

40. It is proposed that the lack of GR (see paragraph 9) might explain some classification discordances in the comparison with the ICCVAM list, e.g. for corticosterone, potentially favouring the results of the AR-STTA assay.

***Charge question 7: All data supporting the assessment of the validity of the test method should be available for expert review***

41. The Panel agreed that this criterion had been met. However, for clarity and ease of evaluation the presentation of PC-, IC- and EC-values should be improved. Furthermore, some additions are suggested (i.e. a list of the substances and the respective primary data used to generate Table 9, Figure 9, Table 11 and Figure 11).

***Charge question 8: Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of Good Laboratory Practice (GLP)***

42. The Panel agreed that this criterion had been partly met.

43. The pre-validation was not conducted according to GLP. However, this is in general considered acceptable. In this context, it was also mentioned that the fact that the pre-validation data have been published in a peer-reviewed journal should be considered.

44. It was mentioned that elsewhere, i.e. not in the review document, the information that the pre-validation was conducted according to GLP principles or in the spirit of GLP has been provided. Therefore, it is recommended to include this information in the validation report.

## **Recommendations**

45. The Panel agrees that this report provides a summary of their views on the status of the validation of the Androgen Receptor Mediated Stably Transfected Transcriptional Activation (AR-STTA) assay to detect androgenic and anti-androgenic activities of chemicals, as detailed in the responses to the questions posed to the Panel and based on the information related to the test method validation provided to the Panel.

46. The report of the Panel, along with the provided review document on Androgen Receptor Mediated Stably Transfected Transcriptional Activation (AR-STTA) assay should form the basis for decisions on whether the validation meets the OECD principles for validation. The Panel recommends that the OECD considers the Panel report as guidance for recommending additional work required to fully meet all OECD principles.

47. Future work should focus especially on the following aspects. Most importantly, it is strongly recommended

- c) to carry out a dedicated inter-laboratory study testing substances covering a broad range of activity, especially including non-active substances as well as weak agonist and antagonists with the final test protocol (see paragraphs 25 and 29).

Furthermore, discussion of several issues, which so far have not or only partially been addressed, should be added to the validation report. These issues are

- d) potential advantages of the AR-STTA assay over similar AR activation assays (see paragraphs 8 and 31),
- e) the possibility of interference of partial agonists with antagonist effects, eventually proposing solutions to elucidate such interference (see paragraph 9),

- f) the potential impact of differences in the protocols of the pre-validation study and the inter-laboratory study (see paragraph 18),
- g) the lack of a cytotoxicity measurement in the agonist assay (see paragraph 13).

Regarding the protocol, the major recommendations were:

- h) to add acceptance and assessment criteria for the positive control DHT (see paragraph 16) and
- i) to more precisely define the decision criteria for classification, especially considering cytotoxic effects, e.g. by introducing the option of equivocal/not conclusive results, since cytotoxicity can interfere with the detection of androgenic and especially anti-androgenic responses (see paragraph 22).

In addition:

- j) the biological and statistical appropriateness of the PC10 should be explored in more detail (see paragraph 35)
- k) and a list of proficiency chemicals for both the androgenic and the anti-androgenic assay should be proposed.

Finally, the recommendations in the paragraphs 10, 11, 12, 14, 17, 19, 20, 23, 26, 27, 32, 34, 36-41 and 44 should be addressed.

## ANNEX 1

## Members of the peer review panel

<b>Panel member</b>	<b>Affiliation</b>
Miriam Jacobs	European Food Safety Agency (EFSA), Parma, Italy
Tsutomu Nishihara	Hyogo University of Health Sciences, Kobe, Japan
Michael Oelgeschläger	Federal Institute for Risk Assessment (BfR), Berlin, Germany
Vickie Wilson	United States Environmental Protection Agency (US EPA), Research Triangle Park, North Carolina, US

## ANNEX 2

## Charge questions

**PRP Charge questions:****The eight OECD principles and criteria for test method validation**

- 1. A rationale for the test method should be available, including a clear statement of scientific basis and the regulatory purpose and need for the test method.**
- 2. The relationship between the test method endpoint(s) and the biological effect and to the toxicity of interest should be described, addressing also limitations of the test methods.**
- 3. A detailed protocol for the test method should be available.**  
The protocol should be sufficiently detailed and should include, *e.g.*, a description of the materials needed, such as specific cell types or construct or animal species that could be used for the test (if applicable), a description of what is measured and how it is measured, a description of how data will be analysed, decision criteria for evaluation of data and what are the criteria for acceptable test performance.
- 4. Within- and between-laboratory reproducibility of the test method should be demonstrated.**  
Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.
- 5. Demonstration of the test method's performance should be based on testing of representative, preferably coded reference chemicals.**
- 6. The performance of test methods should have been evaluated in relation to existing relevant toxicity data as well as information from the species of concern.**
- 7. All data supporting the assessment of the validity of the test method should be available for expert review.**  
The detailed test method protocol should be readily available and in the public domain. The data supporting the validity of the test method should be organised and easily accessible to allow for independent review(s), as appropriate. The test method description should be sufficiently detailed to permit an independent laboratory to follow the procedures and generate equivalent data. Benchmarks should be available by which an independent laboratory can itself assess its proper adherence to the protocol.
- 8. Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of Good Laboratory Practice (GLP).**  
Aspects of data collection not performed according to GLP should be clearly identified and their potential impact on the validation status of the test method should be indicated.

