

For acute effects the US predictions corresponded to the EC results between 78-88% of the time. Eye irritation had the lowest correspondence between predicted and measured value, and dermal irritation had the highest. Nonetheless, irritation and sensitisation are not judged to be particularly amenable to SAR analysis except for general classes; furthermore the tests for these effects are, in general, inexpensive. It seems reasonable that if understanding of these effects is an important consideration under a given scheme, then the submission of data is preferable to prediction. For acute toxicity, the predictive approach worked reasonably well and is judged to be acceptable for screening purposes (i.e. qualitative assessment).

Overall, for mutagenicity the US predictions corresponded to the EC results 94% of the time. Out of 144 data sets available for mutagenicity, 21 initially were in disagreement between the US prediction and the EC results. Further analysis of the 21 revealed that three of the disagreements were due to the use of inappropriate analogues by the US, two were due to lack of positive analogue data and weak or marginal positive responses reported in the EC data, and four were due to the absence of analogue mutagenicity data upon which to base SAR decisions. The remaining twelve may be MPD "false negatives" caused by testing in assay systems known to be insensitive to specific classes of chemicals. These twelve were called positive by the US due to analogue data reporting positive results in assay systems known to be sensitive to chemicals in the specific classes. Six chemicals with positive results were predicted "low" because of the lack of data on analogues and an absence of structural features suggestive of mutagenic activity. These false negatives, while small in number, were of concern and suggest that testing for this end-point should be considered in cases for which data on analogues are unavailable and exposures are projected to be at moderate or higher levels.

For long-term and sub-chronic effects, the US routinely predicts systemic toxicity as well as developmental and reproductive toxicity, neurotoxicity, and oncogenicity. The EC "base set" data includes only a 28-day repeat-dose study which does not address the latter concerns. In order to analyse the results of the study, systemic toxicity was assessed and then the concerns that fall outside of the 28-day study were folded into the analysis to achieve an overall analysis of the US predictions.

Systemic toxicity, exclusive of developmental and reproductive toxicity, neurotoxicity, and oncogenicity, was analysed by comparing the US predictions (concern levels)[3] for systemic toxicity only with the MPD data; both were also scored according to severity of effect, which was predicted/observed. The results of this analysis showed that for 57% of the 138[4] chemicals assessed the scores were identical and for 43% the scores disagreed. Further analysis revealed that the US tends to underpredict systemic toxicity (effects and/or severity) as observed in the MPD's 28-day study (which, in itself, is judged to provide a reasonable approximation of sub-chronic toxicity for most chemicals). For 27% of the chemicals, the US predicted a "low" concern whereas the MPD 28-day study supported a "low-moderate" or greater concern level. For 3% of the cases, the US predicted some concern (i.e. low-moderate or greater) while the MPD results supported a higher level of concern. For 14% of the cases, results of MPD testing supported a lower level of concern than was predicted by the US; in 11% of the cases the MPD supported a "low" concern whereas the US predicted low-moderate or greater concern. Note, however, that while the comparison study suggests a clear tendency to underestimate rather than overestimate the potential for systemic toxicity, the magnitude of the difference between the US and EC calls was relatively small. For example, in 23 of the 41 cases for which the US underpredicted the concern level, the MPD supported a "low-moderate" concern whereas the SAR-based call was for "low" concern while in three additional cases where the US predicted "low-moderate" or greater concern, the MPD supported a one-step increase in the concern level (e.g. "low-moderate" concern to "moderate" concern). This, nonetheless, is interpreted as indicating that the US needs to exercise caution in interpreting systemic toxicity predictions and should consider requiring a repeat dose test in cases where the projected exposures are at moderate or higher levels.

When concerns not addressed by the MPD (i.e. developmental and reproductive toxicity, neurotoxicity, and oncogenicity) were folded into the analysis, the US level of concern scores were identical to the MPD scores 78% of the time. The chemicals for which non-MPD health concerns were identified by the US were analysed to determined the nature and frequency of their occurrence. Of the 143 chemicals, 66 had concerns identified by the US that suggested one or more health effects beyond the scope of the MPD. The breakdown by predicted effect revealed that 32% of the chemicals had developmental toxicity concerns, 23% had oncogenicity concerns, 15% had neurotoxicity concerns, and 9% had reproductive toxicity concerns.

