

*39th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals,  
Pesticides and Biotechnology  
15-17 February 2006*

**Focus Session: Experiences using integrated approaches to fulfil information requirements  
for Testing and Assessment**

**Contribution by Environmental NGOs  
(Environmental Defense Perspective on Integrated Approaches to  
Chemical Testing and Assessment)**

**Environmental Defense Perspective on  
Integrated Approaches to Chemical Testing and Assessment  
Focus Session, 39<sup>th</sup> Joint Meeting, February 2006**

**Richard A. Denison, Ph.D., Senior Scientist  
John M. Balbus, M.D., M.P.H., Health Program Director  
Environmental Defense**

## **Introduction**

Interest in promoting so-called “integrated approaches” to chemical testing and assessment is motivated by a desire to: gain efficiencies in assessing new chemicals prior to market introduction as well as chipping away at the huge backlog of un- or under-assessed chemicals already on the market; reduce the costs associated with traditional testing; and reduce unnecessary use of laboratory animals.

We appreciate and share interest in achieving these objectives, and we are both supportive of and engaged in promoting the development and use of many of the alternative methods that comprise more integrated approaches. At the same time, it is critical that an appropriate balance be struck with other equally important objectives: assuring full protection of human health and the environment; basing decisions on scientifically sound and defensible information; ensuring that all assessment information used to make such decisions is independently verifiable and reproducible; and maximizing transparency in communicating the basis for decisions to stakeholders and the general public.

To achieve all of these objectives, we believe the following “guiding principles” need to be followed:

*Avoid over-reliance.* Precisely because of the large benefits of reducing costs to government and industry and reducing animal use, there can be a strong incentive to *over-rely* on alternative methods. This potential must be acknowledged and tempered, through:

- the creation of clear, scientifically sound guidance on the appropriate and inappropriate uses of each alternative method;
- requirements for justifying and documenting both use of alternative methods and decisions based on such information; and
- careful independent expert review.<sup>1</sup>

*Avoid selective use and reporting (“double standard”).* A corollary concern is the potential for alternative methods to be used, especially by industry, not only as the option of first resort, but also under a “double standard.” We have already seen some indications of companies, for example, arguing that (Q)SAR results are sufficient when they are favorable, i.e., “exonerate” their chemical, and proceeding to do actual testing only in cases where the (Q)SAR results indicate a hazard. While we consider the latter response wholly appropriate, it begs the question as to who decides whether and in what settings (Q)SAR results are deemed sufficiently reliable. Safeguards to prevent selective use and reporting are needed; for example, there should be a requirement that all results derived using all methods employed be reported to regulatory officials.

*Screening vs. other uses.* Another related question is the use to which information derived from alternatives to direct testing is put. As a general matter, we regard such methods – especially when used

---

<sup>1</sup> Given these needs, we also note that the prospect raised by some that integrated approaches will result in “the simplification and streamlining of existing Test Guidelines and associated testing strategies” is likely overly optimistic.

individually – to entail sufficient uncertainty as to be most appropriate for priority-setting or screening-level assessments, and less so to serve as the basis for more full-blown risk assessments or for risk management decisions. Exceptions may arise when a “critical mass” is reached, e.g., a sufficient number of mutually-supporting or corroborative results derived using alternative methods. In such cases, there may be a sufficient degree of confidence to support a more definitive conclusion or decision; this situation is akin in some ways to a weight-of-evidence approach, which we discuss in more detail below.

*Transparency.* An approach that relies on information generated through multiple, diverse methods carries with it an added burden of transparency. The nature, source and means of derivation of each data value needs to accompany it in any subsequent presentation or communication of the data, and should be an integral part of the justification provided for any conclusions or decisions based on such data. Some assessment of the degree of confidence in or reliability of the data is another prerequisite to transparency, and any resulting uncertainty should be captured and communicated through a clear articulation of appropriate qualifications or limitations that apply to conclusions or decisions based on such information.

