An 11-point 1:2 serial dilution should be used if the resulting concentration range will encompass the full range of responses based on the concentration response curve generated in the range finder test. Otherwise, use a 1:5 dilution.

- If a substance exhibits a biphasic concentration response curve in the range finder test, both phases should also be resolved in comprehensive testing.

Comprehensive Tests

29. Comprehensive testing consists of 11-point serial dilutions (either 1:2 or 1:5 serial dilutions based on the starting concentrat*ion for comprehensiv*e testing criteria) with each concentration tested in triplicate wells of the 96-well plate (see Figure 2). Comprehensive testing uses 11 concentrations of E2 (Table 2) in duplicate as the reference standard. Four replicate wells for the DMSO control and three replicate wells for the methoxychlor control (9.06 x 10^{-6} M) are included on each plate.

Figure 2 Comprehensive Test 96-well Plate Layout

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	TS1-1	TS1-2	TS1-3	TS1-4	TS1-5	TS1-6	TS1-7	TS1-8	TS1-9	TS1-10	TS1-11	VC
B	TS1-1	TS1-2	TS1-3	TS1-4	TS1-5	TS1-6	TS1-7	TS1-8	TS1-9	TS1-10	TS1-11	VC
C	TS1-1	TS1-2	TS1-3	TS1-4	TS1-5	TS1-6	TS1-7	TS1-8	TS1-9	TS1-10	TS1-11	VC
D	TS2-1	TS2-2	TS2-3	TS2-4	TS2-5	TS2-6	TS2-7	TS2-8	TS2-9	TS2-10	TS2-11	VC
E	TS2-1	TS2-2	TS2-3	TS2-4	TS2-5	TS2-6	TS2-7	TS2-8	TS2-9	TS2-10	TS2-11	Meth
F	TS2-1	TS2-2	TS2-3	TS2-4	TS2-5	TS2-6	TS2-7	TS2-8	TS2-9	TS2-10	TS2-11	Meth
G	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8	E2-9	E2-10	E2-11	Meth
H	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8	E2-9	E2-10	E2-11	Meth

Abbreviations: TS11-1 to TS1-11 = concentrations (from high to low) of test substance 1; TS2-1 to TS2-11 = concentrations (from high to low) of test substance 2; E2-1 to E2-11 = concentrations of the E2 reference standard (from high to low); Meth = p,p' methoxychlor weak positive control; VC = DMSO (1% v/v) EFM vehicle control

30. Repeat comprehensive tests for the same chemical should be conducted on different days, to ensure independence. At least two comprehensive tests should be conducted. If the results of the tests contradict each other (e.g., one test is positive, the other negative), or if one of the tests is inadequate, a third additional test should be conducted.

Measure of Luminescence

31. Luminescence is measured in the range of 300 to 650 nm, using an injecting luminometer and with software that controls the injection volume and measurement interval. Light emission from each well is expressed as RLU per well.

ANALYSIS OF DATA

EC_{50} Determination

32. The EC_{50} value (half maximal effective concentration of a test substance is determined from the concentration-response data. For substances that are positive at one or more concentrations, the concentration of test substance that causes a half-maximal response (EC_{50}) is calculated using a Hill function analysis or an appropriate alternative. The Hill function is a four-parameter logistic mathematical model relating the substance concentration to the response (typically following a sigmoidal curve) using the equation below:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log IC50 - X)\text{HillSlope}}}$$

where Y = response (i.e., RLUs); X = the logarithm of concentration; Bottom = the minimum response; Top = the maximum response; log IC_{50} (or log EC_{50}) = the logarithm of X as the response midway between Top and Bottom; and Hillslope describes the steepness of the curve. The model calculates the best fit for the Top, Bottom, Hillslope, and EC_{50} parameter. For the calculation of EC50 values, appropriate statistical software should be used (e.g. Graphpad Prism^{R} statistical software.

Determination of Outliers

33. Good statistical judgment could be facilitated by including (but not limited to) the Q-test (see protocol (7) for determining "unusable" wells that will be excluded from the data analysis.

34. For E2 reference standard replicates (sample size of two), any adjusted RLU value for a replicate at a given concentration of E2 is considered an outlier if its value is more than 20% above or below the adjusted RLU value for that concentration in the historical database.

Collection and Adjustment of Luminometer Data for Range Finder Testing

35. Raw data from the luminometer should be transferred to a spreadsheet template designed for the test method. It should be determined whether there are outlier data points that need to be removed. (See Test Acceptance Criteria for parameters that are determined in the analyses.) The following calculations should be performed:

Step 1 Calculate mean value for the DMSO vehicle control (VC).
Step 2 Subtract the mean value of the DMSO VC from each well value to normalize the data.
Step 3 Calculate the mean fold induction for the reference standard (E2).
Step 4 Calculate the mean EC_{50} value for the test substances.

Collection and Adjustment of Luminometer Data for Comprehensive Testing

36. Raw data from the luminometer should be transferred to a spreadsheet template designed for the test method. It should be determined whether there are outlier data points that need to be removed. (See Test Acceptance Criteria for parameters that are determined in the analyses.) The following calculations are performed:

Step 1 Calculate mean value for the DMSO VC.
Step 2 Subtract the mean value of the DMSO VC from each well value to normalize the data.
Step 3 Calculate the mean fold induction for the reference standard (E2).
Step 4 Calculate the mean EC_{50} value for E2 and the test substances.
Step 5 Calculate the mean adjusted RLU value for methoxychlor.