**ANNEX 3**

**Collated initial comments from the peer review panel  
assessing the  
Androgen Receptor Mediated Stably Transfected Transcriptional Activation (AR-STTA) Assay  
to Detect Androgenic and Anti-androgenic Activities**

ISSUES	COMMENTS AND RECOMMENDATIONS
<p><b>General comments</b></p>	<p><b>Reviewer #1:</b> The AR-STTA appears to be a suitable assay for the identification of androgenic as well as anti-androgenic activities. The Chemicals Evaluation and Research Institute (CERI) has done a tremendous amount of work testing a large number of different chemicals to provide evidence that this assay generates relevant information. Unfortunately the method was modified for the inter-laboratory validation study. During the validation study cytotoxicity was determined using the Renilla activity of the AR-EcoScreen™ cell line instead of the cLuc-EcoScreen™ cell used during pre-validation. In addition, different concentration ranges were tested in the pre-validation and validation studies. Although the rationale behind these modifications is explained and understandable, the sole purpose of a validation study is to test for the reliability and relevance of the method established during a pre-validation phase. In particular, the identification of acceptable variations of the test protocol should not be part of a validation study testing coded chemicals in multiple labs. In addition, the inter-laboratory study has a limited informative value since only 5 chemicals were tested. Multi-lab studies with coded chemicals are of central importance for the validation of a test method. This part of the validation needs to be improved and might require additional testing of coded chemicals. There are also additional points that need to be clarified, in particular: The description of the Renilla construct and its use as an endpoint for cytotoxicity (including the lack of cytotoxicity measurements in the agonist assay), the criteria for positive responses and some inconsistencies in the presented data.</p> <p><b>Reviewer #2, #3, #4:</b> -</p>

VALIDATION CRITERIA	
<p><b>1. Rationale for the test Method</b></p>	<p><b>Reviewer #1:</b> The rationale for the test method is clearly stated. The scientific background and its regulatory purpose are well described. From a regulatory perspective this method is intended to be used as a screening procedure to detect androgenic as well as anti-androgenic activities of chemicals and to provide mechanistic information as a level 2 in vitro assay. As minor point: The text does not reflect the current situation and should be updated, e.g. in paragraph 18 it is stated that “Reach is expected to enter into force in 2007”.</p> <p><b>Reviewer #2:</b> The cell is appropriate, the AR construct is full length, and the renilla luciferase vector is an optimum vector to utilise for the AR. An AR ag/antagonist assay is very much needed by regulatory authorities globally; no validated equivalent is so far available. However editorially some of the paragraphs are out of date or redundant (e.g. paragraph 18: REACH entered into force some years ago now; the tests under Reprotect are ‘prevalidated’ only; or paragraph 30), and the report generally needs better editing for clarity in English. Paragraph 78 is incomplete.</p> <p><b>Reviewer #3:</b> The rationale and a clear statement of the scientific basis, and the purpose and need are clearly identified within the text of the report.</p> <p><b>Reviewer #4:</b> The technique used for the proposed assay system has been well-known and well-established in the variety fields. The regulatory purpose and needs were well documented in the validation report.</p> <p><b>The Panel agreed that this criterion has been met.</b></p>
<p><b>2. Relationship between the test method endpoint(s) and the biological effect and to the toxicity of</b></p>	<p><b>Reviewer #1:</b> Limitations of the test method are not sufficiently described. The activation of a stably transfected human androgen receptor and its subsequent nuclear localisation results in binding of the ARE and activation of luciferase expression. This process represents the best understood mechanism of androgen activity, although the transcriptional activation of endogenous genes in vivo might involve additional factors and mechanisms. Thus, an important limitation of this test (like any test that is based on transcriptional activation of an artificial promoter in an “artificial” cellular</p>

