Independent Scientific Peer Review Panel Report
Evaluation of the LUMI-CELL® ER (BG1Luc ER TA) Test Method

May 2011

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences (NIEHS)
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

This document is available electronically at
http://iccvam.niehs.nih.gov/docs/endo_docs/EDPRPRept2011.pdf

When referencing this document, please cite as follows:
Interagency Coordinating Committee on the Validation of Alternative Methods. 2011. Independent Scientific Peer Review Panel Report: Evaluation of the LUMI-CELL® ER BG1Luc ER TA Test Method. Research Triangle Park, NC:National Institute of Environmental Health Sciences.

The findings and conclusions of this report are those of the Independent Scientific Peer Review Panel and should not be construed as representing the official views of ICCVAM or its member agencies.

Table of Contents

List of Abbreviations and Acronyms

BG1Luc ER TA	LUMI-CELL® BG-1Luc4E2 ER TA test method
BPA	Bisphenol A
BRD	Background review document
CERI	Chemicals Evaluation and Research Institute (Japan)
CV	Coefficient of variation
DMSO	Dimethyl sulfoxide
EAC	Endocrine-active chemical
EC_{50}	Half–maximal effective concentration
ECVAM	European Centre for the Validation of Alternative Methods
ED	Endocrine disruptor
EDSP	EPA Endocrine Disruptor Screening Program
EDSTAC	EPA Endocrine Disruptor Screening and Advisory Committee
EDTA	Endocrine Disruptor Testing and Assessment (OECD)
EPA	U.S. Environmental Protection Agency
ER	Estrogen receptor
ERE	Estrogen-responsive element
FBS	Fetal bovine serum
FIFRA	U.S. Federal Insecticide, Fungicide, and Rodenticide Act
GCCP	Good Cell Culture Practices
GLP	Good Laboratory Practices
IC_{50}	Concentration of the test substance that inhibits the reference estrogen response by 50%
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ISO	International Organization for Standardization
JaCVAM	Japanese Center for the Validation of Alternative Methods
KoCVAM	Korean Center for the Validation of Alternative Methods
NICEATM	U.S. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	U.S. National Institute of Environmental Health Sciences
NIH	U.S. National Institutes of Health
OECD	Organisation for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides and Toxic Substances
QA	Quality assurance
QSAR	Quantitative structure–activity relationship
REACH	Registration, Evaluation and Authorisation of Chemicals
RLU	Relative light unit
RUC	Rat uterine cytosol
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods

SD	Standard deviation
SOP	Standard operating procedure
STTA	Stably transfected human estrogen receptor-α transcriptional activation
TA	Transcriptional activation
TG	Test Guideline (OECD)

Members of the Independent Scientific Peer Review Panel

John G. Vandenbergh, PhD (Panel Chair), Professor Emeritus, Department of Biology, North Carolina State University, Raleigh, NC

A. John Bailer, PhD, Distinguished Professor and Chair**,** Department of Statistics**,** Miami University**,** Oxford, OH

Christopher Borgert, PhD, President and Principal Scientist, Applied Pharmacology and Toxicology, Inc., Gainesville, FL

Grantley Charles, PhD, DABT, Senior Scientist, Toxicology, Drug Safety Evaluation, Allergan Inc., Irvine, CA

Daniel Desaulniers, PhD, Research Scientist, Hazard Identification Division, Health Canada, Ottawa, ON, Canada

J. Charles Eldridge, PhD, Professor of Physiology and Pharmacology, Wake Forest University Health Sciences, Winston-Salem, NC

William R. Kelce, PhD, Principal, Exponent®, Cary, NC

Hyung Kim, PhD, Associate Professor, College of Pharmacy, Pusan National University, Busan, Republic of Korea

Steven L. Levine, PhD, Science Fellow and Product Safety Manager, Monsanto Company, St. Louis, MO

Alberto Mantovani, DVM, Head of Food and Veterinary Toxicology Unit, Italian National Health Institute (ISS - Istituto Superiore di Sanità, viale), Rome, Italy

Ellen Mihaich, PhD, DABT, President and Principal Scientist, Environmental and Regulatory Resources, LLC, Durham, NC

Hiroshi Ono, PhD, Scientific Advisor, Food and Drug Safety Center, Hatano Research Institute, Hadano, Japan

Sherry L. Ward, PhD, MBA, Consultant, BioTred Solutions, New Market, MD

Marc Weimer, PhD, Biostatistician, Department of Biostatistics, German Cancer Research Center, Heidelberg, Germany

James Wittliff, PhD, MD hc, FACB, Professor, Department of Biochemistry and Molecular Biology, Director, Hormone Receptor Laboratory, Director, Institute for Molecular Diversity and Drug Design, University of Louisville, Louisville, KY

James Yager, PhD, Professor and Director of Toxicology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

Preface

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) convened an international independent scientific peer review panel (Panel) meeting on March 29-30, 2011, at the Natcher Conference Center in Bethesda, Maryland. The Panel, which included 16 expert scientists from six countries, evaluated the LUMI-CELL® estrogen receptor (ER) transcriptional activation (TA) test method (BG1Luc ER TA), an *in vitro* TA assay used to identify chemicals that can interact with human ERs.

During the March 2011 public meeting, the Panel discussed the test method, listened to public comments, and developed conclusions and recommendations for ICCVAM. The Panel focused on the following areas: (1) review of the ICCVAM draft background review document (BRD) for completeness and identification of errors or omissions of existing relevant data or information that should be included, (2) evaluation of the information in the draft BRD to determine the extent to which each of the applicable ICCVAM criteria for validation and acceptance of toxicological test methods had been appropriately addressed, and (3) consideration of the ICCVAM draft test method recommendations and commentary on the extent to which they are supported by the information provided in the draft BRD for the following:

- Proposed test method uses and limitations
- Proposed recommended standardized protocols
- Proposed test method performance standards
- Proposed future studies

This report details the Panel's independent conclusions and recommendations. ICCVAM will consider this report and all relevant public comments as it develops final test method recommendations. The ICCVAM final test method recommendations will be forwarded to U.S. Federal agencies for their consideration in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. 285*l*-3).

The Panel gratefully acknowledges the efforts of NICEATM staff for an outstanding effort in coordinating the logistics of the Panel meeting and in the preparation of materials for its review. Finally, as Panel Chair, I want to thank each Panel member for her or his thoughtful and objective review of this test method.

John G. Vandenbergh, PhD
Chair, BG1Luc ER TA Test Method Peer Review Panel
May 2011

This page intentionally left blank

Executive Summary

This report describes the conclusions and recommendations of an international independent scientific peer review panel (Panel). The Panel was charged by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) with evaluating the validation status of the BG1Luc estrogen receptor (ER) transcriptional activation (TA) test method according to established Federal and international criteria (ICCVAM 1997). The Panel also commented on ICCVAM draft recommendations regarding the usefulness and limitations of the test method and proposed performance standards.

The Panel considered the results of an international interlaboratory validation study that included laboratories in the United States, Italy, and Japan. Based on their evaluation of these data, the Panel agreed with ICCVAM's draft test method recommendation that the BG1Luc ER TA test method can be used to identify substances with *in vitro* estrogenic and anti-estrogenic activity. Based on results of concordance analyses for a limited number of substances, the Panel further concluded that the BG1Luc ER TA test method could be considered as a replacement for other *in vitro* assays that may provide substantially similar information, specifically the Chemicals Evaluation and Research Institute stably transfected transactivation assay (CERI STTA) and the rat uterine cytosol (RUC) ER binding assay. The Panel noted that additional analysis could further support this recommendation, particularly regarding the RUC ER binding assays.

The Panel endorsed the draft ICCVAM-recommended test method protocols and noted several advantages provided by this assay over the currently accepted test method for this endpoint, including the robust test method protocol, the validated testing range, and the ability to detect substances with *in vitro* anti-estrogenic activity. However, the Panel also noted that careful analysis of cytotoxicity is critical for correctly interpreting results. The Panel expressed a preference for using quantitative approaches for such a measurement. The Panel recommends that a potency endpoint, such as the half-maximal effective and/or inhibitory concentration (EC/IC_{50}), be included in each study report and that the uncertainty associated with these estimates should also be reported. The Panel considered the descriptive approach for evaluating test method reliability acceptable but suggested additional statistical analyses that could be performed to better characterize and understand variability.

The Panel agreed with the draft ICCVAM-recommended future studies and suggested additional studies that should be conducted to expand the usefulness of the BG1Luc ER TA test method. The Panel recommended additional evaluations of the utility of the current categorical assessment of cytotoxicity and advocated for the implementation of a quantitative method for its replacement. The Panel also recommended studies to add *in vitro* metabolism (compound activation or inactivation) to the test method. This addition could expand the utility of this and other ER TA test methods. The Panel recommended that additional efforts focus on expanding the reference substance list, and subsequently the BG1Luc ER TA test results, with additional negative agonist and positive antagonist test substances.

Finally, the Panel concurred that the draft ICCVAM performance standards could be used to evaluate the validation status of test methods that are functionally and mechanistically similar to the BG1Luc ER TA test method. The Panel considered the list of performance standards reference substances to be adequate. The Panel noted that ideally more negatives should be included but recognized that data on such substances are not currently available. When evaluating test method accuracy, the Panel strongly supported quantification of relative agonist and antagonist activity in addition to the dichotomous call of positive or negative. In addition, the Panel concluded that the potent estrogens on the reference list should not be misclassified, but there could be some tolerance for discordance for the weakly active reference substances. Discordant results need to be discussed in terms of the ability of the test method to detect a similar range of potencies and intrinsic activities compared to current

validated test methods. Discordant results for particular chemicals or product classes also need to be discussed.

Overview

Use of the BG1Luc ER TA Test Method to Identify Substances as Potential *In Vitro* Estrogen Receptor Agonists or Antagonists

The overall question that the international independent scientific peer review panel (Panel) considered is whether the validation status of the BG1Luc estrogen receptor (ER) transcriptional activation (TA) test method has been adequately characterized for its intended purpose and whether it is sufficiently accurate and reliable to be used to identify substances with estrogen agonist and/or antagonist activity.

The Panel discussed the intended use of this assay and the potential for its inclusion in a regulatory testing battery. Panel members agreed that the BG1Luc ER TA test method and the Chemicals Evaluation and Research Institute stably transfected transactivation assay (CERI STTA) are similarly capable of assessing *in vitro* estrogen receptor (ER) agonist activity. In addition, the BG1Luc ER TA test method is capable of detecting *in vitro* estrogen antagonists. Clarification of the intended use of these assays in regulatory decision making, particularly in the context of the U.S. EPA's Endocrine Disruptor Screening Program (EDSP), would enable a better understanding of the relative merits of the various screening assays for their intended purpose.

In the absence of clear regulatory guidance, the Panel recommends that the BG1Luc ER TA test method be endorsed as a scientifically valid method for assessing the *in vitro* estrogen agonist and antagonist activity of compounds within a test battery or tiered testing scheme. The Panel recommends that the BG1Luc ER TA method be considered as a replacement for other *in vitro* assays that, in combination, may provide substantially similar information, specifically the CERI STTA assay. This is supported by the following findings:

- The concordance of the BG1Luc ER TA test method with the CERI STTA assay suggests that the BG1Luc ER TA test method and the CERI assay produce similar results.
- The thoroughness and transparency of the BG1Luc ER TA method validation process compare favorably with other *in vitro* assays.
- The detailed BG1Luc ER TA agonist and antagonist protocols permit ease of use.
- The detailed and publically available BG1Luc ER TA data permits thorough evaluation of the performance of the method.
- The endogenous expression of both ERα and ERβ in BG1Luc4E2 cells allows *in vitro* activity through both receptors to be assessed in the BG1Luc ER TA test method. Endogenous expression of the receptor and its related endogenous cellular machinery may be an advantage over receptors that are stably transformed into an immortal cell line and constitutively expressed at high levels.

I. Review of the BG1Luc ER TA Test Method BRD for Errors and Omissions

The Panel noted typographical errors in the draft background review document (BRD), which are detailed in Appendix A of this report.

The Panel also cited relevant information that should be included that would improve the BRD. These include:

- In Figure 1-1 of the BRD, Phase 2 should be changed to Laboratory Proficiency Phase because the laboratory qualification should already have been demonstrated.
- Ongoing improvements were made to the protocol(s) during the course of these validation studies. A better assessment of their impact on the final study results is needed.
- The Panel has recommended additional reliability analyses; therefore, Section 6.0 of the BRD and Lines 43-49 and 76-82 of the draft recommendations of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) may need to be revised pending their outcome.
- Data quality sections for the reference test method data were not included in the draft BRD. The Panel recommends they be added for CERI STTA, ICCVAM literature-based reference data, and RUC ER binding assay.
- Sections II and III of this peer review panel report detail additional information gaps cited by the Panel.

