



Peer Review Results for the Aromatase Assay

Prepared for:

U.S. Environmental Protection Agency
Exposure Assessment Coordination and Policy Division
Office of Science Coordination and Policy
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

Prepared by:

Eastern Research Group, Inc.
14555 Avion Parkway
Suite 200
Chantilly, VA 20151-1102

22 January 2008

TABLE OF CONTENTS

	Page
1.0	INTRODUCTION 1-1
1.1	Peer Review Logistics..... 1-2
1.2	Peer Review Experts..... 1-3
2.0	PEER REVIEW COMMENTS ORGANIZED BY CHARGE QUESTION..... 2-1
2.1	Overall General Comments..... 2-1
2.2	Comment on the Clarity of the Stated Purpose of the Assay..... 2-2
2.3	Comment on the Biological and Toxicological Relevance of the Assay as Related to its Stated Purpose 2-3
2.4	Provide Comments on the Clarity and Conciseness of the Protocol in Describing the Methodology of the Assay such 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 Results..... 2-6
2.4.1	Comprehend the Objective 2-6
2.4.2	Conduct the Assay 2-7
2.4.3	Observe and Measure Prescribed Endpoints..... 2-12
2.4.4	Compile and Prepare Data for Statistical Analyses 2-13
2.4.5	Report Results..... 2-14
2.4.6	Provide Any Additional Advice Regarding the Protocol 2-15
2.5	Comment on Whether the Strengths and/or Limitations of the Assay Have Been Adequately Addressed..... 2-19
2.6	Provide Comments on the Impacts of the Choice of a) Test Substances, b) Analytical Methods, and c) Statistical Methods in Terms of Demonstrating the Performance of the Assay..... 2-22
2.6.1	Test Substances..... 2-23
2.6.2	Analytical Methods..... 2-24
2.6.3	Statistical Methods in Terms of Demonstrating the Performance of the Assay 2-25
2.7	Provide Comments on Repeatability and Reproducibility of the Results Obtained with the Assay, Considering the Variability Inherent in the Biological and Chemical Test Methods..... 2-25
2.8	Comment on Whether the Appropriate Parameters were Selected and Reasonable Values Chosen to Ensure Proper Performance of the Assay, with Respect to the Performance Criteria 2-27
2.9	Comment on the Clarity, Comprehensiveness and Consistency of the Data Interpretation with the Stated Purpose of the Assay 2-28
2.10	Please Comment on the Overall Utility of the Assay as a Screening Tool, to be used by the EPA, to Identify Chemicals that have the Potential to Interact with the Endocrine System Sufficiently to Warrant Further Testing..... 2-30
2.11	Additional Comments and Materials Submitted..... 2-33

TABLE OF CONTENTS (Continued)

	Page
3.0	PEER REVIEW COMMENTS ORGANIZED BY REVIEWER..... 3-1
3.1	Scott Belcher Review Comments 3-1
3.2	Laura Kragie Review Comments..... 3-14
3.3	Marion Miller Review Comments 3-23
3.4	Safa Moslemi Review Comments..... 3-30
3.5	Thomas Sanderson Review Comments 3-37
Appendix A:	CHARGE TO PEER REVIEWERS A-1
Appendix B:	INTEGRATED SUMMARY REPORTB-1
Appendix C:	SUPPORTING MATERIALC-1

1.0 INTRODUCTION

In 1996, Congress passed the Food Quality Protection Act (FQPA) and amendments to the Safe Drinking Water Act (SDWA), which requires EPA to:

“...develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by naturally occurring estrogen, or other such endocrine effect as the Administrator may designate.”

To assist the Agency in developing a pragmatic, scientifically defensible endocrine disruptor screening and testing strategy, the Agency convened the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC). Using EDSTAC (1998) recommendations as a starting point, EPA proposed an Endocrine Disruptor Screening Program (EDSP) consisting of a two-tier screening/testing program with in vitro and in vivo assays. Tier 1 screening assays will identify substances that have the potential to interact with the estrogen, androgen, or thyroid hormone systems using a battery of relatively short-term screening assays. The purpose of Tier 2 tests is to identify and establish a dose-response relationship for any adverse effects that might result from the interactions identified through the Tier 1 assays. The Tier 2 tests are multi-generational assays that will provide the Agency with more definitive testing data.

One of the test systems recommended by the EDSTAC was the placental aromatase assay. Its purpose in the Tier-1 battery is to provide a sensitive in vitro assay to detect chemicals that may affect the endocrine system by inhibiting aromatase, the enzyme responsible for the conversion of androgens to estrogens. Alterations in the amount of aromatase present or in the catalytic activity of the enzyme will alter the levels of estrogens in tissues and dramatically disrupt estrogen hormone action. EPA has chosen to validate two versions of the aromatase assay. The first version uses microsomes isolated from the human placenta. The other uses a human recombinant microsome.

Although peer review of aromatase assay will be done on an individual basis (i.e., its strengths and limitations evaluated as a stand alone assay), it is noted that the aromatase assay along with a number of other in vitro and in vivo assays will potentially constitute a battery of

complementary screening assays. A weight-of-evidence approach is also expected to be used among assays within the Tier-1 battery to determine whether a chemical substance has a positive or negative effect on the estrogen, androgen or thyroid hormonal systems. Peer review of the EPA's recommendations for the Tier-1 battery will be done at a later date by the FIFRA Scientific Advisory Panel (SAP).

The purpose of this peer review was to review and comment on the aromatase assay for use within the EDSP to detect chemicals that may affect the endocrine system by inhibiting aromatase. The primary product peer reviewed for this assay was an Integrated Summary Report (ISR) that summarized and synthesized the information compiled from the validation process (i.e., detailed review papers, pre-validation studies, and inter-lab validation studies, with a major focus on inter-laboratory validation results). The ISR was prepared by EPA to facilitate the review of the assay; however, the peer review was of the validity of the assay itself and not specifically the ISR.

The remainder of this report is comprised of the unedited written comments submitted to ERG by the peer reviewers in response to the peer review charge (see Appendix A). Section 2.0 presents peer review comments organized by charge question, and Section 3.0 presents peer review comments organized by peer review expert. The Integrated Summary Report is presented in Appendix B and additional supporting materials are included in Appendix C.

The final peer review record for the aromatase assay will include this peer review report consisting of the peer review comments, as well as documentation indicating how peer review comments were addressed by EPA, and the final EPA work product.

1.1 Peer Review Logistics

ERG initiated the peer review for the aromatase assay on December 12, 2007. ERG held a pre-briefing conference call on January 3, 2008 to provide the peer reviewers with an opportunity to ask questions or receive clarification on the review materials or charge and to

review the deliverable deadlines. Peer review comments were due to ERG on or before January 9, 2008.

1.2 Peer Review Experts

ERG researched potential reviewers through its proprietary consultant database; via Internet searches as needed; and by reviewing past files for related peer reviews or other tasks to identify potential candidates. ERG also considered several experts suggested by EPA. ERG contacted candidates to ascertain their qualifications, availability and interest in performing the work, and their conflict-of-interest (COI) status. ERG reviewed selected resumes, conflict-of-interest forms, and availability information to select a panel of experts that were qualified to conduct the review. ERG submitted a list of candidate reviewers to EPA to either (1) confirm that the candidates identified met the selection criteria (i.e., specific expertise required to conduct the assay) and that there were no COI concerns, or (2) provide comments back to ERG on any concerns regarding COI or reviewer expertise. If the latter, ERG considered EPA's concerns and as appropriate proposed substitute candidate(s). ERG then selected the five individuals who ERG determined to be the most qualified and available reviewers to conduct the peer review.

A list of the peer reviewers and a brief description of their qualifications is provided below.

- **Scott Belcher, Ph.D.**, is currently an Associate Professor in the Department of Pharmacology and Cell Biophysics, and Co-Director of the Molecular, Cellular and Biochemical Pharmacology Graduate Program at the University of Cincinnati College of Medicine. He received his Ph.D. in Molecular Genetics from the University of Texas Southwestern, Dallas, TX in 1993. Dr. Belcher's research interests include the functional role of steroid hormones and their receptors in the developing nervous system and in brain tumors, signaling mechanisms of endogenous and environmental estrogens, and the molecular and developmental pharmacology and toxicology of ethanol and endocrine disruptors. He is a current member of the Internal Advisory Committee for the Ohio Interdisciplinary Women's Health Career Training Program and the University of Cincinnati's Cancer Center. He is also an investigator in the Signal Transduction Core at

the University of Cincinnati Center for Environmental Genetics. A few of the professional societies of which he is a member include the American Society for Pharmacology and Experimental Therapeutics, the Endocrine Society, the Federation of American Societies for Experimental Biology, the Society for Neuroscience, and the Society of Toxicology. Dr. Belcher has published peer reviewed research articles in journals such as *Developmental Brain Research*, *Endocrinology*, the *Journal of Comparative Neurology*, the *Journal of Neuroscience*, and the *Journal of Pharmacology and Experimental Therapeutics*.

- **Laura Kragie, M.D.**, is a licensed physician and scientist specializing in Translational Medicine. Her career experience encompasses academic medicine, pharmaceutical development, and FDA regulatory review. She received her B.S. in Biology/Biochemistry/Psychology at University of Illinois C-U, and her M.D. at University of Iowa. Postgraduate training in Translational Medicine included areas of Internal Medicine, Endocrinology, Molecular Biology and Biochemical Pharmacology at SUNY Buffalo, and Psychiatry and Clinical Pharmacology faculty appointments at Harvard Medical School and Georgetown University Medical School. She was a Medical Officer at FDA/CDER Divisions Cardio-Renal Drug Products and Anesthesia, Critical Care, & Addiction Drug Products. As a Medical Officer, Dr. Kragie identified pediatric cardiovascular safety issues in ADHD drugs in 1995. Dr. Kragie reviews manuscripts for numerous peer-reviewed science journals, is an associate editor for *Endocrine Research*, is a member of IRB and IACUC committees for The Institute for Genomic Research / J Craig Venter Institute, and is a member of NIH Study Section Special Emphasis Panel for BioInformatics and In Vivo Imaging. She is author to publications assessing azole antifungals impact on aromatase and assessing their reproductive toxicity impact as reported in the literature. In 2005, she was Co-chair & Speaker, ASCPT Workshop Orlando FL, “Interaction of Drugs with Estrogen Endocrinology.” She currently heads her consulting firm, BioMedWorks (www.biomedworks.com).
- **Marion Miller, Ph.D.**, is currently a Professor in the Department of Environmental Toxicology, University of California, Davis; Director of the Western Region IR-4 Project (USDA): A National Agricultural Program to clear pest control agents for minor use

crops; and Associate Director of the University of California's Toxic Substances Research and Teaching Program. She received her Ph.D. in Pharmacology from the Medical University of South Carolina in 1982. In April 2004 Dr. Miller served as an ad hoc member for the Special Emphasis Panel on Innovative Toxicology Models for the National Cancer Institute, and in 2006 and 2007 served as an ad hoc member of the Integrative Clinical Endocrinology and Reproduction (ICER) Study Section for the National Institute for Health (NIH). Professional peer reviewed journals in which she has published research articles include, *Biology of Reproduction*, *Fundamentals of Applied Toxicology*, *Reproduction Toxicology*, and *Toxicology of Applied Pharmacology*.

- **Safa Moslemi, Ph.D.**, is currently an Assistant Professor in Biochemistry at the University of Caen, a researcher in the Biochemistry and Molecular Biology Laboratory studying estrogens and reproduction, and a researcher in the Extra- cellular Matrix and Pathology Laboratory at the Institut de Biologie Fondamentale et Appliquée (I.B.F.A.). He received his Ph.D. in 1993 from Ecole Nationale Supérieure des Industries Agricoles et Alimentaires (E.N.S.I.A), and in 1998 a diploma of the capacity for research management from University of Caen, France. Dr. Moslemi has published articles on aromatases and aromatase inhibitors since 1993 in journals such as, *Comparative Biochemistry and Physiology*, the *Journal of Endocrinology*, the *Journal of Enzyme Inhibition*, the *European Journal of Biochemistry*, *Molecular and Cellular Endocrinology*, and the *Journal of Steroid Biochemistry and Molecular Biology*.
- **Thomas Sanderson, Ph.D.**, obtained his bachelors degree (B.S. 1989) from the Faculty of Chemistry and Pharmacochimistry, Free University of Amsterdam, the Netherlands. He went on to complete a Ph.D. degree (Ph.D. 1994) in Pharmacology and Toxicology at the Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada. His doctoral research focused on the toxic effects of dioxins and PCBs on various wild and domestic avian species. This avian toxicology work was continued during his postdoctoral research training (1994-1996) at the National Food Safety and Toxicology Centre at Michigan State University, MI, USA. It is here that his research interests turned towards endocrine disruption and steroid hormone synthesis and metabolism. During his assistant professorship (1997-2005) at the Institute for Risk

Assessment Sciences, University of Utrecht, the Netherlands, he established a research program to study the effects of xenobiotics on the steroid biosynthesis pathway in humans and wildlife. Key research accomplishments are the identification of aromatase, the enzyme responsible for the conversion of androgens to estrogens, as an important target for endocrine disruptors. His work demonstrated that induction or inhibition of aromatase activity or expression by various pesticides, medicinal drugs and naturally occurring phytochemicals posed an alternate non-receptor mediated mechanism by which chemicals can cause pro- or antiestrogenic/androgenic effects in humans and wildlife. As associate professor (2005) at the Institut Armand-Frappier in Montréal, QC, Canada, Thomas Sanderson is focusing his research on the effects of chemicals on the regulation of expression and catalytic activity of several key enzymes involved in the biosynthesis of potent steroid hormones, such as aromatase, steroid 5-alpha reductase and steroid 17-alpha-hydroxylase/17-20-lyase.

2.0 PEER REVIEW COMMENTS ORGANIZED BY CHARGE QUESTION

Peer review comments received for the aromatase assay are presented in the subsections below and are organized by charge question (see Appendix A). Peer review comments are presented in full, unedited text as received from each reviewer.

2.1 Overall General Comments

Scott Belcher: Below are the prepared comments and suggestions addressing the issues and questions raised in the Charge to Reviewers document for the independent review of the aromatase assay as a potential screen in the Endocrine Disruptor Screening Program (EDSP) Tier-1 Battery. This review is focused upon the scientific work the United States Environmental Protection Agency (EPA) performed as a validation of the Endocrine Disruptor Screening Assay for Aromatase, as presented in the Draft Integrated Summary Report (ISR) on Aromatase, December 11, 2007. However, additional reference material provided as background, including the laboratory reports for the studies, and the Detailed Review Paper (DRP) on aromatase were also reviewed and used extensively as supporting documentation to supplement the information in the ISR. As charged, this review is focused upon the scientific work EPA performed to validate the Aromatase Assay as contained in the draft ISR, and is not a review of the draft ISR. However, some comments regarding critical information contained within, or lacking from the ISR draft will be addressed below. In some cases suggestions for addressing the specific comments are also presented.

Safa Moslemi: This report « Integrated Summary Report or ISR on Aromatase » summarizes and synthesizes the information compiled from the validation process in order to propose a protocol on Aromatase Assay as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery. Both the human placental microsomal assay and the recombinant assay using human recombinant microsomes from Gentest (Human CYP19 + P40 reductase SUPERSOMES) were validated and their equivalence demonstrated. You find here after the answers and comments to the charge questions.

2.2

Comment on the Clarity of the Stated Purpose of the Assay

Scott Belcher: The stated purpose of the aromatase assay and how it fits into the overall EDSP Tier-1 Battery is felt to be adequately described to a well-informed scientist.

However, without detailed scientific understanding of endocrinology and the entire androgen/estrogen/aromatase system, and aromatase's role in peripheral tissues (especially during development), the complete grasp of the purpose and significance of the assay is not possible. Currently superficial information is stated in Background paragraph 1, although 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. This information should be followed by a sentence or two stating that in vitro, microsomal aromatase activity is considered a good surrogate indicator of enzyme activity in aromatase expressing cells and sensitive tissues in vivo. Acceptable communication of this would be possible in two or three well-crafted sentences. Currently a concrete understanding of the relevance and purpose of the assay must be extracted from various portions of ISR including the background sections of the Executive summary and Sections 2.0 – 3.0 of the ISR.

