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

25-Jul-2008

English - Or. English

ENV/JM/MONO(2008)21
Unclassified

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

SERIES ON TESTING AND ASSESSMENT
Number 94

REPORT OF THE VALIDATION PEER REVIEW FOR THE 21-DAY FISH ENDOCRINE SCREENING ASSAY AND AGREEMENT OF THE WORKING GROUP OF THE NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP OF THIS REPORT

JT03249201

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 94

REPORT OF THE VALIDATION PEER REVIEW FOR THE 21-DAY FISH ENDOCRINE SCREENING ASSAY 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 2008

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, *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (2008)*

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 for the 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 (2007)*

No. 86, *Report of the OECD Validation of the Rodent Hershberger Bioassay: Phase 2: Testing of Androgen Agonists, Androgen Antagonists and a 5 α-Reductase Inhibitor in Dose Response Studies by Multiple Laboratories (2008)*

No. 87, *Report of the Ring Test and Statistical Analysis of Performance of the Guidance on Transformation/Dissolution*

of Metals and Metal Compounds in Aqueous Media (Transformation/ Dissolution Protocol) (2008)

No.88 *Workshop on Integrated Approaches to Testing and Assessment (2008)*

No.89 *Retrospective Performance Assessment of the Test Guideline 426 on Developmental Neurotoxicity (2008)*

No.90 *Background Review Document on the Rodent Hershberger Bioassay (2008)*

No.91 *Report of the Validation of the Amphibian Metamorphosis Assay (Phase 3) (2008)*

No.92 *Report of the Validation Peer Review for the Amphibian Metamorphosis Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report (2008)*

No.93 *Report of the Validation of an Enhancement of OECD TG 211: Daphnia Magna Reproduction Test (2008)*

No.94 *Report of the Validation Peer Review for the 21-Day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2008)*

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 developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

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.

**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 report of the peer review of the validation of the 21-day fish screening assay, performed by Japan in 2007. It is preceded by a statement from the Working Group of National Coordinators of the Test Guidelines Programme concerning the outcome of the peer review and the follow-up work.

Report of the Validation Peer Review for the 21-day fish endocrine screening assay and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the Follow-up of this Report

The Peer Review Panel Summary Report of the validation of the 21-day fish endocrine screening assay was submitted for information to the Working Group of National Coordinators of the Test Guidelines Programme (WNT) in February 2008. Following the recommendations from the Report, and taking note of the general views of the Validation Management Group for Ecotoxicity Testing (VMG-eco) expressed at its last meeting in January 2008, the WNT agreed that:

i) additional guidance on the standard operating procedures of the test method should be made available to improve repeatability of the assay, and that

ii) a thorough literature review should be conducted to collect existing data on the three fish species to further support the validation of the assay, or to identify, if necessary, the most relevant chemical(s) to be further tested.

The WNT requested that the VMG-eco and its Fish Drafting Group address technical issues identified by the Peer Review Panel or by the WNT and propose solutions to solve them, and this should include consideration for inclusion of additional validated endpoints, as appropriate.

Provided that the recommendations of the Peer Review Panel are addressed and considering the benefit of the 21-day Fish Endocrine Screening Assay for the detection of substances that have estrogen agonist, androgen agonist and aromatase inhibiting activities, the WNT agreed to proceed to the development and finalization of the draft Test Guideline in a reasonable timeframe.

The Test Guideline should include at minimum the VTG and secondary sex characteristics endpoints, and as soon as the fecundity and gonad histopathology are recognized as robust endpoints by the VMG-eco and its Fish Drafting Group, these endpoints should be included in the draft Test Guideline or a revised Test Guideline, as they broaden the scope and applicability of the Test Guideline.

Report of the Validation Peer Review for the 21-day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report

Background

1. The Working Group of National Coordinators (WNT) agreed to establish a special activity to address the issue of endocrine disruption and develop guidelines to evaluate these effects. This task was assigned to the Task force for Endocrine Disruptor Testing and Assessment (EDTA) as a responsible body. EDTA has begun to pursue efforts to develop and validate test guidelines in the fields of toxicity testing, eco-toxicity testing and *in vitro* testing. A Validation Management Group (VMG) was established to undertake practical management of these projects. This was further divided into three sub-groups: VMG-mammalian; VMG-eco; and VMG-non animal. The Fish Screening Assay is managed by VMG-eco.

2. Under the overview of the VMG-eco, validation work of fish screening assay has been conducted in three different phases: I, II and III. Within these phases, three different fish species were evaluated with up to 11 different chemicals.

Peer Review Process

3. A peer review panel was formed in July 2007 to provide a review of the validation process for the 21-day Fish Screening Assay. Seven candidates from different sectors were nominated with one withdrawal. Therefore, the peer review was conducted by six reviewers.

4. The reviewers were asked to consider the specific criteria that are set out in "Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, GD34" developed by OECD (Organisation for Economic Cooperation and Development). These criteria are as follows.

5.
1) *The rationale for the test method should be available.*
2) *The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.*
3) *A detailed protocol for the test method should be available.*
4) *The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.*
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.*
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.*
7) *Ideally all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.*
8) *All data supporting the assessment of the validity of the test should be available for expert review.*

6. The peer reviewers submitted separate reports. However, as an aid to the participating laboratories and the assessment panel, the comments have been sorted and compiled. All comments that were specifically related to the eight criteria are provided in Annex 1. Other more general comments (in

some cases not specifically related to a particular criterion) are provided in Annex 2. Also, some specific comments on the draft test guideline itself were submitted by some of the reviewers. These are provided in Annex 3. A list of the documents that were used by the peer reviewers is provided in Annex 5.

7. The peer review panel was requested to report their views concerning the validation process on 21-day Fish Screening Assay to VMG-eco which is responsible for the validation process. This report will be further submitted to EDTA, followed by the submission to WNT. On the basis of the peer review report, EDTA and WNT may recommend actions for the further development of an official test guideline.

8. This report reflects the consensus view of the peer review panel. However, some different opinions/views were expressed by reviewers on some criteria. These differences are described in the report and will be the basis of the consideration by EDTA and WNT for their decisions to the next step on this project.

Summary of comments for each criterion

Criterion 1: ***The rationale for the test method should be available.***

9. The reviewers all agree that the rationale is clearly set out. However, from some of the comments relating to other criteria ('fecundity/spawning and histology should be brought back as endpoints', 'histopathology should be measured', 'spawning should be used as an endpoint'), it is obvious that, while all reviewers acknowledge that it is 'available', some do not 'agree with its scope/contents'. It was unclear to one reviewer if the method was developed only for screening of pure chemicals or if it would be applied to screening of sewerage effluents. This should be taken into consideration.

10. Notwithstanding the above comments, the peer review panel agrees that this criterion is met.

Criterion 2: ***The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.***

11. Concerning this criterion, several, sometimes conflicting opinions/views were submitted by the reviewers.

12. Reviewer 6 produced three very useful Tables that summarized the results in each species. There was general consensus that vitellogenin (VTG) induction in male fish was a robust measure of estrogens (both strong and weak) in all three species - with a caveat, from reviewer 1, that the state of maturation of the males might affect VTG production.

13. The results showed that a decrease in VTG in female fish could be brought about by both an aromatase inhibitor and a strong (non-aromatizable) androgen. Reviewers 2 and 3 pointed out that the mechanism whereby the latter compound caused this reduction was not clear. Reviewer 3 also pointed out that there were likely to be at least two other classes of compounds (estrogen receptor antagonists and ArH receptor agonists) that could also reduce VTG levels in females. However, such compounds have not yet been tested. Reviewer 6 makes the point that toxicity of certain compounds could also be predicted to be the cause of a reduction in VTG concentrations in females.

14. Reduction in secondary sexual characteristics (SSC) in males was not found at all in medaka and

to weak estrogens, an aromatase inhibitor and potassium permanganate (negative control – toxic effect) in FHM. Neither species showed any response to an anti-androgen. On the basis of these results, none of the reviewers supported SSC reduction in males as being a useful endpoint.

15. An increase in SSC in females was found in response to a strong (non-aromatizable) androgen in medaka (but not reliably in FHM). An increase in SSC in females was also found in response to an aromatizable androgen in FHM, but was not tested in medaka. Although this endpoint found favour with four of the reviewers (because the appearance of SSC in females is 'unequivocal'), their message is nonetheless confusing, as reviewer 3 suggested that the appearance of SSC in females was reliable only for detecting strong (non-aromatizable) androgens such as trenbolone while reviewer 2 suggested it was useful only for detecting aromatizable androgens.

16. Reviewer 5 questioned the utility of measuring the size of tubercles in the FHM.

17. Reviewer 6 suggested the elimination of zebrafish as a test species as this fish did not have useful SSC (and males had high background level of VTG – see later).

18. Reviewer 2 (see Appendix 2) suggested that medaka should be the 'species of choice' (i.e. both zebrafish and FHM should be dropped) as it had SSC and required the least number of fish for testing.

