Peer Review Panel Comments on the Aromatase Assay

Comment	Commenter	Comment	EPA Response
	Topic: Purpose and relevance of the assay.		
1	LK, MM, SM, TS	The purpose of the assay is well presented and clear.	No response needed.
2	SB	A full understanding must be assembled from several different locations in the ISR and the entire significance of the assay must be constructed from those disconnected pieces of information. Because the purpose of the assay and significance could be lost on even a "non-expert" scientist; suggestions for additional clarification of this issue would be to include a succinct statement in the Executive Summary (prior to the background) indicating the purpose of the assay is to identify compounds capable of influencing aromatase, a key regulatory enzyme involved in androgen/estrogen metabolism and biosynthesis which is believed to be an important regulator of hormone action in some hormonally sensitive tissues throughout life of both males and females.	Other reviewers did not seem to experience this problem, noting that the purpose of the assay was clearly stated. However, the Executive Summary will be revised to improve the statement for the purpose of the assay.
3	All	As an individual component of the Tier 1 screening battery, the aromatase assay is determined to be biologically relevant for the purpose of identifying environmental endocrine disruptors (EDC) that may act via inhibition of aromatase activity. The importance of the aromatase enzyme for estrogen formation and function in the mammalian organism is well reviewed in the ISR.	No response needed.
4	SB, SM	The direct toxicological relevance of the assay is limited. The aromatase assay assesses an influence on a relevant enzyme activity that could potentially impact the metabolism of androgens and the synthesis of estrogens. However the assay does not detect a toxicological endpoint. As a result, the toxicological impact of aromatase inhibitors is implicit and would require additional specific	EPA recognizes the limitations and advantages of this in vitro assay. However, since the assay is an inexpensive, rapid in vitro method to detect chemicals that inhibit aromatase, an enzyme responsible for the conversion

		toxicological assessment. As a component of the Tier I battery the ability to reliably identify candidate EDCs for assessment of endocrine disruptive toxicity in vivo is critical, and thus the aromatase activity is felt an essential component of an integrated assessment of EDC actions and toxicity.	of androgens to estrogens, it is conducive to screening large numbers of chemicals for their potential effects on steroidogenesis. As part of the Tier I Screening Screening Battery, this assay will complement data from three in vivo assays by providing valuable mechanistic information.). As noted by EPA and concurred by a majority of peer reviewers, the assay is fulfilling the role that is anticipated within the T1S Battery.
5	TS	Estrogens, particularly in woman are not only available from the conversion of androgens and all its steroid precursors by aromatase (and all its precursor enzymes). Estrogens are also present as a pool in the form of (post-aromatase) estrogen-sulfates (eg in the mammary gland), which under conditions of reduced estrogen levels may be converted to free estrogen by sulfatases (Pasqualini, 2004). No consideration for this is given in the documentations provided and one should be cautioned that dependent on the tissue of interest a modest degree of aromatase inhibition may have relatively little affect on steady-state estrogen levels if compensatory release by estrogen/aromatic sulfatases occurs.	The aromatase assay, along with three in vivo assays, will be used to detect chemicals that alter steroidogenesis in the Tier 1 Screening Battery. The in vitro aromatase assay is included to provide mechanistic information (e.g., an inhibition of aromatase activity). It was selected for use in the T1S Battery because (1) it has been well documented that environmental chemicals can alter steroidogenesis and ultimately, reproductive function, via this mode of action; and (2) the assay is conducive to rapidly screening large numbers of chemicals. The three in vivo assays are designed to cast a broader net by taking into account absorption, metabolism, distribution and excretion of the test chemical, as well as to detect chemicals that disrupt steroidogenesis through a

			variety of MOAs including the sulfatase pathway.
	Topic: Protocol		
6	All	The protocol does a good job of describing the assays and is generally clear allowing the reader to conduct the assay	No response needed.
7	SB, MM	Some editorial corrections /clarifications to the protocol are needed as follows: In addition to he provided information, information regarding the stability of each chemical should be provided. 2.1.1 μCi/mmol should be mCi/mmol 2.1.3 The buffer is used to make the stock solutions should be clarified. What does "record the weight of each component added refer to?" 2.4.1.2 Bullet 3 and 5: wash volume should be specified Bullet 6: guidelines for volume of butter for resuspending pellet should be specified. Bullet 7: specific guidelines regarding aliquot volumes and minimal acceptable stock protein concentrations should be specified. 2.4.1.3 Timing for use of microsomes should be defined rather than recommended. Specific information regarding the staring amounts of placental tissue and expected ranges of microsomal yield would be helpful. Why is propylene glycol added to the assay? 4.0 Second bullet: "are presented in Table 3" should be corrected to "…Table 2." Check on tenses. 5.0 The minimum activity level for the placental microsomes in 0.03 nmol/min/mg protein. 6.0 The reference to Table 6 should read Table 5.	**EPA will make these revisions to the final protocol.**
8	SB	There should be clear criteria or directions for the preparation of stock formulations. The options for chemical formulation in "buffer, absolute ethanol, or DMSO" is problematic.	The aromatase assay will be used to screen a wide range of structurally diverse chemicals. As such, the solubility

