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

27-Jul-2011

English - Or. English

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

PEER REVIEW REPORT FOR THE VALIDATION OF THE FISH SEXUAL DEVELOPMENT TEST AND DRAFT AGREEMENT OF THE WORKING GROUP OF NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP OF THE PEER REVIEW

Series on Testing and Assessment

No. 143

JT03305532

ENV/JM/MONO(2011)24
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 143

PEER REVIEW REPORT FOR THE VALIDATION OF THE FISH SEXUAL DEVELOPMENT TEST AND DRAFT AGREEMENT OF THE WORKING GROUP OF NATIONAL CO-ORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP OF THE PEER REVIEW

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

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

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2011

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 Uterotrophic 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)*

No. 95, *Detailed Review Paper on Fish Life-Cycle Tests (2008)*

No.96, *Guidance Document on Magnitude of Pesticide Residues in Processed Commodities (2008)*

No.97, *Detailed Review Paper on the use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors (2008)*

No. 98, *Considerations Regarding Applicability of the Guidance on Transformation/Dissolution of Metals Compounds in Aqueous Media (Transformation/Dissolution Protocol) (2008)*

No. 99, *Comparison between OECD Test Guidelines and ISO Standards in the Areas of Ecotoxicology and Health Effects (2008)*

No. 100, *Report of the Second Survey on Available Omics Tools (2009)*

No. 101, *Report of the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox, 15-16 May 2008, Utrecht, the Netherlands (2009)*

No. 102, *Guidance Document for using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals (2009)*

No. 103, *Detailed Review Paper on Transgenic Rodent Mutation Assays (2009)*

No. 104, *Performance Assessment: Conparsion of 403 and CxT Protocols via Simulation and for Selected Real Data Sets (2009)*

No. 105, *Report on Biostatistical Performance Assessment of the draft TG 436 Acute Toxic Class Testing Method for Acute Inhalation Toxicity (2009)*

No. 106, *Guidance Document for Histologic Evaluation of Endocrine and Reproductive Test in Rodents (2009)*

No. 107, *Preservative treated wood to the environment for wood held in storage after treatment and for wooden commodities that are not cover and are not in contact with ground. (2009)*

No. 108, *Report of the validation of the Hershberger Bioassay (weanling model) (2009)*

No. 109, *Literature review on the 21-Day Fish Assay and the Fish Short-Term Reproduction Assay (2009)*

No. 110, *Report of the validation peer review for the weanling Hershberger Bioassay and agreement of the working of national coordinators of the test guidelines programme on the follow-up of this report (2009)*

No. 111, *Report of the Expert Consultation to Evaluate an Estrogen Receptor Binding Affinity Model for Hazard Identification (2009)*

No. 112, *The 2007 OECD List of High Production Volume Chemicals (2009)*

No. 113, *Report of The Focus Session on Current and Forthcoming Approaches for Chemical Safety and Animal Welfare (2010)*

No. 114, *Performance Assessment of Different Cytotoxic and Cytostatic Measures for the In Vitro Micronucleus Test (MNVIT): Summary of results in the collaborative trial (2010)*

No. 115, *Guidance Document on the Weanling Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti) Androgenic Properties (2009)*

No. 116, *Guidance Document on the Design and Conduct of Chronic Toxicity and Carcinogenicity Studies, Supporting TG 451, 452 and 453 (2010)*

No. 117, *Guidance Document 117 on the Current Implementation of Internal Triggers in Test Guideline 443 for an Extended One Generation Reproductive Toxicity Study, in the United States and Canada (2011)*

No. 118, *Workshop Report on OECD Countries Activities Regarding Testing, Assessment and Management of Endocrine Disrupters Part I and Part II (2010)*

No. 119, *Classification and Labelling of chemicals according to the UN Globally Harmonized System: Outcome of the Analysis of Classification of Selected Chemicals listed in Annex III of the Rotterdam Convention (2010)*

No. 120, *Part 1: Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism-based Structural Alerts for the Identification of DNA Binding Chemicals*

No. 12. *Part 2: Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism-based Structural Alerts for the Identification of DNA Binding Chemicals*

No. 121, *Detailed review paper (DRP) on Molluscs life-cycle Toxicity Testing (2010)*