Data Interpretation Criteria

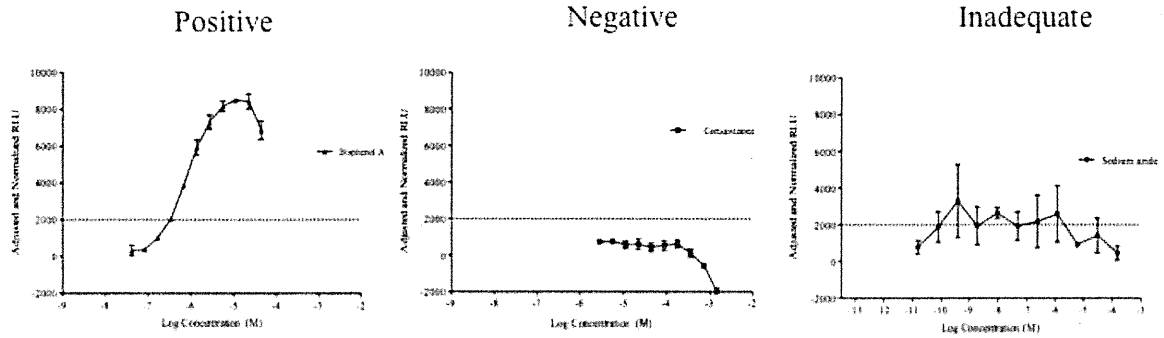
37. The BG1Luc ER TA is intended as part of a weight of evidence approach to help prioritize substances for ED testing *in vivo*. Part of this prioritization procedure will be the classification of the test substance as positive or negative for ER agonist activity. The positive and negative decision criteria used in the BG1Luc ER TA validation study is described in Table 1.

Table 1: Positive and Negative Decision Criteria

Positive	• All test substances classified as positive for ER agonist activity should have a concentration–response curve consisting of a baseline, followed by a positive slope, and concluding in a plateau or peak. In some cases, only two of these characteristics (baseline–slope or slope–peak) may be defined. • The line defining the positive slope must contain at least three points with nonoverlapping error bars (mean ± SD). Points forming the baseline are excluded, but the linear portion of the curve may include the peak or first point of the plateau. • A positive classification requires a response amplitude, the difference between baseline and peak, of at least 20% of the maximal value for the reference estrogen (i.e., 2000 RLUs when the maximal response value of the reference estrogen is adjusted to 10,000 RLUs). • If possible, an EC_{50} value should be calculated for each positive substance.
Negative	The average adjusted RLU for a given concentration is at or below the mean DMSO control RLU value plus three times its standard deviation.
Inadequate	Data that cannot be interpreted as valid for showing either the presence or absence of activity because of major qualitative or quantitative limitations are considered inadequate and cannot be used to determine whether the test substance is positive or negative.

38. Data interpretation criteria are shown in Table 4. Positive results will be characterized by both the magnitude of the effect and the concentration at which the effect occurs, where possible. Examples of positive, negative and inadequate data are shown in Figure 3.

Figure 3: Examples: Positive, Negative and Inadequate Data



Dashed line indicates 20% of E2 response, 2000 adjusted and normalized RLUs.

39. The calculations of E C_{50} can be made using a four-parameter Hill Function (see protocol for more details (7)). Meeting the performance standards indicate the system is operating properly, but it does not ensure that any particular run will produce accurate data. Duplicating the results of the first run is the best assurance that accurate data were produced.

Test Report

40. See paragraph 20 of the Common Elements to all methods.

LITERATURE

(1) ICCVAM, ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (BG1Luc ER TA) Test Method An *In Vitro* Method for Identifying ER Agonists and Antagonists. 2011, National Institute of Environmental Health Sciences: Research Triangle Park, NC.

(2). Monje P, Boland R. 2001. Subcellular distribution of native estrogen receptor α and β isoforms in rabbit uterus and ovary. J Cell Biochem 82(3): 467-479.

(3) Pujol P, Rey JM, Nirde P, Roger P, Gastaldi M, Laffargue F, et al. 1998. Differential expression of estrogen receptor-alpha and -beta messenger RNAs as a potential marker of ovarian carcinogenesis. Cancer Res 58(23): 5367-5373.

(4) Weihua Z, Saji S, Mäkinen S, Cheng G, Jensen EV, Warner M, et al. 2000. Estrogen receptor (ER) β, a modulator of ERα in the uterus. Proceedings of the National Academy of Sciences of the United States of America 97(11): 5936-5941.

(5) Balls, M., et al., The importance of good cell culture practice (GCCP). ALTEX, 2006. 23(Suppl): p. 270-273.

(6) Coecke, S., et al., Guidance on good cell culture practice: a report of the Second ECVAM Task Force on good cell culture practice. Alternatives to Laboratory Animals, 2005. 33: p. 261-287.