19. Reviewer 6 suggested that the dropping of fecundity as an endpoint meant that the possibility was now open, if supported by information provided through the peer-reviewed literature or new studies, to cut the duration of the test to 14 (or even 7 days) and to abandon the use of replicates (and the need to maintain separate spawning groups). This would make it far more economically viable as a screening test.

20. Reviewer 1 wished to re-incorporate fecundity as an end-point (on the basis that it is important in an ecological context). However, reviewer 2 considered that, as an endpoint, this was neither consistent nor specific for screening endocrine active substances (which is the stated purpose of the screen). Reviewer 3 also considered that the rational for removing these endpoints has been clearly stated and the test now appeared more appropriate for screening purposes.

21. Reviewer 3 suggested that direct measurement of aromatase activity would be the best endpoint for measuring the activity of aromatase inhibitors.

22. The consensus of the peer review panel was that this criterion has not been fully met.

Criterion 3: ***A detailed protocol for the test method should be available.***

23. Concerning this criterion, many opinions/views were submitted from peer reviewers. These are précised below:

24. Lack of clarity on necessary numbers of fish/replicates (reviewers 1, 2, 3). Not enough information on blood sampling and handling protocols (reviewers 1, 3, 6). No information on 'acceptability' criteria for VTG levels (reviewers 1, 2, 3). Incorrect use of power table and a lack of consideration for 3R (reviewer 2). Lack of clarity concerning use/non-use of solvents (reviewer 3). Clearer advice needed on range of tested concentrations (reviewers 2, 3). Advice needed on how to measure SSC in live fish without stressing them (reviewer 3). Lack of clarity about how to apply a positive control (reviewer 3). No clear description of genetic background of fish (reviewer 4). No protocol for breeding and handling of fish (reviewer 4). For negative substances, number of fish in some cases too small to draw firm conclusions (reviewer 5). No information on spawning substrate

and water conditions (reviewer 6). No information of how to score SSC in medaka (reviewer 6). Lack of a complete draft protocol (reviewers 2, 5, 6).

25. Reviewer 6 had some suggestions for streamlining the tests (exposure for only 7 or 14 days, no replicates, VTG in males and females, and SSC in females only) if these tests were truly to be considered as "screening" assays.

26. From this list, it can be seen that the peer review panel consider that this criterion has not been fully met.

Criterion 4: ***The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.***

27. There were also a large number of comments on this criterion:

28. Four reviewers considered inter-lab variability was mostly acceptable – i.e. results were in general consistent/reproducible for the vitellogen induction endpoint (reviewers 1, 2, 3, 6), though the way of dealing statistically with 'below detection' levels of VTG by one laboratory was considered unacceptable (reviewer 2)*. Reviewer 5 was of the opinion that inter-lab variability was unacceptable – i.e. the large CVs indicated that only large changes could be detected. Reviewer 4 suggested that fish with the same genetic background should be used by all laboratories. Reproducibility for SSC was generally higher than for VTG (reviewer 1). Reproducibility of VTG was lower in zebrafish than other species (reviewers 1, 3). More acceptable design would be to have fewer labs testing more compounds (reviewer 2). Intra-lab variability remains to be established (reviewers 2, 6). Plots provided for annex 2 phase 2 were poorly presented (reviewer 2). Some labs were unable to quantify size of tubercles in FHM and thus only appearance and number of tubercles should be made mandatory (reviewer 5). The fact that androstenedione was only tested in FHM means that the reproducibility of this endpoint is not yet established (reviewer 6).

29. Although the response of the peer reviewers was generally constructive, it nevertheless seems clear that some problems still need to be cleared up – including an assessment of intra-laboratory variability (that could be possibly be gained by analyzing data separately from duplicate tanks – see comments by reviewer 2).

30. The peer review panel considers that the criterion has not been fully met.

** one of the review compilers would like to point out that Stata provide a statistical tool, TOBIT, specifically for ANOVA of 'left-censored data' (i.e. where some of the observations fall below level of detection). This might perhaps provide a solution to analyze the data.*

Criterion 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.***

31. Three reviewers (1, 4 and 5) were satisfied the criterion was met or had no comment. However, the following three points were made by the remaining reviewers:

32. The fact that more chemicals have been added at each phase means that the test has been more 'a series of optimization steps' than a validation (reviewer 2). The testing of only two types of weak estrogens (that were from the same 'family' of chemicals), only one AR agonist, one aromatase

inhibitor and no ER antagonists is a weakness as it means that it is impossible to determine if the present biological responses are typical of such compounds (reviewer 3). The effect of aromatizable androgens in FHM (increase in VTG in females and increase in SSC in males) could challenge interpretation of results – especially if the test is ever applied to environmental samples (e.g. water treatment plants) (reviewer 6).

33. Unless these points can be answered, it is considered that this criterion is not fully met.

Criterion 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.***

34. Two reviewers stated that this had been met, two had no comments and two thought it had been partially met. The comments were:

35. Table 4 in guidance document 34 showed satisfactory concordance (reviewers 1, 3). There are several papers on FHM that should have been quoted in support of findings (reviewer 2). Reviewer 2 suggests retrospective analysis of data in the literature would aid the validation. Based on published data on medaka reproduction, experiments should have been carried out with 12 week old fish as opposed to 16 to 24 week old fish (reviewer 4). Negative test substances have not been sufficiently tested in medaka and zebrafish.

36. Overall, the peer review panel agrees that this criterion has only been partially met.

Criterion 7: ***Ideally all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.***

37. All peer reviewers answered either yes to this criterion or had no comment. Although not many tests were formally certified as GLP studies, most participating laboratories appear to have conducted their studies according to GLP principles.

38. Overall, the peer review panel considers that this criterion has been met.

Criterion 8: ***All data supporting the assessment of the validity of the test should be available for expert review.***

39. The general consensus was that all data (i.e. results) were available, although reviewer 2 thought that it varied greatly in quality, legibility and reliability. Reviewer 1 complained of the lack of detailed information on procedures to collect and process biological samples for VTG determination, and reviewers 3 and 6 indicated that the methodology was overall not detailed enough to permit other to follow the procedures and generate equivalent data.

40. Thus, the peer review panel considers that this criterion has only been partially met, and that more effort needs to be put into producing clear, detailed protocols.

Conclusions

41. As a whole, the following conclusions are obtained.

Criterion 1: Considered fully met – but there may be a problem in acceptance of the rationale by some reviewers.

Criteria 2: There are some very fundamental suggestions/questions about the course of development of the screen (e.g. abandoning the zebrafish; re-introducing fecundity and histopathology as endpoints; reducing SSC endpoint to induction in females only). Testing of more compounds is also suggested (see also criterion 5).

Criteria 3 and 8: It seem very obvious from the number of comments that the reviewers felt starved of methodological information. All laboratories obviously need to provide far more detailed protocols.

Criterion 4: Intra-assay variability still needs to be resolved and some questions related to inter-assay variability need to be answered.

Criterion 5: More chemical testing is requested - mainly to resolve problems of interpreting concomitant changes in VTG and SSC. Taking all suggestions from all reviewers, the compounds that are recommended for inclusion are: androstenedione, octylphenol, pcp and methoxyethanol in medaka; another weak estrogen (that is not related to octylphenol), another non-aromatizable androgen, another aromatizable androgen, an AhR agonist and an ER antagonist in FHM and medaka.

Criterion 6: Better use of the existing literature for FHM.

Criterion 7: No major problem.

.

Criterion 8: See statement for Criterion 3 above.

42. This report should form the basis for decisions on whether the validation exercise meets the OECD principles for validation for development of the test guideline. In this consideration, TF and WNT should note the various views of peer reviewers. The peer review panel recommends that the TF and WNT consider this report to decide any further work to finalize the validation activity which links to the development of a new OECD test guideline.

Annex 1

Comments from peer reviewers are summarized by each criterion.

Criterion 1) The rationale for the test method should be available.