No. 122, *Guidance Document on the determination of the Toxicity of a Test Chemical to the Dung Beetle Aphodius Constans (2010)*

No. 123, *Guidance Document on the Diagnosis of Endocrine-related Histopathology in Fish Gonads (2010)*

No. 124, *Guidance for the Derivation of an Acute Reference Dose (2010)*

No. 125, *Guidance Document on Histopathology for Inhalation Toxicity Studies, Supporting TG 412 (Subacute Inhalation Toxicity: 28-Day) and TG 413 (Subchronic Inhalation Toxicity: 90-Day) (2010)*

No. 126, S*hort Guidance on the Threshold approach for Acute Fish Toxicity (2010)*

No. 127, *Peer review report of the validation of the 21-day androgenised female stickleback screening assay (2010)*

No. 128, *Validation Report of the 21-day Androgenised Female Stickleback Screening Assay (2010)*

No. 129, *Guidance Document on using Cytotoxicity Tests to Estimate Starting Doses for Acute Oral Systemic Toxicity Tests*

No. 131, *Report of the Test Method Validation of Avian Acute Oral Toxicity Test (OECD test guideline 223) (2010)*

No. 132, *Report of the Multi-Laboratory Validation of the H295R Steroidogenesis Assay to Identify Modulators (2010)*

No.133, *Peer Review Report for the H295R Cell-Based Assay for Steroidogenesis (2010)*

No.134, *Report of the Validation of a Soil Bioaccumulation Test with Terrestrial Oligochaetes by an International ring test (2010)*

No.135, *Detailed Review Paper on Environmental Endocrine Disruptor Screening: The use of Estrogen and Androgen Receptor Binding and Transactivation Assays in Fish (2010)*

No. 136, *Validation Report of The Chironomid Full Life-Cycle Toxicity Test (2010)*

No. 137, *Explanatory Background Document to the OECD Test Guideline On In Vitro Skin Irritation Testing (2010)*

No. 138, *Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories (2011)*

No. 139, *Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism Based Structural Alerts for the Identification of Protein-binding Chemicals (2011)*

No. 141, *Report of the Phase 1 of the Validation of the Fish Sexual Development Test for the Detection of Endocrine Active Substances (2011)*

No. 142, *Report of the Phase 2 of the Validation of the Fish Sexual Development Test for the Detection of Endocrine Active Substances (2011)*

ABOUT THE OECD

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

The Environment, Health and Safety Division publishes free-of-charge documents in 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 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 (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 Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. UNDP is an observer. 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

E-mail: ehscont@oecd.org

FOREWORD

This document presents the peer review report for the validation of the Fish Sexual Development Test. The Fish Sexual Development Test (FSDT) covers a life-stage where sexual development is particularly sensitive to perturbation caused by endocrine active chemicals. The chemical exposure lasts for about 60 days, at the end of which endpoints of ecological relevance like the sex ratio of the exposed fish is calculated and the biomarker endpoint vitellogenin is measured in individual animals.

In 2003, Denmark, on behalf of the European Nordic countries, proposed a new project to develop a Test Guideline on the fish sexual development test to the Working Group of the National Coordinators of the Test Guidelines Programme (WNT). The project was included on the Test Guidelines work plan in 2003, and extensive validation of the test method was carried out until 2009. Two validation reports on the Fish Sexual Development Test are available in the Series on Testing and Assessment, as Nos. 141 and 142, and a draft Test Guideline was approved by the WNT at its meeting held on 12-14 April 2011. It is proposed to include the FSDT at level 4 of a draft revised Conceptual Framework for Disrupters Testing and Assessment of Endocrine Disrupters.