Laura Kragie: Yes, the stated purpose of the assay is clear. This is a screening tool to initially assess chemical compounds for their impact on estrogen formation.

Marion Miller: The purpose of the assay is clearly stated. The aromatase assay is one of a battery of assays developed for the Endocrine Disruptor Screening Program. The purpose of the assay is to screen for chemicals which have the capability of inhibiting aromatase, the enzyme responsible for conversion of androgens to estrogens. This assay is an alternative assay in the Tier 1 screening battery and is designed to detect chemicals that would inhibit estrogen

biosynthesis. It is an *in vitro* assay that allows for rapid and relatively inexpensive screening of chemicals.

Safa Moslemi: The purpose of assay is well presented and consist to propose a validated test using human placental microsomal preparation or human recombinant microsomes to evaluate interference of chemicals with endocrine system by inhibiting aromatase (Tier I) the key enzyme responsible of irreversible conversion of androgens in estrogens. Synthesis of estrogens occur, besides gonads and placenta, in many non reproductive tissues of several vertebrate and species of both sexes. Chemicals testing positive in Tier I would be further evaluated in Tier II which will aim to characterize the adverse effects resulting from that interaction and the exposures required to produce them.

Thomas Sanderson: The Integrated Summary Report (ISR) states that as part of a battery of *in vitro* and *in vivo* tier 1 screening tools, the placental and recombinant microsomal aromatase assays are intended to determine whether chemicals have the ability to inhibit the catalytic activity of aromatase. Within this limited definition of the purpose of the assay, the stated purpose is clear.

2.3 Comment on the Biological and Toxicological Relevance of the Assay as Related to its Stated Purpose

Scott Belcher: As an individual component of the Tier 1 battery of tests, the aromatase assay is felt to be biologically relevant for the purpose of identifying environmental endocrine disruptors (EDC) that may act via inhibition of aromatase activity. The ability to reliably detect subsets of chemicals that influence activity of the aromatase enzyme (in this case, limited to inhibitors of aromatase enzyme activities), and thus potentially impact androgen/estrogen sensitive hormonal systems is a critical and biologically relevant component of the Endocrine Disruptor Screening Program (EDSP).

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 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.

Laura Kragie: The importance of the aromatase enzyme for estrogen formation and function in the mammalian organism is well reviewed in the ISR. The relevance of aromatase to reproductive function and assessment of toxic effects are described in the cited references Kragie et al 2002 and Kragie 2002.

The placental form of aromatase may be different from other isoforms that occur in tissues other than placenta, but it does suffice for this purpose of crude screening. Once the chemicals are classified in one of three categories, then more definitive studies can be performed by researchers to elucidate the compounds impact on biology. Some of these other tests may be included in the Tier 1 Battery.

It is essential, however, to select an assay that is cost-effective for **screening** the proposed 10,000 chemicals. A cell-based assay or HPLC-based assay would be prohibitively costly (about 10X higher than a non-chromatography based assay) if it were pursued instead in an attempt to achieve a higher confidence in the results. Therefore, I recommend that the very initial crude screening phase be done using the High Throughput BD Supersomes Aromatase Assay, which uses a fluorescent enzyme substrate (DBF), microtiter plates, fluorescence detection and perhaps 3 concentrations of assessed chemical: 1, 10, 100 micromolar. (Compounds are rarely relevant in the millimolar range, and their solvents become a dominant effect in that range.) The compounds identified to be inhibitors would then go on for assessment with this EPA validated tritiated water method, using a full concentration curve to better define the IC50 value. The most potent inhibitors should be assessed first.

Marion Miller: Aromatase catalyzes the conversion of androgens to estrogens. The rationale for inclusion of this assay as an alternative Tier 1 assay is based on the likelihood of a differential sensitivity of males and females to aromatase inhibition. Although both males and females require estrogen for reproductive health, the female is viewed as more susceptible to loss

of estrogen biosynthetic capabilities due to the importance of estrogen in normal female reproduction. It is indicated (p 14, Table 2.4-2) that if studies were conducted in the male only, male animals may not be sufficiently sensitive to aromatase inhibition and decreased estrogen levels to allow effects to be detected. Although the extent of this gender difference is not documented in detail, the assay provides a useful first level *in vitro* screen for chemicals capable of inhibiting aromatase. Because the studies are conducted *in vitro*, confounding effects of whole animal physiology and feedback mechanisms as well as the absorption, distribution, metabolism and excretion characteristics of the individual chemical are not considered. This can be viewed both as a strength and a weakness since a direct effect on the enzyme will be readily measured but the relevance of that effect in the whole animal is not tested. An additional application for the assay is to supply more detailed mechanistic information about effects on steroidogenesis which may have been detected in the Tier 1 *In Vitro* Steroidogenesis Assay. However, in the absence of details about the *In Vitro* Steroidogenesis Assay the utility of the aromatase inhibition assay to provide additional information is not clear.

Safa Moslemi: Since estrogens are involved in the homeostasis of many of tissues and organs in different species, therefore, evaluation of its synthesis is relevant of both reproductive and non reproductive systems. However, the proposed protocol could not give any information on toxicological system. To reach this goal, toxicological test using *in vitro* cell culture should be carried out.

Thomas Sanderson: To a degree it is. The rationale for concern about chemicals that inhibit aromatase is that such chemicals would result in reductions in endogenous estrogen concentrations in exposed organisms. As outlined in the ISR, biological and toxicological consequences would be numerous, including disruption of reproductive cycle and pregnancy in females, sperm production/capacitation in males and possible behavioral effects in both sexes. Estrogens are important for bone homeostasis, growth and differentiation of numerous tissues and have modulatory effects on the immune system, and many other systems in the human body regardless of sex. It should be pointed out strongly that, although discussed to a certain degree in the ISR, estrogens are not strictly female hormones and in fact have very crucial functions in both sexes, whether it concerns sexual development or numerous basic functions unrelated to sex. Dependent on the sex and on the extent of inhibition of the aromatase enzyme deleterious

effects can be very diverse. Something not addressed in the documentation is the following: 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.

2.4 **Provide Comments on the Clarity and Conciseness of the Protocol in Describing the Methodology of the Assay such 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 Results**

Marion Miller: (For this section the reviewer specifically evaluated the assay protocol in Appendix A as this represents the summation of findings from protocol development)

Safa Moslemi: Protocol is well described and the methodology presented in a comprehensible manner allowing the reader to follow easily all steps cited above.

2.4.1 **Comprehend the Objective**

Scott Belcher: The object of the protocol is straight forward; the current phrasing should be corrected to make this clearer to the reader.

A-6 1.0 Objective: currently reads: “*The objective of this protocol is to describe procedures for conduct of the aromatase assay as a Tier 1 screen using either human placental or recombinant microsomes.*”

Suggested that the highlighted phrase be changed to read: “**to conduct the aromatase assay**”

Laura Kragie: Yes, the objective is clear. The assay will assess any compound, that upon acute exposure, will reduce the production of product (estrone) via detection of the reaction's product, water (scintillation count of tritium). This detection will occur regardless of mechanism of enzyme inhibition.

Marion Miller: Objective to measure aromatase activity is indicated

Thomas Sanderson: The objective is clearly outlined in sections 1 and 2 of the ISR.

2.4.2 Conduct the Assay

Scott Belcher: Generally the protocol does a good job of describing the assays. It is estimated that with some modification, the protocol (appendix A) would allow a laboratory to conduct the assay.

There are a few typos and some technical difficulties that must be addressed. Importantly there are a few important details missing, and in places it is felt that there is too much "flexibility" allowed in the current protocol. Details regarding difficulties associated with the protocol and some points in the ISR are described, along with recommendations for correction, if applicable, are listed below.

A-6 2.1.1, sentence 3: "...is usually supplied at a specific activity of 20-30 μ Ci/mmol."

This statement is believed to be inaccurate due to a typo that can be corrected by deletion of the highlighted μ Ci/mmol is believed to be the correct unit).

Preferred alternative - the information regarding how the radiolabeled androstenedione is "usually supplied" could be deleted, there is no utility of this information to a specific protocol. Minimal specific activity and purity requirements are stated, thus making this statement irrelevant.

A-7 2.2 Test Chemicals

In addition to the provided information for each test chemical, information regarding stability and date of expiration should be provided.

Test Chemical formulation/4-OH ASDN formulation – the options for chemical formulation in “buffer, absolute ethanol, or DMSO” is problematic. There are no set criteria for preparation of stock formulations. This point is not well addressed in either the ISR or the detailed protocol (Appendix A).

Importantly - 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 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.

Two recommendations are suggested:

- 1) The **Full Activity Control** should be described explicitly as containing the same vehicle and concentration of vehicle as the 4-OH ASDN positive control in the case of the positive control experiment described in Table 4 (A-13). It is also important to clarify that this corresponds to the highest concentration of vehicle present in the concentration/response treatments (“Sample Type Conc 1-8).
- 2) An additional negative control group (i.e. **test chemical vehicle control**) should be added to every experimental assay (e.g. Table 5; A-15). This control would be composed of all assay components plus an amount of “test-chemical vehicle” equivalent to the highest concentration of vehicle present in any of the test-chemical treatment groups (Test Chem. Conc 1-8).

It is suggested that a “*universal solvent*” be adopted which is useful for the majority of anticipated test chemicals (as well as positive or negative control treatments as appropriate). Candidate chemicals incompatible with the “universal solvent” should be identified prior to analysis and a prescribed substitute solvent be used with the addition necessary controls.

As a related comment, 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.

Section 2.4.1 Human Placental Microsomes

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 be addressed.

Suggestions to address the potential influence of genetic variability on the finding from the placental aromatase assay:

1) The genotype (and potentially haplotype) of the CYP19 gene present in each placental preparation should be characterized – isolation and archiving of placental DNA, and sequencing of the CYP19 gene would be straight forward and rapid. **The collection, analysis, and archiving of this material (genomic DNA) and sequence information for each microsomal preparation is considered vital.**

2) Section 6.0 (A-14) paragraph 2 sentence 2 states “*A chemical shall be tested in three independent runs*”. While not done previously, three truly independent runs require the use of 3 different preparations of microsomes. Thus, it is suggested that microsomes from three different placental preparations be used. Regardless, the meaning of “independent run” as used in this sentence must be clearly defined.

General comments: In review of the ISR, DRP and other reference information, specific information regarding the starting amounts of placental tissue used was not found. It would be helpful in the protocol to have information regarding an acceptable scale (amount of starting material) for each preparation. Addition of information specifying an acceptable range of typical tissue wet-weights is considered useful as an aid for preparation planning and assay preparation standardization. Additional information regarding expected ranges of microsome yields, etc., should also be considered for inclusion. Those guidelines would aid in achieving predictable and consistently useful yields of microsomes and aromatase activities.

2.4.1.2, bullet 3 and 5: wash volume should be specified.

2.4.1.2, bullet 6: guidelines for volume of buffer for resuspending pellet should be specified.

2.4.1.2, bullet 7: specific guidelines regarding aliquot volumes and minimal acceptable stock protein concentrations should be specified.

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 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.

Note: The above suggested practice is used for the Human Recombinant microsomes (2.4.2.2, A-10), which are aliquoted into individual use vials based on estimates of protein content.

Laura Kragie: The method protocol is generally clear. However see advice given in Question 4.

Marion Miller: Overall the assay is well described but some points could be clarified.

- 1) Page A -6. Section 2.1.1. Androstenedione is usually supplied with a specific activity of mCi/mmol rather than uCi/mmol as indicated This also explains the contradiction between sentences 4 and 5 in this section.
- 2) What buffer is used to make stock solutions (section 2.1.3)? – presumably the 0.1M phosphate buffer indicated in section 2.5.1 but this could be specified.
- 3) In section 2.1.3 what does “record the weight of each component added” refer to?
- 4) Timing for use of microsomes should be defined rather than recommended (section 2.4.1.3).
- 5) Why is propylene glycol added to the assay?

Thomas Sanderson: The protocol is clearly described in section 4 and appendix A of the ISR.

2.4.3 Observe and Measure Prescribed Endpoints

Scott Belcher: Sections 4.0 through 6.0 are clear and detailed, the measurement of $^3\text{H}_2\text{O}$ using liquid scintillation spectrometry is straight forward and is adequately described.

4.0, pg. A-12, second bullet: "...are presented in Table 3" should be corrected to "...are presented in Table 2".

6.0, last sentence paragraph 2, pg. A-14: the reference to Table 6 should read "Table 5".

Laura Kragie: The full activity point of 100% is clearly understandable and achievable. However, the 0% point (bottom) is more difficult to establish. The scintillation counts are progressively diminished and therefore a reduction of the signal to noise ratio. In the assessed chemical inhibition curve, the lower half is more difficult to clearly establish due to the signal to noise issue and also the problems associated with solubility of chemicals at high concentrations, and the nonspecific effects of the high chemical concentrations.

Marion Miller: The method for measurement of aromatase activity is well described in detail. (Section 4, Pages A-11 & 12.) Typographical errors with confusion of the present and past tense could be corrected. In section 5, Page A-13 where the positive control assay is described, it is indicated that the minimum level of aromatase activity in the full activity control will be 0.100 nmol/min/mg protein. However, this value refers to the minimum level only for the recombinant microsomes. The minimum level for the placental microsomes is 0.03 nmol/min/mg protein and this should also be indicated.

Thomas Sanderson: 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?

2.4.4 Compile and Prepare Data for Statistical Analyses

Scott Belcher: 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).

7.2 Model Fitting and 7.6 Statistical Software – 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 (see comments in #7 below regarding “goodness of fit” obtained with each sigmoidal-model and incomplete dose/responses).

It is strongly recommended that the most recent version of a single statistical software package is adopted (e.g. Prism ver. 5). Convenience of use is not an acceptable justification for selection of a software package to use for critical data analysis – as noted in the ISR there are important differences in regression model fitting algorithms and capabilities between Prism ver. 5 and earlier versions of the software. Those changes directly impact the non-linear regression model fitting used for the aromatase assay.

Laura Kragie: The proposed statistical method is standard for these assays and is appropriate. Because of the problems discussed in 3. (c), more leeway is given to the bottom parameter to establish the sigmoid curve that determines a chemical’s IC50 value.

Marion Miller: Methodology appears appropriate. Use of % control data reduces variability between preparations yet still provides data about the potency of the inhibitor. Spread sheet is supplied. Commercially available statistical software is recommended.

Thomas Sanderson: 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 and included as part of the 'assay package'.

2.4.5 Report Results

Scott Belcher: 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.

Laura Kragie: Yes, the three categories for classification (inhibitor, equivocal, noninhibitor) are a feasible presentation.

Marion Miller: The IC₅₀ values are generally reported as log numbers. Reporting these as linear values would give a better appreciation of relative potencies.

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. However, 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.

Thomas Sanderson: Straightforward, other than several minor items in the comments that follow below.

2.4.6 Provide Any Additional Advice Regarding the Protocol

Scott Belcher: None – detailed advice and recommendations were presented above.

Laura Kragie: Econazole is metabolized by CYP450s in the placental microsomes. It is known CYP4503A inhibitor with submicromolar range IC₅₀. It was the most potent inhibitor of the tested series and any slight change in econazole concentration due to CYP3A metabolism will cause variability in the amount of tritiated aromatase activity measured. The variability should be less when using the recombinant aromatase microsomal preparation because recombinant aromatase P450 is enriched in those microsomes relative to other CYP450s. (SUPERSOME activity is catalyzed by human CYP19 that is expressed from human CYP19 cDNA using a baculovirus expression system. Baculovirus infected insect cells were used to prepare these microsomes. These microsomes also contain cDNA-expressed human P450 reductase. A microsome preparation using wild type virus is used as a control.) Econazole IC₅₀ value determined from this validation series was very consistent with the value determined using the HT SUPERSOME assay using DBF for enzyme substrate. See Kragie et al 2002. Very potent inhibitors require more precise assay procedure and practice; e.g., time, temperature, concentrations, and buffer. HT screening assay also recommends using insect cell protein to reduce the nonspecific binding of drug to apparatus that depletes the effective drug concentration exposed to aromatase enzyme.