Reviewer	Answer	Explanation
1	Yes.	There is clear rationale to develop a standard test methodology for the screening of endocrine active substances in fishes because (a) anthropogenic contaminants found in the aquatic habitat can affect the endocrine system of fishes; (b) changes in the endocrine system of fishes can indicate the occurrence of deleterious health effects or serve as early warning signs for these effects; (c) information collected with other vertebrata taxa cannot be directly extrapolated to fishes; and (d) there are currently no OECD test guidelines to detect endocrine active substances in the aquatic environment and more specifically those that affect fishes.
2	Criteria 1 are met.	There is a very strong scientific rationale for this test method as shown by the number of high quality publications in the scientific literature, and the rationale presented by the Secretariat paper as of April 2007.
3	Fully met	The rationale for proposing the test method is clearly stated with regard to the scientific basis and regulatory purpose. Under the conceptual framework, it would provide information about certain endocrine mechanism as a level 4 in vivo assay. This latter point could be discussed and the test method could be introduced at Level 3 of the actual OECD conceptual framework. The title of the assay method should indicate the limitation of the test (Fish Screening Assay for certain Endocrine Active Substances)
4	Yes	The rationale for the test method is well established and reasonable because it is well documented in the scientific literature that the production of vitellogenin in the liver in females and development of the secondary sex characteristics in males are controlled by sex hormones. However, these characteristics are only indicators of the activity levels of sex hormones, and should not be considered as directly affecting the reproductive activity of fish. In tests of activities of various chemicals on the endocrine system, which affects the reproductive activity of fish, the meaning of these endpoints is weaker in comparison to the assessments of direct endpoints such as spawning activities and gonad histology. It would therefore be desirable to have these direct characters adopted as endpoints for the assays. The establishment of conditions that make it possible to assess endpoints such as these in the assay is expected to constitute future work.
5	Criteria 1 is met	The purpose of the validation was the identification by the OECD of the need for developing testing guidelines to assess the effects of potentially endocrine disrupting substances in different test systems. In the field of ecotoxicity, fish were identified to be one of the target organisms for experimental testing. In order to provide appropriate test methods for different tiers of a hazard and risk assessment, the requirement for a screening test was identified and, subsequently, a specialized task group (fish drafting group) was formed by OECD to direct and monitor the validation of a respective test method. The basis of the test method development came from assay proposals made by several institutions in the US and Europe, and after intensive discussions in the drafting group, a concept of a 21-days screening test in adult fish was determined to be probably

		most appropriate for the purpose. In general, all aims of the validation have been achieved and the reports support the generation of the respective test guideline. Therefore, my general remark is that the assay can successfully be used for the purpose it was designed
6	Criteria 1 is MET	The rationale for the test method is adequately described in the Introduction section of Preliminary Draft and thoroughly described in the Secretariat Paper supplied for this review. The Secretariat Paper includes both the scientific and regulatory basis for the test method.

Criterion 2) The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.

Reviewer	Answer	Explanation
1	Yes, with qualifications.	Upon completion of Phase 3 of the validation process, the recommended core endpoints were narrowed down to two: vitellogenin (Vtg) levels in plasma (fathead minnow and zebrafish) or liver (medaka); and secondary sex characters (SSC) for fathead minnow and medaka. It was already known prior to the conduct of the present validation studies that hepatic Vtg production and secondary sex traits in teleost fishes are generally under the control of sex steroids. The results generated by the present studies have corroborated this earlier knowledge. Namely, Vtg is under the stimulatory control of estrogen (E2) in all three model species and is normally produced in high levels only in females; and, in fathead minnow (nuptial tubercles) and medaka (papillary processes), male SCC are under the control of androgen (11KT). Thus, as noted in the report, there is a mechanistic relationship between the endpoints (Vtg, SSC) and the biological phenomenon of interest (disruption of sex steroid endocrine system) in the present test method. However, although the present report has generally succeeded in describing the relationship between the test method's endpoints and the biological response, there are some caveats that should be considered. For example, the relationship between exogenous estrogenic substances and Vtg production in males may not be as straightforward as commonly believed. There is some evidence suggesting that 11KT, which although by itself may not directly induce Vtg production, can enhance the Vtg response to exogenous estrogens both in male and female fish (e.g., Mori et al. 1998, Journal of Steroid Biochemistry and Molecular Biology 67: 133-141; Asanuma et al. 2003, Fish Physiology and Biochemistry 28: 383-384.). Whether this effect is due to androgen aromatization into estrogens by the liver or not, it suggests that the reproductive status of males (in terms of endogenous 11KT production levels) can affect their response to estrogenic substances in the water. This scenario may at least partly explain the apparent variability in the baseline levels of Vtg reported for males by the present studies. This reviewer agrees with VMG-eco recommendation that additional endpoints should be added to the test method once some of the problems of reproducibility are resolved. In particular, an effort should be made to incorporate fecundity to the battery of endpoints because this would make the test results more relevant in an ecological context.
2	Overall criteria 2	It is fully met for the primary test method endpoint of vitellogenin (VTG) induction in male fish as an indication of the estrogenicity of a test substance. Similarly the decrease in the VTG induction in female fish, as an indication of the inhibition of

	are partially met.	aromatase is also met. The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest is well described for VTG, for the detection of estrogenic chemicals via the measurement of VTG in male fish for all three test species. However the biological relevance of a decrease in VTG production in female fish is not established for substances acting via non aromatisable androgenic substances. Secondary sex characteristics endpoints can only be used for two species: fathead minnow and medaka. But the data generated appears to be vulnerable to confounding, especially with respect to false positives in males. Additionally, as the fish are also vulnerable to handling, which can affect changes in skin colour, the risk of confounding and subjective observation appears to be quite high. It is therefore recommended that secondary sexual characteristics are only considered for female fish where the nuptial tubercles are unequivocally formed following exposure to androgenic substances but not non aromatisable androgenic substances. Both gonad histopathology and fecundity are inconsistent, fecundity is not an appropriate specific endpoint for endocrine disruption and both endpoints were not successfully validated.
3	Partially met	The assay measures two core biomarker endpoints as indicators of endocrine activity : vitellogenin, an hepatic ER-regulated protein and Secondary Sex Characters which are under the control of androgens in FHM and MDK. Vitellogenin in male and female. - estrogenic properties of chemicals are assessed by measuring the induction of an ER-regulated protein, the vitellogenin, in male fish. The measurement of Vtg in male is relevant. The assay provides mechanistic information about the substances. - In the assay, Vtg level is measured in female to assess the endocrine effect of substances that act as aromatase inhibitors or as androgen agonist. For androgen there is no mechanistic link between AR and vitellogenin down-regulation. Decrease of Vtg synthesis is also believe to be a consequence of the decreasing circulating estrogens due to the inhibition of estrogen biosynthesis. It is likely that decreasing Vtg level in female will occurred in case of exposure to ER-antagonist compounds (not tested in the validation process) as well as compounds that are able to interfere negatively with the ER signalling pathway (e.g. AhR compounds). Reduction of circulating Vtg level in female will therefore not provide clear information concerning the mechanism of action of the test substance. - The interest and the relevance of Vtg measurement in female within the context of the fish screening assay for EAS is appropriate but it could specify in the text that decrease of Vtg can occurred after exposure of female to compounds having various mode of action. - In the specific case of aromatase inhibitors, direct measurements of aromatase activities in target tissues will be probably more suitable and will allow assessment of substance that can also induce aromatase. In this regards, compounds that have been described as aromatase inhibitor based on in vitro assays can also induce ovarian aromatase activity in vivo in female fish (Villeneuve et al., Aquatic Toxicology, Vol 76, Issues 3-4, 353-368). Furthermore, vitellogenin decrease is not systematically associated with decreasing concentration of estradiol. In this regards, Ankley et al., (Toxicological Sciences, 2005, 86:300-308) showed that fenarimol decreased plasma vitellogenin concentrations in female fathead minnows but

		increased plasma estradiol concentrations. Secondary Sex Characters in FHM and MDK During the validation process, SSC appeared as not reliable endpoints for detecting EAS acting through various mode of action. SSC are therefore used to assess androgen agonist compounds only and appear suitable to detect very potent androgen agonist.
4	No comment	No comment
5	Criteria 2 is met.	Criteria 2 is met. However, some improvements are suggested: Sec. 2: It may be useful to indicate the limitation of the assay in saying that "it was not demonstrated that other mode of actions such as antiandrogenic activity could be detected". Sec.15:A further acceptance criteria should be "absence of signs of general toxicity, such as unspecific morphological changes, behaviour, food uptake or mortality" Sec. 28: The weight of the fish in the test is not relevant for the determination of the endpoints, but the state of maturity and the appearance of secondary sex characteristics. Those are sufficient to determine a valid test population. Since fathead minnow males and females differ in weight, a sex dependent subgroup would have to be weighed separately, which is unnecessary for the validity of the test. Sec.40,41: Changes in those endpoints should principally invalidate the test (see sec.15). Sec.44: It is not clear, whether the observation on tubercles in fathead minnow should be used qualitatively or quantitatively. The results of Phase 1 a and b as well as 2 are inconclusive in this endpoint, since some laboratories were not able to quantify size of tubercles and variation was large. Therefore, I suggest the mandatory observation is appearance and number of tubercles, and size is optional endpoint. Annex 3, 23: Gonadal histology should be marked as an optional endpoint in accordance with the main test guideline.
6	Criteria 2 is PARTIAL LY MET	The original intent of this method was to include five core endpoints which included plasma vitellogenin expression, alteration of secondary sex characteristics, fecundity, gonado-somatic index and gonad histopathology. After subsequent evaluations it was found only two apical endpoints, vitellogenin expression for detection of estrogenic activity and aromatase inhibition, and secondary sex characteristics for detection of androgenic activity were maintained in the final test suite. The biological relevance of plasma vitellogenin expression in the blood of male fish exposed to estrogenic compounds and the changes in secondary sex characteristic in females exposed to androgens is well described under the protocol section, Initial Considerations and Limitations. However, the discussion concerning the use of zebrafish in part 10 of this section is confusing. It is unclear if the entire paragraph or just the first statement is referring to zebrafish. I would recommend that statements in this section be specific for fathead minnow and medaka with exclusion of the zebrafish. Issues, comments and recommendations concerning the use of zebrafish in this assay will be covered below under Section B, 'Specific Comments Concerning the 21-day Fish Screening Assay'.