---

[3] The concern levels employed by the US in assessing new chemicals (and used in this study) are as follows: low, low-moderate, moderate, moderate-high, and high.

[4] Five of the chemicals were not tested in a 28-day study due to physico-chemical properties (e.g. pyrophoric) that rendered them unsuitable for testing.

The large number of chemicals that were predicted to have effects not addressed by the MPD raises the issue of possible improvements to the MPD. Although it may not be feasible to address oncogenicity directly, the developmental, reproductive and neurotoxicity concerns could conceivably be screened by use of a modified testing scheme. Thus, in designing a "base set" of testing, it may be appropriate, given the relative frequency with which these potential effects were identified in this study, to include testing to screen for these effects.

When overall level of concern scores for health effects are considered (i.e. a bottom-line assessment considering all effect areas), the trend towards underprediction rather than overprediction (which was observed in the analysis of systemic toxicity outcomes) is still apparent. If the overall level of concern scores are analysed similarly to the systemic toxicity scores, 11% of the chemicals were identified by the US as being of low concern whereas the MPD supported a low-moderate or greater concern based on the MPD data, while an additional 8% were identified as being of low-moderate or greater concern by the US while the MPD supported a higher level of concern. In contrast, for only 4% of the cases did the MPD support an overall lower level of concern than had been projected by EPA. However, the scores for overall level of concern for health effects indicate a higher concordance between the US and EC than scores that were seen in the systemic effects analysis, which is due in part to the inclusion of concerns expressed for other MPD end-points (e.g. mutagenicity) as well as effect end-points outside the scope of the MPD "base set".

5.1.2.4. Ecotoxicity

When the EPA-predicted fish and daphnid acute toxicity levels of concern were compared to the levels of concern assigned to the MPD-measured acute values, the agreement (± 1 order of magnitude) for fish acute toxicity was 82% (107 chemicals) and for daphnid acute toxicity 71% agreed (90 chemicals). The number of chemicals in the EC data sets having fish and daphnid toxicity differed from each other, with 139 chemicals tested for fish toxicity and 137 chemicals tested for daphnid toxicity. For fish toxicity, the US tended to overpredict toxicity rather than underpredict (11% versus 7%); for 7% of the chemicals, the US predicted a "moderate" level of concern[5] whereas the MPD data

---

5 For aquatic toxicity the concern levels are expressed as "high," "moderate," and "low" according to the following criteria:

- Acute toxicity values <1 mg/l and/or chronic toxicity values <0.1 mg/l receive a **high concern**.
- Acute toxicity values from 1 to 100 mg/l and/or chronic toxicity values from 0.1 to 1 mg/l receive a **moderate concern**.
- Acute toxicity values >100 mg/l, chronic toxicity values >1mg/l, and cases where the solubility is severely limited and no effects are anticipated at saturation receive a **low concern**.

set supported a "low" concern, for 4% of the chemicals the US predicted a "high" concern and the MPD data set supported a "low" concern, and for 5% of the chemicals the US predicted a "high" level of concern and the MPD data set supported a "moderate" level of concern. Underprediction resulted in 6% of the chemicals having their fish toxicity scores raised from a "low" concern to a "moderate" concern and 1% going from a "moderate" concern to a "high" concern.

In contrast, for daphnid toxicity over- and underprediction of toxicity values occurred at about the same rate (16% versus 13%). The greatest percentage of chemicals (15%) where the US prediction was not supported by MPD data occurred with chemicals the US considered as "low" concern, while the MPD data supported a "moderate" concern level. In only 3% of the cases were the daphnid concern scores raised from a "low" concern to a "high" concern.