The peer review was managed by an independent consultant identified by the OECD Secretariat and members of the peer-review panel were proposed by the WNT. The peer-review was carried out in August-September 2010, and the report was agreed by the peer-review panel. The report was endorsed by the WNT at its meeting held on 12-14 April 2011. It is preceded by the WNT agreement on the follow-up to the peer review, including the development of a draft Test Guideline on the Fish Sexual Development Test. The Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 22 June 2011.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up of the validation peer review of the Fish Sexual Development Test

The validation peer review of the Fish Sexual Development Test was carried out in August-September 2010. The draft Test Guideline and the validation report (Phase 2) have been revised by the lead country to address the comments from the peer reviewers. On 18 October 2010, the peer review report was submitted to the Working Group of National Coordinators to the Test Guidelines Programme, the validation management group for ecotoxicity testing (VMG-eco), and the fish drafting group, for information. At the same time, the VMG-eco was requested to comment on the revised draft validation reports. The fish drafting group reviewed the revised draft validation reports and some technical issues of draft Test Guideline at its meeting held in Tokyo on 9-10 February 2011. Comments, requested from the WNT in October and December 2010, have also been addressed in the final draft Test Guideline.

Responses to the main peer review comments follow, including some changes to the validation report (phase 2) and to the draft Test Guideline:

Negative substance:
Results obtained with the testing of the negative substance ammonia have been added to the validation report (Phase 2);

Data on hatching and survival in laboratory 2 (validation phase 2)
Laboratory 2 experiments are not valid as exposure experiments due to the lack of the intended increase of exposure concentrations, but the data on behaviour and survival have been included in Table 7 and 8 of the validation report. Abnormal appearance (hereunder delayed hatch) and behaviour have been discussed in the Phase 2 validation report under "control animal performance". Embryonic development is impossible to observe in hundreds of individuals within 24-48 h before hatch;

Not every laboratory have tested all chemicals with all fish species
This is not practically possible when 4 species and 4 compounds should be tested and it frequently happens that not all participating laboratories conduct all combinations of studies when validation studies are performed. The results from the validation can however directly be compared to a published paper on zebrafish in relation to prochloraz (Kinnberg et al 2007) and indirectly compared for the estrogenic and androgenic mode of action (Holbech et al 2006);

Relationship between the test methods endpoints and the biological phenomenon of interest:
The draft Test Guideline section on initial considerations and limitations was extended to provide a more detailed description of this relationship, in particular with the addition of a table providing information on the mode of action depending on the increase/decrease of VTG and sex ratio;

Histological criteria for sex ratio:
A description of the histological criteria for validated species has been added to the draft Test Guideline section on procedure, under "Test Acceptance Criteria";

Information on sexing and staging fish gonads:
Information has been included in the draft Test Guideline section on sex determination and in Annex 7; however, it should be noted that staging is optional. Information is also available in the OECD Guidance Document on the Diagnosis of Endocrine-related Histopathology in Fish Gonads.

Variability in measured test concentration:
The actual test substance concentration was measured in all experiments. For the experiments with zebrafish, these concentrations are much below the nominal concentrations, which is not satisfactory. Especially the exposure concentrations in the Laboratory 2 experiments are outliers, and the results of these two experiments cannot be compared to the other results, when related to nominal concentrations. The endpoint responses are though strongly connected to the actual substance concentrations, which can be seen on the NOEC/LOEC values of Table 12 and Table 13 of the validation report and therefore the experiments are recognised as valid. The results confirm the necessity of regular chemical analysis of the exposure water and raise the question about exposure system design.

Variability, particularly of the VTG endpoint:
It was clarified in the draft Test Guideline section on VTG measurement that VTG results alone should be interpreted carefully due to the relative high variability of this parameter in fish of different size and at different developmental stage. VTG should be seen in connection with the sex ratio because the skewing of the sex ratio where genetic sex is changed phenotypically can affect the VTG. For example it is significantly different between genetic males and phenotypically sex reversed females;

Considering the above, and also considering the proposed use of the FSDT in the possible Fish Testing Strategy for long term exposure scenarios, in the (draft) Fish Testing Framework under WNT review, the WNT agreed to proceed to the finalization of the draft OECD Test Guideline for the Fish Sexual Development Test.

References

Kinnberg,K., H.Holbech, G.I.Petersen, and P.Bjerregaard (2007), "Effects of the fungicide prochloraz on the sexual development of zebrafish (Danio rerio)", Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 145, pp. 165-170.

Holbech,H., K.Kinnberg, G.I.Petersen, P.Jackson, K.Hylland, L.Norrgren, and P.Bjerregaard (2006), "Detection of endocrine disrupters: Evaluation of a Fish Sexual Development Test (FSDT)", Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 144, pp. 57-66.

