Unclassified **ENV/JM/MONO(2007)34**

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

27-Nov-2007

English - Or. English

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

ENV/JM/MONO(2007)34
Unclassified

SERIES ON TESTING AND ASSESSMENT
Number 85

REPORT OF THE VALIDATION PEER REVIEW FOR THE HERSHBERGER BIOASSAY, AND THE AGREEMENT OF THE WORKING GROUP OF THE NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP OF THIS REPORT

English - Or. English

JT03236933

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

OECD Environment Health and Safety Publications

Series on Testing and Assessment

No. 85

REPORT OF THE VALIDATION PEER REVIEW FOR THE HERSHBERGER BIOASSAY, AND AGREEMENT OF THE WORKING GROUP OF THE NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP OF THIS REPORT

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

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

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2007

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (in preparation)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Draft Guidance Document on Reproductive Toxicity Testing and Assessment (in preparation)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work Sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay fort he Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay : Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work Towards the Validation of the Rodent Utertrophic Assay - Phase 1 (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional data supporting the Test Guideline on the Uterotrophic Bioassay in rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the follow up of this report (2006)*

No. 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on the Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay - Procedure to Test for Antioestrogenicity (2007)*

No. 72, *Guidance Document on Pesticide Residue Analytical Methods (2007)*

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Toxicity Testing (2007)*

No. 75, *Guidance Document on the Honey Bee (Apis Mellifera L.) Brood test Under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 - Optimisation of the Test Protocol (2007)*

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances (2007)*

No. 79, *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa - Phase 1 (2007)*

No. 80, *Guidance on Grouping of Chemicals (2007)*

No.81, *Summary Report of the Validation Peer Review for the Updated Test Guideline 407, and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up of this report (2007)*

No.82, *Guidance Document on Amphibian Thyroid Histology (2007)*

No.83, *Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No.84 *Report on the Workshop on the Application of the GHS Classification Criteria to HPV Chemicals, 5-6 July Bern Switzerland (2007)*

No.85 *Report of the Validation Peer Review for the Hershberger Bioassay, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report*

© OECD 2007
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North 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 ten different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides and 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 the 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/ehs/).

This publication was produced within the framework of the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC).

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, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. 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.

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

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 presents the peer review report for the validation of the Hershberger Bioassay, preceded by the agreement of the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) on the follow up of the PRP report.

The peer review was managed by an independent consultant. The National Coordinators proposed peer reviewers but the final composition of the independent peer review panel was decided by the consultant. The panel included 5 peer reviewers who had been proposed by Germany, the European Commission and the United States, and one observer involved in the validation. The role of the observer was to respond to technical questions related to the validation. The peer reviewers were requested to send declaration of interests to the consultant. The documents submitted to the peer review panel were posted on the OECD public website and consisted of the validation reports (phase 1 report, draft phase 2 report and draft phase 3 report) and their annexes, a draft Background Review Document, the preliminary draft Test Guideline for the Hershberger Bioassay, references to peer reviewed publications and a Secretariat document explaining how the 8 OECD validation criteria and principles had been addressed.

This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

Contact for further details:
Environment, Health and Safety Division
Environment Directorate
Organisation for Economic Co-operation and Development
2, rue André-Pascal
75775 Paris Cedex 16, France

Tel : 33-1-45-24-16-74
Fax : 33-1-45-24 16-75
E-mail : env.edcontact@oecd.org

Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow up of the Peer Review Report

The peer review summary report of the validation of the Hershberger bioassay was submitted for information at the 19th meeting of the Working Group of National Coordinators of the Test Guidelines Programme (WNT) in March 2007. To follow up on the peer review, the WNT agreed that some recommendations for additional work/guidance need to be addressed, such as characterization of the rate of false positives, differentiation of a positive from a negative test, guidance on dissection. The WNT requested that the Validation Management Group for Mammalian Testing addresses technical issues, raised by the Peer Review Panel (PRP) or by the WNT, and proposes solutions to solve them.

