INTEGRATED SUMMARY REPORT

for

Validation of an Estrogen Receptor Binding Assay using Rat Uterine Cytosol as Source of Receptor as a Potential Screen in the Endocrine Disruptor Screening Program Tier 1 Battery

March 2009

U.S. Environmental Protection Agency
Office of Science Coordination and Policy
and
Office of Research and Development

Table of Contents

List of Appendices

List of Tables

I. Introduction

A. Purpose of the Endocrine Disruptor Screening Program

Section 408(p) of the Federal Food Drug and Cosmetic Act (FFDCA) requires the U.S. Environmental Protection Agency (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 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 committee of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement its Endocrine Disruptor Screening Program (EDSP). The EDSP is described in detail at http://www.epa.gov/scipoly/oscpendo/.

Upon recommendations from the EDSTAC (1998), the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormonal systems as well as wildlife.

B. Tiered approach to screening

The EPA accepted the EDSTAC's recommendations for a two-tier screening program in a Federal Register Notice in 1998 (USEPA (1998)). The purpose of Tier 1 is to identify the potential of chemicals to interact with the estrogen, androgen, or thyroid (EAT) hormonal systems. A negative result in Tier 1 would be sufficient to put a chemical aside as having low to no potential to cause endocrine disruption, whereas a positive result would require further testing in Tier 2. The purpose of Tier 2 is to confirm the interaction, identify and characterize any adverse effects, and to provide information that will be useful in risk assessment based, in part, on dose-response relationships. Tier 2 is expected to comprise multigeneration tests in various taxa (i.e., mammals, birds, fish, amphibians, and invertebrates).

C. The Tier 1 battery of assays

The EDSTAC concluded that a Tier-1 battery should be comprised of a suite of complementary screening assays having the following characteristics:

- Maximum sensitivity to minimize false negatives while permitting an as yet undetermined, but acceptable, level of false positives.
- Range of organisms representing known or anticipated differences in metabolic activity and include assays from representative vertebrate classes to reduce the likelihood that important pathways for metabolic activation or detoxification of parent substances or mixtures are not overlooked.
- Capacity to detect all known modes of action (MOAs) for the endocrine endpoints of concern. All chemicals known to affect the action of EAT hormones should be detected.
- Range of taxonomic groups among the test organisms. There are known differences in endogenous ligands, receptors, and response elements among taxa that may affect the endocrine activity of chemical substances or mixtures.
- Diversity among the endpoints and within and among assays to reach conclusions based on "weight-of-evidence" considerations. Decisions based on the screening battery results will require weighing the data from several assays.
- Inexpensive, quick, and easy to perform.

To detect chemicals that may affect the EAT hormonal systems through any one of the known MOAs — interruption of hormone production or metabolism, binding of the hormone with its receptor, interference with hormone transport, etc. — the EDSTAC recommended the *in vitro* and *in vivo* assays shown in Table 1 for inclusion in the Tier-1 screening battery.

Table 1. Tier-1 screening assays recommended by the EDSTAC

Assays	Reasons for consideration
Estrogen receptor (ER) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the ER.
Androgen receptor (AR) binding or transcriptional activation	A sensitive *in vitro* test to detect chemicals that may affect the endocrine system by binding to the AR.
In vitro steroidogenesis	A sensitive *in vitro* test to detect chemicals that interfere with the synthesis of the sex steroid hormones.
Uterotropic (rat)	An *in vivo* assay to detect estrogenic chemicals. It offers the advantage over the binding assay of incorporating absorption, distribution, metabolism, and excretion (ADME)
Hershberger (rat)	An *in vivo* assay to detect androgenic and anti-androgenic chemicals. It offers the advantage over the binding assay of incorporating ADME and differentiating between AR agonists and antagonists.
Pubertal female (rat)	An assay to detect chemicals that act on estrogen or through the hypothalamus-pituitary-gonadal (HPG) axis that controls the estrogen and androgen hormone systems. It also detects chemicals that interfere with the thyroid system.
Frog metamorphosis	A sensitive assay for detection of chemicals that interfere with the thyroid hormone system.
Fish screen	Fish are the furthest removed from mammalians among vertebrates both from the standpoint of evolution—their receptors and metabolism are different from mammals—and exposure/habitat, since they would be subject to exposure through the gills, whole body, and diet. Thus, the fish assay would augment information found in the mammalian assays and would be more relevant than the mammalian assays in triggering concerns for fish and perhaps other non-mammalian taxa.

In addition, the EDSTAC recognized there were other combinations of screening assays that may be suitable and therefore recommended that the EPA validate the alternative screening assays shown in Table 2.

Table 2. Alternative assays recommended by the EDSTAC for the Tier-1 Screening Battery

Assays	Reasons for consideration
In vitro placental aromatase	The aromatase assay detects chemicals that inhibit aromatase and would be needed if either of the two following assays using males were substituted for the female pubertal assays. The male is not believed to be as sensitive to alterations in aromatase as the female and would not therefore be sufficient to detect interference with aromatase in the screening battery.
Pubertal male (rat)	The assay detects chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It ialso detects chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.
Adult male (rat)	The assay is also designed to detect chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also designed to detect chemicals that interfere with the thyroid system..

D. Validation

As noted, Section 408(p) of the FFDCA requires the EPA to use validated test systems and other scientifically relevant information. Validation has been defined as *"the process by which the reliability and relevance of a test method is evaluated for a particular use"* (OECD (1996); NIEHS (1997)).

Reliability is defined as the reproducibility of results from an assay within and between laboratories.

Relevance describes whether a test is meaningful and useful for a particular purpose (OECD (1996)). For Tier-1 EDSP assays, relevance can be defined as the ability of an assay to detect chemicals with the potential to interact with the estrogen, androgen, and/or thyroid hormonal systems.

The EDSTAC considered the ER binding assay to have "gained sufficient general acceptance within the field of endocrine toxicology to be considered *de facto* validated (reliable *and* relevant)" (EDSTAC 1998, Appendix R). Nevertheless, it continued, "variations in protocols for [this screen] can produce disparate results. Therefore, standardization of the protocol … should be accomplished by EPA before [this assay is] implemented as [a] screening [requirement] for endocrine activity or disruption." (Ibid.) As a result, EPA began efforts to standardize the assay. These standardization efforts are described in this document, along with two additional validation studies.

Federal agencies are also instructed by the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) Authorization Act of 2000 to ensure that new and revised test methods are valid prior to their use.

In general, the EPA has followed a five-stage validation process outlined by the ICCVAM (NIEHS (1997)) for validation EDSP assays. The stages of the process outlined by the ICCVAM are as follows:

First Stage - *Test Development*, an applied research function which culminates in an initial protocol. As part of this phase, the EPA prepares a Detailed Review Paper (DRP) to explain the purpose of the assay, the context in which it will be used, and the

scientific basis upon which the assay's protocol, endpoints, and relevance rest. The DRP reviews the scientific literature for candidate protocols and evaluates them with respect to a number of considerations, such as whether the candidate protocols meet the assay's intended purpose, the costs and other practical considerations. The DRP also identifies the developmental status and questions related to each protocol; provides the information needed to answer the questions; and, when possible, recommends an initial protocol for the initiation of the second stage of validation.

Second Stage - *Standardization and Optimization,* in which the protocol is refined, optimized, standardized and initially assessed for transferability and performance. Several different types of studies are conducted during this second phase depending upon the state of development of the method and the nature of the questions that the protocol raises. The initial assessment of transferability is generally a trial in a second laboratory to determine that another laboratory besides the lead laboratory can follow the protocol and execute the study.

Third Stage - *Inter-laboratory Validation* studies are conducted in independent laboratories with the optimized protocol. The results of these studies are used to determine inter-laboratory variability and to set or cross-check performance criteria.

Fourth Stage - *Peer Review*, an independent scientific review by qualified experts. The EPA has developed extensive guidance on the conduct of peer reviews because the Agency believes that peer review is an important step in ensuring the quality of science that underlies its regulatory decisions (USEPA (2007)).

Fifth Stage - *Regulatory Acceptance*, adoption for regulatory use by an agency.

Criteria for the validation of alternative test methods *(in vitro* methods designed to replace animal tests in whole or in part) have generally been agreed upon in the United States by the ICCVAM, in Europe by the European Centre for the Validation of Alternative Methods (ECVAM), and internationally by the Organisation for Economic Co-Operation and Development (OECD). These criteria, as stated by ICCVAM (NIEHS (1997)), are as follows:

1. The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.

2. The relationship of the endpoints determined by the test method to the *in vivo* biologic effect and toxicity of interest must be addressed.
3. A formal detailed protocol must be provided and must be available in the public domain. It should be sufficiently detailed to enable the user to adhere to it and should include data analysis and decision criteria.
4. Within-test, intra-laboratory and inter-laboratory variability and how these parameters vary with time should have been evaluated.
5. The test method's performance must have been demonstrated using a series of reference chemicals preferably coded to exclude bias.
6. Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test to that of the test it is designed to replace.
7. The limitations of the test method must be described (e.g., metabolic capability).
8. The data should be obtained in accordance with Good Laboratory Practices (GLPs).
9. All data supporting the assessment of the validity of the test methods including the full data set collected during the validation studies must be publicly available and, preferably, published in an independent, peer-reviewed publication.

The EPA has adopted these validation criteria for the EDSP as described elsewhere (USEPA (2007)). Although attempts have been made to thoroughly comply with all validation criteria, the various *in vitro* and *in vivo* screening assays are not replacement assays (Validation Criterion No. 6). Many of them are novel assays; consequently, large data bases do not exist as a reference to establish their predictive capacity (e.g., determination of false positive and false negative rates). It is expected that a review of results from the testing of the first group of 50 to 100 chemicals, which was recommended by the Scientific Advisory Panel (SAP) (USEPA (1999)), will allow a more complete assessment of the performance of the Tier-1 screening battery.

For technical guidance in developing and validating the various Tier-1 screens and Tier-2 tests, the EPA chartered two federal advisory committees: the Endocrine

Disruptor Methods Validation Subcommittee, or EDMVS (from 2001 to 2003), and the Endocrine Disruptor Methods Validation Advisory Committee, or EDMVAC (from 2004 to 2006). These committees, composed of scientists from government, academia, industry, and various interest groups, were charged to provide expert advice to the EPA on protocol development and validation. The EPA also cooperates with member countries of the OECD to develop and validate assays of mutual interest to screen and test for endocrine effects.

Even though assays are being developed and validated individually and peer reviewed on an individual basis (i.e., their strengths and limitations are being evaluated as stand-alone assays), the Tier-1 assays will be used in a battery of complementary screens. An individual assay may serve to strengthen the weight of evidence in a determination (e.g., positive results in an ER binding assay in conjunction with positive results in the uterotropic and pubertal female assays would provide a consistent signal for estrogenicity) or to provide coverage of MOAs not addressed by other assays in the battery. Information supporting the validation of individual assays was used at the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) SAP for peer review of the EPA's recommendations for a Tier-1 battery (USEPA 2008). The Tier-1 battery peer review focused, in part, on the extent of coverage and overlap that the suite of assays will have with one another in detecting endocrine-related effects associated with the EAT hormonal systems.

E. Purpose of this report

The purpose of this Integrated Summary Report is to provide an historical summary of the standardization and validation of a protocol for the estrogen receptor (ER) binding assay using rat uterine cytosol (RUC) as source of receptors. The reasoning and judgments leading to the various studies, and conclusions concerning the strengths and weaknesses of the assay in its current form, are presented.

II. Purpose and brief description of the assay

A. Purpose of the assay

This assay will be used to determine the ability of a compound to interact with the ERs isolated from rat uteri. It will be used in conjunction with other *in vitro* and *in vivo* assays in the EDSP to determine whether there is a potential for the chemical to interact with the estrogen hormonal system. Because it is intended for screening for potential interaction of any sort, the assay is not being standardized for distinguishing one-site competitive binding from other kinds of interaction. In addition, the assay is not being standardized for use in Quantitative Structure-Activity Relationship models.

As explained in section I.B, information from Tier 1 screens such as the ER binding assay will be used as indications of the need for further testing in Tier 2. It is only after Tier 2 results are available that risk assessment will be undertaken. Generating dose-response information for use in risk assessment is not one of the purposes of this assay.

B. Overview of the assay

This assay evaluates the inhibition of radiolabeled estradiol binding to rat ER by a test chemical. Rat uterine cytosol (RUC) is the source of the estrogen receptor. The assay consists of two sets of experiments: a saturation binding assay to characterize receptor activity, followed by competitive binding assays that measure the competition of test compounds and control chemicals (the native ligand, 17β-estradiol, as the reference ("standard") chemical; a weak positive control, norethynodrel; and a negative control, octyltriethoxysilane) against radiolabeled 17β-estradiol for the receptor.

The purpose of the saturation binding assay is to characterize the specificity and activity of the cytosol preparation and ensure that the ER activity is sufficient for the competitive assay. The saturation assay measures the affinity of the receptor for its natural ligand (17β-estradiol, radiolabeled), quantified by the dissociation constant (K_d, nM); and the concentration of active receptor sites, quantified by the maximum specific binding number (B_{max}, fmoles of estradiol/100 µg of protein). (See Section E for further discussion of these quantities.) The saturation binding assay tests eight increasing concentrations of labeled estradiol across two orders of magnitude, each in the

presence of a 100-fold higher concentration of unlabeled estradiol. Both total and nonspecific binding are measured and specific [^{3}H]-estradiol binding is calculated by subtracting non-specific from total. K_d and B_{max} are calculated through nonlinear regression to a one-site binding model. (The Scatchard plot is not used for quantification although it is recommended as a visual aid in evaluating the performance of the assay since the linearity of the plot is a useful indicator of a well-performed run.)

The competitive binding assay measures the affinity of an unlabeled chemical (i.e., reference "standard", weak positive control, or test chemical) in competition with high affinity radioligand (tritiated 17β-estradiol) for the estrogen receptor. It is quantified by the concentration of competitor which inhibits 50% of the binding of the radioligand (IC_{50}) and frequently by relative binding affinity (RBA, % relative to estradiol). The competitive assay measures the binding of [^{3}H]-estradiol at a fixed concentration in the presence of a wide range (eight orders of magnitude) of test chemical concentrations. The data are then fit, where possible, to the Hill Equation (Hill 1910), which describes the displacement of the radioligand by a one-site competitive binder. The extent of displacement of the radiolabeled estradiol is used to characterize the test chemical as interacting, not interacting, or generating an equivocal response. (See Section F.1 for further discussion.)

C. Review of literature

The EPA asked the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to prepare a review of the literature on ER binding and ER transcriptional activation (as well as their androgen receptor counterparts). ICCVAM identified 72 publications with an appropriate level of detail, containing data on 638 substances from 14 different ER binding assays over four species (rat, mouse, rabbit, and human). Relatively few chemicals were tested more than once in the same assay or in multiple assays, and no formal validation studies to assess the reliability or performance of ER binding assays were found. The review found no published guidelines for conducting *in vitro* ER binding studies.

The Background Review Document (BRD) that documents the review provided procedural standards for *in vitro* ER binding assays and proposed that the rat uterine cytosol (RUC) assay, which has been the most widely used method for identifying

substances with ER binding activity, be used as the standard against which new assays for ER binding activity be evaluated. The BRD further recommended that "[b]ased on a consideration of such factors as relative performance, elimination of animal use, the use of the ER from the species of interest, and the use of alternatives to radioactive substanced, the [human recombinant ER alpha (hrERα), human recombinant ER alpha – fluorescence polarization, and human recombinant ER beta] assays should have the highest priority for validation as screening assays for human health-related issues, while the GST-rtERdef assay might be preferred when screening for substances that pose a hazard to wildlife." EPA is currently chairing an OECD-sponsored validation effort for the hrERα binding assay. If the hrERα assay is validated, EPA may consider using it to replace the ER-RUC assay in the Tier 1 battery.

ICCVAM's BRD became available in 2002 (ICCVAM 2002). It serves as the Detailed Review Paper (DRP) referred to in the description of the validation process above (section I.D). The Executive Summary and the Conclusions of the BRD are attached as Appendix 2.

D. Assay components other than test chemical

The following is a general description of the major components of the assay other than the test chemical. Brief descriptions of conduct of the assay are given in Sections E and F. The full protocol is attached as Appendix 1.

1. Solvent

The best solvent for the test chemical is chosen from among dimethylsulfoxide (DMSO), ethanol, or water. The solvent used for a test chemical must also be used for the reference chemical (inert 17β-estradiol) and the control chemicals (norethynodrel and octyltriethoxysilane) unless the solvent is water. If the test chemical is run in water, the controls are run in ethanol since they are not soluble in water. The total volume of ethanol allowed is no more than 3% of the total assay volume; DMSO, if used, may not exceed 10%. (See Section III.H for further discussion of solvent maxima.)

2. Reference estrogen – 17β-Estradiol

17β-Estradiol is the native ligand that binds with high affinity to the estrogen receptor of rat uterus. In the saturation assay, the final unlabeled estradiol

concentrations in the assay tubes are 3, 6, 8, 10, 30, 60, 100, and 300 nM. Unlabeled estradiol concentration is 100x the tritiated estradiol concentration to bind all the high-affinity ER binding sites so that the tritiated estradiol competes only at the low-affinity sites, thus providing a measure of non-specific binding. For the competitive assay, the seven concentrations for estradiol include serial dilutions at each log unit (with half-log spacing around the log(IC_{50})) from 10^{-7} to 10^{-11} M. This establishes a standard reference curve whose characteristics can be compared to performance criteria to ensure that the binding assay is working correctly.

3. Marker/tracer– Radiolabeled 17β-estradiol

Tritiated estradiol is used as the marker/tracer in the assay. Before preparing the dilutions of the [^{3}H]-17β-estradiol, the specific activity is adjusted for decay over time since certification by the manufacturer. In the saturation assay, the final [^{3}H]-estradiol concentrations in the assay tubes are 0.03, 0.06, 0.08, 0.1, 0.3, 0.6, 1, and 3 nM. For competitive assays, the final [^{3}H]-estradiol concentration is a constant 1 nM.