<p><b>interest addressing also limitations</b></p>	<p>environment) is that it does not necessarily reflect transcriptional responses in vivo. In addition, the assay will not detect effects mediated by AR mechanisms that are independent of direct interactions with DNA / AREs, e.g. activation of intracellular signalling pathways by membrane-associated AR. For the antagonist assay, cytotoxicity is determined using a stably transfected Cho-K1 cell line that constitutively expresses firefly luciferase (cLuc-EcoScreen) or the determination of constitutive Renilla luciferase expression of the AR-EcoScreen cell line. However, for the purpose of the test system the Renilla system might generate problems. In particular, firefly luciferase specific inhibitors have been described in the literature (see Thorne et al. 2010 and references therein) that affect the catalytic activity of the enzyme. In addition, chemicals might stimulate post-transcriptional effects, in particular, on protein stability as has been shown for the “superinduction” of luciferase luminescence by genistein (Sotoca et al., 2010). In this respect, the comparison of regulated and constitutive firefly luciferase activity in distinct cellular systems, as done in the extensive pre-validations studies, might be more appropriate to identify and exclude AR unrelated effects. AR-independent effects might have played a role in the effects observed for fluoranthene. In the antagonistic assay, fluoranthene significantly stimulated cell viability (luciferase activity in the cLuc-EcoScreen cell line) at concentrations of 10<sup>-5</sup> M (Fig. 10.2) that apparently also stimulated a positive response in the agonist assay (Fig.8). In addition the “superinduction” of luciferase luminescence by medroxyprogesterone (not tested in the antagonist assay) might reflect an ARE independent activity comparable to the direct effects of genistein on the firefly luciferase protein. A major limitation of the agonist assay is the complete lack of cytotoxicity measurements. Thus, it might be not always clear, if the chemical is a true negative, or if one cannot detect agonistic activities, because the substance is cytotoxic at concentrations below the concentration needed to elicit an agonistic response. This problem is described for hydroxyflutamide (paragraph 98) but should be discussed in more general terms, since it is not clear, if the cytotoxicity data from the antagonist assay could also be used for the interpretation of the agonist assay, if the tests are done in parallel.</p> <p><b>Reviewer #2:</b> Some areas require improvement regarding potential limitations or the comparative lack of limitations compared to similar assays with the same endpoints, i.e. provide more clarifying information on the lack of GR in the cell line. This is a potential issue regarding metabolic capacity and with GR cross talk interference with AR in other cell lines developed for AR agonism/antagonism possibly being implicated in differences in low dose/higher/biphasic dose responses of ligands that are perhaps due to different receptor mechanisms occurring in the same cell with both agonistic or antagonistic activity in the AR at one dose and agonistic or antagonistic activity in the GR at another. Should this be the case it might explain lack of literature concordance with for e.g. corticosterone, and might perhaps suggest the AR STTA assay is more robust in this respect than other similar AR assays. In any case, it would be helpful to discuss the fact that partial agonists can interfere with antagonist effects and how this can be elucidated more clearly. Of particular merit is the Renilla luciferase vector. Not only does this act as a highly effective internal cytotoxicity control, but in reviewing the literature, is probably one of the best vectors to insert into an androgenic cell line, especially PC3, with AR. Certainly,</p>
--	---

	<p>according to the literature, for the AR performing far better and more accurately than the luciferase or beta galactosidase construct used for many transient and stably transfected cell lines.</p> <p><b>Reviewer #3:</b> This assay was designed to evaluate the ability of test compounds to impact androgen-mediated gene transcription. This is a well characterized pathway of androgen action through the androgen receptor. The assay uses a cell line which has been engineered to stably express both human AR and an androgen responsive promoter/luciferase reporter construct. In theory, binding of the compound to the AR induces dimerisation of 2 ligand-bound AR which can then bind and activate the ARE promoter and stimulates production of the reporter gene, luciferase (a protein not normally found in the parent cell line). Luciferase activity should be proportional to the potency of the test compound. Conversely antagonist activity measures the ability of a compound to reduce the normal luciferase activity of a well-characterized agonist such as DHT. Some strengths of the assay include that it is an in vitro cell based assay so no animals are used for the assay, is a cost effective alternative to in vivo assays for prioritization and/or grouping compounds for further chemical testing and, unlike in vitro binding assays, has the ability to distinguish agonist from antagonist responses. Some limitations include that results are limited to the compound tested unless known metabolites are also tested and that test concentrations are limited by the solubility and cytotoxicity of the test compound.</p> <p><b>Reviewer #4:</b> There is a comparison with results of AR binding assay and that of the proposed assay. This comparison demonstrates good relationship as mechanism basis. The comparison with in vivo assay systems, such as Hershberger assay (in vivo) may not be appropriate since Hershberger assay still has a complex biological system other than AR mediated effects that can alter weights of male accessory reproductive organs.</p> <p><b>The Panel agreed that this criterion has been partly met.</b></p>
<p><b>3. Availability of a detailed test method protocol</b></p>	<p><b>Reviewer #1:</b> For most parts, the test protocol is clearly described, although the description of the two different protocols that have been used during pre-validation and the multi-lab study makes it somewhat difficult to follow. There are some points that need to be addressed in more detail, i.e. the description of the Renilla expression construct and the criteria for positive responses.</p> <p>The AR-EcoScreen cell line needs to be described in more detail, in particular the construct that was stably transfected to allow constitutive expression of the Renilla luciferase. It is somewhat confusing that the cLuc-EcoScreen was used for the cytotoxicity measurements during pre-validation, since constitutive Renilla expressing vectors are commonly used to standardize for transfection efficiency or unspecific effects on transcriptional or post-transcriptional events. Since</p>



