UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

March 4, 2011

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

Wendelyn Jones, Ph.D.
Senior Director, Human Health Policy
CropLife America
1156 15th Street NW, Suite 400
Washington. DC 20005

Dear Dr. Jones:

We are writing in response to CropLife America's February 12, 2010, submission to Dr. Steve Bradbury and Frank Sanders. CropLife America's correspondence, resubmitted in July 2010, contained several questions and requests for clarification regarding the Series 890 Test Guidelines. In October 2010, representatives from CropLife America, the Endocrine Policy Forum, and EPA met in Research Triangle Park, North Carolina to further discuss these issues.

Attached, please find the Agency's response to these questions. We will be including a copy of this letter and our responses to the questions in the docket for the Endocrine Disruptor Screening Program (EDSP) as well as posting them on the EDSP website.

Please let us know if you have any questions.

Sincerely yours.

Steven M. Knott

Steven M. Knott
Deputy Director
Office of Science Coordination and Policy

Richard P Keigwin, Jr

Richard P. Keigwin, Jr.
Director, Pesticide Re-evaluation Division
Office of Pesticide Programs

Corrections and clarifications
on technical aspects of the Test Guidelines for
the Endocrine Disruptor Screening Program Tier 1 Assays
(OCSPP Test Guideline Series 890)

March 3, 2011

Introduction

This document was prepared in response to questions raised by Test Order recipients and others about technical aspects of the conduct of Tier 1 assays for the Endocrine Disruptor Screening Program. In some cases, these questions pointed out simple but significant errors in test guidelines: for example, incorrect references to tables and misplaced decimal points. In other cases, questions pointed out ambiguous language that required clarification; and in yet other cases, questions were raised about whether alternative techniques had been considered by the Agency and rejected, or whether they might be acceptable to use. This document addresses all three categories.

Errors in the published test guidelines

Receptor Binding Assays

1. The equation for calculating K_d and B_{max} accounting for ligand depletion via the method of Swillens (1995) contains an error. The equation in Section j(6)(ii) (page 26) of the ER Binding Test Guideline and Section f(2)(i) (page 21) of the AR Binding Test Guideline) should read as follows:

$$Y = \frac{B_{max} * X}{X + K_d} + (\alpha * X)$$

where Y = total binding, α = the ratio between nonspecifically bound ligand and free ligand, and X = concentration of free radioligand.

2. An error has been pointed out in Section (j)(1)(iii) (pages 17 through 19) of the Estrogen Receptor Binding Test Guideline, concerning the procedure for adjusting the volume of radiolabeled estradiol added to compensate for radioactive decay. The section through step ❷ should read as follows. (The remaining steps in the section are correct as written in the Test Guideline.)

(iii) **Preparation of [^{3}H]-17β-estradiol.** Prepare on the day of the assay. Store [^{3}H]-17β-estradiol at -20 °C in the original container. [^{3}H]-17β-Estradiol is usually shipped from vendor in ethanol. Prepare dilutions of the [^{3}H]-17β-estradiol in TEDG + PMSF buffer to achieve the concentrations noted in column E of Table 5. Note that an adjustment for radioactive decay (i.e., decline in specific activity and specific concentration) should be included. Siliconized or silanized glass tubes should be used when preparing serial dilutions.

To calculate the amount of stock [^{3}H]-17β-estradiol to add to buffer to make the stock dilutions (Column E) necessary for the final concentration in Column F:

❶ Convert the specific activity from Ci/mmole to nM. The manufacturer usually packages a specific concentration of Ci/ml and will give this information on the package (for example, often 1.0 mCi/ml in ethanol). If SA = X Ci/mmole, and Y = specific concentration of radiolabel, then X Ci/mmole is converted to nM by the following conversion:

(Y mCi/ml / X Ci/mmole) * 1 Ci/1000 mCi * 10^6 nmole/mmole * 1000 ml/L = (Y/X) * 10^6 nM

❷ Prepare a primary stock in TEDG + PMSF buffer. For example, since the highest concentration in Column E is 30 nM, a stock concentration that is 300 nM would be appropriate.

In this example, one ml was chosen as the amount of stock solution to prepare. A different volume could have been chosen.

How many µl of radioligand at (Y/X) *10^6 nM stock concentration will equal 300 nM in 1 ml? Use the equation:

$$Z\ \mu l * ((Y/X) * 10^6\ nM * FIR) = 1000\ \mu l\ (300\ nM).$$

$$\text{Therefore, } Z\ \mu l = 1000\ \mu l\ (300\ nM) / ((Y/X) * 10^6\ nM * FIR)$$

where

- FIR = Fraction of Isotope Remaining = $SA_{adjusted}$ / SA
- $SA_{adjusted}$ = Adjusted Specific Activity = $SA*e^{-Kdecay*Time}$
- SA = Specific Activity provided by the supplier
- K_{decay} = the decay constant for tritium = $1.54x10^{-4}$ /day

For example, if Y=1.0 mCi/ml, the specific activity is X=140 Ci/mmole, and the fraction of isotope remaining is 0.95, then Z=44. 2 µl [^{3}H]-17β-estradiol plus sufficient TEDG + PMSF buffer to bring to 1 ml will yield 300 nM [^{3}H]-17β-estradiol.

(Dilution calculations can be double-checked on the "QuickCalcs" webpage from GraphPad: http://www.graphpad.com/quickcalcs/ChemMenu.cfm)

3. In the AR binding Test Guideline section (c)(4)(ii), the volume of 200 mM EDTA should be 750 µl, not 7.50 µl.

4. In the AR binding Test Guideline, the unit for the cold initial concentration in the header of Table 3 (page 10) should be µM, not nM.

5. In the AR binding Test Guideline, the concentration of triamcinolone acetonide shown in section (e)(2)(third bullet) should be "(60 µM working solution)", not "(60 mM stock)".

6. In the AR Binding Test Guideline, potassium chloride may be taken out of the list of equipment and materials (section (c)(3)(ii)) and the list of stock preparations (section (c)(4)(i)).

Aromatase Assay

1. The example for preparation of substrate solution (section (e)(3)(iii), 3rd bullet on page 4) should read "2.6 mL buffer" rather than "2.7 mL buffer".

2. The parenthetical remark in the last sentence of section (e)(5) on page 6 should read "(0.1 – 10,000 nM; Table 3)". That is, the correct lowest concentration for the positive control substance is 0.1 nM, not 0.01 nM; and the reference should be to Table 3, not Table 4.

3. The word "soted" in the second sentence of section (e)(6)(i) on page 6 should be "stored".

4. In section (g) on page 8, the first parenthetical statement should read "(as shown in Table 3)", not "(as shown in Table 5)".

5. In section (h) on page 9, the first paragraph should refer to "(Table 4)", not "(Table 2)".

6. The second bullet in section (i)(1) on page 11 should read "The recommended mean background control activity is ≤15% of the full activity control." (not "≤10%").

Policy considerations inferred by the Agency from technical questions

Performance criteria

It was inferred from certain specific technical questions that Test Order recipients are concerned about how strictly the performance criteria included in each Guideline will be applied. That is, particularly for assays with numerous endpoints will the Agency reject studies that do not meet every performance criterion?
In general, the Agency does not intend to apply the performance criteria rigidly. The performance criteria are usually aimed at providing confidence that an assay is sensitive to weakly active chemicals. A study may be judged to be adequately sensitive even if one endpoint did not meet its performance criterion, particularly if there is another endpoint in the assay that provides similar information. In some cases the Agency may give less weight to an endpoint or study that has missed a performance criterion rather than reject the entire study. How much weight an endpoint or study carries in a weight-of-evidence determination of interaction with the endocrine system is likely to depend on many factors. It should be recognized, however, that the Agency may require that an assay be redone if the performance criteria are not met.

