Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

24-Aug-2015

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

PERFORMANCE STANDARDS FOR THE HUMAN RECOMBINANT ESTROGEN RECEPTOR BINDING ASSAY

(INTENDED FOR THE DEVELOPERS OF NEW OR MODIFIED SIMILAR TEST METHODS)

Series on Testing & Assessment
No. 222

JT03380965

ENV/JM/MONO(2015)29
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 222

PERFORMANCE STANDARDS FOR THE HUMAN RECOMBINANT ESTROGEN RECEPTOR BINDING ASSAY

(*Intended for the developers of new or modified similar test methods*)

IOMC INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2015

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document contains the Performance Standards (PS) for the human recombinant estrogen receptor (hrER) binding assay. These PS accompany the Performance-Based Test Guideline (PBTG) for human recombinant estrogen receptor *in vitro* assays to detect chemicals with ER binding affinity (TG 493). The PS are intended for the developers of new or modified test methods, similar to the validated reference methods.

The PS were reviewed and discussed by the Validation Management Group for Non Animal testing (VMG NA), during a meeting that was held on 2-4 December 2014. Comments from the Working Group of National Coordinators of the Test Guidelines Programme (WNT) were requested on the draft PBTG and on the draft PS in December 2014. These comments were addressed and the PS slightly revised at a teleconference of the VMG NA in February 2015.

The present document was approved by the WNT in April 2015, declassified and published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology on 10 July 2015.

INTRODUCTION

1. The following Performance Standards (PS) accompany the Performance Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) Binding In Vitro Assays (TG 493) (1). This document is intended as a guide to developers of new test methods that are analogous to existing, fully validated test methods in that they are based on similar scientific principles and predict the same effect (colloquially referred to as "me too" tests). Prior to the acceptance of a new test method for regulatory testing applications, validation studies are conducted using scientifically sound principles to establish its reliability (i.e., the extent of intra- and inter-laboratory reproducibility over time when performed using the standardized protocol), and its relevance (i.e., the ability of the test method to correctly predict or measure the biological effect of interest) (2) (3) (4) (5). The purpose of the PS is to communicate the basis by which new proprietary (i.e. copyrighted, trademarked, registered) or non-proprietary test methods can be determined to have sufficient accuracy (i.e., agreement between a test method result and an accepted reference value) and reliability (i.e., extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol) for a specific testing purpose. Thus, this provides an avenue to demonstrate that a newly developed test method based on similar scientific principles has comparable or better performance capabilities than those from which the existing PS were derived, and may allow a more timely use of the new test method. New test methods ("me too" tests) can be added to TG 493 after OECD review and agreement that performance standards are met. A new test method developed under this PS will be covered by TG 493 which will be updated to add the new test method.

2. Performance standards are based on an adequately validated test method(s) and provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar (2) (3). The three elements of performance standards are:

   - Essential test method components: These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed test method that is considered to be mechanistically and functionally similar to the fully validated method. Essential test method components include unique characteristics of the test method, critical procedural details, and quality control measures.

   - A list of reference substances: Reference substances are used to assess the accuracy and reliability of a proposed mechanistically and functionally similar test method. These substances are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method, and are the minimum number that should be used to evaluate the performance of a proposed mechanistically and functionally similar test method.

   - Accuracy and reliability performance values: These are the standards for accuracy (i.e., sensitivity, specificity, false positive/negative rates) and reliability (i.e., degree to which the test method can be performed reproducibly within and among laboratories over time) that the proposed test method should meet or exceed when evaluated using the minimum list of the reference substances.

3. The fully validated reference test methods that provide the basis for this PS are:

    - The Freyberger-Wilson (FW) In vitro Estrogen Receptor alpha (ERα) Binding Assay using a Full Length Human Recombinant ERα (6)

    - The Chemical Evaluation and Research Institute (CERI) In Vitro Estrogen Receptor (ER) Binding Assay Using a Human Recombinant ERα Ligand Binding Domain Protein (6)

ESSENTIAL TEST METHOD COMPONENTS AND OTHER VALIDATION CONSIDERATIONS

4. *In vitro* estrogen receptor (ER) binding assays are based on a direct interaction of the chemical with a specific receptor ligand binding site that regulates the gene transcription. Binding of the natural ligand, 17β-estradiol, is the initial step of a series of molecular events that activate the transcription of target genes and ultimately, culminates with a physiological change *in vivo* (7). Thus binding to the ERα is considered to be the molecular initiating event and one key mechanisms of ER mediated endocrine disruption (ED).

5. Certain principles are important in delineating the essential test method components that determine whether modified or new ER binding tests are functionally and mechanistically similar. *In vitro* ER binding assays are designed to identify substances that might have the ability to bind to human ERα

6. The following test method components may vary, so this PBTG does apply to test methods that may differ in:

    - ERα characteristics (full length or partial)
    - marker/tracer (with possible minor adjustments to the protocol)
    - method for production of the receptor

    These elements should be clearly described in each test method.

7. Essential test method components for *in vitro* ER binding protocols should include the following:

    - A saturation binding assay and a competitive binding assay;

    - A strong reference estrogen, preferably 17ß-estradiol should be used to demonstrate the adequacy of the method for detecting substances that are capable of binding to the ER;

    - 17ß-estradiol should be used as ligand for competitive binding;

    - A weak positive control with a potency (e.g., binding) three orders of magnitude lower than the reference estrogen should be included to provide another quality control measure by which to judge the acceptability of the method for detecting a weak binder, and by which to evaluate the reproducibility of the test method;

    - A negative control using a non-binder (e.g. octyltriethoxysilane) should be used. Absence of binding ability should have been demonstrated in the range of used concentrations;

    - A vehicle control (e.g. DMSO, ethanol, or water) that is miscible with the assay buffer and does not interfere with the test system should be used;

    - At least eight concentrations of the test chemical, spaced at decadic logarithmic (log10) intervals and tested up to the limit of solubility, but not to exceed 10^{-3}M.;

- All concentrations of the controls (e.g., solvent (vehicle), buffer, weak binder, negative(s), or the reference estrogen), and the test chemical should be tested in triplicate, and repeated in a minimum of three independent runs.