			of these chemicals in the assay buffer will vary depending upon physical properties, and will require some flexibility as to the solvent used (e.g, water, ethanol, DMSO). The final protocol will be revised to include a procedure for selecting the appropriate solvent for each chemical, and will provide limits for the maximum concentration of each solvent that can be successfully used in the assay.
9	SB	There was no justification identified for using ethanol as the solvent of 4-OH ASDN. Without justification, the use of ethanol as a solvent for the positive control for the entire aromatase assay stands out as atypical and arbitrary. This fact can be readily seen in Table 7.4-1 (page 46) where dimethyl sulfoxide (DMSO) was used for 9 of 11 test substances, with only 4-OH ASDN and ketoconazole prepared in ethanol solvent.	Since steroids are readily soluble in ethanol, it was selected as the solvent of choice for 4-OH-ASDN (see response to Quiery #8). The revised protocol will clarify the need to use different solvents depending upon the physical/chemical properties of a test chemical. In addition, the revised protocol will require that full aromatase activity be determined for every solvent used.
10	SB	There is a lack of a negative vehicle control within the proposed assay test groups, or at least it is difficult to find an explicit statement in the ISR that adequately describes how test chemical vehicle effects will be assessed. There is additional confusion created because in some places it is implied that the "Full Enzyme Activity Control", as described in Table 4.6-2 (pg 28), is considered a proper negative control group. In Table 4.6-2, The Full Enzyme Activity group is described as the ***complete assay components plus inhibitor vehicle***, but it is not indicated clearly there, or in the corresponding text, whether this is positive control vehicle or test chemical vehicle. Further, this	There were three different control groups used consistently throughout the assays. The full activity control is the test system without any inhibitor: the reaction substrate, enzyme, and all other assay components and, thus, gives 100% activity. I believe the confusion is over the wording "inhibitor vehicle control" which means the solvent that the inhibitor or test compound is dissolved in is added

		descriptor is not included in Table 4 of the protocol (Appendix A-13). It should be clarified whether this group is a negative control for the positive control inhibitor, 4-OH ASDN vehicle (e.g. absolute ethanol) or test chemical vehicle (test inhibitor). While unclear, it was interpreted to indicate the 4-OH ASDN ethanol vehicle. As a result there is no true negative control for test chemical vehicle.	to this control group. The Full Enzyme Activity Control functions as the negative control. **This will be clarified in the revised protocol**
11	SB	There is concern related to human genetic variation, which is not addressed at all in the ISR. To date, it appears that only two or three different placental preparations were used for validation of the Aromatase Assay (the recombinant system represents a single CYP19 variant). It is well known that there are numerous variants and haplotypes of the CYP19 gene, some of which have been linked to changes in hormonal levels and endometrial cancer for example (for a review see Olson et al., 2006). Thus, there is much evidence for a high level of variation in CYP19, and its resulting aromatase activity. The anticipated variation representative of human populations is not acknowledged in the ISR, and the fact that the aromatase assay is unable to inform on normal human variation is lacking. While using a single preparation of microsomes from a single individual to assay a number of different compounds as inhibitors of aromatase activity is considered scientifically acceptable, it is felt critical that the potential for normal genetic variation to impact (limit) the conclusions possible from results obtained with the aromatase assay should addressed.	EPA recognizes genetic variation in the CYP19 gene may occur among human placental microsomes. To date there are no data to demonstrate that this would impact the detection of chemicals that inhibit aromatase activity when using human placental microsomes as the source of enzyme. In general, tissue- and species-differences have only been observed for the induction of CYP19 gene expression. The reviewer's recommendation to collect genomic information from each placental microsomal preparation is excessive, and would be beyond the Agency's needs for a semi-quantitative screen to identify inhibitors of aromatase. . Nevertheless, the final protocol for the aromatase assay will use human recombinant microsomes as the source of enzyme and thus will provide a standardized genetic composition.
12	SB	In light of the demonstrated rapid decrease in aromatase activity during the short period of time required to prepare samples and run an individual assay, the practice of storing microsomes in multiple use	The protocol requires that microsomes be stored in aliquots, and to never be thawed and refrozen.