4. Positive control – Norethynodrel

Norethynodrel is used as the weak positive control in the competitive binding assay. It was chosen because it was one of the weakest positive chemicals that reliably produced a full binding curve through 10^{-4} M in preliminary studies. Norethynodrel final concentrations are 8 dilutions that cover log units between 10^{-4} to $10^{-8.5}$ M, with half-log units around the IC_{50}.

5. Negative control – Octyltriethoxysilane

Octyltriethoxysilane is used as the negative control in the competitive binding assay. It was chosen because it reliably showed no competition with the radiolabeled estradiol, and little variability across the entire range of concentrations tested, in multiple laboratories. The eight concentrations cover each log unit from 10^{-3} to 10^{-10} M inclusive.

6. Rat uterine cytosol

Uteri are collected from Sprague-Dawley rats (85 to 100 days of age) ovariectomized seven to ten days prior to being humanely killed. Weighed and trimmed uterine tissues are placed in ice-cold buffer prepared with Tris(hydroxymethyl)amino-methane, Ethylenediaminetetraacetic acid, Dithiothreitol, and Glycerol (TEDG) with

phenylmethylsulfonyl fluoride (PMSF). The final extraction volume has a ratio of 0.1 g of tissue per 1.0 ml buffer. The tissues are homogenized and the cytosol pooled, aliquoted, and stored at -80 ºC. The protein content for each batch of cytosol is determined using a method compatible with buffers that contain DTT. Typical protein values are 1 to 4 mg/ml. Care is taken to thaw only amounts needed and to discard rather than refreeze unused portions.

E. Saturation binding assay

Estrogen receptor saturation binding experiments measure total and nonspecific binding of increasing concentrations of [^{3}H]-estradiol, at equilibrium. Three replicate data points are collected at each concentration. Total binding is calculated by converting the disintegrations per minute (dpm) from samples containing [^{3}H]-estradiol (no radioinert estradiol). Nonspecific binding is calculated by converting the dpm from tubes containing [^{3}H]-estradiol + 100-fold molar excess of radioinert estradiol, assuming that the excess of radioinert estradiol will occupy all of the available estrogen receptor binding sites. Specific binding is calculated as the difference between the nonspecific binding and total binding at each of the tested concentrations. The saturation assay conditions are:

Source of receptor		Rat uterine cytosol
Concentration of radioligand (as serial dilutions)		0.03 – 3 nM
Concentration of inert ligand (100 x [radioligand])		3 – 300 nM
Concentration of receptor		50 µg protein/tube*
Temperature		4º C
Incubation time		16-20 hours
Assay buffer	Tris	10 mM (pH 7.4)
	EDTA	1.5 mM
	Dithiothreitol	1 mM
	Glycerol	10 %
	Phenylmethylsulfonyl fluoride	1 mM

*Protein concentration may need to be adjusted to minimize ligand depletion while maintaining an adequate radioactivity signal.

Reversible binding between 17β-estradiol and the estrogen receptor can be described using a single site binding model:

$$\text{Receptor} + \text{Ligand} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Receptor} \bullet \text{Ligand}$$

where k_{on} and k_{off} represent the rates of the binding and dissociation events respectively. The dissociation constant (K_d) is defined as:

$$K_d \equiv k_{off}/k_{on}$$

At equilibrium, the ratio of free to bound estradiol is constant and the fraction bound can be calculated:

$$\text{Fraction of receptor bound} = \frac{[\text{Ligand}]}{[\text{Ligand}] + K_d}$$

Thus the fraction of receptor bound starts at 0 when ligand concentration is zero and has an upper asymptote of 1 as ligand concentration grows to greatly exceed the receptor concentration.

Ligand binding is conceptually represented as specific binding and nonspecific binding. Specific binding is binding to the ligand binding domain of the receptor while nonspecific binding is binding to sites other than the ligand binding domain, including the reaction tube walls or other components of the reaction mixture. The greater the concentration of free radioligand, the higher the nonspecific binding will be. Nonspecific binding is adjusted for by subtracting the disintegrations per minute (dpms) of the nonspecific binding tubes corresponding to each total ligand concentration, from the total binding dpms for that concentration.

The degree of nonspecific binding is determined by including a parallel set of tubes that contain the same concentrations of radioligand as the total-binding tubes, plus a sufficiently large concentration of unlabeled substance that will bind with all the receptor binding sites and leave no receptors remaining to bind with the radioligand. Any bound dpms must then necessarily correspond to nonspecific binding of the radioligand.

In the current assay, three parallel tubes are run for each total concentration of radioligand and the averages of the radioactive decay (dpm) among the three total binding tubes and among the three total added tubes are determined. The total free radioligand is determined by subtracting the total bound dpm from the average total

added dpm. The specific bound decay is determined by subtracting the average nonspecific binding decay from each of the total bound decay determinations.

The specific bound ligand concentration is related to the total free ligand concentration by the nonlinear regression relation model:

$$Y = \frac{B_{max} X}{X + K_d} + \varepsilon$$

which is based on the law of mass action (i.e., the law that reaction rate is proportional to reactant concentration for reactions with a single mechanistic step). In this relation X is the free radioligand concentration and Y is the concentration of radioligand specific-bound to the receptor. B_{max} is the maximum concentration bound as the concentration of free radioligand goes to ∞. K_d is the equilibrium dissociation constant discussed above; it corresponds to the radioligand concentration at which half the receptor binding locations are filled. ε represents the random variation about the model and is often modeled as independently distributed with mean 0 and constant variance σ^2.

The model relies on the assumption that there is such an excess of free radioligand available that its concentration does not change appreciably when some of it binds to receptor. There appears to be no general agreement in the literature about the extent of "ligand depletion" that can be tolerated, but a value of 10 to 20% is often considered high. Reducing the amount of cytosolic protein that is used in the assay reduces the amount of receptors in the mix and thus reduces ligand depletion (all else being equal) but it also reduces the number of bound dpms available for measurement and can therefore add variability to the measurement. EPA has chosen to aim at a level of protein that binds 25% to 35% of the total radiolabeled estradiol that is added to the tube, and to rely on the method developed by Swillens (1995) to compensate for ligand depletion when calculating K_d and B_{max}. Carter et al. (2007) provide evidence that the Swillens correction "appears to be the most appropriate method for estimating ligand affinity in situations of ligand depletion."

In general, when evaluating data from ER saturation assays, the following points should be considered:

- As increasing concentrations of [^{3}H]-estradiol were used, does the specific binding curve reach a plateau? Maximum specific binding must be reached, indicating saturation of ER with ligand.
- Does the data produce a linear Scatchard plot (a plot of bound/free ligand as a function of specific binding)? Non-linear plots generally indicate a problem with the assay such as ligand depletion or incorrect assessment of non-specific binding.
- Is the K_d within an acceptable range? Literature values for K_d using rat uterine cytosol preparations have varied from 0.05 to 0.5 nM. The variation in K_d may be a reflection of different laboratories using radiolabeled estradiol with a wide range of specific activity ([^{3}H]-17β-estradiol versus [^{125}I]-17β-estradiol). In addition, lower K_d may be observed when assay conditions minimize ligand depletion, and slightly different K_d values exist for ERα and ERβ. Rat uterine cytosol prepared using this protocol will typically yield a K_d of 0.03 to 1.5 nM.
- Are runs consistent? That is, are the standard errors for the K_d or B_{max} excessive?.
- Is nonspecific binding excessive? In general, the value for nonspecific binding should be less than 50% of the total binding at the highest concentration.

F. Competitive binding assay

The competitive binding assay measures the binding of a single concentration of [^{3}H]-estradiol in the presence of increasing concentrations of a test substance. If the test substance interacts with the receptor, it inhibits the binding of increasing amounts of radiolabeled estradiol. EPA requires three concurrent replicates per run at each concentration, and three non-concurrent runs, to characterize the potential of a test substance to interact with the estrogen receptor.

Control samples are included for each assay run. These include:

- Graded concentrations of unlabeled 17β-estradiol. The behavior of unlabeled estradiol in competing with labeled estradiol is well known and provides a standard by which to measure performance of the assay. The highest concentration of this series (100 nM) is 100 times the concentration of the radiolabeled estradiol (1 nM) and serves as the measure of non-specific binding.
- Graded concentrations of a positive control (norethynodrel). The behavior of the weak binder norethynodrel in competing with labeled estradiol is well known and provides assurance that similarly weak binders can be detected.
- Graded concentrations of a negative control (octyltriethoxysilane). This chemical does not interact with the estrogen receptor and provides assurance that a well-performed run does not falsely classify negatives as positives.

- Solvent control. Binding of the radiolabeled estradiol in the absence of any competitor is the baseline condition (100% binding) to which displacement by competitor can be compared.

The competitive binding assay conditions are:

Source of receptor		Rat uterine cytosol
Concentration of radioligand		1.0 nM
Concentration of receptor		50 µg protein/tube*
Concentration of test substance (as serial dilutions)		100 pM – 1 mM
Temperature		4º C
Incubation time		16-20 hours
Assay buffer	Tris	10 mM (pH 7.4)
	EDTA	1.5 mM
	Glycerol	10 %
	DTT	1 mM
	Phenylmethylsulfonyl fluoride	1 mM

* Receptor concentration may need to be adjusted for each batch of cytosol.
** Range and spacing of test substance concentrations may need to be adjusted depending on solubility and strength of interaction, if any.

Receptor concentration is adjusted to keep ligand depletion below 15% (although it is recommended that the protein concentration not be reduced below 35 µg per assay tube since this can result in the loss of centrifuge pellets during the separation of bound estradiol from free estradiol). This adjustment is required for each batch of cytosol prepared.

The test substance is initially tested at concentrations from 1 mM to 100 pM (i.e., 10^{-3} to 10^{-10} M inclusive), in ten-fold (i.e., log) increments. Ethanol, DMSO, or water may be used as solvent. If the highest concentration cannot be prepared in any of these solvents (e.g., because there is precipitate in the stock solution or forms upon addition to 4º C assay buffer, and adding more solvent would cause the final solvent concentration in the tube to be greater than the acceptable limit of 3% ethanol or 10% DMSO), that concentration may be omitted. Evidence must be provided in the report showing measures taken at each highest-concentration-attempted to obtain full solubility, such as gentle heating or using a different solvent. If the test substance is such a strong binder that a full curve is not obtained in the default range of

concentrations tested, additional dilutions are required so that the curve is adequately characterized.

Tubes are loaded (on ice, to prevent degradation of the receptor) as follows:

Volume (μL)	Constituent
390	Master mixture (TEDG + PMSF assay buffer + [^{3}H]-17β-estradiol)
10	Unlabeled 17β-estradiol, weak positive control, negative control, or test substance
100	Uterine cytosol (at concentration determined to be appropriate for that batch of cytosol)
500	Total volume in each assay tube

The tubes are vortexed and incubated at 4° C for 16 to 20 hours. Hydroxyapatite slurry (60% in cold TEDG+PMSF buffer) is added to each assay tube and the mixture vortexed at 5 minute intervals for 15 minutes (kept cold between vortexes), then centrifuged at 4° C for 10 minutes at 1000 x g. After centrifugation, the supernatant is decanted and the pellet containing the bound [^{3}H]-17β-estradiol is re-suspended in cold buffer. The wash is repeated twice more in the same manner. The final pellet is suspended in ethanol and allowed to come to room temperature, centrifuged at 1000 x g for 10 minutes, and a measured aliquot of the supernatant is counted in a scintillation counter for determination of dpms/vial.

1. One-site binding model

Although the nature of the interaction, if any, between a test substance and the estrogen receptor is not usually known beforehand, data from this screening program will be fit to a one-site competitive binding model for the sake of standardization. One-site competitive inhibition of the native ligand estradiol is a mechanism by which many pharmaceuticals interact with the estrogen receptor *in vivo*, and a test substance that displays a good fit to the one-site competitive model will likely be of interest. A substance that interacts with the receptor but is not a true one-site competitive inhibitor may not fit the model well but is still likely to demonstrate behavior that will allow classification as interacting with the receptor when this model is used. Such

compounds would also likely be of interest for further exploration of the compound's estrogenic properties.

If the radioligand and the inhibitor both bind reversibly to the same single binding site on the receptor, then specific binding at equilibrium follows a four parameter relation between percent bound (Y) and logarithm of inhibitor concentration (X). The concentration response relation is described by a sigmoid curve (a variation of the commonly used Hill equation):

$$Y = B + \frac{(T - B)}{1 + 10^{\beta(\log IC_{50} - X) + \log_{10}\left[\left(\frac{T-B}{50-B}\right)-1\right]}} + \varepsilon$$

The parameters in the equation represent the following quantities:

- B is the bottom plateau, i.e., the least expected percent bound.
- T is the top plateau, i.e., the greatest expected percent bound.
- β is the "Hill slope," i.e., the steepness with which the curve declines. Since the curve declines with increasing X, β is necessarily negative.
- $\log_{10}(IC_{50})$ is the logarithm of the concentration at which the expected value of Y = 50%. $\text{Log}_{10}(IC_{50})$ always corresponds to the same percentile of the concentration response and so can be directly compared between the test compound and the standard.
- ε is the random variation about the concentration response relation, with mean 0 and variance a function of the expected value of Y (often modeled as a constant, σ^2).

For an ideal response by a one-site competitive binder,

B = 0,

T = 100,

and β = -1.

The competitive binding assay is functioning correctly if all of the criteria in Table 3 have been met. The criteria apply to each individual run. If a run does not meet all of the performance criteria, the run must be repeated. Results for test chemicals in

disqualified runs are not used in classifying the ER interaction potential of those chemicals.

Table 3. Performance criteria for competitive binding, reference and weak positive controls.

Parameter	Unit	Estradiol		Norethynodrel		Octyltriethoxysilane	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
$Log_e(S_{yx})$ (i.e., Log_e(Residual Std.Dev)	--	NA	2.35	NA	2.60	NA	2.60
Bottom plateau level	% binding	-4	1	-5	1	NA	NA
Top plateau level	% binding	94	111	90	10	NA	NA
Hill slope	$log_{10}(M)^{-1}$	-1.1	-0.7	-1.1	0.7	NA	NA

These performance criteria reflect the fact that estradiol and norethynodrel are one-site competitive binders for the estrogen receptor and thus should display behavior consistent with one-site competitive binding in each run. Specifically, the curve fitted to the data points should descend from 90 – 10% over approximately an 81-fold increase in concentration (i.e., this portion of the curve will cover approximately 2 log units). A binding curve for either of these two standards that drops dramatically (e.g., from 90 - 0%) over one order of magnitude should be questioned, as should one that is U-shaped (i.e., percent bound is decreasing with increasing concentration of competitor but then begins to increase again). In both cases, something has happened to the dynamics of the binding assay and the reaction is no longer following the law of mass action. The values shown for the performance criteria are based on data generated using this ER-RUC assay and judged to be acceptable runs by EPA.

2. Statistical analysis

For each test run the one-site competitive binding model is fit to the data by nonlinear regression analysis. The model fits result in parameter estimates and associated standard errors as well as estimates of residual variability.

Nonlinear regression analysis can be carried out using PRISM 5 software (Motulsky 2003, 2007) or general purpose statistical systems such as SAS (2003). Prism, however, does not have a model for estimating log(IC_{50}), and it must be entered by the user. EPA is supplying a Prism template that includes the manually-entered formula. It is important not to use log(EC_{50}), which is the value Prism supplies, as if it were log(IC_{50}). The EC_{50} (effective concentration, 50%) is the concentration at which

50% of the *effect* of the chemical is seen – that is, the concentration halfway between the top plateau and bottom plateau. Log(EC_{50}) values are generally not comparable across chemicals.

III. Assay standardization and optimization

In 1998 the EDSTAC already considered the ER binding assay to be validated for use in the Endocrine Disruptor Screening Program but recommended standardization of the assay before implementation as a screening tool. (See Section I.D above.) Therefore, EPA standardized and optimized several experimental conditions including buffer composition, extraction method and radiolabel concentration. Some of the parameters were standardized and optimized in response to recommendations made in the ICCVAM Background Review Document.

A. Buffer composition and receptor concentration

Experiments were performed to determine the utility of adding 1 mM phenylmethylsulfonyl fluoride (PMSF) and/or 10 mM sodium molybdate. PMSF is a protease inhibitor and serves to protect the receptor from degradation by native enzymes present in the cytosol preparation. Sodium molybdate is thought to have protein stabilizing activity to prevent denaturing of the receptor at high (physiological) temperatures. The effect at the lower temperature used in this protocol was unclear. Its use had been suggested in the ICCVAM BRD but without explanation.

Also as part of this standardization and optimization, the effect of receptor concentration was evaluated. The original design called for 100 mg of protein per assay tube. Protein concentrations from 25 to 100 mg/tube were evaluated. A series of saturation runs (Table 4) were performed.

Table 4. Saturation assays comparing buffer composition and receptor concentration

Buffer	TEDG Only				
Run	344 J	345 L	350-J	355-J	351-L
Protein (ug/tube)	100	100	50	50	25
K_d (nM)	0.050	0.055	0.048	0.058	0.061
B_{max} (fmole/100 ug)	17.71	21.41	14.51	15.06	11.20

Buffer	TEDG + sodium molybdate + PMSF					
Run	346-J	347-L	348-J	352-J	353-L	349-L
Protein (ug/tube)	100	100	50	40	40	25
K_d (nM)	0.089	0.098	0.041	0.031	0.037	0.032
B_{max} (fmole/100ug)	38.85	40.33	32.33	25.53	27.33	18.05

Based on the linearity of the Scatchard plots and ligand depletion (data not shown), the optimal conditions were 50 µg protein/tube.

The receptor concentrations of cytosol preparations are likely to vary with batch and the 50 µg value will not necessarily be optimal for all batches. These data established a reasonable starting point around which laboratories should determine their own optimum for each cytosol preparation.

