Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

12-Nov-2007

English - Or. English

ENV/JM/MONO(2007)32
Unclassified

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

SERIES ON TESTING AND ASSESSMENT
Number 83

Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report

JT03235654

English - Or. English

OECD Environment Health and Safety Publications

Series on Testing and Assessment

No. 83

Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among UNEP, ILO, FAO, WHO, UNIDO, UNITAR and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2007

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (in preparation)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Draft Guidance Document on Reproductive Toxicity Testing and Assessment (in preparation)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work Sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay fort he Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay : Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work Towards the Validation of the Rodent Utertrophic Assay - Phase 1 (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional data supporting the Test Guideline on the Uterotrophic Bioassay in rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the follow up of this report (2006)*

No. 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on the Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay - Procedure to Test for Antioestrogenicity (2007)*

No. 72, *Guidance Document on Pesticide Residue Analytical Methods (2007)*

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Toxicity Testing (2007)*

No. 75, *Guidance Document on the Honey Bee (Apis Mellifera L.) Brood test Under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 - Optimisation of the Test Protocol (2007)*

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances (2007)*

No. 79, *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa - Phase 1 (2007)*

No. 80, *Guidance on Grouping of Chemicals (2007)*

No.81, *Summary Report of the Validation Peer Review for the Updated Test Guideline 407, and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up of this report (2007)*

No.82, *Guidance Document on Amphibian Thyroid Histology (2007)*

No.83, *Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides and Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and the Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/ehs/).

This publication was produced within the framework of the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC).

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are FAO, ILO, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

**OECD Environment Directorate,
Environment, Health and Safety Division**

**2 rue André-Pascal
75775 Paris Cedex 16
France**

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document presents the summary report of the assessment by an independent Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals.

The need for cost-efficient and rapid screening tools for the detection of chemicals with endocrine modulating properties was confirmed at the 6th meeting of the OECD Endocrine Disrupter Testing and Assessment (EDTA) Task Force in 2002. The OECD conceptual framework for testing and assessment of potential endocrine disrupters was also agreed at the same meeting (OECD, 2002). The conceptual framework is organised into 5 levels corresponding to differing levels of biological complexity. The framework is intended to be used as a toolbox containing a variety of test methods that can contribute to the identification of chemicals with endocrine modulating properties. The Stably Transfected Transcriptional Activation Assay is proposed for inclusion in the conceptual framework at Level 2, comprising *in vitro* assays providing mechanistic information.

There are currently no *in vitro* screening assays for estrogenic activity that have been validated and peer reviewed for inclusion in OECD Test Guidelines. At the first OECD validation management group for non-animal testing (VMG-NA), it was agreed that Japan would take the lead in developing the Stably Transfected Transcriptional Activation Assay. The assay underwent a pre-validation phase within one laboratory, followed by inter-laboratory validation. A draft report of the pre-validation and validation work, together with a draft test guideline, were submitted by the lead laboratory, the Chemicals Evaluation and Research Institute, Japan, in October 2006. These two documents were reviewed by the Peer Review Panel between November 2006 and March 2007.

The peer review was managed by an independant consultant contracted by the OECD Secretariat. The National Coordinators proposed peer reviewers but the final composition of the independent peer review panel was decided by the consultant. Members of the panel are listed in Annex 1 of the summary report; they were requested to send declaration of interest to the consultant. As agreed by the Working Group of the National Coordinators of the Test Guidelines Programme (WNT), the WNT agreement on the follow-up of the peer review panel report is attached to this summary report. This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of its member countries.

Agreement of the WNT on the Follow-up of the Peer Review Panel Report of the Stably Transfected Transcriptional Activation (STTA) Assay

The Peer Review Panel (PRP) report of the validation of the Stably Transfected Transcriptional Activation (STTA) Assay was submitted for information to the 19th meeting of the Working Group of National Coordinators of the Test Guidelines Programme (WNT) in March 2007. In accordance with the recommendations of the peer review panel, the WNT supported the development of a Test Guideline based on the STTA Assay. The WNT requested that the Validation Management Group for Non-Animal testing propose criteria for positive responses and acceptable test performance and it should be noted that at this point the assay can only be used for estrogen agonist testing. Considering the above, the WNT agreed to proceed to further development, refinement and evaluation of a proposed Test Guideline for a STTA Assay.

Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation (STTA) Assay for Detecting Estrogenic Activity of Chemicals

The peer review process

1. The Peer Review Panel (Panel) was constituted in November 2006, to provide an independent review of the validation of the Stably Transfected Transcriptional Activation (STTA) Assay for Detecting Estrogenic Activity of Chemicals. The assay is intended to be used for identifying and prioritising substances that have the potential to act as estrogen receptor (ER) agonists binding to ER-alpha. The work of the Panel was coordinated by a consultant manager from outside OECD. Panel members were chosen by the consultant manager from among candidates nominated by the Working Group of the National Coordinators of the Test Guidelines Programme (WNT). The members of the Panel are listed in Annex 1.

2. The Panel was asked to evaluate the data collected on the assay, and to answer specific questions posed to the Panel in the charge provided by the sponsoring organization, the Organisation for Economic Co-operation and Development (OECD). Panel members were provided with the draft report of the pre-validation and inter-laboratory validation of the STTA assay and the draft OECD guideline for the STTA assay, submitted in October 2006 by the lead Japanese laboratory. As background, they were also provided with the OECD Guidance Document on the Validation and International Acceptance of New or Updated Methods for Hazard Assessment, Series on Testing and Assessment, Number 34, 2005. The charge to the Panel was whether the 8 OECD validation criteria set out in the OECD Guidance Document had been met. A summary of the Panel's responses to the individual questions is presented in paragraphs 8-21 below and more detailed comments from the Panel are provided in Annex 2. In addition, Panel members were asked to give their overall impression of the assay together with any other general comments.

3. Each Panel member provided written responses on the charge questions to the consultant manager and a number of issues were followed up by email correspondence between the manager and the Panel. The Panel also held a teleconference during the evaluation process coordinated by the manager. This report presents the combined responses of the Panel on their overall impression of the assay and to each of the charge questions.

Overall impression of the Panel on the STTA assay

4. The overall impression of the Panel on the STTA assay, from the data generated in the pre-validation and validation phases, is of a robust assay providing similar results to other *in vitro* transcriptional activation assays. The Panel recognised the tremendous amount of work carried out by the participating laboratories and the importance of pressing ahead with the validation of such an assay. The Panel agreed that the proposed assay has been sufficiently tested to meet most of the OECD validation criteria. The Panel also agreed that the validation study provides a basis for a standard protocol for screening for *in vitro* estrogenic activity. The Panel considered that although the draft test guideline will need some refinement, the results provide a good foundation for OECD to continue with the development of this assay.

