36. Data interpretation criteria are shown in Table 5. Positive results will be characterised by both the magnitude of the effect and the concentration at which the effect occurs. Expressing results as a concentration at which a 50% (PC50) or 10% (PC10) of PC values are reached accomplishes both of these goals. However, a test chemical is determined to be positive, if the maximum response induced by the test chemical (RPCMax) is equal to or exceeds 10% of the response of the PC in at least two of two or two of three runs, while a test chemical is considered negative if the RPCMax fails to achieve at least 10% of the response of the positive control in two of two or two of three runs.

37. The calculations of PC10, PC50 and PCMax can be made by using a spreadsheet available with the Test Guideline on the OECD public website[2].

38. It should be sufficient to obtain PC10 or PC50 values at least twice. However, should the resulting base-line for data in the same concentration range show variability with an unacceptably high coefficient of variation (CV; %) the data may not be considered reliable and the source of the high variability should be identified. The CV of the raw data triplicates (*i.e.* luminescence intensity data) of the data points that are used for the calculation of PC10 should be less than 20%.

39. Meeting the acceptability criteria indicates the assay 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 insurance that accurate data were produced, see paragraphs 41 and 42.

40. Where more information is required in addition to the screening and prioritisation purposes of this TG for positive test compounds, particularly for PC10-PC49 chemicals, as well as chemicals suspected to over-stimulate luciferase, it can be confirmed that the observed luciferase-activity is solely an ERα-specific response, using an ERα antagonist (see Appendix 1).

TEST REPORT

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

[2] [http://www.oecd.org/env/testguidelines]

LITERATURE

1. CERI (2006), Draft validation report of TA assay using HeLa-hER-9903 to detect estrogenic activity. [Available at: http://www.oecd.org/document/62/0,3343,en_2649_34377_2348606_1_1_1_1,00.html]
2. Escande, A., Pillon, A., Servant, N., Cravedi, J.P., Larrea, F., Muhn, P., Nicolas, J.C., Cavailles, V. and Balaguer, P. (2006), Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta. *Biochem. Pharmacol.,* 71, 1459-1469.
3. Kuiper, G.G., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., van der Burg, B. and Gustafsson, J.A. (1998), Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. *Endocrinol.,* 139, 4252-4263.
4. Spaepen, M., Angulo, A.F., Marynen, P. and Cassiman, J.J. (1992), Detection of bacterial and mycoplasma contamination in cell cultures by polymerase chain reaction. *FEMS Microbiol Lett.* 78(1), 89-94.
5. Kobayashi, H., Yamamoto, K., Eguchi, M., Kubo, M., Nakagami, S., Wakisaka, S., Kaizuka, M. and Ishii H (1995), Rapid detection of mycoplasma contamination in cell cultures by enzymatic detection of polymerase chain reaction (PCR)-products. *J. Vet. Med. Sci.* 57(4), 769-71.
6. Dussurget, O. and Roulland-Dussoix D. (1994), Rapid, sensitive PCR-based detection of mycoplasmas in simulated samples of animal sera. *Appl. Environ. Microbiol.* 60(3), 953-9.
7. De Lean, A., Munson, P.J. and Rodbard D. (1978), Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am. J. Physiol.* 235, E97-El02.

Appendix 1

False positives: Assessment of non-receptor mediated luminescence signals

1. False positives might be generated by non-ER-mediated activation of the luciferase gene, or direct activation of the gene product or unrelated fluorescence. Such effects are indicated by an incomplete or unusual dose-response curve. If such effects are suspected, the effect of an ER antagonist (*e.g.* 4-hydroxytamoxifen (OHT) at non-toxic concentration) on the response should be examined. The pure antagonist ICI 128780 may not be suitable for this purpose as a sufficient concentration of ICI 128780 may decrease the VC value, and this will affect the data analysis.

2. To ensure validity of this approach, the following needs to be tested in the same plate:

- Agonistic activity of the unknown chemical with / without 10 μM of OHT
- VC (in triplicate)
- OHT (in triplicate)
- 1 nM of E2 (in triplicate) as agonist PC
- 1 nM of E2 + OHT (in triplicate)

3. ***Data interpretation criteria***

Note: All wells should be treated with the same concentration of the vehicle.

- If the agonistic activity of the unknown chemical is NOT affected by the treatment with ER antagonist, it is classified as "Negative".
- If the agonistic activity of the unknown chemical is completely inhibited, apply the decision criteria.
- If the agonistic activity at the lowest concentration is equal to, or is exceeding, PC10 response the unknown chemical is inhibited equal to or exceeding PC10 response. The difference in the responses between the non-treated and treated wells with the ER antagonist is calculated and this difference should be considered as the true response and should be used for the calculation of the appropriate parameters to enable a classification decision to be made.