Potential reasons for the under- and overprediction in both species were investigated and appeared to be largely the same. These reasons include: reported LC50 above water solubility, use of nominal concentrations for chemicals having significant volatility from water, water solubility enhancement with a solvent, impurities, and apparent poor solution preparation. When the EC chemicals having questionable data were removed from the data set, the agreement between the US predicted values and the EC measured values is 87% for fish acute toxicity and 79% for daphnid acute toxicity.

One advantage of the US SAR methods over the MPD data set is that the US SAR analysis evaluates all of the potential effects and concerns of a chemical, e.g. acute and chronic toxicity to fish, aquatic invertebrates, and green algae, including benthic organisms, aquatic insect, and submerged aquatic vegetation. In addition, potential effects to terrestrial organisms, e.g. birds, earthworms, insects, vascular plants, and soil microbes, are evaluated. The MPD for environmental effects is restricted at present to fish and daphnid acute toxicity tests. If the overall EPA level of concern is compared with the level of concern for acute fish toxicity as measured by the MPD data set, there is concordance in 54% of the chemicals. Further analysis of these data reveals that in 28% of the non-concordant cases the driving concern was for algal toxicity, and in 8% of the cases chronic effects were the major concern; these effects are not included in the MPD data set. Comparing the overall EPA level of concern with the level of concern supported by the MPD data for each chemical, the trend towards overprediction of toxicity becomes clear (42% or 59 chemicals). However, recall that if only fish toxicity levels of concern are compared, the overprediction falls to 16%.

If the overall EPA level of concern is compared with the level of concern for acute daphnid toxicity 24-hr EC50 values as measured by the MPD data set, there is concordance in 54% of the chemicals. Further analysis of these data reveals that in 14% of the non-concordant cases, the driving concern was for algal toxicity, in 6% of the cases chronic effects were the major concern, and in 9% of the cases the predicted value was for a 48-hr EC50 instead of the MPD 24-hr EC50. Again, as with the fish values, if the overall EPA level of concern for daphnid toxicity is compared with the level of concern supported by the MPD data, the trend towards overprediction of toxicity is again apparent (37%, 51 chemicals). As with the fish acute values, if only the daphnid toxicity levels of concern are compared, the overprediction falls to 23%.

These analyses demonstrate that in a significant number of cases the driving concern for the US was an effect outside of the MPD data set; this suggests that the MPD data set may be improved by expanding the end-points included in the MPD. The addition of the algal toxicity test would allow the MPD data set to identify chemicals which show their greatest effects toward algae and plants, while the addition of the daphnid reproductive toxicity test would give the MPD a greater chance of identifying chemicals causing chronic toxicity.

5.1.2.5. Other considerations

Several additional factors, specifically chemical purity, classes of chemicals included in the MPD set, and the summary nature of the MPD data, may have added uncertainty to the study that was not possible to quantify.

Unlike the US, which requires pre-manufacture notification, the EC requires pre-marketing notification. For US pre-manufacture notification, the notified chemical is most often submitted as a "pure" compound (i.e. 95% or greater purity), while for EC pre-marketing notification the notice pertains to the substance "as marketed," which is often a formulated product (i.e. a mixture containing other chemicals or solvents). This distinction has important implications for the predictability of physico-chemical properties, biodegradation, and potential hazard concerns. In the US, the new chemical and any impurities reported by the submitter and/or identified as being likely contaminants by the EPA are considered when assessments are performed. In the EC, the submitter is required to provide purity information for the product as marketed and any test data pertaining to this product. Although in only one case did this distinction result in a large disparity in predicted systemic toxicity versus experimentally determined systemic toxicity, more subtle disparities may not be easily discerned. Clearly, in the physico-chemical properties exercise this difference in chemical substances played a not insignificant role in differing results between predicted values and experimental values. The study, however, suggests that the US should consider requiring purity tests for PMN chemicals which are subjected to EPA-required testing. The purity analysis should be conducted on the new chemical as produced via commercial production processes (i.e. characterize the commercial chemical, not a research and development (R&D) sample, which may differ significantly from the commercial substance).