5. The Panel did however emphasise that a number of points still need to be clarified and these are elaborated further below. A strong criticism from all Panel members was that the criteria for a positive response were unclear and need to be further discussed and clearly defined. There is also insufficient

guidance on the criteria for acceptable test performance (i.e. acceptance criteria) by laboratories performing this assay. This is a critical point, because acceptance criteria must be met to demonstrate validity of individual assays performed on each test day.

6. The Panel was in agreement that, even though a protocol for anti-estrogenic activity had been included in the draft report, reproducibility and acceptable performance has only been demonstrated for estrogen agonists. If the assay were to be used for estrogen antagonists, further testing would need to be done.

Panel responses to the charge: Have the eight OECD principles and criteria for test method validation been met?

7. The Panel reached consensus on all the questions posed.

Rationale for the test method should be available

8. The Panel agreed that this criterion had been fully met. The rationale for the test method is clearly stated with regard to the scientific basis and regulatory purpose. Under the OECD conceptual framework, it would provide mechanistic information as a Level 2 *in vitro* assay.

The relationship between the test method endpoint and the biological phenomenon of interest should be described

9. The Panel agreed that this criterion had been fully met. The Panel concurred with the comments in the draft validation report and the draft test guideline that while the assay will provide mechanistic information and may allow comparison of ER selectivity (ERα versus ERβ), it will not inform on any downstream estrogenic responses of test chemicals, which would need to be further explored by *in vivo* methods. However, the Panel noted that there was good concordance between the rank order of test substances in the STTA assay and in the immature rat uterotrophic assay.

A detailed protocol for the test method should be available

10. The Panel agreed that this validation criterion has only been partially met.
In particular, it has not been met with respect to the important issue of clearly defining and justifying the criteria for a positive response. It also has not been met with respect to the need to define criteria for acceptable test performance.

11. Concerning criteria for a positive response, the Panel concluded that there is insufficient clarity on what would constitute a "positive" response in this assay. Similarly, it is unclear what would define a "negative" response, such that no further testing would be necessary. The Panel emphasised that an appropriate trend analysis should be used to evaluate for a significant dose-response relationship and then an appropriate pair-wise test could be used to evaluate for a significant effect at the different test substance concentrations. The Panel agreed that a test substance having a quantifiable PC50 should certainly be regarded as positive. However, the Panel was equally strong in its view that a PC10 value is not appropriate to be used as a criterion of a positive response because it represents a marginal response of doubtful biological predictive value and is likely to generate a high rate of false positives. The Panel discussed whether some intermediate value between PC10 and PC50 (e.g. PC25) might also be considered as a positive response, but noted that there were no data on which to base such a decision and considered that further statistical advice might be needed on this point.

12. The Panel considered that criteria for acceptable test performance that would need to be met by a new laboratory performing the assay for the first time should be outlined more clearly. At the very least, EC50 values for positive controls should meet a specified tolerance interval or range, or fall within 95% confidence limits from historical data.

13. Concerning materials and methods, the Panel raised some questions in relation to sourcing of some materials, including the availability of the cell line. The Panel was also not convinced that "edge effects" would not be a problem and recommended that there should at least be a comment in the draft test guideline on the need to be vigilant regarding the possibility of such effects and to modify the plate dose assignment if edge effects do occur.

14. The Panel commented that non-receptor-mediated luciferase activation needs to be taken into consideration, noting that the proposed protocol does not include any control to assess the specificity of the luciferase induction. The Panel agreed with the suggestion that emerged during the preliminary validation discussions that co-exposure of test substances with a pure anti-estrogen and the use of the ER-negative, luminescent HeLa-9903 control cell line could be included to confirm the ER agonist property of tested chemicals, when they are found positive.

15. The Panel agreed that statistical analysis of the results should utilise the simplest reliable and predictive methods available that would be compatible with the characteristics of the particular index value being calculated (e.g. EC50 or PC50) and with the way in which the calculated index values are to be used in decision-making on a positive or a negative outcome. For example, the Panel considered that if the calculated index values are used as a mean for qualitative expression of the intensity for endocrine disrupting potential, the method that provides a more precise estimate, i.e. the Hill equation-based nonlinear regression method, should be used. However, if it is decided only to classify test chemicals into one of three categories - positive, negative or equivocal - then such precise values are not needed and a simpler approach could be used.

The intra- and inter-laboratory reproducibility of the test method should be demonstrated

16. The Panel agreed that this criterion had been met. The Panel concluded that the design of the inter-laboratory studies was generally appropriate and sufficient to establish reproducibility of the protocol for estrogen agonists and that the data in the draft report show an inter- and intra-laboratory variability that is, overall, acceptable. The Panel did note that there were some systematic differences in results from one of the laboratories, but nevertheless concluded that the protocol is transferable.

17. While the Panel would not advocate that further validation studies are essential, should any further validation studies be conducted, it would offer the opportunity to use additional laboratories and test compounds to assist in establishing acceptance criteria for the assay and to further demonstrate the specificity and performance of the assay.

Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used

18. The Panel agreed that this criterion had been met. A logical subset of chemicals from the list compiled by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) had been chosen covering a range of affinities, 7 of which are on the ICCVAM list and the remaining 3 are adequate replacements, representing 6 different chemical classes. While these 10 chemicals do not cover all chemical classes, a sufficient number of other chemical classes were covered by the lead laboratory which tested 46 compounds from the ICCVAM list, demonstrating a strong indication that the protocol and cell line responds appropriately to estrogenic compounds.

The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data

19. The Panel agreed that this criterion had been met. The Panel considered that a good effort had been undertaken to compare the results of the assay to other existing testing methods. Overall, the data provided show a satisfactory concordance between this assay and other *in vitro* TA, binding and *in vivo* uterotrophic assays.

Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of Good Laboratory Practice (GLP)

20. The Panel agreed that this criterion had been met. Although conduct under GLP is ideal, the fact that only the inter-laboratory validation was conducted under GLP was acceptable in this case.

All data supporting the assessment of the validity of the test method should be available for expert review

21. The Panel agreed that this criterion had been partially met. The draft validation report was, for the most part, sufficiently clear to allow independent review, but the Panel raised some specific suggestions and queries (see Annex 2).