	<p>obviously the same cell line was used in all studies the reasons why Renilla was not used from the very beginning should be explained / discussed.</p> <p>Different doses are tested in the pre-validation and inter-lab studies. The rationale behind these changes should be explained in detail.</p> <p>As mentioned before the different possibilities of unspecific modulation of luciferase activities need to be addressed. In addition, one should consider the possibility that mild cytotoxic effects might not be detected in the Renilla or the Luciferase cytotoxicity control. In general, viral promoters like the SV40 promoter that is also used to express the human AR protein in the AR-EcoScreen cell line, are very robust and might not be as sensitive as the ARE-reporter. Sodium azide and 17<math>\alpha</math>-ethinylestradiol were dissolved in water or ethanol. It is stated in paragraph 51 that 0.1% ethanol did not affect the cells and, thus, ethanol should be included in the SOP.</p> <p>A common list of proficiency chemicals for the androgenic and anti-androgenic assay might be useful, since both assays would have to be performed in parallel anyway.</p> <p>The presentation of the data and the data analyses is somewhat confusing. PC10, PC50, EC50, lin.IC30 or lin.IC50 values are provided and discussed. It is also not always clear if data is based on single or several studies and if the agonist and antagonist studies were performed in parallel or independent of each other.</p> <p>In general, the data interpretation procedure should be stated more clearly and should be consistent throughout the report. For the agonist assay, the sensitivity of method is only acceptable, if PC10 values are used (Table 9). Although it seems from table 9 that this did not increase the false positive rate, some of the data provides evidence that the use of PC10 values may actually create problems.</p> <p>Actinomycin was recommended by ICCVAM as a control for substances that indirectly interfere with the reporter gene activity since it's the only RNA synthesis inhibitor in the ICCVAM list. Following Table 8, Actinomycin was considered positive in the agonist assay, because it induced a very slight increase in luciferase activity at concentrations that might be cytotoxic.</p> <p>As mentioned before, Fluoranthene stimulate a positive response in the agonist assay at concentrations that also stimulated Renilla activity in the antagonist assay (10<sup>-5</sup>M) and therefore it cannot be excluded that this effect might be due to ARE-independent activity.</p> <p>In addition, to classify substances as negative in the agonist assay because they never induce the luciferase activity strongly enough although EC50 values can be calculated might be problematic, since low inductive activities at low concentrations could still be biologically significant. This should be discussed in more detail.</p> <p>For the antagonist assay lin.IC30 as used in the validation study. The reason for using IC30 is not clear. In particular Table 11 shows no differences between lin.IC50 and lin.IC30. In general, the false negative rate is quite high in both cases (almost 17%).</p> <p>Some information about the use of the analysis software should be included. Can additional software be used besides</p>
--	--

GraphPad? Are there specific adjustments of the software preferences necessary?

In general the decision criteria for positive or negative results should be more clearly defined and more attention should be paid to cytotoxic effects. For example, ketoconazole is considered positive in the antagonist assay and an IC<sub>50</sub> of 6,98x10<sup>-6</sup>M is provided in table 12. However, at 10<sup>-6</sup> M no effect on luciferase activity was observed in the antagonist assay and at 10<sup>-4</sup> M ketoconazole is obviously highly cytotoxic (Fig. 10-2). The IC<sub>50</sub> value seems to be based on a single data point (10<sup>-5</sup> M) that revealed a significant change in luciferase activity with a cell viability >80% and, thus, is not really convincing.

Since cytotoxicity can interfere with the detection of androgenic or anti-androgenic responses it might be more appropriate to regard data that is likely to be affected by cytotoxicity (like ketoconazole, but also DDT and methoxychlor) as equivocal / not conclusive or outside of the applicability domain of these assays.

**Reviewer #2:** The following clarifying improvements need to be made with regard to the approximation of the IC<sub>50</sub>, EC<sub>50</sub> and linear and non linear IC's 30 and 50, especially when the dose response curve is not complete. It cannot be considered a robust measure, only a less robust approximation. Also with respect of performance criteria and acceptable test performance that will need to be developed for the draft TG: for the agonist assay, there may well be problems with R1881 as it is an anabolic steroid with potentially highly limiting import restrictions and therefore import restriction for laboratory use also in some jurisdictions. This is probably because it is essentially the same as trenbolone, but as it has undergone 17alpha alkylation it can remain active after oral administration. It is also anticatabolic. Would MDHT be an appropriate alternative?

**Reviewer #3:** Most of the protocols are very clear and adequate; however, there are some minor edits/additions that I suggest be made for ease of use. A graph of the expected DHT dose response would be helpful. How was concentration of DHT used in antagonist assays determined? It is stated in the report that it was about a PC80-100 but that is not stated in the protocol. My reservation here is that the report implies that sometimes 0.5 nM may be PC80 and at others a PC100. If this varies from plate to plate, then the PC10 and PC50 of test compounds may vary as much as 20% also. Inclusion of criteria for the DHT positive control is suggested. Each lab should run DHT dose response to confirm that they are getting correct response and to verify concentration of DHT needed for antagonist assay.