		stocks is strongly discouraged. It is suggested that microsomal suspensions be stored as single use aliquots. This alternative is supported in section 2.4.1.3 (sentence 1, pg. A-10) of the ISR which discourages the practice of refreezing, and suggests dividing into aliquots following initial freeze/thaw cycle. Because of the acknowledged loss of aromatase activity, it seems most reasonable to initially divide the preparation into single use aliquots and not allow re-freezing. Section 2.4.1.3 (sentence 1, pg. A-10) could then be deleted from the protocol. If single use aliquots are not used, a maximum number of allowed freeze-thaw cycles for each stock aliquot should be determined experimentally and specified.	
13	TS	It would be useful to have a better description of the type of quench correction used to convert cpms to dpms. How was the quench curve prepared? What were the counting settings?	The revised protocol will require that the methods used to determine counting efficiency and correct for quenching be reported.
14	SB	As described in section 7.1 (A-15-16), the compilation of data is well described with the exception of the transformation to percent control. It should be noted that %-control values for each of the replicates (including the Full Activity controls) are to be calculated and the mean %-control of the replicates calculated. In this way the variance of the full activity controls of the test run are properly retained for the experiment(s).	No response needed.
15	SB, LK, MM	The approaches used for model fitting are reasonable, straight-forward and applicable to most cases when a full and classical concentration response (sigmoidal) is observed.	No response needed.
16	TS	The relevance of doing a Hill Plot analysis (usually applied in receptor binding studies) could be explained more clearly, as well as the meaning of deviations from a slope of -1. If a test chemical inhibits aromatase with a Hill plot of -2.0, what would that mean? Some inhibitors are known to inhibit competitively as well as allosterically/non competitively….these situations should be explained	**This will be clarified in the data interpretation section of the final protocol.**

		and included as part of the 'assay package'	
17	SB	The reporting of results is felt to be poorly described. Section 10.0 (A-18) of the protocol is extremely general, and must be made much more specific. Importantly, Table 6. Data Interpretation Criteria is not referred to at all in the protocol.	**This will be clarified in the revised protocol.**
18	LK	It is possible that the enzyme reaction product estrone may be further metabolized to another component that may not be detectable using RIA. Estrone concentration in solution is dependent upon the redox state. Under reduced conditions (this assay) it converts to estradiol. Ideally, you would want the estrone product converted to its reduced form as estradiol, because that eliminates end product inhibition and helps to drive the enzyme reaction with mass action effect. Be aware of oxidation and keep tubes capped	EPA concurs with the reviewers on that the human placental microsomes retain some ability to convert estrone to estradiol and further metabolic by-products. Thus, the radioimmunoassay used to measure product (estrone) was not an appropriate comparison with the 3H- water produced. This is not a major issue when using the human recombinant microsomes since the concentration of contaminating P450s is minimal to non-existant.
	TS	The observation that the tritiated water-release assays produces aromatase activities (amounts of ^3H_2O) that are three times higher than aromatase activities based on the measurement of the formation of the product estrone is likely explained by the presence of 17-beta hydroxysteroid dehydrogenase (17HSD). This enzyme is highly expressed in placenta and is present as two subtypes, 1 and 2. 17HSD1 is NADPH dependent, converts estrone to estradiol and is very likely to be responsible for the apparent loss of estrone from the reaction medium. 17HSD2 converts estradiol back to estrone, but is dependent on NADH which is not added to the reaction medium (Vihko *et al.*, 2003; Mindnich *et al.*, 2004) Also, in theory, tritiated water release could also be due to other reactions than aromatization, such as 1-beta-hydroxylation of the tritiated substrate. In rat liver microsomes this is known to occur by the enzymes CYP3A1 and 2B1 (Waxman, 1988).	
19	LK	Regarding the effect of more enzyme activity at the beginning vs end: it is likely due to starting the reaction with the pipetting of microsomes and stopping with quenching or transfer to cold. The speed is faster	The revised protocol will emphasize the need to be attentive to timing and temperature while conducting the assay.