Although the EC chemicals provided a wide range of chemical classes, the number of chemicals in each class and the classes themselves were not wholly representative of the numbers and classes that are typically reviewed by the US. For example, the EC does not routinely review polymer chemicals, so few polymers were included in the study. On the other hand, the EC scheme includes pesticide active ingredients and pharmaceuticals. In the US new chemicals scheme, such chemicals are reported under TSCA only if they have TSCA uses (e.g. industrial or consumer uses). Thus, pesticides and pharmaceuticals occurred with greater frequency in the MPD set of chemicals than would be expected in a typical equivalent set of US new chemicals. Thus, the experience and expertise of the US new chemical assessors was not a "perfect fit" for some of the EC chemicals and the skewed frequency of the classes of chemicals may have affected the US performance in this study.

Lastly, the data from the EC were available to the US only in summary form. The original data were reviewed and a summary was prepared by the Competent Authority in the EC country of origin. These summaries varied widely in the level of detail, so the US assessors were limited in their ability to interpret results independently. While most likely not a limiting factor in the interpretation of overall (qualitative) levels of concern, it may have been a factor in the quantitative determination of the level of toxicity.

5.1.3. Summary

Looking at the overall results of the MPD/SAR study, it is interesting to note that overall the physico-chemical properties appear to be the most difficult to predict accurately, but are among the most inexpensive to measure. On the other hand, predicting of health hazards appears reasonably good although there is an issue, as discussed above, with the prediction of systemic toxicity. Targeted testing may offer a cost effective alternative to use of a standard test battery. US ecotoxicity predictions appear to be reasonably accurate in assessing acute toxicity for fish and daphnia.

The MPD/SAR study provided a unique opportunity to gain insight into the strengths and weaknesses of the SAR approach used by the US versus the MPD approach of the EC in assessing the potential fate and effects of new chemicals. Analysis of the results of this study have shown that while the SAR approach has largely been successful in identifying chemicals of concern, the process could be improved by selectively incorporating specific testing schemes into the process. Results from such schemes would serve two purposes: to gain insight into chemical toxicities and to improve our predictive capabilities. Improving predictive capabilities would result in better hazard assessment for new chemicals by providing a richer data base upon which to base predictions as to their fate and effects. These enhanced capabilities would also serve to avoid questionable testing requirements and thus spare manufacturers the cost of such testing while not compromising worker, consumer or environmental safety. Such a focussed effort would provide valuable data while not presenting large overall cost implications.

5.2. Conclusions: EC perspective

5.2.1. Introduction

This study has provided many useful insights into the strengths and weaknesses of the notification scheme for new chemicals established under Directive 67/548/EEC as amended. The results will be taken into account in the preparation of any future modification to the MPD or "base set" used for the notification of chemicals marketed in quantities in excess of one tonne per annum. In addition to the direct benefits which will result from the project, the study also allowed the Commission and the national authorities in the Member States to obtain a better understanding of the PMN system as applied in the United States under TSCA. While the benefits which accrue from such improvements in mutual understanding are less tangible and difficult to

quantify, they are nonetheless real and will certainly facilitate the development of a more global approach to chemicals control in line with the objectives set out in Chapter 19 of Agenda 21 of UNCED.

5.2.2. Synopsis

5.2.2.1. Physico-chemical end-points

Of the three end-points which were adequately explored, the SAR methods performed best in relation to log P_{OW}. However, even for this end-point, the predictive methods could not be used with confidence for all chemical groups. Given the relatively low cost of carrying out these tests, the results of this project do not constitute a persuasive argument for introducing SAR into the "base set" as an alternative to testing.

5.2.2.2. Biodegradation

The SAR methods performed extremely well in relation to this end-point, and at the next revision of the "base set" consideration should be given to allowing, under defined conditions, the estimation of biodegradation using SAR.

5.2.2.3. Health effects

The SAR methods are not sufficiently developed in relation to the estimation of eye/skin irritation or sensitisation. As knowledge about these end-points is an essential part of the EC notification scheme, testing for these parameters will continue. SAR techniques were, in contrast, relatively successful in providing qualitative assessments of acute lethal toxicity, and the opportunity for building SAR into a future battery of approaches - including SAR, in vitro tests and non-LD50 animal tests - should be explored.