Appendix 1 – section 1.8.2 and 1.8.3: The text says 175 cm<sup>2</sup> flask, but I believe it should be 75 cm<sup>2</sup>. This appears to be a typographical error as 75 cm<sup>2</sup> is correctly used through other parts of the protocol.

Appendix 3, Section 4.5: Need to state the volume of lysis buffer to be added to each well as this would affect assay performance. I could not find this information in the SOP or any of the supporting protocols.

Support protocol (SP) 1 – How long should cells be cultured before they can be used in an assay after reconstituted from frozen? It is my experience that usually 1 or 2 weeks of standard culture are required to allow full cell recovery and best

	<p>performance in similar assays. What is required for this assay?                  Support protocol 2: Information on expected doubling time would be helpful. Clarify when or at what density, e.g. only at confluence, should cells be sub-cultured.                  Support protocol 4: Similar to mentioned above (for SP1), how long should cells be cultured before being used in an assay? Should only cells from confluent flasks be used in the assay? How often can cells be trypsinized without affecting assay performance? Perhaps cells for an assay should plan to be plated at the same time as propagation?                  Appendix 4 – (page 95) What is the purpose of adding Zeocin and Hygromycin? I am assuming they are selection antibiotics but I did not see it specifically stated anywhere.                  Page 98 – mid-page: What is meant by 56 nM DHT 0.1%DMSO solution? Should this be 56 nM DHT in DMSO? It is assumed so as then the 2 subsequent dilutions would result in 0.5 nM DHT in assay.                  Page 99 section 5.2 The referenced figure should be Fig 3 (not Fig 2).</p> <p><b>Reviewer #4:</b> Only the decision criteria for evaluation of data are not clearly stated. Parameters used in the proposed assay system is not commonly used but considered to be effective for the purpose of the assay (prioritizing or screening). However, its calculation may not be easy for laboratories where newly introduce the assay. Therefore, it would be helpful to provide the validated calculation program (e.g. spreadsheet) as in ERTA assay (TG455).</p> <p><b>The Panel agreed that this criterion has been partly met.</b></p>
<p><b>4. Demonstration of within- and between-laboratory reproducibility</b></p>	<p><b>Reviewer #1:</b> Intra- and inter-laboratory reproducibility has been demonstrated. For the agonist assay there was some variation in the Hill slope, but the calculated PC10, EC50 for the standard chemical (DHT) and a comparable strong agonist (R1881) were highly reproducible with very low CV values. The other three chemicals were identified correctly as negative. There were single experiments that resulted in PC10 values for Hydroxyflutamide and Diethylhexylphthalate and were not reproducible and therefore discarded but also demonstrate a high sensitivity of the system. The use of Hydroxyflutamide in the inter-laboratory study might have not been the best choice because of the problems with this substance in the agonist assay as described in paragraph 98, although in figure 10-2 no significant cytotoxicity can be observed up to concentration of 10<sup>-5</sup> M (not 10<sup>-7</sup> M as stated in paragraph 98). For the antagonist assay the CV values are somewhat higher but potential positives and negatives were correctly identified and the calculated IC50 and IC30 values reproducible in each lab and between labs. Only the one laboratory consistently detected anti-AR activity at lower concentrations which might be due to the lower Spike-in/PC DHT ratio. However, the reason for the lack of cytotoxicity data for the R1881 testing in all participating labs should be explained.</p>

**Reviewer #2:** Only for 5 substances for between laboratory reproducibility. Although this is generally acceptable for in vivo validations, for in vitro validations more substances are usually expected. Therefore, further demonstration of reproducibility in the naïve labs would be preferred.

It would be particularly helpful to have information for a replacement for R1881, and a weak agonist and antagonist for which the dose response curve is not complete at 10uM so, up to cytotoxicity/limit of solubility testing (consideration of the solubility properties of the chemical)/ 100uM, plus another negative for both agonist and antagonist. These five examples of additional test substances would be very helpful for the development of the performance criteria for the draft TG.

Testing at a max. concentration of 10uM may be too low to detect weakly active compounds, thus testing to the limit of solubility, the onset of cytotoxicity or higher doses may be required for pre-screening and subsequent screening if the pre-screening indicates that this is more correct and necessary.

**Reviewer #3:** Generally CVs were low and reproducibility was quite good. However, a limited range of affinities of compounds was tested between labs. In the agonist assay, CVs for DHT run by three labs on three different days ranged from 0 to 8.0%. Reported overall CVs for each lab with DHT were 3.6 to 4.5% which is very good. CVs for coded compounds were also low. My reservation for the agonist assays is the fact that only strong agonists and negatives were tested in the interlaboratory study. No weak agonists were tested. Historically, variability often increases when weaker agonists are tested in vitro as the weak agonists are often a greater challenge. Variability in the pre-validation in one lab was acceptable for weaker agonists, which is sufficient as a larger number were tested in the assay, but no data is available for weak agonists from the inter-lab study to assess inter-lab reproducibility. In antagonist assays, only 2 antagonist were assessed, HF and BPA. Reproducibility for these 2 coded compounds was demonstrated and variability was acceptably low. Again, however, these were 2 of the more potent antagonists based on the pre-validation study results. My preference would have been to have at least one or two additional very weak antagonists evaluated in the inter-lab study. This is an important point as most environmental compounds identified to date that impact the AR are antagonists and many of them are weak antagonists in vitro.