Provided that the recommendations of the PRP are addressed and considering the benefit of the Hershberger bioassay for the screening of compounds that have androgen agonist, antagonist, or 5α-reductase inhibitor activity, the WNT agreed to proceed to the development and finalisation of the draft Hershberger bioassay Test Guideline, and agreed that both the immature and the castrate model should be included in the Test Guideline.

Summary

1. The peer review panel (PRP) was constituted in September 2006 to provide a review of the test method validation of the Hershberger Assay, to evaluate the data collected, and to address specific criteria outlined in the OECD GD34 document entitled "Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment". On page 23 of OECD GD34 there are eight (8) specific criteria that must be addressed by the laboratory charged with validating the Hershberger Assay. These criteria were developed by an Expert Group of the National Coordinators of the Test Guidelines Programme. The PRP consisted of 5 international experts qualified to review the Test Method Validation of the Hershberger Assay. In addition, the PRP was augmented by an observer, who assisted the PRP in their review and responded to any technical questions that were raised by the panel members.

2. The Hershberger bioassay is intended to detect androgen agonists, antagonists, and 5α-reductase inhibitors. Validation studies of this test method have been performed under the OECD Test Guidelines Programme. The peer review package submitted to the panel included the 3 validation reports, the draft test guideline, and a Secretariat document to support the peer review panel. In addition, panel members were provided with a draft background document entitled "Draft Hershberger Background Review Document" that was developed for Dr. Gary Timm of the US EPA. The panel members were given approximately 2 months to review the appropriate material and provide written responses to each of the eight (8) questions. The answers were compiled and several versions of the written responses were reviewed by each panel member prior to finalizing this report. Conference calls were arranged so that panel members could discuss various aspects of draft versions of this document and address areas requiring revision. This report presents the combined PRP responses to each of these charge questions.

3. While not specifically requested, several members of the PRP elected to provide some comments on the Test Guideline itself, in addition to answering the specific charge questions. While it is recognized that the Test Guideline is only a preliminary draft, these comments are included and summarized in a separate section at the end of this report.

4. The document provided to the PRP included a description and comprehensive summary of the study results for Phase-3 of the OECD validation of the rat Hershberger bioassay. It contained background on how the validation study was organised and performed, the standardised protocols used, detailed summaries and statistical analyses of the data, and the conclusions drawn from the studies. Phase-3 consisted of coded studies with two androgen agonists (at two dose levels each), three androgen antagonists (one of them at one dose level and the other two at two dose levels), and two negative reference chemicals (at one dose level each). A single protocol was used for the agonist studies based on direct administration of the agonists and statistically significant increases in the target tissues versus an untreated control. Similarly, a single protocol was used for the antagonists based on coadministration with a reference androgen and statistically significant decreases in the target tissues versus the reference androgen group as the control. The negative reference chemicals were tested by both protocols. The laboratory-testing portion of this phase was conducted between the 2nd and 4th quarter of 2004 and the last reports were obtained in the 1st quarter of 2005. As noted above, the PRP was also provided with a background document entitled "Draft Hershberger Background Review Document" that contained literature for 15 validation studies. However this document was sent to the PRP late in the overall review process (January, 2007), and was not reviewed by all panel members.

5. Regarding the overall validation exercise, the final conclusions and the views of the PRP are divided into broad groups. There were some differences expressed regarding the various components of the project. The PRP was not able to reach consensus on every aspect of the issue of the validation status of the Hershberger Assay. For example, there was general agreement that the validation exercise was successful

in detecting strong agonists. The lack of inclusion of weak agonists in the validation exercise was seen as a weakness by one panel member. However, strong agonists were identified at low dose levels. The use of only 2 negatives in the validation study was also viewed as a weakness by several panel members; however some members noted the extensive presentation of negative data in the background document that was provided as part of the review package. All panel members felt that the document describing the validation could have benefited from inclusion of historical data presented in tabular form for all agents, but especially those that tested as negatives in the assay. This information was available in the background document, which had not been peer reviewed and was not in a form that was easily accessible or utilized by all panel members.