The aromatase activity in the absence of any test substance was used as the benchmark (100 percent) activity. However, often there is need for a vehicle blank using the same solvent dilution of DMSO or ethanol, if greater than 1% final concentration in the assay.

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.

Redox conditions are sensitive to oxidation. Be aware of oxidation and keep tubes capped.

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 with stop procedure as compared to reaction start. Also, the last microsomes pipetted may be cooler in temperature than the initial aliquot pipetted. Technician needs to pay attention to timing and temperature.

Marion Miller: 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.

Safa Moslemi: In order to improve protocol, the following advices are proposed :

- 1) Add the substrate androstenedione 4 μM 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 choc 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 tow supplementary extractions by solvent, this help to reduce the time of experimentation.
- 4) Add the formula for the calculation of the specific aromatase activity in $\text{nmol.mg protein}^{-1} . \text{min}^{-1}$ by expressing all parameters used such as, background radioactivity,

specific activity of the substrate, time of incubation, protein concentration and finally the correction for the % of the specific radio-labelling at beta position of C1 of substrate.

Thomas Sanderson: I have several comments that should be considered concerning the protocol as described in Appendix A of the ISR.

Final solvent concentration:

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.

Microsome preparation:

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°C for up to 3 years. Has the necessity of the glycerol and dithiothreitol (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?

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.

Protein determination:

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.

It is surprising the assay should be performed using such large volumes, quantities and in traditional cuvettes. The general availability of absorbance plate readers has allowed for

dramatic miniaturization of such assays. The assay could easily be performed using volumes 5-10 times less than those described in the protocol, thus allowing for the use of spectrometer-suitable multi-well plates of anywhere from 24-96 well formats. This would greatly enhance the efficiency of the assay (faster and using less material).

Aromatase assay:

On main question here is why the tritiated water-release protocol was 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 aromatase assay as described is performed in test tubes. I would have thought that the assay could easily be down-scaled to far smaller 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

Is the addition of propylene glycol necessary? It increases the organic solvent burden of the reaction mixture disproportionately compared with all the other components including solvent used for test chemicals and may not be essential to the performance of microsomal enzyme assays.

Semantically it is more appropriate to express the catalytic activity of aromatase when determined using the tritiated water assay as *pmoles of androstenedione converted per time unit per quantity of protein*, rather than amount of *estrone* formed, because estrone is not measured.

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).

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 analyses.

2.5 Comment on Whether the Strengths and/or Limitations of the Assay Have Been Adequately Addressed

Scott Belcher: Strengths: Although general, the most focused descriptions of the strengths of the assay are addressed in Section 1.0 Executive Summary, Background paragraph 1 (pg 1) and Section 3.3 (pg. 21-22). There does not seem to be an attempt to specifically highlight the strengths of the aromatase assay beyond the fact that it is a well-established and reliable assay. Thus, 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.

Limitations: The limitations of the assay are addressed in summary form in section 3.4 (pg. 22).

The limitations of the assay's ability to only identify inhibitory effects, and the inability of the assay to distinguish the "nature" of inhibition are acknowledged.

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.

The final sentence of the Section 3.4 is felt to be irrelevant. This is not a limitation of the assay as it is proposed to detect EDC effects on aromatase; the activity of any other metabolic enzyme is clearly not being considered, and is thus not a limitation of the assay related to the proposed goals of these studies.

Laura Kragie: In general, yes the strengths and limitations are addressed and discussed.

Marion Miller: The limitations of the assay are indicated (p 22, 232) as inability to detect induction of the aromatase enzyme, lack of information about the nature of any inhibitory response, denaturation of the enzyme (Page 232, “receptor” should be replaced with “enzyme”) with subsequent identification of a false positive, inability to test chemicals that have limited water solubility, and the lack of xenobiotic metabolizing activity with consequent inability to detect metabolites with inhibitory activity. To detect either enzyme induction or the presence of a biologically active metabolite would require the use of whole cell systems or whole animals. The level of complexity would be significantly increased and the goals of a rapid, inexpensive screen would be more difficult to meet. The summary document recognizes these limitations and addresses them adequately. 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. Overall, the strength of the assay is that it is quick, simple, inexpensive, relatively robust, gives reproducible data, and allows detection of chemicals that directly inhibit aromatase *in vitro*.

Safa Moslemi: The availability of human placenta, the facility to prepare microsome as source of aromatase activity and the high sensitivity, reproducibility and rapidity of tritiated water assay all these points are well quoted in IRS and place the proposed protocol as the most relevant for in routine evaluation of chemicals on aromatase activity, a crucial target of endocrine disruption. Besides limitations cited in ISR such as; false negative, lack of metabolizing enzymes, and lack of induction and/or inhibition of aromatase expression, this assay , as conducted, can not show chemicals acting in synergism when incubated in combination. Indeed, we recently showed that substances that have no visible aromatase inhibition alone, at 20 μM , become aromatase disruptors (or even cytotoxic) up to 50% inhibition in combination from 4-10 μM demonstrating that substances together may more easily clutter up the active site or alter the enzyme. Chemicals also often present a possible bioaccumulation, and/or indirect actions on signalling pathways. Since organisms are always exposed to mixtures of chemicals, all these issues become crucial in order to evaluate their effects and actions on human health (Benachour et al, Toxicol Appl Pharmacol , 2007).

Thomas Sanderson: The strengths and weaknesses of the placental microsomal assay have been discussed thoroughly in reference 1 (Final Detailed Review Paper on Aromatase). In section 5 of this document the assay is compared to several cell-based assay systems and deemed a more straightforward and better characterized assay than the cell-based ones. The weaknesses described are exhaustive and have taken all aspects of the placental microsomal assay in to consideration. 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 literature that would be equally suitable as tools for screening inhibitors but would also have the added possibility to

detect inducers. Regardless of the difficulties in the interpretation of the relevance of inductions of aromatase activity/expression in cell-based systems, the crude observations would be readily obtained during screening at no extra effort and this information would be available to future investigators to be further studied if deemed of importance.

2.6 **Provide Comments on the Impacts of the Choice of a) Test Substances, b) Analytical Methods, and c) Statistical Methods in Terms of Demonstrating the Performance of the Assay**

Laura Kragie: Yes, the test substances, analytical methods, and statistical methods chosen are appropriate for appraisal of assay performance. These methods are SOP for clinical diagnostic assay assessments.

Safa Moslemi: The choice of chemicals should be better justified (see blow, editorial comments, point 2). The parameters used for analysis and run comparison such as IC50 (or EC50) values, minimum inhibition (top), maximum inhibition (bottom), and Hilleslope all are appropriate and comprehensible for the majority of scientists. Also, selected statistical parameters such as SD, SEM, CV and analysis of variance (ANOVA) are among the most used by biochemists and biologists which offer high performance and reflect assay efficacy, sensitivity and variability between runs and chemicals when comparison performed. However, it is generally preferable to evaluate K_i 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 (K_i 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). Indeed, IC50 values depend on parameters used for its determination which are different in literature (temperature, protein and substrate concentration, time of incubation etc). For instance, if you chose to use a lower concentration of substrate, it will take a little concentration of inhibitor to compete for 50% of the activity. That's why one should be certain to work in saturation condition of substrate (at least at $10 \times K_m$) for the determination of IC50 value and this is not the case in proposed protocol which use 100 nM of androstenedione while its K_m value is about 39 nM. Actually, IC50 might be under estimated in the present protocol.

Thomas Sanderson: The wide range of compounds selected was an appropriate choice for validation of the assay. The test is conducted in such a manner that no enzyme inhibition kinetic properties can be determined. In other words the nature of the inhibition, competitive *versus* non-competitive or mixed-type inhibition will not be known. The protocol could, however, easily be adapted to obtain such information if desired. Inhibition curves would need to be produced in the presence of various concentrations of ASDN substrate.

It should be pointed out that in section 5.1 on page 29 of the ISR competitiveness is erroneously equated to reversibility. Inhibitors that bind to sites other than the catalytic site may produce non-competitive or mixed-type inhibition kinetics, but this does not mean that the inhibition is irreversible, only that increasing the substrate (ASDN) will not restore catalytic activity by displacing the inhibitors from the catalytic site. Inhibition can still be reversible with restoration of original enzyme activity once the inhibitor is eliminated through, for example, metabolism/elimination, as long as the interaction with the ‘other site’ is not covalent (=mechanism based inhibition = irreversible).

Concerning the estrone formation analyses - Section 7.5.3:

The observation that the tritiated water-release assays produces aromatase activities (amounts of $^3\text{H}_2\text{O}$) 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).

2.6.1 Test Substances

Scott Belcher: the reference chemicals (Table 6.1-1 and Table 12.1) used for validation of the aromatase assay is considered appropriate. As a set, they represent a number of different

chemical classes with reasonably predictable effects on aromatase activity based on the supporting literature.

Marion Miller: Substances selected as test substances represented a diverse range of chemical structures as well as applications. 10 test substances (originally there were 11 but lindane, a negative control was dropped) were used in an interlaboratory validation study. An additional 16 substances were tested in the lead laboratory for aromatase inhibition. The selected substances were comprised of both inhibitors and non inhibitors of aromatase and provided information about the reliability of the assay and its ability to detect aromatase inhibitors.

2.6.2 Analytical Methods

Scott Belcher: the analytical methods used to assess aromatase are well established and straight forward.

Marion Miller: The three most important measurements before addition of test substance are the full enzyme activity, background activity and the inhibitory response to the positive control, 4-hydroxyandrostenedione. The purity of the radiolabelled androstenedione substrate is important particularly in this assay as values for background activity would be expected to change dependent on purity of the starting radioactive material. 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.

Microsomal cytochrome P450 levels were initially measured to give an indicator of total P450. However, this adds little information about aromatase activity in a microsomal preparation as the aromatase isoform represents only one of many P450 isoforms. For the aromatase assay, the methods used were established in a rigorous manner with measurement of the protein- and time-dependence of enzyme activity to ensure appropriate substrate concentrations and incubation times.

2.6.3 Statistical Methods in Terms of Demonstrating the Performance of the Assay

Scott Belcher: the statistical methods comparing variability in performance of the aromatase assay and the performance of the assay in different laboratories were appropriate to demonstrate the performance of the assay.

Marion Miller: Less stringent criteria for coefficient of variation in the lowest percent control values are appropriate when the levels measured are so low as to be hardly above background. Use of simple % inhibition to classify chemicals as inhibitors or non-inhibitors instead of using confidence intervals which may overlap when the data is of poor quality, provides an approach where there is a decreased likelihood of misclassification.

2.7 Provide Comments on Repeatability and Reproducibility of the Results Obtained with the Assay, Considering the Variability Inherent in the Biological and Chemical Test Methods

Scott Belcher: Regarding experimental variation, the assay is considered sufficiently repeatable and reproducible. However, it is clear that there were important differences in the quality of the results obtained from different laboratories. In particular, the results obtained from one of the laboratories were consistently much more variable than the other labs. From the documentation available, it appears that those problems resulted from undefined QC/QA problems within that laboratory. Because the four selected laboratories represent a population of “acceptable” laboratories (section 11.0) there is concern with the criteria used to establish “acceptable” laboratory designation.

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. The modified Lowry assay used for determining protein concentration is well-known as rather inaccurate and variable. As a result, the fact 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 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.

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.

Laura Kragie: These results are consistent with the variability inherent in bioanalytic assays that are used routinely in clinical practice. The results will improve with experience and practice of the technician.

Marion Miller: A major source of variability in the interlaboratory validation seems to have arisen from use of a protein standard curve where the values for the protein sample to be measured fell below the levels measured by the standard curve. From an analytical perspective, the optimal situation is where the unknown sample is bracketed by standards and it is somewhat surprising that extrapolated values were used to obtain the protein concentration. This does to some extent explain variability seen in results from one of the laboratories. However, it should be noted that overall the assay was quite robust and despite interlaboratory variability, IC₅₀ values were generally similar.

Safa Moslemi: Absolutely, the assay is sufficiently reproducible as demonstrated by statistical analysis and by fixing about 15% of CV and 95% of confidence intervals for the fitted curve and estimated parameters.

Thomas Sanderson: I have good confidence in the repeatability, reproducibility and overall reliability of the placental microsomal aromatase assay as a test system for aromatase inhibitors. An overall coefficient of variation of less than 30% is quite acceptable for an *in vitro* bioassay. Within-run variabilities reported to be as low as 5-15% are also very respectable.

2.8 **Comment on Whether the Appropriate Parameters were Selected and Reasonable Values Chosen to Ensure Proper Performance of the Assay, with Respect to the Performance Criteria**

Scott Belcher: The performance criteria for the assay are generally considered appropriate and reasonable based on the presented data.

11.2 Performance criteria for test chemicals (pg. 182) – regarding the examination of test data for outliers – a specific method and criterion for determining outliers should be established.

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.

Laura Kragie: The adjustment of the bottom criteria is a necessary improvement.

Marion Miller: Performance criteria are a key aspect of establishing consistent a methodology and generating reproducible data. Minimum enzyme activity in the placental and the recombinant preparations are defined. For this assay the full enzyme activity control is particularly important . This value is used as the 100% value relative to which the effect of the inhibitor is compared and the concentration dependence of the enzyme inhibition is plotted graphically. Performance criteria are $\pm 10\%$ for the full enzyme activity and represents a reasonable variability for measurement of enzyme activity. If full enzyme activity is highly variable such that IC50 values cannot be obtained, the assay should be repeated. For the 4-

hydroxyandrostendione positive control the outlined performance criteria (Page A-13) are reasonable to ensure that the assay is functioning as it should. 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.

Safa Moslemi: The performance criteria for the full activity control (0.100 nmol/mg.min) and the background control (1% of full activity control) provide reference values to the testing laboratory to ensure the detection of both strong and weak inhibitors. Determination of 80% of tolerance interval with 95% confidence guarantee an acceptable variation of data. However, I think that specialists are in position to comment this portion of study.

Thomas Sanderson: The performance criteria are reasonable and based on common sense and practice. 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.

2.9 **Comment on the Clarity, Comprehensiveness and Consistency of the Data Interpretation with the Stated Purpose of the Assay**

Scott Belcher: 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

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.

Laura Kragie: The ISR Dec 11 2007 version still needs a thorough proofing, especially the Tables, for accuracy and completeness. See general comment section for some instances of misinformation and typos.

Marion Miller: As indicated earlier, the data interpretation criteria for inhibitor classification defined an inhibitor when more than 50% inhibition was achieved and as a non-inhibitor when inhibition was not greater than 75% inhibition. This allowed for easy classification of inhibitor and non-inhibitors. However, the equivocal situation where inhibition is 50-75% is not adequately addressed 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. If the highest concentration data points are to be used, there is a greater possibility that enzyme denaturation or other non specific effects rather than enzyme inhibition has occurred. The limitations of this default approach should be addressed.

Safa Moslemi: It is not always easy to understand all these interpretations criteria. This is because I am not statistician. However, if one look at the table 11.3-2 for the adopted data interpretation criteria (page 185), every thing being fortunately clear and comprehensible for the classification of inhibitors.

Thomas Sanderson: They are clear. 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 that will be interpreting the aromatase assay results.

2.10 **Please Comment on the Overall Utility of the Assay as a Screening Tool, to be used by the EPA, to Identify Chemicals that have the Potential to Interact with the Endocrine System Sufficiently to Warrant Further Testing**

Scott Belcher: Overall, the aromatase assay is considered a **critical** *in vitro* screening tool for use by the EPA to identify chemicals that potentially interact with the androgen/estrogen endocrine system.

Laura Kragie: This *in vitro* aromatase assay meets the criteria for a screening tool to identify chemicals that may potentially interact with the endocrine system. The emphasis here is on **screening tool**, and that this method should not be used to definitively categorize a compound as reproductively toxic. It should be a first step in the evaluation process, because of its ease of use, short time course and overall safety and cost. 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.

The real utility of this screening tool is its use with a Rank Order of test compounds in a series of known potent and weak inhibitors. The Rank Order takes into account general variability of assay conditions that apply universally to the overall test set. The relative relationship of the compound to the known and previously tested substances, is the crucial information.

Marion Miller: 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.

Safa Moslemi: 1) As cited in IRS, one of the weakness of the proposed protocol that it can not predict metabolising 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 inhibit aromatase in JEG3 cells (Nativelle-Serpentini et al, 2003). So, a false negative should not systematically be deleted from the next step of evaluation.