4. ***Data analysis***

Check the performance standard.
Check the CV between wells treated under the same conditions.

1. Calculate the mean of the VC
2. Subtract the mean of VC from each well value **not** treated with OHT
3. Calculate the mean of OHT
4. Subtract the mean of the VC from each well value treated with OHT
5. Calculate the mean of the PC
6. Calculate the relative transcriptional activity of all other wells relative to the PC.

Appendix 2

Preparation of Serum treated with Dextran Coated Charcoal (DCC)

1. The treatment of serum with dextran-coated charcoal (DCC) is a general method for removal of estrogenic compounds from serum that is added to cell medium, in order to exclude the biased response associated with residual estrogens in serum. 500 mL of fetal bovine serum (FBS) can be treated by this procedure.

Components

2. The following materials and equipment will be required:

Materials
Activated charcoal
Dextran
Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)
Sucrose
1 M HEPES buffer solution (pH 7.4)
Ultrapure water produced from a filter system

Equipment
Autoclaved glass container (size should be adjusted as appropriate)
General Laboratory Centrifuge (that can set temperature at 4°C)

Procedure

3. The following procedure is adjusted for the use of 50 mL centrifuge tubes:

[Day-1] Prepare dextran- coated charcoal suspension with 1 L of ultrapure water containing 1.5 mM of $MgCl_2$, 0.25 M sucrose, 2.5 g of charcoal, 0.25 g dextran and 5 mM of HEPES and stir it at 4°C, overnight.

[Day-2] Dispense the suspension in 50 mL centrifuge tubes and centrifuge at 10000 rpm at 4°C for 10 minutes. Remove the supernatant and store half of the charcoal sediment at 4°C for the use on Day-3. Suspend the other half of the charcoal with FBS that has been gently thawed to avoid precipitation, and heat-inactivated at 56°C for 30 minutes, then transfer into an autoclaved glass container such as an Erlenmeyer flask. Stir this suspension gently at 4°C, overnight.

[Day-3] Dispense the suspension with FBS into centrifuge tubes for centrifugation at 10000 rpm at 4°C for 10 minutes. Collect FBS and transfer into the new charcoal sediment prepared and stored on Day-2. Suspend the charcoal sediment and stir this suspension gently in an autoclaved glass container at 4°C, overnight.

[Day-4] Dispense the suspension for centrifugation at 10000 rpm at 4°C for 10 minutes and sterilise the supernatant by filtration through 0.2 μm sterile filter. This DCC treated FBS should be stored at -20°C and can be used for up a year.

22 December 2011

Annex 3

BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists

INITIAL CONSIDERATIONS AND LIMITATIONS (See also General Introduction)

1. This assay uses the BG1Luc4E2 cell line. It has been validated by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (1). The BG1Luc cell lines endogenously predominantly express ERα and a minor amount of ERβ (2) (3) (4).

2. This assay is applicable to a wide range of substances, provided they can be dissolved in dimethyl sulfoxide (DMSO; CASRN 67-68-5), do not react with DMSO or the cell culture medium, and are not cytotoxic. The demonstrated performance of the BG1Luc ER TA agonist test method suggests that data generated with this test method may inform upon ER mediated mechanisms of action and could be considered for prioritization of substances for further testing.

3. This test method is specifically designed to detect hERα and hERß-mediated TA by measuring chemiluminescence as the endpoint. Chemiluminescence use in bioassays is widespread because luminescence has a high signal-to-background ratio (10). However, the activity of firefly luciferase in cell-based assays can be confounded by compounds that inhibit the luciferase enzyme, causing both apparent inhibition or increased luminescence due to protein stabilization (10). In addition, in some luciferase-based ER reporter gene assays, non-receptor-mediated luminescence signals have been reported at phytoestrogen concentrations higher than 1 μM due to the over-activation of the luciferase reporter gene (9) (11). While the dose-response curve indicates that true activation of the ER system occurs at lower concentrations, luciferase expression obtained at high concentrations of phytoestrogens or similar compounds suspected of producing phytoestrogen-like over-activation of the luciferase reporter gene needs to be examined carefully in stably transfected ER TA assay systems (see Annex 2)

4. The general introduction, performance results from the validation of the TA assays and the common elements for all test methods should be read before using this test method for regulatory purposes. Definitions and abbreviations used in this TG are described in Annex 1.