6. In general, the PRP agreed that Hershberger assay could be effectively used as a screening assay for detecting compounds that have androgen agonist, antagonist, or 5α-reductase inhibitor activity. The transferability of the methods among laboratories has been documented, and it appears that differences among laboratories will not impact the ability to detect compounds with androgen receptor agonist or antagonist activity. The test was considered very sensitive for detecting strong antiandrogens, although there was a fairly high incidence of false-positive responses for the negative controls tested. However, at least one panel member noted that the positive results in some of the tissues for the negative reference chemicals analyzed in the assay do not preclude successful validation of the assay. One panel member noted that the stated purpose for this test is to serve as a screen. In that regard, screens are characterized by high false positive rates because their purpose is to identify compounds that require further analysis for the question raised. Screens serve as an initial evaluation and, therefore, it is preferable to falsely include a compound in the group requiring further analysis rather than falsely exclude it. Inclusion of historical data for agents that tested negative would have strengthened the validation document.

7. The PRP also noted that high dose exposures were able to yield results with excellent agreement among the laboratories. There was some disagreement among the PRP regarding the necessity for sensitivity at low doses. One member felt that sensitivity at low doses was a primary requirement since it was his understanding, based on statements in the OECD document, that the need for assay validation arose “from concerns that ambient levels of natural and industrial chemicals may interact with the endocrine system” (first paragraph (i), page 6 of summary of the surgical castrate model protocol, phase 3). In contrast, the other panel members felt that the screening nature of the assay allowed for the use of doses as high as the MTD, and thus did not feel that sensitivity at low doses was essential. Additionally, these panel members did not consider that the purpose of the assay was to address ambient levels of natural and industrial chemicals. One panel member noted that the less than optimal agreement among laboratories for the lower doses did not present an obstacle to validating this test because there was good agreement for the high doses. The panel member noted that the purpose of this test is not to make a quantitative evaluation, but a qualitative one and, therefore, the test is able to achieve its goals.

8. Overall, the majority of the PRP felt that the Hershberger assay was considered to be likely the best available *in vivo* assay for detecting strong androgen receptor agonists and antagonists, and when coupled with *in vitro* AR binding and transactivation assays, should allow a high confidence of identifying compounds that are potent or weak-acting androgen receptor agonists and antagonists. The panel cautioned, however, that *in vitro* binding and transactivation assays do not take into account absorption, distribution, metabolism and elimination processes that may be critical for a positive outcome *in vivo*. However, if a compound is positive in the Hershberger assay and also positive in *in vitro* assays this does add to the weight of evidence. One panel member felt that the absence of weak agonists in the validation test precluded the ability of this assay to be considered validated for weak agonists. Other panel members noted that there is a lack of environmental agents that have weak agonist activity and that weak steroid agonists have been tested successfully numerous times. These panel members pointed out that this information was available in the background document. There was unanimous agreement that data on strong, weak, and negative testing agents should have been extracted from the background document and

presented in a matrix form in the validation document. Inclusion in the validation document of this information would have clearly highlighted the extensive historical data base for this assay and the extent to which assay results for individual agents have been confirmed by other tests. This information would then have supplemented the limited number of agents tested in the validation procedure.

Background

9. The National Coordinators (WNT) agreed to establish a special activity to address the issue of endocrine disruption and develop new Test Guidelines as appropriate. The responsible body was the Task Force for Endocrine Disruptor Testing and Assessment (EDTA). The EDTA agreed to initially pursue efforts to develop and validate Test Guidelines for the uterotrophic assay and the Hershberger assay, and to evaluate enhancements to the current Test Guideline 407. A single Validation Management Group (VMG) was established to manage these projects. Subsequently, the EDTA has begun activities concerning ecotoxicity testing and *in vitro* or non-animal testing also related to the endocrine disruption issue. As a result, the original VMG is now divided into the VMG for mammalian effects assessment (VMG-mammalian), the VMG for ecotoxicity testing (VMG-eco) and the VMG for non-animal testing (VMG-NA) to manage the diverse activities and work loads.

10. The need to validate test methods for the detection of chemicals interfering with the endocrine system arises from the concerns that ambient levels of natural and industrial chemicals may interact with the endocrine system and as a consequence possibly elicit reproductive, developmental, and other adverse effects in humans and wildlife. The Hershberger bioassay is a leading candidate for a level 3 *in vivo* screening assay of the OECD Conceptual Framework to identify potential androgens and antiandrogens. It is rapid and efficient and there is strong evidence for its sensitivity and specificity from the literature.