(7) NICEATM, 2011, Detailed protocols for the BG1Luc Estrogen Receptor Transcription Activation Assay (agonist and Antagonist)

(8) Rogers JM, Denison MS. 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, 2000;13(1):67-82.

(9) Escande A, Pillon A, Servant N, Cravedi JP, Larrea F, Muhn P, Nicolas JC, Cavaillès V, Balaguer P. Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta. Biochem Pharmacol2006 May 14;71(10):1459-69.

(10) Thorne N, Inglese J, Auld DS. Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology. Chemistry and Biology2010;17(6):646-57.

(11) Kuiper GG, Lemmen JG, Carlsson B, Corton JC, Safe SH, van der Saag PT, van der Burg B, Gustafsson JA. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology1998 Oct;139(10):4252-63.

22 December 2011

Draft Performance Standards for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Agonists (For TG 455)

INTRODUCTION

1. The following Performance Standards (PS) accompanies the Performance Based Test Guideline for Transfected Transactivation *In Vitro* Assays to Detect Estrogen Agonists (for TG 455). This document is intended as a guide to developers of new test methods that are analogous to existing, fully validated test methods in that they are based on similar scientific principles and predict the same effect (colloquially referred to as "me too" tests) (1). Prior to the acceptance of a new test method for regulatory testing applications, validation studies are conducted using scientifically sound principles to establish its reliability (i.e., the extent of intra- and interlaboratory reproducibility over time when performed using the standardized protocol), and its relevance (i.e., the ability of the test method to correctly predict or measure the biological effect of interest) (1) (2) (3) (4). The purpose of the PS is to communicate the basis by which new proprietary (i.e. copyrighted, trademarked, registered) or nonproprietary test methods can be determined to have sufficient accuracy (i.e., agreement between a test method result and an accepted reference value) and reliability (i.e., extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol) for a specific testing purpose. Thus, this provides an avenue to demonstrate that a newly developed test method based on similar scientific principles has comparable or better performance capabilities than those from which the existing PS were derived, and may allow a more timely use of the new test method. New test methods ("me too" tests) can be added to TG 455 after OECD review and agreement that performance standards are met. A new test method developed under this PS will be covered by TG 455 only after TG 455 has been updated to add the new test method.

2. Performance standards are based on an adequately validated test method(s) and provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar (1) (2). The three elements of performance standards are:

- Essential test method components: These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed test method that is considered to be mechanistically and functionally similar to the validated method. Essential test method components include unique characteristics of the test method, critical procedural details, and quality control measures.

- A list of reference chemicals: Reference chemicals are used to assess the accuracy and reliability of a proposed mechanistically and functionally similar test method. These chemicals are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method, and are the minimum number that should be used to evaluate the performance of a proposed mechanistically and functionally similar test method.

- Accuracy and reliability performance values: These are the standards for accuracy (i.e., sensitivity, specificity, false positive/negative rates) and reliability (i.e., degree to which the test method can be performed reproducibly within and among laboratories over time) that the proposed test method should meet or exceed when evaluated using the minimum list of reference chemicals.

3. The fully validated reference test methods that provide the basis for this PS are:

- The Stably Transfected TA assay (STTA) using the human (h) ERα-HeLa-9903 cell line (5) and
- The BG1Luc ER TA assay (6) using the BG1Luc-4E2 cell line which predominately expresses hERα with some contribution from hERβ (7) (8).

ESSENTIAL TEST METHOD COMPONENTS AND OTHER VALIDATION CONSIDERATIONS

4. Certain principles are important in delineating the essential test method components that determine whether transactivation (TA) tests are functionally and mechanistically similar. *In vitro* estrogen receptor (ER) TA assays are designed to identify substances that might interfere with ER-mediated cellular processes *in vivo*. The interaction of estrogens with cellular ER initiates a cascade of events leading to the expression of specific genes in multiple target tissues.

5. The *following test method components may vary*, so this PBTG does apply to test methods that may differ in
   - cell type (e.g. mammalian, fish, yeast)
   - cell line (tissue type)
   - characteristics of the cell line including presence of other receptors and metabolism
   - culture conditions
   - plating density
   - plate layout (including how controls are incorporated)
   - ERα characteristics (full length or partial, species of origin); if other ER proteins are present, ERα should predominate and the relative expression of each receptor should be known
   - reporter gene construct (promoter, receptor binding elements, reporter)
   - method of determining cytotoxicity.

   These elements should be clearly described in each test method.

6. Essential test method components for *in vitro* ER TA protocols should include:

   - The use of a strong reference estrogen, perferably 17ß-estradiol, to demonstrate the adequacy of the method for detecting ER agonists;

   - A weak positive control with a potency (e.g., PC_{50}, EC_{50}) two to five orders of magnitude lower than the reference estrogen should be included to provide another quality control measure by which to judge the acceptability of the method for detecting a weak agonist, and by which to evaluate the reproducibility of the test method.

   - A vehicle control (e.g., DMSO, EtOH, or H_2O) that is miscible with cell culture media at concentrations that are not cytotoxic and do not otherwise interfere with the test system.

   - A minimum limit concentration and at least seven concentrations spaced at decadic logarithmic (log_{10}) intervals should be tested up to the limit concentration.

   - In the absence of solubility or cytotoxicity restraints, the maximum concentration may be 1 mM or even up to the limit of solubility if appropriate.