While the SAR methods displayed a tendency to underestimate sub-chronic 28-day, repeated dose toxicity, in most cases this involved an underestimate of the severity of the effects rather than true "false negatives". At the present time, it is unlikely that the testing requirements for sub-chronic/repeated dose toxicity in the "base set" will be modified. However, it is clear that the SAR techniques provide an excellent additional tool for informing decisions about further testing either immediately post "base set" or at level 1/level 2, as foreseen in the Directive.

With regard to mutagenicity, the results of this project would suggest that SAR could, in a future revision of the "base set", usefully be incorporated into a battery of approaches for evaluating the mutagenic potential of a new chemical. In particular, the issue of the apparent "false negatives" given by the current "base set" testing package needs to be addressed.

The proportion of substances in the test sample which were predicted as being of concern in relation to end-points not covered by the sixth amendment "base set", e.g. reproductive toxicity, developmental toxicity, carcinogenicity and

neurotoxicity, is a considerable source of disquiet. The seventh amendment to the Directive does foresee the introduction into the "base set" of a screening test for reproductive toxicity. In the light of this project, consideration should also be given to addressing the other "missing" end-points.

5.2.2.4. Ecotoxicity

The SAR methods performed extremely well in predicting acute toxicity to fish and daphnia. They also provided estimates of toxic effects, e.g. algal toxicity, not addressed in the "base set" of the sixth amendment. As part of any future revision, the conditions under which SAR predictions of acute toxicity to aquatic organisms could be integrated into the "base set" should be explored.

5.2.3. Overview

As indicated in the preceeding section, this project has identified a number of possibilities for making greater use of SAR as part of the "base set" testing package applied to new chemicals marketed in the European Community. These possibilities will be explored in the preparation of any future revision to the legislation. However, in contemplating any such revision, there are a number of factors which should also be taken into account.

1) The EC system is operated in a decentralized manner across twelve different national authorities: this figure will shortly be increased to 16 when the EFTA countries join the scheme in the context of the Enlarged European Economic Area. This means that any approach to notification has to be transparent and objective. Thus, while some SAR methods may be used successfully by a group of highly skilled experts working together over many years in one Agency, such an approach could not work in the decentralized system applied in the EC. This means that opportunities for the (consistent) systematic introduction of SAR into the EC scheme could only be considered where the predictive models could be applied objectively by all agencies working within the decentralised system.

2) The EC Directive puts great importance on the classification of a chemical. The emphasis given to classification is frequently misunderstood because the term classification is almost invariably linked with the term labelling, thereby giving the impression that labelling is the only purpose for which substances are classified: this impression is entirely false.

   Classification means the allocation of a substance to one of a number of danger categories on the basis of its intrinsic properties. The decision to allocate substances to a particular category is based on a series of agreed and published criteria. Classification is therefore synonymous with the term hazard/risk identification. Within the EC, classification is consequently the foundation for hazard assessment and the recently agreed Commission Directive laying down the general principles for the risk assessment of new chemicals, recognises

classification as providing the starting point for hazard/risk assessment. Secondly, classification may also be the basis for risk reduction: substances classified as carcinogens under the EC scheme are, for example, subject to severe restriction in the workplace under separate EC legislation. Finally, classification is also the basis for the system of hazard communication by means of standardized labels which has been developed in the EC.

Given the critical importance of classification for the entire EC policy on chemicals, it is essential that the current approach to classification on the basis of objective, transparent criteria is not put into question by allowing the possibility of using SARs instead of test data. Essentially, this would mean that SARs could be only admitted:

- if they were objective and reliable, and

- if they were able to generate precise quantitative estimations/ predictions of test results which could be incorporated into classification schemes, or

- if notifiers accepted the principle that classification on the basis of SARs would be admitted but escape from classification, i.e. non-allocation to a danger category would not be allowed.