		with stop procedure as compared to reaction start. Also, the last microsomes pipettted may be cooler in temperature than the initial aliquot pipetted. Technician needs to pay attention to timing and temperature.	In addition, the maximum number of tubes, as well as a time limit will be recommended.
20	MM	If a test substance causes inhibition that is classified as equivocal and there are no solubility or enzyme denaturation limitations, it could be recommended that the assay be repeated at higher dose levels so that an IC50 can be obtained from data that reflect the full dose response curve.	See response to comment number 44.
21	SM	In order to improve protocol, the following advice is proposed : 1. Add the substrate androstenedione 4 uM during microsomes preparation to preserve active site of aromatase. This showed, by experience, to increase aromatase half life during storage and ameliorate its stability during the assay. This may reduce the significant difference observed in enzymatic activity of control between the beginning and the end of assay but also after repeated freeze-thaw cycles of microsomes. 2. On the day of use, microsomes should be thawed at 4°C instead of 37°C in order to avoid the thermal shock which could provoke a denaturation of proteins in general and aromatase particularly. 3. The three fold extraction by chloroform or by methylene chloride (be sure to use one of these two solvent in the final report) is useful when solvent is recovered and an analysis of estrogens formed is realized in parallel with the formation of tritiated water during assay validation. However, for the routine work, extraction could be made by chloroform followed by an extraction by charcoal/dextran mixture (7: 1.5%) instead of two supplementary extractions by solvent, this help to reduce the time of experimentation. 4. Add the formula for the calculation of the specific aromatase	Suggestions for improvements will be included in the revised protocol. Two options will be included for the extraction of 3H-water: (1) the original method using methylene chloride; and (2) chloroform/charcoal method.

		activity in nmol.mg protein^{-1}. min^{-1} by expressing all parameters used such as, background radioactivity,	
22	TS	The protocol states that solvent concentrations for the test chemical should not exceed 1%. Dependent on the type of solvent used I would argue that this may be on the high side for solvents such as DMSO which one commonly wants to keep in the 0.1-0.5% range. Also a concentration of 5% propylene glycol is already present.	The revise protocol will include recommendations for limit concentrations of DMSO (0.1 – 0.5%) and ethanol (1.0%). Solvent controls will be required for each experiment to determine full aromatase activity.
23	TS	Microsomes are finally frozen in a resuspension buffer containing 0.25 M sucrose, 20% glycerol and 0.05 mM dithiothreitol. A protocol that uses only 0.25 M sucrose is also commonly used and microsomes prepared in such a manner are stable at -80^{o}C for up to 3 years. Has the necessity of the glycerol and dithiotreitol (which are supposed stabilizing factors) been investigated, and has the influence of these components on the catalytic activity of aromatase and the potency of its inhibitors been studied?	**EPA is following up on this comment.**
24	TS	Is the rehomogenization step really necessary? Generally microsomes are briefly vortexed prior to conducting an enzyme assay, and pottering may introduce unnecessary additional degradation of protein	This step was added to the protocol by the lead contract lab to ensure homogeneity of the microsomes. However, this is not a necessary step for use with the human recombinant microsomes. This will be clarified in the revised protocol.
25	TS	The term extrapolation is used under section 3.1 (page A-11). This suggests that protein concentration are determined by extrapolating the protein standard curve which should never be done. It would be more correct to use the term 'read' from the standard curve or 'superposed' onto the standard curve to avoid the impression that the protein sample reading falls outside the obtained standard curve.	The ISR will be revised as suggested.
26	TS	The aromatase assay as described is performed in test tubes. I would have thought that the assay could easily be down-scaled to far smaller	Although the assay was validated in a large-scale format, it can be adapted for