*Continuing need for generation of experimental data.* Another essential point is that development and improvement of many alternative methods is highly dependent on having a robust and expanding underlying dataset of values derived from *in vivo* testing. Such data are necessary either to provide for the development and refinement of the algorithms that underpin mathematical predictive models ((Q)SARs), and to allow correlations to be established between *in vivo* results and those of other systems (*in vitro* testing, toxicogenomics). In short, at least in the near term, these alternatives will only be as good as the *in vivo* data that underpin them; without continued commitment to enhance databases derived from *in vivo* testing, the applicability and reliability of such alternative methods will not progress to the point where they can fully replace *in vivo* test systems.

Integrated approaches will only be as good as the “sum of their parts.” The appropriate and inappropriate uses of each method need to be clearly understood, and their limitations acknowledged and reflected, in choosing the uses to which such approaches are put and in documenting any resulting conclusion or decision.

In the remainder of this paper, we first briefly discuss the closely related issue of weight-of-evidence (WOE) approaches, and then address the more specific issues, opportunities and limitations associated with each of the identified alternative approaches that may comprise the elements of a more integrated approach:

- (Quantitative) Structure-activity relationships [(Q)SARs]
- Read-across methods (using chemical category and analog approaches)
- *In vitro* tests
- Toxicogenomics (and related emerging technologies)
- Exposure information

## **Weight of Evidence**

Virtually by definition, integrated approaches imply that a weight-of-evidence (WOE) approach is to be used, that is, a variety of information is considered in conducting an assessment or reaching a decision. WOE approaches are, of course, not new and have been used implicitly or explicitly in a variety of settings, especially in the risk assessment arena.

The largest concerns about the application of WOE, and by extension, integrated approaches, are the absence of a rigorous definition of what constitutes WOE, or clear guidance and standards for the use of

WOE and associated documentation and communication needs. A recent paper<sup>2</sup> by Douglas L. Wood of the US National Cancer Institute provides empirical evidence for such concerns. He conducted an extensive survey of the published risk assessment literature, finding the following wide diversity of types of “uses” of WOE:

- “(1) metaphorical, where WOE refers to a collection of studies or to an unspecified methodological approach;
- (2) methodological, where WOE points to established interpretative methodologies (e.g., systematic narrative review, meta-analysis, causal criteria, and/or quality criteria for toxicological studies) or where WOE means that “all” rather than some subset of the evidence is examined, or rarely, where WOE points to methods using quantitative weights for evidence; and
- (3) theoretical, where WOE serves as a label for a conceptual framework.”

Clearly, if integrated approaches that rely on WOE are to meet even basic tests for transparency, objectivity and accountability, addressing this lack of consistency must be a first priority. As noted by Dr. Wood, among the problems to be remedied are:

- the multiplicity of definitions and uses;
- the multiplicity of weighting schemes and criteria for applying them; and
- defining the role of judgment in applying WOE approaches.

Wood goes on to offer an important recommendation: “The WOE concept and its associated methods should be fully described when used. A research agenda should examine the advantages of quantitative versus qualitative weighting schemes, how best to improve existing methods, and how best to combine those methods. ... The goal of this approach is to work toward a consensus on the meaning and methods of weight of evidence, such that a recognizable standard can be created and accepted.”

We consider this recommendation, extended to integrated approaches to chemical testing and assessment, to be an appropriate starting point for the OECD’s further consideration of such approaches.

### **(Quantitative) Structure-activity relationships [(Q)SARs]**

While development and use of (Q)SARs holds considerable promise to reduce testing needs, at present there are significant limitations to their use. Reliable models are available for only a subset of relevant endpoints, and are in particular lacking for most human health-related endpoints, especially chronic ones. The question of validation continues to be a contentious one, with no clear agreement on what constitutes sufficient validation. Public access to underlying algorithms and training datasets has yet to be assured for many (Q)SARs, despite such access being identified as key to providing needed transparency and accountability in the application of (Q)SAR approaches, especially in regulatory contexts. Existing (Q)SARs have limited “domains” of applicability, with many common types of chemicals falling outside. Finally, the accuracy and reliability of (Q)SAR-derived estimates vary from one (Q)SAR and endpoint to another, and the estimates they generate often vary considerably from available experimental values.