PRINCIPLE OF THE TEST METHOD (See also General introduction)

5. The assay is used to indicate ER ligand binding, followed by translocation of the receptor-ligand complex to the nucleus. In the nucleus, the receptor-ligand complex binds to specific DNA response elements and transactivates the reporter gene (*luc*), resulting in the production of luciferase and the subsequent emission of light, which can be quantified using a luminometer. Luciferase activity can be quickly and inexpensively evaluated with a number of commercially available kits. The BG1Luc ER TA utilizes an ER responsive human ovarian adenocarcinoma cell line, BG-1, which has been stably transfected with a firefly *luc* reporter construct under control of four estrogen response elements placed upstream of the mouse mammary tumor virus promoter (MMTV), to detect substances with *in vitro* ER agonist activity. This MMTV promoter exhibits only minor cross-reactivity with other steroid and non-steroid hormones (8). Criteria for data interpretation are

described in detail in paragraph 37. Briefly, a positive response is identified by a concentration-response curve containing at least three points with nonoverlapping error bars (mean ± SD), as well as a change in amplitude (normalized relative unit [RLU]) of at least 20% of the maximal value for the reference substance (17β-estradiol [E2; CASRN 50-28-2].

PROCEDURE

Cell Line

6. The stably transfected BG1Luc4E2 Cell line should be used for the assay. The cell line is available with a technical licensing agreement from the University of California, Davis, California, USA[3], and from Xenobiotic Detection Systems Inc., Durham, North Carolina, USA[4].

Stability of the Cell Line

7. To maintain the stability and integrity of the cell line, the cells should be grown for more than one passage from the frozen stock in cell maintenance media (see paragraph 9). Cells should not be cultured for more than 30 passages. For the BG1Luc4E2 cell line, 30 passages will be approximately three months.

Cell Culture and Plating Conditions

8. Procedures specified in the Guidance on Good Cell Culture Practice (5) (6) should be followed to assure the quality of all materials and methods in order to maintain the integrity, validity, and reproducibility of any work conducted.

9. BG1Luc4E2 Cells are maintained in RPMI 1640 medium supplemented with 0.9% Pen-Strep and 8.0% fetal bovine serum (FBS) in a dedicated tissue culture incubator at 37°C ± 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air.

10. Upon reaching ~80% confluence, BG1Luc4E2 cells are subcultured and conditioned to an estrogen-free environment for 48 hours prior to plating the cells in 96-well plates for exposure to test substances and analysis of estrogen dependent induction of luciferase activity. The estrogen-free medium (EFM) contains Dulbecco's Modification of Eagle's Medium (DMEM) without phenol red, supplemented with 4.5% charcoal/dextran-treated FBS, 1.9% L-glutamine, and 0.9% Pen-Strep. All plasticware should be free of estrogenic activity [see detailed protocol (7)].

Acceptability Criteria

11. Acceptance or rejection of a test is based on the evaluation of reference standard and control results from each experiment conducted on a 96-well plate. Each reference standard is tested in multiple concentrations and there are multiple samples of each reference and control concentration. Results are compared to quality controls (QC) for these parameters that were derived from the historical databases generated by each laboratory during the demonstration of proficiency. The historical databases are updated with reference standard and control values on a continuous basis. Changes in equipment or laboratory conditions may necessitate generation of updated historical databases.

---

[3] UC Davis, Office of Research Technology Transfer Services 1850 Research Park Drive, Suite 100 Davis, CA 95618. (530) 754-8649 e-mail address to be added

[4] Xenobiotic Detection Systems, Inc. 1601 E Geer St # S Durham, NC 27704. (919) 688-4804

Range Finder Test

12. Acceptability criteria for the range finder test are as follows:

- Induction: Plate induction should be measured by dividing the average highest E2 reference standard relative light unit (RLU) value by the average DMSO control RLU value. Five-fold induction is usually achieved, but for purpose of acceptance, induction should be greater than or equal to four-fold.
- DMSO control results: Solvent control RLU values should be within 2.5 times the standard deviation of the historical solvent control mean RLU value.
- An experiment that fails either acceptance criterion should be discarded and repeated.

Comprehensive Test

13. It should include acceptability criteria from the range finder test and the following:

- Reference standard results: The E2 reference standard concentration-response curve should be sigmoidal in shape and have at least three values within the linear portion of the concentration-response curve.
- Positive control results: Methoxychlor control RLU values should be greater than the DMSO mean plus three times the standard deviation from the DMSO mean.
- An experiment that fails any single acceptance criterion will be discarded and repeated.