		volumes (Sanderson *et al.*, 2000), so that the assay could be performed in multi-well plates (incubation step) and 1.5 ml eppendorf vials (extraction steps) and ultimately using 4 ml liquid scintillation tubes. This would dramatically reduce cost and the amount of waste produced	use with a smaller volume in multi-well plates. be run as EPA will provide the option to do so as long as performance criteria are met.
27	TS	Why was the tritiated water-release protocol altered from its original (Lephart and Simpson, 1991) by extracting 3x with methylene chloride instead of 1x chloroform followed by clean-up 1x with dextran-coated charcoal solution? Throughout the documentation I was not able to find a rationale for this decision. The original method would appear more efficient as it uses less solvent and fewer steps. Also, the use of dextran-coated charcoal aides the removal of traces of solvent in the aqueous phase, which is important as chloroform is a potent quencher. As methylene chloride is also a strong quencher of weak beta-emitters such a tritium, I am wondering if quenching was ever a problem in the performance of the experiments. I could not find this information in the documents. Despite the above comments, it nevertheless appears that the changes to the original protocol did not deleteriously affect the assay	The assay may be conducted using either methylene chloride or chloroform/charcoal for the extraction of the 3H-water. Since the use of one extraction method versus the other may depend on safety rules/regulations for a given laboratory, the revised protocol will include both as options for the assay.
28	MM	While an HPLC method is described to establish radiochemical purity the frequency of purity check is not indicated. Tritium exchanges with water and if this occurs to a significant extent, background control activities would increase as tritiated water will not be extracted by methylene chloride. A recommended time for re-analysis could be suggested.	Per the recommendation of the vendor, the purity of the radioactive substrate should be confirmed every 6 months. This recommendation will be included in the revised protocol.
29	TS	Is the addition of propylene glycol necessary? It increases the organic solvent burden of the reaction mixture disproportionally compared with all the other components including solvent used for test chemicals and may not be essential to the performance of microsomal enzyme assays.	**EPA is following up on this comment.**
30	TS	Semantically it is more appropriate to express the catalytic activity of aromatase when determined using the tritiated water assay as *pmoles of*	**The revised protocol will be corrected to reflect this recommendation.**

		androstenedione converted per time unit per quantity of protein, rather than amount of *estrone* formed, because estrone is not measured	
31	TS	A mid-log concentration would be e.g. $10^{-3.5}$, not $10^{-3.3}$ as suggested in section 6.0 on page A-14. Given that the inhibition curves are plotted as log-concentrations it makes sense to choose concentrations as follows: 0.1, 0.3, 1.0, 3.0, 10 etc. micromolar as these points will be equidistant in the concentration-response curves and other analysis.	**The revised protocol will be corrected to reflect this recommendation**
	Topic: Strengths and limitations of the assay		
32	LK,MM, TS	The strengths and limitations of the assay are adequately addressed in the ISR and the DRP.	No response needed.
33	SB	The strengths of the assay are considered inadequately addressed. The entire body of work represented in the ISR confirms the many strengths and reliability of the aromatase assay and it is felt that the Executive Summary should contain a section specifically dedicated to summarizing the assay's strengths. The final paragraph of section 3.3 (pg. 22) which consists of a stand-alone sentence is a ***strong statement of opinion*** that is considered not well supported. It is considered unnecessary and should be deleted	**A statement summarizing the strengths and limitations of the assay will be placed in the conclusions section of the Executive Summary.**
34	SB	The fact that the assay, as described, is limited in its ability to assess the effect of chemicals on only a single variant of aromatase is not discussed. This is felt to be a significant omission (please see comments regarding known Cyp19 variation above). Further, the fact that the aromatase inhibitory dose-response properties of a chemical are likely different in some individuals and populations is a point that should be addressed as a significant limitation of the assay. As a result of that initial limitation, the specific data/conclusions obtained using the aromatase assay can not be generalized to any individual or specific human population. This lack of generalizability of assay results should also be addressed.	As noted in one of Dr. Sanderson's comments, inhibition-wise species-differences and tissue-differences in response to aromatase inhibitors are relatively small. When it comes to potential induction, differences among species, tissues and even times of year (especially in fish, frogs, birds), are qualitatively and quantitatively very different. These key issues are of great importance to our concern about environmental endocrine disruptors and require urgent attention and considerable

			additional research in order to develop the key bioassays suitable for the identification of endocrine disruptors that act via the disruption (induction) of the aromatase enzyme in human and wildlife tissues.
35	MM	Another limitation of the assay that is addressed at various places in the document but not specifically mentioned in the limitations section is the longevity of activity in the enzyme preparation. It is well known that cytochrome P450 metabolism in microsomal preparations declines with time and loss of aromatase activity occurs with time. The nearly consistent decline in aromatase activity in the samples at the end of the run compared to the beginning suggest some activity loss with time. However, although this was reported to be statistically significant in most of the runs, the effect was sufficiently small to minimally affect the data. However, it could be recommended that assays be conducted within a defined (eg 2 hour) time frame (Page A-9) and the importance of timing could be more strongly emphasized for all points in the assay: tissue preparation, time on ice, pre-incubation etc to minimize variability.	The revised protocol will emphasize the need to be attentive to timing and temperature while conducting the assay. In addition, the maximum number of tubes, as well as a time limit will be recommended.
36	TS	By far the greatest weakness of the placental microsomal aromatase assay is its limitation to only be able to detect inhibitors of aromatase. Lacking, however, is a thorough discussion of the implications of this constraint on the validity not so much of the assay as a technique *per se*, but of the relevance as tool to determine affects on aromatase when only one half of the picture can be investigated. It is comparable to wanting an assay for potential interferences with the function of the estrogen receptor, but then proceeding to develop an assay that can only detect antagonists. There are potential assays described in the	EPA has recognized this limitation and is currently validating the H295R cell-based assay to identify inhibitors of aromatase activity in addition to chemicals that induce aromatase (CYP 19) mRNA that results in increased aromatase activity.