II. Evaluation of the Validation Status of the BG1Luc ER TA Test Method

1. Introduction and Rationale for the Test Method

a. Is the historical background provided for the BG1Luc ER TA test method and the rationale for its development adequate?

While the historical background provided in the BRD is adequate, it would be improved by including a more thorough discussion of the specific advantages of this transactivation assay relative to other *in vitro* ER assays (see Rogers and Denison 2000). Accordingly, each of the following important points should be emphasized in the Background section of the BRD:

- Transactivation assays provide advantages compared to ER binding assays in that ER transcriptional activation (TA) assays have the potential to assess both *in vitro* estrogen agonist and antagonist activity. They can also detect endocrine-active substances that elicit effects on estrogen-regulated pathways through non–receptor-mediated mechanisms.
- BG-1 cells endogenously express both ERα and ERβ and, consequently, possess the full transcriptional machinery required for estrogen responsiveness.
- Stably transfected cell lines have advantages over transiently transfected cell lines, including long-term utility without the need to transiently transfect cells for each assay.
- The BG1Luc ER TA test method has demonstrated *in vitro* responsiveness to estrogens and limited cross-reactivity with ligands of other steroid hormone receptors.
- A discussion should be added to detail the four copies of the estrogen-responsive element (ERE) and their orientation, which mimics that in the fish vitellogenin promoter and thereby indicates a native system.

b. Are the current regulatory testing requirements and ICCVAM prioritization criteria adequately discussed and up to date?

ICCVAM's prioritization criteria are adequately discussed. If possible, the BRD should compare the phased Organisation for Economic Co-operation and Development (OECD) conceptual framework approach with the EPA Tier 1 battery approach, as discussed at the March 2011 Panel meeting. If possible, as new/other worldwide regulatory agencies develop endocrine disruptor (ED) testing strategies, this information should be updated via addenda.

c. Are the purpose and mechanistic basis of the BG1Luc ER TA test method adequately described?

The purpose of this test method as a screen for *in vitro* estrogen receptor agonist and antagonist activity has been adequately stated. The mechanistic basis of the test method also is adequately described.

d. Is the description of the use of the proposed test method in an overall strategy of hazard or safety assessment adequate?

The use of the proposed test method in the overall strategy of hazard or safety assessment of endocrine-active chemicals (EACs) is unclear. There has been no clear regulatory guidance on how the method will be used in the EPA Endocrine Disruptor Screening Program (EDSP).

Because the BG1Luc ER TA method has been validated to assess compounds for *in vitro* estrogen agonist and antagonist activity, the BG1Luc ER TA method should be considered as a replacement for the CERI STTA (which detects only agonist activity) and the RUC ER binding assays.

If, in the future, a receptor binding assay is still deemed necessary, the Panel recommends an evaluation of recombinant (human and animal) ER binding assays as replacements for the RUC ER binding assay.

The BRD should also propose how the BG1Luc ER TA method could be used in screening, prioritization, and safety assessments for other regulatory testing programs (see comment above regarding the OECD conceptual framework).

2. Test Method Protocol

a. Are the protocols complete and adequate in detail for a laboratory to conduct the study, including (1) a description of the material and equipment needed to conduct the test and (2) a description of what is measured and how the data are used to identify positive and negative results?

The protocols in Annexes E and F of the BRD appear complete and adequate in detail for a laboratory to conduct the study. Although the protocols state that alternative vendors for technical and cell culture equipment and supplies may be used, this should be more clearly stated in the BRD body.

The approaches to evaluate data were reviewed and the Panel's recommendations on quantification of activity are presented below in the response to Section II Question 4b.

b. Overall, are critical aspects of the test method protocol, as outlined in the ICCVAM Submission Guidelines, adequately justified and described in the BRD?

Critical aspects of the test method protocol are adequately justified and described in the BRD. However, the subjective nature of the visual assessment method for determining cytotoxicity raised concerns about its application with regard to future usage of the test method protocols.

These concerns are:

- Visual assessment of cytotoxicity is subject to operator inconsistencies and may not always accurately reflect the viability of cells in culture. This is particularly relevant as the use of the method expands to other laboratories.
- The current cytotoxicity evaluation is in line with currently accepted practices (ISO 2009). The International Organization for Standardization (ISO) document compared the use of the subjective visual assessment of cytotoxicity with the objective MTT assay with regard to medical devices. This type of comparison has not but should ideally be completed in the future for chemicals evaluated in the BG1Luc ER TA.
- An accurate, objective cytotoxicity method is most critical for the antagonist assay because it measures a loss of function mediated through the ER. Therefore, an ER-mediated decrease in activity must be delineated from that resulting from cytotoxicity alone.

For future test method use, the Panel recommends:

- Testing of a wider set of substances with known mechanisms of cytotoxicity to further validate the qualitative cytotoxicity method
- Use of quantitative cytotoxicity methods when developing new *in vitro* ER assays, which could also allow for normalization of relative light unit (RLU) responsiveness

i. Do you agree with reference standards and controls proposed for the agonist and antagonist protocols? Are there other reference standards and/or controls that you would consider more appropriate?

The reference standards and controls proposed for the agonist and antagonist protocols are appropriate. Future studies should consider including confirmation assays using a pure ER antagonist (e.g., ICI 182,780), which would confirm ER binding behavior.

ii. Do you agree with the plate acceptance criteria as defined in the agonist and antagonist protocols? Are there additional criteria that should be routinely used?

The plate acceptance criteria defined in the agonist and antagonist protocols appear adequate. However, the criteria for acceptance based on the dimethyl sulfoxide (DMSO) controls could be too stringent if, for example, luminometer replacement is necessary. In such a situation, baseline luminometer values might change. The expectation is that the laboratory would perform adequate repeat baseline and positive and negative controls to assure reproducibility but would not repeat an entire validation study. A suggestion was made to use an acceptable range in RLUs induced by a specific concentration of 17β-estradiol. Such criteria would be independent of equipment and of the choice of vehicle; however, this approach would require validation.

iii. Do you agree with the proposed decision criteria for identifying a positive or negative response in the agonist and antagonist protocols?

The proposed decision criteria for identifying a qualitative positive or negative response in the agonist and antagonist protocols are acceptable (see response to Section II Question 4b).

However, potency and intrinsic activity relative to a known endogenous ligand are critical components in determining whether a substance is truly positive, and this is not addressed in the current decision criteria. This assessment could include a relative potency approach by comparing to known reference substances. Including this component would result in dramatic improvement in the utility of this assay.

The Panel recommends that the half-maximal effective and/or inhibitory concentrations (EC/IC_{50}s) (or potency endpoints such as EC/IC_{20} or the concentration associated for a particular fold induction) be included in each study report and discussed in the Conclusions section in association with the qualitative dichotomous positive or negative calls. Uncertainty associated with these estimates should also be reported (e.g., confidence intervals).

3. Substances Used for the Validation Study

a. Do you consider the database for the BG1Luc ER TA test method representative of a sufficient range of chemical classes and physicochemical properties and that it would be applicable to any of the types of chemicals and products that are typically tested for estrogenic activity? If not, what are the relevant chemical classes/properties (other than those that are identified as limitations in the previous ICCVAM BRD [ICCVAM 2003, 2006]) that should be tested with caution, or not evaluated using this test method? What chemicals or products should be evaluated to fill this data gap?

The chemicals tested in this validation exercise represent a broad range of different chemical classes and physicochemical properties and represent a census of available information. The list follows ICCVAM guidance (ICCVAM 2003, 2006) and is largely applicable to chemicals and products that would be screened to evaluate potential *in vitro* estrogenic activity. Testing an adequate range of activities and a structurally diverse group of chemicals is important. In this validation study, estimated EC_{50}s ranged seven orders of magnitude and varied in terms of their coverage of chemical

classes. However, several chemical classes in Table 3-7 of the BRD are represented by as few as one substance; therefore, in such cases, no conclusions on usefulness or limitations specific to these classes can be made.

As new chemicals are tested in both agonist and antagonist protocols (e.g., data generated from ongoing EDSP Tier 1 testing and/or new ER TA protocols), the Panel recommends including a subset of the newly tested chemicals in any future validation studies. In particular, inclusion of data from chemicals in under-represented classes, compounds with known surface-active properties that could perturb the cell system, and compounds that are known to be negative for agonist activity are necessary. This last recommendation concerning negative compounds is particularly important because less than 25% of the agonist substances used for the accuracy analysis are negative. While the Panel acknowledges that great effort went into identifying substances that would fit into this category, it is difficult to investigate false positives in a new test system if the majority of test chemicals are positive.

b. Do you agree with the methodology used to establish the consensus reference classification that was assigned to each reference substance?

The use of the majority classification criteria among study results (i.e., >50%) to establish the consensus reference classification for each reference substance is a reasonable strategy. However, the criteria used to evaluate the quality of the data obtained from the literature to determine the reference classification needs to be described in the BRD. A ranking method such as Klimisch criteria (Klimisch et al. 1997), which focuses primarily on the reliability of the data, would provide clarity on the relative quality of the reference data and strengthen the classification. In addition, a sensitivity analysis could be performed to assess how the results of the validation study vary based on the method of reference classification. Alternatively, the concordance assessment could be repeated with a different classification criterion, such as declaring a chemical positive if at least one laboratory declared a positive result.

4. Data and Results

a. Have all known data for all studies used to evaluate the accuracy and reliability of the BG1Luc ER TA test method been provided?

Data for all studies used to evaluate the accuracy and reliability of the BG1Luc ER TA test method appear to have been provided. The interlaboratory studies of the BG1Luc ER TA test method conducted by NICEATM, JaCVAM, and ECVAM have been included in the accuracy and reliability assessments provided in the BRD. The Panel is unaware of additional studies available for comparison to the BG1Luc ER TA test method for accuracy and reliability.

b. Are the statistical and nonstatistical approaches to evaluate the data resulting from studies conducted with the BG1Luc ER TA test method appropriate? What other approaches could have been used?

Approaches for assessing test method accuracy

The Panel strongly supports the quantification of activity as a complement to the dichotomous call of a positive or negative response. Interpretation of the results should not rely solely on statistics but also on scientific judgment and should incorporate consideration of the nature and shape of the dose-response relationship and, if needed, the reproducibility of the response in independent experiments.

At a very general level, the Panel recommends that the following questions be explicitly addressed in the BRD:

1. What is the role of statistical modeling in the quantification of activity in this assay?
2. What aspects of the dose-response relationship are of greatest interest? For example, potency evaluation may be of interest.
3. Would characterization of a chemical's estrogen receptor activity be most relevant and best utilized for comparison to that of a reference compound's estrogenic response?
4. Are there statistical criteria that can be used to support the decision of a positive call?
    - Should statistical tests of trend be included in the evaluation of a positive call?

ICCVAM-recommended criteria (ICCVAM 2003) incorporating appropriate statistical methods and sound scientific judgment for classifying a substance as an ER agonist or antagonist are essential for ensuring the credibility of the results. However, the evaluation of data from studies conducted with the BG1Luc ER TA test method would benefit from specific attention to the following detailed comments.

The criteria for making a positive call
As defined in the BRD, the criteria for a positive call are based on comparisons to experimental results from the current set of test chemicals and sound scientific judgment. While the Panel views this as an acceptable strategy, a statistical test for trend is a reasonable approach for making a positive call (Bretz and Hothorn 2003).

The criteria for estimating potency
Potency estimation requires an adequate concentration–response model (Ritz 2010). If a particular model has been selected for biological or statistical reasons, then some processing of data may be required. For example, a Hill model was fit to the RLU responses in the current BRD. This model assumes a positive or neutral slope in the dose-response relationship. Points that deviated from a positive/neutral slope were excluded from the analysis. The rules for such preprocessing of data must be clearly stated.

The estimated values for EC_{50} and IC_{50} presented in the BRD were point estimates without any error associated with them. Uncertainty associated with these estimates should be reported (e.g., confidence intervals).

Potency estimates can be broader than just EC_{50} and IC_{50} parameters. For example, the concentration associated with a specific fold induction may be a relevant potency endpoint.

Additional considerations
The BRD should explicitly define all data transformations and normalizations. For example, the procedure to calculate the adjusted RLUs should be clearly described. The variability in the DMSO control responses appears to be ignored for the background correction of the adjusted RLUs. An alternative is to include the control response to help estimate the baseline in the concentration response model.