B. Specific Comments Concerning the 21-day Fish Screening Assay

Background

Three fish species, the fathead minnow, Japanese medaka and the zebrafish were evaluated for use in a fish screening assay for endocrine active substances. As stated previously, the original intent of this method was to include five core endpoints in this assay including plasma vitellogenin expression, alteration of secondary sex characteristics, fecundity, gonado-somatic index (GSI) and gonad histopathology. The low sensitivity of GSI to strong estrogens and androgen agonist in Phase 1A lead to its removal from consideration as an endpoint. Findings from Phase 2 and 3 studies indicated endpoints such as gonadal histopathology and fecundity were susceptible to general systemic toxicity and could produce false positives unrelated to endocrine-specific effects and were thus dropped from further consideration. The final draft protocol retained measures of vitellogenin levels (blood/plasma in fathead minnow and zebrafish; liver homogenate for medaka) and secondary sex characteristics (SSC) in fathead and medaka as indicators of exposure to estrogenic, aromatase inhibiting and androgenic chemicals.

Note on flutamide studies

As outlined in the Phase 1B document as well as in the literature, flutamide appears to cause weak, or perhaps no antiandrogenic activity in most species of fish tested. Because of the inconsistent results obtained with flutamide no conclusion can be drawn from the exposures conducted in Phase 1B. Therefore, it is recommended this test protocol not be indorsed as a screen for anti-androgens until further evaluations can be conducted.

Tabular Synopsis of Results

To assist with the response assessment for each species and endpoint the following tables containing a synopsis of results were generated.

Species: Fathead minnow

MOA	Chemical	VTG		SSC	
		Male	Female	Male	Female
Estrogenic	estradiol	SI	SI	NS	NS
	4-TPP	SI	SI	SD	NS
	octylphenol	SI	SI	SD	NS
Aromatase inhibitor	prochloraz	NS	SD	SD	NS
Androgenic (non-arom)	trenbolone	NS	SD	NS	NS*
Androgenic (aromatizable)	androstenedione	SI	SI	NS	SI
Anti- androgen	flutamide	NS	NS	NS	NS

Negative control	K permanganate	NS	NS	TD	NS
	n-octynol	NS	NS	NS	NS
	pcp	NS	NS	NS	NS
	methoxyethanol	NS	TI	NS	NS

SI= statistically significant increase; SD= statistically significant decrease
NS= not statistically different from control; TD= decrease due to intoxication
TI= increase due to intoxication; --- = not tested
*significant difference not found, however nuptial tubercles appeared in several studies.

Species: Medaka

MOA	Chemical	VTG		SSC	
		Male	Female	Male	Female
Estrogenic	estradiol	SI	SI	NS	NS
	4-TPP	SI	SI	NS	NS
	octylphenol	---	---	---	---
Aromatase inhibitor	prochloraz	NS	SD	NS	NS
Androgenic (non-arom)	trenbolone	NS	SD	NS	SI
Androgenic (aromatizable)	androstenedione	---	---	---	---
Anti- androgen	flutamide	NS	NS	NS	NS
Negative control	K permanganate	NS	NS	NS	NS
	n-octynol	NS	NS	NS	NS
	pcp	---	---	---	---
	methoxyethanol	---	---	---	---

SI= statistically significant increase
SD= statistically significant decrease
TD= decrease due to intoxication
TI= increase due to intoxication
NS= not statistically different from control
--- = not tested

Species: Zebrafish

MOA	Chemical	VTG	
		Male	Female
Estrogenic	estradiol	SI*	SI
	4-TPP	SI	NS
	octylphenol	---	---
Aromatase inhibitor	prochloraz	NS	NS
Androgenic (non-arom)	trenbolone	NS	SD
Androgenic (aromatizable)	androstenedione	---	---
Anti- androgen	flutamide	NS	NS
Negative controls	K permanganate	NS	NS
	n-octynol	NS	NS
	pcp	---	---
	methoxyethanol	---	---

SI= statistically significant increase
SD= statistically significant decrease
NS= not statistically different from control
TD= decrease due to intoxication
TI= increase due to intoxication
--- = not tested
*Responses appeared stronger at day 14 than at day 21 of exposure.

Fathead minnow
The most complete results from the validation studies (Phase 1A, 1B, 2 & 3) were provided for the fathead minnow. Vitellogenin expression was consistent in male fish exposed to estradiol and estrogen agonists and performed as expected in female fish exposed to a non-aromatizable androgenic chemical and an aromatase inhibitor. Not surprisingly, the aromatizable androgen androstenedione caused significant elevation of vitellogenin in both male and female fatheads. In general, exposures with negative control substances were favorable but care must be taken in interpretation of significant decreases in vitellogenin levels of exposed females where symptoms of intoxication are present and may lead to false positives.
Induction of nuptial tubercles in females fathead minnows exposed to trenbolone was not statistically significant although some tubercles did appear at the highest concentration tested in two of the three participating laboratories. Although later tests with androstenedione did result in statistically significant increases in nuptial tubercles in female fathead minnows the conflicting

results with two androgens (an aromatizable and non- aromatizable) raises concern for the robustness and reliability of the endpoint and suggest further testing with a second non-aromatizable androgen (i.e.,dihydrotestosterone or methyldihydrotestosterone) may be advisable. Exposure to toxic levels of potassium permanganate caused a decrease of tubercles in male fish not associated with an endocrine mode of action. These results again demonstrate that caution must be taken with the interpretation of results obtained at exposure levels where intoxication and systemic toxicity occur.

Medaka

Expression of vitellogenin in the livers of medaka exposed to the model endocrine compounds and negative control substances mirrored the results observed with fathead minnows. Unfortunately, octylphenol and androstenedione were not tested in medaka which would have provided a greater measure of the reliability of the test method. Significant increases in papillary processes were observed in female medaka exposed to trenbolone, but SSC responses in male medaka to estrogen agonists and an aromatase inhibitor demonstrated poor performance. Decreases in the number of nuptial tubercles in male fathead minnow appear to be a more sensitive measure of weak estrogens than was observed for papillary processes in male medaka.

Limitations Concerning Zebrafish

Of the three fish species under consideration in this protocol, the zebrafish is the only member which is not sexually dimorphic. The adults are difficult to sex which may lead to misidentification during loading of individuals into test concentrations and misinterpretation of results. With secondary sex characteristics lacking in this species, plasma vitellogenin remains as the sole determinant endpoint. Of concern are results observed with estradiol (Phase 1A) where vitellogenin expression in male fish was pronounced at day 14 but significantly less on day 21. In contrast good results were obtained in exposures with the weak estrogen agonist, 4-tert-pentylphenol (Phase 1B). This raises the question of robustness and transferability of vitellogenin measures in this species. This evaluation would have benefited from the additional information provided by an exposure with octylphenol. Further, investigations with androgen agonists were not sufficiently represented with only the non-aromatizable androgen agonist, trenbolone tested. Taken together, the information above suggests that zebrafish may not be an optimal test species for inclusion in the protocol.

Summation and Recommendations:

Differential expression of vitellogenin in male and female fish combined with appearance of SSC in female fish appears to be the most reproducible and readily transferrable components of the protocol. Using these three endpoints the following can be determined:

MOA	Endpoints
1. Estrogen agonists	Significant increases in male VTG coupled with no change in female SSC (Secondarily, increases in female VTG could be monitored)
2. Aromatase Inhibitors	Significant decreases in female VTG coupled with no change in female SSC
3. Androgen agonists	Significant decreases in female VTG coupled with significant