OECD (2010), Guidance Document on the Diagnosis of Endocrine-Related Histopathology in Fish Gonads, Series on Testing and Assessment No. 123, ENV/JM/MONO(2010)14, OECD, Paris

PEER REVIEW REPORT FOR THE VALIDATION OF THE FISH SEXUAL DEVELOPMENT TEST (FSDT)

BACKGROUND:

In 2003, Denmark, on behalf of the Nordic countries proposed a project to develop a test method for the detection of endocrine active chemicals during a sensitive life stage in fish. The project was accepted on the Test Guidelines Programme work plan and the method became the so called 'Fish Sexual Development Test' (FSDT).

The FSDT is a modified version of OECD guideline 210 adopted in 1992, the Fish Early-Life Stage Toxicity Test, with added end-points for the detection of endocrine disrupters. The main endpoints measured are the vitellogenin level in exposed animals, their sex ratio (sex determined via histological determination) and secondary sex characteristics. FSDT fits into the OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupters, discussed and agreed at the sixth meeting of the Endocrine Disrupters Testing and Assessment (EDTA) Task Force.

Validation studies of the FSDT have been conducted in a step-wise fashion, and overseen by the Validation management Group for Ecotoxicity Testing (VMG-eco). The purpose of the validation is to develop a robust, relevant and reliable test method for the assessment of chemicals acting as estrogens, anti-estrogens, androgens, anti-androgens and aromatase inhibitors. It is also the purpose of the validation to understand and define the area of application of the assay and any limitations of its use.

In Phase 1 validation, two chemicals were tested: prochloraz (aromatase inhibitor) and 4-tert-pentyl phenol (oestrogen). Phase 1 was also the opportunity to optimise the test design in relation to the subsequent statistics analysis (analysis of the variance and regression analysis have different requirements). Five laboratories participated in Phase 1 using one or two fish species (fathead minnow *Pimephales promelas* and/or zebra fish, *Danio rerio*) depending on resources available. The results of Phase 1 of the validation were presented to the Fish Drafting Group in 2008.

Based on the results and analysis of Phase 1, the VMG-eco recommended the inclusion of another oestrogenic chemical (4-tert-octylphenol) and an androgen (dihydrotestosterone, DHT) in Phase 2. It was also decided that other OECD candidate fish species, such as Japanese medaka (*Oryzias latipes*) and three spined stickleback (*Gasterosteus aculeatus*), should be able to be used in the test. Based on the statistical analysis of the sex ratio and the variability of this endpoint in the control data, the VMG-eco also recommended using the analysis of variance (for a NOEC/LOEC determination) for the statistical analysis rather than the regression analysis (for an EC_x determination). Ten laboratories participated in Phase 2 of the validation. Phase 2 started in early 2009 and was completed in mid 2010.

The purpose of this work was to administer an independent peer review of a report summarising the OECD validation, both Phase 1 and 2, of the Fish Sexual Development Test (FSDT). This document contains the peer review report of the validation of the FSDT.

THE PEER REVIEW PROCESS:

A peer review panel was formed in July 2010 to provide review of the validation process for the FSDT. A panel of five reviewers were selected based on a list of reviewers provided by OECD. Potential reviewers were screened and selected for their expertise, independence, availability and absence of conflict of interest. Each reviewer completed a Declaration of Interest, verifying that no actual or potential conflicts existed. The members of the peer review panel and their affiliations are as follows:

- Dr Petra Kunz, Ecotox Centre, Swiss Centre for Applied Ecotoxicology, Eawag/EPFL, Dübendorf, Switzerland (PK)
- Professor Tohru Kobayashi, Laboratory of Molecular Reproductive Biology, Institute for Environmental Sciences, University of Shizuoka, Shizuoka, Japan (TK)
- Dr Reinhard Laenge, Bayer Schering Pharma AG, Berlin, Germany (RL)
- Dr Aristocle Ndayibagira, Environmental Assessment Directorate, Pest Management Regulatory Agency, Health Canada, Ottawa, Ontario, Canada (AN)
- Dr Joanne Parrott, Priority Substances Effects Section, National Water Research Institute, Environment Canada, Burlington, Ontario, Canada (JP)

The reviewers were asked to consider whether (a) the 8 specific criteria that are set out in the 'Guidance Document 34 on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment' and developed by OECD are met, partially met or not met, and (b) whether the reasons why a criterion is 'not met' are acceptable. If the reasons are not acceptable, the peer reviewers were asked to suggest recommendations on how to solve the problem.