8. A commonly accepted approach for analysing data obtained from *in vitro* saturation and competition ER binding assays has been developed using the reference test methods with a well-defined method for classifying a positive and a negative response (i.e. binder vs. non-binder).

9. To ensure that a proposed *in vitro* ER recombinant binding test method possesses characteristics similar to other validated test methods, the reference substances for testing ER binding listed in Table 1 should be used to demonstrate the reliability and accuracy of the new test method. The 23 recommended reference substances, representing chemical classes commonly associated with ER activity, have been classified as ER binders or non-binders based upon published reports, including *in vitro* assays for ER binding and transactivation, and the *in vivo* uterotrophic assay (7) (8) (9) (10) (11) (12) (13). If a reference substance is no longer commercially available, a substance with the same classification and, comparable potency, mode of action and chemical class can be used. Additional chemicals not included in the reference substance list may be used to demonstrate an improvement (e.g. improved reproducibility and/or accuracy with regard to accepted reference data) of the new test method as compared with the fully validated test methods.

10. New similar test methods should not be developed solely on the basis of the 23 reference substances, but rather on a sufficiently large test development set. Reference substances should be preferentially used to determine equivalence of performance compared to the validated reference test methods.

11. As this assay is a cell free assay, metabolism of the test chemical within the assay is unlikely. However, the combination of this cell free assay with pre-metabolised chemicals for example could be of additional benefit. Such considerations are highly relevant when considering results of such in vitro test methods in the context of QSAR modeling approaches, as it may not be the compound under investigation that is actually responsible for the observed downstream biological response, but rather the metabolites formed.

12. All chemicals should be tested in a coded/blinded manner. When evaluated using these reference substances, the reliability and accuracy (i.e. sensitivity, specificity, false positive rates, and false negative rates) of the proposed ER binding test method should approximate the defined reliability and accuracy values as described below. If a chemical is not available, a suitable chemical with a similar binding affinity may be substituted.

Table 1: List of reference substances (23) for evaluation of ER binding assay performance and accuracy.

No.	Chemical Name	CAS RN	Expected Response #	Concentration Range Tested(M)	FW Assay Classification	CERI Assay Classification	MESH Chemical Class	Product Class
1	17β-Estradiol	50-28-2	*Strong Binder*	$1x10^{-11}$ – $1x10^{-6}$	Binder	Binder	Steroid	Pharmaceutical, Veterinary Agent
2	17α-ethynyl estradiol	57-63-6	*Strong Binder*	$1x10^{-11}$ – $1x10^{-6}$	Binder	Binder	Steroid	Pharmaceutical, Veterinary Agent
3	Diethylstilbestrol (DES)	56-53-1	*Strong Binder*	$1x10^{-11}$ – $1x10^{-6}$	Binder	Binder	Hydrocarbon, (Cyclic), Phenol	Pharmaceutical, Veterinary Agent
4	Meso-Hexestrol	84-16-2	*Strong Binder*	$1x10^{-11}$ – $1x10^{-6}$	Binder	Binder	Hydrocarbon (cyclic), Phenol	Pharmaceutical, Veterinary Agent
5	Zearalenone	17924-92-4	*Strong Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Hydrocarbon (heterocyclic), Lactone	Natural Product
6	Tamoxifen	10540-29-1	*Strong Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Hydrocarbon, (Cyclic)	Pharmaceutical, Veterinary Agent
7	Norethynodrel or (Norehindrone)[a]	68-23-5 (68-22-4)[a]	*Moderate Binder*	$3x10^{-9}$ – $30x10^{-4}$	Binder	Binder	Steroid	Pharmaceutical, Veterinary Agent
8	Genistein	446-72-0	*Moderate Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Hydrocarbon (heterocyclic), Flavonoid	Natural Product
9	Equol	531-95-3	*Moderate Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Phytoestrogen Metabolite	Natural Product
10	Butyl paraben (n butyl-4-	94-26-8	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Paraben	Preservative

No.	Chemical Name	CAS RN	Expected Response #	Concentration Range Tested(M)	FW Assay Classification	CERI Assay Classification	MESH Chemical Class	Product Class
	hydroxybenzoate)							
11	Nonylphenol (mixture)	84852-15-3	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Alkylphenol,	Intermediate Compound
12	o,p'-DDT[d]	789-02-6	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Organochlorine	Insecticide
13	5α-dihydrotestosterone	521-18-6	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Steroid, Nonphenolic	Natural Product
14	Bisphenol A (BPA)	80-05-7	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Phenol	Chemical Intermediate
15	4-n-heptylphenol	1987-50-4	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Equivocal	Binder	Alkylphenol	Intermediate
16	Kepone (Chlordecone)	143-50-0	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Hydrocarbon, (Halogenated)	Pesticide
17	Enterolactone	78473-71-9	*Weak Binder*	$1x10^{-10}$ – $1x10^{-3}$	Binder	Binder	Phytoestrogen	Natural Product
18	* Di-n-butyl phthalate (DBP)	84-74-2	*Non-binder*	$1x10^{-10}$ – $1x10^{-4}$	Non-Binder[*†]	Non-Binder[*†,]	Hydrocarbon (cyclic), Ester	Plasticizer, Chemical
19	Octyltriethoxysilane	2943-75-1	*Non-binder*	$1x10^{-10}$ – $1x10^{-3}$	Non-Binder	Non-Binder	Silane	Surface modifier
20	Corticosterone[C]	50-22-6	*Non-binder*	$1x10^{-10}$ – $1x10^{-3}$	Non-binder	Non-Binder	Steroid	Natural Product
21	Benz(a)anthracene	56-55-3	*Non-Binder[b]*	$1x10^{-10}$ – $1x10^{-3}$	Non-Binder	Non-Binder	Aromatic Hydrocarbon	Intermediate
22	Progesterone[C]	57-83-0	*Non-binder*	$1x10^{-10}$ – $1x10^{-4}$	Non-Binder	Non-Binder	Steroid	Natural Product
23	Atrazine[C]	1912-24-9	*Non-binder*	$1x10^{-10}$ – $1x10^{-4}$	Non-Binder	Non-Binder	Heterocyclic compound	Herbicide