Deviations from test guidelines

Also of significant concern is the degree to which deviations from a Test Guideline will be the basis for disqualification of a study. Certain items specified in a Test Guideline may be difficult for laboratories to meet without major changes in facilities or training, and if such changes are not critical to the conduct of the study it would be helpful for the laboratories to know that.

The Agency is unable to provide general guidance on this matter as this requires case-by-case consideration. As can be seen from the remainder of this document, certain details are considered critical to ensure the scientific validity of the resulting data. For other aspects, acceptable alternatives may exist that were not assessed as part of the validation process; however, because these were not included as part of the validation of the Test Guidelines, EPA cannot definitively conclude whether they will affect the validity of the study without reviewing the actual data.. The Agency is providing explanations in the responses below to help laboratories understand the Agency's concerns behind specific items and recommendations in the Test Guidelines. Where laboratories choose to deviate from the Guidelines, EPA recommends that laboratories describe the deviation, the reason for the deviation, and demonstrate that the deviation does not alter the reliability or sensitivity of the assay (including scientific references).

Clarifications

Estrogen Receptor (ER) and Androgen Receptor (AR) Binding Assays

1. Q: May other data analysis packages besides GraphPad Prism be used?

   A: Yes. It is the analysis not the software that matters. Prism was cited as an example for laboratories that may have minimal experience with receptor binding data and may therefore not be aware of relevant software packages.

2. Q: Can the plateau for the IC_{50} calculation be constrained to represent 100% binding?

   A: EPA did not intend to preclude laboratories from conducting analyses in addition to the ones specified in the Test Guidelines may be included. If such additional analyses are included, in order for EPA to evaluate their adequacy EPA expects that laboratories would explain why the alternative analyses are thought to be more appropriate. In general, constraining the top, bottom, slope, or IC_{50} is unlikely to be considered appropriate because it may mask problems with the conduct of the assay or may incorrectly characterize the competitive characteristics of the test chemical.

3. Q: How many saturation runs should be made per cytosol batch for the ER and AR assays?

   A: EPA recommends that three adequate saturation binding runs, each containing three replicates at each concentration, be made initially to characterize a cytosol batch, for both the ER and AR binding assays. In EPA's experience during interlaboratory validation, the variability seen in some laboratories both within and between runs suggests that such multiple runs are required to ensure that the laboratory is producing reliable results. Additional saturation binding runs may be necessary to characterize a particular batch of cytosol later if there are any questions about the integrity of the receptor due to handling, storage, or age.

4. Q: In the AR binding assay, can the prostates be frozen for later cytosol preparation?

   A: Just as uteri may be immediately frozen for later preparation of cytosol for the ER binding assay, prostate may be immediately and rapidly frozen in liquid nitrogen for storage at -80°C for no more than three months without jeopardizing the validity of the experiment. (See also the response to Question 10 below.) EPA does not recommend long-term frozen storage because the integrity of the receptor over a long duration is unknown.

5. Q: Can the same approach for combining the results of several competitive binding assay runs into a summary of the potential of a test chemical to interact with the receptor that is outlined in the ER binding assay be used for the AR binding assay?

   A: The classification method described in section (k)(7)(iv) (pages 48 and 49) of the ER binding Test Guideline would be equally applicable to the AR binding assay as well. Note that classification is a convenient method for summarizing results but will not substitute for consideration of all relevant information from a binding assay when considering the weight of evidence of the potential for a test chemical to interact with the endocrine system, because the summary may not adequately reflect information about the conduct of the assay such as variability, number and placement of unusable data points, and assumptions used in analysis that are important to the evaluation.

6. Q: In the AR binding assay (and also in the ER binding assay), must the specific activity of the tritiated reference standard be adjusted for decay?

   A: Radioactive decay of tritium is expected to be negligible in most cases due to the long half-life of tritium but if it is necessary to use tritiated R1881 (in the AR binding assay) or tritiated estradiol (in the ER binding assay) that is older than one year, the radiochemical should be reanalyzed for purity per the recommendations of the vendor, and the specific activity should be adjusted for decay. The procedure described in the ER binding assay Test Guideline, as corrected in this document (see the section on "Errors in the published test guidelines"), would be scientifically acceptable for adjusting the specific activity.

7. Q: Can recombinant androgen receptor be used in the AR binding assay?

   A: The conditions described in the Test Guideline are specific to androgen receptor in rat prostate cytosol. A separate protocol optimized for recombinant androgen receptors is being developed and validated but is not available at this time.

8. Q: May other sources of estrogen receptors besides rat uterine cytosol be used to determine ER binding ability?

   A: EPA is participating in an international effort to validate an assay using human recombinant estrogen receptor (hrER). The conditions are substantially different from the conditions of the assay using rat uterine cytosol. Until an hrER binding assay has been validated, rat uterine cytosol is the only recommended source of ER.

9. Q: Is the specific activity of 70-87 Ci/mmol for [^{3}H]-R1881 that is listed in section (c)(3)(ii) of the AR binding Test Guideline a requirement?

A: Specific activity lower than the range shown will raise concerns regarding the validity of the data, although small deviations may be acceptable. It is necessary to have sufficient radioactivity counts in order to detect changes in binding to the AR in the presence of test chemical. The assay was validated for the range of specific activities shown and consequently, EPA recommends that laboratories explain why a value outside this range, if used, should be considered acceptable.

10. Q: Can cytosol be used past 90 days?

A: Laboratories are responsible for ensuring that the receptor is handled appropriately at all times and does not degrade prior to use in the binding assay. To this end, it is appropriate to monitor the binding data for the positive and negative controls over time and determine if there is a decrease in binding efficiency that can be attributed to degradation of the receptor. In the Agency's experience, aliquots of uterine cytosol that are kept frozen (-80 °C) and otherwise handled appropriately do not degrade significantly over a 90-day period but a decline in maximal binding was noted at 127 days. For this reason, if it is necessary to use cytosol that is more than 90 days old, it is recommended that a saturation binding assay be conducted to check the K_d and B_{max} of the receptor to ensure that the receptor is performing as expected.

11. Q: Isn't the lowest concentration of estradiol (0.03 nM) in the ER saturation binding assay too high to allow proper characterization of the K_d, which may be as low as 0.03 nM?

A: Laboratories may attempt to prepare a more-dilute concentration of estradiol. It is unlikely, however, that reliable binding values can be gathered when the radioligand concentration is much lower than 0.03 nM for two reasons: (1) the specific-bound radioactivity count (dpms) at 0.03 nM is already quite low due to the low specific activity of tritiated radioligand; and (2) ligand depletion becomes more of an issue as the concentration of radioligand is lowered. There is a lower limit to the protein concentration (ER) that is recommended for the assay, to ensure the integrity of the HAP-pellet during the separation of bound from free radioligand. At this concentration of protein, ligand depletion is at the limit of acceptability and a lower concentration of radiolabeled estradiol would have even greater % of ligand depletion.

12. Q: May a different cocktail of protease inhibitors be used?

A: If a laboratory demonstrates that there are no adverse impacts on results of the assay, protease inhibitors in addition to the EDTA, DTT, and PMSF already specified in the Test Guidelines would be scientifically acceptable. In general, EPA has not found it necessary to add additional protease inhibitors as long as the cytosol is handled properly (kept cold on ice during preparation, etc.). Given the expense and potential toxicity to laboratory personnel of some of the other protease inhibitors, EPA chose not to recommend use of additional inhibitors.

Estrogen Receptor Transcriptional Activation (ERTA) Assay

1. Q: Can the criteria for a positive response be harmonized between the ER binding assay and the ERTA assays?

   A: No, they are fundamentally different assays. One measures competitive binding; the other measures a downstream consequence of binding. The response scales are not the same and they may in fact give somewhat different results for a given chemical.

2. Q: Can the validity criterion for minimum induction from exposure to 1 nM of estradiol be changed from 4-fold to 5-fold induction?