		literature that would be equally suitable as tools for screening inhibitors but would also have the added possibility to detect inducers.	
	Topic: Choice of test substances and analytical methods to demonstrate assay performance		
37	All	The wide range of test substances, analytical methods and statistical methods chosen are appropriate for appraisal of assay performance.	No response needed.
38	SM	Although IC50 is appropriate, it is generally preferable to evaluate Ki which is the dissociation constant of enzyme/inhibitor complex and reflect enzyme affinity for inhibitor (or chemical) and gives more precision than IC50 value. IC50 value, showing inhibition efficiency, is used because it is more rapid and easy to perform. Thus, comparison between constants (Ki values) of different laboratories is easier to make than comparison of IC50 values (see page 60 for the variability of IC50 values between ISR and literature for ketoconazol and econazole)	EPA agrees that the Ki is preferable for comparison of potencies. However, Ki determination is a more complicated procedure and is not necessary for a screen to identify of aromatase inhibitors.
	Topic: Reproducibility of results		
39	All	The assay is considered to be sufficiently repeatable and reproducible and the results are consistent with the variability inherent in bioanalytic assays. Despite interlaboratory variability , the IC50 values were generally similar.	No response needed.
40	MM, SB	The single most important source of experimental variability appears to be associated with accurately determining the concentration of microsomal protein (as a surrogate estimate indicator of active aromatase protein) present in each assay.	Within a single run of the assay, all tubes contain the same volume of microsomes obtained from a common preparation. Thus, within any given experiment (e.g., run) all tubes contain the same protein concentration and are directly comparable. .
	SB	The modified Lowry assay used for determining protein concentration is well-known as rather inaccurate and variable. As a result, the fact	As indicated in Query # 40, each experiment (e.g., run) uses the same

		that estimation of microsomal protein concentration is the major source of assay variability is not surprising. Some analysis was performed using Cytochrome P450 spectral analysis as an additional relative measure for normalization of microsomal proteins. Because of the critical effects that inaccurately determining the amounts of active aromatase protein present in each assay can have has upon the results, further consideration of complimenting the total protein concentration determination with Cytochrome P450 spectral analysis (or another accurate surrogate assessment for active aromatase) is encouraged	volume per tube from a common microsomal preparation. Thus, all tubes within each experiment are directly comparable. While accurate measurement of protein concentration for each microsomal preparation is necessary for comparison between different batches of microsomes, EPA does not believe that it is necessary to measure both protein concentration and CYP P450s spectral analysis.
41	SB	It is notable that throughout the study of inter-lab variation (Section 8) the overall task means comparing inter-lab variation are calculated in an unacceptable fashion that greatly reduces the CV%. This point is demonstrated by the included supplemental information (**review pg. 13**) where the "overall task means" and their associated variance are recalculated in two ways (mean of all replicate assays vs mean of the mean) for Fig. 8.2-2 of the ISR. By taking the simple mean of the mean values reported by each lab, the variance associated with each observation (replicate) is disregarded. It is strongly suggested that the data also be presented in a fully transparent fashion that takes all data points into consideration. This will avoid any suggestion that attempts were made to minimize the apparent variability of the assay.	An important objective of the EDSP is to develop assay systems that provide reproducible results among laboratories. Thus a principal objective of the inter-laboratory statistical analyses is to evaluate the variability of reported results among laboratories. This is important because it predicts how reproducible the assay results will be among different laboratories after the assay is certified for production, when submissions are made from a variety of laboratories. **[** Each laboratory will report a mean value across replicate determinations and an associated within-laboratory standard error. Thus the among-laboratory coefficient of variation, which reflects the extent of variation in average results among laboratories, is the physically appropriate most measure of the