Additionally, a series of competitive assays were run using the estradiol and norethynodrel standards to compare the buffer systems. The IC_{50}s of the standard curves, and the IC_{50}s and RBAs of norethynodrel are presented in Table 5. The RBAs obtained in the different buffers were observed to be in the same range. There did not appear to be any greater variation in IC_{50} values between buffer systems than that observed within buffer systems. The dpm values of the 100% binding tubes in these assays ranged from 1 to 4 percent of those of the hot tubes, indicating that ligand depletion was not significant.

Table 5. Competitive assays comparing buffers

Buffer	TEDG Only		
Run	364 J	367 J	
Estradiol IC_{50} (nM)	0.94	1.47	
Norethynodrel IC_{50} (nM)	4362	5630	
RBA (percent)	0.022	0.026	

Buffer	TEDG + sodium molybdate + PMSF		
Run	361 L	362 J	366 J
Estradiol IC_{50} (nM)	1.44	1.02	0.67
Norethynodrel IC_{50} (nM)	8650	5627	5371
RBA (percent)	0.017	0.018	0.012

The results of these experiments demonstrated that the combination of molybdate and PMSF improved the assay performance. After a review of literature regarding the effect of molybdate on the estrogen receptor (e.g., Mauck et al. (1982); Murayama and Fukai (1985); Pettersson et al. (1985)), it was felt that since this protocol was performed at low temperature the addition of molybdate to prevent thermal degradation of the receptor was not necessary and could perhaps lead to unpredictable effects. For this reason, and to simplify the number of steps needed in the protocol, only the protease inhibitor PMSF was included in the final protocol.

B. Separation technique (HAP vs. DCC)

Separation of free and bound radioactivity was compared using dextran-coated charcoal (DCC) and hydroxyapatite (HAP) using the TEDG buffer with PMSF. A series of competitive assays were run using DCC and HAP for estradiol, norethynodrel and bisphenol A (Figure 1). IC_{50}s are presented in Table 6. Ligand depletion was less than 10% in all assays. The RBAs obtained in the different separation systems were observed to be in the same range. There did not appear to be any greater variation in IC_{50} values between the two separation systems than that observed within separation systems. Given indications in the literature of potential problems such as dilution-induced, dramatic, and variable loss of apparent estrogen receptor content with use of DCC with cytosol-derived receptor (Pettersson et al. (1985)), EPA chose to use HAP as the separation medium in the final protocol.

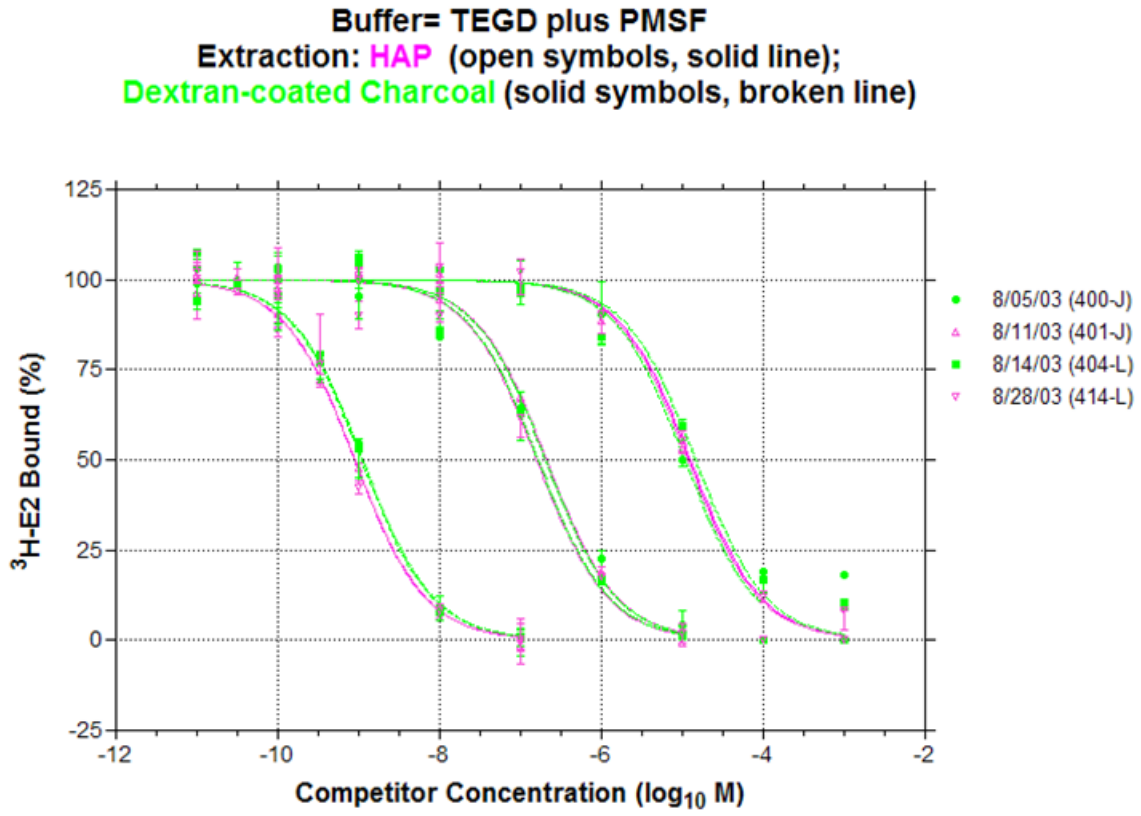


Figure 1. Competitive assay comparing separation media for estradiol (left), norethynodrel (center) and bisphenol A (right).

Table 6. Comparison of separation systems (DCC vs. HAP) in competitive assays

	Standard Curve	Positive Control	Bisphenol A
	$\log_{10}IC_{50}$	$\log_{10}IC_{50}$	$\log_{10}IC_{50}$
8/05/03 (400-J)	-8.949	-6.689	-4.967
8/11/03 (401-J)	-9.046	-6.675	-4.928
8/14/03 (404-L)	-8.964	-6.781	-4.835
8/28/03 (414-L)	-9.051	-6.795	-4.902

C. Post-incubation temperature

In initial assays using DCC as the separation medium, it was observed that there was drift in the values obtained for the 100% tubes placed at the beginning vs. the end of a run. According to the DCC protocol, assay tubes are refrigerated (4°C) or held on ice during the performance of the assay. The addition of the DCC terminates the reaction. Following the DCC addition step the tubes are mixed well, incubated for 15 minutes at 4° C and centrifuged at 4000 RPM for 15 minutes at 4° C. 300 µl of

supernatant is transferred from each assay tube to counting vials. It is at this step, when the tubes were not held on ice, that a drift in counts was observed. In an assay with many tubes, there would be an opportunity for the samples to increase in temperature.

To determine if temperature was the cause, the following experiment was conducted. Two separate assays composed of a series of 100% tubes were prepared in triplicate. On the first day, l[^{3}H]-estradiol (10µl), 100 µl of cytosol in buffer (100 µg of protein), 10 µl of absolute ethanol and 380 µl of buffer were added to each tube, mixed and incubated for 18-20 hours at 4° C on a rotating mixer. The total volume was 500 µl. On the second day, 300 µl of DCC suspension was added to tubes that were then mixed and incubated for 15 minutes at 4° C and centrifuged. 300 µl of each supernatant was removed from tubes at >5, 20, 40, 60, 140 and 160 minutes after centrifugation. One set of tubes remained on ice and the other set of tubes were held at room temperature. A decrease in supernatant counts (Figure 2) was observed over time for the tubes held at room temperature when compared to tubes held on ice. At 1 hour the counts had dropped ~5% and by 2 hours it was at ~10 % when compared to the "< 5 minutes" time point.

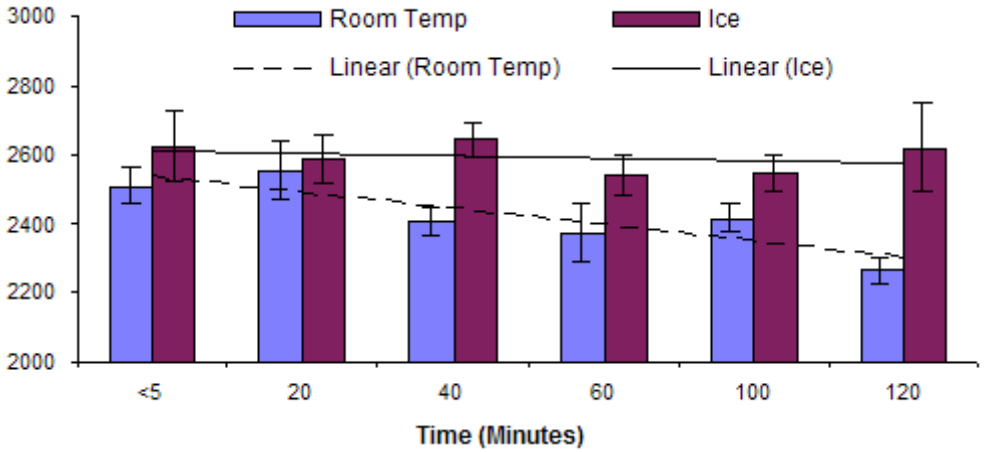


Figure 2. Influence of post-incubation temperature on bound-dpm counts when DCC is used

Although this information was not used in the interlaboratory validation studies, which were conducted using HAP rather than DCC, in retrospect it may be appropriate to investigate whether post-incubation temperature may have been the cause of the drift in solvent control tubes seen in the second interlaboratory validation study.

D. Assay Volume

A set of experiments compared the assay volume at 300 and 500 µL using cytosol at 0.2, 0.1, and 0.05 mg protein per assay tube in saturation assays. The results in both volumes were similar: the B_{max} decreased about 50% with decreased protein concentration, and the K_d also decreased with each reduction in amount of protein used, in both volumes. The reduction of the B_{max} is normal but the reduction in the K_d probably reflects a better estimate of the K_d as the assay approached linearity. At 300 µL the Hill coefficients improve from approximately 2.0 to 1.5 with decreasing protein concentration as would be expected if the receptor concentration was initially too high, but at 500 µL the Hill coefficient improved even further, to 1.2 at the lowest concentration of protein. Saturation assay results using a different batch of protein at 0.065 mg per tube at assay volumes of 300 and 500 µL showed equivalent B_{max} (0.02 and 0.022 pmole) and K_ds (0.097 and 0.095 nM) at the different volumes, but improvement of the Hill coefficient from 1.4 at 300 µL to 1.08 at 500 µL.

The competitive assay results across the two volumes are shown in Table 7. The test chemical was 17β-estradiol. IC_{50} values for this chemical using rat uterine cytosol as source of ER range from 1 nM to 8 nM in the BRD summary of the literature (ICCVAM 2002) but protein concentrations and assay volumes were not reported in that summary.

Table 7. Comparison of total assay volume, competitive binding assay

Total Volume (µL)	Cytosol Protein (mg/ml)	Protein (mg/tube)	IC_{50} (M)	95% Confidence Interval IC_{50} (M)
500	4.905	0.200	0.88E-9	8.3470e-010 to 9.2870e-010
500	4.905	0.150	0.87E-9	7.6990e-010 to 9.7870e-010
500	4.905	0.100	0.89E-9	7.9520e-010 to 1.0050e-009
300	4.905	0.200	0.97E-9	9.1940e-010 to 1.0130e-009
300	4.905	0.150	0.93E-9	8.5630e-010 to 1.0120e-009
300	4.905	0.100	0.93E-9	8.3290e-010 to 1.0420e-009

Given the clearly superior results in the saturation binding assay and no significant difference in results in the competitive binding assay, an assay volume of 500 µL was chosen as the most appropriate assay volume.

E. Cytosol Source

A comparison of cytosols from retired breeders or 80-90 day old female virgin rats was conducted. A set of saturation binding assays was run with assay tube volumes of 500 µl per tube and 50 µg protein per assay tube. In comparing values obtained for K_d, B_{max} or Hill coefficients, no significant differences were seen for these cytosols (Table 8). EPA chose to specify use of younger animals based on other considerations such as more-consistent general health but recognizes that retired breeders may be acceptable sources of receptor.

Table 8. Age of animal source of receptor

Cytosol Source	Protein Con. (mg/ml)	Protein Con. (mg/tube)	pmole/mg protein	B_{max} (pmole)	K_d (nM)
Retired Breeder	3.990	0.050	0.24	0.012	0.06
Retired Breeder	4.295	0.050	0.42	0.021	0.07
Retired Breeder	4.295	0.065	0.46	0.030	0.08
Retired Breeder	4.295	0.050	0.40	0.020	0.05
Retired Breeder	4.075	0.050	0.44	0.022	0.06
80-90 d virgin	2.927	0.050	0.40	0.020	0.06
80-90 d virgin	2.927	0.065	0.40	0.026	0.07
80-90 d virgin	2.927	0.050	0.34	0.017	0.05

F. Cytosol shelf life

Originally the assay included an arbitrary 30-day storage life for cytosol. In an effort to determine whether this could be extended, a series of competitive ER receptor assays were conducted with the same cytosol for a period of time ranging from 17 days to 127 days. The assays included standard curve, positive controls and a weak binder (bisphenol A). In general, the activity of the cytosol was very stable and no apparent loss of activity was observed over the measured time period (17 – 127 days). The cytosol was stored at –80° C for this period and it was not thawed and refrozen. Based on these data, a recommended storage limit of 90 days at –80°C was adopted.

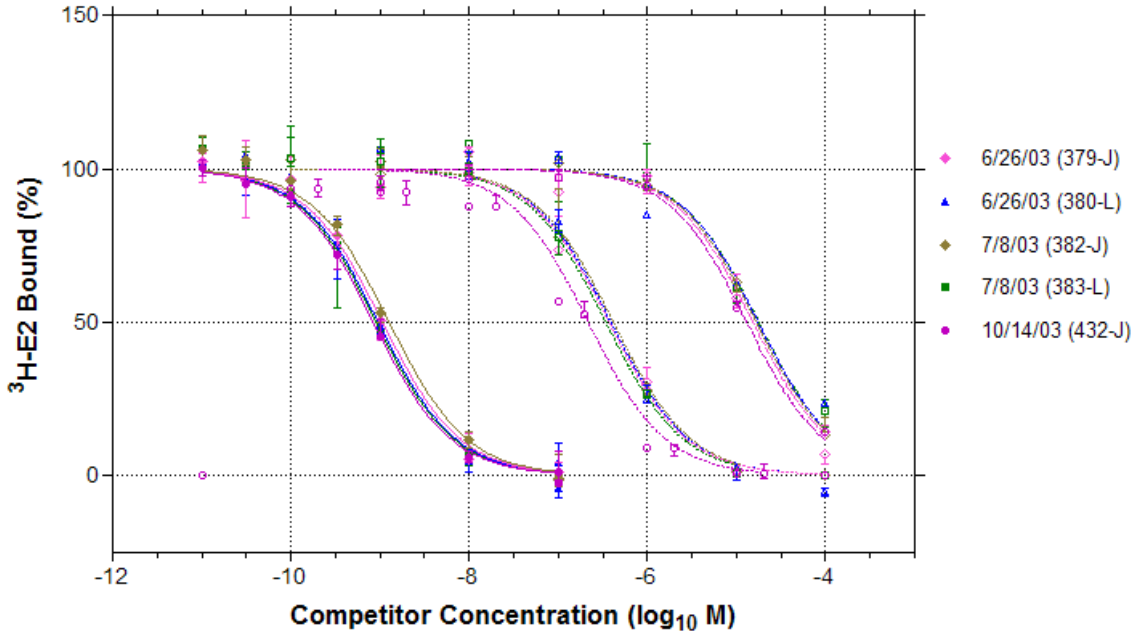
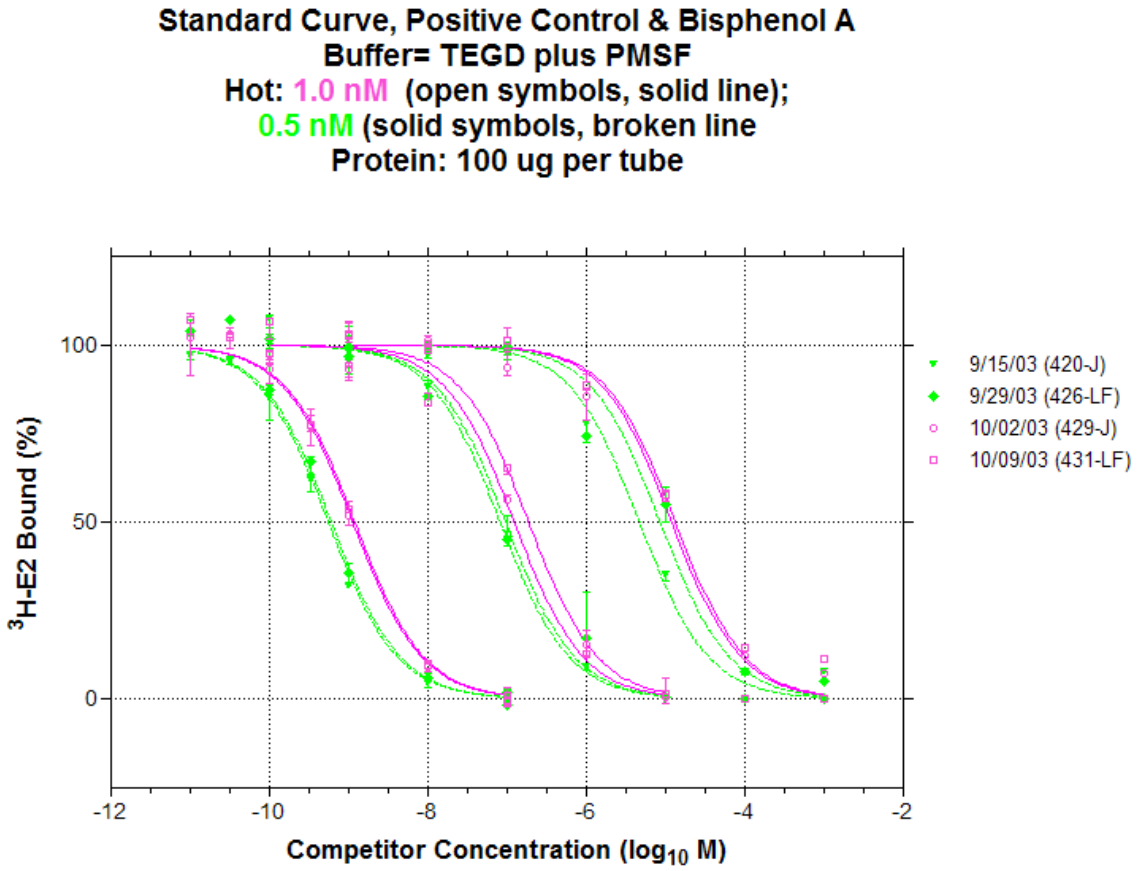


Figure 3. Competitive assay comparing cytosol storage time for estradiol (left), norethynodrel (center) and bisphenol A (right).