		<table><tr><td>(non-aromatizable)</td><td>increases in female SSC</td></tr><tr><td>4. Androgen agonists (aromatizable)</td><td>Significant increases in male and female VTG coupled with significant increase in female SSC</td></tr></table> Significant decreases in nuptial tubercles observed in male fathead minnows exposed to estrogen agonists were determined to be consistent with acceptable levels of reproducibility. However, the method remains difficult and time consuming to perform. Further, the SSC endpoint in male medaka performed poorly in response to estrogen agonists. Therefore, for the detection of estrogen agonist, male SSC scores appear to provide no real advantage over measuring vitellogenin levels in male fish. Perhaps of greatest concern is the economic feasibility of applying this protocol in the context of a "screen' for endocrine active substances. As currently described the protocol requires performing a 21 day flow-through test with 2 to 4 replicates for each of 3 treatments plus controls which clearly exceeds the definition of a screening assay. Many of the original endpoints (fecundity, GSI, gonad histopathology) which required an extended period of exposure to manifest significant changes have now been eliminated from the protocol. It is understandable but unfortunate that information on vitellogenin levels and SSC developed within a validation framework is not available for exposures of 14 days or less with weak estrogen agonists. Reduction in the length of exposure would make the protocol more practicable, affordable and thus available to a larger number of laboratories. It is doubtful the protocol as currently written will be widely used as a screening tool. Based on the information provided the panel may wish to consider the following suggested changes to the existing protocol: 1. Eliminating zebrafish from the protocol. The protocol could suggest zebrafish as an alternate species for vitellogenin analysis but no detailed instructions on their use would be provided. Rationale: Difficult to distinguish sexes; no SSC; untreated males appear to have extreme variability and higher background levels of vitellogenin. 2. Reducing the significant endpoints for analysis to: a) vitellogenin measures in male and female fish; b) SSC in female fish only. Comparisons between these three endpoints will provide the information necessary to determine and classify a chemical with estrogenic, androgenic or aromatase inhibiting properties. 3. Reducing the duration of the test from 21-days to 14 days or possibly 7 days if supported by information provided through other available studies, the peer-reviewed literature or by conducting new studies. 4. Consider reducing or eliminating replicate treatments. Since female fathead minnows and female medaka normally do not display SSC the semi-quantitative measure of nuptial tubercles and papillary processes in these species would be an adequate screen for androgenic activity requiring no replicate treatments. Measures of vitellogenin levels typically demonstrate high variability but are quite sensitive and provide a robust estrogen-responsive biomarker. With fecundity dropped from consideration there is no longer a need to maintain separate spawning groups. Equal numbers of male and female fish (e.g., 20 fish per sex) can be pooled into a single tank for each concentration and control exposure. Statistical analysis can be conducted on the basis of individuals rather than the unit of replication greatly simplifying the execution and interpretation of the assay.

Criterion 3) A detailed protocol for the test method should be available.

Reviewer	Answer	Explanation
1	Yes, with qualifications.	(A) Test method conditions. The husbandry, exposure conditions and quality control measures for the assay protocols were generally well described. During phases 1B, 2 and 3 of the validation effort, the fathead minnow protocol was changed to reduce the number of males in the test vessels from 5 to 2, and females from 5 to 4; and the number of test vessel replicates was increased from 2 to 4. Valid rationale for this change was presented. However, the description of the experimental conditions for Phase 3 protocols (Phase 3, Appendix 1), as well as the Preliminary Draft Test Guideline (Annex 3) still indicate that 5 males and 5 females of fathead minnow are to be used per test vessel, with two duplicates. (B) Endpoint measurements. The method for Vtg purification for the purpose of providing calibration standard for the Vtg assays is well explained (Phase 3 Appendix 7). However, the method (SOP) for biological sample (blood, liver) collection and processing for Vtg analysis does not appear to have been included in the documents available for review. This information appears to have been provided to the laboratories participating in the studies according to statements made in p. 31 and 32 of the Phase 1B report, but the referenced appendix (Annex 6) and tables (Table 6) are either not found in the publicly available report or do not contain the information mentioned. Thus, this reviewer is unable to determine if the protocol for Vtg measurement provided to the participating laboratories had the necessary detail. (C) Acceptability criteria. Baseline levels for Vtg in males and females are not yet well defined. In females, baseline levels are likely to vary with the stage of the reproductive cycle in individual fish; thus, normal baseline levels could vary widely. In males, the present studies seem to have found inconsistent baseline values which, perhaps, may at least be partly due to insufficient knowledge of the role of endogenous androgens in determining the response to xenoestrogens (see answer to Question 2), or to unknown environmental/husbandry variables.
2	provisionally met at this stage	By phase 3, the protocol is relatively well developed, so critera 3 can be considered to be provisionally met at this stage, although four caveats must be addressed for this criterion to be fully met: 1. There appear to be inconsistencies in the description of the protocol for phase 3, concerning the numbers of fish used per test, shown in appendix 1, compared to page 8 of the phase 3 report. The overall design and protocol states that for zebra fish 5 males and 5 females were used in 2 replicate tanks, but this differed for fathead minnow where 4 X 2 males and 4 females per replicates (in total 8 males and 16 females) were used. In appendix 1 it states that 5 males and 5 females were used for all species (including medaka). This needs to be clarified. 2. In addition the acceptability criteria for the endpoints, particularly the main endpoint VTG in control males and females need to be defined. This should have been included in the peer review package. 3. Table 1 given in the secretariat document (p6) concerning the power to detect a significant effect at the highest dose tested should be conducted also for 5 fish (as given for medaka and zebrafish), as the power is not necessarily linear, and so cannot be extrapolated from reading this table. 4. It would also be appropriate to include an animal consumption curve so that it is evident if one more animal is added how

		much power is subsequently added to the results, such that a 3 R's statistically appropriate decision can be taken on the replicate numbers.
3	Partially met	The protocol is well described but some questions/specification remain. Solvent. It is indicated that the use of solvent should be avoid. If solvent is required, information concerning the nature of the solvent and its final concentration in aquaria could be indicated (It can be of interest given that some studies report endocrine effect of certain solvent) Range of tested concentrations. In the protocol, it is recommended to test 3 concentrations but there is no clear indication concerning the range of concentrations to test. A dilution step of 10 can be adapted for a preliminary evaluation of a chemical. In a second step, to confirm the activity of a chemical, a lesser dilution step (e.g. 1/3) could be more adapted to refine the dose-response curve. Blood sampling and measurement of SCC: The protocol should include clear and detailed procedures especially for 1) blood sampling (FHM and ZF) and 2) measurement of SCC in the MDK. Theses procedures are not provide. Appearance and observations of SSC (p7). It is indicated that physical appearance of the fish should be made over the course of the test and at conclusion of the study. I am not familiar with the FHM and the MDK. Does it means that fish have to be removed from their aquaria to record changes in physical appearance? Maybe these observations can stress the fish if they have to be manipulated. Acceptability criteria: There is a lack of acceptability criteria regarding vitellogenin levels in male and female fish (e.g., is the test rejected if one or more individuals present "abnormal" Vtg concentrations in control groups?). During the exposure period, spawning in the control vessel is recorded to check that fish are in spawning condition. In the data reporting, it is not specify to report the spawning status of control fish. Does the spawning status of fish is a criterion of acceptability? Is the test rejected if a "low" or "abnormal" spawning activity is recorded in control fish? P5. Paragraph 25. Positive control. It is indicated that a positive control should be run regularly to ensure that the test system is performing correctly. The meaning of this sentence could be precise. It is recommended to test E2 at 100 µg/L which is a very high concentration. The frequency to test positive control should be precise. What is (are) the criterion (criteria) of acceptability for the positive substance? Is it based on Vtg measurement? One should consider that a significant induction of Vtg in male fish exposed to E2 is not enough to ensure that the test system is performing correctly. Rather a fold induction rate above control fish in agreement with "historical" data could be used. This remark reinforces the need of acceptability criteria for vitellogenin levels (in control and fish exposed to reference substances). Number of replicates: The number of 2 replicates for the ZF and MDK protocol and 4 for the FHM appears as appropriate to detect estrogenic compounds through measurement of Vtg induction. The higher number of replicates (and the higher number of female fish) in the FHM protocol appears more appropriate to detect significant decrease of Vtg in female as compared to the ZF and MDK protocol. The rationale for using 2 replicates in the ZF and MDK protocol is not clearly justify regarding Vtg analysis in female.

4	Partially met	Genetic background of the fish used for the assays should be defined. No clear description of the genetic background of the fish material was recorded in any of the studies. For example, the orange-red variety of medaka is mainly used in those laboratories that used medaka for the Phase 1B assays. While this variety can easily be obtained in pet shops in Japan, the origin and the genetic background of the variety are not defined. In addition, although another medaka strain was used in the assays, no description of the strain appeared in the report (Phase 1B, 4.3 Test fish). This is important because, as an organism, fish have complex systems and in vivo studies using fish stocks with different genetic backgrounds often exhibit different results from each other. Fish strains with defined genetic backgrounds, ideally inbred strains, should be used in the assays to obtain reproducible and reliable results. In medaka, several inbred strains have already been established, but most of these strains have been established without consideration for their use in endocrine screening assays and it is therefore unknown whether they are suitable for endocrine screening or not. Development of strains with a defined genetic background and which are suitable for the assays is necessary to improve the assay system. Protocol for breeding and handling of the fish material should be established for each fish species. The breeding conditions employed for the fish material are not described in detail in any of the reports. Breeding conditions affect fish in a variety of ways, including growth, sexual maturation, health, and the number and quality of eggs. Assays using fish exhibiting differences in these aspects will result in differences in the experimental results and consequently difficulties in the reproducibility and reliability of the work. For example, in the Phase 2 report, a high level of variability in the degree of maturation was observed among control females. Reproducibility and reliability cannot be expected in assays employing such materials.
5	Criteria 3 is met.	A detailed protocol was used for the different validation phases and is reflected by the draft guideline.
6	Criteria 3 is PARTIALLY MET	The peer-review package contained a preliminary draft of the test guideline listed as item number 6. This draft lacked Annex 4-6 containing information concerning spawning substrate (Annex 4), characteristics of acceptable water (Annex 5), and procedures for SSC measurement (Annex 6). Information concerning characteristics of acceptable water and procedures for SSC measurement in the fathead minnow were located in the Phase-3 OECD report (item #1 of the peer-review package). However, no information was provided on the following procedures: 1) collection, preparation and measurement of vitellogenin in medaka using liver tissue; 2) scoring of SSC in the medaka. For a comprehensive review of the 21-day assay to be adequately conducted a complete draft protocol should have been provided. Therefore, my comments are based on the incomplete preliminary draft guideline provided as item number 6 in the peer-review package, the Secretariat paper (item number 7) and from sections provided in the five reports.