Abbreviations: DES, diethylstilbestrol; HPTE, 2,2-bis(p-hydroxyphenyl)-1,1,1-trichloroethane, an intermediate by-product of methoychlor; o,p'-DDT, 1,1,1-trichloro-2-2-[o-chlorophpenyl]-2-[p-chlorophenyl]ethane

#The expected response for each chemical was based upon published data from *in vitro* studies and were reviewed by a Chemical Advisory Board whose members were not directly associated with the validation study for the FW and CERI hrER Binding Assays. Chemicals were selected to represent multiple chemical classes and cover a range of

binding affinity potencies commonly associated with ER agonist activity (6). When tested in the FW or CERI hrER Assays, the $LogIC_{50}$ for strong binders typically ranged from -9 to -7 (M), moderate -7.1 to -6.0 (M) and weak < -5.9 (M) (see Reference 6, Table 40).

[*]The use and classification of Di-n-butyl phthalate (DBP) as a non-binder was based on testing up to 10^{-4} M because the chemical was observed to be insoluble at 10-3M (e.g. turbidity) in some laboratories during the pre-validation studies. When DBP was tested up to 10^{-3}M as a coded chemical, it was classified as 'equivocal' due to displacement of (^{3}H)17β-estradiol at highest in 3/5 laboratories using the CERI assay (6) and 5/6 laboratories using the FW assay (6).

[a]Norethindrone is provided as an alternate for the control weak binder for cases when norethynodrel is unavailable.

[b] During the validation study, benz(a)anthracene was reclassified as a non-binder (i.e., negative) based on published literature demonstrating that the *in vitro* estrogenic activity reported for this chemical (14) is primarily dependent upon its metabolic activation (15) (16). Enzymatic metabolic activation of the chemical would not be anticipated in the cell free hrER assays as used in this inter-validation study. Thus, the correct classification for this chemical is a 'non-binder' when used under the experimental conditions for the FW and CERI assays.

[c]Chemicals were observed to be insoluble at 10-3M (e.g.,turbidity) in some laboratories during the validation study.

[d]: Optional where o,p'-DDT is prohibited by regulatory authorities when replaced by chemical with comparable binding affinity.

DEFINED RELIABILITY AND ACCURACY VALUES

Within-laboratory reproducibility

13. For the assessment of within-laboratory reproducibility, the concordance of classifications (ER binders and non-binders) obtained in three independent consecutive test runs should be 100% for each laboratory when using a subset of 8 chemicals selected from Table 1 which includes at least 1 chemical from those designated as strong binders, 1 chemical from moderate, 3 chemicals from weak, and 3 non-binders. Three independent consecutive runs are required to fulfil the criteria for acceptance. If, for example, runs 2 and 3 are inconsistent with run 1, one additional run (run 4) will be sufficient to show within-lab reproducibility if run 4 is consistent with runs 2 and 3. If run 4 is consistent with run 1 instead, then at least two additional consecutive runs (runs 5 and 6) showing consistency with run 4 will be required to fulfil the requirement for three consecutive independent runs that have at least 100% concordance of classifications.

Between-laboratory reproducibility

14. To assess between-laboratory reproducibility, the remaining 15 reference substances should be tested at least once in two or (preferably) three laboratories. All the data available on the 23 chemicals (8 tested three times; the other 15 tested once in each laboratory) should be utilised. Concordance of classifications for the ER binders (strong and weak affinity) and non-binders between laboratories should be used as a measure to describe between-laboratory reproducibility. To be considered acceptable, a test method should show concordance of 85 % or greater.

Predictive capacity (accuracy)

15. The accuracy (sensitivity, specificity, and overall accuracy) of the proposed test method should be comparable to that demonstrated for the fully validated test methods, i.e. the FW in vitro ERα binding assay using a full length hrERα and the CERI in vitro ER binding assay using a hrERα ligand binding domain protein (6). On the basis of the performance values (sensitivity / specificity) of the validated reference methods for substances in the validation test set, as well as other empirical data from these methods (see Annex 2), the target values for sensitivity, specificity, and overall accuracy to be obtained when testing the reference chemicals (Table 1) are set to be greater or equal to 93%.

16. Although it is not realistic to expect test methods to perform identically, discordant results should be discussed in terms of the ability of the test method to detect a similar range of potencies and chemical/product classes, as demonstrated by the fully validated test methods (6).

LITERATURE

1. OECD (2015), Test Guideline 493: Performance-Based Test Guideline for Human recombinant estrogen receptor (ER) in vitro assays to detect chemicals with ER binding affinity, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Paris.

2. OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, Environment, Health and Safety Publications, Series on Testing and Assessment No.34. OECD, Paris.

3. ICCVAM (1997), Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), NIH Publication No. 97-3981, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

4. ICCVAM (2003), ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test methods, NIH Publication No. 03-4508, Research Triangle Park, NC, National Institute of Environmental Health Sciences.

5. OECD (1996), Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, ENV/MC/CHEM/TG(96)9, OECD, Paris.

6. OECD (2015), Integrated Summary Report Validation of Two Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrERα). Environment, Health and Safety Publications, Series on Testing and Assessment No. 226, OECD, Paris.

7. ICCVAM (2002), Background Review Document. Current Status of Test Methods for Detecting Endocrine Disruptors: *In Vitro* Estrogen Receptor Binding Assays. (NIH Publication No. 03-4504). http://ntp.niehs.nih.gov/go/40428: National Institute of Environmental Health Sciences, Research Triangle Park, NC.

8. ICCVAM (2003), ICCVAM Evaluation of *in vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays. http://ntp.niehs.nih.gov/go/40106. http://ntp.niehs.nih.gov/go/40106.

9. ICCVAM (2006), ICCVAM Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays http://ntp.niehs.nih.gov/go/40428.

10. Akahori Y. et al. (2008), *Relationship between the results of in vitro receptor binding assay to human estrogen receptor alpha and in vivo uterotrophic assay: comparative study with 65 selected chemicals*, Toxicol. In Vitro, 22(1): 225-231.

11. OECD (2007), *Additional Data Supporting the Test Guideline on the Uterotrophic Bioassay in Rodents*, Series on Testing and Assessment No. 67, OECD, Paris.

12. Takeyoshi, M. (2006), *Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity - The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line*, Chemicals Evaluation and Research Institute (CERI): Japan. p. 1-188.

13. Yamasaki, K; Noda, S; Imatanaka, N; Yakabe, Y. (2004), *Comparative study of the uterotrophic potency of 14 chemicals in a uterotrophic assay and their receptor-binding affinity*, Tox Letters 146: 111-120.

14. Kummer, V; Maskova, J; Zraly, Z; Neca, J; Simeckova, P; Vondracek, J; Machala, M. (2008), Estrogenic activity of environmental polycyclic aromatic hydrocarbons in uterus of immature Wistar rats. Toxicology Letters 180: 213-221.

15. Gozgit, JM; Nestor, KM; Fasco, MJ; Pentecost, BT; Arcaro, KF. (2004), Differential action of polycyclic aromatic hydrocarbons on endogenous estrogen-responsive genes and on a transfected estrogen-responsive reporter in MCF-7 cells. Toxicology and Applied Pharmacology 196: 58-67.

16. Santodonato, J. (1997), Review of the estrogenic and antiestrogenic activity of polycyclic aromatic hydrocarbons: relationship to carcinogenicity. Chemosphere 34: 835-848.

ANNEX 1

Definitions and Abbreviations

Acceptability criteria: Minimum standards for the performance of experimental controls and reference standards. All acceptability criteria must be met for an experiment to be considered valid.

Accuracy: The closeness of agreement between a test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (1).

DMSO: Dimethyl sulfoxide

ER: Estrogen receptor

hERα: Human estrogen receptor alpha

Estrogenic activity: The capability of a chemical to mimic 17β-estradiol in its ability to bind to and activate estrogen receptors. hERα-mediated binding activity can be detected with this PBTG.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results.
Inter-laboratory reproducibility is determined during the pre-validation and validation processes, and indicates the extent to which a test method can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (2).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as Within-laboratory reproducibility (2)

PBTG: Performance-Based Test Guideline.

PC: Positive control; a strongly active substance, preferably 17ß-estradiol, which is included in all tests to help ensure proper functioning of the assay.

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (i) essential test method components; (ii) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals (2).

Proficiency chemicals (substances): A subset of the Reference Chemicals included in the Performance Standards that can be used by laboratories to demonstrate technical competence with a standardized test method. Selection criteria for these substances typically include that they represent the range of responses, are commercially available, and have high quality reference data available.

Proficiency: The demonstrated ability to properly conduct a test method prior to testing unknown substances.

Reference chemicals (substances): A set of chemicals to be used to demonstrate the ability of a new test method to meet the acceptability criteria demonstrated by the validated reference test method(s). These

chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative.

Reference estrogen (positive control, PC): 17ß-estradiol (E2, CASRN 50-28-2).

Reference test methods: The test methods upon which the PBTG is based.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (2).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (2).

Sensitivity: The proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (2).

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of attest method (2).

Substance: Used in the context of the UN GHS as chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Validation: The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.

VC: Vehicle (solvent) control, the solvent that is used to dissolve test and control chemicals is tested solely as vehicle without dissolved chemical.

Weak positive control: A weakly active substance selected from the reference substances list that is included in all tests to help ensure proper functioning of the assay.

ANNEX 2

Supplementary Information for

The Freyberger-Wilson (FW) *In vitro* Estrogen Receptor alpha (ERα) Binding Assay using a Full Length Human Recombinant ERα

and

The Chemical Evaluation and Research Institute (CERI) *In Vitro* Estrogen Receptor (ER) Binding Assay Using a Human Recombinant ERα Ligand Binding Domain Protein

Summary of the Reliability and Accuracy Values Obtained during the Validation Study: FW and CERI i*n vitro* hrER Binding Assays

1. The validation study report for the FW and the CERI i*n vitro* hrER Binding Assays provides comprehensive descriptions of the data used to develop the reliability and accuracy values for each of these assays (6). Additional information is provided in this document to facilitate the review of data used to develop the performance standards for the FW and CERI hrER Binding Assays. The following is a summary of the estimates of intra- and inter-laboratory reproducibility and predictive capacity (accuracy) for each of these fully validated test methods.