2) 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).

3) Androstenedione is one of the aromatase substrate (others are : 16 α -hydroxytestosterone, testosterone, 19-norandrogens) and considered as the preferential one in human. However, in some species, other substrates being used preferentially by aromatase as we previously showed that 19-norandrogens are aromatised 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.

Thomas Sanderson: As a bioassay to identify compounds that have the capability, at least *in vitro*, and in a very simplified enzymatic preparation, to inhibit aromatase activity, this protocol fulfills its objective. As a critical reviewer, I would say that within the very limited constraints of the objective of the bioassay it is useful, but have some concerns about its limitations. The rationale for developing a bioassay for effects of chemicals on aromatase is the fact that this enzyme plays a key role in the local production of estrogens in many tissues in the body and is involved in many essential processes throughout (unborn) life. However, the assay only covers inhibitors. It is well established that increased aromatase expression and estrogen production in tissues is associated with various pathologies including endocrine cancers. This entire facet of

the endocrine disruption paradigm and interest in aromatase as a target for endocrine disruptors is eliminated from the final proposed bioassay and is an important loss.

A considerable amount of resources has been spent on developing and validating the placental microsomal aromatase assay. It would have seemed within the bounds of possibility to develop one or more cell-based assays that would cover a more diverse range of aspects of aromatase function to more fully describe the ability of chemicals to interfere (inhibition/induction/downregulation) with this important endocrine endpoint.

The ISR and in more detail the Final Detailed Review Paper on Aromatase discusses several other potential candidate bioassays for the detection of interferences with aromatase but deems them too complicated, uncharacterized or otherwise limited to be useful as screening tools. It is my opinion that this is an unnecessarily missed opportunity.

Firstly, the perceived limitations of cell-based assays for effects on aromatase mention low basal aromatase expression, potential cytotoxicity at high concentrations of test chemicals and possible biotransformation to (in)active metabolites. A fresh look at, and interpretation of these perceived limitations could equally well transform them into advantages. For example, the fact that certain chemicals are cytotoxic at higher concentrations may be an indicator that the limit of physiological relevance has been reached. The fact that higher concentration (above the 100 micromolar range) may be attained in microsomal fractions is generally of not of great interest on a toxicological level. The additional concern in cell-based assays that some compounds may be too lipophilic to cross cell membranes could be seen differently. Would it not in fact be of importance and very relevant to know whether a chemical that appears to inhibit aromatase in microsomes could even enter a cell in the first place, is soluble in cell culture medium (crystallization is easily observed under a microscope) or is not rapidly metabolized to more or less potent metabolites? The future of relevant bioassays is one that provides an integrated more fully developed picture of the biological/toxicological activity of a chemical. Incorporating bioactivation, bioavailability, different mechanisms of action on the endpoint in question (aromatase activity) are essential components of such an approach. It is my fear that the placental microsomal assay for the screening of effects on aromatase activity (strictly limited to inhibition in isolated microsomal fractions) may become outdated in a fairly short term.

I do not underestimate the complexities involved in the full validation of cell-based assays compared with the proposed microsomal assay(s). However, the greater quality and diversity of the information derived from such assays would, in my opinion, outweigh the concerns about their complexity. It has been made clear in the supporting documentation that cell-based assays perform well on their ability to identify inhibitors of aromatase, albeit with somewhat less sensitivity than the proposed placental microsomal assay. This is readily explained by the fact that chemicals encounter more barriers to reach their target in intact cells than in microsomal fractions that have been membrane-disrupted, concentrated and treated with co-solvents such as propylene glycol. Additionally, metabolism may play a greater role in cell-based assays. The question to be considered is: are these aspects of cell-based assays really disadvantages or do they, in fact, help us in providing a far more relevant representation of what endocrine disrupting chemicals may do to the enzyme aromatase in exposed organisms?

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.

2.11 Additional Comments and Materials Submitted

Scott Belcher: Background is rather dated and has been carried over from previous reports included in the background material, some of those previous reports were completed many years ago – the addition of some updated background information regarding currently accepted understanding of fundamental hormone and EDC mechanisms and modes of action are encouraged.

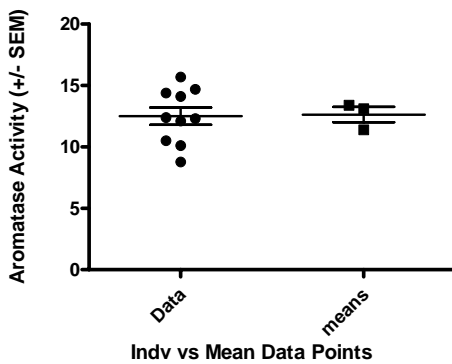
Section 8 and 9 (beginning around section 8.5, page 84 and continuing into section 9.0) contains a large number of typographical errors including Cyp19 being referred to as CYPL9.

Table 8.5-3 would be more valuable if the values for % inhibition mentioned in the text were included.

Supplemental information 1:

Recalculations of data presented in Table 8.2.2. (Prism Ver. 5.0)

Data	Means (reported)
14.40	13.1
14.70	13.4
10.10	11.4
12.10	
12.30	
15.70	
12.40	
14.10	
10.50	
8.78	



	Data	Means
Number of values	10	3
Minimum	8.780	11.40
25% Percentile	10.40	11.40
Median	12.35	13.10
75% Percentile	14.48	13.40
Maximum	15.70	13.40
Mean	12.51	12.63
Std. Deviation	2.238	1.079

Std. Error	0.7076	0.6227
Lower 95% CI of mean	10.91	9.954
Upper 95% CI of mean	14.11	15.31
Coefficient of variation	17.89%	8.54%
Sum	125.1	37.90

Reference: Olson, Sara H., Bandera, Elisa V., and Orlow, Irene Orlow. Variants in Estrogen Biosynthesis Genes, Sex Steroid Hormone Levels, and Endometrial Cancer: A HuGE Review. *Am J Epidemiol* 2007; **165**:235–245.

Laura Kragie: REFERENCES

Kragie L. Turner SD, Patten CJ, Crespi CL, Stresser DM. 2002 Assessing pregnancy risks of azole antifungals using a high throughput aromatase inhibition assay. **Endocrine Research** 28 (3): 129 –140

Kragie L. 2002 Aromatase in primate pregnancy: a review. **Endocrine Research** 28 (3): 121-128

EDITING CORRECTIONS ISR version Dec 11 2007

SECTION 1.0

P 16 Table 1.0 -01. IC50 values for positive control need concentration terms, or logIC50 terms (table 1.0-5 lists -7.3 to -7.0)

P 18 Table 1.0-3. Check these figures

Prochloraz

Placental 20.2 +/- 0.001.8 nM 026.9 +/- 0.003.1 nM

Dicofol

Placental 62.91 +/- 35.86 µM 24.14 +/- 3.72 µM 501 +/-489 µM 53.13 +/-16.56 µM 29.13 +/- 8.62 µM

P 10 Table 1.0-4. Full Activity and Background Control Criteria*

Need to clarify table to state it is for recombinant activity. Legend states different value for placental microsomes.

P 10 Supplemental Testing

“Ten chemicals are clearly noninhibitors, and no concentration-dependent response was observed for any of the noninhibitors These are Vinclozolin, Bisphenol A, Tributyltin, Diethylhexyl phthalate, Methoxychlor, Aldicarb, Flavone, Triadimefon, Imazalil, Apigenin, Ronidazole, Ronidazole, Genestein, p,p’-DDE, Alachlor, Nitrofen, and Trifluralin.” This section needs correcting- 17 chemicals are listed, one repeated, and it includes the six chemicals listed as inhibitors.

P 11 “The percent of control values for each reference chemical run and tube, along with the mean, SD, SEM, and CV of the percent of control across tubes within a run.” Correct sentence fragment.

P 26 correct concentration: **Table 4.6-1. Aromatase Assay Conditions**

[3H]ASDN 100 μ L 100 nM placenta 100 mM recombinant

p 79 missing concentration terms for IC50: “Based on the curve-fit of the percent of control aromatase activity values across six concentrations of 4-OH ASDN, the calculated IC50 values by run and laboratory are summarized in **Table 8.2-5**. The average \pm SEM IC50 values for Laboratories A, B, and C were 57.9 ± 5.9 , 47.3 ± 2.6 , and 81.1 ± 5.5 ; the percent CV values were 17.7, 9.6, and 13.4 percent, respectively. The overall task mean \pm SEM IC50 value was 62.1 ± 10.0 and the percent CV was 27.8 percent.” Missing also from table 8.2-5.

p 83 missing concentration terms for IC50: **Table 8.3-2. Aminoglutethimide IC50 and slope values**

p 85 missing concentration terms for IC50: **Table 8.3-4. Chrysin IC50 and slope values**

p 87 missing concentration terms for IC50: **Table 8.3-6. Econazole IC50 and slope values**

p 90 missing concentration terms for IC50: **Table 8.3-8. Ketoconazole IC50 and slope values**

Marion Miller: Specific Comments by Section

Executive Summary.

Page 7 Table 1.0-3 The value reported for In Vitro with Dicofol and the placental microsomes is much higher than for the other 3 laboratories. Is this a typographical error? Also the overall task group mean for Dicofol does not reflect the mean for the 4 labs even if the In Vitro measurement is changed from 501 to 50.1 or 5.01 or 0.501 assuming that the value in the table represents a misplaced decimal point.

Page 8. Use of terms can be somewhat confusing when they are used without definition. The use of portion to represent beginning and end is not clear in the executive summary although it is apparent later in the document. Similarly, for the use of “full enzyme activity”. Also, the data and treatments in figures 1.0-1 and 1.0-2 are not adequately explained in the executive summary.

Main Document

Page 27. Table 4.6-1. Tritiated androstendione should be the same concentration (100nM) for both assays

Page 61. Isotope is I₁₂₁

Explain abbreviations, for example QC and RE on page 85

Page 90, line 2 units should be ug not g/ml

Page 92 and 135, Define CR and RC

Page 127, Table 9.6-3. Econazole IC₅₀ values in the table are in uM with lots of decimal places whereas in the text nM values are indicated.

Page 175. Define AG and SNLR.

Page 176. What is WA 4-17?

Minor editorial changes are not commented upon

Safa Moslemi: References

Benachour N, Moslemi S, Sipahutar H, Seralini GE, Cytotoxic effects and aromatase inhibition by xenobiotic endocrine disrupters alone and in combination. Toxicol Appl Pharmacol. 2007, 222(2):129-40.

S. Moslemi, P. Auvray, P. Sourdain, M. A. Drosdowski and G.-E. Seralini, Structure-function relationships for equine and human aromatases : a comparative study. *Annals of the New York Academy of Sciences*. 1998, 839 : 576-577.

Nativelle-Serpentini C, Richard S, Seralini GE, Sourdain P, Aromatase activity modulation by lindane and bisphenol-A in human placental JEG-3 and transfected kidney E293 cells. *Toxicol In Vitro*. 2003, 17(4):413-22.

Editorial comments :

- 1) Page 2, the last sentence of first paragraph in “data analysis”, “*Both types of inhibitors have been tested in the validation program*” is not appropriate. Indeed, only nonylphenol was tested for kinetic inhibition which demonstrated a competitive inhibition with Lineweaver- Burk representation while the secondary plots indicated a small contribution of another inhibition type. Therefore, only a competitive inhibitor was demonstrated and this sentence should be modified.
- 2) Page 2, the first paragraph of “reference and supplementary chemicals”, the choice of the compounds is not clearly justified. It would be better to add “aromatase inhibition” in the sentence “*Eleven reference chemicals were selected on the basis of data on aromatase inhibition in the scientific literature and*”
- 3) Page 3, in “protocol optimisation” 6th sentence “*..... cytochrome P450 content*” should be deleted since this evaluation will be omitted from the final protocol.
- 4) Page 5 table 1.0-1, add unity nM for IC50.
- 5) Page 5, table 1.0-2 and other tables in overall ISR the presence of thick horizontal or vertical lines is not justified, please correct them.
- 6) Page 7, Table 1.0-3, please check the values for Prochloraz obtained by RTI and BATTELLE in placental preparation, the values 20.2+/-0.001.8 nM and 026.9+/-0.003.1 nM are unrealistic.
- 7) Page 10, in “testing the methods-supplementary testing” there is a confusion between activity and inhibition. For example, in the 4th line of paragraph “*In no case was the most **activity** (instead of inhibited) level between 75 percent and 50 percent of control, which Five of the six inhibitors exhibited maximal **activity** (instead of inhibition) of*

less than 20 percent of control; the other inhibitor (nitrofen) had a maximal **activity** (instead of inhibition) of about 32 percent of control”. Moreover, in the same paragraph in sentence “Ten chemical are clearly noninhibitors, and.....These are Vinclozolin, BisphenolA, Tributyltin, Diethylhexyl phthalate, Methoxy chlor, Aldicarb, Flavone, Triadimefon, Imazalil, Apigenin, Ronidazole, Genestein, p,p’-DDE, Alachlor, Nitrofen, and Trifluralin”. Should be deleted from this sentence the six inhibitor chemicals (BisphenolA, Flavone, Triadimefon, Imazalil, Apigenin, and Nitrofen) and Ronidazole which was repeated twice.

- 8) Page 12, in “Tiered approach” the sentence “A *negative result in Tier I would be sufficient to put a chemical aside as having low to no potential to cause endocrine disruption*” should be revised since some chemicals considered as negative with placental microsomal system (as Lindane) might have an effect in other system like in JEG3 cells (Nativelles-serpentini et al, 2003) or even their effect being visible when they incubated in combination (Benachour et al, 2007). So, once again, the false negative is really a significant problematic of microsomal system.
- 9) Page 25, in section 4.3.1 Human placental microsomes, should be precised in the first sentence “*Placenta **From healthy and non- smoking women are obtained***” These two conditions (healthy and non-smoking) are quoted in some provided materials (references 1-14) and should be specified in final report as a guarantee for technicians working on aromatase assay and to provide a highly aromatase specific activity in prepared microsomes since it is known that smoking could alter aromatase.
- 10) Page 27, Table 4.6-1, correct final concentration of [3H]ASDN in the recombinant assay to 100 nM instead of 100 mM.
- 11) Page 34, table 6.1-1, for Lindane, in “*basis for selection*”, delete “not” since this substance inhibit aromatase in JEG-3 test for the cited reference (Nativelle-Serpentini et al, 2003). That’s why I think that Lindane should be deleted as negative control in overall IRS. Indeed, the value of background is sufficient for this purpose. In the same table, please add year which was omitted for certain references. In same table, for Chrysin in “test system”, what means “H adip”? Please add comments in legends of this table.
- 12) Page 40, in “Linear production of product” 4th line, correct **0.004** mg/mL instead of **0.005** mg/mL (human recombinant microsomes).

- 13) Page 68 (table 8.2-5), page 72 (table 8.3-2), page 74 (table 8.3-4), page 76 (table 8.3-6), page 79 (table 8.3-8), add unity for IC50.
- 14) Delete page 89 from paragraph “Recombinant microsoms (Human)” until page 92 table 9.0-4 because of repetition (see page 84, from 8.5-1 protein concentration until page 87 section 8.5-4).
- 15) Page 103, tables 9.2-3 to 9.2-6 are not necessary and the text gives enough information.
- 16) Page 115, last sentence “The overall task group mean +/- (instead of ::) SEM”
- 17) page 121, tables 9.5-3 to 9.5-6 are not necessary and the text gives enough information.
- 18) Page 146, line 2 delete “L” from μ LM.
- 19) Page 162, at the beginning of third paragraph, Table **10.3-1** instead of 10.3-2.
- 20) Page 175, last paragraph and at the beginning of page 176, check the mean of Km and Vmax values (37.1 nM and 0.334 nmol/mg/min) which are different from those reported in table 10.6-1 (38.9 nM and 0.351 nmol/mg/min). This is also the case of Ki for nonylphenol (6.83 μ M in the text different from 8.63 μ M in the table).

