iii. CERI provided raw fold induction data for the positive controls and for the chemicals assessed under the (pre) validation stage from data generated by the CERI laboratory. The provision of such data from the other laboratories was not possible. The panel required this information to assess the extent of the variation in fold induction over time.

iv. Dr. Yutaka Aoki (US EPA) provided information on proposed methods for between- and within-variation estimation to the whole group (Appendix 2) and consulted directly with CERI on how to proceed.

v. CERI conducted an internal audit of data transcribed.

vi. CERI provided raw data on edge effects from the CERI laboratory.

- Teleconference 3

vii. CERI submitted the antagonist assay protocol (SOP) and raw data for consideration by the panel. (See appendix 1).

viii. For the negative substances used, information and justification was provided by CERI on solubility and the maximum concentration used.

ix. Data analysis proposal from Dr Yutaka Aoki and subsequent discussion from and response to NICEATM consultant statistician Dr Joe Haseman, and Dr Sebastian Hoffman (ECVAM). (Appendix 3)

x. Assistance from Dr Aoki to CERI in conducting statistical estimations of between- and within-run (laboratory) variation (provisionally in June 2006).

TELECONFERENCE SUMMARIES

2. THE FIRST TELECONFERENCE WAS HELD ON 6 FEBRUARY 2006.

The meeting opened with a presentation from CERI summarising the validation of the reporter gene assay using the hER-HeLa-9903 cell line to detect estrogenic activity. During the presentation there were a number of queries, to which the following clarifications were given:

2.1 Coefficients of variation (CV) analysis to evaluate intra- and inter-laboratory reproducibility were based on log EC_{50} values, not EC_{50} values.

2.2 Requests were made for clarification as to the nature of the PC_{50} and PC_{10} values, how they are calculated, why there was no CV for the PC_{10} of 17β estradiol (E2), and whether PC_{50} and PC_{10} values were calculated within or across experiments. The PC_{50} and PC_{10} values are defined as the concentration of chemical estimated to cause 50% or 10%, respectively, of activity of the positive control response on a plate by plate basis. This measure is not the same as % maximum induction of the positive control, and is not the same as an EC_{50}. It was not always possible to calculate EC_{50} values. 100pM E2 was the single positive control for both PC_{50} and PC_{10} values. No CV could be calculated for the PC_{10} of E2 due to the fact that the lowest concentration tested was 10^{-12} M, at which concentration ERα activation was still high. CERI did not try to increase the concentration of the chemicals for which an EC_{50} could not be obtained at a dose range from 10pM to 10μM, to obtain an EC_{50} value.

2.3 When selecting substances from the ICCVAM List of Reference Substances, CERI excluded substances that had excessive cost or limited commercial availability.

2.4 During prevalidation testing, a historical database was established using three substances, E2, bisphenol A (positive) and methyl testosterone (negative) which were tested 13 times over a four month period.

2.5 The first phase of the inter-laboratory testing study used two substances, E2 and bisphenol A to determine assay transferability. In this phase, it was determined that the sensitivity of the luminometer could be a limiting factor in a laboratory's ability to duplicate the results of other laboratories. This problem was overcome by the use of a more sensitive luminescence system in some of the participating laboratories.

2.6 Of the 10 substances used in the inter-laboratory validation phase, seven were selected because they were positive in the uterotrophic assay and three were selected because they were negative. 10 chemicals were tested to keep within the cytotoxicity and solubility range for each chemical. Of the 46 substances used by CERI to examine concordance between CERI uterotrophic and ICCVAM data, 10 compounds were problematic in terms of cytotoxicity or limited solubility. Further discussion as to the nature of PC_{10} and PC_{50} values ensued.

2.7 In general practice, substances are determined to be positive by CERI based on their PC_{50} values, with a substance being considered positive if a PC_{50} could be calculated. However, for the examination of concordance between CERI and ICCVAM data, CERI considered substances to be positive if a PC_{10} could be calculated for that substance. Concern was expressed that the PC_{10} value could give rise to many false

positives. Potential metabolism issues also need to be addressed with respect to the metabolism of substances such that they do not reach the cellular target.

2.8 Participants requested that CERI send additional copies of raw data for this assay for examination. Raw data spreadsheets were sent out to the original panel of participants, prior to the meeting and to ICCVAM subsequently. Additional raw data was required to assess fold activation/induction, so as to clarify the variation of fold induction and enable comparison with fold induction data from comparable assays. Assessment of this data was favourable, and the variation observed was considered acceptable by the panel.

2.9 The following questions were asked regarding the calculation of PC_{10} and PC_{50} values:

i. Why is only one concentration (100 pM) used for the calculation of PC_{10} and PC_{50} values?.

ii. Might it be more statistically valid to use at least three concentrations (for instance, 10 pM, 100 pM and 1 nM) for this calculation? This would define a range of acceptability which could include historical and concurrent data which one could use to tease out performance criteria.