2. Three control substances were used concurrently during each run of the assay. In addition, the controls were also tested as coded test chemicals in at least 3 runs. The results of the controls when tested as a test chemical (TC) were included in the analyses below, as well as the results of the controls when used as controls. For the controls or the test chemicals, when more than 3 runs were available; the first 3 sequential runs for a given chemical were used for the analysis of reproducibility.

I. Intra-laboratory (within-laboratory) reproducibility:

The closeness of agreement between test results obtained within a single laboratory when the procedure is performed using the same substance under identical conditions within a given time period.

3. The intra-laboratory reproducibility was assessed using 3 controls (a strong binder, a weak binder and a non-binder), 15 substances that covered a range of affinities for the ER (8 uncoded and 7 coded) and 5 substances that were non-binders (1 uncoded and 4 coded). The reference estrogen and weak binder (17β-estradiol and norethynodrel, respectively) as well as the non-binder (DBP) controls were included with each run conducted for the other test substances, both uncoded and coded. Three technical replicate were requested for each of the 8 concentrations of test substance within a run, and three independent runs were to be conducted for each test substance, on a different date.

a) **CERI *in vitro* hrER Binding Assay validation study**: Each substance was tested in 3 separate and acceptable experiments by each of 4 laboratories and classified as an ER binder, non-binder, or equivocal. Agreement on substance classifications within each laboratory ranged from 81 to 96% (Table PS-5 – Annex N of reference 6).

Table 1: Intra-laboratory reproducibility for the CERI hrER Binding Assay

	Ceetox	Bayer	Japan CERI	U. Missouri
Agreement Within Laboratory	24/27[a,b] (89%)	24/25 (96%)	23/24 (96%)	22/27 (81%)
All runs positive	18/20	19/19	16/17	16/20
All runs negative	6/7	5/6	7/7	6/7
Discordance Within Laboratory	3/27 (11%)	1/25 (4%)	1/24 (4%)	5/20 (19%)
Not all runs positive	2/20	0/19	1/17	4/20
Not all runs negative	1/7	1/6	0/7	1/7

[a] Ratio of the number of chemicals with concordance among all 3 runs as compared with total number of chemicals tested within each laboratory.

[b]Total number of chemicals represent (1) those chemicals for which 3 runs submitted by each laboratory, (2) include the reference estrogen and control weak binder and non-binders tested under both uncoded and coded assay conditions in all laboratories (i.e. n=26 maximum), and (3) the alternate control weak binder, norethindrone, that was tested only in the Ceetox and U. Missouri laboratories (i.e. n=27 maximum chemicals).

4. Table 1 was developed using the results presented in Table PS-5 (Annex N of reference (6)), derived from Annex B of Appendix H of the ISR (Table B-5), after re-analysis of the data that included additional runs that had been excluded in the initial analysis of data from the validation study (6). Table 1 includes the results for test chemicals that were submitted from 4 laboratories that were (1) acceptable by the submitting laboratory, (2) contained data for at least 6 concentrations of the test chemical, and (3) were tested in 3 independent runs. In addition, the results reported in Table 1 take into account any changes in classification of the run after the 10% rule was applied (see PBTG, Annex 4), when it was judged necessary to be applied.

5. When tested as a coded test chemical, DBP-TC was classified as a positive in all 3 runs conducted in the Freyberger (Bayer) laboratory and was thus excluded from the intra-laboratory analysis. Heptylphenol and elactone were also excluded from this analysis for the JapanCERI laboratory as both were classified as equivocal in all 3 runs. These chemicals were considered the more challenging among the whole set of chemicals used in the validation study.

b) **FW hrER Binding Assay:** Each substance was tested in 3 separate and acceptable experiments by each of 4 laboratories and classified as binder, non-binder, or equivocal. Agreement on substance classifications within each laboratory ranged from 85 to 96% (Table PS-6 - Annex N of reference 6).

Table 2: Intra-laboratory reproducibility for the FW hrER Binding Assays

	Ceetox	Bayer	CERI	U. Missouri
Agreement Within Laboratory	22/26[a,b] (85%)	22/26 (85%)	23/25 (92%)	24/25 (96%)
All runs positive	18/19	16/19	18/19	19/19
All runs negative	4/7	6/7	5/6	5/6
Discordance Within Laboratory	4/26 (15%)	4/26 (15%)	2/25 (8%)	1/25 (4%)
Not all runs positive	1/19	3/19	1/19	0/19
Not all runs negative	3/7	1/7	1/6	1/6

[a] Ratio of the number of chemicals with concordance among all 3 runs as compared with total number of chemicals tested within each laboratory.

[b] Total number of chemicals represent (1) those chemicals for which 3 runs submitted by each laboratory, (2) include the reference estrogen and control weak binder and non-binders tested under both uncoded and coded assay conditions in all laboratories (e.g., n=26 maximum), and (3) the alternate control weak binder, norethindrone, that was tested only in the Ceetox and U. Missouri laboratories (e.g, n=27 maximum chemicals).

6. Table 2 was developed according to the results presented in Table PS-6 (Annex N of reference (6)), derived from Annex B of appendix H of the ISR (tables B-6), after re-analysis of the data that were excluded in the initial analysis (6). Table 2 includes the results for test chemicals that were submitted from 4 laboratories that were (1) acceptable by the submitting laboratory, (2) contained data for at least 6 concentrations of the test chemical, and (3) were tested in 3 independent runs. In addition Table 2 takes into account any change in classification of the run after the 10% rule was applied (See PBTG, Annex 4), when it was judged necessary to be applied.