3) The EC notification scheme is directed towards the substance as marketed, including impurities but excluding separable solvents and any non-essential stabilizers. The notification scheme is not concerned with purified substances nor is it concerned with formulated products (preparations). While it is clear that the SARs used in this study have in many cases performed very well, such predictive models are, in the most part, based upon pure substances. For SARs to be used in a systematic way in the context of the EC notification scheme would require this important issue of impurities to be addressed.

ANNEXES 1-15

ANNEX 1

US EPA AND EC EXPERTS ON SAR WITHIN THIS JOINT PROJECT

Experts of the US EPA:

Charles Auer

Bob Boethling

Michael C. Cimino

Richard G. Clements

Diana Darling

Mary Henry

Leonard C. Keifer

Asa Leifer

J. Vincent Nabholz

Pauline Wagner

EC Experts:

Herbert Baumann, Bundesgesundheitsamt, Germany

Peter Bougeard, Health and Safety Executive, UK

Andreas Gies-Reuschel, Umweltbundesamt, Germany

Kornelia Grein, Commission of the European Communities

Petra Greiner, Umweltbundesamt, Germany

Björn Hansen, Commission of the European Communities

Jim Hart, Commission of the European Communities

Joop Hermens, Research Institute of Toxicology, Netherlands

Derek James, Health and Safety Executive, UK

Patricia Koundakjian, Health and Safety Executive, UK

Patrick Murphy, Commission of the European Communities

Jay Niemela, Danish Environmental Protection Agency, Denmark

Hans Opdam, TNO - Medical Biological Laboratories, Netherlands

Christine Reteuna, Ministère de l’Environnement, France

Martine Reynier, INRS, France

John Vosser, Health and Safety Executive, UK

ANNEX 2

COMPANIES WHICH GAVE PERMISSION TO USE THEIR SUBSTANCE'S DATA WITHIN THIS PROJECT

3M
AGFA GEVAERT
AGFA GEVAERT AG
AH MARKS & Co Ltd
AKZO CHEMIE
AKZO CHIMICA SPA
AVONDALE CHEMICAL COMPANY
BASF
BENTONE-CHEMIE GmbH
BOEHRINGER INGELHEIM KG
BOEHRINGER MANNHEIM GmbH
BUSH BOAKE ALLEN Ltd
CECA
CHEMISCHE FABRIK STOCKHAUSEN GmbH
CHIMEX
CIBA GEIGY
CIBA GEIGY A/S
CIBA GEIGY DANMARK
CIBA GEIGY D&C
CIBA GEIGY GmbH
CIBA GEIGY INDUSTRIAL CHEMICALS
CIBA GEIGY MARIENBERG GmbH
CIBA-GEIGY PLASTICS
CONTINENTAL PHARMA
CYANAMID BV
DEUTSCHE EXXON CHEMICAL GmbH
DEVELOP DR. EISBEIN GmbH&Co
DOMUS IND. CHIM.
DOW CORNING Ltd
DSM CHEMICALS
DSM RESINS B.V.
DU PONT DE NEMOURS
DU PONT DE NEMOURS (DEUTSCHLAND) GmbH
E. MERCK
ENICHEM SYNTHESIS
EPSON DEUTSCHLAND GmbH
EPSON FRANCE
ERGAM RONEO
FARCHEMIA SRL
FORMICA
FRAT. LAMBERTI
FUJI HUNT
FUJI PHOTO FILM BV
GALVANOCOR (GB) Ltd
GOODYEAR CHEMICALS
GRACE SERVICE CHEMICALS GmbH