Approaches for assessing reliability
The appropriateness of methods used for reliability assessment is discussed in Section 6.0 (below).

c. For each set of data, is the discussion of whether coded substances were tested and whether experiments were conducted without knowledge of the identity of the substances being tested adequately documented?

The discussion of whether coded substances were tested and whether experiments were conducted without knowledge of the identity of the substances being tested is adequately documented. Section 3.9 of the BRD describes procurement, coding, and distribution. Section 4.4.2 of the BRD

states that substances in Phases 2, 3, and 4, which were critical for the evaluation of reliability and accuracy, were coded.

The only additional information that should be provided is a confirmation of the coding for the controlled substances purchased regionally. Specifically, BRD Section 3.9 states that ECVAM and JaCVAM procured "controlled substances" from their regional suppliers. This raises the question of how they ensured that the laboratories did not know the identity of the tested substances—this should be clarified in the BRD.

5. Test Method Accuracy

a. The current accuracy analysis is based primarily on overall concordance with the ICCVAM reference consensus classification, which is based on results from *in vitro* ER TA test methods. Are these data adequate for assessing the accuracy of the test method?

The ICCVAM reference consensus classification was the primary reference method used for assessing BG1Luc ER TA test method accuracy. Some additional measures of agreement were performed in comparison to the CERI STTA, RUC ER binding assay, and the uterotrophic assay. As described in Section 5.5 of the BRD, median estimated EC_{50} and IC_{50} values were compared with reference data using regression. If analysis of "agreement" between the two data sources is a major concern, statistical approaches are available that might provide more insight into the strength of agreement between assay results (e.g., Bland-Altman Plots and Limits-of-Agreement [Assessing Agreement 2007; Bland and Altman 1986, 1999]).

The endorsement of a screening assay cannot be based on its strict agreement with any one other method but should be judged on a weight-of-evidence approach that includes all performance comparisons, the repeatability results, and scientific judgment regarding the biological relevance of the test system.

b. Do you agree that accuracy of the BG1Luc ER TA test method should be based only on those substances for which an unequivocal reference classification can be assigned?

The accuracy of the BG1Luc ER TA test method should be assessed using unequivocal reference classifications from other human ER TA tests. This increases confidence in the results by eliminating questionable responses. One caution is that validation study results based upon unequivocal classifications may result in overly optimistic assessments of test method performance. Substances that result in equivocal reference classifications may provide additional insights into aspects of the test method.

c. Other concordance analyses included in the BRD are based on direct comparisons to *in vivo* (uterotrophic) or *in vitro* (CERI STTA, ER binding assay) test methods. How much emphasis should be placed on these comparisons relative to the comparisons to the ICCVAM classification?

Comparing the new test method with other methods assessing the same mode of action is reasonable. Indices of agreement should be calculated only with data that meets certain quality control measures. Comparison of data from other methods could be confounded if alternative methods are not measuring an endpoint based upon the same biological mechanism. Accordingly, emphasis should be placed on describing the different purpose of the various tests, as well as their advantages and disadvantages.

As noted above, data quality is an uncertainty for all of the reference methods used in the BG1Luc ER TA BRD.

d. Based on your responses to Questions 5 a-c, has the relevance (e.g., accuracy/concordance, sensitivity, specificity, false positive, and false negative rates) of the BG1Luc ER TA test method been adequately evaluated? If not, what other analyses should be performed?

The interpretation of the estimated error rates should be put into context. The first observation is that this set of test chemicals necessarily represented a limited set of substances. As with all previous validation studies of this type, the performance of this assay with new classes of chemicals will be difficult to predict. This is not seen as any reason for not endorsing the method for use in a test battery or other weight-of-evidence approach, but it is an important issue for users to understand.

The second observation is that confidence intervals for the calculated performance statistics, including the false positive and false negative error estimates, should be provided. Additionally, the contingency tables used to generate the summary statistics should be included in the document, including tables showing results from the individual laboratories.

The criteria for declaring a chemical as positive changed after Phase 3 was completed and were retrospectively applied to all of the previous data. The Panel recognizes that the original criteria were inadequate. This should not be a precedent going forward because optimal decision criteria should be selected *a priori*. However, for this study, the Panel affirms that the changes made were performed in an appropriate manner.

e. Do you agree that this test method is effective for generating data in an amount and of a scientific value that is at least equivalent to the data generated from existing tests (i.e., U.S. EPA Endocrine Disruptor Screening Program Test Guideline OPPTS 890.1300: Estrogen Receptor Transcriptional Activation [EPA 2009])?

The BG1Luc ER TA test method is effective for generating data that are functionally equivalent to the data generated from existing tests (i.e., U.S. EPA Endocrine Disruptor Screening Program Test Guideline OPPTS 890.1300: Estrogen Receptor Transcriptional Activation [EPA 2009]).

In fact, the BG1Luc ER TA protocol could be considered an improvement over the CERI protocol given the extent to which the BG1Luc ER TA protocol was standardized and optimized. The Panel recommends that the CERI protocol be revised to meet the standards established in the BG1Luc ER TA protocol.

f. Based on comparison of results from the ER TA with results from RUC ER binding assays, are there sufficient data to conclude that positive and/or negative results from BG1Luc ER TA can adequately identify ER binding potential compared to the RUC? If not, what other analyses and/or additional information would be necessary to draw such a conclusion?

RUC ER binding assays indicate agreement with the BG1Luc ER TA test method results and suggest that the outcomes of the stably transfected ER TA test method can provide insight regarding the biological effect of chemicals mediated by ER–ligand interactions. There appears to be sufficient data to consider replacing the rat uterine cytosol ER binding assay with the BG1Luc ER TA test method, especially if ER confirmation assays are completed as part of the BG1Luc ER TA test method procedure (see Section II.2.b.i). Therefore, the Panel recommends that NICEATM and ICCVAM identify and use other available ER binding data for these comparisons (i.e., substances beyond the 78 tested in the BG1Luc ER TA validation study). As an additional activity, the Panel recommends an evaluation of recombinant (human and animal) ER binding as a replacement for the RUC ER binding assays.

6. Test Method Reliability (Intra- and Interlaboratory Reproducibility)

a. Has the intra- and interlaboratory reproducibility of the BG1Luc ER TA test method been adequately evaluated? If not, what other analyses should be performed? Are any limitations apparent based on this intralaboratory reproducibility assessment?

The Panel combined the responses to both intra- and interlaboratory reproducibility evaluations because they are similar.

In general, insufficient data are presented to ensure that a thorough analysis of intra- and interlaboratory reproducibility has been performed. While reproducibility has been addressed on three data levels (raw data [e.g., DMSO control measurements], derived endpoints [e.g., EC_{50}], and prediction of estrogenic activity), the analysis is descriptive. No formal inference has been carried out. The descriptive approach is acceptable; however, some additional analyses could be performed to better understand the components of assay response variability.

Many of the figures in the BRD do not adequately present the variability associated with the test method and in fact may confuse variability with outliers. For example, it is better to display a scatter plot of data instead of means and standard deviations when only three data points are present at each concentration. Sources of variability could be explored using appropriate methods such as descriptively comparing coefficients of variation (CVs) or using more formal statistical methods to estimate variance components.

The potency discussion in the BRD did not capture the variability in these quantities. This could be captured by calculating the mean and standard deviation of the log EC_{50} for each chemical in the BG1Luc ER TA validation study. These quantities could then be compared to the analysis of other established test methods considered acceptable that also include these calculations (e.g., CERI STTA).

The rationale for selection of substances used to evaluate intra- and interlaboratory reproducibility should be more clearly described in the BRD. It should be noted that excluding substances from the analysis of reproducibility for which definite results were not determined may produce an assessment of reproducibility that is too optimistic.

7. Data Quality

a. Is the extent of adherence to national and international Good Laboratory Practice (GLP) guidelines for all submitted test data and the use of coded substances and coded testing adequately presented?

Validation guidance states that “Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with Good Laboratory Practice (GLP) guidelines. Aspects of data collection not performed according to GLPs must be fully described, along with their potential impact” (ICCVAM 1997). Two out of the three laboratories in this study followed GLPs, and the third laboratory was ISO 9000 certified. Therefore, the extent of adherence to national and international GLP guidelines is adequately presented. The use of coded substances and coded testing in the validation studies is adequately described. Deviations and alterations to original protocols including use of coded substances were relatively well described, such as acquisition of controlled agents (e.g., androgens, barbiturates, etc.) for testing.

However, while training of laboratory personnel was reportedly conducted, details that would be included in formal training records were not available to confirm compliance with GLPs and Good Cell Culture Practices (GCCP).

An essential subsection is missing from BRD Section 7.0 that should include a description of the quality of the reference data. Several reference methods were used to assess BG1Luc ER TA test method accuracy and concordance. Therefore, the "quality" of the reference data for the following methods should be summarized in the BRD: CERI STTA, data used to generate ICCVAM reference classifications, uterotrophic assay, and RUC ER binding assay. If this information is incomplete or unavailable, it should be explained in the BRD. As discussed in Section 3 of the BRD, regarding the ICCVAM reference classifications, consideration could be given to assigning greater weight to certain reference data, and a thorough description of the weighting methodology should be provided.

Section 7.2 of the BRD states that an independent quality assurance (QA) review was conducted at each of the participating laboratories. This should have included a process audit of the assay at each laboratory (at the time of testing) to ensure that standard operating procedures (SOPs) were followed. Any data transfers made during calculations and statistical analysis also need to be audited to ensure that errors are caught and corrected. The responsible QA personnel in each lab should also be identified in this section. Based on the information provided in the BRD, it is not clear if all of these steps were performed. All audits should be documented in an appendix to the BRD. Ideally, an independent audit of all data from all laboratories should be conducted upon the conclusion of an interlaboratory validation study.

The following statement is included in Section 7.4 of the BRD: "since the updated classification system was developed after testing was complete, these substances were not retested." This seems problematic from a validation study perspective. However, as previously discussed in the response to Question 5d, the Panel affirmed that the changes made were performed in an appropriate manner and recommended that the explanation be added to the BRD.

Section 7.5 of the BRD should include a statement about the availability of audit results, statistical evaluations, and methods for calculations. Additionally, the time requirements and location(s) for study data retention/storage should be described, as well as the secondary location for storage of backup copies of study data.

8. Consideration of All Available Data and Relevant Information

a. Based on available information contained in the draft BG1Luc ER TA BRD, have all the relevant data identified in published or unpublished studies that employ this test method been adequately considered? Are there other comparative test method data that were not considered in the draft BRD but are available for consideration? If yes, please explain how to obtain such data.

To the extent the Panel can determine, all the relevant data identified in published studies that employ this test method have been adequately considered. The Panel is not aware of any existing unpublished studies.

Quantitative structure-activity relationship (QSAR)-based predictions for ER binding on some of the substances tested in this study are likely available in the literature, and a comparison to BG1Luc ER TA and RUC ER binding results would be of value.

9. Animal Welfare Considerations

a. Is the extent to which the BG1Luc ER TA test methods will reduce, refine, or replace animal use adequately characterized and discussed? If not, then what should be added?

The extent to which the BG1Luc ER TA test methods will reduce, refine, or replace animal use requires further discussion. In order to fully understand how this method will impact the 3Rs, the

context into which this test method would fit into the overall EPA EDSP Tier I screening battery is needed. For example, if the BG1Luc ER TA test method is just added to the battery of individual screening assays, the BRD should simply state that it would not reduce animal use at this time. However, additional consideration of strategic testing schemes provides possible approaches by which a validated BG1Luc ER TA test method could contribute to reducing animal use. For example, these could include:

- Defining whether the BG1Luc ER TA test method will be performed prior to or simultaneously with the other *in vitro* and *in vivo* assays
- Defining the implications for possible Tier II testing, which will likely incorporate significant animal use

Comments provided in the EPA's 2011 budget (EPA 2011) state the EPA's intent to phase out the EDSP and to include high-throughput ED-detecting assays in the ToxCast screening battery, which could lead to reductions in animal use for regulatory testing over the long term. The BG1Luc ER TA test method could contribute to the development of this screening battery.

A major limitation of *in vitro* EAC testing and the EDSP screening battery is the lack of an *in vitro* method to evaluate metabolism. The inclusion of a metabolism component as part of any *in vitro/in silico* EAC test battery will provide a more biologically relevant assessment of EAC activity. Because there are many *in vitro* metabolism methods being used in drug development and being used for *in vitro* testing of other toxicity endpoints (e.g., genotoxicity), the importance of including metabolism in all future assays for *in vitro* EAC activity needs to be discussed in the BRD.