22. The draft test guideline is sufficiently detailed to permit others to perform the assays. However, the Panel considered that the draft test guideline is incomplete and requires revision. The Panel's suggestions for additions and amendments to the draft test guideline are given in Annex 2, Appendix 1 . In particular the Panel emphasised the need for more detail in the guideline on (i) the criteria for acceptable test performance, (ii) the criteria for a positive response, and (iii) inclusion of controls for non-specific luciferase activity.

Recommendations

23. The Panel agrees that this report provides a summary of their views on the status of the validation of the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, as detailed in the responses to the questions posed to the Panel and based on the information on the validation exercise provided to the Panel.

24. The report of the Panel, along with the information developed on the validation of the STTA assay, should form the basis for decisions on whether the validation exercise meets the OECD principles for validation for development of this test method into an OECD Test Guideline. The Panel recommends that the OECD consider the Panel report, along with the validation information, to decide on additional work needed to finalise the validation exercise for the purposes of developing an OECD Test Guideline.

Reports provided to the Peer Review Panel

Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity. The Human Estrogen receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line. Ver.2006. Oct.06. Masahiro Takeyoshi, Ph.D., Chemicals Evaluation and Research Institute (CERI), Japan.

Draft OECD Guideline fir the Testing of Chemicals. Stably Transfected Transcriptional Activation (TA) Assay for Detecting Estrogenic Activity of Chemicals. The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line. Ver.2006. Oct.12.

References

ICCVAM (2002). Proposed Substances for Validation of Estrogen Receptor (ER) and Androgen Receptor (AR) Binding and Transcriptional Activation (TA) Assays. October 16, 2002. Available at (accessed 7 March 2007):
http://iccvam.niehs.nih.gov/docs/endo_docs/final1002/sub_list.pdf

OECD (2002). OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals. Organisation for Economic Co-operation and Development, Paris. Available at (accessed 7 March 2007):
http://www.oecd.org/dataoecd/17/33/23652447.doc

OECD (2005). OECD Guidance Document on the Validation and International Acceptance of New or Updated Methods for Hazard Assessment, OECD Series on Testing and Assessment, Number 34. ENV/JM/MONO(2005)14, 18 August 2005, Organisation for Economic Co-operation and Development, Paris. Available at (accessed 7 March 2007):
http://appli1.oecd.org/olis/2005doc.nsf/43bb6130e5e86e5fc12569fa005d004c/4a5c0da258915e7fc1257062002d090d/$FILE/JT00188291.PDF

ANNEX 1

MEMBERS OF THE PEER REVIEW PANEL

Panel member	**Affiliation**
Sélim Aït-Aïssa	Research Biologist, National Institute for Industrial Environment and Risks (INERIS), Verneuil-en-halatte, France
Patrick Balaguer	Institut National de la Santé et de la Recherche Médicale (INSERM), Montpellier, France
Thomas Hartung	Head, European Centre for Validation of Alternative Methods, Ispra, Italy
William Kelce	Vice President, Preclinical Development, Pozen Corporation, Chapel Hill, NC, USA
Steven Levine	Bioassay Team Lead, Ecological Technology Center, Monsanto Company, St. Louis, MO, USA
Craig McArdle	Professor of Molecular Pharmacology, University of Bristol, UK
Vickie Wilson	Research Biologist, US Environmental Protection Agency, Research Triangle Park, NC, USA
Isao Yoshimura	Biostatistician, Faculty of Engineering, Tokyo University of Science, Japan

ANNEX 2

Collated Comments From The Peer Review Panel On The Stably Transfected Transcriptional Activation (Ta) Assay To Detect Estrogenic Activity

ISSUES	COMMENTS AND RECOMMENDATIONS
General comments	The overall impression of the Panel is of a robust assay providing similar results to other *in vitro* TA assays. The tremendous amount of work by the participating laboratories was recognised as was the importance of pressing ahead with the validation of such an assay. The proposed assay has been sufficiently tested to meet most of the OECD validation principles. The validation study provides a basis for a standard protocol for screening for *in vitro* estrogenic activity. The draft test guideline needs some refinement but these results provide a good foundation for OECD to move forward. A number of points however still need to be clarified (see below). A strong criticism from all Panel members was that the criteria for a positive response were unclear and need to be further discussed and decided. There is also insufficient guidance on the criteria for acceptable test performance by laboratories using this assay for the first time. Some Panel members commented that it will be important for OECD also to consider other TA assays that are in development, since there may be differences in scope, cost and availability of materials between various assays.

VALIDATION CRITERIA	
1. Rationale for the test method	The Panel agreed that the rationale for the test method is clearly stated, with the scientific basis, regulatory purpose and need well described. From a regulatory perspective, the test is intended to be used for identifying and prioritising substances that have the potential to act as estrogen receptor (ER) agonists binding to ER-alpha. Under the OECD conceptual framework, it would provide mechanistic information as a level 2 *in vitro* assay. **This Panel agreed that this validation criterion has been fully met.**
2. Relationship between the test method endpoint and the biological phenomenon of interest	The endpoint of the assay is transactivation of a firefly luciferase reporter gene under the control of an estrogen-sensitive hormone response element (ERE). In the case of an estrogen agonist, the process would involve first the binding of the ligand to the estrogen receptor ER-alpha, followed by the ligand-bound receptor binding to and activating the hormone response element on the DNA to produce the gene product, luciferase. These are steps in gene activation currently accepted for the ER in a biological system, even though the actual gene product may differ. A limitation of the test, which is pointed out in both the draft validation report and the draft guideline, is that while it will provide mechanistic information and may allow comparison of ER selectivity (ERα versus ERβ), it will not inform on any downstream estrogenic responses of test chemicals; these would need to be further explored by *in vivo* methods. This is because the test employs an artificial ERE construct (vitellogenin ERE driven by a mouse metallothionein promotor TATA element) that is not expressed endogenously in HeLa cells and so it is not known how faithfully it would mimic the response of endogenous estrogen-responsive genes at similar exposure levels. It was also noted that while this same limitation is shared by other ER assays based on HeLa cell lines, there are cell lines, such as MELN, ER-CALUX and Lumi-cells, which retain most of the classical estrogen-responsive genes. However, the Panel noted that there was good concordance between the rank order of test substances in this TA assay and in the immature rat uterotrophic assay. Other limitations of the assay, such as the fact that it would not detect interference with endogenous hormone metabolism or activity due to production of estrogenic metabolites are acknowledged in the draft report (paragraph 128, p.50) but perhaps need to be highlighted earlier in the Introduction, as is done in para 4 of the draft protocol.