   A: EPA cannot unilaterally modify the validity criterion. The ERTA test guideline is an Organisation for Economic Cooperation and Development (OECD) test guideline and modifications can be made only by OECD. Unilateral modifications run the risk of rejection of data under the Mutual Acceptance of Data Treaty. The values in the test guideline were developed by CERI during the validation program. Requests to modify the test guideline should be addressed to the OECD.

Steroidogenesis Assay

1. Q: Can the Minimal Basal Production (MBP) be lower than 40 pg/ml?

   A: To be consistent with the Test Guideline, if the MBP for estradiol is lower than 40 pg/ml, the assay would likely need to meet the other performance criteria specified in the guideline in order to be considered adequately sensitive. In this case, basal E2 production that is consistently 2.5 times the minimum detection limit (MDL) would be considered as meeting the criteria (Table 5, page 24). It should be noted that when plating the cells, it is very important to add fresh supplemented medium at both an initial and 24-hour period in order to maximize MBP. In addition, care should be taken to avoid plastic and glassware that allow estradiol to adhere to the surface since such adherence can lower the MBP measurement. EPA recommends using low-retention (e.g., siliconized/silanized) tubes to achieve this.

2. Q: Is a LIVE/DEAD '-cidal' assay the best measure of cytotoxicity in the steroidogenesis assay?

   A: The Live/Dead® Cell Viability/Cytotoxicity Kit is recommended in the guideline as a method for evaluating cytotoxicity, but the choice of a cytotoxicity assay is left to the performing laboratory. The MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide] test has been shown to work equally well.

3. Q: Is it necessary to perform an extraction on samples to be analyzed by the ELISA or RIA method?

   A: Whether or not an extraction is performed depends on the assay system that is in use. If a testing laboratory can demonstrate that the assay functions without extraction, there is no need for extraction.

4. Q: Can a solvent other than methanol be used if not using the live/dead assay?

   A: Methanol is only used in the quality control (QC) plate (Table 4, page 24) as a positive control for cytotoxicity as recommended by the manufacturer of the live/dead assay. Any departure from methanol would require documentation that the alternative was providing 100% cell death and is suitable for use with the cytotoxicity test of choice.

5. Q: The Test Guideline mentions use of 24-well plates. May other sized plates be used?

   A: As noted in the OECD draft test guideline, 48-well plates may be used:

   > "The assay is usually performed under standard cell culture conditions in 24-well culture plates. Alternatively, 48-well plates

can be used for conducting the assay; however, seeding and experimental conditions should be adjusted accordingly.

6. Q: For the laboratory proficiency test (section (g)(2)(iii), page 22 of the Test Guideline), it appears that dosing is to be accomplished by placing the dose (in DMSO) directly into the well. Would it be acceptable to place the dose (in DMSO) into medium first, and then exchange the blank medium with dosed medium?
   A: The alternative procedure is analogous to the one described in section (h)(2)(ii) Step 3 (pages 28-29) for other test chemicals.

Aromatase Assay

1. Q: Is a supplier's statement of radiochemical purity acceptable for documenting the [^{3}H]-androstenedione purity requirement in section (e)(3)(ii)?

   A: The laboratory is responsible for ensuring that the radiolabeled androstenedione is sufficiently pure for use in the assay. In cases where chain of custody documentation ensures that the material has been handled and stored appropriately and the material has not aged significantly, it can be acceptable to trust the supplier's statement of purity. However, the Agency encourages purity to be checked experimentally, particularly if there is any reason to believe that degradation may have occurred, or at intervals recommended by the vendor.

2. Q: May other methods than what is described in section (e)(3)(ii) be used to analyze the purity of the [^{3}H]-androstenedione?

   A: High performance liquid chromatography and liquid scintillation counting is the preferred method of analysis of purity. It may be appropriate in some cases to use a different column and/or mobile phase from what is specified in the Test Guideline if there is reason to believe another system would be better. EPA recommends that information supporting the choice of an alternative separation system be included in order to allow appropriate evaluation.

3. Q: Can human placental microsomes be used in place of human recombinant microsomes?

   A: Yes, but the Agency prefers that recombinant microsomes be used. The Agency validated the aromatase assay using human placental microsomes as well as human recombinant microsomes as source of the enzyme. However, reviewers have noted that human placental microsomes may be less consistent than recombinant microsomes due to variations between individuals from whom the placental tissue was obtained. There are also other concerns such as safety issues associated with using human placental tissue. If human placental microsomes are used as the source of enzyme, it is important to recognize that the optimized assay conditions are different from those that apply for recombinant microsomes. Appropriate guidelines and performance criteria that are specific to use of human placental microsomes are provided in Chapter 4 of the Integrated Summary Report: Aromatase (http://www.epa.gov/endo/pubs/aromatase_isr.pdf). The Agency has not validated any cell lines for use in this assay.

4. Q: Is it possible to use the original extraction method using chloroform/charcoal as opposed to methylene chloride?

A: It would be appropriate to demonstrate that any alternative extraction method used, such as the chloroform/charcoal method, is equivalent to the methylene chloride method, using a range of chemicals to include at a minimum the following chemicals: OH-androstenedione (positive control) and proficiency chemicals (Table 6). Full concentration curves should be provided for each chemical with 3 runs and triplicate tubes at each concentration.

5. Q: Must each new technician test 4 proficiency chemicals over three test runs?

   A: The Test Guidelines recommend that each new technician demonstrate proficiency in conducting the assay. Performing multiple runs allows evaluation of the ability to reproduce results and provides a measure of variability between runs.

6. Q: What is the highest concentration of proficiency chemicals that must be tested?

   A: Proficiency chemicals (Test Guideline section (i)) should be tested at the same concentrations as described for other test chemicals (section (j) and Table 7). Thus the highest concentration tested should be 10^{-3} M unless the chemical is insoluble at that concentration. The Guidelines include directions in section (j) for determining the highest concentration that is soluble if 10^{-3} M cannot be reached.

7. Q: If a laboratory's performance criteria for the positive control are just outside the recommended range for one of the four parameters associated with the positive control (*i.e.* Top, Bottom, log IC_{50}, Hill slope), but the positive control curve is a good fit with the model, will these data still be accepted?

   A: The performance criteria for these four parameters will not be applied as rigid boundaries, but will be evaluated on a case-by-case basis. For example, if results show clearly that a test chemical interacts with the aromatase enzyme but the performance criterion for log IC_{50} for the positive control is slightly out of range, the study is unlikely to be rejected. Performance criteria become more important when a test chemical presents equivocal or variable data but even in such cases EPA will evaluate the performance criteria, taking into account the individual circumstances.

8. Q: Will a range for the protein standard curve be accepted other than the range of 0.13 to 1.5 mg protein/mL listed in section (f)(1) on page 7?

   A: The Guidelines were not intended to preclude the use of an alternative range, but in order to provide the most accurate determination of the actual aromatase concentration the range should be such that the expected aromatase concentration is approximately centered within the range of the BSA standard curve, the range spans no more than approximately 2 orders of magnitude, and the range is covered by 6 concentrations.

9. Q: Isn't use of a multi-tube vortex after adding the microsomes in section (h) (third bullet on page 10) too vigorous?

   A: The current instruction says to "Initiate the assay by adding 1 mL recombinant microsomal preparation to each assay tube and vortex gently using a multi-tube vortex." If there is any concern that the multi-tube vortex may shake too vigorously and possibly denature proteins, an alternative would be to place the tubes in a gently shaking warm water bath.

Hershberger Bioassay

1. Q: The guideline provides maximum allowable coefficients of variation (CVs) for organ weights. Are these CVs appropriate for treated animals considering variability in organ weight?