G. Concentration of radiolabeled estradiol

The radiolabeled estradiol concentration was tested at 0.5 and 1.0 nM to determine if sensitivity of the competitive binding assay is improved at lower [^{3}H]-estradiol concentrations. The TEDG + PMSF buffer and the HAP separation system were used for these studies. The maximum bound dpms in the 100% tubes ranged from 6000 to 6500 for 0.5 nM [^{3}H]-estradiol and 6700 to 7500 for 1.0 nM [^{3}H]-estradiol. The bound counts using 0.5 nM represent a ~12-15 % decrease from data obtained using 1.0 nM. The observed RBAs for both the positive control and bisphenol A were comparable across the two concentrations of radiolabeled estradiol. In both assay conditions (0.5 and 1.0 nM [^{3}H]-estradiol) ligand depletion did not exceed 10%. Given that the RBAs were similar under the two conditions, the [^{3}H]-estradiol was kept at 1.0 nM in order to maximize the number of counts available (and thus reduce variability).



	Standard Curve	Positive Control	Bisphenol A
	$log_{10}IC50$	$log_{10}IC50$	$log_{10}IC50$
9/15/03 (420-J)	-9.255	-7.088	-5.330
9/29/03 (426-LF)	-9.217	-7.021	-5.068
10/02/03 (429-J)	-8.950	-6.908	-4.919
10/09/03 (431-LF)	-8.922	-6.726	-4.873

Figure 4. Competitive assay comparing radioligand concentration for estradiol (left), norethynodrel (center) and bisphenol A (right).

H. Maximum solvent concentration

The effect of increasing concentrations of absolute ethanol and of DMSO on the binding of 17β-estradiol in the competitive binding assay was examined (Eldridge 2007, attached as Appendix 3). The IC_{95} was approximately 5% ethanol and 20% DMSO. To be conservative, EPA chose 3% ethanol and 10% DMSO to use as the maximum solvent concentrations.

IV. Interlaboratory validation

Following assay optimization, the protocol was tested for transferability to laboratories and for reliability across laboratories. Due to wider than expected intralaboratory variability in the first interlaboratory study -- which resulted in dropping two of five laboratories as well as higher coefficients of variation than expected for RBAs -- the protocol was modified and a second interlaboratory validation study was undertaken.

A. First interlaboratory study

1. Selection of laboratories

Four independent laboratories were recruited that had experience in reliable performance of *in vitro* receptor binding assays using biological materials that they obtained from appropriate biological tissues, although the experience did not have to be specifically with estrogen receptor binding assays. The fifth laboratory was the laboratory that had produced the optimization data described above.

2. Study design

Four steps were taken in this study to evaluate the intra- and inter-laboratory variability of results among five independent laboratories:

a. Saturation and competitive assays using centrally supplied cytosol (reference and weak positive chemical only)
b. Competitive assays using centrally supplied cytosol (9 test chemicals)
c. Saturation and competitive assays using cytosol prepared by individual laboratory (reference and weak positive chemical only)
d. Saturation and competitive assays using cytosol prepared by individual laboratory (5 test chemicals)

[^{3}H]-estradiol and dilutions of 17β-estradiol and norethynodrel were centrally supplied. Cytosol was either centrally supplied (series "a" and "b") or prepared by the individual laboratories (series "c" and "d"), to examine the laboratories' proficiency in preparing cytosol and running the complete assay. The centrally supplied cytosol was prepared by Lab E.

The results of this study are shown for historical purposes only. Competitive binding data from the second interlaboratory validation study (Section IV.B.4) were analyzed differently from the data from this first validation study, so the two validation

studies are not directly comparable. Specifically, in the first study, the tops and bottoms of competitive binding curves were constrained to 100% and 0% respectively, while for the second study the tops and bottoms were not constrained when fitting the data.

3. Results

a. Saturation binding

Saturation data for this study were fit to the one-site binding model as described in Section II.E. The fit provided estimates of the dissociation constant ($\log(K_d)$) and maximum number of receptors (B_{max}). The intra- and inter-laboratory means and coefficients of variation were evaluated to ensure that each laboratory was using the rat uterine cytosol preparations correctly and could reliably measure the K_d and B_{max}. The goodness of fit to the one site model was also calculated.

The laboratories first ran the saturation assay with cytosol provided by the lead laboratory (series "a" and "b"). The protein concentration was 2.52 and 3.10 mg/mL for series a and b cytosol preparations, respectively. The intra-laboratory variability is reported in Table 9.

The mean goodness-of-fit to the one-site binding equation was 88% for the 27 runs. The range of B_{max} (fmole/100 μg) was 12.30 to 52.5 with a mean value of 36.5 for series a, and 20.9 to 56.7 with a mean of 40.1 for series b.. The range of $\log(K_d)$ values was -10.37 to -9.36 with a mean value of -9.99. The inter-laboratory variability was 11% for K_d.

The purpose of the series a and b saturation assays was to compare each lab's ability to accurately and reproducibly conduct the saturation assay on a standard cytosol preparation. Lab B was not able to successfully produce saturation assays consistent with the other laboratories and did not participate beyond the saturation assays of series a. The other laboratories were relatively consistent in these saturation assays.

For the series c and d saturation assays, the goal was to establish the variability of results among the remaining independent laboratories when cytosol was prepared by each laboratory. As noted, Lab B did not participate due to difficulties in the series a saturation assay. Lab A also did not participate due to difficulties in the series a and b competitive assays (discussed below). Lab E did not participate in series c studies as its cytosol had been characterized previously in series a and b.

The protein concentration of the cytosol preparations is shown in

Table 10. The protein concentration varied by a factor of approximately 2 between the laboratories.

The mean goodness-of-fit to the one-site binding equation was 96% for the 15 runs. The range of B_{max} (fmole/100 μg) was 23.89 to 55.6 with a mean value of 43.2 for series c, and 28.3 to 58.0 with a mean of 39.0 for series d. The range of log(K_d) values was -10.15 to -9.48 with a mean value of -9.85. The inter-laboratory variability was 11% for K_d.

Table 9. Intra-laboratory variability of the saturation assay with centrally supplied cytosol preparation (series a and b)

Statistic	Run	Lab A	Lab B	Lab C	Lab D	Lab E
log K_d (M)	1a	-10.18	-9.74	-10.14	-10.06	-9.90
	2a	-10.14	-10.37	-10.15	-9.96	-9.77
	3a	-10.20	-10.07	-9.36	-10.21	-9.85
	1b	-10.18	-	-10.08	-10.20	-9.65
	2b	-10.02	-	-10.10	-10.18	-9.63
	3b	-9.97	-	-9.86	-10.22	-9.57
Mean log K_d (M)		-10.11	-10.06	-9.95	-10.14	-9.73
CV K_d		17%	11%	8%	17%	13%
Mean B_{max} (fmole/100 µg)	a	35.1	16.2	46.0	49.1	35.9
CV B_{max}	a	13%	22%	17%	4%	4%
Mean B_{max} (fmole/100 µg)	b	25.1	-	54.2	48.5	32.6
CV B_{max}	b	23%	-	5%	12%	11%
Average Goodness of Fit		83%	77%	92%	83%	98%

Table 10. Intra-laboratory variability of the saturation assay using individual laboratory cytosol preparation (series c and d)

Statistic	Assay	Lab C	Lab D	Lab E
Protein Conc. (mg/mL)	c	1.60	2.83	
	d	1.79	3.52	3.55
log K_d (M)	1c	-9.97	-9.77	-
	2c	-10.03	-9.96	-
	3c	-9.99	-9.93	-
	1d	-9.82	-9.83	-9.48
	2d	-9.83	-10.00	-9.51
	3d	-9.83	-10.15	-9.65
Mean log K_d (M)		-9.91	-9.94	-9.55
CV K_d		16%	15%	13%
Mean B_{max} (fmole/100 µg)	c	54.5	32.0	-
CV B_{max}	c	3%	37%	-
Mean B_{max} (fmole/100 µg)	d	56.2	30.7	30.0
CV B_{max}	d	3%	6%	10%
Average Goodness of Fit		96%	95%	99%

b. Competitive binding - standards

The competitive assays in the first validation study followed the same structure as the saturation assays described in the previous section. The laboratories first ran the competitive assay with cytosol provided by the lead laboratory for the reference standards (series a and b). Each of the five participating laboratories conducted three independent competitive binding runs using a standard and a weak positive control, with three replicates at each concentration. The data for Laboratory B is not included in the analysis due to that lab's inability to complete sufficient acceptable runs in the time allotted in series a.

The estimated log(IC_{50}) for the standard (estradiol) and weak positive (norethynodrel) are shown in Table 11 for both series a and b. For series a, the range of log(IC_{50}) values for the standard was -9.12 to -8.80 with a median value of -8.90. The range of log(IC_{50}) values for the weak positive control was -6.78 to -6.39 with a median value of -6.59. The resulting RBAs ranged from 0.29% to 0.66% with a median value of

0.46%. The intra-laboratory CVs for RBA ranged from 15% to 36% with a median of 20%.

The inter-laboratory variability of the three competitive binding measurements was 0.4%, 1.6%, and 22.7% for the standard and weak positive log(IC_{50}) values and RBA, respectively. The variability in these measurements was fairly small as can be inferred from the small variability in the fitted one-site competitive curves (Figure 5).

For series b, the range of log(IC_{50}) values for the standard (estradiol) was -9.25 to -8.79 with a median value of -8.96. The range of log(IC_{50}) values for the weak positive control was -9.49 to -6.25 with a median value of -6.45. Lab A, Run 2, with the weak positive had the poorest fit to the one-site competitive curve and produced the smallest log(IC_{50}) value (Figure 6). The percentage bound for this run dropped to less than 20% at a log M concentration of -9 and continued to drop for two more concentrations. The percentage bound then increased at log M concentration of -6 and decrease in a pattern consistent with the two other runs. The weak positive results for this run were removed from the statistical analysis. Lab C, Run 2, weak positive results produced the second smallest log(IC_{50}) value. The standard and weak positive results for Lab D and Lab E were similar to each other. The resulting RBAs for the weak positive ranged from 0.21% to 0.57% with a median value of 0.27%.

Table 11. Intra-laboratory variability of the competitive binding assay with centrally supplied cytosol

Series "a"		Lab A	Lab C	Lab D	Lab E
log(IC_{50}) Standard	Mean	-8.94	-8.98	-9.17	-8.89
	CV	2%	1%	1%	1%
log(IC_{50}) Weak Positive	Mean	-6.57	-6.47	-6.57	-6.29
	CV	2%	4%	1%	1%
RBA	Mean	0.38%	0.33%	0.25%	0.25%
	CV	9%	38%	14%	11%

Series "b"		Lab A	Lab C	Lab D	Lab E
log(IC_{50}) Standard	Mean	-8.99	-8.93	-8.92	-8.91
	CV	2%	1%	1%	1%
log(IC_{50}) Weak Positive	Mean	-6.67	-6.53	-6.45	-6.67
	CV	2%	1%	1%	0%
RBA	Mean	0.51%	0.40%	0.35%	0.58%
	CV	36%	15%	25%	16%

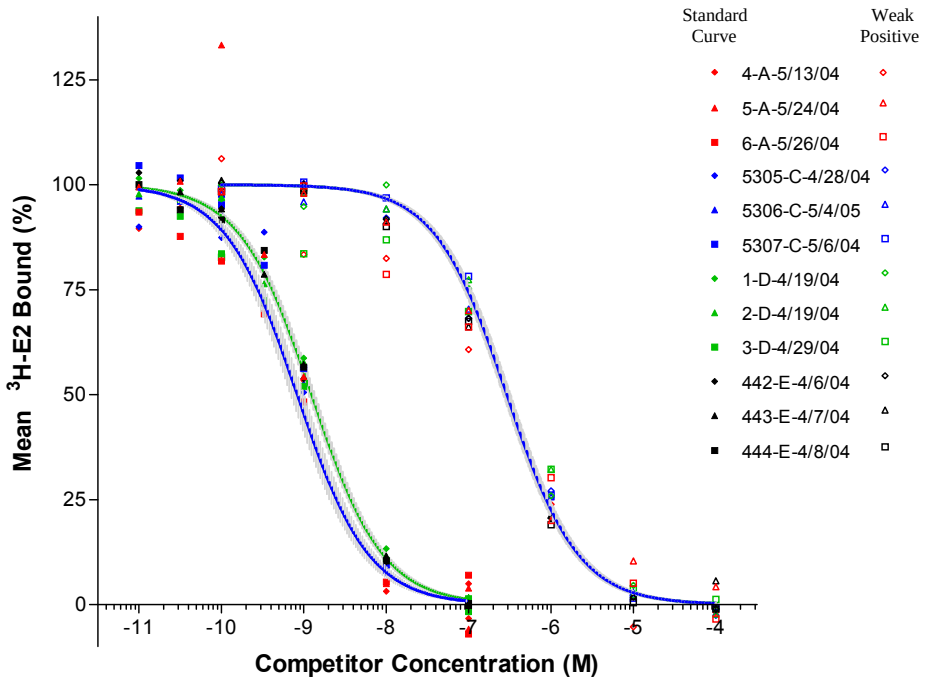


Figure 5. Inter-laboratory variability of standard and weak positive and associated 95% confidence bands (light gray) for Labs A, C, D and E (series a).

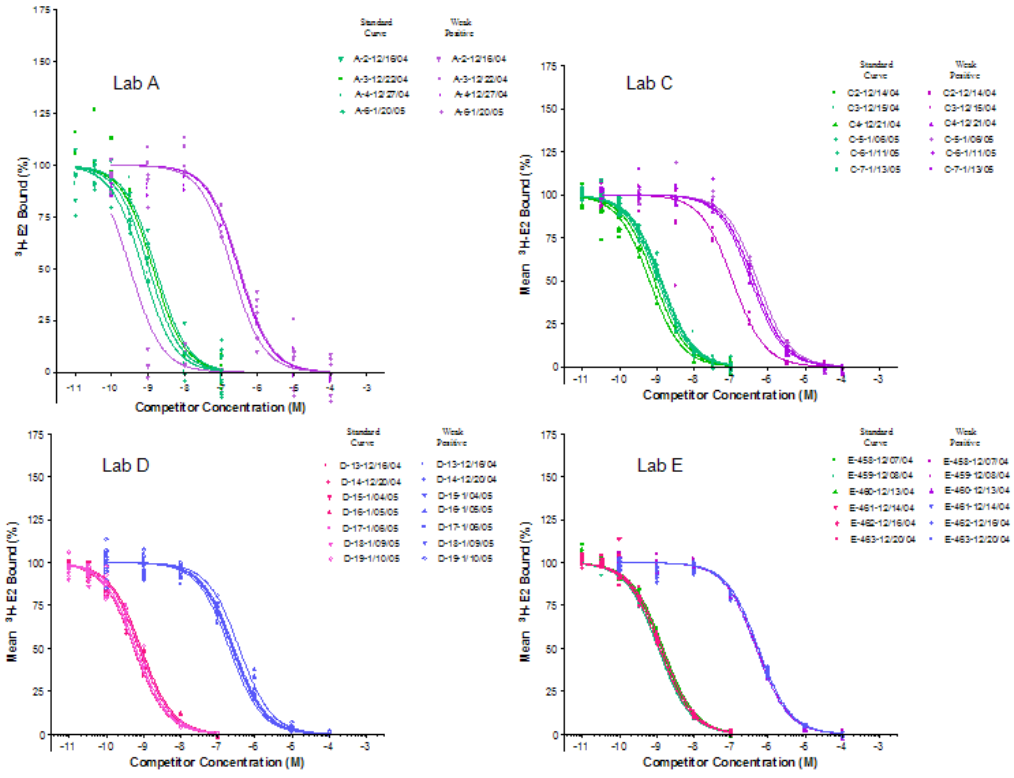


Figure 6. Intra-laboratory variability of the standard and weak positive chemicals in the competitive binding assay (series b).

As can be seen from Figure 6 the data from Lab A on estradiol and norethynodrel display somewhat more intralaboratory variability than the data from other laboratories, but it was the variability of the data on the test chemicals using centrally-supplied cytosol, discussed in section IV.A.3.c below, that led Lab A to be dropped from the remainder of the study.

For series c and d, Labs C and D conducted the competitive binding assay with the individual-laboratory-prepared cytosol. One run from Lab D was removed from statistical analysis because it did not meet certain protocol criteria. No other data were removed from statistical analysis.

The results for series c and d standard and weak positive are shown in Table 12 and Figure 7 and Figure 8. For Lab C, the mean log(IC_{50}) values were similar for both series for the standard and weak positive, with correspondingly similar RBA values. For Lab D, the mean log(IC_{50}) values differed between series but the RBAs were not extraordinarily dissimilar from the RBAs produced by the other laboratories.

