4. Results

4.1. Introduction

For this project, the test set of chemicals was comprised of 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 predefined 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 the 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.