   A: The maximum CVs are appropriate for this test guideline and were determined from the performance of the laboratories in the validation studies. Exceeding the CVs is cause for concern and indicates that the sensitivity of the assay may be compromised. But EPA would consider other factors, such as EPA's interpretation of the study design and results, before rejecting a study on this basis. Additionally, it may be helpful to provide the Agency with historical control data for measures that demonstrate high variability.

2. Q: Cowper's glands are tiny in castrated rats and also surrounded by adipose tissue/connective tissue which can look like this gland. Is it possible to ensure that collected tissue is really Cowper's glands?

   A: One can look at the glands under a dissecting microscope to better determine whether the collected tissue is Cowper's glands.

3. Q: What is the rationale for including the MANOVA analysis in the Hershberger assay when ANOVA analyses with 2 significant organ weights provides a clear criterion for a positive result?

   A: In general, EPA agrees that ANOVA analyses with two significant organ weights would usually be sufficient. MANOVA provides additional information that may be valuable in the weight-of-evidence decision about the chemical. Note that a study in which MANOVA is not provided may not be considered sufficient for OECD purposes.

4. Q: Is it necessary to conduct the androgenic portion of the Hershberger assay as well as the anti-androgenic portion?

   A. Yes, the Hershberger assay consists of both the test for androgenicity and the test for anti-androgenicity.

Uterotrophic Bioassay

1. Q: Does anti-estrogenic activity need to be characterized in the uterotrophic bioassay?

   A: The uterotrophic assay to detect anti-estrogenic activity is not required by the test order and has not been validated. If one chooses to perform this assay, there is an OECD guidance document that outlines a method for doing so.

2. There have been recent reports of anomalous rapid development, prolonged estrus cycles, and disrupted fertility in rats from a certain supplier. Does EPA have guidance on how laboratories should deal with this situation?

   A: If a laboratory believes that its results have been influenced by changing patterns in reproduction from a particular animal source/breeder, EPA recommends that the laboratory provide the Agency with historical control data for the strain, source, age and timeframe to facilitate data interpretation.

Male and Female Pubertal Assays

1. Q: Is it acceptable to conduct a clinical chemistry assessment as part of the probe study? Would it also be required as part of the definitive study?

   A: The clinical chemistry assessment is intended for use in establishing whether the Maximum Tolerated Dose (MTD) has been reached or exceeded, and as such would probably be appropriate to conduct in a probe study. Such clinical chemistry assessments are not particularly useful in the pubertal assay if MTD has been defined by a different endpoint. However, if a clinical chemistry endpoint is used to define an MTD from a probe study it would be appropriate to conduct the clinical chemistry assessment in the pubertal assay itself.

2. Q: Must laboratories use pregnant dams or may pups of a specified age be ordered from suppliers instead?

   A: There are several reasons that the Agency strongly recommends that pregnant dams rather than pups of a specified age be used as the starting point for the pubertal assays. First, it is important to minimize genetic effects (that is, litter effects) on the endpoints of interest where possible. It is important to know which pups are from the same dam in order to be able to randomize distribution of siblings across treatment groups. Second, it is important to cull to 8 to 10 pups per litter within 3 or 4 days after birth. This helps to minimize variability in body weights across a litter, and thus minimize variability in day at vo/pps and in other endpoints such as organ weights that may be related to body weight. Third, it is important to know the day of birth of each pup accurately. Knowing the day of birth accurately is critical for accurate determination of age at pps/vo and keeping the coefficients of variation (CVs) low. Keeping the CVs as low as possible is important for maximizing sensitivity of the pubertal endpoints.

   Thus, if pups of a specified age are ordered rather than pregnant dams, EPA recommends that submitters document that all of the necessary steps have still been taken to allow randomization of litters across treatment groups, that standardization of litters (including exclusion of litters with fewer than 8 total pups per litter as well as litters not delivered by GD 23, reduction of litter size to 8-10 pups per litter between post-natal days 3 and 5 and not allowing cross-fostering) has been done.

   It should be noted that several peer reviewers of the pubertal interlaboratory validation studies recommended strongly that only animals bred in-house be used in the pubertal assays, to avoid any stress associated with shipping. The Agency has shown that use of timed pregnant animals still allows endocrine-active compounds to be identified correctly. It has no such assurances when pups of a specified age are received from suppliers.

3. Q: Is it necessary to use a low phytoestrogen diet?

A: As stated in the Test Guidelines, the genistein-equivalent content of genistein plus daidzein (aglycone forms) of each batch should be below 350 µg/g and preferably below 300 µg/g. Diets marketed as low phytoestrogen diets are likely to be below this level, but specific batches of other diets may also be acceptable. Although there is still considerable uncertainty over the possible effects of phytoestrogens in feed on the endpoints in the pubertal assays, concerns have been raised about the potential for compromising the study (e.g., resulting in false negatives) if the phytoestrogen level is allowed to be too high.

4. Q: Is only deionized water acceptable for consumption?

   A: Deionized water was specified because of concerns by the Endocrine Disruptor Methods Validation Subcommittee at its Dec. 10, 2001 meeting that tap water is too variable: it may contain potentially endocrine active substances such as disinfectant byproducts and perchlorate. A member of the Subcommittee noted that use of deionized water would help reduce such concerns. There was also, however, recognition that it would not be possible to standardize water across all pubertal studies. Other acceptable sources of water include double-distilled water and charcoal-filtered water. Other sources may also be acceptable. However, the presence of soluble organic chemical contaminants such as natural and artificial hormones have the potential to introduce variability into (and potentially compromise) the results (e.g, result in false negatives or false positives). Consequently, if an alternative source of water has been used, EPA recommends that the laboratory document that such contaminants have been removed from the drinking water..

5. Q: Are heat-treated laboratory-grade wood shavings required for bedding?

   A: EPA has not tested all potential bedding materials and cannot state which products will and which products will not interfere with the assay. Cedar shavings are not recommended due to their potential to affect liver function, and corn cob bedding is not recommended due to the potential for influence on the endocrine system as footnoted in the Test Guidelines. Heat-treated laboratory-grade wood shaving has been successfully used in pubertal assays. If other bedding materials are substituted, the EPA recommends that data showing that such materials do not interfere with endocrine assays be provided as support.

6. Q: Is a 14:10 light:dark cycle required?

   A:The 14:10 light:dark is the traditional lighting cycle used for endocrinology studies in female rats and is preferred. However, a pubertal study is unlikely to be deemed scientifically invalid with regard to lighting conditions as long as the continuous light-on duration is between 12 and 14 hours inclusive and is consistent throughout the study.

7. Q: If a method for blood collection other than from the decapitated trunk is used, (e.g., cardiac puncture, vena cava collection, etc.), are other anesthetics permitted in conjunction with these procedures?

   A: While EPA still prefers decapitation without the use of anesthetics as a humane method of kill and little interference with hormone levels if performed correctly, it has previously determined that the use of anesthetic (injectable or inhalational) and exsanguination via other methods would not necessarily be a basis on which to reject a study, subject to the following caveats:
   - Dose levels of anesthetic must be such that the majority of animals reach deep anesthesia within 2 minutes. For animals not reaching deep anesthesia within 2 minutes, either decapitate immediately or record the time until deep anesthesia is achieved and mark the animal as a deviation. Examine whether the additional time resulted in differences in hormone levels, and either use or exclude the information, as appropriate, for further analyses.
   - The amount of blood collected via the method chosen is sufficient to perform the necessary hormonal and blood chemistry work. Use of a method that frequently does not yield sufficient blood for the necessary analyses is not likely to be acceptable.

   The main concern with use of anesthetic or asphyxiant is the induction of stress, which may affect hormone levels within a short period of time. Use of injectable anesthetic is preferred due to better delivery control and thus potentially shorter times to induce deep anesthesia than typically occurs with inhalational anesthetics.

8. Q Do laboratories need to adopt the methodology of Smith et al. (1991) to select five sections per ovary for evaluation?

