Attachment 3. A Summary of Concerns regarding Specific Test Methods Proposed by EPA to be included in the EDSP Tier 1 Battery[1]

In selecting assays for the Tier 1 battery, the following criteria need to be considered: 1) the assays should clearly identify the endocrine mode of action activity under evaluation (estrogen, androgen and thyroid); 2) these assays should not utilize or minimally utilize apical endpoints that could be affected by numerous modes of action; 3) the assays must be validated and have the necessary level of specificity and sensitivity; 4) the assays should provide results that are clearly interpretable and if possible permit assignment of results as either positive or negative; 5) the screening assays should be relatively simple, rapid to perform and cost effective; and 6) the assays must provide sufficient sensitivity to identify substances with an endocrine mode of action, differentiate these from substances that produce toxicity via other modes of action and do so with acceptable false positive / false negative rates.

Many of EPA's proposed Tier 1 assays fall short in meeting the ICCVAM criteria[2] for test method validation. The American Chemistry Council provided extensive comments to EPA in March 2008 expressing concerns about the Tier 1 Test Battery and the validation status of each of the proposed Tier 1 test methods.[3] Specifically, the fish screening assay, the amphibian metamorphosis assay and the male and female pubertal assays all have significant validation issues:

- The pubertal assays have not been shown by EPA to possess the requisite degree of specificity needed to distinguish a response elicited secondary to systemic toxicity from a response caused by primary interaction with one or more components of the endocrine system.
- The amphibian metamorphosis assay has not been shown to be specific for endocrine modes of action. In fact, lab studies indicate that a purported "negative control" test article actually yielded a positive response and a "positive control" test article actually yielded a negative response.
- The EPA's proposed fish screening assay endpoint of fecundity is not specific for endocrine modes of action and is highly variable.

Both the peer reviewers and EPA's Science Advisory Panel (SAP) have expressed concerns regarding the relevance, reliability, sensitivity and specificity of these test methods. For example, the March 2008 SAP report[4] states:

> *"A negative control substance(s) has not been identified for the pubertal assays. This stands as a major limitation to the Tier 1 battery and more compounds should be tested."*

---

[1] Prepared by Richard A. Becker, Ph.D., DABT, American Chemistry Council, 5/18/09.

[2] Interagency Coordinating Committee on the Validation of Alternative Methods [ICCVAM]. 1997. *Validation and Regulatory Acceptance of Toxicological Test Methods.* "NIH Publication No 97-3981". Research Triangle Park, N.C.: National Institutes of Health,

[3] March 20, 2008 comments from, the American Chemistry Council, Docket EPA-HQ-OPP-2008-0012 FIFRA Scientific Advisory Panel Meeting to Consider and Review EPA's Endocrine Disruptor Screening Program (EDSP) Proposed Tier-1 Screening Battery.

[4] http://www.epa.gov/scipoly/sap/meetings/2008/march/minutes2008-03-25.pdf

"The Panel also recognized the significant degree of effort required to develop and validate the AMA for inclusion in the Tier I battery, as is true for all of the assays presented. However, there are limitations in the assay that bear mentioning. The sophisticated design of the flow through systems recommended by EPA may limit the transferability of the assay from lab to lab and the peer review of the assay also noted this concern. Furthermore, the sensitivity of the assay is currently unclear."

Redundancy within the battery is .essential to limit both false positives and false negatives. The question, however, is: what degree of redundancy is necessary to ensure confidence in the battery of assays? Given an initial evaluation of the preliminary data there may be a tendency lacks specificity. If the latter, it is assumed the battery will need to be refined to eliminate potential false positives likely due to toxicities other than EAT.

A good example of this may arise with the fish short-term reproduction screen. As demonstrated, both vitellogenin (VTG) and secondary sexual characteristics (SSC) endpoints for this assay are well accepted, both within EPA and the OECD. Both peer review and public comments questioned the interpretation of the fish fecundity endpoint. As described, the Panel was concerned over the variability of results and possibility of a false positive scoring based upon alterations of fecundity by mechanisms other than those involving EAT activities. EPA should be alert to possible non-endocrine mediated refinements of the fish short-term reproduction assay to ensure the reductions in fecundity are truly representative of EAT mechanisms and not generalized toxicity…."

"It bears mention that in vivo screens will be much more costly and more difficult to repeat with confidence than in vitro screens. False positive findings could result in a substantial burden of expense and interpretation. The regulatory process also faces mandates to "reduce, replace and refine" protocols that consume test animals, yet in vivo screens could be fraught with errors that lead to unfortunate repetition. The Panel recommended that substances not be administered near the maximum tolerated dose (MTD), that appropriate negative controls be identified and utilized in the screens, and that confident results from in vitro screens be given primacy when conducting "weight of evidence" assessment of a battery of results."

"An additional concern related to the handling of weak versus strong agonists and antagonists for each assay. As demonstrated, there is an inherent differential degree of sensitivities for each of the assays to weak agonists/antagonists. It was unclear to the Panel how these differences would be interpreted in relation to the "weight of evidence" approach for evaluating the assays."

Despite such comments and criticisms from the public, peer reviewers and the SAP concerning the lack of reliability and specificity of these test methods, EPA has still, more than a year later, not taken the requisite actions to address these serious deficiencies.