2.10 The comment from the participating statistician was that a single point would probably be sufficient for making this evaluation, but that ideally, this single concentration should be run in several additional wells, to stabilise the titrations and thus improve precision of the calculation. Also, an option worth considering is to introduce a relative index comparing a test chemical to a standard. In this approach one would calculate a ratio of (PC_{10} for the standard) to (PC_{10} for the test chemical) utilizing the data concurrently obtained for the standard. Similar to this ratio is relative binding affinity (RBA), which is already in use for receptor binding assays. Intuitively the use of relative index of this sort would result in more efficient cancellation of day-to-day (batch-to-batch) variation common to the standard and test chemical. Appendix 1 includes the statistical evaluation advice and discussion provided to the Panel by Dr Yutaka Aoki.

2.11 Additional questions were raised regarding the table from the presentation showing five to 15-fold induction of 100 pM E2 over a four-month period:

i. Are the hER-HeLa-9903 cells stable for longer than the four-month period used by CERI?

ii. Why is there so much variability in fold induction?

iii. Is there a risk of "false positives" showing up when the induction is 15 fold that would not appear when the induction was five fold?

iv. Are there upper and lower limit "cut-offs" for fold induction?

In response, CERI stated that the cells are stable for longer than the four month period, but that they do not use the cells longer than this period. The lower cut-off for induction is five fold for 100 pM E2, but there is no upper limit of induction used as a cut-off.

2.12 This led to the question of controlling for cell number. In particular, it was asked whether knowledge of cell number would allow for normalisation of induction. The conclusion was that although this could be done, it would not necessarily prove to be of any use. Luciferase reporter gene systems normally have varied degrees of response (i.e., varying fold inductions) that are not related to cell number in a linear fashion (i.e., on some days, the cells just respond better than on other days). In particular, there should not be any risk of seeing an increase in "false positives" on days where there is a higher than usual induction because what usually happens in these cases is that the response is elevated for all cells. However, it was also decided that this issue would require additional thought and consideration.

2.13 Information on the test cell line characterisation was requested. As a cervical carcinoma cell line it is possible that there may be intrinsic metabolism occurring via for example P450, other receptors such as the Progesterone receptor and the Pregnane X receptor and cellular transporters such as Pgp.

2.14 Cytotoxicity evaluation was conducted by examining baseline induction. If a substance causes luciferase activity to fall below baseline, the substance is considered to be cytotoxic. The panel were concerned that this method was open to confounding, because if a substance is an antagonist, it could suppress luciferase activity below basal levels, without killing cells. A request was made for CERI to provide more information on this issue and QA controls generally (see paragraphs 3. 7 and 4.11).

2.15 It was recommended that CERI compare their submission to the guidelines in the OECD Guidance Document 34 and *ICCVAM Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors* (NIH Pub. No. 03-4503).

3. THE SECOND TELECONFERENCE MEETING WAS HELD ON 17 MARCH 2006.

3.1 This meeting opened with a presentation from CERI summarising where the validation principles in OECD GD34 were met, partially met, or not met (Table 1) and whether the validation principles from the Minimum Standard Procedure recommended by ICCVAM were met, partially met, or not met (Appendix 4, Table 4.1). Table 2 gives the 10 core coded compounds tested in the inter- laboratory testing phase of the validation study.

Table 3.1. Checklist to assess whether the validation principles in OECD GD34 were met, partially met, or not met by the Japanese multi-laboratories validation study of a stably transfected ER alpha mediated reporter gene assay in Japan.

Principles	Met/Not met	Explanation and Justification
a) The rationale for the test method should be available.	MET	The proposed test method is used to provide mechanistic information and used for the purposes of prioritizing or grouping substances that has a potential estrogenic activity mediated estrogen receptor alpha.
b) The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.	MET	The endpoint is a luciferase activity that is produced as a result of transcriptional activation of the reporter gene. Stimulation of reporter gene expression in response to ER agonists, is thought to be mediated by direct binding where E2-liganded ER binds directly to estrogen responsive element (ERE) and interacts directly with coactivator proteins and components of the RNA polymerase II transcription initiation complex resulting in enhanced transcription.
c) A detailed protocol for the test method should be available.	MET	This is provided in the draft report appendices. Further statistical discussions on data analysis and decision criteria are provided in paragraphs 3.11 and 4.10 and appendices 2 and 3.
d) The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.	MET	Demonstrated.
e) Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used. A sufficient number of the reference chemicals should have been tested under code to exclude bias.	NOT FULLY MET	Reference chemicals are necessary to establish the relevance and reliability of the proposed test and should include a minimum number of chemicals possessing expected range of response (strong, moderate, weak and negative). There was not consensus that this requirement was met. A minority view expressed concerns that the requirements specified by the ICCVAM ED to test 78 specified chemicals were not met. This opinion was attached as appendix 1 in the summary of the third teleconference and is attached to this report as appendix 4. 10 coded chemicals (Table 3.2) possessing expected ranges of response were tested under the inter-laboratory validation, and relevance and reliability were demonstrated. However while a sufficient number of chemicals were not

