Table 1 Response to comments on the ESTROGEN RECEPTOR BINDING ASSAY USING RAT UTERINE CYTOSOL

Comment Number	Reviewer	Comment	Response
1. Clarity of purpose of the assay			
1.1	PB	Yes, the stated purpose of the assay is very clear. The purpose is to provide a test that enable to identify the potential of chemicals to interact with rat uterine estrogen receptors.	Agree. No response needed.
1.2	IP	In my opinion the test regarding the purpose of the assay should be modfied. The Estrogen Receptor Binding Assay is in itself only able to to identify compounds that bind to the estrogen receptors ERα and/or ERβ. However, compounds can interact with with the estrogen receptor system through alternative mechanisms that may not involve direct binding to the estrogen receptors.	EPA will not be relying on this assay alone to determine the potential of a substance to interact with the entire estrogen-related system. The assay will be used only to provide information on the ability of the substance to interact with the estrogen receptor. Other *in vitro* (estrogen receptor [ER] transcriptional activation) and *in vivo* (uterotrophic and female pubertal) assays in the Tier 1 Battery will provide information on ER function.
1.3	SS	The rat uterine cytosol estrogen receptor (RUC-ER) competitive binding assay is one in a battery of assays aimed at providing validated strategy for the screening of endocrine disruptors (e.g., xenobiotics and environmental chemicals) that interact with the ER in target tissues and alter their normal functions. The purpose of the RUC-ER binding assay in this extensive project is clearly stated. The background information on the ER binding assay and the various options and choices of ER preparations from different animal species	Agree. No response needed.

		and recombinant human ER are well-presented. The strengths and shortcomings of the RUC-ER binding assay in view of numerous previous studies were taken into account in selecting and including this assay in the collection of other cell-based- and *in vivo* assays.	
1.4	MLS	Yes. It is clearly stated that the estrogen receptor rat uterus cytosol (ER-RUC) binding assay is one of the tests of a battery of complementary screens, included in the endocrine disruptor screening program Tier-1 battery. It is also clearly stated that the aim of the ER-RUC is to detect an interaction with the estrogen receptor, not to identify the mechanism of action (stated i.e. p.8, 65, 69-70 of the ISR, and page 5 of Appendix 1). What is less clear is the weight to give to the result obtained for an unknown chemical using the ER-RUC assay (interactive or not with the ER) within the battery of the Tier-1 program. It should be interesting to give, in the introduction of the integrated summary report (ISR) (page 2, under C. "The Tier1 battery of assays"), a description of the strategy that will be used to classify a	The "weight of evidence" evaluation of the Tier 1 Battery will depend on the specific data and circumstances for a specific chemical, taking into consideration, for example, *in vitro/in vivo* discrepancies (if any), metabolism, and route of exposure. No general statement can be made about the weight to be given to the ER-RUC assay.

		chemical as negative or positive after the Tier 1 screening, that includes various in vitro and in vivo assays (ISR, page 3, table 1), and to give the weight of each assay in the final decision of the Tier 1 screening.	
1.5	WW	The assay is described in a succinct and clear manner. The word "specificity", which is used on page 8 with respect to the saturation binding assay, should be defined in this context.	The sentence in question is as follows: "The purpose of the saturation binding assay is to characterize the specificity and activity of the cytosol preparation and ensure that the ER activity is sufficient for the competitive assay." The protocol has been changed to remove the word "specificity" from Section 8, "Demonstrating acceptable performance in cytosol preparation and laboratory techniques", and to add clarification of the purpose of the saturation binding assay. The sentence now reads as follows: "Conduct a saturation radioligand binding assay to demonstrate ~~ER specificity and saturation~~ *that the estrogen receptor is present in reasonable concentrations and is functioning with appropriate affinity for the native ligand.*"
2. Relevance of the assay to its purpose			

2.1	PB	Yes, the assay is biologically and toxicologically relevant to the stated purpose. It enables to identify binder, equivocal or non binders. However, it does not allow to quantitative structure-activity relationship development for the moment. … The estrogen receptor binding assay using rat uterine cytosol will represent a validated assay for simple screening for interaction with the estrogen receptor in the context of a battery of in vitro assays. EPA has optimized and standardized the most important parameters of this assay and has shown that the resulting protocol is transferable to others laboratories and enable to identify chemicals that have the potential to interact with the endocrine system. . Analysis of compounds binding curves in appendix II of Appendix 5-overall report, enable easily to identify the nature of the compounds. Among very strong and strong binders (17b-estradiol, 17-ethynylestradiol, DES, meso-hexestrol, zearalenone and tamoxifen), zearalenone and tamoxifen are not strong binders (IC50s around 100 nM).	Agree. No response needed.

		Among moderate binders (genistein, norethynodrel, equol), genistein (IC50 around 100 nM) is a stronger binder than norethynodrel an equol (IC50s around 1000 nM). Among predicted weak binders (butyl paraben, nonylphenol, o,p'-DDT, 5alpha-dihdrosterone, bisphenol A, 4-n-heptylphenol, kepone, benz(a)thracene, enterolactone), it was very easy to identify real weak binders in the test (butyl paraben, nonylphenol, bisphenol A, 4-n-heptylphenol, kepone, enterolactone) and non binders (o,p'-DDT, 5alpha-dihydrosterone, benz(a)thracene). Finally among predicted negative compounds (atrazine, corticosterone, octyltriethoxysilane, progesterone and R1881), it was easy to identify weak binders (corticosterone and R1881) and real negative compounds (atrazine, octyltriethoxysilane and progesterone). Curiously, corticosterone was negative when the tested optional chemicals. Concerning R1881, it is likely that the expectation of non-binding is not correct. When the analysis was developed, the EPA expectation was that standardization of the	

		assay would allow precise and replicable quantitative analysis of log(IC50)s and Relative Binding Affinities. The expectation was also that precise, standardized methods of analysis would contribute to reproducibility and therefore use in other applications such as structure-activity relationship models. The variability of the results of the assay may not support for the moment use for quantitative structure-activity relationship model development. Even if strenght of binding could be evaluated, high variability of compounds RBAs were observed. In the face of the variability encountered, EPA is assessing whether such analysis could be replaced with a simpler analysis and still meet the needs of the Screening Program. However, it should be remembered that while intralaboratory variability was disappointingly high for at least one laboratory in this study, such variability is not expected to be as much of a problem for laboratories that demonstrate the ability to meet the required performance criteria. The limited time available to run this large study on 23 chemicals apparently did not allow development of the proficiency necessary to obtain precise runs in all	

		laboratories. The fact that results were almost all in accord with expectations when screening for interaction despite the variability in quantitative values shows that the assay is robust for this use. Furthermore, use of recombinant receptor rather than receptor obtained from whole animals will enable to use purified receptor which would decrease the variability of the binding assay. Finally, small ameliorations of the protocol, fine adjustement of performance criteria and increase of experience by the laboratories should allow the test to reach the expected purpose which is to determine the potential of a substance to interact with the endocrine system.	

2.2	IP	Only partially. In my opinion the assay does not take into full account some issues. Recent experiments have shown that one of the estrogen receptor isoforms namely ERβ is under circadian control in the mouse. The circadian system is well conserved so it therefore likely that this is occurs also in the rat model. Depending on the timepoint of cytosol preparation, the levels of ERβ expression may be very low. These low ERβ levels may result in cytosol preparations that fail to detect compounds that preferentially interact with the ERβ isoform and thus may considered "safe". This point should be taken into account to avoid missinterpretation of obtained results.	The Agency notes that the research on circadian control of ERβ has been carried out only in male mouse lung (Cai W, Rambaud J. et al. Mol Cell Biol 28(2): 784-93 and that there is insufficient information on which to base selection of an optimal time point, if any, at which to collect uteri from rats. If further research substantiates the importance of this effect in rat uterus and allows determination of an optimal time point for collection of tissue, future versions of this protocol may be adjusted.
2.3	IP	A second point regading the biolgical relevance of the assay is the role of the molecular chaperone hsp90 and its role in maintaining the estrogen receptors in a ligand binding state. Previous work with hsp90 associated receptors, in particular receptors like the GR, has demonstrated an important role for the hsp90 complex. In the case of the estrogen receptor ERα experiments have demonstrates that hsp90	The Agency recognizes the importance of heat shock proteins in maintaining the binding activity of ERα. A saturation binding assay is required for each batch of cytosol to provide assurance that the receptor is functioning as expected. If in the future the Agency decides to further refine the ER-RUC assay, the effect of molybdate may be an appropriate topic for further study, as may be the interaction of hsp90 with ERβ.