- A qualitative or quantitative evaluation of cytotoxicity and how it is applied to the test method should be included in each study. Concentrations of test substances that clearly reduce viability should not be considered in the analysis of the data.

- All concentrations of the controls (e.g., vehicle, weak positive(s), or negative(s)), the reference estrogen, and the test substance should be tested in more than one replicate well.

7. No standardized statistical methods for analyzing data obtained from *in vitro* ER TA agonist assays have been developed. Each test method should establish a well-defined method for classifying a positive and a negative response. Where possible, positive results should be characterized by both the magnitude of the effect and the concentration at which the effect occurs (e.g., an EC_{50}, PC_{50}, % max, etc.).

8. To ensure that a proposed *in vitro* ER TA test method possesses characteristics similar to other validated test methods, the reference chemicals for testing ER agonists listed in Table 1 should be used to demonstrate the reliability and accuracy of the new test method. The twenty two recommended Reference Chemicals, representing chemical classes commonly associated with ER agonist activity, have been classified as ER agonists or negatives based upon published reports, including *in vitro* assays for ER binding and TA, and the *in vivo* uterotrophic assay (6) (9) (10) (11) (12) (13) (14) (15). The Reference Chemicals were tested in both the STTA and BG1Luc ER TA test methods (6) (9); the classifications (16 positive, 6 negative) were 100% concordant between the two test methods and consistent with the classifications as ER agonists or negatives, and the group of chemicals cover the potency range of known ER agonists (i.e., EC_{50} 1×10^{-12} M) to very weak (i.e., PC_{10}, EC_{50} 1×10^{-5} M) to negative for ER agonist activity. Supplementary information including the full listings of chemicals tested in both the STTA and the BG1Luc ER TAs, as well as additional chemicals tested in each test method during the respective validation studies, is provided in Annex 2 (Tables 1, 2 and 3). Additional chemicals not included in the reference chemical list may be used to demonstrate an improvement of the new test method as compared with the fully validated test methods.

Table 1. List of Reference Chemicals (22) for Evaluation of ER Agonist Accuracy[1]

Chemicals[1,2]	CASRN	Expected Reponse[1, 3]	BgILuc EC_{50} Value[3,4,5] (M)	STTA and BG1Luc ER TA Results[4,5,7]	STTA ER TA[6,7]	
					PC_{10} Value (M)	PC_{50} Value (M)
Ethyl paraben	120-47-8	POS	2.48×10^{-5}	POS	5.00×10^{-6}	-
Kaempferol	520-18-3	POS	3.99×10^{-6}	POS	1.36×10^{-7}	1.21×10^{-6}
Butylbenzyl phthalate	85-68-7	POS	1.98×10^{-6}	POS	1.14×10^{-6}	4.11×10^{-6}
p,p'- Methoxychlor	72-43-5	POS	1.92×10^{-6}	POS	1.23×10^{-6}	-
19-Nortestosterone	434-22-0	POS	1.80×10^{-6}	POS	9.64×10^{-9}	2.71×10^{-7}
Bisphenol A	80-05-7	POS	5.33×10^{-7}	POS	2.02×10^{-8}	2.94×10^{-7}
Kepone	143-50-0	POS	4.91×10^{-7}	POS	7.11×10^{-7}	7.68×10^{-6}
4-Cumylphenol	599-64-4	POS	3.20×10^{-7}	POS	1.49×10^{-7}	1.60×10^{-6}
Genistein	446-72-0	POS	2.71×10^{-7}	POS	2.24×10^{-9}	2.45×10^{-8}
Coumestrol	479-13-0	POS	1.32×10^{-7}	POS	1.23×10^{-9}	2.00×10^{-8}
4-*tert*-Octylphenol	140-66-9	POS	3.19×10^{-8}	POS	1.85×10^{-9}	7.37×10^{-8}
17α-Estradiol	57-91-0	POS	1.40×10^{-9}	POS	7.24×10^{-11}	6.44×10^{-10}
Norethynodrel	68-23-5	POS	9.39×10^{-10}	POS	1.11×10^{-10}	1.50×10^{-9}
Diethylstilbestrol	56-53-1	POS	3.34×10^{-11}	POS	$<1.00 \times 10^{-11}$	2.04×10^{-11}
meso-Hexestrol	84-16-2	POS	1.65×10^{-11}	POS	$<1.00 \times 10^{-11}$	2.75×10^{-11}
17α-Ethinyl estradiol	57-63-6	POS	7.31×10^{-12}	POS	$<1.00 \times 10^{-11}$	$<1.00 \times 10^{-11}$
Atrazine	1912-24-9	NEG	-	NEG	-	-
Corticosterone	50-22-6	NEG	-	NEG	-	-
Linuron	330-55-2	NEG	-	NEG	-	-
Spironolactone	52-01-7	NEG	-	NEG	-	-
Ketoconazole	65277-42-1	NEG	-	NEG	-	-
Reserpine	50-55-5	NEG	-	NEG	-	-

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; EC_{50} – half maximal effective concentration; NEG = negative; PC_{10} (and PC_{50}) = the concentration of a test chemical at which the response is 10% (or 50% for PC_{50}) of that induced by the positive controls (E2, 1nM); POS = positive.