**Reviewer #4:** Provided in the validation report.

**The Panel agreed that this criterion has been partly met.**

<p><b>5. Demonstration of the test method's performance based on testing of representative reference chemicals</b></p>	<p><b>Reviewer #1:</b> In the multi-lab study five coded chemicals were tested in four labs. The results correctly identified the expected positives and negatives in the agonist and antagonist assays. However, for the agonist assay the positives were DHT itself and R1881 that have comparable / identical activities to the control DHT and are structurally very similar / identical. Thus one can hardly conclude from this study that the assay will detect different kinds of chemicals that activate the AR. In addition these two chemicals have obviously very low PC50/PC10 values and it would have been more informative if chemicals with lower activities (higher PC10) values would have been tested. Testing these two substances only recapitulates the reproducibility of the agonist assay with the positive control substance as shown in table 15. For the antagonist assay it looks more convincing since two structurally distinct chemicals with anti-AR activity were identified correctly with similar but distinct IC50 / IC30 values. As pointed out in GD 34 "For the assessment of inter-laboratory reproducibility a subset of test substances used to assess accuracy might be appropriate, provided that the subset adequately represents an appropriate range of responses and physical / chemical properties for which the test method is proposed to be appropriate". In this respect, the data provided in this report do not allow a clear judgement of the relevance and reliability of the test method.</p> <p><b>Reviewer #2:</b> <i>see respective answer to charge question 4</i></p> <p><b>Reviewer #3:</b> For the data shown the test method performed well in the pre-validation and inter-lab studies. In the inter-lab validation, compounds evaluated were tested using coded compounds to avoid unintentional bias. In the agonist assays 5 compounds were evaluated and 2 in the antagonist assay. The problem is that only relatively strong agonists and antagonists were evaluated. No weak agonists and no relatively weak antagonists were represented. Since one might expect, based on previous reports, that environmental compounds would have relatively weak activity, this is a limitation in the demonstration of assay performance. While a large number of compounds were not evaluated in the inter-laboratory study, I believe that testing a large number of compounds would be less valuable than testing a small number of compounds over a broader range of affinities.</p> <p><b>Reviewer #4:</b> The performance was well-demonstrated by testing 45 chemicals including coded chemicals. Also, the test chemicals were basically selected from the well-known source (ICCVAM). The number of chemicals tested is not enough but in the range that can be tested, and the selection is reasonable, under the limited money and time.</p> <p><b>The Panel agreed that this criterion has been partly met.</b></p>
--	--

<p><b>6. Test methods evaluation related to existing relevant toxicity data</b></p>	<p><b>Reviewer #1:</b> An obvious problem is the lack of a gold standard to compare the results with and to evaluate the relevance of the assay. It might be worth mentioning that the biological relevance of reference data might be for some substances not necessarily clear cut. In any case, the comparison of the data with the ICCVAM list (summarizing published results from other assays) and the additional data obtained from human AR-binding studies is an acceptable and in particular feasible way to evaluate the assay. The AR-Binding data is not discussed sufficiently, in particular the absence of significant binding of chemicals that elicit a response in the AR EcoScreen assay. It is argued in paragraph that this might be due to lower concentrations tested in the binding assay. This argument is not convincing because for the inter-lab study the maximum concentration tested was also 0,1M and, as a consequence, one would have to go back to the testing concentrations of the pre-validation studies (up to the available solubility or 1 M). Here it would be worthwhile to evaluate the relevance of positive responses in the AR EcoScreen assay that are either not detectable in the in vitro binding assay for technical reasons (comparison with published binding data) or mediated by other mechanisms than direct AR-binding. For example, for vinclozolin it has been shown that the substance itself, in contrast to its metabolites that are generated by hydrolysis, does hardly bind the AR receptor and weak in vitro ligand displacement activity for nonylphenol had also already been reported. However, there are also inconsistencies in the presented data. In Paragraph 116 and Table 14 (AR binding data) it is stated that 4-Hydroxytamoxifen were not detected as an AR agonist or AR antagonist despite significant binding to the AR receptor. However in table 8, it is listed as an antagonist in the AR EcoScreen, in accordance with published data.</p> <p><b>Reviewer #2:</b> It is important to note that the ICCVAM lists were published in 2003, so are 8 years old, and I understand for the AR are due to be updated and corrected. In addition to comparison with AR binding assay data given in the report, comparison with the Hershberger TG would be helpful, and discussion of problems with comparison with the Hershberger, especially with the possible selection of additional test substances for the development of performance criteria.</p> <p><b>Reviewer #3:</b> There is lack of a good “standard” for AR-TA assay data which makes this evaluation challenging. Comparisons were made to the ICCVAM list and to AR competitive binding data. These may or may not be the best data for such comparison. For agonist assays, concordance, sensitivity and specificity comparisons were most favourable when PC10 values from the AR-Eco-screen assay were compared to the ICCVAM list. There are two issues here: the use of PC10 values as criteria for a positive is questionable and questions exist with respect to the proper classification of some compounds on the ICCVAM list as positive or negative. With respect to the first issue, using Fig 8B as an example, actinomycin D, a prototypical inhibitor of gene transcription which I would expect to be a negative, was classified as a positive based on PC10 response. Further, there is no dose response – just a relatively flat line at or near the PC10 value.</p>
---	---