		tested in all participating laboratories, according to ICCVAM recommendations, data were collected at the lead laboratory for further comparison with 46 chemicals selected from the ICCVAM list, and these data give a strong indication of relevance of the proposed test method. While the ICCVAM list of 78 chemicals does span a broad range of chemical classes and, for that reason, may be useful for identifying the limitations of the assay it also states that EC50 and IC50 data area available for 18 (23%) and 10 (13%) of these 78 recommended substances for agonism and antagonism, respectively. Qualitative data are available for 27 (35%) and 10 (13%) of these 78 recommended substances for agonism and antagonism, respectively. Thus, there is incomplete information regarding how all 78 of the recommended substances will respond in *in vitro* ER TA agonism and antagonism assays utilizing mammalian cell reporter gene systems. In which case testing only 10 of the 78 substances in multiple laboratories and the remainder in the lead laboratory is not a significant flaw in this validation effort. The limitations of the assay can be adequately determined by testing the remainder of the 78 chemicals in one or more laboratory/s. This could be considered to be consistent with ECVAM's proposed modular approach to validation (Hartung et al 2004), where core, better characterised coded sets of chemicals are tested in all participating laboratories, but further chemicals being tested for the prediction model are split or staggered between the three different laboratories. Such an approach is intended to improve the efficiency, reduce costs and speed up the validation process to meet pressing European and international regulatory requirements.
f) The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data.	MET	Relevant information obtained from the ICCVAM ED list, and results for selected chemicals were compared with this list. All data used for this comparison were produced at the lead laboratory. Additionally a data comparison was conducted with the proposed test method and the hERalpha Binding assay (and data from the immature rat uterotrophic assay) with good concordance.
g) Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.	NOT FULLY MET	The pre-validation and data collection for comparison with ICCVAM list or hERalpha binding assay were not conducted to GLP. However the inter laboratory validation was conducted to GLP. There was consensus from the panel that although GLP is ideal, for practical purposes, the fact that components of this validation and data comparison was not always to GLP was acceptable.

h) All data supporting the assessment of the validity of the test method should be available for expert review.	MET	A detailed test protocol is available, and data is available for independent review (including that prepared by this pre-peer review). Benchmark: The responses of positive control (E2) and vehicle control (DMSO) wells in each assay plate act as a benchmark such that reproducible results can be obtained when generating PC_{10} and PC_{50} values normalized by the positive control response.

Table 3.2. The 10 coded chemicals possessing expected ranges of response tested under the inter-laboratory validation.

Chemical Name	CAS	Category	Chemical Class
17b-Estradiol	50-28-2	Strong ER and AR agonist; AR antagonist	Steroid, phenolic, Estrene
17a-Estradiol	57-91-0	ER agonist	Steroid, phenolic, Estrene
Genistein	446-72-0	Weak ER agonist and antagonist	Flavonoid; Isoflavone; Phenol
Bisphenol A	80-05-7	ER agonist	Diphenylalkane; Bisphenol; Phenol
17a-Methyltestosterone	58-18-4	ER and AR agonist	Steroid, non-phenolic; Androstene
4-*tert*-Octylphenol	140-66-9	ER agonist	Alkylphenol; Phenol
p-tert-Pentylphenol	80-46-6		Alkylphenol; Phenol
Hematoxylin	517-28-2	Negative	
Di(2-ethylhexyl)phthalate	117-81-7	Negative. ER binder	Phthalate
Benzophenone	119-61-9	Negative	Benzophenone

3.2 Cell line Characterization -hER-HeLa-9903. The host cell line was checked for the following nuclear receptors, Estrogen Receptors α and β (ERα, ERβ respectively), Thyroid Receptors α and β (TRα and TRβ respectively) and the Androgen Receptor (AR). This was confirmed by a mock transfection assay with each hormone responsive reporter construct. No mycoplasma infection was detected.

3.3 It was emphasised that although there might be further applications, the assay was primarily designed to provide mechanistic information. There was also concern expressed by CERI with respect to the number of chemicals tested, and whether these were sufficient or not, and that the triplicate tests were not always repeated for the same chemicals.

3.4 Discussion began first with concerns regarding the number of chemicals tested. For statistical purposes it was recommended that in order to a get good grasp of assay reliability for each class of chemicals, a minimum of three chemicals is required for each class. Also that 'difficult' chemicals should be included

(but were not), to address for example, solubility and cytotoxicity responses. A total of 10 chemicals were tested, therefore with 2 to 3 chemicals for each *class* of chemicals for 4 classes. Dr. Akahori pointed out that 5 chemicals classified as "weak" were tested and so were 3 "negatives." While for these particular classes the number of chemicals satisfies the minimum requirement, for the remaining classes, the chemical class-specific information on reliability is somewhat limited. It is therefore suggested that the best course of action at this point is to assess the repeatability of each chemical class by obtaining the chemical class-specific estimation of between- and within-laboratory variations, and examining any statistical evidence that they differ across chemical classes. If they do not, then estimate the variations common to all applicable chemical classes, assuming the true levels of variations are comparable across chemical classes. Before doing this, the intra (within-lab) and inter (between-lab) laboratory variations require reassessment as currently they are overestimated in the draft report.