	The Panel agreed that this validation criterion has been fully met.
3. Test method protocol	The Panel agreed that main test method protocol and support protocols are clearly described in most respects. However some questions remain, particularly concerning (i) the criteria for acceptable test performance, and (ii) the criteria for a positive response. **a. Materials and methods** ***Sourcing of materials*** It is unclear whether there are adequate commercial sources of dextran-coated charcoal-treated fetal bovine serum (DCC-FBS) that can be used or whether the Japanese research institute (CERI) expects to be the sole supplier. The Panel considered that, unless there was a special reason for using the particular DCC-FBS produced by CERI, it would be preferable for the protocol to include guidance on how to prepare it or indicate more than one supplier. The draft test guideline indicates that the hER-HeLa-9903 cell line can be obtained from Sumitomo Chemicals Co. The Panel considered that availability of the cell line is important and that the cell line should be deposited with the American Type Culture Collection (ATCC) to ensure that scientists have ready access to it. The Panel considered it would be helpful to include in the draft test guideline some comment about avoiding the use of unsuitable plastic ware, since some plastic materials are known to have ER-agonist activity. ***Edge effects*** An appreciable effort has been made to evaluate edge effects, as discussed in Appendix 4 of the draft validation report, and the authors of the draft report conclude that there is no edge effect that would affect the results for practical purposes. The Panel was not convinced this is the case. From the dose assignment positioning proposed in the draft test guideline, the baseline and vehicle control (DMSO) are potentially vulnerable to edge effects. This can be seen, for example, in experiment 1; although no edge effect was noted with the reference chemical, 17β-estradiol (E2), some significant differences were reported on baseline luciferase expression (DMSO-treated cells) in row H compared to rows A, B, C, D and F (see Appendix 4, p.88-89). This raises the question of whether to include the vehicle control wells in row H of the recommended plate

	layout. This issue of edge effects, which are frequently observed in many laboratories, was also raised in the preliminary validation discussions (see paragraph 4.6, p.158 of the draft validation report). The Panel considered that there should at least be a comment in the draft test guideline on the need to be vigilant regarding the possibility of edge effects and to modify the plate dose assignment if edge effects do occur. ***Magnitude of top dose*** The assay is sensitive enough that a top dose of 10μM, as recommended in the draft protocol, should be acceptable, rather than 1mM as recommended by ICCVAM. The Panel commented that higher concentrations are uncalled for and often result in non-sigmoidal reduced effects due to squelching of available coactivators. ***Replication*** The number of wells per test chemical concentration was different between the pre-validation study (n=4) and the inter-laboratory validation study (n=3) but no rationale for this was provided. Why are 3 replicates, rather than 2, reasonable to achieve the high-throughput assay format (p.15, paragraph 57)? ***Support protocol No. 6-1*** A minor clarification is needed. Under reagents it states "Cell lysis reagent (4.5x): Dilute 10 ml of 5X… reagent with 45 ml…". However, 10 ml of a 5X solution diluted to 55 ml total does not produce a 4.5X final solution (instead is closer to a 1 X solution). **b. What is to be measured** The luciferase produced is measured with a luminometer. The level of sensitivity of a reporter assay is dependent of the sensitivity of the luminometer, the luciferase assay reagent and the general health and characteristics of the cell line. The Panel agreed with the recommendation in paragraph139 (p.52) of the draft validation report that each laboratory undertaking this assay should optimise the best combination of luciferase assay reagent (Glo-type or Flash-type) and luminometer in preliminary testing. Concerning the variation in positive control values for E2 during the validation study, it would have been useful to see raw data for fig. 6 (p.27 of the draft validation report) to know whether the fold increase varied over time as a consequence of variation in the vehicle control – if so, it might make sense to normalise to an internal control maximum response rather than the vehicle control.

	Non-specific luciferase activation The Panel commented that non-receptor-mediated luciferase activation needs to be taken into consideration. For example, luciferase activity is likely to be influenced, either positively or negatively, by cytotoxic events such as oxidative stress, especially at high concentrations (i.e. >1μM). The Panel noted that the proposed protocol does not include any control to assess the specificity of the luciferase induction. As already suggested in the preliminary validation discussions (see Appendix 7 of the draft validation report), co-exposure of test substances with a pure anti-estrogen (e.g. ICI 182,780) and the use of the ER-negative, luminescent HeLa-9903 control cell line could be included to confirm the ER agonist property of tested chemicals, when they are found positive. Use of the HeLa-9903 cell line to assess cytotoxicity would be superior to other cytotoxicity assays (e.g. MTT) as it would also account for substance-induced, non-specific effects on protein (luciferase) expression and luciferase activity. However, a question was raised about the sensitivity of the HeLa-9903 cell line to detect small changes in cytotoxicity. Constitutive RSV driven luciferase expression is usually very robust in HeLa cells, so there is some concern that this cell line may not be as sensitive to cytotoxicity effects which can manifest in the test cell line. In the development of this construct and this cell line, was the level of expression designed to be similar to that induced in the test cell line with 25pM estradiol stimulation? **c. How the results should be analysed** The Panel agreed that statistical analysis of the results should utilise the simplest methods available that would be compatible with the characteristics of the particular index value being calculated (e.g. EC50 or PC50) and with the way in which the calculated index values are to be used in decision-making on a positive or a negative outcome. The Panel noted that the use of a regression line based on the Hill equation is typically used for competitive binding data. Competitive binding is based on the law of mass action and is a function of the association and dissociation rates for the ligand for the receptor and assumes that an equilibrium rate is reached when the rate of new ligand-receptor complex formation equals the dissociation rate. Whether a Hill equation-based nonlinear regression fit is appropriate for TA data is debatable since a TA assay involves more than just the ligand binding to the receptor. In the case where a full dose-response for the test substance is attained, either method (linear interpolation or nonlinear regression) may give satisfactory results. As with binding data, when a full curve can not be attained for weak estrogens some criteria need to be defined by which compounds can be classified as to response. In this case,