Implications of a validated BG1Luc ER TA test method for use in the European Union and Japan were not discussed in the BRD. An evaluation of potential EACs is required under the EU chemical regulation REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) (Bars et al. 2011; European Commission 2006; Kovarich et al. 2011; Marx-Stoelting et al. 2011), and validated *in vitro* methods may reduce animal use in this kind of testing program. "REACH-type" programs are also being adopted by Asian countries, so the availability of validated *in vitro* and *in silico* methods to screen/prioritize chemicals for these testing programs has further potential to reduce animal use. Following validation, the development of *in vitro* EAC assays into an OECD Test Guideline will broaden their potential for reducing animal use.

The concordance of the BG1Luc ER TA test method with the RUC ER binding assay of 97% (33/34 substances) suggests the BG1 assay could "replace" the rat ER binding assay within the EDSP Tier 1 battery (and similar test batteries). This excellent concordance of the BG1Luc ER TA test method exceeds the "performance" of other methods that have been endorsed as scientifically valid. If necessary, additional retrospective data analyses could be conducted. This topic was extensively discussed by the Panel, and additional Panel recommendations are provided in the Overview and Sections II.1.d and II.5.f of this report.

Based on the concordance of the BG1Luc ER TA test method with the rat uterotrophic assay (92%, 12/13), the BG1Luc ER TA test method is a potential candidate for reducing the use of the uterotrophic assay, which in turn would reduce animal use. The small data set, however, is not sufficient to recommend endorsement of BG1Luc ER TA as a replacement at this time. Therefore, the BG1Luc ER TA should be considered as a high priority for additional studies. A retrospective analysis may be sufficient. However, if necessary, a prospective study could be conducted to further compare these methods, preferably by identifying additional materials already evaluated in the uterotrophic assay. *In vitro* metabolism is an essential component of any prospective study that compares the BG1Luc ER TA to an *in vivo* assay. Also, since the BG1Luc ER TA was already subjected to an extensive interlaboratory study, consideration for an abbreviated assessment (e.g., 1 or 2 laboratories) could be considered to reduce time and costs.

Comparisons of the BG1Luc ER TA test method to the uterotrophic bioassay (or any other *in vivo* test results) would benefit from including analytical methods to measure the extent of free chemical exposure to target organs *in vivo* or to the cells in culture (Gülden and Seibert 2003).

The following text is suggested as a concluding paragraph for BRD Section 9:

> The development of a battery of *in vitro* and *in silico* methods that can replace animal testing for detecting chemicals that have the potential to interact with the endocrine system (i.e., endocrine active chemicals [EAC]) is a biologically complex problem. For example, a method for assessment of metabolites needs to be included with the *in vitro* assays, and assays for assessing the many modes of action of EACs on various tissues and species need to be developed and validated. The experience derived from validating and using the *in vitro* BG1Luc ER TA test method is expected to contribute to our knowledge and promote progress toward this goal. It should lead to the broader use of cell-based methods for EAC screening, and could include the use of cells from other species.

10. Practical Considerations

a. Are the practical considerations associated with the BG1Luc ER TA test method adequately described and are there any points that would seem to preclude its transferability and implementation by other laboratories?

There are a few practical considerations that should be addressed:

- There needs to be better clarification regarding the availability of the BG1Luc4E2 cell line. This cell line has not been placed in a public repository and therefore additional efforts should be undertaken to do so or to otherwise ensure the continued availability of the cell line. It is the understanding of the Panel that the BG1Luc4E2 cell line is available upon request for a reasonable fee. Written confirmation of this needs to be obtained, and indicated as such in the BRD.
- The costs of equipment and supplies are provided, but these costs should be associated with the date they were acquired (i.e., it is not clear if the costs provided are current). This information should be updated in the BRD. Additionally, it should be clarified that the vendors used during this update are suggestions only and that equipment and supplies are not restricted to specific vendors.

b. Are the apparent level of training and expertise required to conduct the BG1Luc ER TA test method reasonable for its wide use?

The level of training and expertise required to conduct the BG1Luc ER TA test method are reasonable for its wide use. In order to minimize variability, the protocol should place greater emphasis on the need for adherence to Good Cell Culture Practices, Good Laboratory Practices, training of personnel and contamination control. Best practices for data analysis, presentation and data illustration (graphs, tables, etc.) should also be used.

III. Draft ICCVAM Test Method Recommendations on the BG1Luc ER TA Test Method to Identify Substances With Estrogen Agonist and/or Antagonist Activity

1. Test Method Usefulness and Limitations

a. Do you agree that the available data and test method performance (accuracy and reliability) support the ICCVAM draft recommendations for the BG1Luc ER TA test method in terms of the proposed test method usefulness and limitations?

Based on overall test method performance as presented in the BRD, the BG1Luc ER TA test method is capable of identifying substances as potential *in vitro* ER agonists or antagonists, as is recommended by ICCVAM, provided the acceptance criteria described in the recommended test method protocol are met.

However, careful analysis of cytotoxicity is critical for correctly interpreting results, particularly with regard to the antagonist protocol. While it appears the qualitative approach can be used successfully for some chemicals, as indicated in the responses to Question II.4, a quantitative approach is recommended as the primary method for evaluating cytotoxicity. In addition, accessibility to the cell line still needs to be resolved, which clearly can impact the future use of this assay, including its use as a screening test in a contract laboratory setting.

The Panel also noted that limitations due to interlaboratory variability may not be fully explained by the data analysis provided in the draft BRD. In addition, test method accuracy statistics (concordance, false positives and negatives, etc.) need to be updated to include the confidence intervals.

If ICCVAM concurs with the Panel's recommendations, then the recommendation that BG1Luc ER TA be considered for replacing the RUC ER binding assay in test batteries and tiered testing schemes such as EDSP Tier 1 battery needs to be moved to this section of the ICCVAM recommendations.

2. Test Method Protocols

a. Do you agree that the available data support the ICCVAM draft recommendations for the BG1Luc ER TA test method procedure in terms of the proposed test method standardized protocols? If not, what recommendations would you make?

The available data support the ICCVAM draft recommendations for the BG1Luc ER TA test method procedure in terms of the proposed test method standardized protocols with the caveat that the visual assessment of cytotoxicity needs to be better standardized and validated for chemicals or replaced by a quantitative method (see related comments on the test method protocol in Section 2b and below relative to future studies). Additionally, appropriate quality control measures should be further emphasized in the protocols to ensure reproducibility (i.e., GLPs, GCCPs, training, and contamination control).

3. Future Studies

Do you agree that the available data support the ICCVAM draft recommendations for the BG1Luc ER TA test method in terms of the proposed future studies? If not, then what recommendations would you make? Please explain your answer.

The available data support the ICCVAM draft recommendations for the BG1Luc ER TA test method in terms of the proposed future studies. Before additional work is recommended, it is important that all efforts have been made to make use of information that is available.

In addition to those proposed studies, recommended future studies include the following:

- Efforts to assess the utility of the current visual assessment of cytotoxicity evaluation for chemicals, or a quantitative method, should be assessed for its replacement.
- Future studies to account for compound metabolism/activation could expand the utility of this and other ER TA methods.
- A search for fully defined media to replace fetal bovine serum is recommended.
- A study to assess protocol components that might impact intra- and/or interlab variability in BG1Luc ER TA test results is recommended.
- An effort to expand the reference substance list, and subsequently the BG1Luc ER TA test results, with additional negative agonist and positive antagonist test substances is recommended. Pazos et al. (2010) is a possible source for additional reference substances.

4. Performance Standards

ICCVAM has developed draft performance standards consisting of essential test method components, a minimum list of reference substances, and expected accuracy and reliability values. These are proposed for evaluating the acceptability of proposed test methods that are mechanistically and functionally similar to the BG1Luc ER TA test method. **The overall question for the Panel is do you consider these performance standards adequate for assessing the accuracy and reliability of test method protocols that are based on similar scientific principles and that measure the same biological effect as the BG1Luc ER TA test method**?

a. Do you agree with the selection and prioritization criteria used to select the performance standards reference substances?

In general the criteria used to select the performance standards reference substances are adequate. As previously noted, the use of the majority classification criteria (i.e., greater than 50%) among results to establish the consensus reference determination that was assigned to each reference substance is a reasonable strategy. However, the criteria used to evaluate the quality of the data obtained from the literature to determine the reference classification could impact the appropriateness of such a strategy. The use of a ranking method such as using Klimisch criteria (Klimisch et al. 1997), which focuses essentially on the reliability of the data, would strengthen the resulting activity determination. In addition, some type of sensitivity analysis could be performed that would be more robust than a simple majority classification.

b. Do you consider the number of substances included in the list of reference substances to be an adequate number upon which to evaluate the performance of functionally and mechanistically similar test methods? If not, how many reference substances should be tested?

Based upon the currently available data, the list of reference substances upon which to evaluate the performance of functionally and mechanistically similar test methods is adequate. Ideally, more negatives should be included, but the Panel recognizes that data on such substances are not currently available. See additional discussion below.

c. Do you consider the types of substances included in the reference substance list, with regard to relative estrogenic activity and physicochemical characteristics to be representative of the overall diversity of substances that are likely to be tested for potential estrogenic activity?

The reference substance list has reasonable overall diversity and reflects the extensive effort to obtain all available relevant information. The list allows for a wide range of estrogenic activity over several orders of magnitude, as well as a few confounders to assess the robustness of the assay and its methodology (e.g., TPA, sodium azide, ammonium perchlorate, cycloheximide, and actinomycin D). However, there may be an opportunity to revisit the list of reference substances and make modifications based on experience gained in the assay subsequent to future testing. For example, more negatives and proportionally fewer positives should be included in the list for agonist testing when possible.

d. Are there other types of information relevant to estrogen agonist or antagonist activity that should be considered in order to demonstrate an adequately diverse reference list? If yes, please explain what additional information should be included.

As noted above, the reference list is adequately diverse, but it would benefit from the inclusion of additional negative substances once they become available.

e. "Discordant chemicals" are also included in the reference list as substances that could be studied to evaluate if the proposed modifications might provide improved performance relative to the BG1Luc ER TA test method. Please comment on the appropriateness of including these specific substances in the reference list. Should more and/or different substances be included? If so, how many more and what are they? Should testing these substances be required?

Discordant chemicals should be included on the reference substance list. These discordant chemicals, which include weakly active or non-active chemicals, transcriptional inhibitors, and general cytotoxicants, are critical for truly characterizing the limitations of the assay. While more compounds of a similar type might have some added value, there are enough to lend sufficient robustness to an assessment. In order to place any "discordant" results into proper context, some metric of potency and intrinsic activity should be included in the accuracy evaluation.

f. Are there any substances on the proposed reference list for which a discordant result would be considered unacceptable and would therefore signal that a proposed test method is not considered scientifically valid, regardless of its overall performance?

The potent estrogens on the reference list should not be missed. There could be some tolerance for discordance for the weakly active reference substances. Therefore, any "discordant" results should be discussed in terms of the ability of the test method to detect a similar range of potencies and intrinsic activity, as well as the chemical/product class. This could also facilitate the interpretation and utility of the data in a possible "weight of evidence" assessment of comparative assays.

g. Do you consider the number of repeat experiments to be adequate to evaluate intra- and interlaboratory reproducibility?

The number of repeat experiments to evaluate intra- and interlaboratory reproducibility appears to be a good start. The intralaboratory and interlaboratory reproducibility assessments allow for a reasonable evaluation of the methodology on the validation reference set of materials.

However, as discussed in Section II.6 of this report, while reproducibility has been addressed on three data levels (raw data [e.g., DMSO control measurements], derived endpoints [e.g., EC_{50}], and

prediction of estrogenic activity), the analysis is descriptive. No formal inference has been carried out. The descriptive approach is acceptable; however, there are some additional analyses that could be performed to better understand these components of assay response variability.

5. Additional Comments on Essential Test Method Components

The justification for some of the essential test method components is not clear. For example:

- The type of cell line does not have to be specified in the performance standards. It is critical to ensure that any cell line used incorporates the appropriate "endogenous cellular machinery," as stated in the BRD, but the specific tissue source, type, and species may not be critical.
- Evaluation of cytotoxicity is necessary. The Panel recommends a quantitative approach be used when developing new ER TA methods, particularly with regard to an antagonist protocol, where false positives can result from cytotoxicity.