Table 12. Competitive assay results for the individual-laboratory prepared cytosol, standard and weak positive chemicals

		Series "c"		Series "d"		
		Lab C	**Lab D**	**Lab C**	**Lab D**	**Lab E**
log(IC_{50}) Standard	Mean	-9.08	-9.02	-8.95	-9.68	-8.89
	CV	0.50%	0.80%	0.30%	1.10%	0.30%
log(IC_{50}) Weak Positive	Mean	-6.62	-7.06	-6.81	-6.77	-6.44
	CV	0.10%	1.90%	3.40%	1.50%	0.40%
RBA	Mean	0.35%	1.16%	0.79%	0.12%	0.36%
	CV	9.50%	38%	44%	10%	4.20%

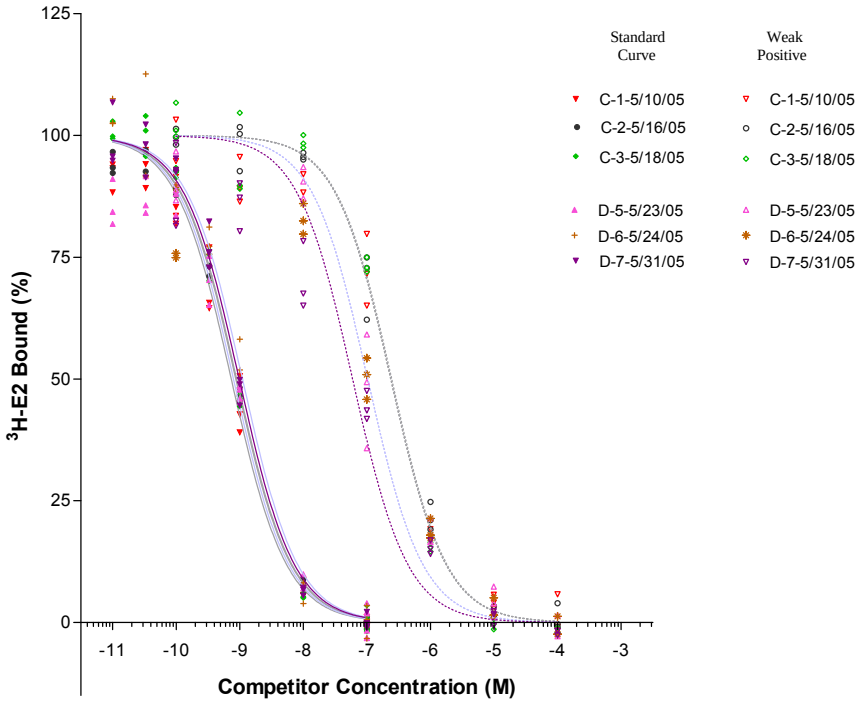


Figure 7. Competitive binding assay for individual-laboratory-prepared cytosol for the standard and weak positive chemicals in Labs C and D (series c)

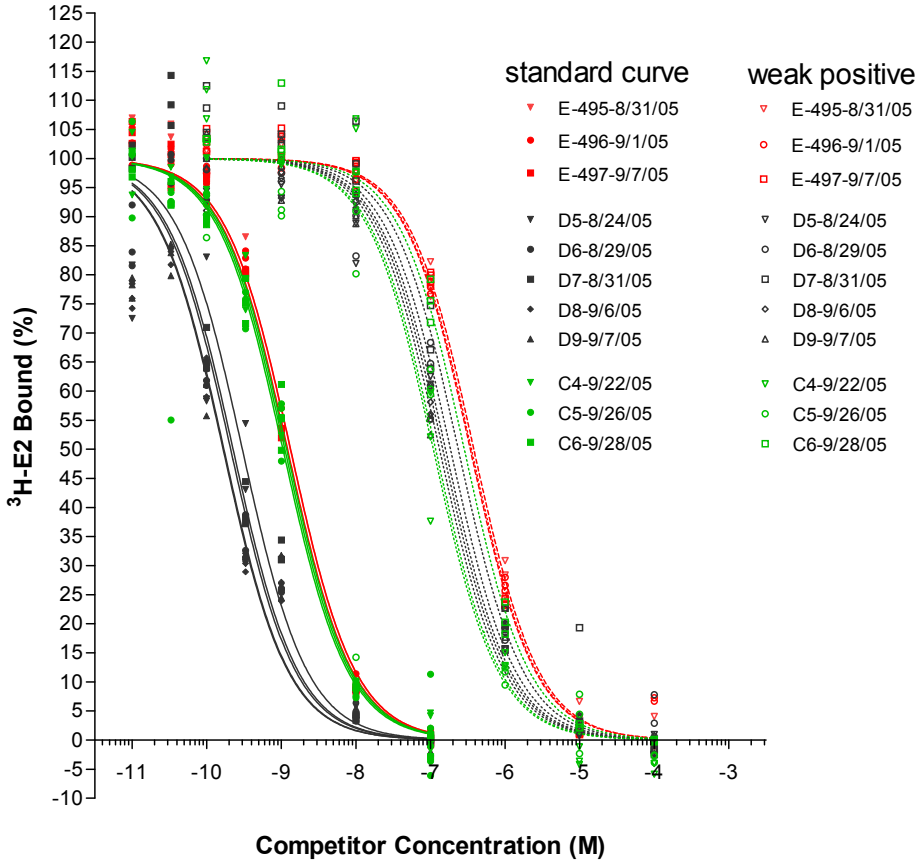


Figure 8. Competitive binding assay for individual-laboratory-prepared cytosol for the standard and weak positive chemicals in Labs C, D and E (Series d)

c. Competitive binding – test chemicals

In addition to the standard chemicals, the laboratories conducted competitive assays with centrally supplied cytosol (series b - 9 test chemicals) and laboratory prepared cytosol (series d - 5 chemicals).

The goodness-of-fit to the one-site competitive equation for the nine test chemicals reflected the characteristics of the chemical being tested. Thus, the lack of convergence did not cause any data to be removed from the analysis. Most of the R^2 values were greater than 80%. However, the progesterone results fit the curve poorly or did not converge (as would be expected from a substance that does not interact with the estrogen receptor). The mean log(IC_{50}), R^2, and RBA and intra-laboratory CVs for the standard, weak positive, and test chemicals are presented in Table 13. The intra-laboratory CVs for the weak positive RBA ranged from 10% to 38% with a median of

12%. The large CV of 38% was directly related to Lab C, Run 2. There was no obvious relationship between the test chemical run and the resulting intra-laboratory CV for the RBA (Figure 9). Each laboratory produced the smallest intra-laboratory CV for a different test chemical.

As can be seen in Table 13, Lab A had higher intralaboratory variability as measured by coefficients of variation, for both log(IC_{50})s and RBAs for most chemicals. While CVs were higher than expected for other labs for many of the chemicals, the difficulties that Lab A experienced were considered sufficient to disqualify it from further participation in the study.

The inter-laboratory CV for the standard and weak positive log(IC_{50}) and RBA in series b was 1.4% 2.0% and 20%, respectively (Table 14). Estrone had the largest inter-laboratory mean RBA (7.59%) of the test chemicals. The inter-laboratory CVs for the test chemical RBAs were similar and averaged 60%.

Table 13. Test chemical intra-laboratory variability using centrally-supplied cytosol (series b)

	Standard			Weak Positive			Bisphenol B			4-Cumylphenol		
Mean	log(IC_{50})	R^2		log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**
Lab A	-8.94	0.97		-6.57	0.96	0.38%	-6.07	0.94	0.18%	-4.85	0.90	0.01%
Lab C	-8.98	0.99		-6.47	0.98	0.33%	-6.16	0.98	0.14%	-5.16	0.97	0.01%
Lab D	-9.17	0.99		-6.57	0.99	0.25%	-6.72	0.97	0.41%	-5.44	0.97	0.02%
Lab E	-8.89	0.99		-6.29	0.99	0.25%	-6.09	0.99	0.16%	-4.64	0.95	0.01%
CV	log(IC_{50})			log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**
Lab A	2%			2%		9.49%	5%		91.0%	2%		45.3%
Lab C	1%			4%		38.1%	1%		49.4%	5%		34.2%
Lab D	1%			1%		14.0%	1%		27.0%	3%		44.4%
Lab E	1%			1%		10.6%	1%		3.41%	1%		12.4%

	Estrone			Coumestrol			Progesterone			Daidzein		
Mean	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**
Lab A	-7.70	0.93	5.38%	-7.50	0.92	4.0%	**No Convergence**			-4.96	0.80	0.01%
Lab C	-7.53	0.99	4.17%	-7.68	0.99	5.8%	**No Convergence**			-5.07	0.84	0.01%
Lab D	-8.30	0.98	11.6%	-8.22	0.98	11.9%	**No Convergence**			-5.64	0.90	0.03%
Lab E	-7.85	1.00	9.16%	-7.34	1.00	2.9%	-2.52	0.02	0.00%	-4.93	0.78	0.01%
CV	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**
Lab A	0%		38.9%	4%		64%	**No Convergence**			4%		57%
Lab C	1%		25.4%	1%		19%	**No Convergence**			1%		19%
Lab D	1%		25.4%	1%		34%	**No Convergence**			3%		32%
Lab E	1%		12.5%	2%		30%	12%		58.3%	2%		12%

	Tamoxifen citrate			4-t-Octylphenol			Bisphenol A		
Mean	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**
Lab A	-5.71	0.92	0.05%	-5.07	0.87	0.01%	-4.87	0.95	0.01%
Lab C	-6.52	0.90	0.40%	-5.29	0.98	0.02%	-5.06	0.98	0.01%
Lab D	-7.32	0.99	1.36%	-5.72	0.97	0.03%	-5.57	0.97	0.03%
Lab E	-6.50	0.97	0.43%	-5.19	0.97	0.02%	-4.84	0.97	0.01%
CV	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**
Lab A	1%		17%	3%		47%	3%		41%
Lab C	1%		7%	1%		8%	1%		31%
Lab D	1%		23%	1%		14%	1%		13%
Lab E	2%		44%	0%		14%	0%		12%

= Poor fit to the competitive binding curve

Table 14. Test chemical inter-laboratory variability (series b and d)

		Centrally supplied Cytosol (series b)		Individual-lab-prepared Cytosol (series d)	
Test Chemical	Statistic	Mean	CV	Mean	CV
Standard	$log(IC_{50})$	-8.99	1.40%	-9.17	4.78%
Weak Positive	$log(IC_{50})$	-6.47	2.00%	-6.68	3.03%
	RBA	0.30%	20%	0.42%	79%
Bisphenol B	$log(IC_{50})$	-6.26	5.00%		
	RBA	0.22%	57%		
4-Cumylphenol	$log(IC_{50})$	-5.02	7.00%		
	RBA	0.01%	56%		
Estrone	$log(IC_{50})$	-7.85	4.20%	-8.02	6.93%
	RBA	7.59%	45.20%	8.92%	48%
Coumestrol	$log(IC_{50})$	-7.69	5.00%		
	RBA	6.15%	65%		
Progesterone	$log(IC_{50})$	-2.52	NA	**No Convergence**	
	RBA	0.00%	NA	**No Convergence**	
Daidzein	$log(IC_{50})$	-5.15	6.40%		
	RBA	0.02%	52%		
Tamoxifen citrate	$log(IC_{50})$	-6.51	10.10%	-7.54	10.90%
	RBA	0.56%	100%	6.69%	149%
4-t-Octylphenol	$log(IC_{50})$	-5.32	5.30%		
	RBA	0.02%	37%		
Bisphenol A	$log(IC_{50})$	-5.09	6.60%	-5.42	12.80%
	RBA	0.01%	66%	0.02%	69%
4-t-Butylphenol	$log(IC_{50})$			-3.83	17.60%
	RBA			0.00%	84%

= Poor fit or lack of convergence

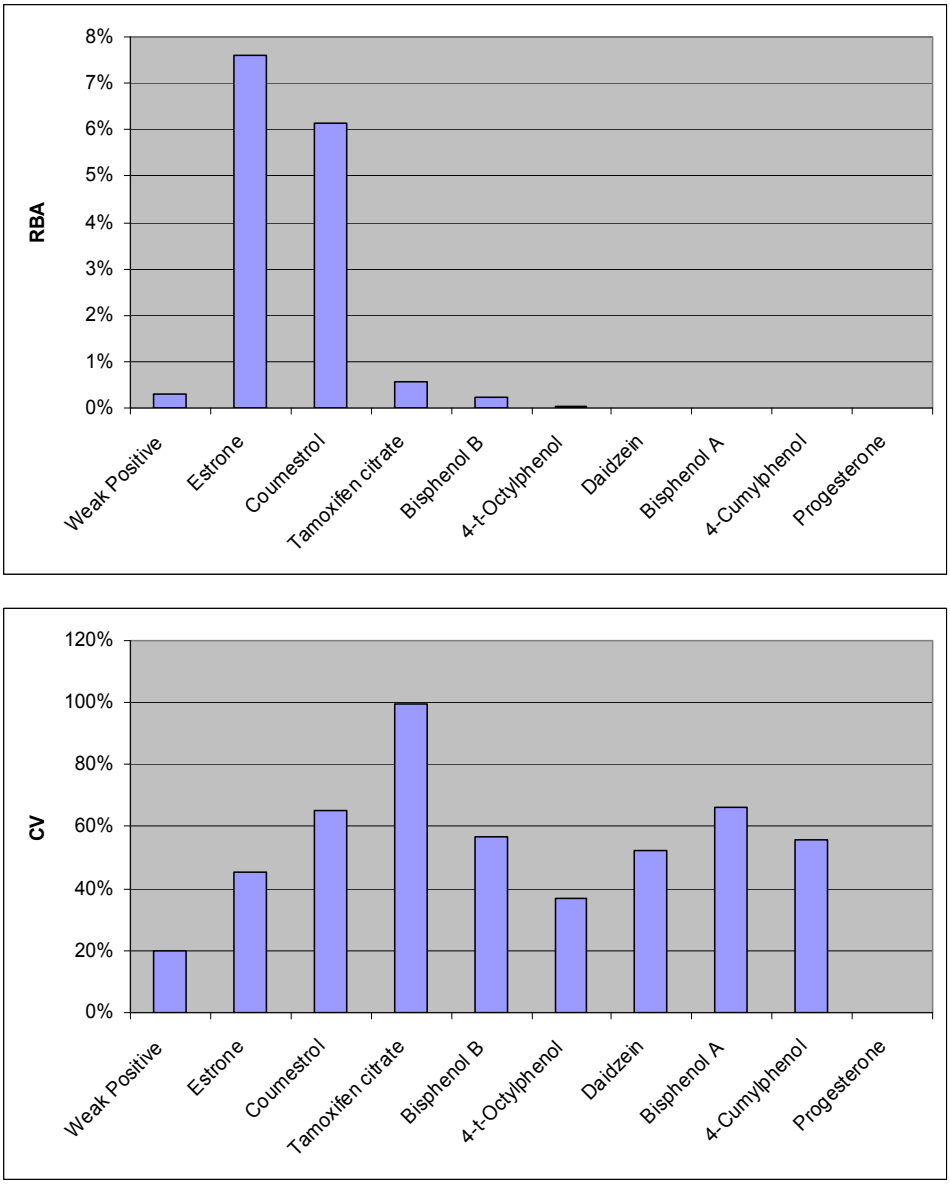


Figure 9. Inter-laboratory mean (top) and CV (bottom) for the weak positive and nine test chemical RBAs in order of greatest to least binding (series b)

For series d, the goodness-of-fit to the one-site competitive equation for the five test chemicals reflected the characteristics of the chemical being tested. Thus, the lack of convergence did not cause any data to be removed from the analysis. Most of the R^2 values were greater than 70%. However, the progesterone results fit the curve poorly or did not converge. The mean log(IC_{50}), R^2, and RBA and intra-laboratory CVs for the standard, weak positive, and test chemicals are presented in Table 15. The intra-laboratory CVs for the weak positive RBA ranged from 4% to 44% with a median of 9.6%. The large CV of 44% was directly related to Lab C, Run 6. There was no

obvious relationship between the test chemical and the intra-laboratory CV for the RBA (Figure 10). Lab E had intra-laboratory CVs consistently less than 11%, and Lab C averaged about 45%.

The inter-laboratory CV for the standard and weak positive log(IC_{50}) and RBA in series d was 4.78% 3.03% and 79%, respectively (Table 14). The greatest difference in RBA for the weak positive was between Lab C and Lab D with Lab E results between the two. Estrone had the largest inter-laboratory mean RBA (8.92%) of the test chemicals. The inter-laboratory CV for the test chemical RBAs was the least for estrone (48%) and the greatest for tamoxifen citrate (149%).