3.5 The use of the concordance, sensitivity and specificity. It was recommended that sensitivity and specificity be the primary endpoints, and that the use of concordance as a summary measure of the sensitivity and specificity should be avoided unless a caveat is included stating that here the term concordance is used to mean the weighted average of sensitivity and specificity, with weights being the prevalence of the substances being evaluated for sensitivity and specificity, and that prevalence is not a well defined concept in this example. The reason for this is that here true positives and true negatives have been chosen arbitrarily, so prevalence does not have real meaning. As such, the concordance here is a function of an arbitrary number (prevalence). Sensitivity and specificity values are far more accurate terms to use, as they are not influenced by an arbitrary level of prevalence.

3.6 Maximum concentrations that can be realistically tested in this test method. ICCVAM's expert panel recommend a maximum concentration of 1mM which is extremely high for cellular systems, and it was agreed that for practicality, one does not really need such a high response if a full dose response curve is obtained at a lower dose, or depending upon the reasons for conducting the assay or if practical reasons such as solubility/cytotoxicity preclude it. However it was agreed that for substances that tested negative in the assay (where negative is defined as no observed transcription), information and or justification should be provided on solubility and the maximum concentration used. This is because in some instances, higher concentrations have been shown to be positive in other cellular assay systems. This should therefore help explain where negative data is discordant with that published in the literature (as seen with nonylphenol for example, which is positive at higher concentrations) and identify limitations of the test, or possibly the literature. Further testing with the antagonist ICI 182 780 which is used to inhibit effects seen, would be useful to verify the ER alpha mediated mechanism.

3.7 Cytoxicity. Questions were raised about the cytotoxicity tests, and whether the control cells were the same as those used for assay purposes. It was explained that the same basal cell line had been used to develop both the ER responsive cell line and that used to evaluate cytotoxicity, and that the cytotoxicity test was not conducted at the same time as the ER test. Concern was expressed with respect to reproducibility of the cytotoxicity assay is when conducted at a different time and using a different (but related) cell line. Cytotoxicity was further discussed during the third teleconference, see paragraphs 2. 15, 4.11.).

3.8 From the summary information provided on the ERα antagonist TA assay (also see appendix 1 for the SOP), it was noted that the vehicle and positive controls were placed on the far edge of the plate. The question was therefore raised about assessment of edge effects, by dosing test plates with all vehicle controls and another with all positive controls to assess any variation for both controls. CERI informed the participants that the plate layout was different for the ERα agonist TA assay. CERI confirmed that they had assessed edge effects and it was not a concern, however this was discussed further at the final teleconference, see paragraph 5.6.

3.9 **GLP.** Although preferable, GLP and GCCP were not considered to be an issue of major concern, so long as the laboratory practice was clear and transparent. However an internal audit was requested, such that all data transcription is double checked by an additional operator (as with QA in GLP), to ensure error reduction.

3.10 Discussion with respect to the Prediction Model (PM) and the Data Interpretation Procedure (DIP). The applicability of the Prediction Model (PM) concept, that is the relevance of the assay to the applicability domain(s) of the chemical universe that the test can be applied to, was not planned for this domestic validation at the outset, and does not appear to be possible on the basis of the chemicals selected for testing in this reporter gene assay domestic validation. This may have implications for similar reporter gene assays that may be taken forward for validation assessment at the OECD level. While GD 34 is the OECD validation guidance for the panel, concern was expressed at setting a precedent that did not comply with the more stringent PM validation requirements considered by both ICCVAM and ECVAM to be an essential component of a successful formal validation exercise. Relevance with respect to the DIP, can however be established on the mechanistic knowledge of the broadly-defined "estrogenic effects", as proposed by CERI, although the panel felt that some supplemental analyses on relevance based on comparison of this assay to other "semi-gold standard" assays would be useful, particulary with reference to sensitivity and specificity.

3.11 Issues regarding log EC50 were presented for information and discussion, see appendix 2. It was agreed that there was a need to continue the discussion on the potential usefulness of $logPC_{10}$ and $logPC_{50}$ over logEC50, relative potency measures and definition of a positive chemical based on these or other measures including a traditional LOAEL. Preliminary suggestions on the best approach for the relative induction potency, is to use difference between logPC10 for estradiol and $logPC_{10}$ for a test chemical, which is similar to logIC50-based RBA for the estrogen receptor (ER). Further statistical concerns included the size of error bars and classification of positives and magnitude of response. CERI confirmed a positive to be a PC_{10} value (three fold increase above vehicle control). It was agreed that this required further exploration, to achieve consensus on an agreed statistical uniformity/consistency with particular respect to reporter gene assays (beyond this study) between countries and individual regulatory bodies, and that this might be the sole subject of a future teleconference.

