## 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-market 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 physical/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 physical/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 (physical/chemical properties, biodegradability, health effects, and ecotoxicity) for discussion purposes and appear below.

##### 5.1.2.1. Physico-chemical properties

The physical/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 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.

---

[1]This study examined hazard (or toxicity) predictions and did not examine 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.

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 $\pm$ 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 program 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 that 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/under estimation can result in an over/under estimation of the exposure associated with a chemical and thus contribute to an over/under estimation 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 program 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 under-estimated by the US. Water solubility contributes to the hazard and exposure portions of the risk assessment for new chemicals and over/under estimation can result in an over/under estimation of the hazard/exposure associated with a chemical and thus contribute to an over/under estimation 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 12 may be MPD "false negatives" caused by testing in assay systems known to be insensitive to specific classes of chemicals. These 12 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 under-predict 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 under-predicted the concern level, the MPD supported a "low-moderate" concern whereas the SAR-based call was for "low" concern while in 3 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.

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.

---

[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 physical/chemical properties (e.g., pyrophoric) that rendered them unsuitable for testing.

When overall level of concern scores for health effects are considered, (i.e., a bottom-line assessment considering all effect areas), the trend towards under-prediction rather than over-prediction (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 ($\pm$ 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 under-predict toxicity rather than under-predict (11% versus 7%); for 7% of the chemicals the US predicted a "moderate" level of concern[5] whereas the MPD data 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. Under-prediction 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 under-prediction 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.

---

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

- Acute toxicity values < 1mg/l and/or chronic toxicity values < 0.1mg/l receive a **high concern**.

- Acute toxicity values from 1 to 100mg/l and/or chronic toxicity values from 0.1 to 1mg/l receive a **moderate concern**.

- Acute toxicity values > 100mg/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**.

Potential reasons for the under- and over-prediction 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 over-prediction of toxicity becomes clear (42% or 59 chemicals). However, recall that if only fish toxicity levels of concern are compared, the over-prediction 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 over-prediction 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 over-prediction 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 suggest 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 physical/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 pertain 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 physical/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 physical/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.