		is important but not crucial, for ligand binding. I am concerned that during cytosol preparation, the hsp90 complex may dissociate which would negatively impact on the receptors ligand binding activity. The presence of molybdate will stabilize the complex but it may not be sufficient. In the case of the second estrogen receptor isoform ERβ, very little is know regarding the putative role of hsp90 and regarding the stability of the complex. Again this may cause problems and in particular may explain some of the interlaboratory variation.	
2.4	IP	I think the assay is suitable to identify compounds that bind to the estrogen receptors ERα and ERβ. Again, the assay should be regarded in the context of being one of several assays. I have however certain reservations to the statement that the assay will cover metabolism. Metabolism and the presence of the p450 enzymes is in many cases a cell and tissue specific process and all the necessary components may not be present in uterine tissue and cells	The Agency did not intend to imply that the assay incorporates metabolism. There is no mention of metabolism in the protocol itself. The Integrated Summary Report makes it clear in several places that metabolism is *not* included in the ER binding assay (page 3 under "uterotrophic", page 68, page 70).
2.5	SS	The stated purpose of the RUC-ER binding assay is clear- it aims at identifying compounds that interact with the ER by testing their potential to compete with the	Agree. No response needed.

		natural ligand, 17β-estradiol, for binding with rat ER. It is clearly acknowledged that this competitive binding assay is not aimed at ascertaining the functional properties (weak/strong agonists, partial/mixed agonists or antagonists) of the test chemicals. However, the goal to categorize these molecules according to their binding affinities with the receptor is feasible and suits the formal objectives of the assay. It should be noted, however, that while there is a high degree of confidence that high affinity binders/competitors may activate or inhibit ER function in *in vitro* and *in vivo* models, it is doubtful whether weak or very weak binders/competitors would substantially interact with the ER and exert biological functions *in vivo*. Thus, this RUC-ER assay (and most likely the hER assay that is evaluated in parallel) allows the classification of potential endocrine disruptors by virtue of their intrinsic binding affinity for the ER. Further analysis of the biological or toxicological effects of these compounds entails independent cell- and animal-based assays.	
2.6	MLS	Yes, it is *[biologically and toxicologically relevant to the stated purpose]*, even if this single assay gives no indication about the toxicity of a chemical.	No response needed.

		The stated purpose of assays involving estrogen receptors is to evidence an estrogenic or an anti-estrogenic activity of the test chemicals. For both estrogenic and anti-estrogenic compounds, in most cases, the first step of their biological activity is the binding to the estrogen receptor (ER), before target genes transcriptional activation (estrogenic compounds) or inhibition of it (anti-estrogenic compounds). There are two isoforms α and β of the estrogen receptor, and some compounds act more specifically on one or another form. As the rat uterus tissue displays both α and β isoforms, the ER-RUC allows detecting both α and β ER ligands, which is an advantage over the binding assays using a recombinant receptor of a single isoform. As the transcriptional activation assay of the Tier-1 battery is specific to the α isoform of the receptor, it is relevant to use the ER-RUC allowing the detection of the binding to both α and β isoforms, in a screening approach using a battery of complementary assays, in order to decrease the rate of false positive results of	

		the Tier-1 testing.	
2.7	WW	The assay is relevant to the stated purpose, i.e., to determine the ability of a compound to interact with the ERs isolated from rat uteri". As discussed in the ISR on page 64, it is not clear whether testing compounds at 1 milliMolar (1 mM) is toxicologically relevant. Although the rationale for testing compounds at this high, physiologically non-relevant concentration is explained adequately (pp. 64-65), the results obtained may be difficult to explain and subject to mis-interpretation and even false conclusions.	On the advice of a previous Expert Group (see ICCVAM DRP), the Agency is including the 1 mM concentration (subject to solubility limitations) but recognizes that interpretation of data that include this high concentration may in some cases be difficult. To the extent that this high concentration helps to define the bottom of the binding curve it may help classify the ability of the test chemical to interact with the ER. However, to the extent that such a high concentration introduces interactions with the ER that may not be physiologically relevant, false positive results may occur. The Agency believes that a bias toward false positives in order to avoid false negatives is, within reason, acceptable for a screening study, particularly when the study is part of a battery of assays that will be considered together to evaluate the potential for interaction with the endocrine system, and when confirmatory studies will be run, as in the EDSP.
2.8	WW	The ER binding assay is fairly adequate as an initial screening tool. However, the ISR should describe to what extent the results (and/or interpretation of the results) of the assay might be influenced by recognition of non-genomic estrogenic signaling effects.	The Agency recognizes that plasma membrane bound ER and at least one transmembrane intracellular (GPR30) ER that initiate rapid nongenomic signaling events have been identified (Revankar et al. 2005. Science 307:1625). While it is likely that environmental contaminants will also have an effect on the function of these ER, there is currently no single *in vitro* method available to screen for interaction of these G protein-coupled receptors with the chemicals of interest. In the future, the Agency may review research developments in this area, and the impact that multiple ER types (genomic vs. non-

			genomic) may have on interpretation of the assays in the Tier 1 Battery. Since the ER RUC assay uses a uterine cytosolic preparation, cross contamination with subcellular components such as the endoplasmic reticulum (where GPR30 has been localized) is minimal. Thus, the ER RUC assay will provide information on the ability of the test chemicals to interact with the classical nuclear ER.
3. Repeatability and reproducibility of the assay			
3.1	PB	Considering the variability inherent in biological and chemical test methods, the results obtained with this assay are relatively repeatable and reproducible. Very strong an strong compounds. All of very strong, strong and moderate (17b-estradiol, 17-ethynylestradiol, DES, Meso-hexestrol, zearalenone, tamoxifen, norethynodrel, genistein and equol) binders were correctly determined to interact with rat uterine ER. The only exception is estradiol in laboratory Z. Binding at the concentrations tested (100 pM to 1 mM) showed clear interaction with the receptor but the laboratory did not adjust the test concentration range to a more dilute range that would have allowed characterization of	No response needed.

		the full binding curve as required by the protocol. Note that laboratory Z misinterpretated (compare table 27 ER-Ruc and table 2 p22 appendix 8) strong affinity compound 1 (estradiol), 3 (DES) and 13 (tamoxifen). Note that tamoxifen is not by itself a strong binder. 4-hydroxy tamoxifen (a tamoxifen metabolite) is the strong binder. When the three laboratories' RBAs were compared, similar values were obtained for very strong and strong compounds. Moderate compounds For moderate compounds, similar results were obtained for norethynodrel, equol and zearalenone while variability between laboratories is observed for genistein and tamoxifen (appendix 6 table 5, appendix 7 table 6 and appendix 8 table 2). Weak binders The weak chemicals had various classifications by the laboratories. Butyl paraben and bisphenol A were classified as positive by all of the laboratories. Nonyphenol (mixture), 4-n-heptylphenol, and enterolactone were positive, negative and equivocal for laboratory X, Y, and Z,	

		respectively. Kepone was positive in two laboratories and equivocal in the third. O,p'-DDT was negative in two laboratories and positive in the third. After analysis of the laboratory results, it was concluded in the ER-RUC ISR that these seven compounds were binders. I agree with the ER-RIC ISR exepted for o,p'DDT for which more experiments are necessary to conclude. 5alpha-dihydrotestosterone was negative in two laboratories and equivocal in the third. Benz(a)anthracene was classified as negative by all of the laboratories. Thus, most of the compounds that were expected to be weak binder were «positive« in their responses. However as mentionned above, the assay does not allow precise and replicable quantitative analysis of log(IC50) and RBA. For the negative chemicals (corticosterone, progesterone, octyltriethoxysilane, atrazine and R1881), laboratory X and Y had corticosterone listed as positive, and laboratory Z had it as equivocal. Curiously, this compound was correctly classified as negative in the optional portion. Progesterone and octyltriethoxysilane were	

		classified as negative by all three laboratories. Atrazine was classified correctly as negative by laboratory T and Z but positive by laboratory X. R1881 that was designed to be a negative control for the assay proved to be positive for the control runs, and as a coded chemical it was positive for laboratory X and Y and equivocal for laboratory Z. Thus, most of the compounds that were expected not to be binder were « negative» or « equivocal » in their responses.	
3.2	IP	In general I would say yes [the results obtained with this assay are sufficiently repeatable and reproducible]. I would suggest that the experimental issues raised previously be taken into account.	The issues raised “previously” in this reviewer’s comments are addressed elsewhere in this document.
3.3	SS	In most cases the answer to this question *[“Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?”]* is positive. The summary Table (Table 27, ISR, p.60) shows that despite a considerable variability in the data obtained by the 3 laboratories, most test compounds were correctly classified. The case of Compound #1 (17β-estradiol) in Lab Z that	Agree. No response needed.