   A: The Smith et al. (1991) paper was referenced to support the use of appropriately prepared random sections as opposed to serial sections. The follicle counts described in that paper are not part of the pubertal assay. Other methodologies for obtaining a representative distribution of sections are acceptable but should be described.

9. Q: Is it necessary to analyze data for the age at puberty onset and organ weights by three different statistical methods – ANOVA, ANCOVA and trend analyses?

   A: EPA validated the assay using a methodology in which the data for age at puberty onset and organ weights are to be analyzed by Analysis of Covariance (ANCOVA) using initial weight as the covariate. For this reason, EPA recommends that laboratories follow this methodology. EPA also recommends that, following this, the laboratory examine the individual group means with pairwise t-tests against the control group if the ANCOVA was significant, or Dunnett's or trend test if the ANCOVA was not significant. This is standard statistical practice for determining differences between treated groups and controls. Where laboratories choose to deviate from the Guidelines, EPA recommends that laboratories describe the deviation, the reason for the deviation, and generally demonstrate that the deviation

does not substantially alter the sensitivity of the analysis (including scientific references). However, because different analyses were not included as part of the validation of the Test Guidelines, EPA cannot determine whether they will affect the validity of the study without reviewing the actual data.

10. Q: Do laboratories need to adopt the 5-point thyroid histopathology scoring system?

A. The 5-point thyroid histopathology scoring system with photomicrographs as examples is a significant contributor to the confidence that can be placed in the sensitivity of the histopathology results. It was used in the interlaboratory validation studies and is a critical component of the assay that EPA validated. EPA therefore cannot determine whether use of a different scoring system would affect the validity of the study without reviewing the actual data. Questions may arise about the sensitivity of the scoring method if a system using fewer gradations is used, or if the pathologist is unaware of the photomicrographs that have been provided in the pubertal Test Guidelines as examples of the subtle changes that can consistently be scored. Consequently, where laboratories choose to deviate from the Guidelines, EPA strongly recommends that laboratories describe the deviation, the reason for the deviation, and generally demonstrate that the deviation does not alter the reliability or sensitivity of the assay (including scientific references). Note that the term "5-point system" includes both ends of the scale as "points" (i.e., both the apparently unaffected end and the maximally affected end), while terms used for scoring systems in other contexts may include only one of the two ends as a "point" (i.e., the maximally affected end). The "5-point system" described in the pubertal Test Guidelines corresponds to a "4-point system" in that terminology.

11. Q: Will studies be rejected because the CVs from the control animals fall outside of the prescribed ranges?

A: The performance criteria for coefficients of variation from control animals are important to achieve in order to have confidence that the assay is sensitive. EPA does not intend to apply the performance criteria rigidly, however. The relationship of an observed CV to its corresponding performance criterion is likely to be only one factor among many in the weight-of-evidence determination of whether a substance interacts with the endocrine system. The CVs for VO and PPS are particularly important and EPA is likely to closely examine the justifications for any deviations in the CVs for the control animals for these endpoints, to determine the extent to which they may have impacted the validity of the results. It may be helpful to provide the Agency with historical control data for measures that demonstrate high variability, but past performance will not necessarily justify deviations from the performance criteria.

12. Q: Is it necessary to report the coefficients of variation from the treated animals?

A: The coefficients of variation for endpoints measured in the treated animals are important indicators of whether there is abnormal variability in a particular group.

Such abnormalities can be important in interpreting the results of a study during the weight-of-evidence determination of the ability of the chemical to interact with the endocrine system. It may be helpful to provide the Agency with historical control data for measures that demonstrate high variability.

13. Q: Do laboratories need to check cages for new births in the morning?

    A: The EPA strongly recommends that cages be checked for new births in the morning. This is because dams may begin to deliver late in the day but not complete delivery until after lights-out. If animals from the beginning of the litter are assigned PND 0 on one day but the remainder of that same litter is assigned PND 0 on the next day, there will be variability in the age at PPS/VO that could have been avoided by proper assignment of day of birth. The variability would decrease the sensitivity of the endpoint.

14. Q: Is it necessary to mark pups on PND 21, then weigh and randomize them on the same day?

    A: Marking does not need to occur on the same day as weighing and randomizing. Marking done prior to PND 21 is not expected to affect the endocrine endpoints in the study. Because randomization is weight-ranking-dependent and weight-ranking can change with time, EPA strongly recommends that randomizing be done on the same day as weight-ranking. These steps may occur on either PND 21 or 22 for the males since treatment does not begin until PND 23, but would need to occur on PND 21 for the females in order for the females to begin treatment on PND 22 as called for in the guideline.

15. Q: Can an intermediate solvent (e.g., ethanol) be used to assist in solubilization if needed?

    A: Use of an intermediate solvent would not be expected to affect the validity of the data significantly, provided the concentration is kept at low concentrations (e.g., 1% or below) and is used across all test groups including control. Use of higher concentrations is likely to cause EPA to question whether the solvent itself interfered with the results. EPA recommends that the study design include the use of a solvent or vehicle control group.

16. Q. Is it necessary to administer chemical doses in 2.5-5 ml vehicle/kg body weight?

    A: EPA prefers that 5 ml/kg not be exceeded, in part to ensure that stress from an excess of volume does not interfere with the study. Studies in 7-8-week-old rats suggest that gavage dosing volume should not exceed 10 ml/kg based on corticosterone levels as indicators of stress (Brown AP, Dinger N, Levine BS. 2000. Stress produced by gavage administration in the rat. Contemp Top Lab Anim Sci 39(1):17-21). Staying well below this volume per kilogram is prudent since the

effects of such stressors on the endocrine endpoints being measured in the pubertal assays have not been investigated, and since it is not known whether tolerance to dosing volume is dependent on age or size in a developing rat. (The animals in the pubertal studies begin dosing when they are significantly younger and smaller than the animals studied.) Caloric effects when corn oil is used may also be a concern. EPA recommends that the study design include a solvent or vehicle control group that receives the same volume of vehicle per kilogram of body weight as the treated groups.

17. Q: Is it necessary to use the gavage needle types and sizes identified in the test guidelines?

    A: As noted in the Guidelines, the "[n]eedle gauge may be optimized to animal size but must be constructed of metal to avoid the potential for absorption by or leaching of substances from rubber or plastic tubing."

18. Q: Can dose time vary from the 0700 and 0900 specified in the test guidelines?

    A: It is important that doses be administered at approximately the same time each day, and on such a schedule that kills begin no sooner than 2 hours following the dosing on the last day and all kills are completed no later than 1:00 p.m. This is to ensure that circadian rhythms that are associated with some of the endpoints are minimized as sources of variation in the results. Variation interferes with the sensitivity of such endpoints.

19. Q: Is it necessary for laboratories to begin evaluating animals for puberty onset as early as PND 22 and PND 30 as specified in the female and male pubertal test guidelines, respectively?

    A: It is critical that, for each animal, the VO/PPS observation be recorded for the day immediately prior to the day on which vaginal opening or preputial separation begins. Labs may choose not to begin monitoring on PND 22 (for VO) or PND 30 (for PPS), but missing the day on which VO or PPS begins for each animal will be considered a serious deficiency in the study because the sensitivity of this endpoint is dependent on the accurate determination of the day of VO/PPS (or in certain cases, day of initiation of VO/PPS if the process is not complete in one day). If the day of VO/PPS is expected to be different for control animals from what is noted in the guideline, EPA recommends that appropriate documentation to support a modification in study design to begin observations later than recommended be provided to the Agency. Note that consideration should be given to the possibility that the test chemical may accelerate VO/PPS; observations for these endpoints should begin substantially before the age at which control animals are expected to reach these endpoints. The objective of the VO/PPS endpoint is to determine quantitatively the difference, if any, in age at VO/PPS between treated groups and controls, not merely to determine that an acceleration or delay has occurred.