Thomas Sanderson: References

- Kao, Y. C., Zhou, C., Sherman, M., Laughton, C. A., and Chen, S. (1998). Molecular basis of the inhibition of human aromatase (estrogen synthetase) by flavone and isoflavone phytoestrogens: A site-directed mutagenesis study. *Environ Health Perspect* **106**, 85-92.
- Lephart, E. D., and Simpson, E. R. (1991). Assay of aromatase activity. *Methods Enzymol* **206**, 477-483.
- Mindnich, R., Moller, G., and Adamski, J. (2004). The role of 17 beta-hydroxysteroid dehydrogenases. *Mol Cell Endocrinol* **218**, 7-20.
- Pasqualini, J. R. (2004). The selective estrogen enzyme modulators in breast cancer: a review. *Biochim Biophys Acta* **1654**, 123-143.
- Sanderson, J. T., Hordijk, J., Denison, M. S., Springsteel, M. F., Nantz, M. H., and Van Den Berg, M. (2004). Induction and Inhibition of aromatase (CYP19) activity by natural and synthetic flavonoid compounds in H295R human adrenocortical carcinoma cells. *Toxicol Sci* **82**, 70-79.
- Sanderson, J. T., Seinen, W., Giesy, J. P., and van den Berg, M. (2000). 2-Chloro-s-triazine herbicides induce aromatase (CYP19) activity in H295R human adrenocortical carcinoma cells: a novel mechanism for estrogenicity? *Toxicol Sci* **54**, 121-127.

Vihko, P., Harkonen, P., Oduwole, O., Torn, S., Kurkela, R., Porvari, K., Pulkka, A., and Isomaa, V. (2003). 17 beta-hydroxysteroid dehydrogenases and cancers. *J Steroid Biochem Mol Biol* **83**, 119-122.

Waxman, D. (1988). Interactions of hepatic cytochromes P-450 with steroid hormones: regioselectivity and stereospecificity of steroid metabolism and hormonal regulation of rat P-450 enzyme expression. *Biochem. Pharmacol.* **37**, 71-84.

3.0 PEER REVIEW COMMENTS ORGANIZED BY REVIEWER

Peer review comments received for the Aromatase assay are presented in the subsections below and are organized by reviewer. Peer review comments are presented in full, unedited text as received from each reviewer.

3.1 Scott Belcher Review Comments

Peer Review of the Aromatase Assay

Below are the prepared comments and suggestions addressing the issues and questions raised in the Charge to Reviewers document for the independent review of the aromatase assay as a potential screen in the Endocrine Disruptor Screening Program (EDSP) Tier-1 Battery. This review is focused upon the scientific work the United States Environmental Protection Agency (EPA) performed as a validation of the Endocrine Disruptor Screening Assay for Aromatase, as presented in the Draft Integrated Summary Report (ISR) on Aromatase, December 11, 2007. However, additional reference material provided as background, including the laboratory reports for the studies, and the Detailed Review Paper (DRP) on aromatase were also reviewed and used extensively as supporting documentation to supplement the information in the ISR. As charged, this review is focused upon the scientific work EPA performed to validate the Aromatase Assay as contained in the draft ISR, and is not a review of the draft ISR. However, some comments regarding critical information contained within, or lacking from the ISR draft will be addressed below. In some cases suggestions for addressing the specific comments are also presented.

1. Is the stated purpose of the assay clear?

The stated purpose of the aromatase assay and how it fits into the overall EDSP Tier-1 Battery is felt to be adequately described to a well-informed scientist.

However, without detailed scientific understanding of endocrinology and the entire androgen/estrogen/aromatase system, and aromatase's role in peripheral tissues (especially during development), the complete grasp of the purpose and significance of the assay is not possible. Currently superficial information is stated in Background paragraph 1, although 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. This information should be followed by a sentence or two stating that *in vitro*, microsomal aromatase activity is considered a good surrogate indicator of enzyme activity in aromatase expressing cells and sensitive tissues *in vivo*. Acceptable communication of this would be possible in two or three well-crafted sentences. Currently a concrete understanding of the relevance and purpose of the assay must be extracted from various portions of ISR including the background sections of the Executive summary and Sections 2.0 – 3.0 of the ISR.

2. Is the assay biologically and toxicologically relevant to the stated purpose?

As an individual component of the Tier 1 battery of tests, the aromatase assay is felt to be biologically relevant for the purpose of identifying environmental endocrine disruptors (EDC) that may act via inhibition of aromatase activity. The ability to reliably detect subsets of chemicals that influence activity of the aromatase enzyme (in this case, limited to inhibitors of aromatase enzyme activities), and thus potentially impact androgen/estrogen sensitive hormonal systems is a critical and biologically relevant component of the Endocrine Disruptor Screening Program (EDSP).

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 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.

3. Does the protocol describe the methodology of the assay in a clear, and concise manner so that the laboratory can:

a) comprehend the objective

The object of the protocol is straight forward; the current phrasing should be corrected to make this clearer to the reader.

A-6 1.0 Objective: currently reads: *“The objective of this protocol is to describe procedures for conduct of the aromatase assay as a Tier 1 screen using either human placental or recombinant microsomes.”*

Suggested that the highlighted phrase be changed to read: *“to conduct the aromatase assay”*

b) conduct the assay

Generally the protocol does a good job of describing the assays. It is estimated that with some modification, the protocol (appendix A) would allow a laboratory to conduct the assay.

There are a few typos and some technical difficulties that must be addressed. Importantly there are a few important details missing, and in places it is felt that there is too much “flexibility” allowed in the current protocol. Details regarding difficulties associated with the protocol and some points in the ISR are described, along with recommendations for correction, if applicable, are listed below.

A-6 2.1.1, sentence 3: *“...is usually supplied at a specific activity of 20-30 μ Ci/mmol.”*

This statement is believed to be inaccurate due to a typo that can be corrected by deletion of the highlighted μ (*Ci/mmol* is believed to be the correct unit).

Preferred alternative - the information regarding how the radiolabeled androstenedione is “usually supplied” could be deleted , there is no utility of this information to a specific

protocol. Minimal specific activity and purity requirements are stated, thus making this statement irrelevant.

A-7 2.2 Test Chemicals

In addition to the provided information for each test chemical, information regarding stability and date of expiration should be provided.

Test Chemical formulation/4-OH ASDN formulation – the options for chemical formulation in “buffer, absolute ethanol, or DMSO” is problematic. There are no set criteria for preparation of stock formulations. This point is not well addressed in either the ISR or the detailed protocol (Appendix A).

Importantly - 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 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.

Two recommendations are suggested:

1) The **Full Activity Control** should be described explicitly as containing the same vehicle and concentration of vehicle as the 4-OH ASDN positive control in the case of the positive control experiment described in Table 4 (A-13). It is also important to clarify

that this corresponds to the highest concentration of vehicle present in the concentration/response treatments (“Sample Type Conc 1-8).

2) An additional negative control group (i.e. **test chemical vehicle control**) should be added to every experimental assay (e.g. Table 5; A-15). This control would be composed of all assay components plus an amount of “test-chemical vehicle” equivalent to the highest concentration of vehicle present in any of the test-chemical treatment groups (Test Chem. Conc 1-8).

It is suggested that a “*universal solvent*” be adopted which is useful for the majority of anticipated test chemicals (as well as positive or negative control treatments as appropriate). Candidate chemicals incompatible with the “universal solvent” should be identified prior to analysis and a prescribed substitute solvent be used with the addition necessary controls.

As a related comment, 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.

Section 2.4.1 Human Placental Microsomes

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 be addressed.

Suggestions to address the potential influence of genetic variability on the finding from the placental aromatase assay:

1) The genotype (and potentially haplotype) of the CYP19 gene present in each placental preparation should be characterized – isolation and archiving of placental DNA, and sequencing of the CYP19 gene would be straight forward and rapid. **The collection, analysis, and archiving of this material (genomic DNA) and sequence information for each microsomal preparation is considered vital.**

2) Section 6.0 (A-14) paragraph 2 sentence 2 states “*A chemical shall be tested in three independent runs*”. While not done previously, three truly independent runs require the use of 3 different preparations of microsomes. Thus, it is suggested that microsomes from three different placental preparations be used. Regardless, the meaning of “independent run” as used in this sentence must be clearly defined.

General comments: In review of the ISR, DRP and other reference information, specific information regarding the starting amounts of placental tissue used was not found. It would be helpful in the protocol to have information regarding an acceptable scale (amount of starting material) for each preparation. Addition of information specifying an acceptable range of typical tissue wet-weights is considered useful as an aid for preparation planning and assay preparation standardization. Additional information regarding expected ranges of microsome yields, etc., should also be considered for inclusion. Those guidelines would aid in achieving predictable and consistently useful yields of microsomes and aromatase activities.

2.4.1.2, bullet 3 and 5: wash volume should be specified.

2.4.1.2, bullet 6: guidelines for volume of buffer for resuspending pellet should be specified.

2.4.1.2, bullet 7: specific guidelines regarding aliquot volumes and minimal acceptable stock protein concentrations should be specified.

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 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.

Note: The above suggested practice is used for the Human Recombinant microsomes (2.4.2.2, A-10), which are aliquoted into individual use vials based on estimates of protein content.

c) observe and measure prescribed endpoints

Sections 4.0 through 6.0 are clear and detailed, the measurement of $^3\text{H}_2\text{O}$ using liquid scintillation spectrometry is straight forward and is adequately described.

4.0, pg. A-12, second bullet: "...are presented in *Table 3*" should be corrected to "...are presented in *Table 2*".

6.0, last sentence paragraph 2, pg. A-14: the reference to Table 6 should read "*Table 5*".

d) compile and prepare data for statistical analyses

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).

7.2 Model Fitting and 7.6 Statistical Software – 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 (see comments in #7 below regarding “goodness of fit” obtained with each sigmoidal-model and incomplete dose/responses).

It is strongly recommended that the most recent version of a single statistical software package is adopted (e.g. Prism ver. 5). Convenience of use is not an acceptable justification for selection of a software package to use for critical data analysis – as noted in the ISR there are important differences in regression model fitting algorithms and capabilities between Prism ver. 5 and earlier versions of the software. Those changes directly impact the non-linear regression model-fitting used for the aromatase assay.

e) report the results?

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.

What additional advice, if any, can be given regarding the protocol?

None – detailed advice and recommendations were presented above.

4. Have the strengths and/or limitations of the assay been adequately addressed?

Strengths:

Although general, the most focused descriptions of the strengths of the assay are addressed in Section 1.0 Executive Summary, Background paragraph 1 (pg 1) and Section 3.3 (pg. 21-22). There does not seem to be an attempt to specifically highlight the strengths of the aromatase assay beyond the fact that it is a well-established and reliable assay. Thus, 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.

Limitations:

The limitations of the assay are addressed in summary form in section 3.4 (pg. 22).

The limitations of the assay's ability to only identify inhibitory effects, and the inability of the assay to distinguish the "nature" of inhibition are acknowledged.

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.

The final sentence of the Section 3.4 is felt to be irrelevant. This is not a limitation of the assay as it is proposed to detect EDC effects on aromatase; the activity of any other metabolic enzyme is clearly not being considered, and is thus not a limitation of the assay related to the proposed goals of these studies.

5. Were the (a) test substances, (b) analytical methods, and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?

a) Test substances: the reference chemicals (Table 6.1-1 and Table 12.1) used for validation of the aromatase assay is considered appropriate. As a set, they represent a number of different chemical classes with reasonably predictable effects on aromatase activity based on the supporting literature.

(b) Analytical methods: the analytical methods used to assess aromatase are well established and straight forward.

(c) Statistical methods: the statistical methods comparing variability in performance of the aromatase assay and the performance of the assay in different laboratories were appropriate to demonstrate the performance of the assay.

6. Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?

Regarding experimental variation, the assay is considered sufficiently repeatable and reproducible. However, it is clear that there were important differences in the quality of the results obtained from different laboratories. In particular, the results obtained from one of the laboratories were consistently much more variable than the other labs. From the documentation available, it appears that those problems resulted from undefined QC/QA problems within that laboratory. Because the four selected laboratories represent a population of “acceptable” laboratories (section 11.0) there is concern with the criteria used to establish “acceptable” laboratory designation.

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. The modified Lowry assay used for determining protein concentration is well-known as rather inaccurate and variable. As a result, the fact 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 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.

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.

7. With respect to performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?

The performance criteria for the assay are generally considered appropriate and reasonable based on the presented data.

11.2 Performance criteria for test chemicals (pg. 182) – regarding the examination of test data for outliers – a specific method and criterion for determining outliers should be established.

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.

8. Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?

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 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.

9. Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be used by the EPA to identify chemicals that have the potential to interact with the endocrine system.

Overall, the aromatase assay is considered a **critical** *in vitro* screening tool for use by the EPA to identify chemicals that potentially interact with the androgen/estrogen endocrine system.

Additional Comments:

Background is rather dated and has been carried over from previous reports included in the background material, some of those previous reports were completed many years ago – the addition of some updated background information regarding currently accepted understanding of fundamental hormone and EDC mechanisms and modes of action are encouraged.

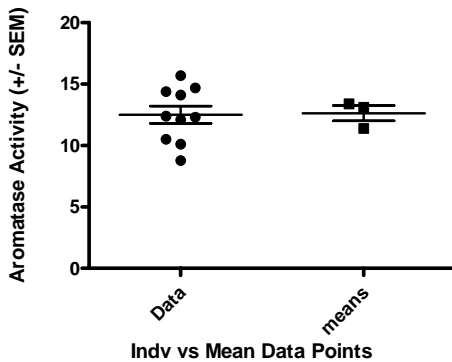
Section 8 and 9 (beginning around section 8.5, page 84 and continuing into section 9.0) contains a large number of typographical errors including Cyp19 being referred to as CYPL9.

Table 8.5-3 would be more valuable if the values for % inhibition mentioned in the text were included.

Supplemental information 1:

Recalculations of data presented in Table 8.2.2. (Prism Ver. 5.0)

Data	Means (reported)
14.40	13.1
14.70	13.4
10.10	11.4
12.10	
12.30	
15.70	
12.40	
14.10	
10.50	
8.78	



	Data	Means
Number of values	10	3
Minimum	8.780	11.40
25% Percentile	10.40	11.40
Median	12.35	13.10

75% Percentile	14.48	13.40
Maximum	15.70	13.40
Mean	12.51	12.63
Std. Deviation	2.238	1.079
Std. Error	0.7076	0.6227
Lower 95% CI of mean	10.91	9.954
Upper 95% CI of mean	14.11	15.31
Coefficient of variation	17.89%	8.54%
Sum	125.1	37.90

Reference: Olson, Sara H., Bandera, Elisa V., and Orlow, Irene Orlow. Variants in Estrogen Biosynthesis Genes, Sex Steroid Hormone Levels, and Endometrial Cancer: A HuGE Review. *Am J Epidemiol* 2007; **165**:235–245.

3.2 Laura Kragie Review Comments

Kragie BioMedWorks

POB 71091
Chevy Chase MD
708 601-1645 cell ph

Laura Kragie, M.D.
President & Chief Scientific Officer
lkragie@biomedworks.com

MEMO

January 9 2008

TO Laurie Waite
Eastern Research Group, Inc. (ERG)
110 Hartwell Avenue
Lexington, MA 02421-3136
Attn: Laurie Waite
E-mail: laurie.waite@erg.com or peerreview@erg.com
Ph 781-674-7362 or 781-674-7324

RE: CHARGE to PEER REVIEWERS for INDEPENDENT PEER REVIEW of the AROMATASE ASSAY as a POTENTIAL SCREEN in the ENDOCRINE DISRUPTOR SCREENING PROGRAM (EDSP) TIER-1 BATTERY

Background:

According to Section 408(p) of the EPA's Federal Food Drug and Cosmetic Act, the purpose of the EDSP is to: "develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate" [21 U.S.C. 346a(p)]. Subsequent to passage of the Act, the EPA formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a panel of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement the EDSP. Upon recommendations from the EDSTAC, the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormone systems as well as wildlife.

One of the test systems recommended by the EDSTAC was the placental aromatase assay. Its purpose in the Tier-1 battery is to provide a sensitive in vitro assay to detect chemicals that may affect the endocrine system by inhibiting aromatase, the enzyme responsible for the conversion of androgens to estrogens. Alterations in the amount of aromatase present or in the catalytic activity of the enzyme will alter the levels of estrogens in tissues and dramatically disrupt estrogen hormone action. EPA has chosen to validate two versions of the aromatase assay. The first version uses microsomes isolated from the human placenta. The other uses a human recombinant microsome.