		was classified 'equivocal' is peculiar and most probably results from deviations from the standard protocol of the competitive binding assay. In fact, it is somewhat surprising that despite considerable variability in datasets obtain by participating laboratories and among them- the final evaluation was powerful enough to categorize most test chemicals according to their rank. Yet, adherence to the protocols and further optimization of the assay will eventually lead to more reproducible results.	
3.4	MLS	The results were sufficiently repeatable and reproducible for the reference compounds (estradiol and norethynodrel), but for test chemicals, some results were too much dispersed (for example, for test chemical n°1, for lab Z, see in Appendix 5, page 1 of Appendix 2).	The Agency agrees that variability between runs and between labs was disappointingly high for some of the test compounds but notes that for purposes of classifying substances as interacting with the estrogen receptor, the other four peer reviewers agree that the assay is adequate. In the case of the specific example cited, the Agency notes that since the ER RUC protocol requires full concentration curves for each chemical it is unlikely that all of these runs would be considered acceptable if submitted in the EDSP due to incomplete definition of tops and bottoms of the binding curves.
3.5	WW	Yes, the results from the assay were extremely repeatable and reproducible especially in view of the variability inherent in biological and chemical test methods.	Agree. No response needed.
4. Clarity of the protocol			

4.1	PB	Yes, the description of the methodology is clear and concise and enable the comprehension of the objective. However the finding that strong binders were considered « equivocal » rather than clearly interactive in one laboratory when tested blindy deserves and need to be noted in the protocol. When binding at the concentrations tested showed clear interaction with the receptor, the laboratory did adjust the test concentration range to a more dilute range that would allow characterization of the full binding curve as required by the protocol. Thus, the protocol need to be adjusted to emphasize that a full curve must be obtained where there are clear indications of binding.	The protocol has been revised to emphasize that even if there is indication of binding at high concentrations, sufficient low concentrations are also to be included to define the top plateau.
4.2	IP	The protocol is technically sounds, and very detailed. In my opinion efforts should be done to make it easier to follow. It is stated that the groups performing the assays should have a suitable background in the field so they are well informed about the method. However I feel that the protocol is difficult to follow. In general I feel that an “Excecutive Summary” would be beneficial in particular to understand the purpose of the assay.	A short, concise description of the assay has been added to the protocol to give the reader an overview of the assay.

4.3	SS	The objective of the RUC-ER binding assay is clear. The background information on the assay and its inclusion in the study are well-discussed. The assay was developed following an in-depth review of the literature, which provides the scientific basis for the assay. The standardization and attempts to optimize the assay following the First Interlaboratory Study resulted in detailed, comprehensive and clear protocols. Experienced laboratories, such as the 3 that participated in the Second Interlaboratory Study, are expected to adapt and conduct the assay with minimal deviations from the protocols. Nevertheless, the results obtained from these 3 laboratories and their comments on problems encountered in the course of the study do indicate the need for further examination and optimization of the protocols.	The Agency has made adjustments to the protocols, based on comments received from the laboratories that participated in the second interlaboratory validation study and from peer reviewers. These minor adjustments are expected to reduce variability. For example, test chemical dilutions will be made in solvent rather than buffer so that the concentration of solvent across test chemical concentrations will be constant. The Agency has avoided making larger changes that in the aggregate might necessitate another validation study, noting that most of the peer reviewers found the current form of the assay relevant to its intended purpose.
4.4	SS	The conduct of the assay is expected to follow the detailed protocols. The two-test approach (the [^{3}H]-17β-estradiol saturation binding analysis, on one hand, and the competitive binding assay, on the other) is very important as it provides intra- and inter-laboratory quality controls and intrinsic assay controls. There are, however, certain problems that were encountered 'on the bench' or when data were collected, analyzed and interpreted. These issues are	The reviewer's comments in section "f" as submitted to the Agency are found as comments 10.7 through 10.19 of this document and are addressed individually there.

		presented and discussed at the end of this section under **f**.	
4.5	SS	The protocols are clear, detailed and meticulously describe all necessary technical aspects- from uteri excision to data acquisition and analyses. By conducting the assay according to protocols each laboratory is expected to observe and measure prescribed endpoints. In some cases along the Second Interlaboratory Validation Assay not all endpoints were achieved- mostly due to deviations from the prescribed protocols and/or due to technical problems.	Agree. No response needed.
4.6	SS	The platform provided to compile and prepare the data for kinetic analyses followed by statistical analyses has served well the participating laboratories and suits the requirements of the assay. Not being an expert in statistics, I cannot comment on the statistical methodology (Appendices 4 and 9). However, Appendix 13, which depicts graphs of acceptable saturation and competitive binding curves, raises some reservations on the acceptability of these graphs: the range of 17β-estradiol IC_{50} values (TRL's and Hammer's curves, Appendix 13/pp. 2-3) and of the weak positive runs (*ibid*, pp. 5-6) covers nearly an entire order of magnitude. It appears that narrower pre-defined limits are required in	The performance criteria that were established for estradiol and norethynodrel are, as recommended by this reviewer, narrower than the ranges shown by the "acceptable" runs graphed in Appendix 13. Performance criteria were established to accept 80% (not 100%) of the values for top, bottom, and slope. A further performance criterion limiting within-run variability of replicates also limits within-laboratory variability.

		these statistical analyses.	
4.7	SS	All reports of the results by the three laboratories are extensive and provide a solid basis for intra- and interlaboratory analyses of the data. Attention should be given to the suggestions related to **improving initial data manipulations** and **curve fitting** (Appendix 5/pp. 16-17)	The following responses address the suggestions in Appendix 5 of the ISR, pp. 16-17. There was a suggestion for EPA to provide a template using Prism software into which raw data could be entered, so that transferring data back and forth between spreadsheet software and Prism is avoided. The "tools" that EPA intends to provide for analysis will be only suggestions that the submitter or others may find helpful as they set up their own analyses. EPA will not vouch for accuracy of the tools it provides, and the tools are subject to change. The submitter of the data will be responsible for defending the accuracy of the analysis it submits. EPA will request only that the raw data be entered into a spreadsheet or spreadsheet-compatible electronic form, to facilitate entry into a centralized database. The suggestion to analyze data together from the same runs and to use shared top and bottom in the fit will not be used. Performance criteria are established for estradiol and norethynodrel separately, not in common, so that valuable information about accurate testing of each chemical separately is preserved. The final suggestion for curve fitting – "If it is clear that high percent bound values are the result of low counts in the total binding samples, then permit the data to be normalized using the "Normalize" function in Prism." – will be taken in part. If the top plateau for estradiol is significantly above the upper performance criterion, , then the top may be normalized using binding of estradiol at the lowest concentration in the reference

			curve as 100%. Normalization of the bottom is neither appropriate nor necessary. It suppresses information that could reflect the performance of the assay.
4.8	MLS	The answer to questions a) to e) *[Does the protocol describe the methodology of the assay in a clear, and concise manner so that the laboratory can:* *a. comprehend the objective;* *b. conduct the assay;* *c. observe and measure prescribed endpoints;* *d. compile and prepare data for statistical analyses; and* *e. report the results?*] is globally YES, but some clarifications are needed (see here below).	No response needed. (Clarifications addressed elsewhere.)
4.9	WW	The protocol adequately describes the assay in a clear and concise manner.	Agree. No response needed.
4.10	WW	In terms of advice, a *glossary of terms* would facilitate clarity between the EPA and the laboratory and to help avoid any misunderstanding or mis-interpretation in meaning of specific terms described in the protocol(s).	Section 4 of the protocol addresses terminology, as do Sections 9.6 and 10.7. Clarifications and additions have been made to the latter in response to specific comments from peer reviewers. (See, for example, comment 10.20.1.)
5. Performance criteria			
5.1	PB	Yes, appropriate parameters were selected and reasonable values were chosen to ensure proper performance of the assay.	Agree. No response needed.