II. Inter-laboratory (between-laboratory) reproducibility:

A measure of the extent to which different qualified laboratories using the same protocol and testing the same substances can qualitatively produce similar results. Inter-laboratory reproducibility is determined during the validation process, and indicates the extent to which a test method can be transferred successfully among laboratories.

7. Inter-laboratory reproducibility was assessed using 23 chemicals (17 positives, 6 negatives), 3 controls (coded and uncoded) and Norethindrone tested in only 2 laboratories. Multiple runs conducted within a laboratory for a chemical were combined by assigning numeric values to each run and averaging across the runs as shown in Table 3. Results for the combined runs within each laboratory are compared with the expected classification for each test chemical.

Table 3. Method for classification of test chemical using multiple runs within a laboratory[a].

To assign value to each run:	
Classification	**Numeric Value**
Binder	2
Equivocal	1
Non-binder	0
To classify average of numeric value across runs:	
Classification	**Numeric Value**
Binder	Average ≥ 1.5
Equivocal	0.5 Average < 1.5
Non-binder	Average < 0.5

[a] Additional information on methods for classification of test chemicals is provided in references (1) (6).

a) **CERI hrER Binding Assay validation study**: There was 85% (23/27) agreement on the classifications for these chemicals among the laboratories (Table 4). This analysis has been performed based on the results from four laboratories: (i.e. Ceetox, Freyberger (Bayer), JapanCERI and U. Missouri) as shown for each individual run of the assay in Table 6.

Table 4: Inter-laboratory reproducibility for the CERI hrER Binding Assay

Results Among Laboratories	Percent Agreement [a]
Agreement Among Laboratories[b]	**23/27 (85%)**
++++ / ++(Nethindrone)	17/27 (63%)
- - - -	6/27 (22%)
Discordance Among Laboratories[b]	**4/27 (15%)**
+++E	1/27 (4%)
++EE	2/27 (7%)
+- - E	1/27 (4%)

[a] Ratio of the number of chemicals with concordance among all laboratories as compared with total number of chemicals tested.

[b] Key to Symbols: ++++ (binder, all 4 laboratories), ---- (non-binder, all 4 laboratories), E (equivocal)

b) **FW hrER Binding Assay validation study**: There was 85% (23/27) agreement on the classifications for these chemicals among the laboratories (Table 5). A comparison of classifications for each chemical among the laboratories is shown in Table 7. This analysis has been performed based on the results from the four following laboratories: Ceetox, Freyberger (Bayer), JapanCERI and U. Missouri.

Table 5: Inter-laboratory reproducibility for the FW hrER Binding Assay

Results Among Laboratories	Percent Agreement [a]
Agreement Among Laboratories[b]	**23/27 (85%)**
++++ / ++(Nethindrone)	17/27 (63%)
- - - -	6/27 (12%)
Discordance Among Laboratories[b]	**3/27 (13%)**
- - EE	2/27 (7%)
+++ E	1/27 (4%)
++E -	1/27 (4%)

[a] Ratio of the number of chemicals with concordance among all laboratories as compared with total number of chemicals tested.
[b] Key to Symbols: ++++ (binder, all 4 laboratories), ---- (non-binder, all 4 laboratories), E (equivocal), --EE (non-binder, 2 laboratories; equivocal, 2 laboratories)

III. Predictive Capacity (accuracy):

A measure of performance (i.e., sensitivity, specificity, positive and negative predictivity), and overall accuracy provide a quantitative assessment of the closeness of agreement (e.g., the proportion of correct outcomes) between test methods results and the values obtained from reference substances.

8. The results from all the participating laboratories were combined to classify a test chemical among the laboratories. Values were averaged across laboratories to obtain a single result for each chemical. Each chemical within a laboratory was assigned a score based on the classification determined above in Table 3. Then the scores for a chemical were averaged across laboratories, and the final classification was set using the ranges for the numeric value as shown in the bottom panel of Table 3. Classifications for each of the reference substances are shown in Table 6 (CERI Assay) and Table 7 (FW Assay) after testing in each of the laboratories along with a final classification using the combined results from all of the laboratories.

Table 6: CERI Assay, Chemical Classification by Individual Laboratories and Overall Simple Average Using Results from all Laboratories.

Chemical	CAS #	Expected binding affinity[a]	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification (Average of Laboratories)
				CeeTox	Freyberger	JapanCERI	Missouri	
Subtask 1 (Controls)								
17β-Estradiol (E2)	50-28-2	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Norethynodrel (NE)	68-23-5	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Di-n-butyl phthalate[b] (DBP)	84-74-2	Negative	Non-binder[b]	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
17β-Estradiol - TC	50-28-2	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Norethynodrel - TC	68-23-5	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Di-n-butyl phthalate - TC[b]	84-74-2	Negative	Non-binder[b]	Equivocal	Binder	Non-binder	Non-binder	Equivocal
Subtask 2 (uncoded chemicals)								
Diethylstilbestrol (DES)	56-53-1	Very strong	Binder	Binder	Binder	Binder	Binder	Binder
17α-ethynyl estradiol (EE2)	57-63-6	Very strong	Binder	Binder	Binder	Binder	Binder	Binder
Meso-Hexestrol (Hexestrol)	84-16-2	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Genistein (GEN)	446-72-0	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Equol	531-95-3	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Butyl paraben (ButylPar)	94-26-8	Weak	Binder	Binder	Binder	Binder	Binder	Binder