Criterion 4) The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.

Reviewer	Answer	Explanation

1	Yes.	The issue of reproducibility of the test method was specifically addressed during Phase 1b of the validation process. It appears that the reproducibility of SCC results was generally higher than for Vtg results. Also, as indicated by measurements of Coefficient of Variation (CV), the reproducibility of the Vtg measurements was lower for zebrafish than for medaka and fathead minnow. However, it should be noted that the method for calculating the CV included biological variability (variability in individual fish responses); thus, some degree of test method variability is expected. The important thing to note is that observations based on the response of the test system to test compounds generally yielded consistent (reproducible) conclusions; namely, the relative effects (increase/decrease/no change) in endpoint values (Vtg and SCC) caused by the test compounds were consistent across laboratories.
2	Partially met for phase 3	A general criticism is that overall many laboratories were involved, testing few compounds. More acceptable validation study design is to have fewer laboratories, testing a large number of compounds. This criterion is partially met for phase 3, but not met for the other phases for the following reasons: Intra-laboratory variability The study design is such that for each laboratory two tanks of the same dose chemical are compared to assess intra-laboratory variability. However this was not done, instead the fish were compared on an individual basis. Under these circumstances the Secretariat state that 'It is uncertain how this can be analysed further'. This is therefore a further failing in the validation exercise, and as such, this criterion is not fully met. Is it possible to go back and reanalyse these data on a vessel comparative basis? Inter-laboratory variability For phase 3 this criteria is demonstrated for 4 out of 5 laboratories. One laboratory reported VTG baselines in males below the ELISA detection limit. In such a case acceptable practice is to note this and omit all data from this laboratory, and conduct the statistical analyses on the remaining laboratories. However this report indicates that a most extraordinary and completely unacceptable step was taken. Page 6: 'A random set of values was generated with a distribution equivalent to other groups and with a mean at one half of the detection limit.' So false data was generated similar to that obtained from the other laboratories! Under the circumstances the data should be recalculated and statistically analysed, then presented in a revised report expressly excluding this false data. Annex 2 in the phase 2 report requires clarification. The plots are unclear and inadequately described. Please provide better quality graphics for these plots together with a clear description of what they are saying.
3	Partially met	I am not competent to review in detail the intra- and inter-lab statistics analyses.The inter-lab studiesis appropriate and sufficient to establish reproducibility of the protocol. The inter-lab variability in vitellogenin responses is high but acceptable regarding the biological marker. It is surprising to observe high variations in the basal concentrations of Vtg in fish from a laboratory to another (in particular in zebrafish phase 1A). However, in phase 3, Vtg analysis has been made under blind observations and results were satisfactory. In practical terms, the protocol transferability and reliability has been shown.
4	Partially met	The reproducibility of results, particularly controls for inter-laboratory reproducibility of results, was not adequately established in the assays, and this is one issue that needs to be resolved in improvements of the guidelines and the robustness of the assay.

		While evaluations of fecundity and gonad histology were included in the earlier stages of the validation studies, reproducibility was not established for these endpoints. Consequently, the recommended endpoints of the assay are restricted to only the vitellogenin level in the blood or liver, and the development of secondary sexual characteristics in the Phase-3 assay. However, even with these restricted endpoints, failures with respect to inter-laboratory reproducibility were observed. One of the main reasons for the decreased reproducibility in endpoints such as fecundity and gonad histology is considered to be the non-establishment of detailed protocols regarding the genetic background of the fish used in the assays, as well as the breeding and handling methods used for testing. This issue should be improved and I will elaborate further on these points below. Please refer to the comments in other criteria.
5	Criteria 4 is met with one exception	Some laboratories were not able to quantify size of tubercles and variation was large. Therefore, I suggest the mandatory observation is appearance and number of tubercles, and size is optional endpoint (see above).
6	Criteria 4 is PARTIAL LY MET	As was clearly stated in the Secretariat paper, the intra-laboratory reproducibility of the two remaining core endpoints, vitellogenin levels and SSC can only be based on a limited amount of information. Of the species tested, fathead minnow had the most complete dataset. Only one androgen, the non-aromatizable agonist, trenbolone was tested with medaka and zebrafish which is troubling. Therefore, the test method was fully validated only for fathead minnow and subsequent exposures should be undertaken to demonstrate the reproducibility of the SSC endpoint to androgen agonists in female medaka.

Criterona 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.

Reviewer	Answer	Explanation
1	Yes.	The chemicals tested during the course of the 3 phases of the validation process included potent and weak estrogens (including difficult weak estrogen), aromatase inhibitors, antiandrogens, androgens, weak aromatisable androgen, and difficult negative substance. The results obtained with the various test compounds showed that the endpoints measured (Vtg, SCC) responded as expected in regards to test compounds' modes of action.
2	No	Each study phase of the 21 day fish screening assay has attempted to improve the optimisation the assay, based upon the findings at each phase. Overall a limited number of compounds were tested, on the basis of modes of action, not chemical class, (11, including 2 positive controls, and not including one chemical tested in 1 laboratory), but the range of chemicals tested was progressively improved such that by phase 3, 2 weaker positives and a more challenging negative had been tested. It is acknowledged that as this is an in vivo study a smaller range of chemicals for testing is appropriate for animal welfare reasons. However the excessive number of laboratories involved, especially in phase 1B, is unnecessary. For animal welfare reasons, but appropriate as this test method is an in vivo study, phase 3 created in response to phase 2 negative

		control problems, and lack of good but potentially problematic negatives, and no weak estrogens tested in phases 1 and 2. Phase 1A: 2 strong compounds: estrogenic: E2, androgenic: 17 beta trenbolone at 3 concentrations Phase 1 B: 3 potent positive compounds: fadrazole, prochloraz and flutamide, 1 weakly acting compound: t-pentylphenol. 2 methoxyethanol tested in one laboratory only. Phase 2: 2 negatives potassium permangate (oxidizer), and n-octylphenol (non polar narcosis) Phase 3: 3 compounds: weak estrogen (octylphenol) weak aromatisable androgen (androstenedione) and one challenging negative (sodium pentachlorophenol). The fact that the protocol was progressively changed to address the problems encountered originally means that the phases cannot be compared adequately, and that this test has followed a series of optimisation steps, but is not yet validated.
3	Partially met	The selected chemicals represented various substances having different mode of action including reference compounds (E2, trenbolone, Flutamide), weak estrogenic compounds (TTP, OP) as well as non active substances. There is a clear rationale for their selection. The selection looks suitable for a validation purpose. Different remarks/comments can be made - weak estrogens: other weak estrogenic compounds could have been tested. For instance, weak estrogenic compounds that have been tested are belonging to the same chemical family and therefore not representative of other chemicals for which the assay will be used. Substances that elicit estrogenic effect after being metabolised into estrogenic metabolites (e.g. methoxychlor) could have been tested notably in the phase 3. - AR agonist: Only one very potent androgen agonist was used and tested during phase 1A and it would have been relevant to assess other AR agonist compounds to determine whether decrease of Vtg concentration in female constitute a "typical" biological response. - Aromatase inhibitors: Only prochloraz has been tested in the three fish species. An other "aromatase inhibitor" should have been tested to better demonstrate the usefulness of the Vtg measurement in female and the robustness of the test design (e.g., number of replicate, number of female fish) used to detect compounds acting via this mode of action. ER Antagonist compounds has not been tested
4	No comment	No comment
5	Criteria 5 is met	A positive control was successfully employed, reference chemicals with different mode of actions were used.
6	Criteria 5 is PARTIALLY MET	Reference chemicals used for the method's performance appear adequate to evaluate chemicals with estrogenic and aromatase inhibiting properties. I have some concern regarding androgens, especially, aromatizable androgens and how their effects could challenge the interpretation of the results obtained for vitellogenin and SSC. It is unclear if this method is to be applied only for screening of pure chemicals or if it could be applied for screening of sewerage effluents. If the latter is true, sewerage may contain aromatizable androgens in high abundance from human and pharmaceutical sources thus more research may be warranted to characterize their effects.

Criterion 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.