At SIAM 21, a “(Q)SAR application pilot project” was proposed (see Document ENV/JM/EXCH/SIAM(2005)9), generally agreed to and forwarded to the Existing Chemicals Task Force. Under the pilot, sponsors would apply selected (Q)SAR models for selected SIDS endpoints to the chemicals for which they are also preparing a SIDS Initial Assessment Report, thereby allowing a direct comparison of experimental and predicted values for the same endpoints. We support this proposal: In addition to facilitating an increase in OECD experience with the application of (Q)SARs, the project would

---

<sup>2</sup> Weed, D.L. (2005) “Weight of Evidence: A Review of Concept and Methods,” *Risk Analysis*, Vol. 25, No. 6, pp. 1545-1556.

help to illustrate both the opportunities for and limitations of using (Q)SARs as an alternative to direct testing.

Having a full understanding of both the appropriate application and the limitations of (Q)SARs is essential to ensuring their appropriate use. OECD has recently focused appropriate attention on the need for validation of (Q)SARs, as a means to gain greater acceptance for their use. Two case studies presented at the 2nd Meeting of the ad hoc Expert Group on (Q)SARs (held 20-21 September, 2004 in Paris) are useful to consider in this context.

The first case study, from the U.S.,<sup>3</sup> described its view of problems arising from too rigidly applying validation principles to (Q)SARs the USEPA uses to assess new chemicals. (Unlike most other OECD countries, U.S. law does not require that chemical manufacturers provide a “base set” of hazard data when notifying authorities of their intent to make a new chemical, and hence the great majority of new chemical notifications submitted in the U.S. lack such data. In addition, EPA is given only 90 days to decide whether a chemical needs any restrictions placed on its manufacture or use; if it fails to act, manufacture can commence. For this reason, EPA has developed and made extensive use of (Q)SARs to predict the hazards of chemicals it reviews that lack actual test data.) EPA argued that whether a (Q)SAR is “valid” depends in part on how and for what purpose it is used; e.g., used as a means to rapidly screen and prioritize many chemicals to identify those in most need of further scrutiny, a higher degree of uncertainty may be accepted than, say, for risk assessment purposes. Hence, EPA argued that both (Q)SAR use and concepts of validity need to be flexible and tailored to the regulatory needs of each country.

While this argument has merit, and the constraints faced by EPA are indeed considerable, I would raise two concerns:

- Where such constraints do not exist (i.e., the rest of the OECD), the larger question needs to remain whether and when (Q)SAR-generated estimates can reliably replace experimental data and hence serve as a scientifically sound alternative to testing.
- (Q)SAR estimates generated by or submitted to regulatory agencies for use in such a context-specific manner cannot be assumed to be “valid” universally, and hence should not simply be adopted for use by other countries that do not face the same constraints and may need or wish to develop a more certain basis for regulating chemicals.

An EU country participant contrasted the EPA’s use of (Q)SARs versus that contemplated under REACH: In the former case, the government develops, applies and interprets (Q)SAR results; under REACH, industry would utilize (Q)SARs, and governments would have to be in a position to be able to judge the validity of the results. This difference, it was argued, suggests a greater need for rigorously validating (Q)SARs in advance, at least for this type of use.

The second case study, provided by Denmark,<sup>4</sup> compared experimental data and (Q)SAR estimates for chemicals assessed at SIAMs 11-18 held over the previous several years. Only five endpoints were able to be compared: biodegradability; acute toxicity to fish, aquatic invertebrates and algae; and mutagenicity. These are the endpoints for which the “best” (Q)SARs exist – those that are based on large sets of experimental data and have been considered to provide the most reliable estimates. The results are summarized below:

- The (Q)SAR models were able to identify 80-90% of the chemicals that actually tested as readily biodegradable, but (depending on the specific model) only 46-80% of the chemicals that actually tested as not biodegradable.

---

<sup>3</sup> “Regulatory Application of (Q)SARs: A U.S. EPA Case Study,” ENV/JM/TG(2004)25/REV2.

<sup>4</sup> Comparison of SIDS Test Data with (Q)SAR Predictions for Acute Aquatic Toxicity, Biodegradability and Mutagenicity on Organic Chemicals Discussed at SIAM 11-18,” ENV/JM/TG(2004)26/REV1.