11. The objective for the Hershberger Bioassay validation programme is to validate a test protocol in order to support the development of a Test Guideline for the detection of chemicals having the potential to act as androgen agonists or antagonists in rats. In the preceding validation Phase-1, a protocol was developed for identification of androgen agonists and antagonists. In Phase-2, the protocol developed during Phase-1 was used to test two further androgen agonists, four weaker androgen antagonists, and a potent 5α-reductase inhibitor. The test materials were supplied uncoded at pre-selected dose levels to obtain a dose-response curve by the participating laboratories. The Phase-2 validation program successfully achieved the goal of demonstrating the reproducibility of the protocol for detecting the weaker androgen agonists and antagonists, as well as the potent 5α-reductase inhibitor.

12. After successful completion of the Phase-2 validation testing, the Phase-3 validation was initiated. Coded substances were tested at one or two predetermined dose levels to exclude possible investigator bias. The same dose levels were used by all participants to further substantiate inter-laboratory reproducibility. The dose levels for the agonists and antagonists had already been used in the previous Phase-1 and Phase-2 test series or were derived from the Phase-2 results. In addition, two chemicals anticipated to act neither as an androgen agonist nor as an androgen antagonist were added to give an indication for the specificity of the Hershberger Bioassay. The overall goal of the Phase-3 validation study was to further assess the robustness and reproducibility of the Hershberger Bioassay in a blinded manner.

13. The charge of the PRP was to address specific criteria outlined in the OECD GD34 document entitled "Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment".

14. To facilitate review of this report, the eight (8) specific criteria posed to the PRP are listed along with the consensus responses. Additionally, responses to the criteria are also provided from each panel member (attached).

CRITERIA FOR TEST METHOD VALIDATION – Combined Comments from Members of the Peer Review Panel

Criterion	Peer reviewers comments on whether these criteria have been met
1) **The rationale for the test method should be available.** This should include a clear statement of the scientific basis, regulatory purpose and need for the test.	In general, there was sufficient rationale for the scientific and regulatory basis for the testing for endocrine-disrupters and the use of the Hershberger assay. For example, a clear statement regarding the rationale for the test is given. The test is designed to specifically identify a chemical as being and androgen receptor agonist or antagonist or 5α-reductase inhibitor. It does not identify chemicals that may interfere with androgen synthesis. Chemicals that interfere with metabolism may be detected through the use of the testosterone treated group. Emphasis on the use of *in vitro* assays as a preliminary requirement for the initial considerations and experimental design, prior to embarking on the Hershberger *in vivo* assay appears well balanced and in concert with the goals of reducing unnecessary animal experimentation. The value of such *in vivo* studies with respect to metabolic capacities, for strong acting metabolites, is also addressed in the validation report. It is stated that any positive outcome in the Hershberger Bioassay should normally be evaluated using a weight of evidence of approach, including *in vitro* assays such as the androgen receptor and *estrogen* receptor binding and transcriptional activation assays, as well as any *in vivo* data that may be available. This is rightly so due to “non-specific” effects that may occur at high doses.