20. Q: Is it necessary to record the initiation of puberty onset?

A: Yes. As explained in the Guidelines, (section (i)) it is critical to record the day of initiation of VO and PPS, not just the completion. The guidelines give instructions for when and how to use the age at initiation vs. completion.

21. Q: Is it necessary for laboratories to score estrous cycle data as first day of one proestrus to the first day of the next proestrus, particularly when one considers that proestrus stage of the cycle lasts 8-12 h and can be missed during once daily evaluations of vaginal smears? How will estrous cycle data be interpreted?

A: As stated in the guideline, "Cycle length may be defined as either the number of days from one proestrus to the next proestrus, or from one diestrus to the next diestrus." (The latter means "the first day of diestrus to the first day of diestrus in the next cycle" and does not include two consecutive days of diestrus.) Estrus to estrus would also provide equivalent information since it should not matter which stage is used as the beginning of a cycle. Interpretation of vaginal cytology data will emphasize a) the time to first estrus, and b) the percentage of animals cycling regularly, with the recognition that the period of observation may be too short even in some control animals to establish a clear pattern of regularity.

22. Q: Will estrous cycle data be used as a "stand-alone" endpoint for the determination of endocrine activity?

A: Interpretation of data will be based on the weight of all of the evidence. EPA recognizes that estrous cyclicity may not be well established within the duration of the pubertal assay even in control animals and thus will generally not rely on small deviations as contributing heavily to the weight of evidence.

23. Q: Is a holding room separate from the room in which the kills and/or necropsies are performed required?

A: The purpose of the holding room adjacent to the kill room is to minimize the stress on animals immediately prior to kill. Stress can affect hormone levels. By keeping the animals undisturbed during holding and minimizing the time between removal of the animal from its cage and the kill, such interference can be minimized. Holding the animals in the kill room is not acceptable as this causes continuous stress. Holding the animals in a hallway close to the kill room is highly discouraged as the disturbance from normal hallway activity can be stressful. If a holding room is not available, EPA recommends that measures be taken to eliminate disturbances that could affect the animals, such as shutting off the hallway to traffic, blocking bright lights, and preventing sudden or continuous noises. EPA recommends that the measures taken be described in the report in order for the EPA to be able to

evaluate the adequacy of the study. As directed in the Guidelines, the final dosing is done in the holding area and is to be completed two hours before kill.

24. Q: May laboratories report fixed pituitary weights in lieu of fresh weights?

A: EPA prefers that fresh weights be reported as this is what was used in the interlaboratory validation study. However, it is unlikely that weights from fixed pituitaries would be rejected as unacceptable simply because they were fixed before weighing.

25. Q: Is it necessary to report weights of the ovaries, uterus (with fluid and blotted), pituitary and adrenal glands to 0.0001 grams? Is it necessary to report thyroid weights to 0.01 milligrams?

A: The test guidelines recommend reporting weights of ovaries, pituitary, adrenals, and uterus to 0.1 mg. They recommend reporting weights of thyroids to 0.01 mg.

26. Q: Is it necessary for laboratories to conduct kidney histopathology if previous toxicity studies have not indicated that the kidney is a target organ for the compound in question?

A: Histological examination of kidneys would generally not be necessary provided that adequate studies previously submitted to the Agency indicate that the kidney is not a target organ for the compound.

27. Q: What other organs may be used to establish that the Maximum Tolerated Dose has been reached but not exceeded? In particular, may changes in the liver or liver enzymes be used as indication that MTD has been reached or exceeded?

A: Judgment of whether the Maximum Tolerated Dose (MTD) has been reached but not exceeded will be made in the context of the overall toxicity profile of the chemical and will not necessarily be based on effects observed for a particular enzyme or organ. The weight of evidence (WoE) will determine whether liver or kidney effects indicate that the MTD has been exceeded. The WoE will include evaluation of clinical signs of toxicity, clinical chemistry, organ weight, histopathology, study design, duration, conduct, and dose spread. The frequency of occurrence and severity of the findings will be factored into the WoE. Therefore the Agency does not recommend relying solely on changes in liver or liver enzymes.

- Clinical signs of toxicity can include minimal or slight behavioral changes (e.g., hyperactivity, transient tremors) or severe behavioral changes (e.g., persistent tremors, salivation, partial paralysis); the latter may indicate that the dose is excessive. Persistent behavioral changes are likely to be considered evidence that

the MTD has been reached or exceeded. Transient changes will be evaluated in conjunction with changes in body weight gains, clinical chemistry, hematology, and histopathology.

- Body weight/body weight gain will be considered with respect to both statistical and biological significance. There is no single method to properly assess body weight data. In the absence of any other indicator of toxicity, a significant decrease in bodyweight or body weight gain compared to controls over the duration of the study will generally indicate that MTD has been reached but not exceeded as long as the difference from controls is not more than approximately 10%.

- Clinical chemistry levels are usually considered adverse when at least two liver parameters have a dose dependent, biologically significant change in albumin; alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, bilirubin, cholesterol or gamma glutamyltransferase. These changes should corroborate each other and be consistent with the known significance of the parameters. With renal toxicity, serum creatinine concentrations tend to parallel changes in BUN. Thus, in well-controlled toxicity studies in rodents, relatively small increases in serum BUN and creatinine concentrations (e.g., ~ 1.5-fold) can be indicative of renal injury but significant and consistent increases in BUN or creatinine above control ranges, including laboratory reference ranges, provide more support for a treatment related effect.

- Organ weight changes in non-endocrine organs should not be the sole criteria used to determine that an adequate dose has been achieved. Significant changes in organ weights may not necessarily be an indication of toxicity. Changes in non-endocrine organ weights should be considered along with other changes (e.g., changes in clinical chemistry and/or histologic effects such as cell proliferation or single cell necrosis) and be biologically relevant.

- Histopathological changes such as hepatocellular hypertrophy, hyperplasia, degeneration or necrosis should be assessed based on incidence and severity. Hepatocellular hypertrophy (and its corresponding increased liver size/weight) may be indicative of adaptation which, by itself, is not necessarily adverse. Generally, a change in one or more histopathological parameters accompanied by one or more changes in hematology, clinical chemistry, or non-endocrine organ weights are likely to be sufficient evidence of attainment of an adequate dose.

In cases where liver damage is thought from prior studies to be the most sensitive adverse effect among the (non-endocrine) endpoints tested, EPA recommends that the laboratory select a fraction of the lowest observed adverse effect level (LOAEL) recognized by regulatory agencies as the highest dose for the pubertal study, and provide a justification for why that fraction was chosen. The Agency recommends that in the absence of other relevant information, a dose no lower than one-half of the LOAEL be used as the highest dose level in the pubertal assay, in order to maximize the challenge to the endocrine system without causing adverse effects.

28. Q: Must every pubertal study that is reported include at least one dose level below the Maximum Tolerated Dose? That is, if at the end of the study it is found that adverse effects are seen at the highest dose level but not at the second-highest dose level, is a third dose level below the second one required?

    A: If at the end of the study it is found that adverse effects are seen at the highest dose level, a third dose level would not be necessary in this situation.

29. Q: Should the listed performance criteria be used as test validity criteria?

    A: The performance criteria are indications of whether the sensitivities of individual endpoints are sufficient to allow conclusions that the test chemical did not affect those endpoints. A study with strongly positive endpoints may be considered valid even if the performance criteria were not met. A study that slightly missed meeting the performance criterion for one out of several potentially-redundant endpoints may, when considered with other available information, be considered acceptable even if that endpoint is statistically not significantly different from controls.

30. Q: Are changes in thyroid hormone levels, without corresponding changes in thyroid weight or histopathology, considered toxicologically meaningful?

    A: The biological/toxicological significance of changes in thyroid hormone levels in the absence of corroborative histopathological changes will be evaluated in the context of the overall toxicity of the chemical using the WoE approach including the thyroid toxicity data available from the amphibian metamorphosis assay.