- The fraction of chemicals for which the (Q)SAR predictions for acute toxicity “agreed” (defined as being within an order of magnitude of the test result) with the experimental data were: 4/5 of the chemicals for fish; 3/4 for invertebrates (specifically, crustaceans); and 2/3 for algae.
- The (Q)SAR models were able to identify 95% of the chemicals that actually tested as negative for mutagenicity, but (depending on the specific model) only 60-80% of the chemicals that actually tested as mutagenic. However, the comparison included so few substances that tested positive that the latter conclusion must be viewed as tentative.

As with studies describing scientific validation efforts for specific (Q)SAR models, these results suggest both the utility of certain (Q)SARs but also some important limitations to their accuracy and reliability. As with the other methods discussed in this paper, appropriate use of (Q)SARs can play an important role in supplementing and extending the base of information available for use in chemical assessment – as long as their shortcomings are kept in mind and clearly communicated. While we believe that relying *solely* on (Q)SAR-derived information will be sufficient only in relatively rare cases, such information considered as part of an integrated approach may well be able to help compensate for weaknesses or resolve conflicting results found in data derived from other methods, thereby strengthening the overall assessment.

### **Read-across methods (using chemical category and analog approaches)**

We support efforts to promote scientifically appropriate use of category approaches, and are pleased to see the enhancements made recently to OECD guidance governing: category definition and justification, the process to be followed for verifying category soundness once data have been developed; and the specific methods to be used to assign specific hazard values to individual members of a category that have not been directly tested. Continuous refinement and enhancement of this guidance will be needed to incorporate experience in real-world application of the guidance.

Equally important, even with clear guidance in place, careful expert review of all category-based assessments is essential, as demonstrated by experience with chemical categories under the US HPV Challenge Program, in which about 80% of all sponsored chemicals are being assessed as members of categories rather than individually. In comments filed on the initial industry submissions for these categories, USEPA and public comments identified concerns or deficiencies in the category justifications about half of the cases. For example, some categories were found to be overly broad or ill-defined, or a whole category or the inclusion of specific chemicals was found not to be supported by available data. While many or most of these concerns were addressed in subsequent revisions, the experience highlights the critical role that expert review plays in applying category-based approaches.

Presentation of the results of applying category-based approaches must be transparent. Assuming that a category is still found to be justified once all data development has been completed and evaluated, the final dataset needs not only to provide all required data elements for each category member, but also to clearly indicate those values that are extrapolated rather than experimentally measured, together with clear explanations as to how each value has been derived.

### ***In vitro* tests**

*In vitro* tests comprise a gamut of different assays, ranging from relatively simple protein or receptor binding assays to the complex simulation of heterogeneous tissues outside of living organisms. In comparison to the similarly wide range of *in vivo* tests, *in vitro* tests offer certain advantages and disadvantages. Advantages include reduced cost, reduced or no sacrifice of animals, generally rapid results, and the ability to perform multiple replications or parallel experiments simultaneously. The primary disadvantage of *in vitro* testing is increased uncertainty in interpreting results, due to difficulties

correlating binding profiles or cytotoxicity with *in vivo* effects, and the inability to account for metabolism or other complex interactions that can moderate or exacerbate toxicity *in vivo*.

While some *in vitro* tests, such as the Ames test, have long been incorporated into predictive toxicology, there are still large knowledge gaps that must be filled before *in vitro* tests can begin to replace *in vivo* tests in most applications. As those knowledge gaps are filled, however, it is likely that there will be specific applications in which *in vitro* tests can form part of an integrated assessment. One such possibility would be the inclusion of high-throughput binding assays for an array of different endocrine receptors, in order to be able to categorize compounds' or mixtures' ability to stimulate various endocrine pathways. Such studies are common in the published literature,<sup>5</sup> but these methods are not yet common in regulatory use.