Table 15. Test chemical intra-laboratory variability using individual-laboratory-prepared cytosol (series d)

	Standard		Weak Positive			4-t-Butylphenol			Tamoxifen citrate		
Mean	log(IC_{50})	R^2	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**
Lab C	-8.95	98%	-6.81	92%	0.79%	-4.01	97%	0.00%	-6.61	97%	0%
Lab D	-9.68	93%	-6.77	99%	0.12%	-4.39	91%	0.00%	-7.88	92%	1%
Lab E	-8.89	99%	-6.44	100%	0.36%	-3.08	84%	0.00%	-8.15	100%	18%
CV	log(IC_{50})		log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**
Lab C	0.27%		3.41%		43.6%	5.76%		43.6%	3.09%		54.3%
Lab D	1.09%		1.45%		9.62%	3.16%		8.37%	0.69%		3.01%
Lab E	0.25%		0.38%		4.15%	0.86%		7.67%	0.50%		10.6%

		Bisphenol A			Estrone			Progesterone		
Mean		log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**	log(IC_{50})	R^2	**RBA**
Lab C		-5.23	0.82	0.02%	-7.56	0.89	4.49%	**No Convergence**		
Lab D		-6.19	0.85	0.04%	-8.64	0.94	13.1%	**No Convergence**		
Lab E		-4.84	0.99	0.01%	-7.85	1.00	9.22%	**No Convergence**		
CV		log(IC_{50})		**RBA**	log(IC_{50})		**RBA**	log(IC_{50})		**RBA**
Lab C		2.37%		27.0%	3.1%		56.2%	**No Convergence**		
Lab D		1.59%		44.1%	4.5%		84.7%	**No Convergence**		
Lab E		0.76%		8.15%	0.3%		7.2%	**No Convergence**		

= Poor fit or lack of convergence to the competitive binding curve

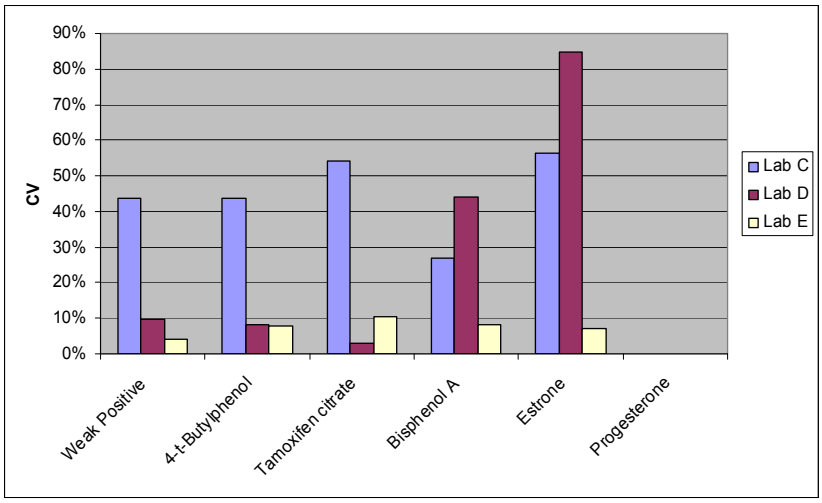


Figure 10. Intra-laboratory CV for five test chemicals and the weak positive RBAs (series d)

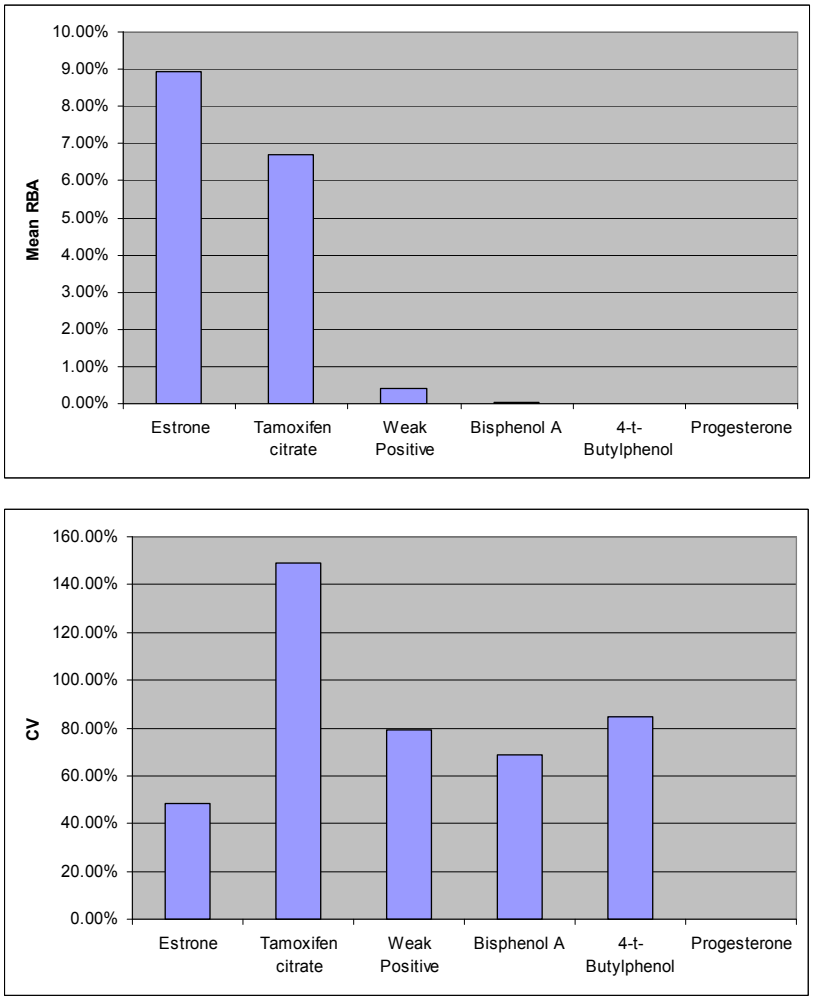


Figure 11. Inter-laboratory mean (top) and CV (bottom) for the weak positive and five test chemical RBAs in order of greatest to least binding (series d)

The intralaboratory variability was higher than expected in most laboratories. Laboratory E produced the most consistent and acceptable results, but this lab had worked closely with EPA on previous ER binding assay studies (most notably the optimization studies) and thus could have benefited from "coaching" that was not written in the protocol. It was expected that most laboratories should be able to produce the low variability exemplified by Laboratory E if the protocol were written in sufficient detail. Thus the protocol was significantly expanded with details and examples, and a second interlaboratory study undertaken.

B. Second interlaboratory study

For this study, significantly more detail was provided in the protocol. For example, specific dilution "recipes" were provided and analysis templates were improved. Attention was drawn to particular steps that need particular care, such as use of multi-tube decanting racks during separation of bound from free tracer in order to minimize the time that receptor is exposed to room temperature. Also, the procedure for making dilutions of test chemical was changed. Dilutions in the first study were made in small quantities of solvent first, and a prescribed amount of each dilution was added to the assay tube to obtain the final concentrations. Since all the dilutions were made separately this might have accounted for a significant amount of variability in the results in the first interlaboratory study. Therefore in the second study one stock solution was made in solvent and this solution diluted sequentially with buffer.

The study was conducted using fewer laboratories (three rather than five), but more chemicals (23 rather than 9). Each laboratory prepared its own cytosol; there was no centrally-prepared cytosol in this study.

An attempt was made to establish a base level of competence in performing the assay before allowing full participation in the study. Laboratories were to qualify by meeting performance criteria that had been established from acceptable runs generated in the first interlaboratory study.

1. Selection of laboratories

Laboratories were required to be competent in laboratory methods relevant to *in vitro* receptor binding assays. They were also required to be mutually independent, and

unconnected with the first interlaboratory validation study so that transferability of the new protocol could be assessed without influence from prior experience with the assay.

Also in order to evaluate the transferability of the protocol, EPA requested that "coaching" of laboratories be minimized and that all questions about the protocol be referred to the EPA rather than discussed between the labs. EPA refrained from visiting labs to observe technique and offer suggestions for improvement for most of the study, but towards the end of the study such a trip was made to two labs to observe technique during the critical stages of separation of bound from free radioligand, in order to try to pinpoint the cause of the variability that had been seen in the data. No obvious cause was found, and no significant changes were recommended.

As noted above, laboratories were to demonstrate proficiency in performing this assay by meeting performance criteria during a qualification step, before generating data in the main study. The performance criteria are shown in Table 16. They were set separately for estradiol and norethynodrel, based on the performance of Labs C, D, and E of the first interlaboratory validation study. They were meant to include 80% of the values from these labs, with 95% confidence. Details of the method used are provided in Appendix 4. Table 17 through Table 20 describe the results of this qualification step.

Table 16. Performance criteria for second interlaboratory study

Parameter	Unit	Estradiol		Norethynodrel		R1881	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
$SD_{within\text{-}run}$ (within-run variation)	% binding	NA	5.0	NA	5.7	NA	10
Bottom plateau level	% binding	-5.0	1.0	-5.0	1.0	NA	NA
Top plateau level	% binding	90.0	110.0	90.0	110.0	NA	NA
(Hill) Slope	$\log_{10}(M)^{-1}$	-1.1	-0.7	-1.1	-0.7	NA	NA

Table 17. Qualification runs: Saturation binding assays

	Lab X	Lab Y	Lab Z
Number of runs	4	5	5
K_d	0.1137 [nM] (CV 13.1%), 0.461 [nM] (CV 2.89%)	0.13 ± 0.006 nM (mean of 5 runs)	0.6588 nM* [CV 29.5%] (last assay) *Lab reported "6.69 E+08", unitless. EPA retrieved number above from data file.
B_{max} (fmole/100µg protein)	51.99 (CV 3.6%), 89.97 (CV 8.3%)	25.85 ± 0.56 (mean)	4918 dpm** [CV 21.5%] (last assay) **Lab reported "5.487 E+15", unitless. EPA retrieved number above from data file.
Protein/assay tube (µg)	50, 46	10 or 25	37, 37, 40, 50

Table 18. Qualification runs: Competitive binding assays

	Lab X	Lab Y	Lab Z
Number of runs	7	5	4
K_i	0.2865 0.3289 0.3565 0.4037		
Log(IC_{50}) for 17β-estradiol (M)	-9.042 -8.982 -8.947 -8.893	-9.04 ± 0.04	-9.7 -8.9 -9.09 -8.92 -8.90
Log(IC_{50}) for norethynodrel (M)	-6.083 -6.042 -5.971 -5.859	-6.01 ± 0.05	-6.11 -5.98 -5.89
RBA (norethynodrel compared with estradiol)	1.099 E-03 1.148 E-03 1.06 E-03 9.247 E-04	9.20E-04 ± 6.63E-05	1.718 E-10 1.216 E-09 6.339 E-04 1.271 E-09
Protein/assay tube (µg)	50, 35, 35, 46, 46, 46, 46	25	First assay 40 µg, next 3 at 50 µg

Table 19. Qualification runs: Summary of estradiol competitive binding data

Lab X					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	-1.1	103	-8.89	1.29	-0.99
	-0.7	100	-8.95	1.12	-0.97
	-1.7	104	-8.98	1.05	-0.91
	-1.4	103	-9.04	0.91	-0.89
Mean	-1.2	103	-8.97	1.09	-0.94
SE	0.5	1		0.08	0.02
Lab Y					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	0.1	88	-9.16	0.69	-1.10
	0.3	101	-9.05	0.89	-1.40
	0.3	103	-8.94	1.15	-0.98
	1.1	119	-9.01	0.98	-1.10
	-0.1	112	-9.05	0.89	-0.90
Mean	0.3	105	-9.04	0.92	-1.10
SE	0.2	5		0.07	0.08
Lab Z					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	-1.6	110	-8.92	1.20	-0.92
	-0.3	110	-9.09	0.81	-1.07
	0.0	117	-8.90	1.26	-0.99
Mean	-0.6	112	-8.97	1.09	-0.99
SE	0.5	2		0.14	0.04

Table 20. Qualification runs: Summary of norethynodrel competitive binding data

Lab X					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	2.8	102	-5.86	1380	-0.84
	2.8	99	-5.97	1072	-1.01
	-2.3	100	-6.04	912	-0.91
	-0.2	94	-6.08	832	-1.04
Mean	0.8	99	-5.99	1049	-0.95
SE	1.3	2		121	0.05
Lab Y					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	-5.9	100	-6.16	687	-1.08
	-5.2	111	-6.11	785	-0.74
	-3.9	116	-5.88	1309	-0.66
	-1.2	113	-5.98	1057	-1.10
	-6.4	104	-5.93	1180	-1.06
Mean	-4.5	109	-6.01	1004	-0.93
SE	0.9	3		117	0.09
Lab Z					
	Bottom	**Top**	**Log(IC_{50})**	**IC_{50} (nM)**	**Hill Coef**
	3.5	110	-5.98	1047	-0.91
	13	110	-5.89	1288	-0.92
	6.5	105	-5.88	1318	-1.2
Mean	7.7	108	-5.92	1218	-1.01
SE	2.8	2	0.03	86	0.10

The highlighted cells show values which did not meet the performance criteria. Lab X met the performance criteria; the other two labs did not. Note that it was primarily the top plateaus that were exceeded.

While these results were disappointing, the study continued with these laboratories. Given the experience in finding these laboratories, finding other laboratories would have delayed the study -- and the Screening Program -- significantly, and the deviations were judged to be marginally acceptable in this context.

2. Selection of test chemicals

Chemicals were selected to cover a wide range of binding strengths, indicated by their Relative Binding Affinities (RBAs) for the estrogen receptor as reported in the literature. Strengths were assigned as follows:

- Very strong: RBA > 100 (where RBA of 17β-estradiol=100)
- Strong: RBA between 100 and 1
- Moderate: RBA between 1 and 0.1
- Weak: RBA < 0.1
- Negative: no RBA achieved

These assignments are only approximate as they are based on the median value of reported RBAs, some of which span two or more orders of magnitude and others of which are based on a single value. See Table 21. Values from human recombinant ER binding assays reported in the literature are also reported in Table 21 for comparison. RBAs from hrERα studies, where available, also vary by an order of magnitude.

The chemicals selected for this study had been agreed upon by an international group of experts that is managing the validation of the human recombinant estrogen receptor (hrER) binding assay under the auspices of the Organisation for Economic Cooperation and Development The EPA used this chemical list so that results could be compared between the ER-RUC study and the parallael hrERα study. A draft list of proposed chemicals was submitted by the OECD group to an independent Chemical Advisory Board (CAB) of three international experts. The CAB generally concurred with the strategy of getting a wide range of strengths and chemical structures but recommended removing estrone, a strong binder, from the proposed list and adding enterolactone and benz(a)anthracene (weak binders), and atrazine (non-binder but estrogen-active). The CAB noted that this places more emphasis on weak binders, increases structural diversity, and provides the possibility of differentiating receptor-based from non-receptor-based modes of estrogenic action. The CAB's recommendations were adopted, and the final list of chemicals is as recommended by the CAB with the exception of the negative control chemical. The ER-RUC study planned to use R1881 as the negative control, while the hrER study had decided upon dibutylphthalate (DBP) as its negative control. DBP was regarded as inappropriate for use in the ER-RUC study because of prior indications that it did not give a clear negative signal in the RUC assay (Zacharewski et al. 1998).

The test chemicals were coded before shipment to the laboratories so as not to reveal their identity, although molecular weight was disclosed so that molar concentrations could be prepared. 17β-Estradiol (the reference standard),

norethynodrel (the weak positive control), and R1881 (the negative control) were included among the blinded test chemicals.

Table 21. Chemicals selected for the second interlaboratory validation study

Code	Chemical	Binding affinity	RUC* Historical RBAs (ICCVAM)	hrERα historical RBAs	Description
			median RBA (range of values)	*median RBA (range of values)*	
1	17b-estradiol	Strong	100 (reference estrogen)	100 (reference chemical)	physiological estrogen
2	17a-ethynylestradiol	Very strong	148 (100-867)	no data	synthetic estrogen
3	DES	Very strong	124 (0.003 - 5000)	236 (66.7 - 468)	synthetic estrogen
4	meso-hexestrol	Very strong	234 (58 - 302)	no data	phenol (bisphenol)
5	genistein	Moderate	0.56 (0.45 - 0.67)	2.36 (0.7 - 5)	flavonoid, phytoestrogen
6	norethynodrel	Moderate	0.22 (0.2 - 0.23)	0.7 (tested once)	steroid, nonphenolic
7	butyl paraben	Weak	0.002 (0.0009 - 0.002)	no data	paraben
8	4-nonylphenol	Weak	p-nonylphenol 0.033 (0.0025 - 0.5)	0.026 (0.05, 0.001)	alkylphenol, intermediate cmpd
9	o,p'-DDT	Weak	0.013 (0.001 - 0.09)	0.055 (0.01 - 0.1)	organochlorine
10	corticosterone	Negative	negative (one study)	negative (one study)	steroid
11	equol	Moderate	0.15 (tested once)	0.33 (tested once)	phytoestrogenic metabolite
12	zearalenone	Strong	44.07 (tested once)	8.5 (7-10)	resorcylic acid lactone, mycotoxin
13	tamoxifen	Strong	3.1 (0.13 - 6)	4 (2.94 - 7)	antiestrogen
14	5a-dihydrotestosterone	Weak	0.0135 (0.001 - 0.26)	0.05 (tested once)	steroid, nonphenolic
15	Bisphenol A	Weak	0.056 (0.008 - 0.1793)	0.01 (0.003-0.05)	phenol
16	4-heptylphenol	Weak	no data	no data	alkylphenol
17	kepone (chlordecone)	Weak	0.03 (0.0035 - 0.2)	0.06 (tested once)	organochlorine
18	benz(a)anthracene	Weak	no data	no data	aromatic hydrocarbon
19	enterolactone	Weak(†)	no data	no data	phytoestrogen, lignan
20	progesterone	Negative	0.0003 (1 positive/8 tests)	no data	steroid
21	octyltriethoxysilane	Negative	no data	no data	silane
22	atrazine	Negative	1/2 studies negative	negative (one study)	triazine (herbicide)
23	R1881	Negative	no data	no data	classic androgen receptor binder

* RUC = Rat Uterine Cytosol (values are from ICCVAM's "Evaluation of *In Vitro* Test Methods for Detecting Potential Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays", NIH document No. 03-4503, May 2003).

† Enterolactone added per recommendation by independent Chemical Selection Board, which characterized it as a weak binder.

3. Preparation of rat uterine cytosol

The rat uterine cytosol (RUC) was prepared by the laboratories according to the protocol. The following table (Table 22) shows similarities and differences between the cytosol preparations. The wide range of protein concentrations is noteworthy.