5.2	IP	In my opinion yes [appropriate parameters were selected and reasonable values chosen to ensure proper performance of the assay]	Agree. No response needed.
5.3	SS	The saturation binding assay of 17β-estradiol and the competitive binding assay were introduced some 40 years ago with the discovery of the estrogen receptor and the availability of radioactive estradiol. The kinetic parameters, K*d*, B*max* and RBA have been standard endpoints in numerous binding assays. The K*d* values for 17β-estradiol binding to the RUC-ER in this study fall within the acceptable range. Obviously, B*max* values can vary widely, depending on the quality and ER concentration of the uterine cytosols.	Agree. No response needed.
5.4	MLS	Performance criteria were relevant, but I would add some criteria such as those presented in the answer to question n°3.	Laboratories will be requested to report the percent of total radiolabeled estradiol added that is bound for both the saturation binding assay and the competitive binding assay, for each concentration of radiolabeled estradiol used.
5.5	MLS	To ensure the performance of the assay to detect ER interactive chemicals, I would add some criteria such as maximum rate of false negative and false positive decisions, measured with known substances. In this validation study, 23 test chemicals were used to check this parameter, but the results are not very conclusive. On 22	The Agency believes it would be unreasonable to require a large number of known positive and negative chemicals to be tested each time a single test chemical is assayed.

		chemicals (if we exclude R1881), respectively 5, 2 and 5 false decisions were taken by labs X, Y and Z (see ISR, page 60, table 27). This rate of false decision will also depend of the weight of the ER-RUC assay in the Tier 1 screening (see my answer to question 1).	
5.6	MLS	Page 19, table 3: "*Performance criteria for competitive binding, reference and weak positive controls*". The figures of table 3 are different from figures of table 16 : … "*Performance criteria for second interlaboratory study*". Is table 3 valid for the 1st interlaboratory study only? As I understood, octyltriethoxysilane was not used as a negative control during the 1st interlaboratory. And in table 16, R1881 is still mentioned as negative control. Please give explanations about the differences between table 3 and table 16. Furthermore, the upper limit for the top plateau level for norethynodrel (% binding) is 110 and not 10.	In the ISR, Table 3 shows the performance criteria that are based on the data from the second interlaboratory study. Table 16 shows the performance criteria that were based on data from the first interlaboratory study, and which were supposed to be used to qualify laboratories to participate in the second interlaboratory study. The value for the upper limit for the top plateau level for norethynodrel in Table 3 of the ISR should indeed by 110, as noted by the peer reviewer.
5.7	MLS	[Suggestions for improvement/correction of the ISR on pages 33-1 to 33-3]	The ISR is an historical document and at this time there are no plans to update or correct it.
5.8	WW	Yes, appropriate performance criteria were selected to ensure proper performance of the assay. These criteria are clearly articulated throughout the ISR	Agree. No response needed.

6. Data interpretation			
6.1	PB	Yes, the data interpretation criteria are clear, comprehensive, and consistent with the stated purpose.	Agree. No response needed.
6.2	IP	The data interpretation is in my opinion clear and consistent with the purpose.	Agree. No response needed.
6.3	SS	The data interpretation follows rigorous kinetic and statistical analyses. Both provide clear representation and analyses of the data and are fully consistent with the stated purpose of the assay.	Agree. No response needed.
6.4	MLS	Yes, *[the data interpretation criteria clear, comprehensive, and consistent with the stated purpose]* but there are some inconsistencies between the different documents. Appendix 5 presents the report of the second inter-laboratory validation of the estrogen receptor binding assay (rat uterin cytosol). First, in Appendix 5, labs are named A, B, C, which is confusing for the reader, because in the ISR, the labs are named X, Z and Y respectively. In Appendix 5, page 10, it is indicated that the labs were required to classify the unknown chemicals as positive, negative or equivocal binders. This is not the classification described in the protocol	The protocol originally called for classifications of "positive", "negative", or "equivocal" for "binding". While the terms "positive", "negative", or "equivocal" could be retained and the word "binding" changed to "interaction", the terms "interactive" and "not interactive" were chosen to emphasize that the assay is not necessarily distinguishing binding from other types of interaction. The discrepancies between the cited tables in the ISR and Appendix 5 are due to the normalization of the values to the lowest concentration of the estradiol curve (explained on page 55 of the ISR), and to the change in assigning classifications. In the original protocol, which was the basis for the table in Appendix 5, laboratories were directed to average Relative Binding Affinities (RBAs) across the three runs to classify a substance as "positive". A chemical for which an RBA existed was classified as "positive".

		(Appendix 1, page 48,10.7.4), where the classification is described as interactive, not interactive or equivocal. The results of the classification of the 23 test chemicals by the 3 labs X, Y and Z are presented in both ISR and Appendix 5 documents, in table 27 and 5 respectively. I expected that the results were the same in both documents, but they are not for 14 chemicals out of 23, without any explanation in the ISR. This is very confusing for the reader!	This method did not deal explicitly with combining runs for which an RBA did not exist for at least one run. Thus the classification criteria described in the ISR (page 58) have been adopted.
6.5	WW	Yes, the criteria for data interpretation are described in a clear, comprehensive and consistent manner.	Agree. No response needed.
7. Appropriateness and completeness of validation			
7.1	PB	Yes, the test substances were appropriately chosen to demonstrate the performance of the assay. The test substances correspond to a wide range of strenghs and chemical structure. They are 3 very strong, 3 strong (including E2), 3 moderate, 9 weak and 5 negative compounds. Among negative compounds, atrazine (non-binder but estrogen-active) was included.	Agree. No response needed.
7.2	PB	Analytical methods and statistical methods were also appropriately chosen to demonstrate the performance of the assay.	Agree. No response needed.

7.3	IP	I think that both the test substances and the methods are appropiate for the purpose [of demonstrating the performance of the assay].	Agree. No response needed.
7.4	SS	Test substances were carefully collected to cover a wide range of compounds that bind to the ER at very strong (higher than 17β-estradiol), moderate or weak affinities, regardless of their biological functions. The analytical methods are appropriate to demonstrate the performance of the assay.	Agree. No response needed.
7.5	MLS	The choice of a panel of strong, very strong, moderate, weak or negative binders seems relevant. However, it would have been better to avoid substances without available data in the literature about their binding capacity (in ER-RUC assays) to the ER.	While the Agency agrees that if there were no other considerations, testing only chemicals for which reliable ER-RUC data already existed in the literature might have been appropriate. However, the Agency designed the study so that the same chemicals would be tested in both the ER-RUC assay and in assays using human recombinant ER (hrER). The list of chemicals for hrER testing was developed by an international group and subjected to review by independent scientists. Thus the selection of chemicals was not based solely on availability of data from ER-RUC studies.
7.6	WW	Yes, these three aspects *[test substances, analytical methos, and statistical methods]* have been amply presented in the ISR. The substances and methods are appropriate to demonstrate the performance of the assay.	Agree. No response needed.
8. Strengths of the assay			

8.1	PB	As an in vitro assay, the ER-RUC provides direct interaction between chemical and ER. The assay provides consistent responses at the simple screening level, across laboratories, and these responses are in line with expectations for those chemicals tested whose ER binding behavior is well-established. Compounds that do not interact with ER consistently test negative in the assay. The assay is short and inexpensive compared to in vivo tests.	Agree. No response needed.
8.2	SS	The strengths of the assay are adequately addressed (ISR, pp.68-9). The major strength was the ability of the RUC-ER assay to classify correctly most test compounds, despite inconsistencies among the participating laboratories and various technical problems. Clearly, once the recombinant hER binding assay is completed, a thorough comparison and analysis of the results of the two independent assays is required for further validation of the RUC-ER as a relatively simple and affordable screening assay	The Agency agrees that a thorough comparison and analysis of the results of the ER-RUC interlaboratory study and the results of the hrER-alpha interlaboratory study will be appropriate. However, the Agency believes that validation of the ER-RUC assay for use as a screening assay in the Tier 1 Battery is complete and sufficient even in the absence of such a comparison. As noted in other peer review comments (see, for example, comments 2.6 and 8.3), the ER-RUC assay covers both alpha and beta isoforms of the ER so it is not clear that the hrER-alpha assay alone will provide an appropriate standard by which to judge the validity of the ER-RUC assay.
8.3	MLS	The strengths and the weaknesses have been adequately addressed. The strengths of the assay are, for the most, relevant for all estrogen receptor binding assays but one important specificity of the ER-RUC is stressed: it is that the rat uterine	Agree. No response needed.

		cytosol contains both isoforms alpha and beta of the estrogen receptor. The high degree of homology between the rat and the human ER ligand binding domain could also be mentioned here.	
8.4	WW	The strengths and limitations of the assay are not clearly specified the ISR. This should be done.	In the ISR, strengths were discussed in Section VI.A., "Strengths", and limitations were discussed in Section VI.B., "Weaknesses". Without further discussion by the reviewer of what is unclear in these sections, the Agency is unable to address this comment further.
9. Limitations of the assay			
9.1	PB	The assay is sensitive to many details of preparation and technique and can show wide variability if not performed exactly as stated in the protocol. It is subjected to problems if the receptor concentration in the cytosol is too low or too high. Another problem is that the tubes are not kept cold at all times during preparation, incubation and separation of bound from free tracer. This problem is certainly responsible of the the top plateaus for the standart chemicals (estradiol and norethynodrel) often exceeded the performance criteria by several tens of percentage points. However, the data suggest that a lab that meets the performance criteria for teh standart and weak positive is likely to generate data that is much less varaiable than laboratories that	Agree. No response needed.