Criterion	Peer reviewers comments on whether these criteria have been met
2) **The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.** This should include a reference to scientific relevance of the effect(s) measured by the test method in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.	The Hershberger Bioassay raises a clearly and narrow defined question that can be answered with the methods used. The assay is relevant for identification of androgen receptor agonists/antagonist, and apparently potent 5α-reductase inhibitors based on the data presented in the validation report. Alterations in androgen-dependent reproductive tissue weights are sensitive and relevant markers for androgen agonists, which will produce an increase in the androgen-dependent tissue weights, or androgen receptor antagonists, which will produce decreases in the androgen-dependent tissue weights in the presence of low-levels of an androgen receptor agonist such as methyl testosterone. For the compounds evaluated, the test is very sensitive for detecting antiandrogens. There were no weak-acting androgen receptor agonists evaluated to effectively evaluate the ability of the assay to detect weak-acting compounds; however, the potent androgen receptor agonists were detected at the "lower" doses. It was the opinion of one panel member that this test guideline may be appropriate to screen for moderate to strong androgen receptor (ant)agonists, when considering all five tissues together, but not weak to moderate androgen receptor (ant)agonists. Other panel members disagreed. One member noted that two weak antagonists, p-p'-DDE and Linuron, were tested and gave a positive result in the assay and felt that the necessity for weak antagonists to be tested a low dose levels if a positive response is seen at a higher dose level is questionable. Another member noted that the Hershberger Bioassay is a screening assay. The objective is not to determine a NOEL. As a screening assay doses up to the maximum tolerated dose should be acceptable. The test methods for this assay are not biologically relevant for questions regarding toxicity. The majority opinion was that the Hershberger assay is likely the best available *in vivo* assay for detecting androgen receptor agonists and antagonists, and when coupled with receptor binding assays, should allow a high confidence of identifying compounds that are potent or weak-acting androgen receptor agonists and antagonists. However, *in vitro* binding and transactivation assays do not take into account absorption, distribution, metabolism and elimination processes that may be critical for a positive outcome *in vivo* and thus, *in vitro* assays alone should not be used to evaluate compounds. The validation report could have made more correlations to the expected effects in humans given the mode of action of the chemicals detected in the Hershberger assay. However, this does not significantly impact the quality of the validation report since there is sufficient data in the scientific literature illustrating correlations between effects in test animals and humans, particularly with pharmaceutical agents. The choice of rat accessory sex organs is biologically relevant and mechanistically sensitive and adequate to detect the pharmacological effects of androgen receptor agonists and antagonists and 5α-reductase inhibitors. The use of animals is consistent with the objective of the assay to determine *in vivo* activity and may be included as one of a battery of tests to screen for endocrine disrupting activity. The strength of the assay lies in the use of several tissues that allows conclusions to be drawn due to the tissue specific growth responses. A slightly broader definition of an androgen antagonist would include the identification of chemicals which interfere with testosterone synthesis. These substances cannot be identified due to the use of a castration model.

Criterion	Peer reviewers comments on whether these criteria have been met
3) **A detailed protocol for the test method should be available.** The protocol should be sufficiently detailed and should include, *e.g.,* a description of the materials needed, such as specific cell types or construct or animal species that could be used for the test (if applicable), a description of what is measured and how it is measured, a description of how data will be analysed, decision criteria for evaluation of data and what are the criteria for acceptable test performance.	In general, the protocol is clearly written and comprehensive and the protocols used for each phase of the validation program were sufficient to evaluate the validation report. There are adequate descriptions of the materials, equipment, methodology, acceptable performance criteria and justification of animal numbers. The description of data analysis is comprehensive in the use of correct statistical approaches. It is not clear if a statistically significant reduction in only one of the tissues weighed would be sufficient to be called a positive signal or outcome? The Test Guideline makes a number of statements with regard to different modes of action having different tissue sensitivities. Clearly, more than one tissue would add to the confidence of the results, but the requirement should be for a statistically significant change in only one organ/tissue weight. One panel member felt that presentation of the statistical analyses could be improved (e.g., in paragraphs 'reproducibility' and 'predictive capacity') and additional analyses such as ANOVA specific post hoc tests (e.g., Dunnett's and Bonferonni tests) and control of type I error (for a multiple test situation) should be considered. In a teleconference discussing this matter, the technical advisor to the panel noted that an ANOVA combined with a post hoc t-test produced outcomes similar to those when a one-tailed Dunnett's test was used. The technical advisor commented that since this is a screening assay, he was concerned by the possible increase in false positives if a 2-tailed Dunnett's test was used. The panel agreed that an ANOVA combined with a post hoc one-tailed Dunnett's test based on an *a priori* hypothesis was probably a reasonable statistical approach. One panel member also commented that with the comparison between observed versus expected effects in phase 3 of the validation study, the data appear to be used very selectively, and without provision of ranges, with respect to numbers of laboratories included and phase comparisons (e.g., TP control [7 labs], flutamide from phase 1 [5 labs], linuron and pp DDE from phase 2 [4 labs]). The panel member questioned why data appeared to be excluded from the other laboratories and other phases.