Table 22. Rat uterine cytosol preparations

	Lab X	**Lab Y**	**Lab Z**
Source of rats	Charles River	Harlan	Taconic
Age of rats (days) at ovariectomy	94/94	88-93	84-90
Days after ovariectomy that uteri were removed	8/7	8	8-10
Strain of rats	Sprague-Dawley	Sprague-Dawley	Sprague-Dawley
Rats ovariectomized at source	yes	yes	yes
Number of cytosol preparations	2	3	5
Date ovariectomized	1/14/08, 3/4/08	2/19/08	2/8/08, 2/13/08, 2/22/08, 5/28/08
Uteri removed at source and frozen	no	yes	no
Dates uteri removed	1/22/08, 3/11/08	2/27/08	2/18/08, 2/21/08, 2/22/08, 3/3/08, 6/6/08
Dates cytosol made	1/23/08, 3/19/08	2/29/08, 4/20/08, 6/12/08	2/18/08, 2/21/08, 2/22/08, 3/3/08, 6/6/08
Protein concentration of cytosol batches (mg/mL)	2.2, 2.3	8.09, 6.67, 2.5	1.16, 3.16, 2.69, 2.75, 2.52

4. Results

The contractor's summary report of the results from the three labs is attached as Appendix 5, and the individual laboratories' reports are attached as Appendix 6 through Appendix 8.[1]

[1] Note, however, that the analyses done by the individual laboratories differ from the uniform analysis on which this ISR is based. Thus the summaries and graphs in the laboratories' reports are different from what is presented here. See Section IV.B.5 below.

a. Saturation binding

Saturation binding assays were performed on each batch of cytosol used in this study. The results are shown in Table 23 (Lab X), Table 24 (Lab Y), and Table 25 (Lab Z):

Table 23. Saturation binding results, Lab X

Assay ID	Cytosol Prep	Plateau reached ?	Linear Scatchard ?	K_d (nM)	RSE of K_d (%)	B_{max} (fmole/ 100 µg)	RSE of B_{max} (%)	NSB acceptable?[a]
Sat 1	1/23/08	Y	Y	0.1508	10.9%	71.03	3.2%	Y
Sat 2	1/23/08	Y	Y	0.1466	12.5%	58.02	3.9%	Y
Sat 3	1/23/08	Y	Y	0.1813	18.9%	59.84	5.6%	Y
Average				0.1596				
Sat 4	3/19/08	Y	Y	0.6457	6.8%	52.09	2.7%	Y
Sat 5	3/19/08	Y	Y	1.1810	28.8%	84.96	13.1%	Y
Sat 6	3/19/08	Y	Y	0.3691	7.0%	57.16	2.4%	Y
Average (4,5,6)				0.7319				
Average (4.6)				0.5074				

[a]NSB is acceptable if it is <50% of total binding.

Table 24. Saturation binding results, Lab Y

Cytosol Prep	µg Protein/tube	K_d (nM)	Bmax (fmol ER/100 µg protein)
3	25	0.042	10.99
3	50	0.061	17.43
4	50	0.138	15.17
5	50	0.235	36.46

Table 25. Saturation binding results, Lab Z

Cytosol Batch ID	Protein Concentration (mg/ml)	Est. K_d (nM)	Est. B_{max} (fmole/100 g protein)
RUC2-18-08	1.16	1.4	146.0
RUC2-21-08	3.16	1.1	85.2
RUC2-22-08	2.69	1.5	107.9
RUC3-3-08	2.75	1.0	41.2
RUC6-6-08	2.52	0.8	113.2

Scatchard plots are provided in the final reports (Appendix 6 through Appendix 8).

b. Competitive binding

Competitive binding experiments were performed by each of the three labs for the estradiol (strong) standard, the norethynodrel (weak) standard, and 23 test chemicals. EPA dropped R1881 from testing as a standard chemical (negative control) when it became clear that it was a weak binder. (This chemical was tested at higher concentrations in this study than previously tested by EPA, and this revealed its weak binding affinity for the ER.)

Each test run consisted of the estradiol and norethynodrel standards and one or more test chemicals. EPA required that all chemicals be run in ethanol as solvent for this study (with the exception of DMSO for benzanthracene) in order to facilitate comparisons across laboratories, but this requirement was not followed by the labs. In addition, apparently labs did not always run the standards in the same solvent as was used for the test chemical – a requirement of the protocol. Since the labs did not always report the solvent used for the standard even when the solvent for the test chemical was reported, EPA was unable to examine solvent effects.

Each laboratory analyzed its own data, evaluated the acceptability of individual runs, and classified the test chemicals as binders, non-binders, or equivocal according to the data interpretation guidelines in the protocol.

The labs did not adhere to the performance criteria for the standards as EPA had required. In particular, the tops of the standard estradiol curves often were significantly higher than allowed (e.g., 130% of binding in the solvent control tubes and higher). This may have been due to an unexplained drift downward in radioactivity counts between solvent control tubes at the beginning of the run and tubes at the end of the run.

5. Analysis

To ensure comparability of results across laboratories, EPA had the data reanalyzed uniformly, normalizing each run to the mean value of binding at the lowest concentration of estradiol standard for each run. (See the discussion in Section IV.B.6.) It is the result of this uniform analysis that is presented in this Integrated Summary Report.

The analysis for the second interlaboratory study was substantially different from the analysis for the first. In the analysis of the second study, tops and bottoms were not

constrained to 100% and 0%, respectively. Not constraining the tops and bottoms allows performance criteria to be used for these quantities, and provides a more realistic estimate of the Hill slope. Outliers were identified and excluded, using the automatic outlier elimination procedure that had not been available for the first study. For these reasons, the analyses of the first and second interlaboratory studies are not quantitatively comparable.

a. Model fitting

The renormalized percent binding values were calculated for each run and transferred to PRISM Version 5 for model fitting. The four-parameter one-site competitive binding model was fitted to the data using PRISM's algorithm for nonlinear least squares with automatic outlier detection. The algorithm is discussed in detail in Motulsky and Brown (2006). Details of the algorithm, free parameters in the algorithm, and PRISM default values are discussed in the detailed statistical methods report (Appendix 9).

Separate model fits were carried out for each test run of each of the standard chemicals and the 23 test chemicals. These fits are shown as Appendix 10 (Lab X), Appendix 11 (Lab Y), and Appendix 12 (Lab Z).

b. Evaluation of runs for acceptability

Since the original performance criteria were not followed, EPA reviewed the individual renormalized fits for the standard chemicals (estradiol and norethynodrel) and judged each run as "acceptable" or "unacceptable", as described in section IV.B.5.d

EPA also reviewed the test chemical runs for acceptability. It would have been better to accept test chemical runs if and only if both of the standard chemicals for that run were judged to be acceptable, but this would have resulted in too few acceptable runs to evaluate.

Test chemicals were evaluated as if they were true unknowns; that is, it was not assumed that these were one-site competitive binders whose slopes, tops, and bottoms were expected to conform to expected behavior. Instead, runs were discarded only if there was extreme scatter across data points; or if a run was markedly different from two runs, judged acceptable for that chemical, that were similar to each other. If unclear whether to keep or discard a run, the run was kept. Tables showing the EPA

evaluations for test chemicals compared to the evaluations of the standards for that run as well as to the labs' own designations of acceptability are included as Tables 14 (Lab X), 15 (Lab Z), and 16 (Lab Y) of the Detailed Statistical Report (Appendix 9).

c. Comparison across laboratories

For each test laboratory the individual test runs of the test chemicals were designated as "binder", "equivocal", "non-binder", "non-testable"[2]. The criteria are given below.

- Binder – A binding curve can be fitted with "slope approximately -1". The lowest point on the response curve within the range of the data is less than 50%.

  The criterion "slope approximately -1" was determined as follows: The set of estradiol standard fits for the runs that were considered by EPA to be acceptable runs were compiled across the test laboratories. There were 134 such runs. The range of slopes within these runs was defined to be the range of slopes that are "approximately -1". This range was -1.78 to -0.67.

- Non-binder – The run is testable (see discussion below concerning "non-testable") and
  a) a binding curve can be fitted (with any slope) and the lowest point on the fitted response curve within the range of the data is above 75%.
  or
  b) a binding curve cannot be fitted and the lowest average of replicates at any concentration is above 75%.

- Equivocal – Any testable run that is neither a binder nor a non-binder is equivocal. The run might or might not have a fitted model. In general, this category covers those compounds which appear to be interacting with the receptor at high concentrations but are so weak that they displace only 25 to 50% of the radioactive estradiol. It also covers those compounds that appear to have precipitous slopes since such slopes may be an artifact of the fitting algorithm rather than a reflection of the behavior of the chemical. Such curves deserve a closer look in the weight-of-evidence determination of interaction with the receptor.

[2] The terminology has been changed in the protocol to "interacting" and "not interacting" instead of "binder" and "non-binder" to reflect that this assay does not fully characterize the interaction. Nevertheless, the old terms are retained here.

- Non-Testable – There are no data points at or above concentration 10^{-6} M and one of the two following conditions hold:
  a) A binding curve can be fitted but the binding curve is not lower than 50% by concentration 10^{-6}.
  or
  b) A binding curve cannot be fitted and the lowest average of replicates at any concentration is above 50%.

The "non-testable" designation is meant to cover chemicals that could not be put into solution at a high enough concentration to determine whether or not they are binders. It does not cover strong binders, which typically reach 50% inhibition of the radioligand by 10^{-6} M.

The classifications for each run are shown in Table D (i.e., Appendix D) of the Detailed Statistical Report (Appendix 9).

Classifications for chemicals were obtained by assigning the following values to each run and averaging:

- Testable runs:
  binder = 2
  equivocal = 1
  non-binder = 0
- Non-testable runs: missing value, not used in averaging

The average score determined the overall binding category for that laboratory and test chemical as follows:

Binder:	Average ≥ 1.5
Equivocal:	$0.5 \leq$ Average < 1.5
Non-binder:	Average < 0.5

The binding categorizations were combined across labs by a majority voting rule. If two or more labs reported the same binding categorization it was taken as the overall binding categorization. If each lab reported a different binding categorization (e.g. Test chemicals 8, 14) the overall binding designation was reported as "Inconclusive".

The "Expected Affinity" of each test chemical is, as explained in Section IV.B.2, based on Relative Binding Affinities for the estrogen receptor reported in the literature.

Table 27 summarizes the classification of each chemical by lab, and compares results across labs. It also compares the majority result to the expected result.

Because there are only three categories in the assay (binder, non-binder, equivocal) but five categories in the expected affinity designation (very strong, strong, moderate, weak, negative), a map between the two was required. The convention adopted is displayed in Table 26.

Table 26. Correspondence between experimentally determined binding category and expected affinity

Binding Category	Expected Binding Affinities Considered Equivalent to the Binding Category
Binder	Very Strong, Strong, Moderate, Weak
Equivocal	Moderate, Weak, Negative
Non-binder	Negative

Table 27. Comparison of classifications across labs, ranked by expected affinity

Chemical Code	Chemical (all were run blind)	Expected Affinity	Majority of Labs	Lab X	Lab Y	Lab Z
2	17-Ethynylestradiol	Very Strong	Binder	Binder	Binder	Binder
3	DES	Very Strong	Binder	Binder	Binder	Binder
1	17beta-Estradiol (blinded)	Strong	Binder	Binder	Binder	*Equivocal*[3]
4	Meso-Hexestrol	Very Strong	Binder	Binder	Binder	Binder
13	Tamoxifen	Strong	Binder	Binder	Binder	Binder
12	Zearalenone	Strong	Binder	Binder	Binder	Binder
11	Equol	Moderate	Binder	Binder	Binder	Binder
5	Genistein	Moderate	Binder	Binder	Binder	Binder
6	Norethynodrel (blinded)	Moderate	Binder	Binder	Binder	Binder
16	4-n-heptylphenol	Weak	Binder	Binder	Binder	*Equivocal*
14	5alpha-Dihydrotestosterone	Weak	Inconclusive	*Equivocal*	*Non-Binder*	*Binder*
18	Benz(a)anthracene	Weak	Inconclusive	*Non-Binder*	*Binder*	*Equivocal*
15	Bisphenol A	Weak	Binder	Binder	Binder	Binder
7	Butyl paraben	Weak	Binder	Binder	Binder	Binder
19	Enterolactone	Weak	Binder	Binder	Binder	*Equivocal*
17	Kepone (Chlordecone)	Weak	Binder	Binder	Binder	Binder
8	Nonylphenol (mixture)	Weak	Binder	Binder	Binder	[4]
9	o,p'-DDT	Weak	Binder	*Equivocal*	Binder	Binder
22	Atrazine	Negative	Equivocal	Equivocal	*Non-Binder*	Equivocal
10	Corticosterone	Negative	Binder	Binder	Binder	Binder
21	Octyltriethoxysilane	Negative	Non-Binder	Non-Binder	Non-Binder	Non-Binder
20	Progesterone	Negative	Equivocal	Equivocal	*Non-Binder*	Equivocal
23	R1881	Negative[5]	Binder	Binder	Binder	Binder

Legend	
Black (standard) font	Agreement with the "Majority of Labs"
Red (italics) font	*Disagreement with the "Majority of Labs"*
Green cell (dark shading)	Agreement with the "Expected Affinity"
Yellow cell (light shading)	Disagreement with the "Expected Affinity"

[3] Lab Z did not test more-dilute solutions than the default and thus did not obtain a full curve. The data that it obtained showed full displacement of the radioligand at higher concentrations. The protocol has been adjusted to emphasize that a full curve must be obtained where there are clear indications of binding.
[4] Lab Z had no acceptable test runs for test chemical 8.
[5] During the validation study, higher concentrations of R1881 were tested than previously used and the chemical was demonstrated to be a binder.

d. Development of performance standards

The results of each individual run for estradiol and norethynodrel were reviewed by EPA and were designated as "acceptable" or "unacceptable". These judgment calls were based on review of fitted curves and the numerical values for slopes, tops, and bottoms as provided by the Prism software, based on data normalized to the binding value of the run's lowest estradiol concentration. Graphs superimposing all of the runs judged "acceptable" for each lab, estradiol separately from norethynodrel, are attached as Appendix 13. EPA's designations were used in the determination of performance criteria for the standards. The performance criteria were defined as the tolerance bounds that include 80% of the acceptable runs with 95% confidence, for each of the binding curve parameters (top, bottom, slope, residual standard deviation; and in the case of the weak positive, the RBA), across all of the laboratories. Graphs showing where each of the runs (both acceptable and unacceptable) falls in relation to the tolerance bounds are included as Figures 1 through 9 of the Detailed Statistical Report (Appendix 9).

The tolerance bounds reflect the variance components among the runs. The laboratories were assumed to be a random sample from the population of acceptable laboratories that might carry out the ER assay in the EDSP. The laboratories were treated as random effects. The variance components incorporated in the tolerance bounds are:

- Within run variation (reflected in the standard errors of the fitted binding curve parameters)
- Run to run variation within labs
- Lab to lab variation

The variance components were estimated by mixed models analysis of variance. The analysis of variance models and fitting methods are discussed in Appendix 9.

The tolerance bounds are normal theory tolerance bounds, constructed to include 80% of the population of parameter values with 95% confidence. The tolerance bounds for top, bottom, Hill slope, and $\log_{10}$ (Relative binding affinity) are two sided bounds. The tolerance bounds for residual variation are one sided upper bounds.

The tolerance bounds are summarized in Table 28 and Table 29.

Table 28. Slope, top, bottom, RBA: Tolerance interval bounds to contain at least 80% of population of test runs with 95% confidence. Outliers deleted.

Chemical	Parameter	Average Estimate	Std Error Estimate	Lower Tolerance Limit	Upper Tolerance Limit
Estradiol	Hill Slope	-1.0006	0.02319	-1.3434	-0.658
	Top	102.43	0.5857	94.3303	110.533
	Bottom	-1.4549	0.1611	-3.1906	0.281
Norethynodrel	Hill Slope	-0.9874	0.03832	-1.4048	-0.570
	Top	99.8434	1.7742	74.0351	125.652
	Bottom	1.4100	1.6191	-19.1099	21.930
	$\log_e$RBA	-2.9256	0.01889	-3.22953	-2.62167

Table 29. Ln(residual standard deviation), residual standard deviation: Tolerance interval bounds to contain at least 80% of population of test runs with 95% confidence. Outliers deleted.

Chemical	Mean Log(Syx)	Std Error Log(Syx)	Upper Tolerance Limit Log(Syx)	Upper Tolerance Limit Syx
Estradiol	1.2999	0.1964	2.34866	10.4715
Norethynodrel	1.6175	0.1927	2.59228	13.3601

The tolerance bounds for estradiol are remarkably similar to the performance criteria established after the first interlaboratory study (Table 16) but the values for norethynodrel are significantly wider. Although the reason for this is not known, the differences might be attributable to solubility problems with norethynodrel. The laboratories in the second study reported consistent difficulty in preparing the highest concentration of norethynodrel and were often forced to drop that concentration from the run. If the highest concentration that ultimately was used was not fully soluble, the serial dilutions prepared from that mixture would not have the expected concentrations, thus shifting the plotted curve. Also, the lack of a value at the highest concentration (10^{-4} M) could have caused the curve to be fit differently than it would have been if that data point had been available. A further difference between the first study and the second was that in the second study dilutions were made serially in assay buffer whereas in the first study they were done serially in ethanol so that all of the concentrations had the same amount of solvent. Labs in the first interlaboratory validation study did not report difficulties in preparing the 10^{-4} M concentration, and a review of the data from that study (not shown) indicates only a few isolated runs where solubility might have been a problem. Pending an ongoing investigation into the solubility of norethynodrel in this assay, EPA is retaining the more restrictive performance criteria established from the first study.

6. Discussion

Results generally matched expectations (Table 27). All of the very strong, strong, and moderate binders were correctly determined to interact with the estrogen receptor, and seven of nine weak binders were also correctly identified as interacting. Most chemicals that were expected not to be binders were “equivocal” in their responses, and in the case of at least one if not both of the chemicals that were expected to be negative but consistently showed interaction (R1881 and corticosterone) it is likely that the expectation of non-binding was incorrect. One chemical (octyltriethoxysilane) produced consistently negative results across all three laboratories indicating that the assay can correctly identify compounds that do not interact with the estrogen receptor.

There are, however, several results in Table 27 that deserve comment. The finding that estradiol was "equivocal" rather than clearly interactive in one laboratory when tested blindly appeared to be a serious deficiency of the assay but is explained by the fact that the laboratory did not adjust the test concentration range to a more dilute range that would have allowed characterization of the full binding curve as required by the protocol. Binding at the concentrations tested showed clear interaction with the receptor. Although the data interpretation procedures could be adjusted to accommodate cases like this where there is clear interaction even though the top is not fully characterized, the Agency finds it more appropriate to require that the full curve be generated than that partial datasets be accepted. This will increase confidence that the chemical has been adequately characterized, by minimizing the effect of variability over a reduced set of datapoints.

The highest concentration at which a chemical is tested in this protocol (1 mM) may help explain why several chemicals that were expected to be negative produced equivocal responses or even showed evidence of interaction with the estrogen receptor. As noted in the Background Review Document, "[h]istorically, the highest dose tested … has ranged generally from 1 to 100 μM, with some tests conducted at doses as high as 1 mM." Thus, some chemicals that were reported in the literature to be negative may have shown evidence of interaction had they been tested at this high concentration.