	the determination of a PC50 value (by linear regression) compared to the positive control gives added useful information. The Panel noted that the independent statistical analyses of the inter-laboratory validation study (Appendix 6 of the draft validation report) concluded that for estimating logPC10, Hill equation-based nonlinear regression has advantages over linear interpolation in terms of accuracy and precision, but that for logPC50, there was not much difference between the two methods. The Panel agreed that if the calculated index values are used as a mean for qualitative expression of the intensity for endocrine disrupting potential, analogous to LD50 for mortality, the method that provides a more precise estimate, i.e. the Hill equation-based nonlinear regression method, should be used. However, if it is decided only to classify test chemicals into one of three categories - positive, negative or equivocal - then such precise values are not needed and a simpler approach could be used. The Panel also noted that the measurements obtained in the pre-validation and inter-laboratory studies provided data of high quality. However, in daily use, various factors such as the skill of operator, time and temperature of propagation, etc, may disturb the monotonic dose-response relationship. This should be taken into consideration in selecting the method of data analysis. The linear interpolation method is probably not sufficiently robust for such disturbances. The Panel raised the issue of partial ER agonists that reach a plateau at a lower level than the maximal effect by E2. In such cases, what should be taken into account: EC50 or PC50? This point is not addressed by the authors. Paragraph 76 (p.24) of the draft validation report and paragraph 20 (p.6) of the draft protocol point out that EC50 is calculated where Hill's equation is applicable but there is no guidance offered on how to decide if it is applicable. **d. Criteria for acceptable test performance** The Panel considered that the performance criteria that would need to be met by a new laboratory performing the assay for the first time should be outlined more clearly. It is not uncommon to include such criteria in OECD test guidelines and they are usually based on the outcome of a large ring test. At the least, EC50 values for positive controls should meet a specified tolerance interval or range. Other possible considerations for acceptable test performance are mentioned below. Consideration should be given to including a positive control substance, with significantly less potency than E2 but still able to produce a full dose-effect curve, in the methodology for the validated assay. However, it should also be considered essential for quality control purposes and to verify the performance of the assay.

Paragraph 75 (p.23-24) of the draft validation report points out that positive control responses were always >4 fold above the vehicle control – would the data be used if it were not?

The plate configuration uses E2 at a single reference concentration. How often should E2 dose-response curves be run to ensure an adequate sigmoid response and ED50 values that are consistent with historical controls? This could be critical to track genetic drift in the cell line and week-to-week cell line variability.

Paragraph 52 (p.15-16) of the draft validation report points out that compounds reducing baseline luciferase activity would be considered cytotoxic – does this mean that such compounds would be entirely excluded from the test? In many systems ER antagonists specifically reduce the base-line luciferase activity (possibly by blocking effects of estrogenic compounds residual in the fetal bovine serum or derived from other sources such as the cells themselves). A concern is that a pure ER antagonist would be defined as a cytotoxin in this assay. This concern was also expressed in the preliminary validation discussions (see Appendix. 2.14).

e. Criteria for a positive response

The Panel was in agreement that there was insufficient clarity on what would constitute a positive response. Equally, it is not clear from a regulatory perspective, what defines a "negative" in this assay, i.e. what level of ER agonist activity is not considered important and would trigger "NO further testing".

In discussing what might be suitable criteria for a positive response, the Panel emphasised the need to first see whether there are significant differences and a dose-response (by running an ANOVA). The presence of a dose-response is an important contributor to confidence in identifying positives. The validation group needs to propose what would be judged as a significant dose-response. Will a trend test be run and if so which one is recommended? Will expert judgment be used to evaluate the shape of the dose-response curve and if so how will it be factored into the analysis?

A test substance having a quantifiable PC50 should certainly be regarded as positive.

The Panel was of the view that the PC10 value is not appropriate to be used as a criterion of a positive response because it represents a marginal response of doubtful biological predictive value and is likely to generate a high rate of false positives (see also further comments below).

The Panel discussed whether some intermediate value between PC10 and PC50 (e.g. PC25) might also be

	considered as a positive response. The Panel acknowledged that although PC25 sounded reasonable, there were no data on which to base a decision to use it. The Panel felt that further statistical advice might be needed on this point. The Panel also made other comments on more detailed aspects of the proposed calculated index values and these are set out below. ***Interpretation of the PC50 value*** The PC50 value is defined as the test chemical concentration eliciting transcription activity equivalent to 50% of the positive control value. A concern is that the PC50 value could be misinterpreted. Sigmoid dose-response curves are normally described in terms of EC50, maximum and Hill co-efficient. The EC50 provides a measure of "potency" (position of the curve along the horizontal axis) whereas the maximum measures "efficacy". This distinction of potency from efficacy is important and informative. The problem with the PC50 value is that is influenced by both potency and efficacy but may be used erroneously as a measure of potency alone (e.g. treated as if it were an EC50). Thus, two compounds having identical potency (e.g. identical EC50) and different efficacy (e.g. different maxima) would have different PC50 values and this could be taken, mistakenly, to reveal different potencies. The text of the draft report is rather vague on this aspect and there needs to be a clearer discussion of the issue. Table 14 (p.34 of the draft validation report) reports the "relative potency in reporter gene assay (E2=100)" but again it is unclear what this means. Is it showing the PC50 of the test compounds normalised to the EC50 of E2? Since PC50 does not actually measure potency (in the pharmacological sense) the PC50/EC50 normalisation is a hybrid of potency and efficacy, not a relative potency. Comparison of (relative) binding affinities and potencies is extremely informative but this comparison is not made in tables 13 and 14 because PC50 is not a measure of potency. Paragraph 92 (p.29 of the draft validation report) states that "the PC50 values can be calculated in the cases of weak estrogenic compounds as the relative estrogenic activity to the natural estrogen" but it is not evident what this means. The Panel noted that the pre-validation and the validation work used two different E2 levels for the PC50 calculations, but it is strange that they both gave similar responses. Furthermore, in the draft test guideline, a single E2 concentration is proposed for the positive control but no rationale is given for choosing this concentration. The protocol should give reasons for the choice of E2 concentration to which the levels of response of test compounds should be compared. Where a maximum cannot be calculated, simpler measures such as minimal effective concentration and maximal