[1]Chemicals, classified as ER agonists or negatives [6, 9, 10-15], were selected to represent the different chemical classes and the range of potency from strong (i.e., EC_{50} 1×10^{-12} M) to very weak (i.e., PC_{10}, EC_{50} 1×10^{-5} M) to negative for ER agonist activity.

[2]See Annex 2 (Table 1) for chemical and product classes as assigned using the U.S. National Library of Medicine's Medical Subject Headings (MeSH), an internationally recognized standardized classification scheme (available at: http://www.nlm.nih.gov/mesh), and the U.S. National Library of Medicine's Hazardous Substances Database (available at: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB).

[3]Expected responses and BG1Luc ER TA data compiled and reported in ICCVAM Test MethodEvaluation Report on the LUMI-CELL® ER (BG1Luc ER TA) Test Method An *In Vitro* Method for Identifying ER Agonists and Antagonists [6].

[4]Mean EC_{50} values were calculated with values reported by the laboratories of the BG1Luc ER TA validation study (XDS, ECVAM, and Hiyoshi).

[5]See draft proposal for new test guideline: BG1Luc Estrogen Receptor Transactivation Test Methods for identifying estrogen receptor agonist and antagonists, Table 4 for definitions of positive and negative classifications.

[6]See OECD TG 455: Stably Transfected Human Estrogen Receptor-alphaTranscriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals, Table 5 for definitions of positive negative classifications [5].

[7]PC10/PC_{50} values reported in 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 [9].

9. Metabolism of the reference chemicals in the cell system under development should be considered when assessing the results when testing the Reference Chemicals (Table 1). The degree of metabolic competence of the cell system may influence the qualitative (positive or negative) or quantitative (EC_{50}/PC_{10}) result. Metabolism of inactive chemicals to active chemicals, e.g., from DEHP (Bis (2-ethylhexyl) phthalate) into MEHP (Mono (2-ethylhexyl) phthalate), or from active chemicals to more active metabolites, e.g, metabolism of methoxychlor to HPTE (2,2-bis(*p*-hydroxyphenyl)-1,1,1-trichloroethane), or the conversion of estrone into 17β-estradiol, may result in lower EC_{50}/PC_{10} values. However, the opposite may occur as well, e.g. inactivation of estradiol by hydroxylation, or in cell lines competent in Phase 2 metabolising enzymes the test chemicals may be metabolised to inactive glucuronide or sulphate conjugates. Ideally, the metabolic capability of the cell line should be characterised. However, the metabolic capabilities of the STTA and BG1Luc cell lines have not been completely characterised, and therefore, this should also be considered. For example, the positive result with benzyl butyl phthalate (Table 1) is thought to result from hydroxylation to monosubstituted phthalates. Thus, a cell line with a hydroxylase deficiency would give a negative result. These considerations are extremely important for QSAR modeling approaches, as it may not be the compound under investigation that is actually responsible for the observed response, but rather the metabolites formed.

10. New similar test methods should not be developed solely on the basis of the twenty two Reference Chemicals, but rather on a sufficiently large test development set. Reference chemicals should be preferentially used to determine equivalence of performance compared to the validated reference test methods.

11. All chemicals should be tested in a coded/blinded manner. When evaluated using these reference chemicals, the reliability and accuracy (i.e. sensitivity, specificity, false positive rates, and false negative rates) of the proposed ER TA test method should approximate the following:

DEFINED RELIABILITY AND ACCURACY VALUES

12. For purposes of establishing the reliability and relevance of the proposed similar or modified test method when transferred between laboratories, all twenty two Reference Chemicals (Table 1) should be tested in at least three laboratories. In each laboratory, all twenty two Reference Chemicals should be tested in three independent runs performed at sufficiently spaced time points.

13. The calculation of the reliability and accuracy values of the proposed new test method should be conducted following the three criteria below, ensuring that the values for reliability and relevance are calculated in a predefined and consistent manner:

1. Only the data of runs from complete run sequences qualify for the calculation of the test method within, and between-laboratory variability and predictive capacity (accuracy).
2. Only the data obtained for chemicals that have complete run sequences in all participating laboratories qualify for the calculation of the test method between-laboratory variability.
3. The calculation of the accuracy values should be done on the basis of the individual laboratory predictions obtained for the twenty two Reference Chemicals by the different participating laboratories.

In this context, a run sequence consists of three independent runs from one laboratory for one test chemical. A complete run sequence is a run sequence from one laboratory for one test chemical where all three runs are valid. This means that any single invalid run invalidates an entire run sequence of three runs.

Within-laboratory reproducibility

14. The assessment of within-laboratory reproducibility, the concordance of classifications (positive/negative) obtained in three independent test runs with each of the twenty two Reference Chemicals within a single laboratory, should be 100%.

Between-laboratory reproducibility

15. An assessment of between-laboratory reproducibility, the concordance classifications (positive/negative) obtained in three independent test runs with all or a subset of the 22 Reference Chemicals between preferentially a minimum of three laboratories, should be in the range of 83 - 100%.