Chemical	CAS #	Expected binding	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification
Nonylphenol (mixture) (NE)	84852-15-3	Weak	Binder	Binder	Binder	Binder	Binder	Binder
o,p'-DDT[a] (DDT)	789-02-6	Weak	Binder	Binder	Binder	Binder	Binder	Binder
Corticosterone (CORT)	50-22-6	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Subtask 3 (coded chemicals)								
Zearalenone (Zear)	17924-92-4	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Tamoxifen (Tamox)	10540-29-1	Strong	Binder	Binder	Binder	Binder	Binder	Binder
5α-dihydrotestosteroneb (DHT)	521-18-6	Weak	Binder	Binder	Binder	Binder	Equivocal	Binder
Bisphenol A (BPA)	80-05-7	Weak	Binder	Binder	Binder	Binder	Binder	Binder
4-n-heptylphenol (heptylphenol)	1987-50-4	Weak	Binder	Binder	Binder	Equivocal	Equivocal	Binder
Kepone (Chlordecone)	143-50-0	Weak	Binder	Binder	Binder	Binder	Binder	Binder
Benz(a)anthracene (BaA)[c]	56-55-3	Weak	Non-binder[c]	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Enterolactone (ELactone)	78473-71-9	Weak	Binder	Binder	Binder	Equivocal	Equivocal	Binder
Progesterone	57-83-0	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Octyltriethoxysilane (OTES)	2943-75-1	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Atrazine (ATR)	1912-24-9	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Subtask 4 (optional, coded chemicals)								

Chemical	CAS #	Expected binding	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification
Norethindrone	68-22-4	Weak	Binder	Binder	-	-	Binder	Binder

[a] The expected response for each chemical was based upon published data from *in vitro* studies and were reviewed by a Chemical Advisory Board whose members were not directly associated with the validation study for the FW and CERI hrER Binding Assays. Chemicals were selected to represent multiple chemical classes and cover a range of binding affinity potencies commonly associated with ER agonist activity (6). When tested in the FW or CERI hrER Assays, the $LogIC_{50}$ for strong binders typically ranged from -9 to -7 (M), moderate -7.1 to -6.0 (M) and weak < -5.9 (M) (see Reference 6, Table 40).

[b]The use and classification of Di-n-butyl phthalate (DBP) as a non-binder was based on testing up to 10^{-4} M because the chemical was observed to be insoluble at 10-3M (e.g, turbidity) in some laboratories during the pre-validation studies. When DBP-TC was tested up to 10^{-3}M as a coded chemical, it was classified as 'equivocal' due to displacement of (^{3}H)17β-estradiol at highest in 3/5 laboratories using the CERI assay (6) and 5/6 laboratories using the FW assay (6).

[c]During the validation study, benz(a)anthracene was reclassified as a non-binder (i.e., negative) based on published literature demonstrating that the *in vitro* estrogenic activity reported for this chemical (14) is primarily dependent upon its metabolic activation (6). Enzymatic metabolic activation of the chemical would not be anticipated in the cell free hrER assays as used in this inter-validation study. Thus, the correct classification for this chemical is a 'non-binder' when used under the experimental conditions for the FW and CERI assays.

Yellow shading: Classification that is not concordant with that expected.

Table 7: FW Assay, Chemical Classification by Individual Laboratories and Overall Simple Average Using Results from all Laboratories.

Chemical	CAS #d	Expected binding affinity[a]	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification (Average of All Laboratories)
				CeeTox	Freyberger	Japan CERI	Missouri	
Subtask 1 (Controls)								
17β-Estradiol (E2)	50-28-2	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Norethynodrel (NE)	68-23-5	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Di-n-butyl phthalate[b] (DBP)	84-74-2	Negative	Non-binder[b]	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
17β-Estradiol -TC	50-28-2	Strong	Binder	Binder	Equivocal	Binder	Binder	Binder
Norethynodrel -TC	68-23-5	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Di-n-butyl phthalate -TC[b]	84-74-2	Negative	Non-binder[b]	Non-binder	Non-binder	Equivocal	Equivocal	Equivocal
Subtask 2 (uncoded chemicals)								
Diethylstilbestrol (DES)	56-53-1	Very strong	Binder	Binder	Binder	Binder	Binder	Binder
17□-ethynyl estradiol (EE2)	57-63-6	Very strong	Binder	Binder	Binder	Binder	Binder	Binder
Meso-Hexestrol (Hexestrol)	84-16-2	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Genistein (GEN)	446-72-0	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Equol	531-95-3	Moderate	Binder	Binder	Binder	Binder	Binder	Binder
Butyl paraben	94-26-8	Weak	Binder	Binder	Binder	Binder	Binder	Binder

Chemical	CAS #d	Expected binding	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification
(ButylPar)								
Nonylphenol (mixture) (NE)	84852-15-3	Weak	Binder	Binder	Binder	Binder	Binder	Binder
o,p'-DDTa (DDT)	789-02-6	Weak	Binder	Binder	Binder	Binder	Binder	Binder
Corticosterone (CORT)	50-22-6	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Subtask 3 (coded chemicals)								
Zearalenone (Zear)	17924-92-4	Strong	Binder	Binder	Binder	Binder	Binder	Binder
Tamoxifen (Tamox)	10540-29-1	Strong	Binder	Binder	Binder	Binder	Binder	Binder
5α-dihydrotestosterone (DHT)	521-18-6	Weak	Binder	Binder	Binder	Binder	Binder	Binder
Bisphenol A (BPA)	80-05-7	Weak	Binder	Binder	Binder	Binder	Binder	Binder
4-n-heptylphenol (heptylphenol)	1987-50-4	Weak	Binder	Equivocal	Binder	Binder	Non-binder	Equivocal
Kepone (Chlordecone)	143-50-0	Weak	Binder	Binder	Binder	Binder	Binder	Binder
Benz(a)anthracene[c] (BaA)	56-55-3	Weak	Non-binder[c]	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Enterolactone (ELactone)	78473-71-9	Weak	Binder	Binder	Binder	Binder	Binder	Binder