	response can be informative. In paragraph 107 (p.37 of the draft validation report) it is argued that is better to use PC50 values than EC50 values (because the latter were not calculable/accurate in some instances), but a calculable PC50 value just means that a maximum response >50% of the E2 maximum was obtained. The comparison with the uterotrophic assay is based on a simple positive or negative response. It would have been simpler to just calculate maximum test response as a percentage of the maximum E2 response and treat this as positive when it exceeds 50%. This is conceptually easier, less prone to misinterpretation and negates the arguments about how best to assess the PC50 value. In the footnote to Table 16 (p. 36 of the draft validation report), it is stated "Positive/Negative based decision of stably transfected TA assay was made based on the PC50 values." It is not clear what this means. Does it imply that a chemical is classified as negative if the potency at the highest dose is below 50%? If so, the evaluation of PC50 is not necessary. A similar question arises from Table 12. Clarification is needed. ***Is the PC10 value informative?*** The Panel had several reservations about the usefulness of the PC10 value. If the PC10 value were to be used as an indicator of a positive response, it may lead to identification of a lot of false positives when testing a broad array of environmental compounds. Does this somewhat "marginal" response have any biological significance, especially when the positive control concentration is at or below its EC50? If it were to be used, should the increase attained by the test chemical also be a statistically significant increase over vehicle control (when analyzed by ANOVA)? A PC10 value is also likely to have little predictive value of *in vivo* ER agonist activity given the levels of endogenous estrogens *in vivo*. The results from the validation study have an unacceptably high false-positive rate when assessing the PC10 endpoint. For example, in Table 6.2 (p.114-115 of the draft validation report) there were false positives observed in 1 of the 3 replicate assays for 3 of the 9 test substances for 3 of the labs. Thus, the results are equivocal for nearly one-third of the substances tested in the validation when using the PC10 endpoint. This is considered to be a reflection of the variability of the test system at low induction levels and raises a serious concern regarding use of the PC10 value as an endpoint. PC10 is also an unreasonably low endpoint to use, particularly considering the nonlinear nature of the response being measured. Should a minimum PCx value no less than a 25% effect level be considered? The minimum value of 25% was recommended by ICCVAM and although this value is not supported by a statistical assessment it reflects expert judgment.

	The Panel did however note that in Table 16 (p. 36 of the draft validation report), three chemicals (4,4'-(Octahydro..)bisphenol, Levonorogestrol and Diphenyl-p-phenyldiamine) were weak estrogens *in vitro* (no PC50 reported) but still positive *in vivo*. For these chemicals, the PC10 had thus higher predictive value than the PC50. **The Panel agreed that this validation criterion has been partially met.** **It has not been met with respect to the important issue of clearly defining and justifying criteria for a positive response.**
4. Intra- and inter-laboratory reproducibility (reliability)	The Panel concluded that the design of the inter-laboratory studies was generally appropriate and sufficient to establish reproducibility of the protocol for estrogen agonists and that the data in the draft report show an inter- and intra-laboratory variability that is, overall, acceptable. The protocol is transferable. However, the Panel noted that some obviously aberrant data were reported in the Table 20 (page 43-44), for genistein (vial 23-3), 17a-MT (vial 26-3) and p-tert-pentylphenol (vial 34-1) and that this was not discussed by the authors. The draft validation report also fails to highlight a potentially important difference between performance in different labs. Table 18 (p.39-40 of the draft validation report) shows that PC50 values were always lower in CERI assays than in Sumitomo assays (7/7 compounds – difference 0.14-0.56 log10 units) and table 20 (p.43-44) shows that EC50 values were always lower in CERI assays than Sumitomo assays (7/7 compounds – differences 0.1-2.41 log10 units). There needs to be some comment on this kind of systematic variation (rather than just random variation) since it could have major implications for protocol transferability. The Panel noted that fewer laboratories had been involved in the validation study than would normally be the case. While the Panel would not advocate that further validation studies are essential, should any further validation studies be conducted, it would offer the opportunity to use additional compounds from the ICCVAM list and to better define the criteria of tolerance for acceptable performance of the assay. **The Panel agreed that this validation criterion has been met.**
5. Testing of reference chemicals	The reference chemicals, including non-active substances, are clearly described and the major rationale for their choice (comparison with existing data) is explained. Although only 10 coded chemicals were used for tests of inter-laboratory reproducibility, since the reliability criteria have been met, it follows that the number is sufficient. A logical subset of chemicals has been chosen, covering a range of affinities, seven of which are on the ICCVAM

	list and the remaining three are adequate replacements, representing six different chemical classes. While these 10 chemicals do not cover all chemicals classes, a sufficient number of other chemical classes were covered by the lead laboratory which tested 46 compounds from the ICCVAM list, demonstrating a strong indication that the protocol and cell line responds appropriately to estrogenic compounds. The fact that not all 78 ICCVAM recommended compounds were tested is not a serious flaw. The compounds included in the ICCVAM list are meant to be used to potentially validate a wide range of assays (i.e. AR, ER and TR mediated and agonists and antagonists in some cases). Therefore, some chemicals are more appropriate for evaluating each type of assay as responses for every chemical on the list have not been well characterized for every type of assay. There would be little point in testing all 78 chemicals listed by ICCVAM when the required comparator (EC50 or IC50) is unavailable. Additional insight into the limitations of the assay could be made by testing further substances, such as more from the ICCVAM list to cover more than the 6 out of 15 classes tested so far, and including substances with difficult physico-chemical characteristics (low solubility) or overtly cytotoxic substances. However, further validation should not hold up progress in developing the assay. Modifications to the protocol and an assessment of the limitations of the protocol and the assay could be made on a continuing basis as new information becomes available. **The Panel agreed that this validation criterion has been met.**

6. Performance of the test method in relation to existing toxicity testing data	The Panel considered that a good effort had been undertaken to compare the results of the assay to other existing testing methods. Overall, the data provided show a satisfactory concordance between this assay and other *in vitro* TA, binding and *in vivo* uterotrophic assays. The Panel agreed that there is no "gold standard" with which to compare these types of results. The most relevant comparison is the ER binding assay comparison with which there was high concordance and good comparative sensitivity and specificity, though a little more detail on how sensitivity and specificity were calculated would have been helpful. Binding assays, however, do not differentiate between agonists and antagonists. Therefore, the comparisons of the reference chemicals in the ICCVAM list to the results of other ER TA assays and to uterotrophic assay results, adds strength to the conclusions on performance of this assay. A number of specific comments and queries were raised: It is not entirely clear what "good concordance" or "acceptable concordance" mean. Table 11 (p.31 of the draft validation report) claims to show good concordance (80%) but 5 of 24 compounds found positive in the ICCVAM (2003) report were negative in the ER/TA assay, and 4 of 22 compounds found negative in the ICCVAM report were positive in the ER/TA assay. The Panel recognises that there is no OECD criterion for what constitutes acceptable concordance but there needs to be some discussion of why this 20% lack of concordance exists, rather than a simple assumption that 80% is good. Although the assay compares well with other *in vitro* TA tests, the comparison is made on a very limited number of chemicals (n=7 or 8), which limits the strength of the assessment. One surprising result is the slope of 0.712 when comparing with HELN-ERa, which could indicate a higher sensitivity of the HELN cells compared with hER-HeLa-9903 cells. This is relatively unexpected since both assays are done in HeLa cells and the HeLa-ER-9903 showed a high sensitivity to E2. This may reflect a bias in the comparison due to a few number of chemicals included in the correlation analyses. Table 10 (p.29 of the draft validation report) refers to only 3 references related to the other *in vitro* tests. More data are available in the literature. For instance, HELN-ERa cells have been well characterised in response to pesticides or phenol compounds (Lemaire *et al.*, 2006, Life Sci, 79, 1160-1169; Paris *et al*., 2002, Mol Cell Endocrinol, 193, 43-49 ; Balaguer *et al*., 1999, Sci Tot Environ 233, 47-56). Also, more