Criterion	Peer reviewers comments on whether these criteria have been met
4) **The intra- and inter-laboratory reproducibility of the test method should be demonstrated.** Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.	The number of compounds evaluated and the number of different laboratories evaluating the compounds was sufficient to illustrate intra- and inter-laboratory variability/reproducibility. The report was more than adequate in presenting the data in a manner that facilitated review of the validation report. The endpoints of this assay are extremely robust and the inter-laboratory reproducibility is very good. The data presented indicate reproducibility and repeatability for the articles tested. This is especially true when high doses were evaluated. Overall, the 3-phase approach of the validation effort was very thorough and provided a sound approach for evaluating the reproducibility among laboratories and over time. There was not complete agreement among laboratories when the lower doses were evaluated. For example in the phase three study, there was 100% agreement among all laboratories when 160 mg/kg/d DDE was evaluated. There was only 20 to 50% agreement among the laboratories when the 16 mg/kg/d DDE dose was investigated. In the final analysis, this is not a problem because the purpose of the test is to serve as a screen (i.e., we are looking for a no or yes answer). Panel members noted the high variability among laboratories for some of the measured organ weights. As pointed out in the validation report, it is likely that differences in training may have accounted for some of this variability. One member noted that the ability of the laboratories to perform the protocol differs, as indicated by the CVs of the tissue weights, especially for the fluid filled tissues and that this will have a particularly important impact upon the detection of weak androgen (anti) agonists. Other panel members emphasized the fact that weak agents will have to be tested at high doses, up to the MTD. Specialised skills are fundamental to the production of accurate results. For example, while the use of immature animals as opposed to mature animals is a refinement, evident problems with the glans penis dissections by individual laboratories indicates that the protocol should give more specific directions in this regard. In view of the report discussion concerning the reduction of this problem, such that the castration date should be well after pnd 42, perhaps the optimum date should also be clearly specified. While there was considerable variability in organ weights among laboratories, the panel members felt that it did not adversely affect the ability of the Hershberger assay to detect androgen receptor agonists and antagonists. The data presented in the validation report suggest that the parameters that were allowed to vary among laboratories (e.g., rat strain, animal husbandry, etc) will not impact the ability to detect chemicals in the Hershberger assay. Appendix 4 would be better presented graphically with the range of variability within each lab clearly shown.

Criterion	Peer reviewers comments on whether these criteria have been met
5) **Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used.** A sufficient number of the reference chemicals should have been tested under code to exclude bias.	Overall, the performance of the Hershberger assay is adequate for use as a screening assay for detecting chemicals that interact with the androgen receptor and modify androgen signalling. The chemicals selected represented the mode of actions that were identified as important for detection by the Hershberger assay, namely androgen receptor agonists and antagonists. One panel member's opinion was that the phase 3 trial admirably demonstrated the protocol's ability to detect weak androgen receptor antagonists and 5α-reductase inhibitors. As stated in the report, phase I and II did not use coded samples, but phase 3 did. This is appropriate given the specific goals of the different phases of the validation effort. In general, the chemicals selected were appropriate, except that a weak androgen receptor agonist (e.g., androstenedione, DEHA) was not evaluated. A weak agonist would have provided a greater range of activities for the androgen agonist evaluations. However, since androgen agonists appear to be rare, this may not be a major short fall. Thus, this criticism does not detract from the validation report since multiple dose levels of testosterone propionate, methyltestosterone, and trenbolone were tested. While there was a sufficient number of androgen receptor antagonists evaluated to illustrate the ability of the assay to detect both androgen receptor antagonists, there were not enough negative control substances evaluated. Therefore, there is not sufficient data to characterize the false-positive rate. As indicated in the report, the false-positive rate for identifying antiandrogens is approximately 14% based on the two chemicals presented. However, a more robust evaluation of negative controls would be helpful. Also, it is not clear how compounds that effect testosterone metabolism via increased hepatic enzyme induction, particularly those that induce transient effects, will be detected in the Hershberger assay. There was unanimous agreement that data on strong, weak, and negative testing agents should have been extracted from the "Draft Hershberger Background Review Document" and presented in a matrix form in the validation document. Inclusion in the validation document would have clearly highlighted the extensive historical data base for this assay and the extent to which assay results for individual agents have been confirmed by other tests. This information would then have supplemented the limited number of agents tested in the validation procedure.