4. THE THIRD TELECONFERENCE MEETING HELD ON 19 MAY 2006.

4.1 This meeting opened with update from Secretariat on the 9th EDTA and 18th WNT regarding the activities of this panel and the preparation of the preliminary validation assessment report. This was followed by an Update from CERI regarding outstanding action points from second teleconference.

4.2 For the negative substances used in the validation study, information and justification on solubility and the maximum concentration used was provided, See Table 5.1.

Table 4.1. Information and justification on solubility and the maximum concentration used for three negative compounds.

	Hematoxylin	Di(2-ethylhexyl)phthalate	Benzophenone
Preparable max. conc. in DMSO	>1 M	100 mM	>1 M
Solubility in assay medium	>1 mM	Not soluble at 100 µM	>1 mM
Control cell assay	Decrease of luminescence more than 100 µM.	-	Not cytotoxic at 1 mM
Overall conclusion of testable concentration	10 µM	10 µM	Could be tested up to 1 mM

Discussion followed with concern again expressed that for some substances classified as negative; they had not been tested at concentrations up to 1mM (solubility depending), so that very weak agonists might not be detected. A counter argument was that such doses may be unrealistic for physiological purposes, even in an extreme exposure situation, as the medium is considered to be equivalent to the *in vivo* situation. In the protocol it could be indicated that it may be possible, and in some situations desirable to test at concentrations higher than 10µM.

4.3 To what extent one needs to identify very weak agonists or antagonists was discussed further. By testing at higher doses, the EC50 can be measured, and is particularly appropriate for prioritizing for testing. However it was pointed out that the US EPA (and other regulatory authorities and agencies) would never prioritize based on just one assay, rather on the basis of a battery of tests.

4.4 It was further pointed out that one cannot control what the test might be used for, and that it would be constructive to consider more long term planning, particularly with the 3R's in mind. A robust and broad testing strategy would be ideal. While this assay falls under level 2 in the EDTA conceptual framework and US EPA ED screening program tier 1 screening, identifying substances for further testing, provision of data evaluating the ability of the test method to predict *in vivo* ED effects would be of great prospective value. It would allow better characterisation of the ability of this test method, and this might potentially lead to a reduction in animal use for ED testing.

4.5 Non-receptor mediated effects upon chemical luminescence. Concern was raised that some substances could also be inducing other non receptor mediated effects at higher concentrations that might impact upon and increase the chemical luminescence. This has been reported for some phytoestrogens (e.g. Escade et al., 2006) and has also been found to be the case in QA contract work conducted by the US EPA. Dr. Nicholas of INSERM reported that they are working on a new cell line containing two reporter genes, one responding to the hormone and a control in order to identify these non-specific effects. All three cell lines are HeLa cells lines, where one is the control and the other two are controlled by the ERα or ERβ.

Escade, et al., state... 'Moreover, at a concentration higher than 1μM, we noticed an over activation of the luciferase reporter gene by genistein, daidzein and biochanin A which was observed not only in HELN-ERα and HELN-ERβ cells but also in the parental HELN cell line....This effect, which was previously reported for genistein (Kuiper, et al., 1998), indicated that luciferase expression obtained at high concentrations of phytoestrogens needs to be examined carefully.'

4.6 Edge effects. CERI provided CERI laboratory data on edge effects (assessed by tested a single concentration of estradiol in all 96 wells) conducted after the 2nd teleconference. CERI considered that there was no edge effect affecting the final results. Data from the other participating laboratories was not available. This was discussed further as follows;

At the edge of the plate the wells may suffer from humidity effects and evaporative loss, and the conditions of incubation at the CERI lab are the same as that generally found in other laboratories. Dr Yutaka Aoki assessed the data and noted higher signals by 3.5% among the edge wells, compared with the inner wells, although it was agreed with CERI that these differences were likely to be trivial, and unlikely to affect the final result. For this reason it would be preferable to document the overall CV, and even conduct a formal analysis to see that these results do not affect the final data. It was agreed that as long as the CV for the whole plate is small, say less than 10%, in a plate with common positive control in all well on one hand, and with clear dose response in a plate with test chemical(s) and standard on the other, then the edge effects could be considered not to affect the final data for practical purposes, and that this should be clearly stated in the protocol and monitored by individual laboratories.

Further, there can be a number of plate effects one might usefully consider, for example:

- There can be effects due to cell respiration and metabolism that can be affected by the buffering capacity of the medium and cell number in each well, such that the greater the cell density required by a protocol, the more unhealthy or depleted the cells in central wells might be due to limited gas exchange, compared to those at the edge.

- Optical differences in position of the different wells of the plates can affect the luminosity readings by a plate reader (as well as observation by the naked eye).