31. Q: Can criteria other than a 10% reduction in terminal body weight be used to establish the Maximum Tolerated Dose (e.g. clinical signs, clinical chemistry, organ histopathology)?

    A: As stated in the pubertal assay guidelines (section (f), one of the conditions for using decrease in weight gain as an indicator that MTD was reached is that "no clinical signs of toxicity associated with the dose level are observed throughout the study." That was meant to imply that clinical signs may (and should) be used to indicate that MTD had been reached or exceeded. The guidelines continue: "In addition, abnormal blood chemistry values at termination...may indicate that MTD was exceeded, even in the absence of a reduction in terminal body weight compared to controls. Histopathology of the kidney (or any other organ where gross observations indicate damage) may be used as evidence that MTD was exceeded."

32. Q: Many IACUCs will not look favorably on decapitation without anesthetic and some labs may be concerned about decapitation resulting in damage to the thyroid and contamination of trunk blood with necropsy debris. Can alternative methods for anesthesia, euthanasia, or blood collection be used if these methods do not significantly alter hormone levels in untreated control animals? What are acceptable alternative methods?

A: EPA is mainly concerned with stress that may be induced by use of anesthetics or CO_2. Such stress may affect hormone levels, particularly testosterone. It is important to obtain the blood from animals before such hormone level changes occur. For that reason, use of CO_2 is strongly recommended to be limited to 60 seconds, after which decapitation is performed even if the animal has not fully succumbed. With inhaled anesthetics, some individuals may attempt to hold their breath for an extended period and this may result in stress. Injected anesthetics are less subject to this voluntary modification but the time it takes to reach deep anesthesia may still be somewhat unpredictable depending on both the test chemical and the individual.

While decapitation remains the preferred method of kill for the pubertal assays (with injectable anesthetic for the females, without anesthetic for the males), the use of anesthetic (injectable or inhalational) followed by aortal exsanguination may provide a reasonable alternative if performed appropriately, for both the male and female pubertal assays. If anesthetic is used, EPA strongly recommends that the individual reach deep anesthesia within 2 minutes. If an animal has not reached a sufficient level of anesthesia for exanguination by 2 minutes, record the time it takes to reach the appropriate level of anesthesia and mark that animal as a deviation. It should then be possible to evaluate whether the extended time affected hormone levels, and to do the necessary analyses excluding those animals if appropriate.

EPA recommends that labs verify that they are able to recover sufficient blood from each individual to do the necessary hormone analyses using the exsanguination method chosen. Systematic lack of blood volume sufficient to provide hormone measurements is likely to be considered a serious deficiency of a study because the hormone measurements are important pieces in the weight-of-evidence determination of interaction with the thyroid and androgen systems and acceptable methods of collection of sufficient volume are readily available.

33. Q: Is housing in accordance with the recommendations contained in Guidelines for the Care and Use of Laboratory Animals acceptable or are only clear plastic containers of the dimensions specified in the TG allowed?

A: Dimensions specified in the pubertal test guidelines are approximate, as stated in the guidelines. The use of clear plastic containers is not critical; other materials that meet the GCULA recommendations are equivalent for the purposes of this assay.

34. Q: Is it necessary to report the diet used at the animal supplier? This information is not typically available.

A: Concerns have been expressed by some scientists about the influence of phytoestrogens in the feed of dams from which animals used in endocrine studies are obtained. It is considered prudent to collect information that might be helpful in this respect at some future time. EPA suggests that laboratories request the identity of the diet that the animals provided by a supplier were fed prior to receipt by the test laboratory. If this information is not forthcoming from the supplier, note this in the report.

Fish Short-Term Reproduction Assay (FSTRA)

1. Q: Since OECD test guideline 229 is considered consistent with test guideline OCSPP 890.1350, would it be accepted in lieu of OCSPP 890.1350?

   A: In general, OECD test guideline 229 would provide much of the same information as test guideline OCSPP 890.1350. However, it should be noted that test guideline OCSPP 890.1350 contains endpoints (e.g., fertility) that are not part of OECD Test Guideline 229. In particular, consideration of egg fertility from control animals is among the performance criteria in OCSPP 890.1350.

2. Q: There are several places in the guideline where very specific instructions are provided. Is it necessary to use all the equipment, suppliers, materials, analytical methodology (e.g., ELISA for vitellogenin), and range-finding methods as specified in the guideline?

   A: A distinction between "test guideline" and "test mandate" should be recognized. The test guideline provides a description of materials and methods that have proven to be satisfactory. EPA recommends that deviations from the guidance provided be explained, but each study will be judged on its own merits.

3. Q: There is guidance concerning the selection of the highest test concentration, but there is no guidance concerning the selection of lower test concentrations. The test guideline for the Amphibian Metamorphosis (Frog) Assay (OCSPP 890.1100) suggests spacing of not less that 3-fold nor more than 10-fold. Would this be applicable to the fish assay as well?

   A: Spacing of not less than 3-fold nor more than 10-fold would be scientifically acceptable. OECD test guideline 229 provides similar guidance: "A range of spacing factors between 3.2 and 10 is recommended."

4. Q: How will endocrine-mediated effects be distinguished from non-endocrine mediated toxicity in the fish screen?

   A: If a given exposure level results in substantial mortality or other overt signs of toxicity, responses for other endpoints may be due to general toxicity, not necessarily mediated primarily via interaction with the endocrine system. To address this, EPA recommends that the lower treatment level(s) be examined for effects outside of the range of general toxicity. If all test concentrations exhibit mortality, then the assay would likely need to be repeated with lower concentrations before inferences about possible endocrine activity can be made. It is recognized that some endpoints may be responsive to non-endocrine stresses in addition to endocrine-mediated pathways, particularly fecundity. Although reductions in

fecundity indicate adverse organismal and, potentially, population level effects (i.e., reproductive toxicity), these cannot be definitively distinguished from direct endocrine-mediated effects by this assay when changes in other core endpoints are not present. Nevertheless, reductions in fecundity are best considered a positive effect in this assay because they may be endocrine-mediated, but should also be considered in concert with results of other assays in the Tier 1 battery. Similarly, responses in secondary measurements (e.g., length, weight) should be considered in light of other results from the battery. More generally, results that are considered to be equivocal for this single assay are best considered indications of potential endocrine activity and further evaluated in light of the weight of evidence from other assays in the Tier 1 battery.

5. Q: Do we use the criteria for an endocrine positive or an endocrine negative test as stated in the guideline?

   A. The criteria for an endocrine positive or an endocrine negative test as stated in the test guideline are provided for interpreting results from this single assay. Therefore, any statistically significant effect in one or more of the core endpoints of this assay (i.e., fecundity, secondary sex characteristics, vitellogenin, GSI, and histopathology) may be indicative of a potential of the test chemical to disturb the hypothalamo-pituitary-gonadal (HPG) axis of fishes. Nevertheless, results from this assay are intended to be used in conjunction with results from other assays in the Tier 1 battery. Determination of whether a chemical has the potential to interact with the estrogen, androgen, or thyroid hormonal systems will be made on a weight-of-evidence basis taking into account data from the Tier 1 assays and other scientifically relevant information available.

6. Q: The guideline states that mean measured test concentrations should be maintained at ≤ 20% CV over the 21 day test. If concentrations are not maintained at ≤ 20% CV, does this necessarily invalidate the test (especially when testing difficult test substances)?

   A. Maintenance of test concentrations with a CV ≤ 20% is consistent with OECD test guideline 229 and other aquatic test guidelines. A CV > 20% in measured test concentrations generally occurs with poor diluter setups or with difficult substances. The test guideline and other existing guidance for dealing with difficult substances (OCSPP 850.1000, Special considerations for conducting aquatic laboratory toxicity studies) offer some flexibility in terms of what is acceptable, provided that documentation of steps taken to optimize solubility and test concentration stability is included.