Reviewer	Answer	Explanation
1	Yes, with qualification.	General information that is relevant to assess the performance of the test method and that was available prior to undertaking the present validation studies include: (1) data from a wide variety of sources that allowed classification of the modes of action for the test chemicals; and (2) data derived from the general fish reproductive biology literature, as well as information specific to the species of concern (medaka, fathead minnow, zebrafish), that allowed prediction of how the endpoints measured should respond to the exposure to test chemicals (based on their modes of action). In the context of this previously available information, it can be concluded that the test method "performed" as expected. Moreover, Table 4 in the document, Application of Guidance Document 34 Criteria to the Validation of the 21-Day Fish Endocrine Screening Assay, provides a side-by-side comparison of the present test method results with partial life cycle and full life cycle test results conducted by a laboratory in Japan (the species was not mentioned in the table) that seem to indicate that the 21-day assay "performs" consistently with the other assays.
2	No met in small part	This test is new, and not designed to replace an existing test guideline. However there are other relevant peer reviewed toxicity testing data where comparative analyses would be relevant and useful. The Secretariat give the example of the key findings of a Japanese study (Table 4) Secretariat paper April 2007. There are others e.g. In a mixtures study fathead minnow were found to correlate well with in vitro and (later) mammalian mixture studies. Environ Health Perspect. 2005 Jun;113(6):721-8.Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals. Brian JV, Harris CA, Scholze M, Backhaus T, Booy P, Lamoree M, Pojana G, Jonkers N, Runnalls T, Bonfà A, Marcomini A, Sumpter JP. Envir on Sci Technol. 2006 Sep 1;40(17):5478-89.Modeling effects of mixtures of endocrine disrupting chemicals at the river catchment scale. Sumpter JP, Johnson AC, Williams RJ, Kortenkamp A, Scholze M. It would appear that it is not possible to fully meet this criterion, we agree with the view expressed by the Secretariat, that the term 'performance' is not (exactly) appropriate, however as part of a retrospective analysis for performance criteria, it might be possible to include this additional data. Table 4, giving an overview of the same chemicals tested by the Japan Ministry of the Environment, the species (presumably Medaka) should be specified.
3	Partly met	It is difficult to answer this question. Based on data provided by Japanese (Table 4 of the GD 34 document) it appears that the data showed a satisfactory concordance and a comparative sensitivity between the Fish Screening assay for EAS and the partial and full life cycle for estrogenic compounds.
4	Partially	Age of fish material

	met	In my experience of breeding of medaka, they reach maturity at 8 weeks post hatching and full maturity at 12 weeks post hatching. However, the protocol used in the assays recommends that 16-week old medaka be used for the assay, and fish older than 16-week old are also used. For example, in Phase 1B, medaka aged 16 to 24 weeks were used as test animals. While 24-week old medaka spawn eggs, the quality of the eggs is not comparable to that which can be obtained from 12 week-old medaka; young, fully mature fish consistently produce eggs of good quality. Reproductive activity decreases and becomes erratic as fish age. Fully mature and sufficiently young test fish should be used in the assays.
5	Criteria 6 is met	Literature information on fish and mammalian toxicity of the relevant substances were used to develop the test protocols.
6	Criteria 6 is PARTIAL LY MET	Evaluation of the method for vitellogenin expression and for SSC in the fathead minnow for identification of estrogenic, aromatase inhibiting and non-aromatizable androgenic chemicals appears to be adequate based on the information supplied by the OECD validation studies and from information available in the literature. For medaka and zebrafish, vitellogenin and SSC (medaka only) endpoints have been evaluated to a lesser extent particularly in regards to negative test substances where intoxication may affect their response profiles.

Criterion 7) Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.

Reviewer	Answer	Explanation
1	Yes, with qualificatio n.	This criterion is difficult to adequately evaluate. Namely, if it is simply "ideal" to collect data in accordance to principles of GLP, data collected without the guidance of GLP would still meet the criterion because strict accordance to GLP is not a requirement. However, the document, Application of Guidance Document 34 Criteria to the Validation of the 21-Day Fish Endocrine Screening Assay, further states that "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." In this context, most participating laboratories appear to have conducted their studies according to GLP principles but not many were formally certified as GLP studies.
2	No comment	That many laboratories were GLP compliant but not GLP certified is not considered a major problem.
3	No comment	It is indicated that inter-laboratory validation was conducted using GLP guidelines for certain laboratories.
4	No comment	No comment
5	No comment	It is not clear from the reports of the different phases of the validation work, whether all studies were perfomed under GLP.
6	Criteria 7 is MET	It appears that Phase 1A, 1B and Phase 2 studies were conducted in compliance with GLP although most participating laboratories were not GLP certified. Phase 3 had no indication of GLP compliance but stated the studies met the test acceptance criteria for test

		procedures and parameters. Departures from protocol and deviations observed in the results were reported providing reviewers an indication of data quality. Overall, the principles of GLP appear to have been met.

Criterion 8) All data supporting the assessment of the validity of the test method should be available for expert review.

Reviewer	Answer	Explanation
1	Yes, with qualification	The reports for Phase 1a, 1b, 2 and 3 are declassified and available for expert review and in the public domain. These reports contain the vast majority of the information that would be required for an independent laboratory to follow the procedures and generate equivalent data. However, there are a few instances where additional critical information would be useful; for example, detailed information on procedures to collect and process biological samples for Vtg determination was not available for review (see answer to Question 3).
2	Almost met	This criteria is almost met, all data is made available, although it varies greatly in quality, graphical quality and legibility, and reliability. All data provided should be legible. Appendix 2, phase 2 plots are illegible.
3	Partly met	The data presented allow the evaluation of the validity of the test Draft test guideline could be more detailed to permit others to follow it and generate equivalent data.
4	No comments	No comments
5	Criteria 8 is met	All data of the different tests performed in course of the validation procedures were available. A few minor omissions in the reports are indicated in the attachment "Peer review, OECD validation of the 21 day fish screening assay and preliminary draft guideline".
6	Criteria 8 is PARTIALLY MET	A detailed test method protocol containing annexes with necessary supplemental information pertinent to the proper execution of the protocol was not provided for expert review. The preliminary draft protocol (item number 6) lacked annex sections. Further comments on the protocol will be provided below under 'Specific Comments Concerning the 21-day Fish Screening Assay'

Annex 2

General comments from peer reviewers are summarized as follows

Reviewer 2
The scientific basis, regulatory purpose and need for the test are evident. This assay is a highly promising test method that once validated, has the potential to screen for endocrine active substances in the aquatic environment. ECVAM agrees and supports its eventual use, once validated, as part of a weight of evidence approach utilizing other sources of information (e.g. physico-chemical properties, (Q)SAR, in vitro data, in vivo mammalian data, scientific literature etc.). However on reviewing the peer review documentation, particularly the Phase 1A, 1B, 2 and 3 reports, it is evident that this test method has been undergoing a series of test optimisation steps, as one would require at the pre validation stage, or even before, and that as the process of validation is incorrect and incomplete, this assay is not yet validated. Only following phase 2 and phase 3 does the protocol appear to be optimised and standardised, such that it is performing well, at phase 3 some coding is in place, but at this phase only three compounds were tested, although they were more challenging and included negatives and weak compounds. Phase 3 is further handicapped by the highly irregular and unscientific statistical analysis for one laboratory. ECVAM therefore recommends that ideally this test method should enter validation according to GD 34, ECVAM, ICCVAM and JaCVAM validation requirements, with the optimised protocol in three laboratories for the three test species, at three concentrations for each test substance, with a minimum of 10 test substances addressing estrogenic, androgenic and aromatase mechanisms. However given the enormous resources and animals required for this, compromises are necessary. So far there has already been excessive and unnecessary animal use, in phase 2 for example 14 laboratories were involved in this optimisation process, showing a lack of consideration of the 3R's. In addition, the 3 R's should be specifically addressed prior to undertaking the validation by consideration of the appropriate power calculations and an animal consumption curve (see question 3 below). A retrospective validation approach could be taken, to support the proposed validation. This should include a thorough critical background review document identifying the protocol differences between the studies conducted here with the optimised protocol derived during this optimisation process, the protocol differences with those published in the recent literature and in peer reviewed databases (e.g. those of the Japanese databases on medaka), such that a robust meta analysis of such data is provided. With respect to predictive capacity, this data could then be assessed in a similar fashion to the retrospective validation for the micronucleus test. Such a retrospective procedure was endorsed by ESAC in November 2006. Important considerations Test Species and reduction: Due to male territoriality, fathead minnow require greater numbers of female to male ratios to yield reliable and robust measurement of reproduction (page 9 phase 2 report), i.e. 16 fish as opposed to 10 fish per group for medaka and zebra fish. With respect to the 3R's, particularly reduction, given that the data generated for VTG induction is comparable between species, it should be a requirement that the species requiring less numbers per group should always be used, where this endpoint and species sensitivity is deemed sufficient. Where secondary sex characteristics such as nuptial tubercles (adult male fathead minnow) and papillary processes (adult male medaka) can be induced in female fathead minnow and medaka on exposure to androgen receptor agonists for example, medaka would be the species of choice, as zebrafish do not have such secondary sex characteristics. Low dose effects: Page 112, paragraph 233; regarding cost efficiency, it is suggested that pathologists initially examine the high dose and negative control groups, and suggest that if no effects are evident, then the mid and low dose groups may not need to be assessed. This is unacceptable. Low dose effects are a real concern and are not necessarily seen at higher doses. All dose groups must be assessed. Pathology: Before meeting together at the Heidelberg meeting, the pathologists were highly