Although peer review of aromatase assay will be done on an individual basis (i.e., its strengths and limitations evaluated as a stand alone assay), it is noted that the aromatase assay along with a number of other in vitro and in vivo assays will potentially constitute a battery of complementary screening assays. A weight-of-evidence approach is also expected to be used among assays within the Tier-1 battery to determine whether a chemical substance has a positive or negative effect on the estrogen, androgen or thyroid hormonal systems. Peer review of the EPA's recommendations for the Tier-1 battery will be done at a later date by the FIFRA Scientific Advisory Panel (SAP).

This peer review will focus on the scientific work EPA performed to validate the assays. Each peer reviewer is asked to focus his/her review on this issue and utilize the Integrated

Summary Report (ISR) as the vehicle for conducting this review. The review is not a critique or peer review of the ISR per se. Laboratory reports of the studies supporting validation and the Detailed Review Paper on aromatase are provided as background information.

Charge Questions to Peer Reviewers:

1. Is the stated purpose of the assay clear?

Yes, the stated purpose of the assay is clear. This is a screening tool to initially assess chemical compounds for their impact on estrogen formation.

2. Is the assay biologically and toxicologically relevant to the stated purpose?

The importance of the aromatase enzyme for estrogen formation and function in the mammalian organism is well reviewed in the ISR. The relevance of aromatase to reproductive function and assessment of toxic effects are described in the cited references Kragie et al 2002 and Kragie 2002.

The placental form of aromatase may be different from other isoforms that occur in tissues other than placenta, but it does suffice for this purpose of crude screening. Once the chemicals are classified in one of three categories, then more definitive studies can be performed by researchers to elucidate the compounds impact on biology. Some of these other tests may be included in the Tier 1 Battery.

It is essential, however, to select an assay that is cost-effective for **screening** the proposed 10,000 chemicals. A cell-based assay or HPLC-based assay would be prohibitively costly (about 10X higher than a non-chromatography based assay) if it were pursued instead in an attempt to achieve a higher confidence in the results. Therefore, I recommend that the very initial crude screening phase be done using the High Throughput BD Supersomes Aromatase Assay, which uses a fluorescent enzyme substrate (DBF), microtiter plates, fluorescence detection and perhaps 3 concentrations of assessed chemical: 1, 10, 100 micromolar. (Compounds are rarely relevant in the millimolar range, and their solvents become a dominant

effect in that range.) The compounds identified to be inhibitors would then go on for assessment with this EPA validated tritiated water method, using a full concentration curve to better define the IC50 value. The most potent inhibitors should be assessed first.

3. *Does the protocol describe the methodology of the assay in a clear, and concise manner so that the laboratory can:*

a) *comprehend the objective;*

Yes, the objective is clear. The assay will assess any compound, that upon acute exposure, will reduce the production of product (estrone) via detection of the reaction's product, water (scintillation count of tritium). This detection will occur regardless of mechanism of enzyme inhibition.

b) *conduct the assay;*

The method protocol is generally clear. However see advice given in Question 4.

c) *observe and measure prescribed endpoints;*

The full activity point of 100% is clearly understandable and achievable. However, the 0% point (bottom) is more difficult to establish. The scintillation counts are progressively diminished and therefore a reduction of the signal to noise ratio. In the assessed chemical inhibition curve, the lower half is more difficult to clearly establish due to the signal to noise issue and also the problems associated with solubility of chemicals at high concentrations, and the nonspecific effects of the high chemical concentrations.

d) *compile and prepare data for statistical analyses; and*

The proposed statistical method is standard for these assays and is appropriate. Because of the problems discussed in 3. (c), more leeway is given to the bottom parameter to establish the sigmoid curve that determines a chemical's IC50 value.

e) *report the results?*

Yes, the three categories for classification (inhibitor, equivocal, noninhibitor) are a feasible presentation.

4. *What additional advice, if any, can be given regarding the protocol?*

Econazole is metabolized by CYP450s in the placental microsomes. It is known CYP4503A inhibitor with submicromolar range IC₅₀. It was the most potent inhibitor of the tested series and any slight change in econazole concentration due to CYP3A metabolism will cause variability in the amount of tritiated aromatase activity measured. The variability should be less when using the recombinant aromatase microsomal preparation because recombinant aromatase P450 is enriched in those microsomes relative to other CYP450s. (SUPERSOME activity is catalyzed by human CYP19 that is expressed from human CYP19 cDNA using a baculovirus expression system. Baculovirus infected insect cells were used to prepare these microsomes. These microsomes also contain cDNA-expressed human P450 reductase. A microsome preparation using wild type virus is used as a control.) Econazole IC₅₀ value determined from this validation series was very consistent with the value determined using the HT SUPERSOME assay using DBF for enzyme substrate. See Kragie et al 2002. Very potent inhibitors require more precise assay procedure and practice; e.g., time, temperature, concentrations, and buffer. HT screening assay also recommends using insect cell protein to reduce the nonspecific binding of drug to apparatus that depletes the effective drug concentration exposed to aromatase enzyme.

The aromatase activity in the absence of any test substance was used as the benchmark (100 percent) activity. However, often there is need for a vehicle blank using the same solvent dilution of DMSO or ethanol, if greater than 1% final concentration in the assay.

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.

Redox conditions are sensitive to oxidation. Be aware of oxidation and keep tubes capped.

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 with stop procedure as compared to reaction start. Also, the last microsomes pipetted may be cooler in temperature than the initial aliquot pipetted. Technician needs to pay attention to timing and temperature.

5. *Have the strengths and/or limitations of the assay been adequately addressed?*

In general, yes the strengths and limitations are addressed and discussed.

6. *Were the (a) test substances, (b) analytical methods, and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?*

Yes, the test substances, analytical methods, and statistical methods chosen are appropriate for appraisal of assay performance. These methods are SOP for clinical diagnostic assay assessments.

7. *Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?*

These results are consistent with the variability inherent in bioanalytic assays that are used routinely in clinical practice. The results will improve with experience and practice of the technician.

8. *With respect to performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?*

The adjustment of the bottom criteria is a necessary improvement.

9. *Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?*

The ISR Dec 11 2007 version still needs a thorough proofing, especially the Tables, for accuracy and completeness. See general comment section for some instances of misinformation and typos.

10. *Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be used by the EPA to identify chemicals that have the potential to interact with the endocrine system.*

This in vitro aromatase assay meets the criteria for a screening tool to identify chemicals that may potentially interact with the endocrine system. The emphasis here is on **screening tool**, and that this method should not be used to definitively categorize a compound as reproductively toxic. It should be a first step in the evaluation process, because of its ease of use, short time course and overall safety and cost. 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.

The real utility of this screening tool is its use with a Rank Order of test compounds in a series of known potent and weak inhibitors. The Rank Order takes into account general variability of assay conditions that apply universally to the overall test set. The relative relationship of the compound to the known and previously tested substances, is the crucial information.

REFERENCES

Kragie L. Turner SD, Patten CJ, Crespi CL, Stresser DM. 2002 Assessing pregnancy risks of azole antifungals using a high throughput aromatase inhibition assay. **Endocrine Research** 28 (3): 129 –140

Kragie L. 2002 Aromatase in primate pregnancy: a review. **Endocrine Research** 28 (3): 121-128

EDITING CORRECTIONS ISR version Dec 11 2007

SECTION 1.0

P 16 Table 1.0 -01. IC50 values for positive control need concentration terms, or logIC50 terms (table 1.0-5 lists -7.3 to -7.0)

P 18 Table 1.0-3. Check these figures

Prochloraz

Placental 20.2 +/- 0.001.8 nM 026.9 +/- 0.003.1 nM

Dicofol

Placental 62.91 +/- 35.86 µM 24.14 +/- 3.72 µM 501 +/-489 µM 53.13 +/-16.56 µM 29.13 +/- 8.62 µM

P 10 Table 1.0-4. Full Activity and Background Control Criteria*

Need to clarify table to state it is for recombinant activity. Legend states different value for placental microsomes.

P 10 Supplemental Testing

“Ten chemicals are clearly noninhibitors, and no concentration-dependent response was observed for any of the noninhibitors These are Vinclozolin, Bisphenol A, Tributyltin, Diethylhexyl phthalate, Methoxychlor, Aldicarb, Flavone, Triadimefon, Imazalil, Apigenin, Ronidazole, Genestein, p,p'-DDE, Alachlor, Nitrofen, and Trifluralin.” This section needs correcting- 17 chemicals are listed, one repeated, and it includes the six chemicals listed as

inhibitors.

P 11 “The percent of control values for each reference chemical run and tube, along with the mean, SD, SEM, and CV of the percent of control across tubes within a run.” Correct sentence fragment.

P 26 correct concentration: **Table 4.6-1. Aromatase Assay Conditions**

[3H]ASDN 100 μ L 100 nM placenta 100 mM recombinant

p 79 missing concentration terms for IC50: “Based on the curve-fit of the percent of control aromatase activity values across six concentrations of 4-OH ASDN, the calculated IC50 values by run and laboratory are summarized in **Table 8.2-5**. The average \pm SEM IC50 values for Laboratories A, B, and C were 57.9 ± 5.9 , 47.3 ± 2.6 , and 81.1 ± 5.5 ; the percent CV values were 17.7, 9.6, and 13.4 percent, respectively. The overall task mean \pm SEM IC50 value was 62.1 ± 10.0 and the percent CV was 27.8 percent.” Missing also from table 8.2-5.

p 83 missing concentration terms for IC50: **Table 8.3-2. Aminoglutethimide IC50 and slope values**

p 85 missing concentration terms for IC50: **Table 8.3-4. Chrysin IC50 and slope values**

p 87 missing concentration terms for IC50: **Table 8.3-6. Econazole IC50 and slope values**

p 90 missing concentration terms for IC50: **Table 8.3-8. Ketoconazole IC50 and slope values**

3.3 Marion Miller Review Comments

Peer Review of Aromatase Assay

Peer Reviewer: Marion G Miller

Charge Questions

1. Is the stated purpose of the assay clear?

The purpose of the assay is clearly stated. The aromatase assay is one of a battery of assays developed for the Endocrine Disruptor Screening Program. The purpose of the assay is to screen for chemicals which have the capability of inhibiting aromatase, the enzyme responsible for conversion of androgens to estrogens. This assay is an alternative assay in the Tier 1 screening battery and is designed to detect chemicals that would inhibit estrogen biosynthesis. It is an *in vitro* assay that allows for rapid and relatively inexpensive screening of chemicals.

2. Is the assay biologically and toxicologically relevant to the stated purpose?

Aromatase catalyzes the conversion of androgens to estrogens. The rationale for inclusion of this assay as an alternative Tier 1 assay is based on the likelihood of a differential sensitivity of males and females to aromatase inhibition. Although both males and females require estrogen for reproductive health, the female is viewed as more susceptible to loss of estrogen biosynthetic capabilities due to the importance of estrogen in normal female reproduction. It is indicated (p 14, Table 2.4-2) that if studies were conducted in the male only, male animals may not be sufficiently sensitive to aromatase inhibition and decreased estrogen levels to allow effects to be detected. Although the extent of this gender difference is not documented in detail, the assay provides a useful first level *in vitro* screen for chemicals capable of inhibiting aromatase. Because the studies are conducted *in vitro*, confounding effects of whole animal physiology and feedback mechanisms as well as the absorption, distribution, metabolism and excretion characteristics of the individual chemical are not considered. This can be viewed both as a strength and a weakness since a direct effect on the enzyme will be readily measured but the relevance of that effect in the whole animal is not tested. An additional application for the assay is to supply more detailed mechanistic information about effects on steroidogenesis which may have been detected in the Tier 1 *In Vitro* Steroidogenesis Assay. However, in the absence of

details about the *In Vitro* Steroidogenesis Assay the utility of the aromatase inhibition assay to provide additional information is not clear.

3. Does the protocol describe the methodology of the assay in a clear, and concise manner so that the laboratory can:

(For this section the reviewer specifically evaluated the assay protocol in Appendix A as this represents the summation of findings from protocol development)

a) comprehend the objective;

Objective to measure aromatase activity is indicated

b) conduct the assay;

Overall the assay is well described but some points could be clarified.

- 1) Page A -6. Section 2.1.1. Androstenedione is usually supplied with a specific activity of mCi/mmol rather than uCi/mmol as indicated This also explains the contradiction between sentences 4 and 5 in this section.
- 2) What buffer is used to make stock solutions (section 2.1.3)? – presumably the 0.1M phosphate buffer indicated in section 2.5.1 but this could be specified.
- 3) In section 2.1.3 what does “record the weight of each component added” refer to?
- 4) Timing for use of microsomes should be defined rather than recommended (section 2.4.1.3).
- 5) Why is propylene glycol added to the assay?

c) observe and measure prescribed endpoints;

The method for measurement of aromatase activity is well described in detail. (Section 4, Pages A-11 &12.) Typographical errors with confusion of the present and past tense could be corrected. In section 5, Page A-13 where the positive control assay is described, it is indicated that the minimum level of aromatase activity in the full activity control will be 0.100 nmol/min/mg protein. However, this value refers to the minimum level only for the recombinant microsomes. The minimum level for the placental microsomes is 0.03 nmol/min/mg protein and this should also be indicated.

d) compile and prepare data for statistical analyses; and

Methodology appears appropriate. Use of % control data reduces variability between preparations yet still provides data about the potency of the inhibitor. Spread sheet is supplied. Commercially available statistical software is recommended.

e) report the results?

The IC₅₀ values are generally reported as log numbers. Reporting these as linear values would give a better appreciation of relative potencies.

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. However, 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.

What additional advice, if any, can be given regarding the protocol?

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.

4. Have the strengths and/or limitations of the assay been adequately addressed?

The limitations of the assay are indicated (p 22, 232) as inability to detect induction of the aromatase enzyme, lack of information about the nature of any inhibitory response, denaturation of the enzyme (Page 232, “receptor” should be replaced with “enzyme”) with subsequent

identification of a false positive, inability to test chemicals that have limited water solubility, and the lack of xenobiotic metabolizing activity with consequent inability to detect metabolites with inhibitory activity. To detect either enzyme induction or the presence of a biologically active metabolite would require the use of whole cell systems or whole animals. The level of complexity would be significantly increased and the goals of a rapid, inexpensive screen would be more difficult to meet. The summary document recognizes these limitations and addresses them adequately. 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. Overall, the strength of the assay is that it is quick, simple, inexpensive, relatively robust, gives reproducible data, and allows detection of chemicals that directly inhibit aromatase *in vitro*.

5. Were the (a) test substances, (b) analytical methods, and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?

Test substances. Substances selected as test substances represented a diverse range of chemical structures as well as applications. 10 test substances (originally there were 11 but lindane, a negative control was dropped) were used in an interlaboratory validation study. An additional 16 substances were tested in the lead laboratory for aromatase inhibition. The selected substances were comprised of both inhibitors and non inhibitors of aromatase and provided information about the reliability of the assay and its ability to detect aromatase inhibitors.

Analytical methods. The three most important measurements before addition of test substance are the full enzyme activity, background activity and the inhibitory response to the positive control, 4-hydroxyandrostenedione. The purity of the radiolabelled androstenedione substrate is

important particularly in this assay as values for background activity would be expected to change dependent on purity of the starting radioactive material. 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.

Microsomal cytochrome P450 levels were initially measured to give an indicator of total P450. However, this adds little information about aromatase activity in a microsomal preparation as the aromatase isoform represents only one of many P450 isoforms. For the aromatase assay, the methods used were established in a rigorous manner with measurement of the protein- and time-dependence of enzyme activity to ensure appropriate substrate concentrations and incubation times.

Statistical methods. Less stringent criteria for coefficient of variation in the lowest percent control values are appropriate when the levels measured are so low as to be hardly above background. Use of simple % inhibition to classify chemicals as inhibitors or non-inhibitors instead of using confidence intervals which may overlap when the data is of poor quality, provides an approach where there is a decreased likelihood of misclassification.

6. Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?