Reference Standards, Positive, and Vehicle Controls

Vehicle Control

14. The vehicle that is used to dissolve the test substances should be tested as a vehicle control. The vehicle used during the validation of the BG1Luc method was 1% (v/v) dimethylsulfoxide (DMSO, CASRN 67-68-5) (see paragraph 21). If a vehicle other than DMSO is used, all reference standards, controls, and test substances should be tested in the same vehicle. .

Reference Standard (Range Finder)

15. The reference standard is E2 (CASRN 50-28-2). For range finder testing, the reference standard is comprised of a serial dilution of four concentrations of E2 (1.84×10^{-10}, 4.59×10^{-11} , 1.15×10^{-11} and 2.87×10^{-12} M), with each concentration tested in duplicate wells.

Reference Standard (Comprehensive)

16. E2 for comprehensive testing is comprised of a 1:2 serial dilution consisting of 11 concentrations (ranging from 3.67×10^{-10} to 3.59×10^{-13} M) of E2 in duplicate wells.

Weak Positive Control

17. The weak positive control is 9.06×10^{-6} M *p,p'*-methoxychlor (methoxychlor; CASRN 72-43-5) in EFM.

Fold-Induction

18. The induction of luciferase activity of the reference standard (E2) is measured by dividing the average highest E2 reference standard RLU value by the average DMSO control RLU value, and the result should be greater than four-fold.

Demonstration of Laboratory Proficiency

19. To demonstrate proficiency with the BG1Luc ER TA test method, a laboratory should compile historical databases with reference standard and control data generated from at least 10 independant experiments, conducted on different days. These experiments are the foundation for reference standards and the historical controls. Future acceptable results should be added to enlarge the database. A successful demonstration of proficiency will be achieved by producing values that are no more than 2.5 standard deviations of the historical controls (see paragraph 11).

20. Once the historical databases are compiled, the proficiency substances listed in Table 2 of the Common Elements for all test methods, should be tested.

Vehicle

21. Test Substances should be dissolved in a solvent that solubilizes the test substance and is miscible with the cell medium. Water, ethanol (95% to 100% purity) and DMSO are suitable vehicles. If DMSO is used, the level should not exceed 1% (v/v). For any vehicle, it should be demonstated that the maximum volume used is not cytotoxic and does not interfere with assay performance. Reference standards and controls are dissolved in 100% solvent and then diluted down to appropriate concentrations in EFM.

Preparation of Test Substances

22. The test substances are dissolved in 100% DMSO (or appropriate solvent), and then diluted down to appropriate concentrations in EFM. All test substances should be allowed to equilibrate to room temperature before being dissolved and diluted. Test substance solutions should be prepared fresh for each experiment.. Solutions should not have noticeable precipitate or cloudiness. Reference standard and control stocks may be prepared in bulk; however, final reference standard, contrl dilutions and test substances should be freshly prepared for each experiment abd used within 24 hours of preparation.

Solubility and Cytotoxicity: Considerations for Range Finding

23. Range finder testing consists of seven point, 1:10 serial dilutions run in duplicate. Initially, test substances are tested up to the maximum concentration of 1 mg/mL (~1 mM). Range finder experiments are used to determine the following:

- Test substance starting concentrations to be used during comprehensive testing
- Test substance dilutions (1:2 or 1:5) to be used during comprehensive testing

24. An assessment of cell viability/cytotoxicity is included in the test method protocol (7) and is incorporated into range finder and comprehensive testing. The cytotoxicity method that was used to assess cell viability during the validation of the BG1Luc ER TA (1) was a scaled qualitative visual observation method, however, a quantitative method for the determination of cytotoxicity can be used (see protocol (7). Data from test substance concentrations that cause more than 20% reduction in viability cannot be used.

Test Substance Exposure and Assay Plate Organization

25. Cells are counted and plated into 96-well tissue culture plates (2 x 10^5cells per well) in EFM and incubated for 24 hours to allow the cells to attach to the plate. The EFM is removed and replaced with test and

reference chemicals and incubated for 19-24 hours. Special considerations will need to be applied to those compounds that are highly volatile since nearby control wells may generate false positive results. In such cases, "plate sealers" may help to effectively isolate individual wells during testing, and is therefore recommended in such cases.

Range Finder Tests

26. Range finder testing uses all wells of the 96-well plate to test up to six substances as seven point 1:10 serial dilutions in duplicate (see Figure 1). It uses four concentrations of E2 (Table 1) in duplicate as the reference standard and four replicate wells for the DMSO control.