inconsistent in their analyses. The inconsistency of their expert approach is transparently demonstrated in the documents provided. However following this meeting the pathologists recommended that formal independent statistical analyses were probably not necessary. This is incorrect, independent statistical analyses and a weight of evidence approach is essential, and therefore an absolutely necessary validation requirement. Feed: As for other in vivo test methods being developed as Test Guidelines, feed should be strictly controlled for endocrine active pollutants and estrogenic/androgenic material e.g. PCBs, pesticides and genistein/soya.
Reviewer 3
My general feeling/ comment is that an extensive work has been done for the three fish species recommended. Duirng the different phases, an important effort has been made on gonadal histology and measurment of reproductive output. These two endpoints have been removed from the final Preliminary draft test guideline. The rationale for removing these endpoints has been clearly exposed/stated and the test appears more appropriate for screening purpose.
Reviewer 4
On the whole, the work required for the validation of the 21-day fish-screening assay for the detection of endocrine active substances was well executed. In addition, the research followed the standard OECD protocol and met the eight criteria mentioned in OECD Guidance Document 34. The considerable and laborious efforts of the laboratories that performed this research are commendable.
Reviewer 5
The reviewer was impressed by the validation process, the extensive experimental work involved and the well structured reporting and reviewing process. He thanks all involved laboratories for their excellent work, the OECD secretariat for the organization of the validation and review process, and the CERI, Japan, for coordinating the peer review process. In general, all aims of the validation have been achieved and the reports support the generation of the respective test guideline. Therefore, my general remark is that the assay can successfully be used for the purpose it was designed. In the following, the final reports of the Phase 1 A and B, Phase 2 (1+2) and phase 3 of the validation of the 21-day fish screening assay and the draft guideline are commented on. The purpose of the validation was the identification by the OECD of the need for developing testing guidelines to assess the effects of potentially endocrine disrupting substances in different test systems. In the field of ecotoxicity, fish were identified to be one of the target organisms for experimental testing. In order to provide appropriate test methods for different tiers of a hazard and risk assessment, the requirement for a screening test was identified and, subsequently, a specialized task group (fish drafting group) was formed by OECD to direct and monitor the validation of a respective test method. The basis of the test method development came from assay proposals made by several institutions in the US and Europe, and after intensive discussions in the drafting group, a concept of a 21-days screening test in adult fish was determined to be probably most appropriate for the purpose. The fist step of validation ended up in the reports of the Phase 1 A and B, which are commented on first. In the first round of validation a few endpoints of the assay concept were validated between 4 laboratories in 3 fish species and with 2 reference substances. In the report of this exercise the following observation are made by the reviewer: Report Phase 1 b The report showed clearly that the results from phase 1 a with potent hormones was reproducible with weaker active compounds, such as the pentylphenol (estrogen) and procloraz (aromatase inhibitor). The study with the potent antiandrogen Flutamide did give inconclusive results. The main issue for further developments was the interlaboratory reproducibility of the VTG assay. The relatively large CVs showed that only large changes can be detected. In further studies, the methods and assays used must carefully be described and positive controls should be included (see review draft guideline).

Report Phase 2 (1): Negative test substances
In essence, the results of the assays with negative reference substances (octanol OC, potassium permanganate PP) demonstrate that, in certain limits, these compounds triggered no response in either VTG, secondary sex characteristics or histological changes. However, this study had some drawbacks, which makes a clear evaluation difficult. First, the number of fish was too small (e.g. medaka exposed to octanol, n=2) in some of the asays to draw any conclusion. Secondly, the top concentration of both, OC and PP, seemed to induce changes of observed/measured parameters possibly due to general toxicity, as discusse din sections 52ff in the report. This has to be taken into account when defining appropriate test concentrations for an unknown compound in order to avoid false positive results (see review draft guideline).
Report Phase 2 (2): Negative test substances CEFIC project
Again, in this study similar observations were made that at high concentrations of the test compounds the "endocrine specific" endpoints were affected by "negative" reference substances. This is very relevant for the selection of test concentration and for considering parameters indicating general toxicity. Fish age was also identified as a major source for invalidation of the observations.
Report Phase 3: Additional test substances
The report of this study supports the finding of the previous test programmes. In table 3 a-c and 4 a-c one had wished to see the "n" on which the means were based. Otherwise, the information as it is presented now, is incomplete.

Annex 3

Some reviewers made more practical comments on the draft preliminary test guideline. These practical comments are summarized below.

Reviewer 2
Page 1-2. Initial Considerations and limitations Include a statement on the reduction of fish numbers used, as far as scientifically and statistically feasible. Test Design Page 5 paragraph 26. There is only one endpoint for Zebrafish, not two. Page 6. paragraph 32. Evaluation for naturally estrogenic foods (e.g. genistein, daidzein, soya etc) should also be evaluated and controlled for. It is not possible to give further comment as much is missing (e.g. Annexes 4-6) from this preliminary draft.
Reviewer 5
Results, page 25, fig 3a-d: Vitellogenin (VTG) induction in males is the most prominent endpoint in exposure to the estrogenic substance (Estradiol, E2). However, the statistical significance of this endpoint is not illustrated in fig 3 a and c. Are the asterix's missing? Otherwise, the conclusion in the text section on p.25, section 36 makes no sense. I recommend to clarify the illustrations. The same holds true for figure 4a and c, page 26. Page 22, table 2: Typo in Lab 4-zebrafish 10 ng/L nominal: It must read 8.65 (instead of 86.5). Sec. 2: It may be useful to indicate the limitation of the assay in saying that "it was not demonstrated that other mode of actions such as antiandrogenic activity could be detected". Sec.14: For Fathead minnow the number of fish must be 8 males and 16 females to be in accordance with sec. 13. Sec.15: A further acceptance criteria should be "absence of signs of general toxicity, such as unspecific morphological changes, behaviour, food uptake or mortality". Sec 19f: I suggest to add a guidance phrase on selection of test concentrations. "Test concentrations should be selected on the basis of knowledge on general toxicity of the compound in fish optimally over extended periods. The selected test compounds must be below a concentration which affects survival or any other unspecific signs of toxicity. It may be selected as e.g. a fraction of the acute LC50 or on the basis of realistic environmental exposure scenarios with appropriate uncertainty factors built in." Sec: 24, 2nd sentence: The data should be statistically analyzed, if possible, in order to define… Sec. 28: The weight of the fish in the test is not relevant for the determination of the endpoints, but the state of maturity and the appearance of secondary sex characteristics. Those are sufficient to determine a valid test population. Since fathead minnow males and females differ in weight, a sex dependent subgroup would have to be weighed separately, which is unnecessary for the validity of the test. Sec.40,41: Changes in those endpoints should principally invalidate the test (see sec.15). Sec.44: It is not clear, whether the observation on tubercles in fathead minnow should be used qualitatively or quantitatively. The results of Phase 1 a and b as well as 2 are inconclusive in this endpoint, since some laboratories were not able to quantify size of tubercles and variation was large. Therefore, I suggest the mandatory observation is appearance and number of tubercles, and size is optional endpoint. Annex 1: I propose to delete Annex 1 from the guideline, since it has nothing to do with the actual performance of the test method, but is part of a regulatory framework. Annex 3, 13: According to sec. 13 the number of fish per treatment in Fathead minnow must be 24 (2m,4f x 4 reps). Annex 3, 23: Gonadal histology should be marked as an optional endpoint in accordance with the main test guideline.

Annex 4

Members of the peer review panel

Panel member	Affiliation
Reviewer 1	
Reviewer 2	
Reviewer 3	
Reviewer 4	
Reviewer 5	
Reviewer 6	

Name and affiliations will be opened later.

Annex 5

Document number	Affiliation
1	Report of Eight 21-day Fish Endocrine Screening Assays With Additional Test Substances for Phase-3 of the OECD Validation Program: Studies with Octylphenol in the Fathead Minnow (Pimephales promelas) and Zebrafish (Danio rerio) and with Sodium Pentachlorophenol and Androstenedione in the Fathead Minnow (Pimephales promelas)
2	Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A)
3	Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1B)
4	Report of Three 21-day Fish Endocrine Screening Assays to Complete CEFIC's Contribution to Phase-2 of the OECD Validation Program
5	Draft Report of the Validation of the 21-Day Fish Endocrine Screening Assay for the Detection of Endocrine Active Substances (Phase 2-Testing Negative Substances)
6	Preliminary Draft Test Guideline: "The Fish Screening Assay for Endocrine Active Substances
7	Secretariat Paper: "Application of Guidance Document 34 Criteria to the Validation of the 21-Day Fish Endocrine Screening Assay

Documents to be used for the peer review are summarized below.