Criterion	Peer reviewers comments on whether these criteria have been met
6) **The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data.** In the case of a substitute test method adequate data should be available to permit a reliable analysis of the performance and comparability of the proposed substitute test method with that of the test it is designed to replace.	Overall, the discussion of the data in light of the known effects of the test materials was sufficient for evaluation of performance of the Hershberger assay. Comparisons were presented between positive pharmacological outcomes in the Hershberger screen and true adverse effects in developmental studies to indicate the appropriate species was used and the sensitivity of the methods appropriate. It is recognized that relevant human data on the test materials is typically not available. However, given the understanding of the effects of androgen receptor agonists and antagonists on reproductive organs in both experimental animals and in humans treated with pharmaceutical agents, the data are presented in a way that allows correlation of the data from the validation report to expected effects in humans. Correlations to rodent studies were presented to adequately show consistent effects between the Hershberger assay and the effects expected for the positive control chemicals.

Criterion	Peer reviewers comments on whether these criteria have been met
7) **Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.** Aspects of data collection not performed according to GLP should be clearly identified and their potential impact on the validation status of the test method should be indicated.	While not all data were collected in accordance with Good Laboratory Practices (GLPs), all appropriate documentation was included, and there appears to be no impact on the validity of the studies. In addition, the results were very consistent among laboratories, and therefore the impact from the laboratories that did not conduct the analyses in accordance with GLPs was not sufficient to raise concerns as to the validity of the validation effort.

Criterion	Peer reviewers comments on whether these criteria have been met
8) **All data supporting the assessment of the validity of the test method should be available for expert review.** The detailed test method protocol should be readily available and in the public domain. The data supporting the validity of the test method should be organised and easily accessible to allow for independent review(s), as appropriate. The test method description should be sufficiently detailed to permit an independent laboratory to follow the procedures and generate equivalent data. Benchmarks should be available by which an independent laboratory can itself assess its proper adherence to the protocol.	The presentation of the report was adequate to make an assessment and all documentation needed to adequately evaluate the performance of the Hershberger assay was publicly available on the OECD website. However, one panel member noted that aspects regarding the presentation of the statistics that appear to be subjective. Terminology such as ‘appropriate’ or ‘consistent’ is regularly used without definition. The statistical presentation could be improved in this regard.

Recommendations

1. It may be more prudent to reserve the Hershberger Assay for chemicals that are negative in *in vitro* assays. In vitro screening should be the first method of choice for screening. Chemicals which are negative in vitro could then be screened in the Hershberger Assay for possible activity *in vivo* due to influences of absorption, distribution, metabolism and elimination. Prevalence must be considered when screening negatives *in vivo*.

2. One important point is that this protocol is not designed to detect substances that interfere with testosterone synthesis.

3. As a screen, this test should not be used to determine No Observed Effect Levels (NOELs) for compounds and it is not suitable to be used in risk assessment for quantitative purposes. The Test Guideline calls for the establishment of a NOEL. When only 2 dose levels (plus control) are used, and one is maximum tolerated dose (MTD), it is not clear why establishment of a NOEL is necessary for a pharmacological screen. One might surmise that all that should be required is the demonstration of an appropriate biological signal – not at what dose level this occurs, or does not occur. If the users of the protocol have ensured appropriate dose selection of the high dose to remove confounding toxicity, the achievement of a NOEL should not be required and may actually detract from obtaining reasonable dose response information for the pharmacological effect. References to the use of this assay to determine NOELs should be eliminated.

4. Additional work should further characterize the rate of false-positives. There were no weak-acting *androgen receptor* agonists evaluated to effectively evaluate the ability of the assay to detect weak-acting compounds; however, the potent *androgen receptor* agonists were detected at the "lower" doses. The laboratory personnel need further training on handling and dissection of small organs.