Predictive capacity (accuracy)

16. The accuracy (sensitivity, specificity, positive and negative predictivity, and overall accuracy) of the proposed new similar or modified test method should be comparable to that demonstrated for these the fully validated test methods (e.g, STTA and BG1Luc ER TAs) (6) (9). On the basis of the twenty two reference chemicals (Table 1) tested in both fully validated reference methods, as well as other empirical data from these methods (see Annex 2), the target values for sensitivity, specificity, and overall accuracy to be obtained on the basis of the twenty two Reference Chemicals are set to be greater or equal to 95%.

17. Although it is not realistic to expect test methods to perform identically, discordant results should be discussed in terms of the ability of the test method to detect a similar range of potencies and chemical/product classes as demonstrated by the fully validated test methods (6) (9).

LITERATURE

1. OECD Guidance Document No. 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. 2005, Paris:OECD Publishing.

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

3. ICCVAM. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. 2003. NIH Publication No. 03-4508. Research Triangle Park, NC:National Institute of Environmental Health Sciences

4. OECD. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test OMethods. Paris:OECD Publishing.

5. OECD, Test No. 455: The Stably Transfected Human Estrogen Receptor-alpha Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals, in OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. 2009, OECD Publishing: Paris. Available at: [http://dx.doi.org/10.1787/9789264076372-en]

6. ICCVAM, *ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (BG1Luc ER TA) Test Method An In Vitro Method for Identifying ER Agonists and Antagonists*. 2011, National Institute of Environmental Health Sciences: Research Triangle Park, NC.

7. Pujol, P., et al., Differential expression of estrogen receptor-alpha and -beta messenger RNAs as a potential marker of ovarian carcinogenesis. Cancer Res, 1998. 58(23): p. 5367-73.

8. Rogers, J.M. and M.S. Denison, 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 Vitro and Molecular Toxicology: Journal of Basic and Applied Research, 2000. 13(1): p. 67-82.

9. Takeyoshi, M., *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*. 2006, Chemicals Evaluation and Research Institute (CERI): Japan. p. 1-188.

10. ICCVAM (2003), ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. Available at: [http://iccvam.niehs.nih.gov/methods/endocrine.htm#fineval]; Addendum, 2006 [http://iccvam.niehs.nih.gov/docs/endo_docs/EDAddendFinal.pdf]

11. ICCVAM, Background Review Document: Estrogen Receptor Binding. Appendix D, Substances Tested in the ER Binding Assay, 2002. NIH Publication Report No. 03-4504. Available at: [http://iccvam.niehs.nih.gov/docs/endo_docs/final1002/erbndbrd/ERBd034504.pdf]

12. ICCVAM, Background Review Document: Estrogen Receptor Transcriptional Activation (TA)

Assay. Appendix D, Substances Tested in the ER TA Assay, 2002. NIH Publication Report No. 03-4505. Available at: [http://iccvam.niehs.nih.gov/docs/endo_docs/final1002/erta_brd/ERTA034505.pdf]

13. Kanno, J, et al., The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase 1. 2001, Environ Health Perspect. 109:785-94.

14. Kanno J, et al., The OECD program to validate the rat uterotrophic bioassay: Phase Two - Dose Response Studies. 2003, Environ. Health Persp. 111:1530-1549

15. Kanno J, et al., The OECD program to validate the rat uterotrophic bioassay: Phase Two – Coded Single Dose Studies. Environ. Health Persp. 111:1550-1558.

ANNEX 1

Definitions and Abbreviations

Acceptbility criteria: Minimum standards for the performance of experimental controls and reference standards. All acceptance criteria must be met for an experiment to be considered valid.

Accuracy: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method.

Agonist: A substance that produces a response, e.g., transcription, when it binds to a specific receptor.

BG-1: An immortalized adenocarcinoma cells that endogenously express estrogen receptor.

BG-1Luc4E2: The BG-1Luc4E2 cell line was derived from BG-1 immortalized human-derived adenocarcinoma cells that endogenously express both forms of the estrogen receptor (ERα and ERβ) and have been stably transfected with the plasmid pGudLuc7.ERE. This plasmid contains four copies of a synthetic oligonucleotide containing the estrogen response element upstream of the mouse mammary tumor viral (MMTV) promoter and the firefly luciferase gene.

Cytotoxicity: the harmful effects to cell structure or function ultimately causing cell death and can be a result of a reduction in the number of cells present in the well at the end of the exposure period or a reduction of the capacity for a measure of cellular function when compared to the concurrent vehicle control.

CV: Coefficient of variation

DMSO: Dimethyl sulfoxide

E2: 17β-estradiol

EC_{50}: The half maximal effective concentration of a test substance.

ER: Estrogen receptor

hERα: Human estrogen receptor alpha

hERß: Human estrogen receptor beta

ERE: Estrogen response element

Estrogenic activity: the capability of a chemical to mimic 17β-estradiol in its ability to bind to and activate estrogen receptors. hERα-mediated specific estrogenic activity can be detected in this PBTG.

HeLa: An immortal human cervical cell line

HeLa9903: A HeLa cell subclone into which hERα and a luciferase reporter gene have been stably transfected

hERα: Human estrogen receptor alpha

hERß: Human estrogen receptor beta

LEC: Lowest effective concentration is the lowest concentration of test substance that produces a threshold response (*i.e.* the lowest test substance concentration at which the fold induction is statistically different from the concurrent vehicle control).