A major source of variability in the interlaboratory validation seems to have arisen from use of a protein standard curve where the values for the protein sample to be measured fell below the levels measured by the standard curve. From an analytical perspective, the optimal situation is where the unknown sample is bracketed by standards and it is somewhat surprising that extrapolated values were used to obtain the protein concentration. This does to some extent explain variability seen in results from one of the laboratories. However, it should be noted that overall the assay was quite robust and despite interlaboratory variability, IC₅₀ values were generally similar.

7. With respect to performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?

Performance criteria are a key aspect of establishing consistent a methodology and generating reproducible data. Minimum enzyme activity in the placental and the recombinant preparations are defined. For this assay the full enzyme activity control is particularly important . This value is used as the 100% value relative to which the effect of the inhibitor is compared and the concentration dependence of the enzyme inhibition is plotted graphically. Performance criteria are $\pm 10\%$ for the full enzyme activity and represents a reasonable variability for measurement of enzyme activity. If full enzyme activity is highly variable such that IC50 values cannot be obtained, the assay should be repeated. For the 4-hydroxyandrostendione positive control the outlined performance criteria (Page A-13) are reasonable to ensure that the assay is functioning as it should. 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.

8. Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?

As indicated earlier, the data interpretation criteria for inhibitor classification defined an inhibitor when more than 50% inhibition was achieved and as a non-inhibitor when inhibition was not greater than 75% inhibition. This allowed for easy classification of inhibitor and non-inhibitors. However, the equivocal situation where inhibition is 50-75% is not adequately addressed 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. If the highest concentration data points are to be used, there is a greater possibility that enzyme denaturation or other non specific effects rather than enzyme inhibition has occurred. The limitations of this default approach should be addressed.

9. Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be used by the EPA to identify chemicals that have the potential to interact with the endocrine system.

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.

Specific Comments by Section

Executive Summary.

Page 7 Table 1.0-3 The value reported for In Vitro with Dicofol and the placental microsomes is much higher than for the other 3 laboratories. Is this a typographical error? Also the overall task group mean for Dicofol does not reflect the mean for the 4 labs even if the In Vitro measurement is changed from 501 to 50.1 or 5.01 or 0.501 assuming that the value in the table represents a misplaced decimal point.

Page 8. Use of terms can be somewhat confusing when they are used without definition. The use of portion to represent beginning and end is not clear in the executive summary although it is apparent later in the document. Similarly, for the use of “full enzyme activity”. Also, the data and treatments in figures 1.0-1 and 1.0-2 are not adequately explained in the executive summary.

Main Document

Page 27. Table 4.6-1. Tritiated androstendione should be the same concentration (100nM) for both assays

Page 61. Isotope is I₁₂₁

Explain abbreviations, for example QC and RE on page 85

Page 90, line 2 units should be ug not g/ml

Page 92 and 135, Define CR and RC

Page 127, Table 9.6-3. Econazole IC₅₀ values in the table are in uM with lots of decimal places whereas in the text nM values are indicated.

Page 175. Define AG and SNLR.

Page 176. What is WA 4-17?

Minor editorial changes are not commented upon

3.4 Safa Moslemi Review Comments

Comments :

This report « Integrated Summary Report or ISR on Aromatase » summarizes and synthesizes the information compiled from the validation process in order to propose a protocol on Aromatase Assay as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery. Both the human placental microsomal assay and the recombinant assay using human recombinant microsomes from Gentest (Human CYP19 + P40 reductase SUPERSOMES) were validated and their equivalence demonstrated. You find here after the answers and comments to the charge questions.

Charge Questions:

1. Is the stated purpose of the assay clear?

The purpose of assay is well presented and consist to propose a validated test using human placental microsomal preparation or human recombinant microsomes to evaluate interference of chemicals with endocrine system by inhibiting aromatase (Tier I) the key enzyme responsible of irreversible conversion of androgens in estrogens. Synthesis of estrogens occur, besides gonads and placenta, in many non reproductive tissues of several vertebrate and species of both sexes. Chemicals testing positive in Tier I would be further evaluated in Tier II which will aim to characterize the adverse effects resulting from that interaction and the exposures required to produce them.

2. Is the assay biologically and toxicologically relevant to the stated purpose?

Since estrogens are involved in the homeostasis of many of tissues and organs in different species, therefore, evaluation of its synthesis is relevant of both reproductive and non reproductive systems. However, the proposed protocol could not give any information on

toxicological system. To reach this goal, toxicological test using in vitro cell culture should be carried out.

- 3. 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?**

Protocol is well described and the methodology presented in a comprehensible manner allowing the reader to follow easily all steps cited above.

What additional advice, if any, can be given regarding the protocol?

In order to improve protocol, the following advices are proposed :

- 1) Add the substrate androstenedione 4 μ M 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 choc 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 tow supplementary extractions by solvent, this help to reduce the time of experimentation.

- 4) Add the formula for the calculation of the specific aromatase activity in $\text{nmol.mg protein}^{-1} . \text{min}^{-1}$ by expressing all parameters used such as, background radioactivity, specific activity of the substrate, time of incubation, protein concentration and finally the correction for the % of the specific radio-labelling at beta position of C1 of substrate.

4. Have the strengths and/or limitations of the assay been adequately addressed?

The availability of human placenta, the facility to prepare microsome as source of aromatase activity and the high sensitivity, reproducibility and rapidity of tritiated water assay all these points are well quoted in IRS and place the proposed protocol as the most relevant for in routine evaluation of chemicals on aromatase activity, a crucial target of endocrine disruption. Besides limitations cited in ISR such as; false negative, lack of metabolizing enzymes, and lack of induction and/or inhibition of aromatase expression, this assay , as conducted, can not show chemicals acting in synergism when incubated in combination. Indeed, we recently showed that substances that have no visible aromatase inhibition alone, at 20 μM , become aromatase disruptors (or even cytotoxic) up to 50% inhibition in combination from 4-10 μM demonstrating that substances together may more easily clutter up the active site or alter the enzyme. Chemicals also often present a possible bioaccumulation, and/or indirect actions on signalling pathways. Since organisms are always exposed to mixtures of chemicals, all these issues become crucial in order to evaluate their effects and actions on human health (Benachour et al, Toxicol Appl Pharmacol , 2007).

5. Were the (a) test substances, (b) analytical methods, and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?

The choice of chemicals should be better justified (see blow, editorial comments, point 2). The parameters used for analysis and run comparison such as IC50 (or EC50) values, minimum inhibition (top), maximum inhibition (bottom), and Hilleslope all are appropriate and comprehensible for the majority of scientists. Also, selected statistical parameters such as SD, SEM, CV and analysis of variance (ANOVA) are among the most used by biochemists and biologists which offer high performance and reflect assay

efficacy, sensitivity and variability between runs and chemicals when comparison performed.

However, it is generally preferable to evaluate K_i which is the dissociation constant of enzyme/inhibitor complex and reflect enzyme affinity for inhibitor (or chemical) and gives more precision than IC_{50} value. IC_{50} value, showing inhibition efficiency, is used because it is more rapid and easy to perform. Thus, comparison between constants (K_i values) of different laboratories is easier to make than comparison of IC_{50} values (see page 60 for the variability of IC_{50} values between ISR and literature for ketoconazol and econazole). Indeed, IC_{50} values depend on parameters used for its determination which are different in literature (temperature, protein and substrate concentration, time of incubation etc). For instance, if you chose to use a lower concentration of substrate, it will take a little concentration of inhibitor to compete for 50% of the activity. That's why one should be certain to work in saturation condition of substrate (at least at $10 \times K_m$) for the determination of IC_{50} value and this is not the case in proposed protocol which use 100 nM of androstenedione while its K_m value is about 39 nM. Actually, IC_{50} might be under estimated in the present protocol.

6. Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?

Absolutely, the assay is sufficiently reproducible as demonstrated by statistical analysis and by fixing about 15% of CV and 95% of confidence intervals for the fitted curve and estimated parameters.

7. With respect to performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?

The performance criteria for the full activity control (0.100 nmol/mg.min) and the background control (1% of full activity control) provide reference values to the testing laboratory to ensure the detection of both strong and weak inhibitors. Determination of 80% of tolerance interval with 95% confidence guarantee an acceptable variation of data. However, I think that specialists are in position to comment this portion of study.

8. Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?

It is not always easy to understand all these interpretations criteria. This is because I am not statistician. However, if one look at the table 11.3-2 for the adopted data interpretation criteria (page 185), every thing being fortunately clear and comprehensible for the classification of inhibitors.

9. Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be used by the EPA to identify chemicals that have the potential to interact with the endocrine system.

1) As cited in IRS, one of the weakness of the proposed protocol that it can not predict metabolising 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 inhibit aromatase in JEG3 cells (Nativelle-Serpentini et al, 2003). So, a false negative should not systematically be deleted from the next step of evaluation.

2) 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).

3) 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 substrates being used preferentially by aromatase as we previously showed that 19-norandrogens are aromatised 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.

References

Benachour N, Moslemi S, Sipahutar H, Seralini GE, Cytotoxic effects and aromatase inhibition by xenobiotic endocrine disrupters alone and in combination. *Toxicol Appl Pharmacol.* 2007, 222(2):129-40.

S. Moslemi, P. Auvray, P. Sourdain, M. A. Drosowski and G.-E. Seralini, Structure-function relationships for equine and human aromatases : a comparative study. Annals of the New York Academy of Sciences. 1998, 839 : 576-577.

Nativelle-Serpentini C, Richard S, Seralini GE, Sourdain P, Aromatase activity modulation by lindane and bisphenol-A in human placental JEG-3 and transfected kidney E293 cells. Toxicol In Vitro. 2003, 17(4):413-22.

Editorial comments:

- 1) Page 2, the last sentence of first paragraph in “data analysis”, “*Both types of inhibitors have been tested in the validation program*” is not appropriate. Indeed, only nonylphenol was tested for kinetic inhibition which demonstrated a competitive inhibition with Lineweaver- Burk representation while the secondary plots indicated a small contribution of another inhibition type. Therefore, only a competitive inhibitor was demonstrated and this sentence should be modified.
- 2) Page 2, the first paragraph of “reference and supplementary chemicals”, the choice of the compounds is not clearly justified. It would be better to add “aromatase inhibition” in the sentence “*Eleven reference chemicals were selected on the basis of data on aromatase inhibition in the scientific literature and*”
- 3) Page 3, in “protocol optimisation” 6th sentence “..... *cytochrome P450 content*” should be deleted since this evaluation will be omitted from the final protocol.
- 4) Page 5 table 1.0-1, add unity nM for IC50.
- 5) Page 5, table 1.0-2 and other tables in overall ISR the presence of thick horizontal or vertical lines is not justified, please correct them.
- 6) Page 7, Table 1.0-3, please check the values for Prochloraz obtained by RTI and BATTELLE in placental preparation, the values 20.2+/-0.001.8 nM and 026.9+/-0.003.1 nM are unrealistic.
- 7) Page 10, in “testing the methods-supplementary testing” there is a confusion between activity and inhibition. For example, in the 4th line of paragraph “*In no case was the most **activity** (instead of inhibited) level between 75 percent and 50 percent of control, which Five of the six inhibitors exhibited maximal **activity** (instead of inhibition) of less than 20 percent of control; the other inhibitor (nitrofen) had a maximal **activity** (instead of inhibition) of about 32 percent of control*”. Moreover, in the same paragraph

in sentence “*Ten chemical are clearly noninhibitors, and.....These are Vinclozolin, BisphenolA, Tributyltin, Diethylhexyl phthalate, Methoxy chlor, Aldicarb, Flavone, Triadimefon, Imazalil, Apigenin, Ronidazole, Genestein, p,p’-DDE, Alachlor, Nitrofen, and Trifluralin*”. Should be deleted from this sentence the six inhibitor chemicals (BisphenolA, Flavone, Triadimefon, Imazalil, Apigenin, and Nitrofen) and Ronidazole which was repeated twice.

- 8) Page 12, in “Tiered approach” the sentence “*A negative result in Tier I would be sufficient to put a chemical aside as having low to no potential to cause endocrine disruption*” should be revised since some chemicals considered as negative with placental microsomal system (as Lindane) might have an effect in other system like in JEG3 cells (Nativelles-serpentini et al, 2003) or even their effect being visible when they incubated in combination (Benachour et al, 2007). So, once again, the false negative is really a significant problematic of microsomal system.
- 9) Page 25, in section 4.3.1 Human placental microsomes, should be precised in the first sentence “*Placenta **From healthy and non- smoking women** are obtained*” These two conditions (healthy and non-smoking) are quoted in some provided materials (references 1-14) and should be specified in final report as a guarantee for technicians working on aromatase assay and to provide a highly aromatase specific activity in prepared microsomes since it is known that smoking could alter aromatase.
- 10) Page 27, Table 4.6-1, correct final concentration of [3H]ASDN in the recombinant assay to 100 nM instead of 100 mM.
- 11) Page 34, table 6.1-1, for Lindane, in “*basis for selection*”, delete “not” since this substance inhibit aromatase in JEG-3 test for the cited reference (Nativelle-Serpentini et al, 2003). That’s why I think that Lindane should be deleted as negative control in overall IRS. Indeed, the value of background is sufficient for this purpose. In the same table, please add year which was omitted for certain references. In same table, for Chrysin in “test system”, what means “H adip”? Please add comments in legends of this table.
- 12) Page 40, in “Linear production of product” 4th line, correct **0.004** mg/mL instead of **0.005** mg/mL (human recombinant microsomes).
- 13) Page 68 (table 8.2-5), page 72 (table 8.3-2), page 74 (table 8.3-4), page 76 (table 8.3-6), page 79 (table 8.3-8), add unity for IC50.

- 14) Delete page 89 from paragraph “Recombinant microsoms (Human)” until page 92 table 9.0-4 because of repetition (see page 84, from 8.5-1 protein concentration until page 87 section 8.5-4).
- 15) Page 103, tables 9.2-3 to 9.2-6 are not necessary and the text gives enough information.
- 16) Page 115, last sentence “The overall task group mean +/- (instead of ::) SEM”
- 17) page 121, tables 9.5-3 to 9.5-6 are not necessary and the text gives enough information.
- 18) Page 146, line 2 delete “L” from μLM .
- 19) Page 162, at the beginning of third paragraph, Table **10.3-1** instead of 10.3-2.
- 20) Page 175, last paragraph and at the beginning of page 176, check the mean of K_m and V_{max} values (37.1 nM and 0.334 nmol/mg/min) which are different from those reported in table 10.6-1 (38.9 nM and 0.351 nmol/mg/min). This is also the case of K_i for nonylphenol (6.83 μM in the text different from 8.63 μM in the table).

3.5 Thomas Sanderson Review Comments

A Peer Review of the Validity of the Aromatase Assay as an Appropriate Tier 1 Screening Tool for Endocrine Disruptors

By J Thomas Sanderson

1. Is the stated purpose of the assay clear?

The Integrated Summary Report (ISR) states that as part of a battery of *in vitro* and *in vivo* tier 1 screening tools, the placental and recombinant microsomal aromatase assays are intended to determine whether chemicals have the ability to inhibit the catalytic activity of aromatase. Within this limited definition of the purpose of the assay, the stated purpose is clear.

2. Is the assay biologically and toxicologically relevant to the stated purpose?

To a degree it is. The rationale for concern about chemicals that inhibit aromatase is that such chemicals would result in reductions in endogenous estrogen concentrations in exposed organisms. As outlined in the ISR, biological and toxicological consequences would be numerous, including disruption of reproductive cycle and pregnancy in females, sperm

production/capacitation in males and possible behavioral effects in both sexes. Estrogens are important for bone homeostasis, growth and differentiation of numerous tissues and have modulatory effects on the immune system, and many other systems in the human body regardless of sex. It should be pointed out strongly that, although discussed to a certain degree in the ISR, estrogens are not strictly female hormones and in fact have very crucial functions in both sexes, whether it concerns sexual development or numerous basic functions unrelated to sex. Dependent on the sex and on the extent of inhibition of the aromatase enzyme deleterious effects can be very diverse. Something not addressed in the documentation is the following: 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.

3. Does the protocol describe the methodology of the assay in a clear and concise manner so that the laboratory can:

a) comprehend the objective?

The objective is clearly outlined in sections 1 and 2 of the ISR.

b) conduct the assay?

The protocol is clearly described in section 4 and appendix A of the ISR.

c) observe and measure prescribed endpoints?