		do not meet performance criteria.	
9.2	PB	The analysis of datasets was relatively complicated. The standardization of the assay does not allow precise and replicable quantitative analysis of log(IC50) and RBA.	The analysis method, while somewhat complicated, is intended to make use of all the data in the run, reducing dependence on any single subset of replicates such as the subset that spans the 50% binding level. The Agency is considering whether a simpler analysis will allow consistent classification of chemicals, particularly chemicals which are equivocal or interact weakly with the receptor. The analysis was originally developed because it was expected that standardization of both the assay and the analysis would allow reasonably precise and repeatable determination of Relative Binding Affinities. This is no longer an objective of the assay.
9.3	PB	Finally, the assay requires the use of animals.	See response to comment 9.6.
9.4	IP	A further limitation of the assay is due to the use of laboratory animals. There are a number of parameters that are difficult to control, such as quality of feed. Has the possibility to use established cell-lines been considered?	See response to comment 9.6.
9.5	SS	The weaknesses of the assay (*ibid*, p.69-70) refer mostly to technical and methodological aspects such as, ER concentration in the assay, insolubility of test compounds, complicated analyses of data and the inevitable use of rats. There are no satisfactory explanations to the lack of adherence of some participating laboratories to the standard protocols.	See responses to comments 9.2 and 9.6

9.6	MLS	The main weakness of the assay comes from the use of animals to prepare the binding fraction, with not only ethical, but also technical consequences (lack of reproducibility in receptor preparations). The solution is the use of recombinant ER binding assays, using both alpha and beta isoforms of the ER.	The Agency agrees that it is appropriate to investigate the use of human recombinant ER as a potential replacement for the cytosol preparation used in the ER-RUC assay. It is currently participating in an international, multi-laboratory effort to validate the hrERα assay for use in screening. However, availability of recombinant ERβ for widespread screening purposes is limited due to patent considerations.
9.7	WW	The strengths and limitations of the assay are not clearly specified the ISR. This should be done.	See response to comment 8.4.
10. Suggestions for improvement			
1. Conduct abbreviated pilot studies			
10.1.1	PB	<u>As suggested in RTI project (appendix 5-overall report, 2n), a first screening of all compounds at high concentrations (1 mM and 0,1 mM for example) would enable to identify binders and indicate which compounds would have benefited from adjustement to lower concentrations. This should significantly cut down on the number of reruns required and the cost of characterizing compounds.</u>	The protocol does not preclude a laboratory from doing such an initial experiment. However, while this approach may prove useful for determining the appropriate concentrations to provide a full curve for chemicals that moderately or strongly interact with the ER, it is not optimal for use with negative or less potent chemicals as it will not provide sufficient information. Based on the results from the validation study, the Agency does not believe that use of only two concentrations (particularly one as high as 1 mM; see comment 2.7) would be sufficient to characterize many of the chemicals that are likely to be screened using this assay. Thus, the EPA will continue to request that a full curve be generated.

10.1.2	SS	RTI scientists listed several suggestions regarding the assay protocol (Appendix 5/p. 15). The suggestion to **conduct an abbreviated pilot study** (screen test compounds at 10 μM and further characterize only those that inhibit [^{3}H]-17β-estradiol binding by 50% or more) sounds reasonable. However, the limit in my opinion should be lower than 50% to eliminate false negative determinations from the assay.	See the response to comment 10.1.1.
2. Do fewer tubes per run.			
10.2.1	PB	Reduction of the number of points by run. In the analyses produced by the individual laboratories in this study, the top plateaus for the standard chemicals often exceeded the performance criteria by several tens of percentage points. The reason for the high plateaus may be related to the solvent control tubes. Such tubes placed at the end of the run (of several test chemicals run simultaneously) often yielded lower dpms than similar tubes placed at the beginning of the run. The average of all solvent control tubes was therefore lower than it would have been had only the first solvent control tubes been included. The lower average could have contributed to the appearance of higher-than-solvent-control values for the estradiol and norethynodrel. Processing this large number of tubes may have increased	The protocol has been changed to recommend that a maximum of three test chemicals be included in a single run, and that this number be reduced further if within-run variability appears to be a problem.

		variability due to such factors as increased duration of exposure to room temperature (and subsequent denaturation of the receptor), and diminished ability to monitor partial pellet loss after centrifugation. This potential source of variability is expected to be less of a factor for laboratories if only one chemical is being tested at a time.	
10.2.2	PB	Consider not running all the standards with every assay. Limit to running E2 and NOR once with each new cytosolic preparation and only E2 with each assay would reduce the number of points by run and would ameliorate the results.	The purpose of including the weak positive and negative controls with each run is to minimize the likelihood of false negative and false positive results on the test chemical. Dissociating the controls from the runs may not adequately control for variables that can change with each run.
10.2.3	SS	I do, however, tend to agree that the norethynodrel standard curve is redundant.	See response to comment 10.2.2.
10.2.4	SS	The suggestion **not to run full standard assays** (17β-estradiol and norethynodrel), but use the compounds at their respective IC_{20} and IC_{80} values is problematic: close examinations of the acceptable standard curves in various runs of each participating laboratory and among them fails to provide absolute concentrations of each compound that induce 20 or 80% displacement of [^{3}H]-17β-estradiol binding. Furthermore, full standard curves and the experimental values of IC_{50} of unlabelled 17β-estradiol do provide important information on the quality and saturability of the assay. Similarly, a Hill	Agree. No response needed.

		coefficient value close to 1 indicates a simple binding mechanism.	
3. Pipette larger volumes than 10 µL.			
10.3.1	PB	Variability due to small amount (10 microliters) of test chemical and standard. Choice of an higher volume (50 microliters) would reduce variability.	Dilutions are prepared in solvent (rather than buffer) in order to keep solvent concentrations constant across all test chemical concentrations. Using a higher volume (e.g., 50 µl) would increase solvent concentration in the final tube to an unacceptable 10% (in the case of ethanol) unless the concentration of initial stock were raised to 5x the current value (i.e., raise 100 mM to 500 mM). Given the diversity of the chemicals that will be evaluated using this assay, the preparation of such high concentrations is not likely to be feasible for many of the chemicals. However, as noted in the protocol, the laboratories are free to attempt this approach as long as they do not exceed the maximum allowed solvent concentration in each assay tube.
4. Reduce the concentration of radioactive estradiol			
10.4.1	PB	Radioactive estradiol concentration could be decreased. The use of radioactive estradiol (3H-E2) 1nM in the assay is perhaps too high to identify weak binders. Due to solubility, tested compounds cannot use at higher concentrations than 1 mM. At this	As described in Section III.G of the ISR, the Agency considered reducing the concentration of radioligand to increase the sensitivity of the assay. A study was conducted to compare 0.5 and 1.0 nM. Although a clear difference in sensitivity was observed and potential problems with solubility could thus be mitigated somewhat as the reviewer noted, the

		concentration, weak binders can be classified as unequivocal (due to lower than 50% displacement). Use of radioactive estradiol 0,1 nM would increase the sensibility of the assay and enable to better characterize weak binders. However use of 0,1 nM 3H-E2 will give lower signal than with 1 nM estradiol and an higher depletion by the receptor.	Agency determined that the reduction in specific binding disintegrations per minute (dpms) would be detrimental. Unexplained problems with low counts by one laboratory during the second interlaboratory validation study confirm that maximizing dpms, within reason, is an appropriate goal.
5. Use a source of estrogen receptors other than rat uterus			
10.5.1	PB	Remark 2. Use of estrogen receptors from another source than rat uterus. Rat uterine cytosol contains both ER alpha and beta. These two receptors have not the same affinity for some natural, industrial or pharmaceuticals compounds like phytoestrogens, biphenols, cosmetics and ethynylestradiol (Kuiper et al, 1995, Paris et al, 2001, Escande et al, 2006, Molina et al, 2008). Variability of relative binding affinities (RBA) results can be induced by variation of the ER alpha /ER beta ratio in the different rat uterus cytosol batches. Thus, the protocol could be improved by using recombinant human ER alpha (or ER beta) receptor (Gangloff et al, 2001, Eiler et al, 2001, Pillon et al, 2005). Recombinant ER is tagged with six histines (6-His) or	The Agency is participating in an international effort to validate a binding assay using human recombinant estrogen receptor alpha (hrERα). A validation effort for hrERβ was considered but use of such an assay appears to be hindered by patent considerations. The Agency will consider the extent to which an hrERα-based assay alone will be an improvement over the ER-RUC assay. However, the hrERα assay is not currently ready for use in the EDSP and the Agency does not believe it is appropriate to delay the testing phase of the Program until the hrERα system has been validated. The Agency has not examined human cell lines as sources of estrogen receptor, but assumes at this time that a separate validation effort would be needed before allowing their use in the EDSP. Additional guidance on the extent of such a validation effort that