	bibliographic data should be found for ER-CALUX. A more accurate comparison between models could thus be obtained by enhancing the number of chemicals in the correlation. Fig. 8-1 (p.30 of the draft validation report) shows the relationship between hER-HeLa-9903 EC50's and Reference EC50's, but the latter are derived from 3 separate assays (mammalian TA, proliferation, yeast reporter). It would be preferable to see this correlation for each of the Reference assays. The correlation coefficient (R^2=0.802) is described as "successful" (para. 93, p.29) what does this mean? In Table 14 (p.34 of the draft validation report), it is not clear whether the chemicals with minus "-" are negative or are positive but the PC50 was not calculated. Adding the PC10 to the table would be helpful to distinguish between these two types of chemicals. For instance, methoxychlor, a well-known weak estrogen (positive in Tables 10 and 16) is described in Table 14 as negative in TA assay but positive in the binding assay. Considering this pesticide as negative may be an artefact in the concordance analysis. It may be informative to know whether or not this particular cell line expresses ABC transporting cassettes like MDR1/PgP that can limit intracellular exposures to some chemicals and lead to differences with other cell lines *in vitro* and differences with *in vivo* ER agonist endpoints. **The Panel agreed that this validation criterion has been met.**
7. GLP	The Panel noted that whilst the inter-laboratory validation was conducted to GLP, the pre-validation and data collection for comparison with ICCVAM were not. However, the Panel agreed with the views of the preliminary validation assessment panel (p.151 of the draft validation report) that although conduct under GLP is ideal, the fact that only the inter-laboratory comparison was conducted under GLP was acceptable in this case. It was also noted that the identity, purity, strength or homogeneity of the test article did not appear to have been characterised by any laboratory. While this should not negate any of the findings as the substances chosen for assessment were stable, for future studies, characterization of the test article is a required part of GLP and should be completed. **The Panel agreed that this validation criterion has been met**

8. Availability of data supporting the assessment of the validity of the test method	**a. Draft validation report** For the most part, the Panel considered that the draft validation report was sufficiently clear to allow independent review. However, some specific suggestions and queries were raised: The raw data in the draft validation report are not as transparent they could be. Provision of graphs (such as histograms with SE bars) of the actual data on test substances and controls from the inter-laboratory study would have been helpful. Since one concentration of E2 was used to determine PC50 and PC10 values for other test compounds, it would have been especially useful to see the complete estradiol (E2) curve so that it could be seen whether a reasonable E2 positive control concentration was used. This type of comparison should be made to the maximal E2 response (i.e. where you get 80 to 100 % of the overall response, but not beyond). Based on the EC50 information for E2, it appears that the positive control concentration used in the pre-validation studies of 100 pM may be too low for making PC50 or PC10 conclusions, raising a reservation about the use of the PC value as an indicator of a positive response in that phase of the work. However, in the inter-laboratory study a positive control of 1 nM was used which would seem to be closer to what would be expected for a maximal E2 response, although there is no information with which to confirm that. In paragraph 43 (p.13 of the draft validation report) it should be clarified that "transient" transfections were used not "mock" transfections, which means something very different. Mock transfections are performed with an "empty" plasmid. Also, if stimulating ligands (i.e. estradiol) were added to the transiently transfected cells to asses the presence/absence for the different receptor types, this should be added to the text. **b. Draft test guideline** The draft test guideline is sufficiently detailed to permit others to perform the assays, especially if the test lab already has some cell culture experience, however, overall the draft test guideline could be better written and is very incomplete. Suggestions for additions and amendments are given in Appendix 1. In particular the Panel wished to reiterate the need for more detail concerning (i) the criteria for acceptable test performance, (ii) the criteria for a positive response, and (iii) inclusion of controls for non-specific luciferase activity. Careful consideration of cytotoxicity is critical to ensure that only responses at non-toxic doses are considered. It was recommended in the ICCVAM report that concurrent with reporter assays, an identically treated parallel "satellite" assay should be run and evaluated for cytotoxicity. Additionally, ICVAAM

	recommended that only dose levels not associated with greater than 10% cytotoxicity should be included in the analysis. The criterion for cell toxicity in the validation report was significantly higher and listed as an 80% effect. There is concern that the proposed method of evaluating basal luciferase activity as an index of cytotoxicity is open to confounding factors. If a substance is an antagonist, it could suppress luciferase activity below basal levels without resulting in cell toxicity. Additionally, since the proposed method considers evaluating PC values that are less than 50% as an endpoint, it is critical that a sensitive, reliable, and low variability method be adopted to assess cell toxicity. There appears to be some confusion between the draft test guideline and the protocols in Appendices 2, 3 and 5 of the draft report with respect to the plate assignment for the test chemicals. It would be preferable to include in the test guideline both the plate dose assignment illustrated in the draft test guideline and Appendix 3 (i.e. including a full E2 dose response on every plate) and the plate assignment for cytotoxicity as described in Appendix 5 (p.98). It is ICCVAM's recommendation that a full dose-response for the positive control should be included with each plate. A sentence could be added to the draft test guideline to indicate how to expose the cells to the chemicals from DMSO stock solutions. **The Panel agreed that this validation criterion has been partially met.**

OTHER ISSUES	
	The Panel was in agreement that, even though a protocol for anti-estrogenic activity is included, **reproducibility and acceptable performance has only been demonstrated for estrogen agonists**. If the assay were to be used for estrogen antagonists, further testing would need to be done. However, this is not a reason to hold up development of the assay for agonists. Alternatively, to maximize the efficiency of this effort, the agonist and antagonist protocols could be co-validated. The rationale for this is related to the strengths of the ER TA assay to both detect and differentiate between agonists and antagonists, whereas an ER binding assay can detect both agonists and antagonists but cannot differentiate between them. However, it is acknowledged that the antagonist assay should be given lower priority since there are very few examples of ER antagonists compared to ER agonists.