GREAT LAKE CHEMICALS (EUROPE)
HAAG TECHNO BV
HERCULES
HIMONT ITALIA SPA
HOECHST AG
INTERNATIONAL PAINT p.i.c.
ISF
ISF SPA
JANSSEN PHARMACEUTICA
KODAK PATHE
KRONOS SA/NV
LAGOR SPA PROD. CH.
LONZA FRANCE
LONZA ITALIA SPA
LONZA-WERKE GmbH WALDSHUT
LUBRIZOL FRANCE
LUBRIZOL Ltd
LUPEROX GmbH/MAP
MERCK, SHARP & DOHME
MINOLTA CAMERA HANDELSGES. mbH
MOBIL OIL Co Ltd
MONSANTO
MONTEDISON SPA
N L ABBEY CHEMICALS Ltd
OLIN HUNT
PALMAROLE
PANASONIC
PANASONIC DANMARK A/S
PANASONIC DEUTSCHLAND GmbH
PANASONIC ITALIA
POLAROID (EUROPA)
PROCTER AND GAMBLE LIMITED
Q O CHEMICALS INC
RHONE POULENC
RICOH EUROPE
RICOH FRANCE
RICOH NEDERLAND
RIEDEL-DE HAËN AG
RWE-DEA AG FÜR MINERALÖL UND CHEMIE
SANDOZ
SANDOZ HUNINGUE
SANDOZ ITALIA
SANDOZ PROD. CHIM. SPA
SANDOZ-QUINN PRODUKTE GmbH
SANDOZ SPA
SANOFI CHIMIE
SCHERING AG
SCHERING AGROCHEMICALS Ltd
SCHLOETTER Ltd
SHELL CHIMI
SHELL NEDERLAND CHEMIE BV
STAUFFER CHEMICAL
TESSENDERLO CHEMIE

TEXACO Ltd
TEXAS ALKYLS BELGIUM
TH. GOLDSCHMIDT AG
WACKER-CHEMIE GmbH
WIGGINS-TEAPE
WINKELHORN A/S
WWE. AUGUST HEYMANNS & Co
YAMANOUCHI IRELAND COMPANY, Ltd

ANNEX 3

GENERIC CHEMICAL DESCRIPTIONS OF THE CHEMICALS IN THIS PROJECT

Chemical	Description
4	alkyl aluminium, halogenated complex
6	aryl dialkyl ammonium clay complex
16	mixture of bis-(hydroxyalkylammonium) salts of fatty acids
17	reaction mixture of unsaturated fatty acids, imino-dialcohol and inorganic acid
21	complex haloaryl alkylamide
23	substituted alkali pyrazoline arylsulfonate
24	phenolic benzopyran derivative
26	substituted spiro bis-indane
37	aryl substituted alkyl dione
44	perhalo polycyclic hydrocarbon
47	alkyl hydroquinone
49	phosphorodithionic aliphatic amine
50	halogenated polymer of polyalkylmethacrylate
53	complex alkyl ester of a diaza-spiro carboxylic acid
54	thioaryl morpholine ketone
61	haloaryl acetanilide
68	halotriazine dye
69	halotriazine azo dye
70	haloaryl anilide
76	mixture of aryl (substituted benzotriazole) esters of polyethylene glycol
78	azo dye
79	aryl organo-nickel complex

Chemical	Description
87	substituted phenol
96	azo dye
99	trialkoxy vinyl silane
101	halotriazine dye
102	bis-(dialkyl)aryl-substituted peralkyl phenol
106	bis-(bicycloalkyl) alkane
107	alkyl substituted siloxy aluminium
108	halogenated alkylaryl silane
107	dialkyl carbonate
113	alkyl alkoxybenzene di-alkyl valerate
118	alkyl amino triazole
124	haloaryl silane substituted triazole
128	haloaryl substituted pyrazole
133	pyrazole substituted with various aryls
144	alkoxy aryl quinoline
148	substituted polyaromatic hydrocarbon
151	bisphenol A derivative
155	mixture of various substituted benzotriazoles
156	alkyl substituted aryl thiocarbamate
164	phosphothioalkylamide mixture
170	N-arylalkylamino acetophenone hydrochloride
173	mixture of aryl tertiary amines
176	alkylamino chain substituted with piperidine and triazine
182	calcium alkyl aryl sulfonate
186	haloaryl substituted triazole