6. Additional Comments on Test Method Performance

The criteria for accuracy included in the performance standards should be based on meeting or exceeding the lowest accuracy that was noted among the participating laboratories. As noted previously, the Panel strongly supports the quantification of activity as a complement to the dichotomous call of a positive or negative response. Interpretation of the results should not rely solely on statistics but also on scientific judgment and should incorporate consideration of the nature and shape of the dose-response relationship and, if needed, the reproducibility of the response in independent experiments. A test of trend might be preferred to the BRD's suggestion of "three points with non-overlapping error bars."

- The BRD should define what statistic corresponds to the error bars included in the criteria for a positive response (i.e., standard deviations, standard errors, half-width of confidence intervals). Additionally, the idea of nonoverlapping confidence intervals is a conservative way to declare difference between parameters and is not clear guidance.
- A metric of potency and intrinsic activity should be included in the accuracy evaluation; appropriate measures of uncertainty should also be included. Therefore, the reporting requirements should include the metric of potency that is used.
- Sensitivity analyses based on EC_{50} (agonist) or IC_{50} (antagonist) should also be considered.

IV. References

Assessing Agreement (Special Issue). 2007. J Biopharm Stat 17(4).

Bars R, Broeckaert F, Fegert I, Gross M, Hallmark N, Kedwards T, et al. 2011. Science based guidance for the assessment of endocrine disrupting properties of chemicals. Regul Toxicol Pharmacol 59(1): 37-46.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1: 307-310.

Bland JM, Altman DG. 1999. Measuring agreement in method comparison studies. Stat Methods Med Res 8: 135-160.

Bretz F, Hothorn LA. 2003. Statistical analysis of monotone or non-monotone dose-response data from in vitro toxicological assays. Altern Lab Anim 31(Suppl) 1: 81-96.

EPA. 2009. Endocrine Disruptor Screening Program Test Guidelines. OPPTS 890:1300: Estrogen Receptor Transcriptional Activation (Human Cell Line (HeLa-9903)). EPA 740-C-09-006. Washington, DC:Office of Prevention, Pesticides and Toxic Substances, U.S. Environmental Protection Agency.

EPA. 2011. FY 2012 EPA Budget in Brief. EPA-190-S-11-001. Washington, DC:U.S. Environmental Protection Agency, 60-62.

European Commission. 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L 396.

Gülden M, Seibert H. 2003. In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations in vitro. Toxicology 189: 211-222.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. Research Triangle Park, North Carolina:National Institute of Environmental Health Sciences.

ICCVAM. 2003. ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. NIH Publication No. 03-4503. Research Triangle Park, North Carolina:National Institute of Environmental Health Sciences.

ICCVAM. 2006. Addendum to ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. NIH Publication No. 03-4503. Research Triangle Park, North Carolina:National Institute of Environmental Health Sciences.

ISO. 2009. ISO 10993, Biological evaluation of medical devices—Part 5:Tests for *in vitro* cytotoxicity. Geneva:International Organization for Standardization.

Klimisch HJ, Andreae E, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental and ecotoxicological data. Regul Toxicol Pharmacol 25: 1-5.

Kovarich S, Papa E, Gramatica P. 2011. QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants. J Hazard Mater 190(1-3): 106-112.

Marx-Stoelting P, Pfeil R, Solecki R, Ulbrich B, Grote K, Ritz V, et al. 2011. Assessment strategies and decision criteria for pesticides with endocrine disrupting properties relevant to humans. Reprod Toxicol 31(4): 574-584.

Pazos P, Pellizzer C, Stummann TC, Hareng L, Bremer S. 2010. The test chemical selection procedure of the European Centre for the Validation of Alternative Methods for the EU Project ReProTect. Reprod Toxicol 30(1): 161-199.

Ritz C. 2010. Toward a unified approach to dose-response modeling in ecotoxicology. Environ Toxicol Chem 29(1): 220-229.

Rogers JM, Denison MS. 2000. Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals. In Vitr Mol Toxicol 13: 67-82.

This page intentionally left blank

Appendix A

Errors and Omissions

The Panel has identified several errors that should be corrected, as well as omissions of existing relevant data or information that should be included. These are detailed below.

General Format

In general, table titles are not explicit enough and should better describe the content of the tables.

Throughout the text the use of the wording "normal" should be discouraged:

- "… the ~~normal~~ function …".
- "… interfere with ~~normal~~ estrogen activity. …"

Preface

Page xvii
Spell out ER, AR and TA at first use; add citations for BRDs (3rd line up from bottom)

Page xvii (last paragraph)
Add "identification of substances with *in vitro* ER agonist or antagonist activity, *as predictive of activity in vivo*".

Page xviii - top line
Add citation

Page xviii - last paragraph statement
"...for identification of substances with ***in vitro*** ER agonist or antagonist activity"
Suggest deleting "*in vitro*" from this and other similar statements in the BRD as this conflicts with your own definition of the scientific basis for the BG1Luc ER TA Luc ER TA on page 1-11, which says the in vitro ER TA assays are to identify... that might interfere with normal estrogen activity ***in vivo***. Including "*in vitro*" also conflicts with the original described goals of the EPA for this kind of assay to be validated "to determine whether certain substances may have an effect in humans...." - see page 1-2.

Page xix - 6th line from top
"these proposed..." should be "the proposed...."

Page xix
The second charge to the panel (3rd line from top) is "assessing the extent that established validation and acceptance criteria have been adequately addressed." To do this NICEATM/ICCVAM selectively provided the peer panel with a list of questions that effectively cover some of the criteria but not others. The second question in this set of "questions to the panel" should list the validation and acceptance criteria, so that the panel members can easily review them and respond effectively to this second charge identified for the panel. The remaining NICEATM questions can then follow.

Executive Summary
The Executive Summary should be a stand-alone document and provide a clear explanation of how the assay is used to identify a substance as positive or negative as well as how substances were classified as positives or negatives. This information will improve the reader's understanding of results.

Page xxi
Spell out ER, AR and TA at first use; add citations for BRDs (first paragraph)

This appendix documents errors and omissions identified by the Panel, as communicated to NICEATM. Editing of Panel comments by NICEATM has been limited to only that necessary to ensure clarity.

Page xxii (end of paragraph just before protocol section)
The ES does not make it clear to a reader, as a stand-alone document, the role of the peer review and the Panel and that this is a draft BRD. Therefore, add statement something like this to the end of the paragraph: "This draft version of the BRD will be reviewed by an international independent scientific peer review panel (Panel) that will meet on March 29-30, 2011. The Panel's comments will be considered by ICCVAM before ICCVAM recommendations on the BG1Luc ER TA test method are finalized."

Page xxii
Protocol section should briefly explain how the assay is used to identify a substance as positive or negative, and why the qualitative endpoint is being used for the proposed validated assay.

Page xxii-xxiii - Substances used
Add citation for identification of 78 substances; add statement explaining how substances were classified as +/- for ER TA activity; the numbers don't add up - make it easier for reader to understand the numbers of positives and negatives identified and used in this study.

Page xxiii - Accuracy section
More clearly define the reference method - "preponderance of published data" - of what kind of data? Average response from 3 labs should not be used. In real testing one lab will be used, therefore each lab should be assessed independently for correlation to reference method(s). First time CERI STTA is mentioned it should briefly be explained as the only currently accepted *in vitro* method.... The numbers in this section don't add up without further explanation - check their correctness and simplify this for the reader.

Page xxvi (last paragraph)
Add "antagonist" to this statement: "....identify substances with estrogen agonist and antagonist activity."

Page xxvi (last paragraph)
Add statement something like this to the end of the page: These ICCVAM draft recommendations will be reviewed by the scientific peer review panel at the March 2011 meeting, and the Panel's comments will be considered by ICCVAM before they are finalized.

1.0 Introduction

The Introduction should provide the advantages of the BG1Luc ER TA test method. Included in this information, this section should include some of the background behind the preference for a TA assay over a binding assay (there are parts to this here and there) but might consider pulling some from the Rogers paper:

Binding assays and caveats leading to why TA assays are better than binding assays (agonist and antagonist activity can be assessed, done at physiological temperatures etc).

Transient transfection assay vs. development of a stable cell line and why a stable cell is better.

For this ER TA cell line the properties that make it best suitable for this use (little metabolic activity, appropriate transcription machinery for hormone responsive cells, no other steroid hormone receptors are expressed to confound the assay, etc.)

Page 1-2: 3rd line
Delete "*in vitro*" for same reason previously discussed

Line 33-35 in the Public Health Perspective section (1.2)
Says "EDs are widespread in our environment…", though most of the substances have not yet been judged to be endocrine disrupting, since the definitive test methods have not been established. Diethylstilbestrol is obviously an endocrine disruptor by its clinical and epidemiological evidence,

and bisphenol-A may deserve to be an ED by many supporting experimental results, but other chemicals are in purgatory so far. Thus, “EDs are widespread …” should be “Potential EACs are widespread…”

Page 1-2: Historical Background
Lists 3 Acts as requiring EPA to develop EAC test methods, and then the next statement refers to only one Act. Please clarify and correct.

Pages 1-2 and 1-13
Remove the Fenner-Crisp and Fisher reference, which appears to be only an abstract and not needed to define EACs.

The BG1Luc ER TA Luc ER TA should be defined as an appropriate screen (section 1.5.2) to be integrated in an accurate, comprehensive and cost-effective *in vitro* testing battery (section 1.5.5)

Section 1.4 describes that the interlaboratory validation study was conducted by three laboratories in three different regions. Why they were restricted to three, and do three suffice the validation? I am not making an objection but want to confirm the background opinion.

Page 1-5
Delete “academia” from 4th paragraph unless you can identify academic labs that conduct high throughput testing of chemicals.

Page 1-6: section 1.5.2
Move this statement “An appropriate screen such as....” to the end of the section, and revise as follows: Depending on how it is used, an appropriate screen such as BG1Luc ER TA Luc ER TA has the potential to limit human and ecological.....

Page 1-7: section 1.5.3
Not sure last statement is correct - how would BG1Luc ER TA assay eliminate need for testing in these [Tier 1?] animal models?

Page 1-3
A section titled “Regulatory rationale for BG1Luc ER TA” should be added to the Introduction section. The EPA EDSP is the motivation for this entire decade-long validation process and relevant information from the EPA EDSP should be summarized in Section 1.0, including a summary of the regulatory use of the current and the proposed EDSP test batteries. Especially relevant are: a) the recent policy document regarding how to use Tier I data - this should be summarized for easy understanding, b) the 11 Tier 1 assays were [pre?]validated by the EPA/OECD, and c) when is Tier 2 testing required? The validation of the CERI ER TA reference method should be described and references to relevant documents provided. Current and future OECD TG activities could be described in this section, as well as EU and Japan regulatory implications.

Section 1.7 describes more precisely the interlaboratory validation study. It is not clear how the laboratory qualification was done. It seems to have been done in the Phase 2 depicted in Fig. 1-1, but it is better to be stated also in the text.

Page 1-8: section 1.5.5
Seems to exaggerate savings; BG1Luc ER TA as part of a battery of in vivo and in vitro tests would not provide cost and time savings; its potential use as a screening assay compared to an animal test might....If BG1Luc ER TA leads to development and validation of a more complete battery of in vitro methods or tiered testing scheme then the potential to reduce animal use would be more likely.

Page 1-9: section 1.7
The statement about KoCVAM should be removed. Other table entries are not described in this degree of detail in the text so it seems out of place.

2.0 Protocol Components

Discuss the sequence of the vector itself and that it has been closely examined for the presence of cryptic activation sites which could confound the data (I assume this was done as with computers it is quite easy to assess – not like the old days). The use of 4 ERE's with varying orientations in the construct design should be discussed.

A section on reference method protocol(s) is missing from the BRD.

Protocol section should clearly explain when over the course of the study changes were made to the protocol and test method evaluation criteria, and why these changes would not affect study results that are being evaluated to assess the scientific validity of the assay. Validation study guidance states that: "Prevalidation is the process by which testing laboratories are selected and demonstrate competence in performing the testing procedures, and during which the test protocols are standardized. It is important that this be established in advance of formal validation procedures" (http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf, page 33).

Section 2.3: Preparation and use of cells
The BRD protocol section does not adequately describe the cell culture procedures (split ratios, number of cells seeded per well, storage conditions, passage numbers acceptable for validated assays, etc.)

Section 2.2.1
Explain how the cell line was characterized including relevant citations, and state what components are critical to the development of a "replacement" cell line (cell line characterization criteria). This information is needed for development of the performance standards.

Section 2.2.1
Should identify source(s) for the cell line.

Table 2-4
Add column showing corresponding % cell viability.