APPENDIX 1

Suggestions for additions and modification of the draft test guideline

Paragraph 1: For validation, it is recommended that in addition to the reference substance 17 β-estradiol that a positive control substance be included on each plate. The positive control would be a substance with significantly less potency than 17 β-estradiol but still able to produce a reproducible dose-response. Having a positive control should be considered optional for the validated assay but should be included in a validation for quality control purposes and to verify the performance of the assay.

Paragraph 5: It is acknowledged that this assay has a deficiency, detection of a non-receptor mediated response at high levels of phytoestrogens. I question whether the assay can be fully validated when responses are observed that call into question the specificity of this assay for ER mediated action.

Paragraph 6: This is a very awkward paragraph and factually incorrect. A revised version of the paragraph is below.

The transcriptional activation assay using a reporter gene technique is an in vitro tool that provides mechanistic data. The assay is used to signal binding of the estrogen with a ligand. Following ligand binding, the receptor-ligand complex translocates to the nucleus where it binds specific DNA response elements and transactivates a firefly luciferase reporter gene. Luciferin is a substrate that is transformed by the luciferase enzyme to bioluminescence and can be quantitatively measured with a luminometer. Luciferase activity can be evaluated quickly and inexpensively with a number of commercially available kits.

Paragraph 11: There needs to be additional information regarding maintenance of the cell culture. For example, recommended sources for plastic-ware used to culture cells, guidance on the effect of passage number on the response which is critical, method to monitor the stability of the stably transfected cell line, maintenance of the stably transfected cell line (i.e., details regarding antibiotic selection requirements needed to maintain the stably transfected cell line).

Surprisingly, there was no guidance on the volume of media required for each well. There needs to be a recommendation (e.g. 100 µL as in supporting material) in light of all of the work done to investigate potential edge effects.

The source of 10% dextran-coated-charcoal (DCC)-treated fetal bovine serum (DCC-FBS) is potentially a large source of variability in the assay. There should be an appendix in the guideline that provides guidance on preparation of the DCC-FBS. This information is in paragraph 70 of the validation report.

Current text: "*Free radioligand was removed by incubation with 0.2% activated charcoal and 0.02% dextran in PBS (pH 7.4) for 10 min at 4°C followed by filtration*."

Paragraph 12: There is no limit level in the draft guideline defining the maximum concentration of vehicle. Typically, the level should not exceed 0.1% (v/v). Suggest the following wording change:

"Test substance should be dissolved in a solvent that solubilizes that test substance and is miscible with the cell medium. Water, ethanol (95% tom 100%) and dimethylsulfoxide (DMSO) are suitable vehicles. If DMSO is used the level should not exceed 0.1% (v/v). For any vehicle, it should be demonstrated that the maximum volume used is not cytotoxic and does not interfere with assay performance."

Additionally, it is highly recommended that the definitive screening assays should be replicated three times on different days rather than only performing them on a single assay day.

Paragraph 19: The analyst is told to wait minimally 5 min before measuring luciferase activity but there is not maximum time listed. A maximum time should be added based on the manufacturer's recommendation (e.g., 2 hours).

Paragraph 20: There needs to be rationale provided for choosing either 1 nM E2 or 100 pM E2 for the positive control. The y-axis is not labelled. It should be labelled "Average Fold Induction Relative to the Vehicle control." It is important to include in the protocol that if dose-response modelling is performed, the responses in the treatment groups should not be divided by the mean response of the replicates from the vehicle controls. This is not clear to the reader. If this normalization is performed, then the 4th parameter from the equation, the y-intercept must be constrained to equal zero. I also recommend to provide a better description of the 4-parameter equation and to cite the correct citation, De Lean et al., 1979.

Paragraph 21: The aforementioned cytotoxicity issue needs to be addressed in this paragraph. How is cytotoxicity to be evaluated in parallel? Will visual observations for cell toxicity be made and what constitutes cytotoxicity based on visual observations.

Paragraph 22: How will stability of the test compounds be judged? Chemicals that are hydrophobic or have high vapour pressure are difficult to evaluate in multi-well culture plates; volatile compounds may generate false positives in nearby control wells (DeCastro et al. 2006).

Recommendations must be provided on how to handle 17β-estradiol as to not contaminate the laboratories with this highly potent ER ligand.

Paragraph 23: An acceptance criterion should be included setting a minimum fold-induction for a positive control and a reference compound. Also, the language in this paragraph needs to be more definite. The current wording is incomplete and inadequate. For example, the term "in general" needs to be revised to give clearer guidance on a positive response. The guideline states that dose-responsiveness should also be considered but provides absolutely no guidance on the topic. Based on the results generated to date, determination of a PC10 is not an acceptable endpoint to judge a response as positive. Further work needs to be done before criteria can be proposed for what is positive if a PC50 cannot be calculated. The group needs to propose what is judged as a significant dose-response? Will a trend test be run and if so which one is recommended? Will expert judgment be used to evaluate the shape of the dose-response curve and if so how will it be factored into the analysis? It is critical to provide criteria that incorporate sound statistical methods and sound scientific judgment for classifying substances to ensure the credibility of the results.

The response for the reference chemical should be within an acceptable historical range.

Recommend changing from: "*In general, when PC50 can be calculated, a test chemical is considered as positive in hERα mediated transcriptional activation. In this case, dose responsiveness should be considered.*"

To the following:

1) The response for the reference substance (E2) and controls should be within the appropriate historical acceptance range. The test method protocol defines acceptance criteria as follows:

- *EC50 of the reference substance (E2) must be within 2.5 standard deviations (SD) of the historical mean established by the test laboratory and have a coefficient of determination value ≥ 0.9 calculated by four parameter Hill equation.*
- *At least one data point on each EC50 plot of the positive control and the test chemical must be 10% - 50% of the maximum response and at least one datapoint must be 50% - 90% of the maximum response.*
- *At least two data points must be < 10% of the maximum response and these points constitute the bottom plateau of the concentration response curve.*

- *The standard deviation of all vehicle controls should not be more than 15% of the mean.*
- *The standard deviation of all negative controls should not be more than 15% of the mean.*

*2) The minimum fold-induction for the positive control must minimally be **x**-fold for an assay to be considered valid.*

3) A substance is considered positive and classified as an ER-α agonist if a PC50 value can be calculated.

4) The study should comply with GLP standards and the impact of GLP deviations must be discussed in the report.

References:

ICCVAM Evaluation of *in vitro* test Methods for detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. (2002) (http://:iccvam.niehs.nih.gov/methods/endocrine.htm)

DeCastro BR, Korrick SA, Spengler JD, & Soto AM. 2006. Estrogenicity of polychlorinated biphenyls present in human tissue and the environment. Environ Sci Technol. 40(8):2819-25.