*In vitro* methods hold great promise for more rapid screening of chemical compounds and environmental samples, but they also present many of the same limitations as the use of QSAR models mentioned above. Because *in vitro* findings are several steps removed from whole animal histopathology, they are more easily discounted when they suggest a problem. Indeed, many in industry argue that *in vitro* methods should not be relied upon because they are invariably more "sensitive" than the corresponding *in vivo* studies; even where this is the case, however, that property might well be *desirable* if such tests are used as a first-line screen for chemicals. More generally, whether *in vitro* methods are always more sensitive remains to be seen; there simply are not enough correlative data with *in vivo* studies to know at this point. Given this, just as many in industry are concerned about an over-reliance on positive findings from *in vitro* studies, an over-reliance on negative results from *in vitro* methods, in the absence of documentation of their sensitivity, could also lead to erroneous decisions and to inadequate public health protection.

As mechanisms of toxicity continue to be elucidated, the utility of *in vitro* testing may well increase, initially for screening of chemicals but ultimately, perhaps, for use in more definitive assessments. In order for this to occur, an intensive effort to determine and map out the relevant mechanisms for a wide variety of types of toxicity is necessary. In addition, regulatory toxicology agencies and laboratories will need to run selected *in vitro* assays in parallel with traditional *in vivo* toxicological tests on a range of chemicals so that databases that correlate *in vitro* findings with relevant adverse health outcomes can be populated. Initial attempts to incorporate *in vitro* assays into larger, integrated assessments should focus on limited applications with the greatest knowledge base, such as endocrine or metabolic disruptors. Confidence gained in these limited applications may foster increased investment and confidence in broader development and use of *in vitro* assays. As with all of the methods discussed in this paper, such confidence will also be dependent on transparency of methods, materials, and interpretation.

### **Toxicogenomics (and related emerging technologies)**

The term toxicogenomics has developed as an umbrella term describing a number of different technologies that measure global or large-scale gene, protein and metabolite expression within biological systems and systematically analyze the resulting data. Thus, DNA and RNA microarrays, proteomic and metabolomics assays, and the bioinformatics systems needed to record and analyze such complex datasets are all part of "toxicogenomics."

To date, most of the development of toxicogenomics methods has taken place in the pharmaceutical industry, driven by a desire to improve the screening and culling of drug candidates early in the research

---

<sup>5</sup> See, e.g., Oh SM, Park K, Chung KH, "Combination of *in vitro* bioassays encompassing different mechanisms to determine the endocrine-disrupting effects of river water." *Sci. Total Environ.* February 1, 2006;354(2-3):252-64. Epub March 19, 2005; or Araki N, Ohno K, Nakai M, Takeyoshi M, Iida M., "Screening for androgen receptor activities in 253 industrial chemicals by *in vitro* reporter gene assays using AR-EcoScreen cells." *Toxicol. In Vitro.* September 2005; 19(6):831-42.

and development process. The requirements for screening of drug candidates differ significantly from the requirements for screening environmental chemicals. These differences include: 1) a relatively selective focus on a few types of toxicity, especially hepatotoxicity, to eliminate unsuitable drug candidates; in contrast, environmental chemicals to which children and other susceptible populations may be exposed must be assessed more comprehensively, including for carcinogenic, developmental and reproductive effects; 2) drug candidates can be usefully screened out based just on acute or sub-acute toxicity; environmental chemicals must be screened for potential chronic toxicity; 3) false negative results in drug screening can be detected in later, rigorous pre-clinical testing regimens, while there is essentially no backstop for false negative results in environmental chemical screens. Thus, even with successful use of toxicogenomics methods to screen for toxicity by the pharmaceutical industry, substantial development of knowledge and data will be necessary before toxicogenomics methods can be safely and broadly applied to screen environmental chemicals. Early efforts to develop this capability are underway in various government and academic research centers around the world.

Despite these limitations, toxicogenomic data may be a very useful adjunct within an integrated assessment framework long before such data can be reliably used as the basis for regulatory decisions. As understanding of toxicity mechanisms continues to accrue, short-term toxicogenomic assays may be able to confirm or exclude the ability of specific chemicals or mixtures to act by certain mechanisms.