3.0 Substances Used

L7: complete the dotted line.

L8: "… EA TA …" should be ER TA.

L10: "… EA TA …" should be ER TA.

L13 and 14
Consistency between both titles "… to assess ..."
Also correct the title accordingly in the text.

L27
Citation order 2002 a to d.

Page 3-9, line 54-55: add citation.

Footnote of table 3-2

L96-97: ER = ~~endocrine~~ <u>estrogen (?)</u> receptor

L100, 101, and 102: (<u>2010</u>) Could this citation be incomplete?

L147: "… EA TA …" should be ER TA.

L156: …~~ and vise versa~~… delete.

L176: "… EA TA …" should be ER TA.

L190-191: "… in conjunction with CERI …"

L194-199: The number of substances listed by the bullets is 13, not 15 (?).

Table 3-3, 3-4, 3-5, 3-6
Add a footnote to explain the column "ICCVAM Consensus Classification". For example: estrogenic activity based on a literature review for effects on ER binding, ER-TA based on CERI, and uterotrophic response.

Consider the following to reduce the number of tables: The only new information in Table 3-5 and 3-6 is the Study Phase. Table 3-5 and 3-6 can be deleted and replaced by a single column inserted in Table 3-3 and 3-4. There is enough space to insert this column in Table 3-3 and 3-4. For example, 17b-estradiol is listed as agonist in Phase 1 and 3, therefore in this new column the number "1,3" can be added beside 17b-estradiol.

A footnote can be added to the table to explain the implication of the Study Phase.

In general the Table titles poorly describe the content of the tables.

L221: "Substances Used to Assess ~~for~~ …". To be consistent with title 3.6.

L232-250
If needed, this is an area where the length of the document can be reduced by presenting Section 3.7 and 3.8 into an annex. This new annex can be cited with the description of Table 3-1. The information presented in section 3.7 and 3.8 is redundant with the information already presented in Table 3-1. The text in both sections (3.7 and 3.8), as well as the "incomplete" titles of Table 3-7 and 3-8, failed at justifying the repetition of the same data but in a different format. It is understood that the idea is to demonstrate that the validation process involved a wide distribution of chemicals based on chemical and product classes, but Table 3-1 was sufficient to achieve this. In addition, the unavoidable fact that the same chemical can be assigned to more than one category is reducing the importance of these tables.

L262: methyl-testosterone

In section 3.9 it says that ECVAM and JaCVAM procured "controlled substances" from their regional suppliers. This raises the question how ECVAM ensured that the identity of the tested substances was not known to the (in-house?) ECVAM laboratory. This should be clarified in the BRD.

4.0 Test Method Data and Results

L20-21: Delete, not useful.

L76-77: "Hiyoshi reported the lowest values …". This is not obvious, is this useful?

L103: "… ~~(i.e., the methoxychlor control must be positive).~~"

L147
Not clear, rewrite adjustment calculation. What is meant by "the maximum Ral response to 10,000…". Perhaps an example could be provided.

L160-161: Not clear. Perhaps an example could be provided.

L200 and L204: Flouranthene should be Fluoranthene.

Table 4-11. "Agonist Range Finder Starting Concentrations ~~Based on Solubility~~ in the culture medium".

Table 4-12 "Antagonist Range Finder Starting Concentrations ~~Based on Solubility~~ in the culture medium".

L224: "… comprehensive testing. This ~~which~~ was …"

L231: The reference to Table 4-11 is not appropriate, it should be 4-13.

L235: The reference to Table 4-12 is not appropriate, it should be 4-14.

L250-251: The table numbers are wrong and should be 4-13, 4-14, and 4-15.

L266: this footnote should refer to IC50 not EC50.

Section 4.4.1
According to table 4-14 substance 17-alpha-estradiol was not "negative" but "inadequate" at Hiyoshi.

Section 4.4.1
According to table 4-14 substance clomiphene citrate was not "negative" but "inadequate" at XDS.

Table 4-14: Coumestrol at ECVAM is N(0/2). Why is it a "negative" if no runs support this result?

In table 4-14
ECVAM classification of p-n-nonylphenol is N (2/3), but in 6.1.7 it says that there were no "inadequate" data generated at any lab during analysis of antagonist reference substances and that there was 100% agreement within each lab for each of the three repeat tests. But does N(2/3) not mean +-- or I--?

Table references in section 4.4.1

5.0 Accuracy

Page 5-2
Would be good to start this section out with the definition of accuracy from validation guidance document and a citation to the guidelines you used (ILS said OECD).

Page 5-2, lines 32-34
Pos/Neg calls were based on average results from 3 labs does not appear valid way to evaluate the data. When an organization submits a sample for testing, it will be submitted to one lab. They need to know what confidence they can have in getting the correct result from one lab (this issue overlaps with inter-lab variability assessment). A justification could be included here explaining that the BG1Luc ER TA is being evaluated as a screening assay rather than a replacement assay, and that there is usually no exact way to assess performance of an assay that is part of a battery or tiered scheme. Additionally, NICEATM/ILS developed the novel approach of developing weight-of-evidence results from the literature, etc.

Pages 5-7 and 5-8
Replace Tables 5-4 and 5-6 with tables presented at meeting showing accuracy calculations for individual labs and combined.

Table 5-2
Discuss why negative substances are likely to be I or NT. Why are there so few negative substances tested and how will this effect the results?

Table 5-7: alpha symbol for estradiol and ethinyl estradiol.

Page 5-7, line 100: after "antagonist testing" add (Table 5-3)

L130: "… listed in Table ~~5-5~~ 5-7."

L136: "… listed in Table ~~5-5~~ 5-7."

L151
The following addition can be considered: "Although EC50 values can differ by a log between methods (even two logs for norethynodrel), this relatively high correlation …"

Table 5-10
The origin of the EC50 values can be added as a footnote (median XDS, ECVAM, Hiyoshi (?)).

L203: Owens and Koëter 2003 (same modification required in the list of references).

In appendix B it says "The model calculates the best fit for the Top, Bottom, HillSlope, and EC50 parameters. See Section 11.6.5 for more details." Where is section 11.6.5?

Classification criteria are given in section 2.7.1.5. Please explain what you mean by "error bars". Standard deviation? Standard error?

Table 5-2
Inconsistencies with table 4-13, e.g. 17-alpha-Estradiol POS(2/2) at Hiyoshi in table 5-2, but P(1/1) in table 4-13.

Table 5-3
Inconsistencies with table 4-14, e.g. 4-Hydroxytamoxifen POS(1/1) at XDS in table 5-3, but P(3/3) in table 4-14.

Table 5-7: explain how BG1Luc ER TA and CERI results were determined.

Tables 5-8 and 5-9
Errors - Table 5-7 lists 27 total test substances (not 26 like used in the tables) and 5 negatives (not 4 like used in these tables). If correct, the calculations shown need to be revised and the numbers reported on lines 117-119 corrected.

Section 5.7
Mention and cite results of additional paper comparing Lumi-Cell assay with uterotrophic assay (Jefferson, Padilla-Banks et al. 2002) that was mentioned in the XDS submission (Annex A).

There have been sufficient toxicity test method validation studies resulting in endorsed alternative methods so that the conclusions for BG1Luc ER TA accuracy should be discussed in terms of how the BG1Luc ER TA performance statistics compare to previously endorsed *in vitro* test methods.

6.0 Test Method Reliability

Section 6.0, lines 19-20
Cite validation guidance document rather than secondary source for definition of reliability.

Section 6.1.3
Add table showing mean, range, and CVs for agonist test substances in the 3 labs (or refer to tables in Section 4.0). These are the real intra-lab variability data to be evaluated in this section of the BRD.

Section 6.1.7
Add table showing mean, range, and CVs for antagonist test substances in the 3 labs (or refer to tables in section 4.0). These are the real intra-lab variability data to be evaluated in this section of the BRD.

Section 6.2.2 (agonist phase 3)
Says that there was discordance among the laboratories for six substances: dicofol, fluoranthene, butylphenol, androstenedione, clomiphene citrat, resveratrol. What about dexamethasone? For this substance definitive results are available in table 4-13 for ECVAM and Hiyoshi, but results are discordant. This would change table 6-9 as well.

Footnotes of table 6-3
Says "each of three replicate tests". But in some cases more than three replicates are available, e.g.

Corticosterone at Hiyoshi (Table 4-13). This means that for different compounds certainty about reproducibility is different. Same issue in table 6-7.

Section 6.1.5, Table 6-5: add units.

Section 6.1.6 line 185: "As indicated in **Table 6-3**…" Is this the right table?

Section 6.2.1 line 236-237: Are you referring to the correct table? There are no such results in Table 4-11 and 4-12 (?).

Section 6.2.1, Table 6-8: footnote "d" is wrong. ~~The substance was classified as negative in the third laboratory~~.

Section 6.2.2, line 269: "(… defined in Section ~~2.7.1.4~~)." should it be 2.7.1.5?

Table 6-9: footnote f an h: corrections needed.

Line 358-359
"… under the revised testing protocol, …". After the word protocol, insert within parenthesis the section where the revised protocol is described.

Section 6.1.7
Says "Although the classifications for some of the test substances differed among the laboratories", but according to table 4-14 and section 6.2.1 "Among the substances tested for antagonist activity, there was 100% agreement among the three laboratories for all 12 substances."

Section 6.1.7
Says "... within each laboratory for each of the three repeat tests", but according to table 4-14 some substances were tested 4 times.

Table 6-2
Plates 2 and 3 of ECVAM have identical means and SDs and the means of plates 4 and 5 are clearly different. In contrast, in figure 6-2 means of plates 2 and 3 are different and means of plates 4 and 5 are similar. Also inconsistencies in other tables? E.g. table L-3, plates 6, 7,10 of Hiyoshi (compare with figure 6-3).

Table 6-2
Mean E2 EC50 of ECVAM is 1.1e-11. Are the horizontal lines in figure 6-4 means? If so, In figure 6-4 the respective ECVAM mean seems to be smaller than 1.0e-11. In Table 6-6 mean Flavone Control Value for XDS is 3774, but in figure 6-10 it looks like the mean is smaller. Are means in the figures calculated without the values excluded to minimze scale distortion?

Footnotes of table 6-9
There are some inconsistencies (e.g. footnote h +--- , …) that should be corrected.

Section 6.2.1, line 237
Reference to tables 4-13 and 4-14?

(p. 6-3, Figure 6-1) - What is the meaning of "plate" here? Should we expect that these values would be the same?

(p. 6-5, Figure 6-3) – All figures have a tremendous amount of wasted space (above 10000 in this figure). If the figures are rescaled, then the variability in points will be clearer. Figure 6-6 could be much improved.

(p. 6-6, Table 6-2) – SD of EC50 estimates is essentially a SE estimate. Are each of the constituent EC50 values estimated with the same precision?

(p. 6-7, Figure 6-5) – Footnotes describe points omitted to "minimize scale distortion" – this omission removes points that are potentially dramatic outliers.

There have been sufficient toxicity test method validation studies that the conclusions of intra and inter-lab variability could be discussed in terms of how the BG1Luc ER TA reliability compares to previously endorsed methods - CVs, % correct classifications, other analysis.

Additional suggestions for criteria to evaluate intra and intra lab variability were provided in Panel comments for Section 6.0.

7.0 Data Quality

Sub-section on Reference Methods Data Quality is missing.

Section 7.5
Add statement on the availability of audit results, statistical evaluations, and methods for calculations; descriptions for the specific length of time that study data will be stored (and where), and where data backups will be stored (a secondary location).

Line 122-124: This sentence does not clearly explain Table 7-4.

Line 145: "… provided in Section 2.7.1.3 …". You probably meant section 2.7.1.4.

Line 169: "… defined in Section 2.12.3." This section does not exist.

Line 182-183:
"… presented in Tables 4-11 and 4-12…". Wrong tables, you probably meant 4-13 and 4-14 (?).

8.0 Other Scientific Reports
No edits.

9.0 Animal Welfare
The organization and wording of this entire section can be improved. A main consideration is that the content of this section does not appear to contradict the results or statements made elsewhere in the BRD.

Additional content is also suggested for this section in the Panel comments for Section 9.0.

Test Method Recommendations

L29: Perhaps the reader can be referred to the glossary for the definition of accuracy and reliability.

L32
Should "definitive results" be in quote, and added to the glossary to indicate that it is the data excluding rejected plates and inadequate data.

Lines 56-57 - numbers possibly incorrect.