		Glutatione-S-transferase (GST) which enables purification by Nickel- or glutathion-Sepharose. Advantages of purified ER are the producion of reproductible, well charaterized batches and a reduced variability of the assay by limiting binding interference with non ER proteins. Alternatively, human cell lines (HeLa, U2OS) expressing ER alpha or ER beta (Escande at al, Sotoca et al, 2008) can also be used as source of estrogen receptor for binding experiments. In these cells, binding experiments could be done in lysed cells or whole-cells (Escande et al, 2006, Molina et al, 2008). Advantages to use whole-cells is a simpler protocol with higher high-throughput screening possibilites. Inconvenients are concentrations of compounds that cannnot exceed 10^{-5}-10^{-4}M when lived cells are used.	is needed may be available when the OECD's effort to define Performance-Based Test Guidelines comes to fruition.
6. Standardize the time of kill for obtaining uteri			
10.6.1	IP	I feel that the assay as first line approach is valid, but would benefit from additional scientific experimentation and validation, in particular characterize some of the issues regarding the circadian expression of the	The Agency is aware of both of these issues (potential circadian expression of ERβ and stability of the hsp90/ER complex). It regards these as research issues that may lead to improvements to the assay. However, performing the research and re-validating

		ERβ isofom and of the potential stability of the hsp90-ERα and hsp90 ERβ complex. I feel this limitations should be addressed in the text and efforts should be undertaken to fill the scientific gap regarding these issues. … The possible effect of the circadian system can be covered by characterizeng the perak of ERβ expression at different time points, the food composition of the animals should be centralized, and stabiltiy of the hsp90 complex needs to be better addressed. A experimental possibility would be to use cell-lines, I understand however that this would lead to less material being available.	the assay at this time would delay use of the ER binding assay in the Screening Program. Since the assay as currently described is adequate as a first-line approach, it would not be appropriate to delay use of this assay. If in the future the Agency decides to refine the ER-RUC assay, optimizing the time of kill to maximize the concentration of ERβ, and stabilizing the hsp90/ER complex may be items to consider.
7. Discard uteri from incompletely ovariectomized animals			
10.7.1	SS	The choice of ovariectomized Sprague-Dawley rats for uteri collection is good. However, the argument made by the RTI scientists to extend the 8-day period after ovariectomy before excision of uteri due to remaining endogenous estrogens (Appendix 5/p.5) is not pharmacokinitically sound: the half-life of 17β-estradiol in Sprague-Dawley rats is about 10 hours (*Petroff and Mizing. Reproductive Biology, 3:131, 2003*). Thus, a full clearance of 17β-estradiol is expected after nearly 20 half-lives within the 8 days	The protocol has been changed to direct that the laboratory carefully examine the uterine tissue for signs that residual ovarian tissue may be present, and discard uterine tissue that is obviously compromised.

		following ovariectomy. It is more likely, however, that the presence of stimulated uteri (due to fluid retention) was the consequence of an incomplete ovariectomy. In fact, RTI scientists confirmed the presence of residual ovarian tissue in some operated rats (*ibid*). It is therefore suggested that in cases where such remnants of ovarian tissue are found or when the excised uteri appear imbibed to discard them from the assay.	
8. Allow a range of cytosol protein amount to be used in the saturation binding assay			
10.8.1	SS	The requirement to perform a saturation binding assay for each RUC preparation prior to the competitive binding assay is mandatory: it allows the determination of the ER concentration in cytosols in term of maximal binding capacity of [^{3}H]-17β-estradiol. The large variation in the total protein content of the various rat uterine cytosols prepared by the 3 participating laboratories, particularly in the three cytosols prepared in Lab Y (ISR, Table 22, p.53) attests to this need. Despite the extraordinary high protein content in 2 of the cytosols in Lab Y, the mean maximal binding (B*max*) reported was 2-3-fold lower that that reported by Lab X, whose cytosols contained significantly lower protein content.	The Agency has clarified the protocol to focus on use of the amount of cytosol protein that results in 25-35% binding of the total amount of radiolabeled estradiol added to the tube. The Agency recognizes that the receptor content of cytosol may vary across cytosol batches and it was never the Agency's intention to specify that a fixed amount of protein be used across all saturation binding assays.

		This is probably due to the addition of a fixed amount of cytosolic proteins to the test tubes, leading inevitably to low content of ER. Therefore, I suggest to pre-determine the range of acceptable protein content in rat uterine cytosols. Large variations in protein content as reported by participating laboratories may complicate the interpretation of results.	
10.8.2	SS	The suggestion to **determine the K*d* values of [^{3}H]-17β-estradiol only once** for each newly prepared cytosol is acceptable, as long as the B*max* values do not significantly change when the cytosol is tested repeatedly in independent assays. Moreover, inclusion of a control for receptor stability (see above) will also warrant similar experimental conditions among assays.	The Agency reviewed the saturation binding data generated during the second interlaboratory validation assay. Where data from several acceptable runs were available using the same batch of cytosol, B_{max} (nM) values varied by less than a factor of 3 across independent runs, and usually by less than a factor of 2. Data were not available to evaluate the change in B_{max} over storage time but the assessment of cytosol storage time as measured by competitive binding assays (described in Section III.F of the ISR) provide some assurance that the receptor is stable over the 90-day storage period allowed by the protocol. Thus the Agency believes that determining the K_d once for each batch of cytosol (from the three runs that constitute a single saturation binding experiment) will be sufficient.
9. Consider reducing protein concentration in the competitive binding assay if slope is significantly greater than 1.			
10.9.1	SS	It is important to keep the total receptor concentration in the binding assay low enough to restrict the binding interaction of	This has been added to the protocol as a suggestion for a laboratory to consider if the performance criterion for slope for estradiol in the competitive binding assay

		17β-estradiol with the receptor to a simple (single and non-interacting binding sites) mechanism (linear Scatchard plot, Hill coefficient=1) and to avoid complex binding kinetics due to receptor dimerization that occur at higher receptor concentration. It is therefore recommended to scrutinize all experiments where the Hill coefficient of 17β-estradiol binding is significantly higher than 1.0 and dilute the cytosol accordingly. Dilution of the ER in the competitive binding assay to 0.5 nM usually suffices the requirement of a simple binding kinetics.	is not met because the slope is too steep.
10. Include "receptor stability" control tubes to compare post-incubation cytosol with freshly thawed cytosol.			
10.10.1	SS	Often, the 'top plateau levels' of competition curves were markedly higher than the expected 100%. There are several technical explanations to these results, such as, underestimation of the control binding, receptor instability or variable concentration of solvents in test tubes. The recommendation to use a fixed volume of solvent (2 % ethanol) in all tubes, including controls, will eliminate the latter cause. It is suggested to design and include 'receptor stability' controls in all assays. The simplest assay requires an incubation of the cytosol under the same assay conditions of the competitive binding assay (16-20 hours at	Comparing post-incubation cytosol to freshly-thawed cytosol from the same batch has been added to the protocol as a suggestion that the laboratories may wish to consider.

		4ºC) with no ligands. This cytosol is then used in a binding assay along with a freshly thawed cytosol. The binding capacity of these two cytosols is compared in a binding assay with saturating concentrations of [^{3}H]-17β-estradiol, an efficient anti-protease cocktail and incubation at high temperature (15-20ºC) for 1-2 hours followed by the HAP treatment. This simple assay eliminates misinterpretation of data due to receptor instability during the long incubation period at 4ºC.	
11. Use the saturation binding data to estimate the appropriate receptor concentration for the competitive binding assay.			
10.11.1	SS	By definition, if the [^{3}H]-17β-estradiol is present at a concentration that saturates the receptor- an equal concentration of unlabeled 17β-estradiol is expected to reduce the specific binding of [^{3}H]-17β-estradiol by 50% due to the 1:1 dilution of the radioactive ligand. In the case of the presence of excessive binding sites (non-saturating conditions) the 50% competition of the labeled 17β-estradiol with the unlabelled 17β-estratiol is observed at a ratio higher than 1:1 for the unlabeled estradiol. This may lead to an inaccurate estimation of IC_{50} values. The range of IC_{50} values is depicted in the Acceptable	The recommendation that the protein concentration be established such that no more than 10 – 15% of the total radioligand added to the tube is bound when no competitors are present ensures that there are far more radioligand molecules than receptor sites in the assay.