Chemical	CAS #d	Expected binding	Expected Classification[a]	Laboratory Classification Using Results from 3 Independent Runs				Overall Classification
Progesterone	57-83-0	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Octyltriethoxysilane (OTES)	2943-75-1	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Atrazine (ATR)	1912-24-9	Negative	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder	Non-binder
Subtask 4 (optional, coded chemicals)								
Norethindrone	68-22-4	Weak	Binder	Binder	-	-	Binder	Binder

[a] The expected response for each chemical was based upon published data from *in vitro* studies and were reviewed by a Chemical Advisory Board whose members were not directly associated with the validation study for the FW and CERI hrER Binding Assays. Chemicals were selected to represent multiple chemical classes and cover a range of binding affinity potencies commonly associated with ER agonist activity (6). When tested in the FW or CERI hrER Assays, the $LogIC_{50}$ for strong binders typically ranged from -9 to -7 (M), moderate -7.1 to -6.0 (M) and weak < -5.9 (M) (see Reference 6, Table 40).

[b] The use and classification of Di-n-butyl phthalate (DBP) as a non-binder was based on testing up to 10^{-4} M because the chemical was observed to be insoluble at 10-3M (e.g, turbidity) in some laboratories during the pre-validation studies. When DBP-TC was tested up to 10^{-3}M as a coded chemical, it was classified as 'equivocal' due to displacement of (^{3}H)17β-estradiol at highest in 3/5 laboratories using the CERI assay (6) and 5/6 laboratories using the FW assay (6).

[c] During the validation study, benz(a)anthracene was reclassified as a non-binder (i.e., negative) based on published literature demonstrating that the *in vitro* estrogenic activity reported for this chemical (14) is primarily dependent upon its metabolic activation (6). Enzymatic metabolic activation of the chemical would not be anticipated in the cell free hrER assays as used in this inter-validation study. Thus, the correct classification for this chemical is a 'non-binder' when used under the experimental conditions for the FW and CERI assays.

Yellow shading: Classification that is not concordant with that expected for the chemical.

1. **CERI hrER Binding Assay validation study:** The predictive capacity was assessed based upon historical documentation in published literature for 23 reference substances (17 positives, 6 negatives) that were tested in subtasks 1-3 and norethindrone (tested in subtask 4) in the CERI protocol (6) – see table 6.

Table 8: Predictive capacity for the CERI hrER Binding Assay

		CERI hrER binding assay			
		Binder	**Non-binder**	**Equivocal[a]**	**Total**
Expected results*	Binder	20	0	0	20
	Non-binder	0	6	1	7
	Total	20	6	1	27

*Based upon historical documentation in published literature - BRD for ER binding Test methods, ICCVAM, 2006, 2003, 2002) and Reference 6.

[a] Equivocal result indicates a binding result where the lowest point on the fitted response curve was between 76 and 51% (Table 10, Reference 6)

CERI hrER Binding Assay		
Overall Accuracy	**96%**	**26/27**
Sensitivity	100%	20/20
Specificity	86%	6/7
False positive	0%	0/7
False negative	0%	0/20
Positive predictivity	100%	20/20
Negative predictivity	100%	6/6

2. **FW *in vitro* hrER Binding Assay validation study:** The predictive capacity was assessed based upon historical documentation in published literature for 23 reference substances (17 positives, 6 negatives) that were tested in subtasks 1-3 and norethindrone (tested in subtask 4) in the FW protocol (6) – see Table 7.

Table 9: Predictive capacity for the FW *in vitro* hrER Binding Assay

		FW hrER binding assay			
		Binder	**Non-binder**	**Equivocal**	**Total**
Expected results*	Binder	19	0	1	20
	Non-binder	0	6	1	7
	Total	19	6	2	27

* Based upon historical documentation in published literature - BRD for ER binding Test methods (ICCVAM, 2006, 2003, 2002) and Reference 6.

[1] Equivocal result indicates a binding result where the lowest point on the fitted response curve was between 76 and 51% (Table 10, Reference 6)

FW hrER Binding Assay		
Overall Accuracy	**93%**	**25/27**
Sensitivity	95%	19/20
Specificity	86%	6/7
False positive	0%	0/7
False negative	0%	0/20
Positive predictivity	100%	19/19
Negative predictivity	100%	6/6

Table 10: Template for Accuracy Analysis

		New Test Outcome			
		Positive	**Negative**	**Equivocal**	**Total**
Reference Test Classification	Positive	a	c	e	a + c
	Negative	b	d	f	b + d
	Total	a + b	c + d	e + f	a+b+c+d

a = positive in both new assay and by reference test classification
b = positive in new assay and negative by reference test classification
c = negative in new assay and positive by reference test classification
d = negative in both new assay and by reference test classification
e = equivocal in new assay and positive by reference test classification
f = equivocal in new assay and negative by reference test classification
Accuracy = ([a+d]/[a+b+c+d])
Sensitivity = (a/[a+c])
Specificity = (d/[b+d])
Positive Predictivity = (a/[a+b])
Negative Predictivity = (d/[c+d])

9. Other files that document the data used to develop the PS are available in Annex N of the Integrated Summary Report (6). It includes Tables PS-5 and PS-6, an Excel file with Control runs used to prepare the 95% confidence intervals for performance criteria and Pdfs with graphs of curves for controls and for test chemicals added back to data analysis.