<u>The principles and criteria for test method validation are as follows:</u>

1) *The rationale for the test method should be available.* This should include a clear statement of the scientific basis, regulatory purpose and need for the test.
2) *The relationship between the test method's endpoint(s) and the biological phenomenon of interest should be described.* This should include a reference to scientific relevance of the effect(s) measured by the test method in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.
3) *A detailed protocol for the test method should be available.* The protocol should be sufficiently detailed and should include, e.g., a description of the materials needed, such as specific cell types or construct or animal species that could be used for the test (if applicable), a description of what is measured and how it is measured, a description of how data will be analysed, decision criteria for evaluation of data and what are the criteria for acceptable test performance.
4) *The intra-and inter-laboratory reproducibility of the test method should be demonstrated.* Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.
5) *Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used.* A sufficient number of the reference chemicals should have been tested under code to exclude bias (see paragraphs on "Coding and Distribution of Test Samples").
6) *The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data.* In the case of a substitute test method adequate data should be available to permit a reliable analysis of the performance and comparability of the proposed substitute test method with that of the test it is designed to replace.
7) *Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.* Aspects of data collection not performed according to GLP should be clearly identified and their potential impact on the validation status of the test method should be indicated.

8) *All data supporting the assessment of the validity of the test should be available for expert review.* The detailed test method protocol should be readily available and in the public domain. The data supporting the validity of the test method should be organised and easily accessible to allow for independent review(s), as appropriate. The test method description should be sufficiently detailed to permit an independent laboratory to follow the procedures and generate equivalent data. Benchmarks should be available by which an independent laboratory can itself assess its proper adherence to the protocol.

The peer-reviewers have submitted separate reports. The comments to each question have been sorted and compiled below, followed by a general summary comment provided by the peer-review manager. The original comments from the peer-reviewers related to the eight criteria above are provided in Annex I.

The peer review package, provided by Denmark, includes the draft reports of Phase 1 and Phase 2 of the validation of the FSDT and the preliminary draft Test Guideline prepared by the Fish Drafting Group. A list of the documents that were used by the peer reviewers is provided in Annex II.

The peer review panel was requested to report their views concerning the validation process of the FSDT to VMG-eco which is responsible for the validation process. This report will be further submitted to the Task force for Endocrine Disrupter Testing and Assessment (EDTA), followed by the submission to the Working Group of National Coordinators (WNT). On the basis of the peer review report, EDTA and WNT may recommend actions for the further development of an official test guideline.

This report reflects the consensus of the peer review panel. In some cases, different opinions and comments were expressed by the review panel on some of the criteria. These are described in the report and will be further considered in the discussions by EDTA and WNT in their decisions to the next step of this validation project.

SUMMARY OF COMMENTS FOR EACH CRITERION:

Criteria 1: The rationale for the test method should be available

All reviewers agree that the rationale for the test method is clearly set out.

- One reviewer (AN) would like to see the usefulness of the test to screen for endocrine disrupting chemicals (EDCs) for regulatory purposes discussed in the rationale. It was also noted that in the draft Phase 2 report, pg 5, paragraph 6 and 7 respectively, references were not provided for the following statements: "the production of vtg is controlled by interaction of oestrogenic hormones with the oestrogen receptor" and "…skewed sex ratio in a fish population exposed to EDCs impact its sustainability", these need to be added.

The review panel agrees that this criterion is met

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

All reviewers were in agreement that this criterion has been at least partially met. Three reviewers agreed that this criterion was partially met (TK, AN, JP) and two (PK, RL) that is has been fully met with a provision attached to it. The majority of reviewers agreed that this criterion is fully met for the endpoint of sex ratio but that for the vitellogenin (vtg) endpoint, the relevance to population-level effects is less clear. A number of comments were made:

- The criterion is fully met with the provision that the method consists of a mechanistic endpoint (vtg), which is a marker endpoint not necessarily quantitatively correlated to a toxicological effect, and an toxicological endpoint (sex ratio) which is environmentally relevant (RL). However, it was felt that since reproduction as a toxicological endpoint is not implied to be addressed by this method, the test method can only be applied in a case-specific context with additional information.