The EPA included the 1 mM concentration based on an analysis in the BRD. The BRD explains that the ability of the assay to identify weakly positive chemicals rises with the highest concentration tested, and notes that "…if testing for ER binding substances requires the ability to detect substances with an IC_{50} that is at least six orders of magnitude lower than that of 17β-estradiol, then the limit dose…should be above 4 mM (e.g., 10 mM) to allow for the detection of an IC_{50} in the concentration range of interest. However, if five orders of magnitude are sufficient for RBA values, then the limit dose would have to be above 400 μM (e.g., 1 mM). Decreasing the limit dose to 100 μM would limit the sensivity of the assay to RBA values that cover approximately four orders of magnitude." The EPA believes that five orders of magnitude is appropriate for this Tier 1 screening assay and thus is requiring 1 mM concentration (where achievable in solvent). The EPA recognizes that the 1 mM

concentration is of questionable relevance to *in vivo* systems, and that other interactions besides one-site competitive binding may be occurring at such levels. Such interactions would be "false positives" for one-site competitive binding even though they are indeed "interactions" with the estrogen receptor. The EPA finds the emphasis on "interaction" rather than one-site competitive binding *per se* to be appropriate for a Tier-1 screening assay, and preferable to attempting to identify the specific mechanism of interaction using this assay. Since the data on octyltriethoxysilane produced unequivocal negative responses in all three laboratories, it is known that the use of 1 mM as the highest concentration will not result in positive results for all chemicals. The assay is specific to interaction with the estrogen receptor.

The protocol calls for a negative control in each run (along with reference chemical and weak positive control) but except for a few runs in the beginning, no negative controls were included in this study. This is because R1881, the chemical chosen to be the negative control, turned out to be a weak binder during the first few runs in this study. As discussed in the previous paragraph, testing at higher concentrations than previously tested may show that a chemical previously thought not to interact is a weak binder. Evidence that R1881 interacts with the estrogenic system (i.e., is not just anti-androgenic) includes a positive uterotropic study (Ojasoo and Raynaud 1978). This suggests strongly that R1881 indeed interacts with the estrogen receptor rather than that the ER-RUC assay falsely characterized a non-interactor as an interactor. EPA has replaced R1881 with octyltriethoxysilane, a compound which tested consistently negative in all laboratories, as negative control.

In the analyses produced by the individual laboratories in this study, the top plateaus for the standard chemicals (estradiol and norethynodrel) often exceeded the performance criteria by several tens of percentage points. The reason for the high plateaus was not determined but may be related to the solvent control tubes. Such tubes placed at the end of the run (of several test chemicals run simultaneously) often yielded lower dpms than similar tubes placed at the beginning of the run. The average of all solvent control tubes was therefore lower than it would have been had only the first solvent control tubes been included. The lower average could have contributed to the appearance of higher-than-solvent-control values for the estradiol and norethynodrel

standards, which were always placed at the front of the runs. To compensate for this phenomenon, EPA requested that the data be analyzed using the binding at the lowest concentration of estradiol, rather than the binding in the solvent control tubes, as the value for 100% binding of the radiolabeled estradiol (i.e., zero displacement by unlabelled estradiol). (The lowest concentration of estradiol (10^{-11} M) had previously been determined to be sufficient to establish the top plateau for estradiol.) The results of this analysis are what was presented in this Integrated Summary Report.

Use of the binding values at the lowest concentration of estradiol to establish 100% binding of the radioligand may have contributed to the variability of results both within laboratories and across laboratories in this study inasmuch as there were only a maximum of three replicates per run to establish the 100% binding value as opposed to the six solvent control tubes. If, as happened in a few cases, one or more of the replicates was unusable, the 100% binding value was likely to have a larger standard error of the mean even though the standard deviations may not have been unusually high.

As explained in the previous section, laboratories had significant difficulty getting the weak positive control chemical, norethynodrel, to stay in solution at the highest concentration (1 mM), and this data point is frequently missing from a run. Loss of this data point can significantly affect the estimate of the bottom plateau and can cause a run to miss the performance criterion for the bottom even though the remaining data points are consistent with a good run. Solubility problems with norethynodrel had not been encountered before in any of the preliminary studies. EPA is currently performing solubility studies to determine whether the solubility of norethynodrel is likely to pose a significant problem in the future. It is also investigating other weakly positive estrogen receptor binders as possible substitutes. Alternatively, it may specify a lower concentration to use as the maximum for this weak positive control, and establish modified performance criteria using that concentration.

Although the influence of the number of chemicals per run on intralaboratory variability of the results was not studied, EPA notes that most laboratories ran 3 or 4 chemicals per run. Each test chemical requires 24 tubes (8 concentrations x 3 replicates) in addition to the 87 tubes for controls, so a run of 3 chemicals consists of

159 tubes. (In this study, such a run consisted of 132 tubes since 27 R1881 tubes were omitted.) Processing this large number of tubes may have increased variability due to such factors as increased duration of exposure to room temperature (and subsequent denaturation of the receptor), and diminished ability to monitor partial pellet loss after centrifugation. This potential source of variability is expected to be less of a factor for laboratories in the EDSP if only one chemical is being tested at a time.

Finally, it should be remembered that while intralaboratory variability was disappointingly high for at least one laboratory in this study, such variability is not expected to be as much of a problem for laboratories that demonstrate the ability to meet the required performance criteria. The limited time available to run this large study on 23 chemicals apparently did not allow development of the proficiency necessary to obtain precise runs in all laboratories. The fact that results were almost all in accord with expectations when screening for interaction despite the variability in quantitative values shows that the assay is robust for this use.

V. Additional considerations

An ER transcriptional activation assay has recently been validated by the OECD for use in screening chemicals. There is overlap between receptor binding assays and transcriptional activation assays inasmuch as binding is the first step in transcriptional activation. Both assays are relatively simple and inexpensive *in vitro* assays, and the EPA is likely to require use of both, at least during the initial stages of the EDSP. The transcriptional activation assay is specific to the α isoform of the estrogen receptor while rat uterus contains both the α and β isoforms. Thus the ER-RUC binding assay may respond to substances that are specific to the β isoform while the transcriptional activation assay cannot respond to such substances.

Several changes have been made to the protocol since the second interlaboratory validation study. Small revisions were necessary to clarify wording that laboratories had found confusing, such as the upper limit for ligand depletion. The three significant changes are: 1) substitution of octyltriethoxysilane for R1881 as the negative control (discussed above); 2) introduction of an optional solubility testing step before running the assay; and 3) adjusting the test chemical dilution scheme so that dilutions are made in solvent rather than buffer. This last change was made in order to keep the

concentration of solvent constant (2%) across all test chemical concentrations. This change is not expected to affect the results of the assay given the limit on solvent concentration included in the protocol and the results of the solvent concentration study (Eldridge 2007) discussed above. The dilution scheme is now similar to the scheme used in the first interlaboratory study.

EPA tested additional chemicals at the weak end of the expected affinity spectrum in a later portion of this study. Those results are not being released at this time in order not to compromise the identity of the substances in the on-going, parallel hrER validation study. However, the pressure of processing 20 additional substances in addition to the 23 chemical reported here, within a defined time period, may have led to greater variability in results than will typically be seen when attention can be focused on one chemical in the Endocrine Disruptor Screening Program.

VI. Summary

The EPA agrees with the 1998 conclusion of the EDSTAC that the estrogen receptor binding assay using rat uterine cytosol is a validated assay for simple screening for interaction with the estrogen receptor in the context of a battery of assays. As recommended by the EDSTAC and ICCVAM, EPA has optimized and standardized the most important parameters of this assay and has shown that the resulting protocol is transferable to other laboratories. The variability of results may not support use for quantitative structure-activity relationship model development at this time, but if further work were to be undertaken to validate for such a use in the future, it may be that only the performance criteria rather than the protocol itself need to be adjusted.

A. Strengths

As an *in vitro* assay, the ER-RUC assay provides direct contact between chemical and the estrogen receptor without modification through absorption, distribution, metabolism, and excretion (ADME) considerations. The assay therefore has the potential of being more sensitive than the *in vivo* assays in the EDSP Tier 1 Battery, which usually involve ADME.

The assay provides consistent responses at the simple screening level, across laboratories, and these responses are in line with expectations for those chemicals tested whose estrogen receptor binding behavior is well-established.

The criteria used for classifying a chemical as interactive or not err on the side of classifying a chemical as interactive, which is appropriate for a screening assay. Since additional assays will be used in Tier 1 screening before a final determination of the potential for interaction is assigned, and additional Tier 2 assays will be performed to confirm the interaction and provide dose-response information before risk is assessed, the bias towards false positives is appropriate.

Despite the bias towards false positives, chemicals that truly do not interact in any way with the estrogen receptor (whether by one-site competitive binding or any other mechanism) consistently test negative in this assay.

The assay is short and inexpensive compared to the *in vivo* assays in the Tier 1 Battery.

Rat uterus contains both α and β isoforms of the estrogen receptor (Kuiper et al. 1997) and unlike current assays using recombinant receptor, which are specific to the alpha isoform, may therefore respond to substances which are specific to either isoform.

B. Weaknesses

The assay is sensitive to many details of preparation and technique and can show wide variability if not performed exactly as stated in the protocol. It is, for example, subject to problems if the receptor concentration in the cytosol is too low or too high, or the tubes are not kept cold at all times during preparation, incubation, and separation of bound from free tracer. However, the data suggest that a lab that meets the performance criteria for the standard and weak positive is likely to generate data that is much less variable than laboratories that do not meet the performance criteria. Thus the data from the EDSP, which requires adherence to the performance criteria, are likely to be of less variability than the data obtained in the second interlaboratory study.

Because of the sensitivity of this assay to technique, it is not consistently possible to characterize the probable mechanism of action of an “interactor” as one-site competitive binding even when the substance is known to be a one-site competitive

inhibitor. Nevertheless, since the purpose for which this assay will be used in the EDSP is only to identify interactors, not the mechanism by which they interact or to develop quantitative descriptors of the interaction such as the log(IC_{50}), this weakness is not acute.

Insolubility of test chemicals can be a significant problem when trying to identify weak interactors. When a high concentration of test chemical in solvent cannot be obtained, it may not be possible to test adequately the ability of the chemical to interact with the estrogen receptor. The choice of three possible solvents (DMSO, ethanol, or water) should help mitigate this potential weakness.

The lack of metabolic activity can cause chemicals which require metabolic activation to test negative. Results from *in vivo* assays in the Tier 1 Battery may not be consistent with results from this assay. It will be important not to tally simple "positives" and "negatives" in the Tier 1 Battery of assays but to evaluate the entire dataset when judging the weight of the evidence for interaction with the estrogen system.

The analysis of datasets might be more complicated than necessary. When the analysis was developed, the expectation was that standardization of the assay would allow precise and replicable quantitative analysis of log(IC_{50})s and Relative Binding Affinities. The expectation was also that precise, standardized methods of analysis would contribute to reproducibility and therefore use in other applications such as structure-activity relationship models. However, in the face of the variability encountered, EPA is assessing whether such analysis could be replaced with a simpler analysis and still meet the needs of the Screening Program.

Finally, the assay requires the use of animals. Although it is an *in vitro* assay, the receptor is obtained from uteri. EPA is cooperating with an OECD effort to validate a binding assay that uses human recombinant estrogen receptor rather than receptor from animals.

C. Conclusion

EPA believes that the standardized estrogen receptor binding assay using rat uterine cytosol as source of receptor has proven to be transferable, sensitive to chemicals known to interact with the estrogen receptor, specific to chemicals which interact, and reproducible in contract laboratories in terms of classifying chemicals as

interacting with the estrogen receptor or not. The assay is appropriate for use in its standardized form in a screening program to identify interaction with the endocrine system even though it may not be appropriate for other uses such as development of quantitative structure-activity relationship models.

VII. References

Carter, C.M.S., Leighton-Davies, J.R., Charlton, S.J. (2007). "Miniaturized receptor binding assays: complications arising from ligand depletion." J Biomol Screen 12(2): 255-66.

EDSTAC (August 1998). Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) final report. U.S. Environmental Protection Agency, Washington, DC. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/finalrpt.htm

Eldridge, J.C. (2007). Final report: Development of a standardized approach for evaluating environmental chemicals with low solubility in the estrogen receptor (ER) binding assay. USEPA Contract number EP-D-05-076. (Attached to this ISR)

FFDCA. (1996). Federal Food, Drug, and Cosmetic Act, as amended by the Food Quality Protection Act of 1996 (Pub. L. 104-170), 21 U.S.C. 346a(p). Section 408(p).

Hartung, J. and Makambi, K.H. (2001). Simple non-iterative t-distribution based tests for meta-analysis South African Statistical Journal 35: 1-17.

Hill, A.V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J. Physiol. (Lond.) (1910) 40, iv-vii.

ICCVAM. ICCVAM Authorization Act of 2000. Interagency Coordinating Committee on the Validation of Alternative Methods. 42 U.S.C. 285. Public Law 106-545, December 19, 2000. http://iccvam.niehs.nih.gov/docs/about_docs/PL106545.pdf

ICCVAM (2002). Background Review Document - Current Status of Test Methods for Detecting Endocrine Disruptors: In Vitro Estrogen Receptor Binding Assays (NIH 03-4504). U.S. Department of Health and Human Services, National Toxicology Program, Interagency Coordinating Committee on the Validation of Alternative Methods. http://iccvam.niehs.nih.gov/docs/endo_docs/final1002/erbndbrd/ERBd034504.pdf

Kuiper, G., Carlsson, B., Grandien, K. et al. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. Endocrinology 138(3): 863-870.

Mauck, L.A., Day, R.N, Notides A.C. (1982). Molybdate interaction with the estrogen receptor: effects on estradiol binding and receptor activation. Biochemistry 21(8):1788-1793.

Motulsky, H. (2003). GraphPad PRISM User's Guide Version 4.0, GraphPad Software, Inc. http://www.graphpad.com/prism/Prism.htm

Motulsky, H.J. and Brown, R.E. (2006). Detecting outliers when fitting data with nonlinear regression˜ a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics, Vol 7, pp 123-142.

Motulsky, H. (2007). GraphPad PRISM Version 5 software, San Diego, CA. http://www.graphpad.com/prism/Prism.htm

Murayama A. and Fukai F. (1985). Effect of molybdate on the molecular organization of the estrogen receptor system of porcine uterus. J. Biochem. 98(4):1083-1090.

NIEHS (October 2002). Background Review Document: Current Status of Test Methods for Detecting Endocrine Disruptors: *In Vitro* Estrogen Receptor Binding Assays, National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences NIH Publication No. 03-4054. Available at: http://iccvam.niehs.nih.gov/docs/endo_docs/final1002/erbndbrd/ERBd034504.pdf

NIEHS (March 1, 1997). Validation and regulatory acceptance of toxicological test methods: a report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (NIH 97-3981). Available at http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf

OECD (August 1996). Final report of the OECD workshop on harmonization of validation and acceptance criteria for alternative toxicological test methods. Organisation for Economic Co-operation and Development. Stockholm, Sweden, January 22-24, 1996.

Ojasoo T., and Raynaud, J.-P. 1978. Unique steroid congeners for receptor studies. Cancer Research 38:4186-4198.

Pettersson, K.S., Vanharanta, R.M., Soderholm, J.R. (1985). Pitfalls in the dextran-coated charcoal assay of estrogen receptors in breast cancer tissue. J. Steroid Biochem. 22(1):39-45.

SAS, Version 9.1.3 (2003). SAS Institute, Inc., Cary, NC.

Swillens, S. (1995). Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis. Molecular Pharmacology 47(6): 1197-1203.

USEPA. (December 28, 1998). Endocrine Disruptor Screening Program; Proposed statement of Policy, Federal Register Notice, vol. 63, issue 248, pp. 71542-71568. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/122898frnotice.pdf

USEPA (July 1999). Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. http://yosemite.epa.gov/sab/sabproduct.nsf/C8ABD410E357DBCF85257193004C42C4/$File/ec13.pdf

USEPA (July 2007). Validation of screening and testing assays proposed for the EDSP. Version 5.4. Available at: http://www.epa.gov/scipoly/oscpendo/pubs/edsp_validation_paper_v%205.4.pdf

USEPA (June 2008). Meeting minutes of the FIFRA Scientific Advisory Panel Meeting held on March 25-26, 2008 to review and consider the Endocrine Disrupter Screening Program (EDSP) Proposed Tier 1 Screening Battery. http://www.epa.gov/scipoly/sap/meetings/2008/march/minutes2008-03-25.pdf

Zacharewski, T.R., Meek, M.D., Clemons, J. H., Wu, Z. F., Fielden, M. R., Matthews, J. B. (1998). Examination of the *in vitro* and *in vivo* estrogenic activities of eight commercial phthalate esters. Toxicological Sciences 46(2): 282-293.

Appendix 1. Protocol (as revised following the second interlaboratory validation study)

Appendix 2. ICCVAM Background Review Document on the Estrogen Receptor Binding Assay. Executive Summary and Conclusions.

Appendix 3. Eldridge CJ. 2007. Final report: Development of a standardized approach for evaluating environmental chemicals with low solubility in the estrogen receptor (ER) binding assay.

Appendix 4. Report on statistical methods for evaluating variability in and setting up performance criteria for receptor binding assays

Appendix 5. Overall report on second interlaboratory validation study

Appendix 6. Final report from second interlaboratory validation study: Lab X

Appendix 7. Final report from second interlaboratory validation study: Lab Y

Appendix 8. Final report from second interlaboratory validation study: Lab Z

Appendix 9. Detailed statistical report

Appendix 10. Curve fits after normalization: Lab X

Appendix 11. Curve fits after normalization: Lab Y

Appendix 12. Curve fits after normalization: Lab Z

Appendix 13. Graphs of acceptable runs for reference standard (estradiol), weak positive (norethynodrel), and test chemicals, by laboratory