Interlaboratory reproducibility: A measure of the extent to which different qualified laboratories using the same protocol and testing the same substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory reproducibility: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Me-too test: A colloquial expression for a test methods that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method.

PBTG: Performance-Based Test Guideline.

PC: Positive control (1 nM of E2)

PC_{10}: the concentration of a test chemical at which the measured activity in an agonist assay is 10% of the maximum activity induced by the PC (E2 at 1nM for the STTA assay) in each plate.

PC_{50}: the concentration of a test chemical at which the measured activity in an agonist assay is 50% of the maximum activity induced by the PC (E2 at the reference concentration specified in the test method) in each plate.

PC_{Max}: the concentration of a test chemical inducing the RPCMax

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (1) essential test method components; (2) a minimum list of reference substances selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals.

Proficiency chemicals (substances): A subset of the Reference Chemicals included in the Performance Standards that can be used by laboratories to demonstrate technical competence with a standardized test method. Selection criteria for these substances typically include that they represent the range of responses, are commercially available, and have high quality reference data available.

Proficiency: The demonstrated ability to properly conduct a test method prior to testing unknown substances.

Reference chemicals (substances): A set of chemicals to be used to demonstrate the ability of a new test method to meet the acceptability criteria demonstrated by the validated reference test method(s). These chemicals should be representative of the classes of chemicals for which the test methods is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative.

Reference estrogen (Positive control, PC): The reference estrogen, 17ß-estradiol (E2, CAS 50-28-2).

Reference standard: a reference substance used to demonstrate the adequacy of a test method. 17β-estradiol is the estrogenic reference standard for the STTA and BG1Luc ER TAs.

Reference test method: The test methods upon which this PBTG is based.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method.

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

RPCMax: maximum level of response induced by a test chemical, expressed as a percentage of the response induced by 1 nM E2 on the same plate

SD: Standard deviation.

Sensitivity: The proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method.

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of attest method.

Stable transfection: When DNA is transfected into cultured cells in such a way that it is stably integrated into the cells genome, resulting in the stable expression of transfected genes. Clones of stably transfected cells are selected by stable markers (e.g., resistance to G418).

STTA: Stably Transfected Transcriptional Activation Assay, the ERα transcriptional activation assay using the HeLA 9903 Cell Line.

Substance: Used in the context of the UN GHS (1) as chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

TA: Transactivation.

Transcription: mRNA synthesis

Transcriptional activation: The initiation of mRNA synthesis in response to a specific chemical signal, such as a binding of an estrogen to the estrogen receptor.

Validated (reference) test method: An accepted test method for which validation studies have been completed to determine the accuracy and reliability of the method for a specific proposed use.

Validation, a process based on scientifically sound principles by which the reliability and relevance of a particular test, approach, method, or process are established for a specific purpose. Reliability is defined as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardised protocol. The relevance of a test method describes the relationship between the test and the effect in the target species and whether the test method is meaningful and useful for a defined purpose, with the limitations identified. In brief, it is the extent to which the test method correctly measures or predicts the (biological) effect of interest, as appropriate (16).

VC (Vehicle control): The solvent that is used to dissolve test and control chemicals is tested solely as vehicle without dissolved chemical.

Weak positive control: A weakly active substance selected from the reference chemicals list that is included in all tests to help ensure proper functioning of the assay.

22 December 2011

ANNEX 2

Supplementary Information

for the

Stably Transfected Human Estrogen Receptor-α Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals using the hERα-HeLa-9903 cell line

And

BG1Luc Estrogen Receptor Transcriptional Activation Test Method for Identifying Estrogen Receptor Agonists and Antagonists

Table 1. Comparison of Results from STTA and BG1Luc ER TA Assays for Chemicals Tested in Both Assays and Classified as ER Agonists or Negatives.