	<p>This compound should inhibit transcription of luciferase and be a negative. Hence, a compound that produced a similar curve to actinomycin-D should likely be questioned. Mifepristone produces a similar lack of dose response. Hence the biological relevance of a PC10 is questioned, especially without additional increase in response as concentrations are increased. Correlation to PC50 values would be preferred but due to solubility and cytotoxicity issues a PC50 values could not always be attained in the AR-Eco-screen, thus the correlation to PC50 was reduced as compared to a PC10 value. I also question whether the PC10 value is a statistically significant response. With respect to the second issue, classification of compounds on the ICCVAM list were determined from the literature and were compiled for data from many types of assays (including cell proliferation assays and yeast-based assays) not just AR-TA assays. Classification is currently based primarily on a weight-of-evidence type approach (number of papers reporting a positive) with little evaluation of the thoroughness of the assessment, quality of the data or methods/performance of the assay. In addition, with respect to antagonists, the ICCVAM list combined data which included the use of various differing “spike-in” compounds and their use at differing concentrations. Logic suggests that these factors would impact the results obtained in an assay and hence the classification of positive or negative. It is suspected that lower correlation of PC50 values may be related to these issues among others. That said, based on experience of the interaction compounds with the AR, the following additional comments should be considered:</p> <p>The approach taken to calculate EC50 and IC50 was the best approach (i.e. top and bottom of the test compound curve used without regard to the whether a full dose response was attained or not) is questioned. Calculation based on comparison to the dose response of the positive control (i.e. DHT) is suggested as percent of control. For example PC50 would be preferred.</p> <p>Poor correlation to the ICCVAM list for antagonists might be expected and may not be the best comparison until the ICCVAM list is updated.</p> <p>AR-Eco-Screen (based on data in Table 14) correlated well with AR binding data - Compared to the AR binding data the relevance of the assay was well demonstrated.</p> <p><b>Reviewer #4:</b> This was done by comparing “human” AR binding assay data showing good performance. (see respective comment to charge question 2) There are no other suitable data sources to compare with.</p> <p><b>The Panel agreed that this criterion has been partly met.</b></p>
--	---

<p><b>7. Availability of all relevant data for expert review</b></p>	<p><b>Reviewer #1:</b> For most parts the report did allow review, although the lack of primary data in some parts and the inconsistent use of different calculated values i.p. PC50, PC10, IC50, IC30 (with and without taking cytotoxicity data into account) caused difficulties to evaluate the data. In particular, a list of the substances and the respective primary data used to generate Table 9, Figure 9, Table 11 and Figure 11 would have been helpful.</p> <p><b>Reviewer #2:</b> See the respective responses to charge questions 3 and 4. In addition, the discussion on PC50, PC10, EC, IC and lin.IC values needs further refinement and improved accuracy.</p> <p><b>Reviewer #3:</b> A detailed test protocol was included.. I suggested a few minor edits above (see question 3) but overall the protocols were clear and easy to follow. Given the success of the multi-lab validation exercise, the protocol was sufficiently detailed to allow other labs to generate equivalent data. Benchmark data was provided by the use of a positive control (DHT) by which other labs could evaluate their performance. I might also suggest a small set of additional proficiency compounds over a range of activities.</p> <p><b>Reviewer #4:</b> Data: Available; Protocols: The detailed protocols used during the multi-lab validation are available so that an independent lab can perform the assay. Also the performance criteria are provided.</p> <p><b>The Panel agreed that this criterion has been met.</b></p>
<p><b>8. GLP (ideally)</b></p>	<p><b>Reviewer #1:</b> The inter-laboratory study was conducted under GLP conditions whereas the pre-validations studies (comprising most of the data) were not. In general that might be acceptable. However in this case only very few chemicals were tested in the inter-laboratory study, different test protocols had had been used for the pre-validation and inter-laboratory studies Since the validity of the test method cannot be concluded from the inter-laboratory study that was performed in accordance to GLP this requirement is not met. <i>(Comment by the Panel chair: The relevance of the reply to this question was discussed with the reviewer. As a result, Reviewer #1 considered this criterion as partly met.)</i></p> <p><b>Reviewer #2:</b> The pre-validation was in the spirit of GLP, and the actual validation was to GLP.</p> <p><b>Reviewer #3:</b> It was clearly identified that data collected in the pre-validation assays were not collected under GLP. While</p>