			variability of the assay among laboratories for the purposes of the EDSP.
	Topic: Performance criteria		
42	SB, MM, TS	The performance criteria for the assay are generally considered appropriate and reasonable based on the presented data. The performance criteria are based on common sense and practice and should ensure that the assay is functioning properly.	No response needed.
43	SB	A specific method and criterion for outliers should be established.	The Prism program v5 has an outlier exclusion feature. This was used in the EPA analysis.
	MM	No specific performance criteria were established for outlier data although test laboratories were cautioned to evaluate data for experimental error. In the hands of experienced personnel this should be sufficient. However, aberrant or outlier data should be examined closely.	
44	SB	Regarding difficulties with the dose/response curves fits of the compounds (nitrofen, BPA) that were found not to be well described by the models – the concentration-response data does not reach a high concentration plateau and are likely incomplete – the shape of the curve is not easily described by a fit of the data with any equation describing a sigmoid. Data should be inspected for both high and low dose completion. If a clear plateau in response is not observed, additional data for higher (if possible) or lower concentrations should be collected and incorporated into the D/R curves.	EPA has specified that the assay be run first of all at a limit dose of 1 mM which was selected based on the recommendations of an expert panel for binding assays. This has been criticized by some as being too high to be of physiological significance and EPA does not intend to require testing at higher doses. To better define the lower portion of the curve, the protocol requires testing at half-log lower concentrations if solubility problems are encountered at the limit dose. **EPA will add a note to the data interpretation portion of the protocol regarding denaturation at high concentrations.**
	TS	One aspect of concern is the relevance of testing concentrations as high as 1 mM. It is reassuring that there has been considerable discussion and awareness in the documentation, including the ISR, concerning solubility problems, surfactant issues (eg. nonylphenol). It is important to keep in mind that a decrease in enzyme activity, particularly at excessively high concentrations, may be due to such artifacts as mentioned above. In fact, the use of microsomal fractions	

		or purified enzyme (supersomes) tends to invite the temptation to test compounds at concentrations well beyond any true biologically relevant exposures. The question still remains whether the protocol in its present form will be able to identify such artifacts as enzyme denaturation under all circumstances. An experiment with a surfactant such as triton X, for example, may provide a 'typical' denaturation-induced inhibition curve that could pose as a template for other compounds with unknown mechanisms of action. In any case, continued awareness of possible artefactual inhibitory effects when interpreting the proposed bioassay is essential	
	Topic: Data interpretation		
45	LK, MM, TS	Data interpretation criteria for classification as an inhibitor uses a simple cut off approach of achieving more than 50% inhibition for an inhibitor and above 75% inhibition for a non-inhibitor. This is a useful approach and allows easy classification of inhibitors and non-inhibitors	No response needed.
46	MM	The equivocal situation where inhibition is 50-75% is not adequately addressed. None of the tested chemicals fell into this category and additional testing approaches are not suggested. In addition, a 4-parameter regression model is proposed to describe the inhibitory effect of the test chemicals yet if the data do not fit the model then the default is to use the average activity of data points collected at the highest concentration. This latter approach makes the more sophisticated software based analysis of concentration dependent inhibition of the enzyme appear redundant. If the highest concentration data points are to be used, there is a greater possibility that enzyme denaturation rather than enzyme inhibition has occurred. The limitations of this default approach should be addressed.	**The data interpretation criteria were revised such that failure to fit the model results in a classification of a chemical as a non-inhibitor.**
	SB	The utility of including the equivocal designation of the inhibitors is not established. Using the prescribed performance criteria and the sigmoidal curve-fitting models, it is unclear whether or not identification of a chemical that acts in an "equivocal" fashion is	The equivocal designation was for chemicals whose behavior could not be clearly discerned as being inhibitors or not. It was suggested by the ICCVAM

		possible. It might be considered useful to computationally model an equivocal-type curve to determine whether the assay performance and analysis criteria even allow equivocal-type identifications. At first blush it seems that such concentration-response relationships might be identified as "failures-to-fit". If this is the case, the equivocal-type category should be eliminated.	expert panel for binding assays which are modeled with the same equation as competitive inhibition. Nitrofen fit the curve well ($R^2 > 0.98$), but it exhibited only a partial curve. However, it did meet the binding criteria. Although partial curves falling short of 50% activity were not seen in the aromatase assay, these were seen in the binding assays. With such curves, EPA was reluctant to extrapolate to an IC50.
47	TS	The example given in table 11.3-1 of the ISR suggests to me that the 95% confidence interval approach is the better approach, although more involved. The discrepancy for dicofol is readily explained in the text, but the discrepancy for genistein occurs only in the best curve fit approach, the 95% CI approach is consistent. Genistein has been investigated on a very detailed level, including various molecular modeling studies which demonstrate that isoflavones (genistein), unlike flavones (chrysin, apigenin) are, due to their stereoisomeric conformation, incapable of interacting with the heme moiety of aromatase to cause aromatase inhibition (Kao *et al.*, 1998). Ironically, and this is a major limitation of the currently presented bioassay, genistein (for example) is a relatively potent inhibitor of tyrosine kinase and phosphodiesterase, the latter effect causing increased gene expression of CYP19 (aromatase) in tissues where its expression is under control of the cAMP-driven pII or I.3 promoters (Sanderson *et al.*, 2004). The microsomal assay as proposed categorizes genistein, together with other (*in vitro*) inducers of aromatase, such as atrazine and vinclozolin as negative, whereas in reality they have an inductive effect on the endpoint (catalytic activity of aromatase) in question, at least in certain systems. This is could be misleading to the regulators	**EPA will investigate this further** but still believes that without adequate controls on the variability of data from test compounds, high variability would result in too may chemicals being classified as equivocal rather than as inhbitors.