5. Clear guidance should be given to define the maximum tolerated dose for use in the screening assay and also define how to differentiate a positive test from a negative test.

6. The Data Interpretation Procedure needs further discussion/expansion. For example, will positive results trigger more definitive *in vivo* studies? It would be valuable to provide additional guidance around what constitutes a positive response (and vice versa for a negative response). For example, if only one organ weight was affected, would the test still be considered positive? Moreover, some additional guidance around the maximum dose to be administered

should be provided. There is some discussion of dosing at a level to reach an MTD; however, no information was provided as to what constitutes reaching an MTD. It will be particularly important when compounds are negative in the Hershberger assay that they have been dosed up to an MTD

7. Dose-response requirements in the designation of a positive response are also an issue. The Test Guideline asks that changes in organ weight in response to the (anti)androgen test articles should be dose–responsive. This needs further discussion in the guidelines. No rationale was provided for these criteria. It is easy to envision circumstances where the low and high dose levels in the screen produce a statistically significant change in organ weights at both levels, but this may be a flat dose-response, or the lower dose may produce a larger response than high dose (for a variety of reasons, including toxicokinetics). This would not negate the finding of a positive signal in the screen, but if a GL requirement for a positive is for dose responsive findings, then such an outcome as described above may not, or would not, be deemed a positive response and this would be an error in a screening battery.

8. There was unanimous agreement among panel members that the validation document could have been greatly improved by presentation of data for other tested agents, especially negatives, within the validation document in a table format that presented outcome in the Hershberger assay compared with outcomes in other in vitro or in vivo tests of androgenic activity. This table would be complete with the appropriate reference for each individual study. Inclusion of the range of doses used would be essential. Thus, potential issues as to the usefulness for the assay in correctly detecting negative, weak positives and strong positives would be readily apparent. This table would also provide an immediate answer to the question as to how many negatives have been run through the assay historically. The use of only 2 negatives in the validation study is viewed as a weakness by all panel members. Inclusion of this table in the validation document would have provided the reviewers with the ability to assess the degree to which the assay was historically able to detect negatives, thus decreasing the need to include negatives in the validation assay itself.

Comments on the Draft Test Guidelines

1. The draft guideline is well written, comprehensive and conveys a high scientific and balanced message. Emphasis on the use of *in vitro* assays as a preliminary requirement for the initial considerations and experimental design, prior to embarking on the Hershberger in vivo assay appears well balanced and in concert with the goals of reducing unnecessary animal experimentation. The

value of such *in vivo* studies with respect to metabolic capacities, for strong acting metabolites, but not weak (ant)agonists or negatives, is also addressed in the validation report.

2. Two key points need to be made and should be included in the draft Test Guidelines. The document does not define what a screening test is (this would be useful to any user of the Test Guidelines) and how this differs from other OECD testing activities. The initial preamble should also state that this guideline is for the measurement of a pharmacological, not a toxicological, response to the action of (anti)androgens.

3. The question with regard to positive and negative outcomes from the screen still requires further work and clarification in the Test Guideline. In particular: 1) why is a NOEL needed in a screen? Indeed, the establishment of a NOEL could be misinterpreted that these data could be used for quantitative risk assessment purposes. This is not a physiological model. 2) How many tissues are required to have a statistically significant change to call a positive response? 3) Why do the data have to show a dose response to be considered a positive, if statistically significantly different from control? The guideline states that we should be trying to eliminate false negatives but can accept some false positives.

4. One panel member felt that the route of administration needs further consideration. This panel member commented that diet administration might be preferable for animal welfare considerations. However, dietary administration can lead to errors in dosing, especially in the group housing conditions preferred. Other panel members did not agree on the dietary route as appropriate in the assay and noted that gavage gives more precision of administered dose in a short term study and obviates the need for palatability studies and other diet analyses (e.g., distribution and stability). In general, diet is a more difficult and costly matrix for chemical analysis than a dosing vehicle.

5. In terms of reducing animal numbers, it might be beneficial to have fewer better trained laboratories in the validation effort, assessing more (negative) compounds. The 2 negative compounds presented are insufficient for proper evaluation.