Chemical	Description
192	mixture of esters of alkane phosphinic acid
194	pyrazole substituted with various aryls
196	halotriazine dye
197	haloalkylphenoxy aminoaryl aniline hydrochloride
200	halo substituted diaryl alkane
204	variously substituted haloacetanilide
214	partially quaternised arene tallow carbamate
216	substituted bis-(cycloalkene) iron
217	aryl pyrrolopyrroledione
218	cycloalkyl alkyl substituted xylene
219	haloalkoxy arene
222	aryl substituted alkylisocyanate
224	phenoxodiazine dye
235	alkyl aminoalkyl substituted benzothiazolethione
237	alkoxy alkyl silane
239	alkylamino arene substituted halophthalide
240	halotriazine dye
241	alkyl pyridinium halide
242	alkyl pyridinium halide
253	thioalkyl cresol
256	amino acid amide
263	chromium azo dye
265	haloacetyl amino acid derivative
267	haloaryl-ketone polymer
268	alkylamino carboxylic acid, Cn(medium chain)halo-alkyl ester

Chemical	Description
269	substituted alkyl styrene polymer
270	alkyl piperidine succinate
271	mixed sodium salts of aminocarboxylic acid
275	nitroaryl azo dye
278	mixed isomers of a terpene carboxylate
281	diaryl ketone
283/429	mixture of perhaloalkyltetraoxodecanates
286	halo alkyl alkoxy aryl sulfonamido substituted pyrazolo-triazol
287	alkyl aryl sulfonamide substituted indole
289	azo nitrobenzoate dye
291	azo dye
292	haloaryl alkyl silyl triazole
300	arylpropionate alkyl ester
307	haloalkoxy nitroaryl
309	diaryl substituted aryl diamine
312	alkyl diol substituted arylamine
318	azo dye, calcium salt
320	nickel complex of oxyiminopolyaryl
321	substituted triazine trione
330	carboxyalkyl amino acid
335	chromium azo dye, alkyl ammonium salt
336	aluminium tris alkylphosphonate
337	haloalkyl phosphate tri-ester
340	cyano-alkyl thiazole
341	thia lactam derivative

Chemical	Description
342	alkylene carbonate
344	arylacetoacetate alkanolamine salt
348	aryl substituted urea
349	aryl substituted anthracenedione
354	alkyl alkoxyaryl carbamate
355	methacrylic acid, aryl ester
360	aryl alkyl carboxylate
361	alkyl imidazolidine substituted halobenzoate
362	aryloxyalkyl tosylate
364	halotriazine azo dye
366	alkenyl substituted polysiloxane
368	alkylalkoxy silane
369	ferric ammonium salt of carboxyalkyl amino acid derivative
370	haloalkene carbonate
373	C10-terpene
376	condensation mixture of alkylphenol, formaldehyde and alkane thiol (alkylthioalkylaryl substituted methylene bis-(alkylaryl))
379	branched alkene
381	substituted phenoxazine pigment
383	aryl triazine trione
386	aryl alkenyl morpholine
393	alkyl amino acridindione
394	potassium salt of substituted amino acid
396	substituted imidazole
398	cycloalkyl alkoxy silane
401	chiral aryl arylamide dibenzoyl tartrate

Chemical	Description
406	aryl glycidyl ether
411	haloaryl azo dye, calcium salt
413	dialkyl ester of alkyl disulfide
414	hexahydro aromatic carboxylate, ammonium salt
415	pyrazole substituted arylsulfonamide
416	aryl substituted naphthyl ketone
417	spiro naphthoxazine
420	halo alkoxy benzophenone
421	aryl aminoalkenyl ester sulfone
425	alkylamino alkanol
431	haloaryl alkyl carbonate
436	alkylammonium alkylphosphonate
437	mixture of substituted thiadiazoles
439	sulfonated styryl biphenyl
441	alkyl substituted heterocyclic amine hydrochloride
442	sulfonated vinylic acetamide
443	aza bicyclo alkane
444	heterocyclic ester of methacrylic acid
445	copolymer of methacrylic acid and heterocyclic ester of methacrylic acid
446	aryl substituted thiazole
451	alkoxy alkyl ester of unsaturated carboxylic acid
472	alkoxyalkyl tetradecanoate