A second potentially useful near-term application of toxicogenomics assays could be in testing the validity of proposed chemical categories being increasingly employed in large-scale chemical screening initiatives such as the OECD HPV SIDS program and the US HPV Challenge, and expected under the European Union's emerging REACH initiative.<sup>6</sup> At present, there are relative few objective data used to justify most proposed category classifications. *In vivo* studies of acute toxicity and 28-day repeat dosing are the most commonly used traditional toxicology tests within the HPV program to support the validity of proposed categories, but results of these tests are often only available for a small subset of chemicals in the category. Submitting *all* members of a category to a short-term gene or protein expression assay, using either large numbers of markers or a specified subset, could allow more rigorous evaluation of proposed categories, and in particular provide a mechanism to identify potential outliers. For this to be feasible, reliable and inexpensive *in vitro* toxicogenomics assays will need to be further developed and validated. Such assays are likely to be limited in scope, with respect to dose range and time course. It must be recognized that assays with such limitations could not be expected to fully exclude the possibility of differential toxicity within a category, and variability due to technical or random biological factors would have to be accounted for. Toxicogenomic data would not be expected to be the sole basis for category validation, but rather would be considered along with other data as part of a weight-of-evidence approach.

### **Exposure information**

As documented in detail elsewhere,<sup>7</sup> we have raised concerns for some time within OECD about the serious limitations of available exposure information as a general matter, as well as the more specific

---

<sup>6</sup> We note and indicate our strong support for the project proposed by the OECD/IPCS Advisory Group on Toxicogenomics, entitled "Molecular Screening for Characterizing Individual Chemicals and Chemical Categories," which would undertake just such a study. See "Progress report on toxicogenomics: prepared for JM39, ENV/JM(2006)8, paragraphs 10-11, Work Item 1 and Annex 1.

<sup>7</sup> Denison, R., "Environmental Defense's perspective on policy issues related to exposure assessment," paper presented at the OECD Chemical Committee's Policy Dialogue on Exposure Assessment, held 6-7 June 2005, at OECD Headquarters, Paris; Denison, R. and Silbergeld, E., memorandum to the Task Force on Existing Chemicals for its 12<sup>th</sup> meeting, "Environmental Defense's concerns regarding use of exposure information in making recommendations regarding further work on HPV chemicals," 15 August 2003; Denison, R., addendum to memorandum to the Task Force on Existing Chemicals for its 12<sup>th</sup> meeting, "Statistics for all chemicals with SIAPs reviewed for SIAMs 15-16 and to be reviewed at SIAM 17 regarding use of exposure information in making

tendency toward over-reliance on extremely limited exposure information in OECD's hazard assessment activities. These concerns, briefly summarized here, apply equally to reliance on exposure information in applying more integrated approaches to testing and assessment.

#### *Key differences between assessing hazard and exposure*

Hazard is largely inherent to a substance, while exposure changes with place, use and time. This means that hazard (and hazard characterization or assessment) is relevant whatever the setting or use, while exposure is highly site/use-specific. Any exposure assessment is necessarily a "snapshot" of current exposure; the next new use or activity alters the picture. Exposure assessment must therefore be ongoing: scope, frequency of measurement must characterize *variation* in as well as *magnitude* of exposure.

Mechanisms for generating and evaluating hazard data are far more advanced and accepted than for exposure data. Extensive international-consensus standards exist for generating hazard data; they also address quality/reliability, interpretation, and reproducibility/verifiability. In contrast, standardized and routine collection of exposure data is rare and infrequent, and public access to such data is even rarer.

Differential access to both exposure data and the means to generate them can severely limit the "reproducibility" of such data. Most exposure data and the means to generate them reside virtually exclusively with industry. Industry's interest in claiming low exposure must be acknowledged, and means that having the ability to independently verify such information is essential. It must also be acknowledged that direct access to exposure "settings" is limited even for government officials. In addition, confidential business information (CBI) restrictions limit public access to exposure-relevant data; in contrast, hazard data are typically ineligible for CBI protection. Finally, supply-chain impediments to sharing exposure-relevant information abound, where for competitive reasons both suppliers and their customers have only limited access to information in the possession of the other party.

#### *Implications for integrated approaches that incorporate exposure information*

There is a critical need to develop international consensus guidelines governing the generation and use of exposure information, addressing:

- scope, completeness and quality;
- means of collection, analysis, QA/QC, verification, validation and reporting/ presentation transparency; and
- representativeness (accounting for both spatial and temporal variability).