Chemical	CASRN	STTA ER TA[1]			BG1Luc ER TA[2]		Data Source For Classification[4]			Chemical Class[5]	Product Class[6]
		ER TA Activity	PC_{10} Value (M)	PC_{50} Value[b] (M)	ER TA Activity	EC_{50} Value[b,3] (M)	Other ER TAs[c]	ER Binding	Uterotrophic		
17-ß Estradiol[a]	50-28-2	POS	$<1.00 \times 10^{-11}$	$<1.00 \times 10^{-11}$	POS	5.63×10^{-12}	POS (227/227)	POS	POS	Steroid	Pharmaceutical, Veterinary Agent
17-α Estradiol[a]	57-91-0	POS	7.24×10^{-11}	6.44×10^{-10}	POS	1.40×10^{-9}	POS(11/11)	POS	POS	Steroid	Pharmaceutical, Veterinary Agent
17-α Ethinyl estradiol[a]	57-63-6	POS	$<1.00 \times 10^{-11}$	$<1.00 \times 10^{-11}$	POS	4.20×10^{-8}	POS(22/22)	POS	POS	Steroid	Pharmaceutical, Veterinary Agent
17-β-Trenbolone	10161-33-8	POS	1.78×10^{-8}	2.73×10^{-7}	POS	7.31×10^{-12}	POS (2/2)	NT	NT	Steroid	Veterinary Agent
19-Nortestosterone[a]	434-22-0	POS	9.64×10^{-9}	2.71×10^{-7}	POS	1.80×10^{-6}	POS(4/4)	POS	POS	Steroid	Pharmaceutical, Veterinary Agent
4-Cumylphenol[a]	599-64-4	POS	1.49×10^{-7}	1.60×10^{-6}	POS	3.20×10^{-7}	POS(5/5)	POS	NT	Phenol	Chemical Intermediate
4-*tert*-Octylphenol[a]	140-66-9	POS	1.85×10^{-9}	7.37×10^{-8}	POS	3.19×10^{-8}	POS(21/24)	POS	POS	Phenol	Chemical Intermediate
Apigenin[a]	520-36-5	POS	1.31×10^{-7}	5.71×10^{-7}	POS	1.60×10^{-6}	POS(26/26)	POS	NT	Heterocyclic Compound	Dye, Natural Product, Pharmaceutical Intermediate
Atrazine[a]	1912-24-9	NEG	-	-	NEG	-	NEG (30/30)	NEG	NT	Heterocyclic Compound	Herbicide
Bisphenol A[a]	80-05-7	POS	2.02×10^{-8}	2.94×10^{-7}	POS	5.33×10^{-7}	POS(65/65)	POS	POS	Phenol	Chemical Intermediate
Bisphenol B[a]	77-40-7	POS	2.36×10^{-8}	2.11×10^{-7}	POS	1.95×10^{-7}	POS(6/6)	POS	POS	Phenol	Chemical Intermediate
Butylbenzyl phthalate[a]	85-68-7	POS	1.14×10^{-6}	4.11×10^{-6}	POS	1.98×10^{-6}	POS(12/14)	POS	NEG	Carboxylic Acid, Ester, Phthalic Acid	Plasticizer, Industrial Chemical

Chemical	CASRN	STTA ER TA[1]			BG1Luc ER TA[2]		Data Source For Classification[4]			Chemical Class[5]	Product Class[6]
		ER TA Activity	PC_{10} Value (M)	PC_{50} Value[b] (M)	ER TA Activity	EC_{50} Value[b,3] (M)	Other ER TAs[c]	ER Binding	Uterotrophic		
Corticosterone[a]	50-22-6	NEG	-	-	NEG	-	NEG(6/6)	NEG	NT	Steroid	Natural Hormone, Pharmaceutical
Coumestrol[a]	479-13-0	POS	1.23×10^{-9}	2.00×10^{-8}	POS	1.32×10^{-7}	POS(30/30)	POS	NT	Heterocyclic Compound	Natural Product
Daidzein[a]	486-66-8	POS	1.76×10^{-8}	1.51×10^{-7}	POS	7.95×10^{-7}	POS(39/39)	POS	POS	Flavonoid, Heterocyclic Compound	Natural Product
Diethylstilbestrol[a]	56-53-1	POS	$<1.00 \times 10^{-11}$	2.04×10^{-11}	POS	3.34×10^{-11}	POS(42/42)	POS	NT	Hydrocarbon (Cyclic)	Pharmaceutical, Veterinary Agent
Di-n-butyl phthalate	84-74-2	POS	4.09×10^{-6}		POS	4.09×10^{-6}	POS(6/11)	POS	NEG		Plasticizer, Chemical Intermediate
Ethyl paraben	120-47-8	POS	5.00×10^{-6}	(no PC_{50})	POS	2.48×10^{-5}	POS ( )		NT	Carboxylic Acid, Phenol	Pharmaceutical, Preservative
Estrone[a]	53-16-7	POS	3.02×10^{-11}	5.88×10^{-10}	POS	2.34×10^{-10}	POS(26/28)	POS	POS	Steroid	Pharmaceutical, Veterinary Agent
Genistein[a]	446-72-0	POS	2.24×10^{-9}	2.45×10^{-8}	POS	2.71×10^{-7}	POS(100/102)	POS	POS	Flavonoid, Heterocyclic Compound	Natural Product, Pharmaceutical
Haloperidol	52-86-8	NEG	-	-	NEG	-	NEG (2/2)	NEG	NT	Butyrophenone	Pharmaceutical
Kaempferol[a]	520-18-3	POS	1.36×10^{-7}	1.21×10^{-6}	POS	3.99×10^{-6}	POS(23/23)	POS	NT	Flavonoid, Heterocyclic Compound	Natural Product
Kepone[a]	143-50-0	POS	7.11×10^{-7}	7.68×10^{-6}	POS	4.91×10^{-7}	POS(14/18)	POS	NT	Hydrocarbon, (Halogenated)	Pesticide
Ketoconazole	65277-42-1	NEG	-	-	NEG	-	NEG (2/2)	NEG	NT		Pharmaceutical
Linuron[a]	330-55-2	NEG	-	-	NEG	-	NEG (8/8)	NEG	NT	Phenylurea	Herbicide
meso-Hexestrol[a]	84-16-2	POS	$<1.00 \times 10^{-11}$	2.75×10^{-11}	POS	1.65×10^{-11}	POS(4/4)	POS	NT	Steroid	Pharmaceutical, Veterinary Agent