	<p>GLP is preferable in all cases, I see no reason to exclude the pre-validation data based solely on this criteria. The Inter-laboratory validation was conducted under GLP and I suggest that this is sufficient.</p> <p><b>Reviewer #4:</b> The data collected during the pre-validation phase were not obtained under GLP. However, it is considered to have no impact on the validation status as long as the all raw data of the assay are available. In addition, the results in pre-validation study had been already published in the peer-reviewed scientific journal.</p> <p><b>The Panel agreed that this criterion has been partly met.</b></p>
--	--

Remark by the panel chair: Note that the initial replies of the reviewers as provided by February 11, 2011, have been collated here (see paragraph 3). Furthermore, the numbering of the reviewers does not correspond to the alphabetical order of Annex 1.

## ANNEX 4a

January 24, 2011

## Response to the inquiry dated Jan. 17, 2011

Masahiro Takeyoshi, Ph.D. , CERJ-Japan

**1. Fluoranthene stimulated Renilla activity in the antagonist assay at concentrations that also seem to stimulate a positive response in the agonist assay (10-5M). At least that is how it looks like comparing figure 8 and figure 10-2. However in table 10, the EC50 value is 1,29x10-8M.**

So far, the cause for the dose-dependent induction of Renilla luciferase is unknown. However this phenomenon would not be relevant with androgenic activity.

Fluoranthene showed clear antagonistic response, it means this chemical interact with androgen receptor. So we think Fluoranthene may be partial agonist.

We have rechecked the values reported in the validation report, article published in Toxicology in vitro and the raw data related. Consequently, there is a mistake in the IC50 values reported in the article, IC50 values for Mifepristone and Fluoranthene were exchanged by mistake. And also we have proofreading error.

We cited the IC50 values from the article to the validation report, and resulted in the mistakes in the validation report.

Correct IC50 values are provided in the following table.

	IC50 reported	Correct IC50 value
Mifepristone	3.23x10-5	1.29x10-8
Fluoranthene	1.29x10-8	3.23x10-5

As a consequence of reinspection, there was no other mistake of IC50 values in the validation report.

**2. DTT and Methoxychlor have been classified as negatives in the antagonist assay (Table 8), but IC50 values are provided in Table 12**

These chemical showed cytotoxicity (below 80% viability) at 10-6M. So these chemicals were decided as negative in the antagonist assay.

IC50 values can be calculated if the cytotoxic effects were observed. So we provided the IC50 values for these two chemicals in table 12, But cytotoxicity of these chemicals were mentioned in Paragraph 111.

Where we have another mistake in Para 111; the cytotoxic dose of DTT and Methoxychlor are shown as “10-9 or 10-8 M”, but correct values are “10-6 or 10-5 M”.

As a consequence of reinspection about this issue, there was no other mistake in the validation report.

**3. Corticosterone was classified as a positive in the agonist assay and in paragraph 103 it is argued that this might be due to the high concentrations tested. However, in table 10 the EC50 value for Corticosterone is  $4,5 \times 10^{-7}$  M.**

EC50 values can be calculated automatically by the calculation softwares, such as GraphPad Prism, if the chemical has not high induction level. Indeed, EC50 can be calculated for corticosterone, but the chemical failed to yield PC50 value, so the chemical was decided as negative.

We have been advocated that EC50 is not appropriate to express the potency of the hormonal effect. Because EC50 is defined as 50% point between top and bottom values, even if the top value is only 3% level of the maximum induction level of positive control substance. This is the critical issue to evaluate the hormonal potency of chemicals by EC50 values. So we emphasize that PC50 value is suitable for evaluating hormonal potency because the actual induction level compared with maximum induction level of the positive control substance is considered in the PC50.

## ANNEX 4b

**Questions (send by e-mail on January 28, 2011):**

Two further issues in relation to the third point were brought up (see Annex 4a.3.):

1. The answer refers to the PC50. But should it not be the PC10?
2. How are the EC50 calculated? Do the 50% relate to the response of the positive control or to the maximum response achieved with the chemical?

**Answers (received by e-mail on February 03, 2011):**

To 1.:

Yes! In the agonist assay, positive response is decided if the chemical has PC10 response. Corticosterone has only PC10, and not has PC50. It means the chemical has low top response. Paragraph 103 refers only to the reason for getting positive response of this chemical.

To 2.:

EC50 is calculated with logistic equation by using Graphpad PRISM etc. This parameter is defined as 50% point between Top and Bottom responses. Please see the attached figure (below); Top values are varied from chemical to chemical. Corticosterone is corresponding the response curve (B). In this case, having low top response, EC50 value would be apparently lower value than that obtained regular response curve (A) having high top value. Meanwhile, PC50 is defined as 50% point of maximum response induced by positive control substance.