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?

d) compile and prepare data for statistical analyses?

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 and included as part of the 'assay package'.

e) report the results?

Straightforward, other than several minor items in the comments that follow below.

What additional advice, if, any, can be given regarding the protocol?

I have several comments that should be considered concerning the protocol as described in Appendix A of the ISR.

Final solvent concentration:

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.

Microsome preparation:

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°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?

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.

Protein determination:

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.

It is surprising the assay should be performed using such large volumes, quantities and in traditional cuvettes. The general availability of absorbance plate readers has allowed for dramatic miniaturization of such assays. The assay could easily be performed using volumes 5-10 times less than those described in the protocol, thus allowing for the use of spectrometer-suitable multi-well plates of anywhere from 24-96 well formats. This would greatly enhance the efficiency of the assay (faster and using less material).

Aromatase assay:

On main question here is why the tritiated water-release protocol was 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 aromatase assay as described is performed in test tubes. I would have thought that the assay could easily be down-scaled to far smaller 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

Is the addition of propylene glycol necessary? It increases the organic solvent burden of the reaction mixture disproportionately compared with all the other components including solvent used for test chemicals and may not be essential to the performance of microsomal enzyme assays.

Semantically it is more appropriate to express the catalytic activity of aromatase when determined using the tritiated water assay as *pmoles of androstenedione converted per time unit per quantity of protein*, rather than amount of *estrone* formed, because estrone is not measured. 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).

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 analyses.

4. Have the strengths and/or limitations of the assay been adequately addressed?

The strengths and weaknesses of the placental microsomal assay have been discussed thoroughly in reference 1 (Final Detailed Review Paper on Aromatase). In section 5 of this document the assay is compared to several cell-based assay systems and deemed a more straightforward and better characterized assay than the cell-based ones. The weaknesses described are exhaustive and have taken all aspects of the placental microsomal assay in to consideration. 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 literature that would be equally suitable as tools for screening inhibitors but would also have the added possibility to detect inducers. Regardless of the difficulties in the interpretation of the relevance of inductions of aromatase activity/expression in cell-based systems, the crude observations would be readily obtained during screening at no extra effort and this information would be available to future investigators to be further studied if deemed of importance.

5. Were the (a) test substances, (b) analytical methods and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?

The wide range of compounds selected was an appropriate choice for validation of the assay. The test is conducted in such a manner that no enzyme inhibition kinetic properties can be determined. In other words the nature of the inhibition, competitive *versus* non-competitive or mixed-type inhibition will not be known. The protocol could, however, easily be adapted to obtain such information if desired. Inhibition curves would need to be produced in the presence of various concentrations of ASDN substrate.

It should be pointed out that in section 5.1 on page 29 of the ISR competitiveness is erroneously equated to reversibility. Inhibitors that bind to sites other than the catalytic site may produce non-competitive or mixed-type inhibition kinetics, but this does not mean that the inhibition is irreversible, only that increasing the substrate (ASDN) will not restore catalytic activity by displacing the inhibitors from the catalytic site. Inhibition can still be reversible with restoration of original enzyme activity once the inhibitor is eliminated through, for example, metabolism/elimination, as long as the interaction with the 'other site' is not covalent (=mechanism based inhibition = irreversible).

Concerning the estrone formation analyses - Section 7.5.3:

The observation that the tritiated water-release assays produces aromatase activities (amounts of $^3\text{H}_2\text{O}$) 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).

6. Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?

I have good confidence in the repeatability, reproducibility and overall reliability of the placental microsomal aromatase assay as a test system for aromatase inhibitors. An overall coefficient of variation of less than 30% is quite acceptable for an *in vitro* bioassay. Within-run variabilities reported to be as low as 5-15% are also very respectable.

7. With respect to the performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?

The performance criteria are reasonable and based on common sense and practice. 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.

8. Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?

They are clear. 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 that will be interpreting the aromatase assay results.

9. Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be use by the EPA to identify chemicals that have the potential to interact with the endocrine system.

As a bioassay to identify compounds that have the capability, at least in vitro, and in a very simplified enzymatic preparation, to inhibit aromatase activity, this protocol fulfills its objective. As a critical reviewer, I would say that within the very limited constraints of the objective of the bioassay it is useful, but have some concerns about its limitations. The rationale for developing a bioassay for effects of chemicals on aromatase is the fact that this enzyme plays a key role in the local production of estrogens in many tissues in the body and is involved in many essential processes throughout (unborn) life. However, the assay only covers inhibitors. It is well established that increased aromatase expression and estrogen production in tissues is associated

with various pathologies including endocrine cancers. This entire facet of the endocrine disruption paradigm and interest in aromatase as a target for endocrine disruptors is eliminated from the final proposed bioassay and is an important loss.

A considerable amount of resources has been spent on developing and validating the placental microsomal aromatase assay. It would have seemed within the bounds of possibility to develop one or more cell-based assays that would cover a more diverse range of aspects of aromatase function to more fully describe the ability of chemicals to interfere (inhibition/induction/downregulation) with this important endocrine endpoint.

The ISR and in more detail the Final Detailed Review Paper on Aromatase discusses several other potential candidate bioassays for the detection of interferences with aromatase but deems them too complicated, uncharacterized or otherwise limited to be useful as screening tools. It is my opinion that this is an unnecessarily missed opportunity.

Firstly, the perceived limitations of cell-based assays for effects on aromatase mention low basal aromatase expression, potential cytotoxicity at high concentrations of test chemicals and possible biotransformation to (in)active metabolites. A fresh look at, and interpretation of these perceived limitations could equally well transform them into advantages. For example, the fact that certain chemicals are cytotoxic at higher concentrations may be an indicator that the limit of physiological relevance has been reached. The fact that higher concentration (above the 100 micromolar range) may be attained in microsomal fractions is generally of not of great interest on a toxicological level. The additional concern in cell-based assays that some compounds may be too lipophilic to cross cell membranes could be seen differently. Would it not in fact be of importance and very relevant to know whether a chemical that appears to inhibit aromatase in microsomes could even enter a cell in the first place, is soluble in cell culture medium (crystallization is easily observed under a microscope) or is not rapidly metabolized to more or less potent metabolites? The future of relevant bioassays is one that provides an integrated more fully developed picture of the biological/toxicological activity of a chemical. Incorporating bioactivation, bioavailability, different mechanisms of action on the endpoint in question (aromatase activity) are essential components of such an approach. It is my fear that the

placental microsomal assay for the screening of effects on aromatase activity (strictly limited to inhibition in isolated microsomal fractions) may become outdated in a fairly short term.

I do not underestimate the complexities involved in the full validation of cell-based assays compared with the proposed microsomal assay(s). However, the greater quality and diversity of the information derived from such assays would, in my opinion, outweigh the concerns about their complexity. It has been made clear in the supporting documentation that cell-based assays perform well on their ability to identify inhibitors of aromatase, albeit with somewhat less sensitivity than the proposed placental microsomal assay. This is readily explained by the fact that chemicals encounter more barriers to reach their target in intact cells than in microsomal fractions that have been membrane-disrupted, concentrated and treated with co-solvents such as propylene glycol. Additionally, metabolism may play a greater role in cell-based assays. The question to be considered is: are these aspects of cell-based assays really disadvantages or do they, in fact, help us in providing a far more relevant representation of what endocrine disrupting chemicals may do to the enzyme aromatase in exposed organisms?

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.

References

- Kao, Y. C., Zhou, C., Sherman, M., Laughton, C. A., and Chen, S. (1998). Molecular basis of the inhibition of human aromatase (estrogen synthetase) by flavone and isoflavone phytoestrogens: A site-directed mutagenesis study. *Environ Health Perspect* **106**, 85-92.
- Lephart, E. D., and Simpson, E. R. (1991). Assay of aromatase activity. *Methods Enzymol* **206**, 477-483.

- Mindnich, R., Moller, G., and Adamski, J. (2004). The role of 17 beta-hydroxysteroid dehydrogenases. *Mol Cell Endocrinol* **218**, 7-20.
- Pasqualini, J. R. (2004). The selective estrogen enzyme modulators in breast cancer: a review. *Biochim Biophys Acta* **1654**, 123-143.
- Sanderson, J. T., Hordijk, J., Denison, M. S., Springsteel, M. F., Nantz, M. H., and Van Den Berg, M. (2004). Induction and Inhibition of aromatase (CYP19) activity by natural and synthetic flavonoid compounds in H295R human adrenocortical carcinoma cells. *Toxicol Sci* **82**, 70-79.
- Sanderson, J. T., Seinen, W., Giesy, J. P., and van den Berg, M. (2000). 2-Chloro-s-triazine herbicides induce aromatase (CYP19) activity in H295R human adrenocortical carcinoma cells: a novel mechanism for estrogenicity? *Toxicol Sci* **54**, 121-127.
- Vihko, P., Harkonen, P., Oduwale, O., Torn, S., Kurkela, R., Porvari, K., Pulkka, A., and Isomaa, V. (2003). 17 beta-hydroxysteroid dehydrogenases and cancers. *J Steroid Biochem Mol Biol* **83**, 119-122.
- Waxman, D. (1988). Interactions of hepatic cytochromes P-450 with steroid hormones: regioselectivity and stereospecificity of steroid metabolism and hormonal regulation of rat P-450 enzyme expression. *Biochem. Pharmacol.* **37**, 71-84.

Appendix A

CHARGE TO PEER REVIEWERS

CHARGE TO PEER REVIEWERS

for

INDEPENDENT PEER REVIEW OF THE AROMATASE ASSAY AS A POTENTIAL SCREEN IN THE ENDOCRINE DISRUPTOR SCREENING PROGRAM (EDSP) TIER-1 BATTERY

December 10, 2007

Background:

According to Section 408(p) of the EPA's Federal Food Drug and Cosmetic Act, the purpose of the EDSP is to:

develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate [21 U.S.C. 346a(p)].

Subsequent to passage of the Act, the EPA formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a panel of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement the EDSP. Upon recommendations from the EDSTAC, the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormone systems as well as wildlife.

One of the test systems recommended by the EDSTAC was the placental aromatase assay. Its purpose in the Tier-1 battery is to provide a sensitive *in vitro* assay to detect chemicals that may affect the endocrine system by inhibiting aromatase, the enzyme responsible for the conversion of androgens to estrogens. Alterations in the amount of aromatase present or in the catalytic activity of the enzyme will alter the levels of estrogens in tissues and dramatically disrupt estrogen hormone action. EPA has chosen to validate two versions of the aromatase assay. The first version uses microsomes isolated from the human placenta. The other uses a human recombinant microsome.

Although peer review of aromatase assay will be done on an individual basis (i.e., its strengths and limitations evaluated as a stand alone assay), it is noted that the aromatase assay along with a number of other *in vitro* and *in vivo* assays will potentially constitute a battery of complementary screening assays. A weight-of-evidence approach is also expected to be used among assays within the Tier-1 battery to determine whether a chemical substance has a positive or negative effect on the estrogen, androgen or thyroid hormonal systems. Peer review of the EPA's recommendations for the Tier-1 battery will be done at a later date by the FIFRA Scientific Advisory Panel (SAP).

This peer review will focus on the scientific work EPA performed to validate the assays. Each peer reviewer is asked to focus his/her review on this issue and utilize the Integrated Summary Report (ISR) as the vehicle for conducting this review. The review is not a critique or peer

review of the ISR per se. Laboratory reports of the studies supporting validation and the Detailed Review Paper on aromatase are provided as background information.

Charge Questions:

Your review and comments shall be directed to each of the following questions:

1. Is the stated purpose of the assay clear?
2. Is the assay biologically and toxicologically relevant to the stated purpose?
3. 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?

What additional advice, if any, can be given regarding the protocol?

4. Have the strengths and/or limitations of the assay been adequately addressed?
5. Were the (a) test substances, (b) analytical methods, and (c) statistical methods chosen appropriate to demonstrate the performance of the assay?
6. Considering the variability inherent in biological and chemical test methods, were the results obtained with this assay sufficiently repeatable and reproducible?
7. With respect to performance criteria, were appropriate parameters selected and reasonable values chosen to ensure proper performance of the assay?
8. Are the data interpretation criteria clear, comprehensive, and consistent with the stated purpose?
9. Please comment on the overall utility of the assay as a screening tool described in the introduction of the ISR to be used by the EPA to identify chemicals that have the potential to interact with the endocrine system.

Appendix B

INTEGRATED SUMMARY REPORT

[Integrated Summary Report for Validation of the Aromatase Assay as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery \(PDF\)](#) (223 pp, 2.7M)

Appendix C

SUPPORTING MATERIALS

[Reference 1. Final Detailed Review Paper on Aromatase Assay \(PDF\)](#) (93 pp, 316K)

[Reference 2. Prevalidation of the Aromatase Assay Using Human, Bovine and Porcine Placental Microsomes and Human Recombinant Microsomes \(PDF\)](#) (224 pp, 1.7M)

[Reference 3. Microsomal Aromatase Prevalidation Supplementary Study: Determine Day to Day and Technician Variability \(PDF\)](#) (110 pp, 40.6M)

[Reference 4. Microsomal Aromatase Prevalidation Supplementary Study: Establish Inhibition Curves and IC50s for Two Reference Chemicals \(PDF\)](#) (138 pp, 34.3M)

[Reference 5. Microsomal Aromatase Prevalidation Supplementary Study: Compare Estrone and Tritiated Water Measurement Methods \(PDF\)](#) (138 pp, 34.3M)

***NOTE: References 4 and 5 are found in the same document (i.e., PDF file).**

[Reference 6. Microsomal Aromatase Prevalidation Supplementary Study: Summation of Findings and Revised Aromatase Protocol \(PDF\)](#) (22 pp, 832K)

[Reference 7. Placental Aromatase Assay Validation: Positive Control Study \(PDF\)](#) (36 pp, 980K)

[Placental Aromatase Assay Validation: Appendix A \(Battelle Memorial Institute Report\) \(PDF\)](#) (175 pp, 4.6M)

[Placental Aromatase Assay Validation: Appendix B \(In Vitro Technologies, Inc. Report\) \(PDF\)](#) (178 pp, 6.2M)

[Placental Aromatase Assay Validation: Appendix C \(WIL Research Laboratories, LLC Report\) \(PDF\)](#) (174 pp, 4.5M)

[Placental Aromatase Assay Validation: Appendix D \(Draft Interlaboratory Statistical Analysis Report\) \(PDF\)](#) (19 pp, 609K)

Reference 8. Placental Aromatase Assay Validation: Multiple Chemicals Studies with Centrally Prepared Microsomes

[Vol. 1 – Draft Final Report & Appendix A \(PDF\)](#) (387 pp, 36.4M)

[Vol. 2 – Appendices B, C, D \(PDF\)](#) (775 pp, 26.9M)

Reference 9. Placental Aromatase Assay Validation: Prepare Microsomes in Two Participating Laboratories

[Vol. 1 – Draft Final Report & Appendix A \(PDF\)](#) (290 pp, 11M)

[Vol. 2 – Appendices B, C, D, E \(PDF\)](#) (620 pp, 25.5M)

Reference 10. Placental Aromatase Validation Study: Conduct Multiple Chemical Studies with Microsomes Prepared in Participating Laboratories

[Vol. 1 - Overall Summary, Appendices A & B \(PDF\)](#) (960 pp, 44.4M)

[Vol. 2 - Appendices C, D, E, F & G \(PDF\)](#) (1,252 pp, 50.3M)

Reference 11. [Human Recombinant Microsomal Aromatase Assay Validation Study: Positive Control Study \(PDF\)](#) (564 pp, 20.8M)

Reference 12. Recombinant Aromatase Validation Study: Conduct Multiple Chemical Studies With Recombinant Microsomes

[Vol. 1 - Overall Summary, Appendices A & B \(PDF\)](#) (962 pp, 42.8M)

[Vol. 2 - Appendices C, D, E, F & G \(PDF\)](#) (1,230 pp, 43.4M)

Reference 13. [Characterization of the Inhibition of Aromatase Activity by Nonylphenol \(PDF\)](#) (158 pp, 5.2M)

Reference 14. [Supplementary Testing of 16 Chemicals in the Recombinant Aromatase Assay \(PDF\)](#) (717 pp, 13.6M)