		that will be interpreting the aromatase assay results.	
	Topic: Utility of the assay		
48	All	This in vitro aromatase assay meets the criteria for a screening tool to identify chemicals that may potentially interact with the endocrine system via inhibition of aromatase. Assay is robust, has a reasonable level of reproducibility, and is a relatively quick and inexpensive screen for an inhibitory effect of a test chemical on aromatase activity. It should be noted that this is a very specific assay carried out *in vitro*, and potential *in vivo* effects on aromatase (eg enzyme induction) would not be detected with this methodology. It should be a first step in the evaluation process, because of its ease of use, short time course and overall safety and cost.	No response needed.
49	LK	I do recommend that the CYP 19 recombinant microsomes (SUPERSOMES) be used preferentially (reasons stated in prior sections), but the placental microsomes are a good alternative in situations where the purchase of SUPERSOMES is prohibitive and placentas are plentiful. The recombinant enzyme preparation was more comparable across labs.	As stated in the ISR, EPA believes that the advantages of the recombinant assay make it the obvious choice.
50	TS	As noted before, an assay that detects both inhibition and induction would be more useful. This assay may become outdated in the near term.	EPA agrees and is validating the H295R assay.
51	SM	As cited in IRS, one of the weakness of the proposed protocol that it can not predict metabolizing chemicals and formation of metabolites which could react with aromatase differently than original substance. In the ISR, Lindane is reported as negative chemical with both microsomal systems while it inhibits aromatase in JEG3 cells (Nativelle-Serpentini et al, 2003). So, a false negative should not systematically be deleted from the next step of evaluation.	Several in vivo assays will be included in the battery to address such issues as metabolism.
52	SM	Androstenedione is one of the aromatase substrate (others are : 16α-hydroxytestosterone, testosterone, 19-norandrogens) and considered as the preferential one in human. However, in some species, other	To date there are no data to indicate that the choice of substrate for this assay is not appropriate for use with microsomes

		substrates being used preferentially by aromatase as we previously showed that 19-norandrogens are aromatized at least at the same efficiency as androgens by equine aromatase (Moslemi et al, New York Academy of Sciences, 1998). So, the use of androstenedione in the proposed protocol should be specified for human and could not be representative of all species	prepared from human , rodent or fish tissues. In general, the inhibition of aromatase activity by a given chemical has been universal when using microsomes from each of the above listed sources, including human recombinant microsomes. Thus, this approach still stands as a valuable tool for screening large numbers of chemicals. EPA does recognize that species- and tissue-differences are likely to occur if a chemical induces aromatase activity, and are currently validating another in vitro steroidogenesis assay (H295R cell line) that has the capability to detect both inhibition and induction of aromatase activity. In addition, the fish assay being proposed for use as a Tier 1 screen has been shown to be sensitive for the identification of aromatase inhibitors and will serve as a complementary assay,.
53	SM	Evaluation of chemicals should also be made in combination especially for those showed false negative since some of them react in synergism way and could have a favourable complementary structures to inhibit more efficiently aromatase activity (Benachour et al, 2007).	Mixtures always present problems for testing and risk assessment. While EDSTAC recommended limited testing of mixtures in the EDSP, the Scientific Advisory Pane (SAP) in 1999 recommended that EPA focus only on testing individual chemicals in the near term.
	Additional comments and editorial corrections--See pages 2-33 through 2-41 of the ERG Peer Review Report		

Peer Review Panel Members:

SB = Scott Belcher
LK = Laura Kragie
MM = Marion Miller
SM = Safa Moslemi
TS = Thomas Sanderson