7. Q: The guideline suggests that unfortified and fortified vitellogenin (VTG) samples be run each day that a VTG ELISA is conducted. Since there is no known commercially available source of fathead minnow VTG, is it necessary to perform this evaluation?

What are the expected performance criteria (variability) of the unfortified and fortified VTG samples in the ELISA?

A. Running unfortified and fortified VTG samples each day that a VTG ELISA is conducted is recommended. Analyses of unfortified and fortified samples of control male plasma provide an additional level of quality control intended to account for potential inter-assay variation. Ideally, VTG used to fortify the sample of control male plasma as well as the control male plasma sample itself should be of a large enough volume such that all samples/ELISAs from a particular study can be run using the same fortifying VTG/control male plasma. This fortifying VTG and control male plasma could be prepared in advance of running the first ELISA and aliquoted for single use. Taking this approach will most likely necessitate that the sample of control male plasma be prepared as a composite from several fish. The VTG used to fortify the plasma sample may be prepared using VTG ELISA standards. The control male plasma sample should be fortified to achieve a VTG concentration between 10 and 100 times (closer to 100 is recommended) the expected VTG concentration of control male fish. A CV of ≤ 20% for measured concentrations of the fortified and unfortified samples between assays is expected.

8. Q: For potential use as negative controls in the FSTRA, what compounds have tested as conclusive negatives in the Fish Short-Term Reproduction Assay?

   A. Negative controls are not used in this assay.

9. Q: The highest concentration to be tested is based on mortality and is inconsistent with the MTD used for other assays. Will concentrations set using sublethal endpoints be allowed?

   A. Determination of concentrations using sublethal endpoints would be consistent with the test guideline, provided that justification for the chosen test concentrations is provided.

10. Q: Is use of the randomized block design required?

    A. The use of a randomized block design for allocation of spawning groups based on pre-exposure spawning performance and physical placement of replicates in the test system as described in the Test Guideline is essential to ensure equitable distribution of these groups across the test concentrations and to minimize variability associated with the experimental environment.

11. Q: Is measuring the dorsal nape pad required?

A. Measuring the dorsal nape pad is not required, although EPA recommends that any unusual changes in dorsal nape pads noted during routine daily observations be recorded.

12. Q: Is a static-renewal test design an equal option to testing using a flow-through test system?

A. A static-renewal test is not recommended for conducting this assay.

Amphibian metamorphosis assay (AMA)

1. Q: Is *Silurana (Xenopus) tropicalis* an acceptable alternative species for the AMA?

   A: Use of *Silurana (Xenopus) tropicalis* would be consistent with the test guideline. However, use of *Xenopus laevis* is recommended. *X. laevis* is currently being used in the validation effort for the Tier 2 Larval Amphibian Growth and Development Assay (LAGDA).

2. Q: Can other endpoints besides mortality be used for determining non-specific toxicity?

   A: Determination of concentrations using sublethal endpoints would be consistent with the test guideline, provided that a justification for the chosen test concentrations has been provided.

3. Q: If dietary iodide is available to frogs undergoing normal metamorphosis, is it necessary to supplement the water with additional iodide?

   A: Previous studies have shown that organisms reared in water containing 0.5 to 2.0 µg/L iodide develop normally with the dietary source of iodide included. Therefore, an iodide concentration within that range would be acceptable.

4. Q: Will studies be rejected if tadpoles are used that did not reach stage 51 in 17 days?

   A: The determination of whether to reject a test in which the tadpoles used did not reach stage 51 in 17 days before being placed on test will depend on how well the test meets the other performance criteria specified in the test guideline.

5. Q: Must all organisms used in a study originate from a single spawn event as long as all tadpoles are uniform in their development at stage 51?

   A: Use of organisms from a single spawn is recommended. The purpose of using a single spawn is to reduce any variability that could be introduced by using multiple spawns. In general, an individual spawn should be sufficiently large to accommodate the testing needs.

6. Q: Will snout-vent length (SVL) based on termination of the abdomen be accepted (in lieu of the cranial aspect of the vent)?

   A: The SVL is a more traditional measurement, but measurement based on termination of the abdomen may be acceptable if employed consistently. EPA recommends that the study report describe which method was used.

7. Q: The statement "Statistically significant developmental delays, in absence of other signs of overt toxicity, indicate that the chemical is thyroid active (antagonistic)" is inconsistent with earlier statements regarding the use of histopathology in conjunction with observations of developmental delay. Should thyroid histopathology be used to confirm potential thyroid involvement in developmental delays?

   A: The test guideline requires thyroid histopathology for all tests except those that indicate advanced development.

8. Q: Do we use the criteria of a positive or negative test (as related to thyroid activity) as stated in the guideline?

   A: The criteria for a positive or negative test as stated in the test guideline are provided for interpreting results from this single assay. Nevertheless, results from this assay are intended to be used in conjunction with results from other assays in the Tier 1 battery. Determination of whether a chemical has the potential to interact with the estrogen, androgen, or thyroid hormonal systems will be made on a weight-of-evidence basis taking into account data from the Tier 1 assays and other scientifically relevant information available.

9. Q: How frequently are bent frogs (tail flex) observed? Will observations of bent frogs result in study rejection because it is interpreted as non-specific toxicity or resulting from chemical exposure?

   A: Tail flex is not an unusual observation, and its origin is not well understood. Tail flex may occur across an entire study, including the controls. Tail flex that occurs in treatment groups at a similar rate as in controls would not typically be identified as a treatment-related effect. However, if the rate of tail flex is increased in treatment groups or is present in treatment groups but absent in controls, it may be identified as a treatment-related effect.

10. Q: For potential use as negative controls in the AMA, what compounds have tested as conclusive negatives in the AMA?

    A: Negative controls are not used in this assay.

11. Q: The guideline asks that special attention should be given as to whether the water is free of copper, chlorine and chloramines and furthers recommends analysis of the dilution water for background levels of fluoride, perchlorate and chlorate. What are the acceptable levels?

A: Specific levels have not been determined for test guideline OCSPP 890.1100 (AMA). In general, these are the same considerations used in most general aquatic toxicology labs.

12. Q: What is the optimum iodide concentration in the food (i.e., in case Sera Micron is not available or a different diet is chosen)?

    A: The optimum iodide concentration in food has not been determined.

13. Q: For interspecies comparison used for MTC evaluation, what species are acceptable and what defines professional judgment?

    A: Small fish data would likely be the most commonly available.

14. Q: Is there another method recommended by EPA for randomized selection of day 7 subsamples that does not involve netting all of the animals?

    A: The only other method for randomized selection would require that individuals be uniquely marked, which is infeasible. During validation efforts, the process of netting did not result in deleterious effects.

15. Q: The guideline states that test concentrations should be maintained at ≤ 20% CV over the 21 day test. If concentrations are not maintained at ≤ 20% CV, does this necessarily invalidate the test (especially when testing difficult test substances)?

    A: Maintenance of test concentrations with a CV ≤ 20% is consistent with test guideline OCSPP 890.1350 (and OECD Test Guideline 229) and other aquatic test guidelines. A CV > 20% in measured test concentrations generally occurs with poor diluter setups or with difficult substances. The test guideline and other existing guidance for dealing with difficult substances (OCSPP 850.1000, Special considerations for conducting aquatic laboratory toxicity studies) offer some flexibility in terms of what is acceptable, provided that documentation of steps taken to optimize solubility and test concentration stability is included.

16. Q: In phase 3, 50%, 0%, and 48% of the organisms from the three studies were > than stage 60. Is there an upper limit of % animals beyond stage 60 where the method can be used?

    A: The test guideline includes a statistical method for addressing the situation when >20% of the animals in a treatment exceed stage 60. No upper limit has been established.