- Stacking of plates: effects on cell metabolism have been observed in plates at the bottom of the pile of stacked plates when a large number, i.e. more than 5 plates have been stacked on top of one another in the incubator in some cells.

- Over spraying of ethanol before placing plates in the incubator.

4.7 Provision of antagonist data. CERI provided the SOP (appendix 1) and presented antagonist data on three substances tested nine times each in-house. It was noted that a concurrent positive control was not included in these experiments. The Secretariat reminded the conference call participants that the focus of the validation effort was on the agonist assay and that more antagonist testing data existed. These data are currently available on the CERI website in Japanese; CERI offered to prepare and make these data available to the panel and for independent scientific review. Although there was no assessment of interlaboratory reproducibility for the antagonist assay, CERI indicated that as the assay is almost identical to the agonist protocol, extrapolation might be possible by consideration of that validated protocol.

4.8 Concern was expressed that from a regulatory standpoint not having the antagonist data would mean that a substance that was negative for agonist activity would need to be tested in, for example, a binding

assay to demonstrate that the substance was not an antagonist. A compromise was suggested such that at a later time point the currently validated protocol could be updated and extended in a catch-up manner, with the validated antagonist protocol as and when such a protocol might be supported and made available (within a year or so). However for the present, the progression of this test should continue, as other similar assays are not so close to being validated and independently scientifically reviewed. It was generally preferred that there is no delay with moving the assay forward now. However, provision of the range of antagonist data in the report for independent peer review submission, which shows that the antagonist assay is working well, would be of great value to the reviewers.

4.9 Internal audit of data transcribed. This was done according to GLP; one error was identified in Table 13, which has now been corrected. A modified Table 13 will be attached in the final report for independent scientific review.

4.10 Statistical data analyses: Proposed methods for estimation of between- and within-run (laboratory) variation. Agreement on future plans on the revision of and addition to the analysis. Dr Yutaka Aoki (US EPA) gave a presentation with a focus particulary on a weighted average approach for assessing between- and within-run (laboratory) variation and the calculation of standard deviation (SD), with a view to refine the estimates of the various sources of variability that contribute to differences in response. Two macros were also included for the panel participants to experiment with. Appendices 2 and 3 provide information on this approach and further discussion which is presently ongoing.

4.11 Cytotoxicity queries: Provision of the criteria for when cytotoxicity is evaluated and how the data are interpreted, together with the provision of such data with respect to the reproducibility of the cytotoxicity assay when conducted at a different time and using a different (but related) cell line (see paragraph 4.7). CERI explained that generally, when the cell viability is below 80% of the solvent control, the test concentration is regarded as a cytotoxic concentration and the data at that concentration is excluded from the antagonist data analysis. CERI does not have data on the reproducibility of the cytotoxicity assay at this point.

5. DISCUSSION

5.1 Overall, the feeling from the Japanese participants for the domestic validation of this ERα reporter gene assay is that they consider that the current status of the assay is sufficient to be taken forward for official independent scientific peer review with respect to pre-screening for ERα mediated ED effects. This recommendation was therefore made to the WNT meeting in May 2006, and endorsed by the ED Task Force and WNT. With the assistance of the Secretariat, the Japanese are therefore now preparing a report for submission for independent scientific peer review.

5.2 Queries with respect to protocol optimisation, chemical selection, data analyses with sufficient statistical power for the assay, and relatively minor and non essential questions regarding inter (or between) laboratory assessment of making up the chemicals in stock solution have or are in the process of being addressed as far as reasonably possible. From a retrospective point of view, taking the validation data generated together with the extensive data set conducted by CERI in-house using this assay (which is generally in concordance with that from other published ERα mediated in vitro assays), the majority view was that this assay was robust. The minority view (Dr Tice, Dr Stokes and Prof. Combes) was attached as an appendix to the Summary of teleconference 3 and is presented in this report as appendix 4.

References

Current Status of Test Methods for Detecting Endocrine Disruptors: In Vitro Estrogen Receptor Transcriptional Activation Assays, The National Toxicology Program (NTP), Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences (NIEHS), 2002.

Earl-Gray L Jr. (1998) Tiered screening and testing strategy for xenoestrogens and antiandrogens. Toxicology Letters 102-103:677-680

Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) (1998) Final report.

URL: http://www.epa.gov/scipoly/oscpendo/history/finalrpt.htm

Escande, A., et al. (2006) Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta. Biochem. Pharmacol. May 14; 71(10):1459-69. Epub 2006 Mar 22.

Hartung et al., (2004) A modular approach to the ECVAM Principles on test validity. ATLA 32, 467-472.

ICCVAM (May 2003) Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays NIH Publication No. 03-4503). URL: http://iccvam.niehs.nih.gov/methods/endodocs/edfinrpt/edfinrpt.pdf

Kuiper, et al., (1998) Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinol 139: 4252-4263.