L59-68 and L83-89
Perhaps these paragraphs can be used to create a separate section numbered 1.1.3 to discuss the limitations of the assay. Note that L61-64 list new experiments (mixtures, volatiles) to be presented with the others in section 1.3. Also, L62-64 (volatiles…) are redundant with L125-127.

L104 and L114
Verify both expressions on both lines, perhaps L114 should have been: "…100% specificity (no false positive)". (?)

L119
"… to ~~replace~~ reduce the need for the uterotrophic bioassay." What would be the alternative test if a substance cannot be tested in the BG1Luc ER TA assay?

Section 1.1.1: Lines 43-49 on test method reliability
May need to be revised following Panel discussions and revision of the draft BRD.

Lines 56-57 - numbers possibly incorrect.

Section 1.1.2
Lines 76-82 on test method reliability may need to be revised following Panel discussions and revision of the draft BRD.

Add Section 1.1.3
Add recommendation that BG1Luc ER TA be endorsed for replacing rat ER binding assay in test batteries and tiered testing schemes such as EDSP Tier 1 battery.

Performance Standards

Performance Standards section (Tab 6, p. xii, lines 272-307)
What is the justification for some of the essential model components? For example, why were these maximal concentrations suggested? Why log10 spacing of 7 concentrations? Evaluation of cytotoxicity is reasonable although modeling of the possibility of such a response could also be considered. Why is 20% of maximal response the cut-point for significant response?

ER TA antagonist testing (Tab 6, p. 6, lines 476-493) – concentration curve defined by a baseline followed by a negative slope – isn't it sufficient to require a negative slope? (will always have a baseline). A test of negative trend might be preferred to the suggestion of "three points with non-overlapping error bars." Error bars corresponding to what? Standard deviations? Standard Errors? Half-width of confidence intervals? The idea of non-overlapping confidence intervals is an incredibly conservative way to declare difference between parameters. This is not clear guidance.

Would Figure 1 (line 470) reflect a real study result? The guideline of log10 spacing of 7 concentrations wouldn't span the -13 to -4 log(conc) range depicted here. Add revised Figures 1 and 2.

Appendices

(Appendix B – Agonist Protocol, p. B-6, line 135)
A four parameter model is suggested here that captures a pattern of growth between two horizontal asymptotes. A couple of concerns: 1) won't "Bottom" = 0 and "Top" = 10000 when you do relative scaling of responses? If so, then you are not working with 2 parameters and not 4; 2) data are known to have a downturn at high doses – this is not consistent with the assumed model. How do you routinely address this? Are you dropping the highest concentration responses in this case? Could you model downturn at the higher concentration levels?

(Appendix B – Figures 12-2 through 12-7)
Better graphics to not have so much wasted white space in the graphs.

(Appendix B – p. B-39, Figure 13-1)
The layout of the test plate is unclear to me. Can more explanation of rows A-H and columns 1-12 be provided?

This page intentionally left blank

Appendix B

Peer Review Panel Member Biosketches

Panel Member Biosketches

A. John Bailer, PhD
Dr. Bailer received a PhD in 1986 from the University of North Carolina at Chapel Hill and is currently the Distinguished Professor and Chair, Department of Statistics at Miami University. He is a member of the American Statistical Association, International Biometric Society, International Statistical Institute (elected 2004), International Association of Statistical Education, Society for Risk Analysis, Society of Environmental Toxicology and Chemistry, and Sigma Xi. Dr. Bailer has served as a reviewer for numerous statistical and toxicological journals including Biometrics, Risk Analysis, Environmental Toxicology and Chemistry, American Journal of Industrial Medicine, American Journal of Public Health, Occupational and Environmental Medicine, Environmental and Ecological Statistics, Environmetrics, Environmental and Molecular Mutagenesis, Environmental Health Perspectives, Fundamental and Applied Toxicology, Biological and Environmental Statistics, Journal of the American Statistical Association, and Journal of Pharmacokinetics and Biopharmaceutics, Cancer Research. He has served as Secretary/Treasurer, Vice-President, and President of the Cincinnati Chapter of the American Statistical Association, and is a member of the International Statistical Institute Council (2009–2013) and is Council liaison, Committee on Statistics in the Life Sciences. Dr. Bailer has authored or coauthored 121 peer-reviewed journal articles, 4 books, and 34 book chapters and technical reports.

Christopher J. Borgert, PhD
Dr. Borgert received a PhD in 1991 from the University of Florida College of Medicine, Department of Pharmacology and Therapeutics. He completed a postdoctoral fellowship at the Center for Environmental and Human Toxicology at the University of Florida Research and Technology Park. He is currently the President and Principal Scientist at Applied Pharmacology and Toxicology, Inc. Dr. Borgert has served on numerous panels and committees focused on toxicological assessment of mixtures and human health impacts and was a member of the OECD Peer Review Panel for Validation of the Uterotrophic Assay, and the EPA Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), the EDSTAC Screening & Testing Workgroup, and was Co-Chair of the EDSTAC Communication & Outreach Workgroup. Dr. Borgert has served as a reviewer for Chemosphere, Environmental Research, Environmental Toxicology and Chemistry, Human and Ecological Risk Assessment, International Journal of Toxicology, International Workshop on Quantitative Structure-Activity Relationships (QSARs) in Environmental Sciences, Journal of Agricultural and Food Chemistry, Nonlinearity in Biology, Toxicology, and Medicine, Regulatory Toxicology and Pharmacology, Toxicological Sciences, and Toxicology and Applied Pharmacology. He is currently a member of Toxicology Forum, Society of Environmental Toxicology and Chemistry, Southeast Chapter, Society of Environmental Toxicology and Chemistry, and the Society of Toxicology, International Society of Regulatory Toxicology & Pharmacology. Dr. Borgert has authored or coauthored 20 peer-reviewed journal articles, 5 book chapters, and has given 35 invited presentations.

Grantley D. Charles, PhD, DABT
Dr. Charles received his PhD from the University of Florida, Gainesville, in Pharmacology and Toxicology and completed a postdoctoral fellowship at the Dow Chemical Company. While at Dow, he received a Dow Chemical special recognition award for research conducted on endocrine disruptors. Dr. Charles is currently a Senior Scientist at Allergan, Inc. He was a member of the expert panel for the ICCVAM/NICEATM review of the Validation Status of *In Vitro* Test Methods for Detecting Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. He also served as a member of the Scientific Advisory Committee on Alternative Test Methods (SACATM) for ICCVAM/NICEATM, and the American Chemistry Council (formerly Chemical Manufacturer's Association) ad hoc *In Vitro* Endocrine Disruptor subcommittee. He is a reviewer for Toxicological Sciences, Journal of Toxicology and Environmental

Health, Toxicology In Vitro, Journal of Pharmacy & Pharmacology, Reproductive Toxicology and Birth Defects Research, Expert Opinion in Pharmacotherapy, Expert Opinion in Drug Metabolism and Toxicology. He is currently a member of the Society of Toxicology and the Southern California Chapter of the Society of Toxicology and a Diplomate of the American Board of Toxicology. Dr. Charles has authored or coauthored 16 peer-reviewed journal articles and given 4 invited presentations.

Daniel Desaulniers, MSc, PhD
Dr. Desaulniers received a Master of Sciences degree in Veterinary Anatomy and Physiology and a PhD in Biomedical Sciences from Montreal University, St-Hyacinthe, Quebec. He is currently a Research Scientist, Health Canada, Healthy Environments and Consumer Safety Branch. His research has included using *in vitro* MCF-7 cell proliferation and the uterotrophic bioassay to evaluate the estrogenic activity of chemicals. He participated in the Mammary Gland Evaluation and Risk Assessment Workshop, was a reviewer for the OECD Test Guideline on the Uterotrophic Bioassay in Rodents, and was a member of the National Advisory Council for Environmental Policy and Technology (NACEPT) Endocrine Disruptor Methods Validation Subcommittee (EDMVS). He has served as a reviewer for 11 scientific journals, including Environmental Health Perspectives, International Journal of Toxicology, Teratology, Toxicology and Industrial Health, and Toxicological Sciences. Dr. Desaulniers has authored or coauthored 21 peer-reviewed journal articles, 5 government reports, and over 90 posters. He has given 16 invited presentations.

John Charles Eldridge, PhD
Dr. Eldridge received his PhD in Endocrinology from the Medical College of Georgia, Augusta and completed post-doctoral work as an Attaché de Recherche, Institut Nationale de la Santé et de la Recherch Mèdicale in Bordeaux, France. He is currently a Professor at the Wake Forest University School of Medicine in the Department of Physiology and Pharmacology. His interests include the endocrinology of reproduction, steroid hormone biochemistry and receptor mechanisms, and hormone and drug assay methodology. His research has focused on endocrine toxicology of xenobiotic chemicals, especially triazine herbicides, and hormones of stress and endocrine aspects of aging. Dr. Eldridge is a program reviewer for the National Institute on Aging, the National Center for Research Resources at NIH, the Health and Environmental Effects Research Laboratory (EPA), and is a member (ad hoc) of the FIFRA Science Advisory Panel (EPA). He is also a consultant (ad hoc) for reproductive and endocrine toxicology for Novartis Animal Health, Inc. and was a member of the EPA Endocrine Disruptors Methods Validation Committee. Dr. Eldridge holds or has held membership in numerous professional societies, including the American Society for Reproductive Medicine, the Endocrine Society (emeritus), Society for the Study of Reproduction (emeritus), the American Society for Andrology, and the Society for Neuroscience (emeritus). He is reviewer for a number of scientific journals including the Journal of Pharmacology and Experimental Therapeutics, Environmental Health Perspectives, the Journal of Toxicology and Environmental Health, Neuroendocrinology, Biology of Reproduction, and the Endocrine Journal. Dr. Eldridge has authored or coauthored 53 peer-reviewed journal articles, 16 books or monographs, 16 book chapters, and 69 abstracts for national or international meetings. He has given 24 invited lectures or seminars.

William R. Kelce, MS, PhD, ATS
Dr. Kelce received his PhD in Physiology/Toxicology from the University of Missouri–Columbia and completed a postdoctoral fellowship in Toxicology at The Johns Hopkins University. He is currently vice-president of preclinical development and clinical operations at POZEN Corporation and is also an Adjunct Associate Professor at the University of North Carolina School of Medicine, Department of Pediatrics and the Laboratories for Reproductive Biology. Dr. Kelce previously served as Director of Developmental and Reproductive Toxicology at Pfizer Global Research and Development. He received the Young Andrologist Award from the American Society of Andrology and the U.S. EPA Gold Medal Award (Scientist of the Year). As a postdoctoral fellow, he received First Place, Young

Investigator Award–Developmental and Reproductive Toxicology Section from the Society of Toxicology. He is a member of the International Life Sciences Institute Developmental and Reproductive Toxicology Section. He has served on numerous panels and committees focusing on endocrine disruptors, including the EPA Endocrine Disruptor Methods Validation Advisory Committee, the EPA Endocrine Disruptor Standardization and Validation Task Force, where he served as Chair of the *In Vitro* Screening Work Group, and as a member of the ICCVAM Endocrine Disruptor Peer Review Panel. He is a member of the Drug Information Association, Society of Toxicology (Editorial Board Member, Toxicology and Applied Pharmacology and member of the Board of Publications for Toxicological Sciences), the Endocrine Society, American Society of Andrology (Editorial Board Member, Journal of Andrology), American College of Toxicology, Teratology Society (Editorial Board Member, Birth Defects Research: Developmental and Reproductive Toxicology), Midwest Teratology Association (Steering Committee Member), and the Academy of Toxicological Scientists. Dr. Kelce has authored or coauthored 46 peer-reviewed journal articles and 13 invited reviews and book chapters. He has given 44 invited presentations.

Hyung Sik Kim, PhD
Dr. Kim received his PhD in Pharmacy from College of Pharmacy, Sung Kyun Kwan University. He is currently an Associate Professor in the Laboratory of Molecular Toxicology, College of Pharmacy, at Pusan National University, in the Republic of Korea, where he conducts research in the development of biomarkers and mechanisms of action of environmental xenobiotics at molecular and cellular levels. Dr. Kim is a member of Society of Toxicology, the American Association for Cancer Research, the Society of Study for Reproduction, the Korean Cancer Association, the Korean Society of Environmental Toxicology, the Korean Society of Applied Pharmacology, the Korean Society of Toxicology (Editorial Board), the Pharmaceutical Society of Korea, and the Korean Society of Food Hygiene Safety. Dr. Kim has authored or coauthored 95 peer-reviewed journal articles and 5 books.