		Standard 17β-estradiol Curves (Appendix 13/pp.1-3): the range reported by TRI is narrow enough and agrees with the assay's terms. However, the ranges shown by TRL and Hammer cover an entire order of magnitude. These wide ranges most probably result from variable receptor levels and non-saturating [^{3}H]-17β-estradiol concentrations. Therefore, I suggest using the data obtained from the standard 17β-estradiol saturation curves to calculate and use the optimal receptor concentration for maximal binding capacity (saturation) of [^{3}H]-17β-estradiol. In Appendix 6 (Appendix D, p.3, 8 and 12) Lab X presents ER binding curves that were saturates with 2-3 nM 17β-estradiol. Lab Y presents a summary of 3 saturation curves in Appendix 7 (p.16) in which maximal binding was observed with 0.5-1.0 nM [^{3}H]-17β-estradiol. Lab Z gives 4 saturation curves (Appendix 8/p. A-17, , C-20, D-17, E-17) which did not reach saturation at 1-2 nM [^{3}H]-17β-estradiol and one curve (B-22) that saturated between 1.3-3.6 nM [^{3}H]-17β-estradiol. These inconsistent data may explain intra- and interlaboratory variations of the calculated kinetic parameters (K*d*, IC_{50}, and RBA). The equivocal results of the competitive binding assay of Test Chemical #1 (17β-estradiol) presented by Lab Z could have been	

		avoided if the assays were conducted with a prescribed concentration ER and saturating concentrations [^{3}H]-17β-estradiol	
12. Use physical methods such as light-scattering spectroscopy to determine solubility			
10.12.1	SS	It is highly recommended that researchers will assess potential precipitation of hydrophobic test compounds in the binding assay mixture by using physical methods such as light scattering spectroscopy. Precipitation problems can often be solved with some organic solvents. Appendix 3 shows the compatibility of the binding assay to 2% ethanol. However, the recommendation to allow DMSO up to 20% in the assay is worrying; the U-shaped displacement curves of various ligands to the ER (Figures16, 18, 22 and 23) indicate that DMSO may affect the receptor and change its binding kinetic properties in a concentration-dependent manner. Indeed, such effects of solvents (*i.e*., dimethylformamide) on the binding kinetics of the ER were previously reported (*Sasson & Notides, J. Steroid Biochem. 29:491-5, 1988*).	In a follow-up solubility study that was unavailable at the time of the peer review, the Agency found that plate-reading spectrophotometers appear to be more sensitive for this use than visual inspection but can lead to false positive results for insolubility if, for example, the test compound itself absorbs at the wavelength chosen. (In the solubility study, 650 nm was used and resulted in false positives for several chemicals.) More-specialized equipment such as nephelometers, which measure light-scattering specifically (as opposed to simple attenuation) are of unproven reliability for this use and can be prohibitively expensive. The Agency requests inspection using 20x magnification (e.g., dissecting microscope) since this method provides significant improvement over unaided visual inspection (i.e., fewer false negatives) but is less susceptible to false positives than spectrophotometry is. Appendix 3 of the ISR shows in Figure 1 and in text (page 10) that concentrations of DMSO up to 16.7% do not affect the estrogen receptor. The protocol restricts DMSO concentration to 10% (not 20%).
10.12.2	SS	Similarly, problems with **solubility of chemicals** are critical and the suggestion to	The protocol has been slightly revised to limit the choice of solvents to DMSO, ethanol, or water, whose

		employ various approaches to solubilize the compounds, other than these specified in the protocols, is sound. Yet, this requires control experiments showing that the solubilizing solvent or other compounds used do not interfere with the binding kinetics of 17β-estradiol.	effects on the estrogen receptor binding assay have been characterized as shown in Appendix 3 of the ISR.
13. Add insulin or IgG to raise protein concentration of low-protein cytosols, to minimize loss of test chemical to test tube walls.			
10.13.1		The absorption of hydrophobic chemicals to test tube walls is not always eliminated by using siliconized glass (borosilicate) tubes. Compounds with a high partition coefficient (*e.g.,* the Log*P* of tamoxifen is 6.58 in comparison with 3.67 of 17β-estradiol) do interact with these seemingly 'inert' surfaces. This absorption problem may be solved by other means than just using solvents at different concentration. It was shown (*ibid*) that a protein content higher than 2.5 mg/ml in the binding assay reduced significantly tamoxifen lose to the tube walls. Various proteins, such as insulin or IgG, which minimally increase the non-specific binding of 17β-estradiol, can be added to low-protein cytosols for this purpose. Yet, such unusual solutions are rare because ethanol is a good solvent for most chemicals and is well tolerated in the binding assay up	The cytosol used in the ER RUC assay contains additional non-ER proteins that serve as carrier proteins and help keep compounds with high partition coefficient in solution. Adding other proteins to the cytosol might require re-optimization and re-validation of the assay. If in the future the Agency decides to refine the ER-RUC assay, the addition of a carrier protein may be considered.

		to a final volume of 2%.	
14. Require internal standards to ensure accurate conversion of cpm to dpm.			
10.14.1	SS	The HAP sedimentation procedure is used for separating free from bound [^{3}H]-17β-estradiol. Following washes the HAP slurry is extracted with ethanol and a 1 ml aliquot is then mixed with 14 ml of scintillation cocktail to measure radioactivity in a β-counter. An accurate conversion of counts-per-minute (cpm) to disintegrations-per-minute (dpm) is of a paramount importance for subsequent data analysis. The counting efficiency may differ among various scintillation cocktails and β-counter settings. Various methods are available to estimate counting efficiency and correctly calculate dpm: internal standards, Channel Ratio (CR) or external standard (γ-radiation). It is imperative that each participating laboratory be able to reliably calculate dpm of radioactive samples. The use of internal standards (that is, fixed amount of μCi of [^{3}H]-17β-estradiol counted with 1 ml ethanol in 14 ml of scintillation cocktail) is often adequate; Erratic CR values identify counting vials in which quenching was irregular.	A note has been included in the final protocol regarding the appropriate selection of scintillate. The recommendation for the calibration of the scintillation counter and routine determination of efficiency for counting [^{3}H]-17β-estradiol is now also included in the protocol and covered under the Good Laboratory Practices for each laboratory.

15.			
10.15.1	SS	**Stability of materials** is very important and I fully support the suggestions regarding this issue.	The protocol was changed to specify that radioactive estradiol must be stored at -20 rather than 4 ºC and routinely evaluated for purity per the recommendations of the manufacturer. Receptor preparation and maximal storage interval are already included in the protocol, and the protocol already specifies that solutions of key labile solutions such as dithiothreitol must be made fresh daily.
16. Include a "suggested supply list" that identifies brands and suppliers of reagents.			
10.16.1	SS	Obviously, **brands of reagents and suppliers** do vary. Thus, the idea to provide a "suggested supply list" is encouraged.	The Agency prefers not to endorse particular products but has noted, in some instances, typical products that are likely to be appropriate. Other products may also be appropriate. In cases where there are significant differences in performance which may be relevant (e.g., forms of hydroxyapatite), such differences are noted.
17. Always pipette minute quantities (10 µL) first, into a dry tube.			
10.17.1	SS	The problem regarding **difficulties in pipetting 10 µL of test chemical and standards** can easily be solved if these test chemicals or standards are pipetted first to dry tube- no immersion of pipette tips in the binding mixture or an incomplete delivery of the tip's content are encountered this way.	A suggestion to pipette the 10 µl of radioligand to a dry tube has been added to the protocol, with a cautionary note to be careful of the potential for evaporation if ethanol is the solvent.

		Accordingly, the order of addition of the assay components should be modified accordingly.	
18. Investigate the potential for dose-dependent mixed agonism/antagonism, as well as sex-based differences in response to estrogen receptor binders, in the EDSP (but not with this assay).			
10.18.1	SS	The case of dual (mixed) estrogenic-antiestrogenic properties of certain ER ligands was not addressed in the study. Some compounds, like estriol, act as weak estrogens when present or administered alone to immature or ovariectomized animals. However, when present in excess over estradiol, these compounds exert potent antiestrogenic properties. This duality in function depends on the concentration of the competing ligand relative to the ambient 17β-estradiol concentration. This phenomenon results from a partial activation of the ER when the test compound is present alone, and to an aberrant dimrization of the ER in the presence of the natural ligand (*Melamed et al., Mol. Endocrinol. 11:1868-78, 1997*). The simple binding assay employed in this study is not aimed at analyzing such complex kinetic interactions. However, these complex interactions may occur in animal studies. Hence, it is recommended that functional assays aimed at determining the properties	This comment applies to the appropriateness of the battery of assays and test conditions used in the EDSP, but not to this assay *per se*.

		of ER binders *in vitro* and *in vivo* would carefully explore their potential to act as mixed agonists/antagonists. Equally important is the potential of such compounds to exert antiestrogenic properties in female with normal ovary function and secretion of 17β-estradiol, while acting as estrogenic agonists in males, due to the absence of significant levels of natural estrogens. This spectrum of functions should be acknowledged and examined carefully in studies that aim at determining biological and toxicological functions of such ER ligands.	
19. When data are available, compare results from ER-RUC to results from hrER assays.			
10.19.1	SS	This assay is one in a battery of assays aimed at identifying endocrine disruptors. The RUC-ER binding assay, which was developed some 40 years ago, provides simple kinetic analyses of ER binding interactions with natural and synthetic agonists and antagonists. The principles of the assay have not significantly changed over this period. A major development is the source of the ER: uteri from various species (mostly rats and calves) have been used extensively for years. The current availability of recombinant human ERα and ERβ now enables studies more relevant to human.	As noted above, the Agency is participating in an international effort to validate an hrERα binding assay, using the same test chemicals as were used in the second interlaboratory validation study for the ER-RUC assay. The Agency agrees that the ER-RUC assay is adequate for first-line screening.