Figure 1 Range Finder Test 96-well Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12
A	TS1-1	TS1-1	TS2-1	TS2-1	TS3-1	TS3-1	TS4-1	TS4-1	TS5-1	TS5-1	TS6-1	TS6-1
B	TS1-2	TS1-2	TS2-2	TS2-2	TS3-2	TS3-2	TS4-2	TS4-2	TS5-2	TS5-2	TS6-2	TS6-2
C	TS1-3	TS1-3	TS2-3	TS2-3	TS3-3	TS3-3	TS4-3	TS4-3	TS5-3	TS5-3	TS6-3	TS6-3
D	TS1-4	TS1-4	TS2-4	TS2-4	TS3-4	TS3-4	TS4-4	TS4-4	TS5-4	TS5-4	TS6-4	TS6-4
E	TS1-5	TS1-5	TS2-5	TS2-5	TS3-5	TS3-5	TS4-5	TS4-5	TS5-5	TS5-5	TS6-5	TS6-5
F	TS1-6	TS1-6	TS2-6	TS2-6	TS3-6	TS3-6	TS4-6	TS4-6	TS5-6	TS5-6	TS6-6	TS6-6
G	TS1-7	TS1-7	TS2-7	TS2-7	TS3-7	TS3-7	TS4-7	TS4-7	TS5-7	TS5-7	TS6-7	TS6-7
H	E2-1	E2-2	E2-3	E2-4	VC	VC	VC	VC	E2-1	E2-2	E2-3	E2-4

Abbreviations: E2-1 to E2-4 = concentrations of the E2 reference standard (from high to low); TS1-1 to TS1-7 = concentrations (from high to low) of test substance 1 (TS1); TS2-1 to TS2-7 = concentrations (from high to low) of test substance 2 (TS2); TS3-1 to TS3-7 = concentrations (from high to low) of test substance 3 (TS3); TS4-1 to TS4-7 = concentrations (from high to low) of test substance 4 (TS4); TS5-1 to TS5-7 = concentrations (from high to low) of test substance 5 (TS5); TS6-1 to TS6-7 = concentrations (from high to low) of test substance 2 (TS2); VC = vehicle control (DMSO [1% v/v EFM.]).

27. The recommended final volume of media required for each well is 200 µL. Only use test plates in which the cells in all wells give a viability of 80% and above.

28. Determination of starting concentrations for comprehensive testing are described in depth in the protocol (7). Briefly, the following criteria are used:

- If there are no points on the test substance concentration curve that are greater than the mean plus three times the standard deviation of the DMSO control, comprehensive testing will be conducted using an 11-point 1:2 serial dilution starting at the maximum soluble concentration.

- If there are points on the test substance concentration curve that are greater than the mean plus three times the standard deviation of the DMSO control), the starting concentration to be used for the 11-point dilution scheme in comprehensive testing should be one log higher than the concentration giving the highest adjusted RLU value in the range finder. The 11-point dilution scheme will be based on either 1:2 or 1:5 dilutions according to the following criteria:

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 comprehensive* 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 (9.06×10^{-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

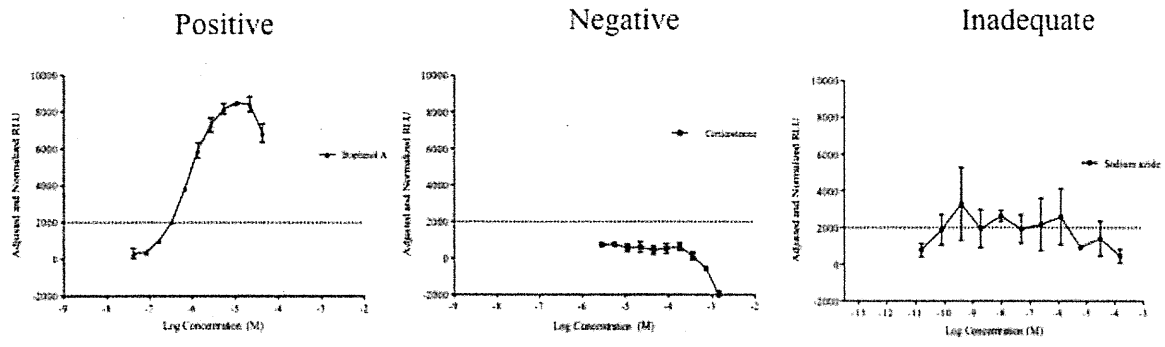
37. The BG1Luc ER TA is intended as part of a weight of evidence approach to help prioritize substances for ED testing *in vivo*. Part of this prioritization procedure will be the classification of the test substance as positive or negative for ER agonist activity. The positive and negative decision 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]	Bg1Luc 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).