Steven Levine, PhD
Dr. Levine received his PhD in Zoology from Miami University (Ohio) and was an NIEHS Postdoctoral Fellow in Toxicology at Pennsylvania State University. He is currently the Product Safety Manager – Ecotoxicology & Risk Assessment; Quantitative Bioassay Team Lead and Science Fellow in the Regulatory Sciences Department at Monsanto Company in St. Louis, MO. Dr. Levine's primary research interests include development of functional assays with insecticidal traits in biotechnology-derived products, aquatic and terrestrial toxicology, molecular mechanisms of steroidogenesis, and approaches for probabilistic risk assessment. He was a BIAC representative to the OECD Working Group of National Coordinators for Test Guidelines and Endocrine Disruption Testing Advisory Committee, chaired Crop Life America's (CLA) Endocrine Disruptor Group and served on the CLA and American Chemistry Council's Ecological Risk Assessment Working Group. He also served on the EPA Endocrine Disruption Methods Validation Advisory Committee. Dr. Levine was the President of the regional Society of Environmental Toxicology and Chemistry (SETAC) and is a member of SETAC and the Society of Toxicology. Dr. Levine has authored or coauthored 21 articles in peer-reviewed journals, 33 abstracts/presentations at national and international meetings, nine of which were invited. He has given eight invited lectures at universities and other nonmeeting venues.

Alberto Mantovani, DVM, MSc
Dr. Mantovani received a DVM from the University of Bologna and a MSc in Veterinary Public Health from the University of Edinburgh. He is currently the head of the Food and Veterinary Toxicology Unit within the Department of Food Safety and Veterinary Public Health of the Italian National Health Institute (ISS). The main research topics of the unit are endocrine disruptors and trace elements. Dr. Mantovani chaired the Endocrine Disrupters Technical Working Group within the SCALE Project to set the scientific bases of the "Environment and Health Action Plan" of the European Commission, chaired the Italian pilot project on endocrine disrupters, and participated in

the EU Working Group "Specialized Experts in the Field of Mutagenic, Carcinogenic and Teratogenic Substances." He was the Italian Expert on toxicology at the Safety of Residues Working Party of the Committee for Veterinary Medicinal Products–European Agency for Evaluation of Medicinal Products; and currently participates on the OECD Working Group on Endocrine Disrupters Testing and Assessment. Dr. Mantovani has authored or coauthored 38 peer-reviewed journal articles since 2001.

Ellen Mihaich, PhD, DABT
Dr. Mihaich received her PhD from the Integrated Toxicology Program in the School of Forestry and Environmental Studies at Duke University. She is a Diplomate of the American Board of Toxicology and is currently the owner and principal scientist of Environmental and Regulatory Resources, LLC, specializing in environmental toxicology, risk assessment, and regulatory services. She is also an Adjunct Professor at Duke University. Dr. Mihaich was formerly Manager of Environmental Toxicology at Rhodia, Inc., and a Senior Environmental Toxicologist at Rhône-Poulenc. She served as a domestic and international consultant on environmental toxicology at both companies. Her responsibilities included management of environmental toxicology-based issues and environmental toxicology programs for chemical enterprises, monitoring and evaluating general and specialty laboratory and field environmental toxicology studies, plant site and chemical-use risk assessment, and interaction with international industry groups and regulatory agencies. Dr. Mihaich served as President of the Society of Environmental Toxicology and Chemistry (SETAC) and was on the Board of Directors for SETAC North America and the World Council. She is currently on the Scientific Advisory Board of the Strategic Environmental Research and Development Program (SERDP) of the Department of Defense. Dr. Mihaich previously served on the EPA Ecological Processes and Effects Committee, Deriving Aquatic Life Criteria for Emerging Contaminants, and ICCVAM peer panels on endocrine disruptors (2002) and the Frog Embryo Teratogenesis Assay (2000). Dr. Mihaich has authored or coauthored more than 25 publications in peer-reviewed journals or government publications and over 40 abstracts at national or international meetings.

Hiroshi Ono, MD, PhD
Dr. Ono received his MD from the Tohoku University School of Medicine and his PhD in Medical Research from Tohoku University. He is currently a scientific adviser to Hatano Research Institute, Food and Drug Safety Center. Dr. Ono has served as a visiting lecturer of General Medical Science at Azabu University School of Environmental Health Science, and toxicology at Yamanashi Medical College and at Tohoku University School of Medicine. He also served as a longtime delegate to the Meeting of National Coordinators of Test Guidelines Programme of Organisation for Economic Cooperation and Development. He is a member of the Japanese Society of Pharmacology, Japanese Society of Toxicology, Japanese Society for Alternatives to Animal Experiments, Japanese Society for Endocrine Disruptors Research, the Society of Toxicology, and EUROTOX-Association of European Toxicologists and Toxicological Societies. He received the Tanabe Prize for the notable publication of the year from the Japanese Society of Toxicology. Dr. Ono has 23 original articles and 49 reviews published in Japanese and 95 original articles published in English.

John G. Vandenbergh, PhD
Dr. Vandenbergh received a Masters degree in Zoology from Ohio University and a PhD in Zoology from Pennsylvania State University. He is currently a Professor Emeritus in the Department of Biology at North Carolina State University. In recent years, Dr. Vandenbergh has focused his research on the effects of endocrine disruptors on development and later reproductive performance in rodents. In 2002 he received the Holladay Medal, NC State University's highest faculty award. He is a founding board member and former chair of the North Carolina Association for Biomedical Research (NCABR) and served as a fellow and former president of the Animal Behavior Society. He was a member of the Institute for Laboratory Animal Research Council of the National Academies of Science and was on the committee that wrote the *Guide for the Care and Use of Animals* (both 1986

and 1996). Dr. Vandenbergh has also served on several additional committees at the National Academies of Science, including as Chair of Committee on Animal Biotechnology: Science Based Concerns. He was named a National Associate of the National Academies of Science in 2003. He has been a member of several scientific review committees at the NSF, EPA and the NIH. Most recently, he served on the National Toxicology Program's Expert Panel on Human Risks of Exposure to Bisphenol A (BPA). Dr. Vandenbergh is a present or past member of several scientific societies and editorial boards. He has edited 2 books, published over 100 chapters and review papers, and given many invited scientific presentations.

Sherry Ward, PhD, MBA
Dr. Ward received her PhD in Biochemistry from Michigan State University, an MBA from the University of Maryland University College (UMUC), and an executive MS in Technology Management from UMUC. She is currently a consultant with BioTred Solutions in New Market, Maryland. Dr. Ward has expertise in *in vitro* toxicology, scientific writing and project management, grant proposal review, and grant writing. She also has experience in biotechnology market research, commercialization, and strategy development. Dr. Ward is a contributing editor to AltTox.org. She is an adjunct faculty member at UMUC in Biotechnology & Project Management. She has animal welfare experience and has served since 2006 on the board of the International Foundation for Ethical Research. As a Staff Scientist at the Gillette Company, she developed, characterized, and drafted patent applications for the first human conjunctival epithelial cell lines and gained experience in bioassay development and validation. Dr. Ward has served on numerous scientific panels and committees and was a panel member and presenter at the ICCVAM symposia on mechanisms of ocular injury and recovery and minimizing pain and distress in ocular toxicity testing held at the NIH in May 2005. She has been actively involved with trade organizations and served on the European Cosmetic, Toiletry and Perfumery Association Eye Irritation Task Force and the ILSI-HESI Alternatives to Animals Task Force. Dr. Ward's experience in models of eye irritation and mechanisms of injury is reflected in 19 publications in peer-reviewed journals, 4 unpublished validation or prevalidation documents related to ICCVAM activities, 17 presentations, 28 abstracts, and a patent. She is a member of the Hopkins Medical and Surgical Association and the Washington Academy of Sciences.

Marc Weimer, PhD
Dr. Weimer received a PhD in Neurophysiology from the University of Hohenheim, Germany, and an MS in Methods and Models from FernUniversität in Hagen, Germany. He joined the Department of Biostatistics, German Cancer Research Center (DKFZ), in Heidelberg in 2006 as a biostatistician. Dr. Weimer's primary areas of work are toxicogenomics and development and validation of alternative methods to animal experiments. As a statistical consultant, he has been involved in national and international research projects aimed at reducing, refining, and replacing animal testing in toxicology. His main interests include dose–response modeling, agreement statistics, and toxicogenomics. Funded by ECVAM, he has been responsible for the statistical evaluation of the quality of in vitro assays developed within ReProTect, a project of the European Union advancing alternative methods in reproductive toxicity. Dr. Weimer has authored or coauthored 19 peer-reviewed journal articles.

James Wittliff, PhD, MD *hc*, FACB
Dr. Wittliff received his PhD in Molecular Biology from the University of Texas at Austin and completed postdoctoral studies in the Biology Division of Oak Ridge National Laboratory in Tennessee. He is currently a Professor of Biochemistry and Molecular Biology in the Graham Brown Cancer Center in the School of Medicine at the University of Louisville with additional appointments as Research Professor of Surgery and Director of the Institute for Molecular Diversity & Drug Design (IMD). He is also the Director of the Hormone Receptor Laboratory at the University and has held numerous professorships at universities in Europe, Asia, and Africa. Dr. Wittliff's research interests include mechanisms and applications of steroid and peptide hormone action in disease, biochemical

techniques and concepts for detection and treatment of cancer, and laser capture microdissection and its use in proteomics and genomics. He was among the first to prove that the appearance of estrogen receptors in breast cancer predicted a patient's response to hormone therapy. Dr. Wittliff has researched the biological properties and cellular roles of estrogen and progestin receptors in human cancers and the actions of estrogen mimics acting as endocrine disruptor compounds (EDC). Formerly at NEN/DuPont, Dr. Wittliff developed the original FDA-approved kits for assessing receptors in biopsies, celebrated as a major contribution to laboratory medicine. He was the Principal Investigator or the Investigator for a number of funded studies on genomic approaches to disease, including a Genomic Approach for Assessing Clinical Outcome of Breast Cancer using Cells Isolated by Laser Capture Microdissection. Dr. Wittliff has served on numerous panels and committees, including the ICCVAM Endocrine Disruptor Peer Review Panel (2002). He is a member of several professional societies including the Endocrine Society, the American Association for Cancer Research, and the American Association for the Advancement of Science. Dr. Wittliff has authored or coauthored over 250 peer-reviewed publications and holds patents on methods and apparatus for measurement of the effect of test compounds on signal transduction at the receptor level, quantitative immunohistochemistry, breast cancer signatures, and gene expression profiles.

James Yager, Jr., PhD

Dr. Yager received his PhD from the University of Connecticut, Storrs Campus, in Cell and Developmental Biology and conducted postdoctoral studies at the McArdle Laboratory for Cancer Research at the University of Wisconsin. He is currently a Professor in Preventive Medicine and Toxicology in the Department of Environmental Health Sciences at the Johns Hopkins Bloomberg School of Public Health with a joint appointment in the Department of Oncology. He has administrative responsibility as the Senior Associate Dean for Academic Affairs. He was formerly a Professor of Anatomy and an Adjunct Professor in the Biochemistry Program at Dartmouth College. Dr. Yager has served as the Program Director and Principal Investigator for the Training Program in Environmental Health Sciences, Director of the Division of Toxicological Sciences, and Director of the Molecular Toxicology Program of the NIEHS-supported Center in Urban Environmental Health. His research interests include mechanisms of promotion of hepatocarcinogenesis by estrogenic xenobiotics, mechanisms of estrogen-induced oxidative DNA damage in liver and human breast epithelium, and the role of genetic susceptibility in human cancer through polymorphisms in biotransformation enzymes involved in estrogen oxidative metabolism. Dr. Yager serves on various committees and task forces for several professional societies, including the American Association of Cancer Research (AACR), the American Society for Investigative Pathology (ASIP, FASEB), and the Society of Toxicology. Dr. Yager serves or has served on various advisory boards and chartered review panels, including the EPA Endocrine Disruptor Methods Validation Subcommittee and the ICCVAM Scientific Review Panel to evaluate the validation status of in vitro estrogen and androgen receptor binding and transcriptional activation assays (2002). He is a peer reviewer for numerous journals, including Biochemical Pharmacology, Cancer Research, Chemical–Biological Interactions, Molecular Carcinogenesis, Science, and the Journal of the National Cancer Institute. Dr. Yager is on the editorial board for the Journal of Environmental Pathology, Toxicology and Oncology; In Vitro-Cell & Developmental Biology; Toxicology Sciences; and Chemical Research in Toxicology. He has authored or coauthored 84 peer-reviewed journal articles, 15 book chapters, 66 abstracts or presentations at national and international meetings; and he and has given over 50 invited presentations.