		Thus, it will be interesting to compare the results of the RUC-ER assay to the recombinant hER project that is run in parallel. Both assays are relatively simple and can be included in the battery of assays to provide a first-line screening of potential ligands to the ER. I believe that the RUC-ER meets these criteria. Nevertheless, the protocols of the assay require some further optimization.	
20. Define the term "total binding" and clarify what is to be graphed.			
10.20.1	MLS	*1. Data for plotting the competitive binding curves : the y axis.* In Appendix 1, page 46, paragraph 10.7, the data analysis is not totally clearly explained. It is indicated that "*The competitive binding curve is plotted as specific [^{3}H]-17β-estradiol binding versus the concentration (log10 units) of the competitor*." If we look at 10.7.1. "*Terminology*", the definition of the specific binding is "*Total binding minus non-specific binding*", but the "*total binding*" is not defined, and thus the way to calculate the specific binding is not indicated. It should be added that :	These clarifications have been made in the protocol.

		Total binding is radioactivity in DPMs in the tube that contains [3H]-17β-estradiol and receptor, in the absence or presence of competitor. Furthermore, it is not clear which figures we have to plot on the y axis of the competitive binding curve. In Appendix 4, page 14, Paragraph 2.1.1. (Input data specification), it is indicated that "*input data for the dependent variables should be standardized and expressed as "% binding of the reference ligand to the receptor*", which I translate as the ratio : $\frac{\text{Total binding (in presence of competitor) - NSB}}{\text{Total binding (in absence of competitor) - NSB}}$ As this is indicated nowhere in the documents (may be it is clear on the excel worksheets, but unfortunately, we didn't receive the excel files, this should be added to the terminology, under "*specific binding*".	
21. Clarify that the percent of radioligand bound at the 0.03 nM concentration in the saturation binding assay should be reported, and be sure that the optimal binding range stated in the protocol is consistent with the column name in the data analysis template.			
10.21.1	MLS	In Appendix 1, page 20, 9.1.5.	These clarifications have been made in the protocol.

		"*Standardization of receptor concentration*", it is indicated that "*For the saturation assay, the optimal protein concentration binds 25 -35% of the total radiolabeled estradiol that has been added to the tube. To ensure that this percent range of radioligand is bound at the lowest concentration of radioligand added to the assay, the 0.03 nM concentration shall be used to make this determination for the saturation binding assay*". Under 9.7. ("*test report*", pages 26 - 27 of Appendix 1), it is not clearly asked to the lab to report the percent of bound radioligand at the 0.03 M concentration. It should be clearly asked to the labs to report this value. Furthermore, I tried to retrieve these data in the raw data of the individual labs, and I found for both lab Z and Y (Appendix 8, page A-2 and Appendix 7, page B-9 respectively), that the percent of radioligand bound at 0.03 M radioinert estradiol was indicated in a column named "*10 Percent rule*", and the criteria was obviously that the percent of radioligand bound should be below 10%. This is not in agreement with the protocol (which requires for 25 - 35 %).	During the second interlaboratory validation study, the recommendation was that the percent of radioligand bound should not significantly exceed 10% of the total added. This recommendation proved difficult to meet, and given the alternative of reducing the amount of receptor added to the tube, which would reduce the signal, the Agency decided to relax the limit somewhat and to rely on the Swillens correction for ligand depletion when estimating K_d and B_{max}.

22. Clarify that the ratio of total binding of radioligand in the absence of competitor to the total amount of radioligand added to the tube must be reported for the competitive binding assay. Also, the target range for this ratio should be the same for the competitive binding assay as for the saturation binding assay.			
10.22.1	MLS	In Appendix 1, Page 47, 10.7.3. "*Performance criteria for the competitive binding assay*", it is indicated that: "*Ligand depletion is minimal. Specifically, the ratio of total binding in the absence of competitor to the total amount of [^{3}H]-17β-estradiol added per assay tube is no greater than approximately 15%*". Again here, it is not asked to the labs to report this ratio, but only the amount of protein added per tube. It should be asked to the lab to report this parameter too. Again here, lab Y used the below "*10% rule*", and generally, the percent of radioligand bound was around 3%, which I find too low (Appendix 7, page D-3). I was not able to find the information in the report of lab X. Furthermore, I see here a contradiction with Appendix 1, page 20, 9.1.5. where it is said that "*For the saturation assay, the*	The protocol has been changed to request reporting of the ratio of total radioligand bound in the absence of competitor to total radioligand added, and an explanation has been added to the protocol for the difference between the target ranges for the saturation binding assay vs. the competitive binding assay.

		optimal protein concentration binds 25 - 35% of the total radiolabeled estradiol that has been added to the tube". This would mean that a different protein concentration should be used for saturation binding assays (25 - 35% binding of the total radioligand added) and for competitive binding assays (15% binding of the total radioligand added), which is not what is done in practice. Please, clarify that point in the protocol.	
23. Conduct a third interlaboratory validation study that reflects the changes made to the protocol since the second validation study, and use the same chemicals that are tested when validating human recombinant ER-based assays.			
10.23.1	MLS	<u>In a validation study, the protocol given to the participating laboratories should be a definitive one. As several changes have been made in the protocol after the second interlaboratory study (we see that by comparing the modified protocol described in Appendix 1 and the original protocol described in Appendix 5), it should be relevant to organize a third interlaboratory study using (and reviewed) a definitive protocol.</u>	The Agency recognizes that the ER-RUC assay is sensitive to small changes in techniques, and that ideally a validation study would be conducted only on the final protocol. However, the Agency also recognizes that the time it takes to conduct such a validation study would delay the Screening Program significantly. In its judgment, the modifications to the protocol are not of such magnitude that a new validation study is required.
10.23.2	MLS	If a third interlaboratory study is organized, it would be interesting to analyze the same list of chemicals with both ER-RUC and recombinant ER binding assays. The	The Agency notes that the same test chemicals were used in the second interlaboratory validation study for the ER-RUC assay and the validation study for the hrERα assay which is currently underway (with the

		comparison could provide different conclusions that could add some weight to the strength of the ER-RUC using both α and β isoforms of ER compared to a recombinant ER binding assay using the α isoform only.	exception of the negative control chemical).
24. Correct the performance criteria table for the lower limit of the top plateau for estradiol.			
10.24.1	MLS	Appendix 1, page 48, table 9 : the lower limit of the top plateau for estradiol is 90% and not 94%.	As noted in the response to comment 5.6, the performance criteria going into the second interlaboratory validation study and subsequent to that study are different. The 94% value is the correct value.
25. Provide references for the historical RBAs cited in the ISR.			
10.25.1	MLS	In ISR, page 52, table 21, page 52, references of "*historical*" RBAs should be given.	The reference for the values cited is given at the bottom of the table. The Agency did not check the primary references cited in the ICCVAM document.
26. Describe how the weight-of-evidence decision will be made for a test chemical when all of the information from Tier 1 assays is available.			
10.26.1	MLS	The introduction clearly indicates that the ER-RUC assay is one of the complementary assays included in Tier-1, listed in table 1, page 3. In page 1, it is indicated that "*A negative result in Tier 1 would be sufficient to put a chemical aside*	The Agency recognizes that there is the potential for apparently non-concordant results between the several assays in the Tier 1 Battery that are relevant to the estrogen hormone system. Interpretation of such results will depend on the particular circumstances of the studies. For example, were each

		as having low to no potential to cause endocrine disruption, whereas a positive result would require further testing in Tier 2 ". And in page 7, we read : "*An individual assay may serve to strengthen the weight of evidence in a determination (e.g., positive results in an ER binding assay in conjunction with positive results in the uterotropic and pubertal female assays would provide a consistent signal for estrogenicity)".* However, it is not clear, how is taken the final decision after Tier-1 (negative result or positive result), if all the complementary assays of the Tier-1 battery don't give concordant results (for example positive results in the ER binding assay but negative result in the in vitro transcriptional assay). In page 2 of the ISR, there is a mention about "*false negatives*" and "*false positives*" resulting from the Tier-1 screening in the sentence "*Maximum sensitivity to minimize false negatives while permitting an as yet undetermined, but acceptable, level of false positives*". It would be necessary to explain how a positive or a negative decision is taken after the Tier-1 screening, or to mention a reference giving this explanation.	of the studies well-conducted with little variability between replicates or were there substantial differences in reliability between studies? Also, these Tier 1 assays do not give sufficient information to determine precise mechanisms of action of a test chemical, so there may be instances where apparent inconsistencies may have a plausible explanation. For example, since the transcriptional activation assay integrates the results of more intracellular processes than simple binding to the receptor, it is possible that a chemical could be positive in the binding assay but negative in the transcriptional activation assay. A complete description of how all of the details available to the Agency will be combined into a weight-of-evidence determination is not feasible. See also the response to comment 1.4.