Equally important is to ensure the capacity exists and is used to provide adequate expert review of any reliance on exposure information, to ensure that resulting conclusions or decisions:

- explicitly assess the information's scope, completeness and quality;
- sufficiently acknowledge limitations and the degree of uncertainty; and
- fully qualify conclusions.

Chemical assessment policies must acknowledge and directly address the variable nature of exposure. This means that exposure must be periodically reassessed to account for changes over time in production, use patterns. A corollary need is that requirements for the prompt reporting of such changes needs to be in place.

---

recommendations regarding further work on HPV chemicals," 26 September 2004; Denison, R., memorandum to the Task Force on Existing Chemicals for its 13<sup>th</sup> meeting, "Further Proposal Re Use of Exposure Information in SIAM Recommendations," 17 September 2004.

With respect to the differential access to exposure-related information, government officials need to be provided with authority and be able to demonstrate their ability to independently verify exposure data submitted by industry. Industry should itself commit to mechanisms such as third-party review and public release of all such data. Steps to de-bottleneck supply-chain flows of exposure-relevant information need to be instituted, by both industry and government. Finally, the allowed scope of CBI claims for such information should be as limited as possible.

Reliance even on reliable and complete exposure information does not preclude the need to develop a hazard characterization for a chemical, which has value independent of exposure and will virtually inevitably be needed as the exposure situation changes.

Slide 1

ENVIRONMENTAL DEFENSE

**Integrated Approaches to Chemical Testing and Assessment:  
Environmental Defense's Perspective**

Focus Session

39<sup>th</sup> Joint Meeting  
15-17 February 2006

ENVIRONMENTAL DEFENSE  
*finding the ways that work*

Slide 2

*Key objectives*

- Assuring full protection of human health and the environment
- Basing decisions on scientifically sound and defensible information
- Ensuring that information used in decisions is independently verifiable and reproducible
- Maximizing transparency in communicating the basis for decisions

Slide 3

*Guiding principles*

- *Avoid over-reliance*, by:
  - creating clear and sound guidance on appropriate / inappropriate uses
  - requiring justification, documentation of both use of alternative methods and decisions based on them
  - ensuring careful independent expert review

Slide 4

*Guiding principles*

- *Avoid selective use and reporting*  
(*“double standard”*)
  - Need for consistent application: can't deem QSAR results sufficient when they “exonerate” a chemical and inadequate when they don't
  - Should require all results derived using all methods employed to be reported to regulatory officials

Slide 5

*Guiding principles*

- Screening vs. other uses
  - In general, alternative methods – especially when used individually – are
    - most appropriate for priority-setting or screening-level assessments, and
    - less appropriate as the basis for more full-blown risk assessments or for risk management decisions.
  - Measure of confidence level needed

Slide 6

*Guiding principles*

- Transparency
  - Relying on information from multiple, diverse methods carries an added burden of transparency:
    - Nature, source and means of derivation of each data value needs to accompany it
    - Any conclusions or decisions based on such data need:
      - clear justification
      - explicit qualifications or limitations to communicate extent of uncertainty

## Slide 7

### *Guiding principles*

- *Continuing need to generate experimental data using in vivo testing*
  - Continued development of alternative methods depends on having a robust and expanding underlying dataset:
    - refinement of QSAR algorithms
    - correlations between *in vivo* results and those of other systems (*in vitro* testing, toxicogenomics)

## Slide 8

### Conclusion

- *Integrated approaches will only be as good as the “sum of their parts”*
  - Need clarity on appropriate / inappropriate uses of each method
  - Need to acknowledge / reflect limitations:
    - in choosing the uses to which such approaches are put, and
    - in justifying and documenting any resulting conclusion or decision

Slide 9

*See our paper for more specifics on:*

- Weight-of-evidence approaches
- (Quantitative) Structure-activity relationships [(Q)SARs]
- Read-across methods (using chemical category and analog approaches)
- *In vitro* tests
- Toxicogenomics (and related emerging technologies)
- Exposure information