OECD conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals URL: http://www.oecd.org/dataoecd/17/33/23652447.doc

Organization of Economic Cooperation and Development. OECD. 2001. 3rd meeting of the validation management group for the screening and testing of endocrine disrupters (mammalian effects). ENV/JM/TG/EDTA (2001). Paris: Joint meeting of the chemicals committee and the working party on chemicals, pesticides and biotechnology .2001.

Sonneveld E, Riteco JA, Jansen HJ, Pieterse B, Brouwer A, Schoonen WG, van der Burg B. (2006) Comparison of in vitro and in vivo screening models for androgenic and estrogenic activities. Toxicol. Sci. Jan;89(1):173-87.

Takeyoshi,M., Yamasaki,K., Sawaki,M., Nakai,M., Noda,S. and Takatsuki,M. (2002) The efficacy of endocrine disruptor screening tests in detecting anti-estrogenic effects downstream of receptor-ligand interactions. Toxicol. Lett. 126, 91-98.

Yamasaki, K.; Takeyoshi, M.; Yakabe, Y.; Sawaki, M.; Imatanaka, N.; and Takatsuki, M. (2002) Comparison of Reporter Gene Assay and Immature Rat Uterotrophic Assay of Twenty-Three Chemicals. Toxicology. 170 (1-2), 21-30.

Appendix 1

Detection of anti-estrogenic activity using reporter gene assay

Description: This document provides a methodology for detecting anti-estrogenic activity of chemicals by reporter gene assay technique using hER-HeLa-9903 cell line.

Materials and methods

1. Test chemicals

Test chemicals should be dissolved in dimethylsulfoxide (DMSO) at a concentration of 10 mM.

2. Competitive substance

17β-Estradiol (E2)

3. Vehicle for chemical stock solutions

Dimethylsulfoxide (DMSO) should be used for the vehicle.

4. Test system and operating procedures

4.1 Cell lines

hERα-*HeLa*-9903 stable cell line (Sumitomo Chemicals Co.) will be used for the assay and 9903-control cell which consistently express firefly luciferase by the RSV promoter without stimulation will be used for evaluating cell-toxic effect of chemicals when anti-estrogenic like effect is observed.

4.2 Cell culture (See support protocols No.1 – No. 4)

Cells should be maintained in Eagle's Minimum Essential Medium (EMEM) without phenol red, supplemented with 10% dextran-coated-charcoal (DCC)-treated fetal bovine serum (DCC-FBS), in a CO_2 incubator (5% CO_2) at 37°C.

4.3 Preparation of chemicals

All chemicals will be dissolved in DMSO at a concentration of 10 mM, and the solutions will be serially diluted with the same solvent at a common ratio of 1:10 to prepare stock solutions with concentrations of 1 mM, 100 μM, 10 μM, 1 μM, 100 nM and 10 nM.

4.4 Preparation of cells

Assay plate will be prepared according to the support protocol No.5

4.5 Reagents for luciferase assay

Commercial luciferase assay reagent, Steady-Glo Luciferase Assay System (Promega, E2510 and its equivalents) or standard luciferase assay system (Promega, E1500 and its equivalents) will be used in this study. A bottle of Luciferase Assay Substrate is dissolved with the Luciferase Assay Buffer. Dissolved substrate should be used immediately or stored below -20ºC.

In the case of using the standard luciferase assay system, Cell Culture Lysis Reagent (Promega, E1531) should be used before adding the substrate.

4.7 Chemical exposure

Each test chemical diluted in DMSO will be added to the wells to final concentrations of 10 μM, 1 μM, 100 nM, 10 nM, 1 nM, 100 pM, and 10 pM (10^{-11}-10^{-5}M) for test in triplicate.

Exact 1.5 μl of 10 mM chemical stock and 6 working solutions will be diluted in serum-free EMEM (500 μl) containing 75 pM of E2.

Then 50 μl of the diluted test samples will be added to each well of assay plate according to the assignment table shown in Figure 1.

Reference control wells (n=6) treated with 25 pM of E2 without any other chemicals and vehicle control wells (n=6) treated with DMSO alone at concentration of 0.2% will be prepared on every assay plate. After adding the chemicals, the assay plates will be incubated in as CO_2 incubator for 20-24 h to induce the reporter gene product.

Figure 1.1 Typical assignment of assay plate for antagonist assay

	Chemical 1				Chemical 2				Chemical 3			
	1	2	3	4	5	6	7	8	9	10	11	12
A	10 μM	→	→	→	→	→	→	→	→	→	→	→
B	1 μM	→	→	→	→	→	→	→	→	→	→	→
C	100 nM	→	→	→	→	→	→	→	→	→	→	→
D	10 nM	→	→	→	→	→	→	→	→	→	→	→
E	1 nM	→	→	→	→	→	→	→	→	→	→	→
F	100 pM	→	→	→	→	→	→	→	→	→	→	→
G	10 pM	→	→	→	→	→	→	→	→	→	→	→
H	VC	→	→	→	→	→	RC					

VC: Vehicle control (DMSO only), RC: Reference control (25 pM E2 only)

In the case that the anti-estrogenic like effect or downward trends in transcriptional activity are noted, cytotoxicity of chemicals should be examined by using HeLa-9903 control cell. Cytotoxicity of chemicals will be evaluated by luciferase activity under existence of test chemicals. The assay will be performed in the same manner to the above mentioned assay procedure except using HeLa-9903 control cell. The plate format should be as shown Figure 2.

Figure 1.2 Typical assignment of assay plate for cytotoxicity

	1	2	3	4	5	6	7	8	9	10	11	12
A	10 μM	→	→	→	→	→	→	→	→	→	→	→
B	1 μM	→	→	→	→	→	→	→	→	→	→	→
C	100 nM	→	→	→	→	→	→	→	→	→	→	→
D	10 nM	→	→	→	→	→	→	→	→	→	→	→
E	1 nM	→	→	→	→	→	→	→	→	→	→	→
F	100 pM	→	→	→	→	→	→	→	→	→	→	→
G	10 pM	→	→	→	→	→	→	→	→	→	→	→
H	VC	→	→	→	→	→	→	→	→	→	→	→

VC: Vehicle control (DMSO only)

4.8 Luciferase assay (See support protocol No. 6)

Luciferase activity will be measured with the luciferase assay reagent and a luminometer according to the manufacturer's instructions.

5. Analysis of data

The luminescence signal data will be processed, and the average and standard deviation for the vehicle control wells will be calculated. The integrated value for each test well will be divided by the average integrated value of the vehicle control wells to obtain individual relative transcriptional activity. Then the average transcriptional activity will be calculated for each concentration of the test chemical. Then 50% inhibitory concentration against mean transcriptional activity induced by reference wells (25 pM of E2), will be calculated, and used for evaluating anti-estrogenic activity of chemicals.

Calculation described above will be made by the commercial software with the Hill's logistic equation showing below;

Y=Bottom + (Top-Bottom)/(1+10^((LogEC50-X)*HillSlope))

*Where, X is the logarithm of concentration. Y is the response and Y starts at Bottom and goes to Top with a sigmoid shape.

In the cytotoxicity test, the luminescence signal data will be also processed, and the average of vehicle control wells will be calculated. The integrated value for each test well will be divided by the average integrated value of the vehicle control wells to obtain individual relative transcriptional activity. When transcriptional activity are reduced less than 80% of the mean transcriptional activity of vehicle control wells, the concentration should be regarded as cytotoxic concentration and excluded for evaluation of anti-estrogenic effect.

SUPPORT PROTOCOLS

No.1 Preparation of medium

Reagents

- *Eagle's Minimal Essential medium without Neutral red (Nissui Pharmaceutical Co.)*
- *10% Sodium bicarbonate ($NaHCO_3$)*

  Dissolve 10 grams of $NaHCO_3$ to a final volume of 100 mL with water. Then the solution should be sterilized using vacuum-driven bottle-top sterilization filter unit and stored in room temperature.

- *3% Glutamine*

  Dissolve 3 grams of glutamine to a final volume of 100 mL with water. Then the solution should be sterilized using vacuum-driven bottle-top sterilization filter unit. Prepared 3% Glutamine should be stored in aliquots under -20 ℃.

- *Dextran-corted charcoal (DCC)-treated Fetal bovine serum (DCC-FBS)*

  Prepared and provided by CERI-Japan.

Preparation of EMEM*

Add following reagent into a 1-L conical glass flask and then make to 1 liter with Milli-Q water.

- 9.4 grams of pre-made powder medium
- 18 mL of 10% Sodium bicarbonate
- 12 mL of 3% Glutamine

Preparation of EMEM containing 75pM of E2

Add 75nM E2 to EMEM at proportion of 1:1000 just prior to use.

Preparation of 10%FBS-EMEM*

Add 56 mL of dextran-corted charcoal (DCC)-treated Fetal bovine serum (DCC-FBS) to 500mL EMEM.

**EMEM and 10%FBS-EMEM should be stored in a refrigerator after sterilized with vacuum-driven bottle-top sterilization filter unit.*

SUPPORT PROTOCOLS

No. 2. Reconstitute of cell from the frozen stock

1. Remove vial from Liquid Nitrogen or freezer and immediately transfer to 37°C water bath.
2. While holding the tip of the vial, gently agitate the vial.
3. When completely thawed, transfer the cell stock into 5 mL pre-warmed 10%FBS-EMEM in 15 mL conical tube.
4. Centrifuge the tube at 1100 rpm (200-300 x g) for 5min, and remove the supernatant carefully.
5. Resuspend the cell with 10 mL of 10%FBS-EMEM and place to 90 mm culture dish.
6. Incubate the cell in 5% CO_2 incubator at 37°C.