- The criterion is fully met provided that the advantages and limitations (i.e. variations among laboratories) of the vtg endpoint are added. An additional section in the protocol that states what to expect from the combined data gained by the vtg and sex ratio endpoints is requested. Of particular interest is information regarding the ability of the test system to differentiate between strong or weak (anti-)estrogenic and (anti-)androgenic substances and aromatase-inhibitors, in addition to information highlighting the limitations of the test system (PK).

- According to two reviewers, clear histological criteria are lacking and therefore requested for the sex ratio endpoint (TK, JP) i.e. 3 or 5 eggs per zebra fish and one egg per medaka testis in order for classification of intersex to be defined, and consistency across the laboratories to be attained. It was also questioned whether one egg in one section of testes would constitute a classification of intersex (JP).

- For the secondary sex characteristics, it was noted that this endpoint is achievable in the medaka only (via anal fin rays) whereas in 60 days post hatch zebra fish this is not a useful endpoint (JP). Thus attention was drawn to the fact that the 3 core endpoints of the FSDT cannot be measured in all fish species tested (i.e. only 2 endpoints in the zebra fish).

- It was pointed out that the presence of positive results in control (vtg: draft report phase 2, pgs 34, 47, 48, 51) should be considered as false positive (AN). A better understanding of the false positive results is needed and it is necessary to specify what should be the acceptable rate of false positive/negative results (AN).
- Additional references describing the relationship between the endpoints of interest and the effects of known estrogens/anti-androgens and androgens/anti-estrogens is suggested i.e. Sholz, S. and Klüver, N. 2009. Effects of endocrine disrupters on sexual, gonadal development in fish. Sex Dev 3 (2-3):136-151 (AN).

The review panel agrees that this criterion is at least partially met (3 reviewers – partially met, 2 reviewers – fully met).

Criterion 3: A detailed protocol for the test method should be available.
The reviewers all agree that the report contained most, if not all, the information needed for a detailed protocol. 3 peer reviewers felt that this criterion was fully met (PK, RL, and JP) whereas two reviewers (TK, AN) were of the view that this criterion was partially met.
3 reviewers noted that there is currently no specification on how sex determination is performed. Thus, guidance (including photographs and micrographs) for sexing and staging fish gonads is strongly requested (TK, AN, JP).
Other comments made:

- Reference to scientific literatures was insufficient in the draft protocol (AN).

- The test acceptance criteria (draft protocol, pg 3) indicate that the concentrations of the test substance in solution must be maintained within ±20% of the mean measured values. The protocol left however unaddressed the difficulties in maintaining measured test concentrations within the acceptable range (draft reports phase 1 and phase 2) (AN). This needs addressing.

- There is a need to clarify whether concentrations higher than 10% of the acute adult LC_{50} or 10 mg/L are limit concentrations (draft protocol, pg 6). It is unclear whether these concentrations should be used in limit test for regulatory purposes. It is therefore felt that the limit test conditions (concentrations) need to be specified (AN).
- The considerations of the statistician regarding the statistical procedures should be taken into account for the description in the guideline (RL).

- Due to the incorrect storage of some water samples in Phase 1 it is suggested that a recommendation on the storage of water samples is added to the draft proposal (PK).

- In medaka, genetic sex is assessed by sex-specific DNA markers, such as *dmy* etc. Therefore, in order to determine the genotypic sex, the method of extraction of genome from medaka, PCR condition and primers should be described (TK).

The review panel agrees that this criterion has been at least partially met (3 reviewers – fully met, 2 reviewers – partially met)

Criterion 4: The intra-and inter-laboratory reproducibility of the test method should be demonstrated.
All reviewers were satisfied that the criterion has been partially met. Notwithstanding this, a number of critical comments on this criterion were made. On the basis of the points raised, one reviewer (PK) requested that the inter-laboratory variation should be examined thoroughly and if necessary, additional experiments should be conducted.

The comments of concern made on the intra- and inter-laboratory reproducibility of the FSDT include:

(i) Analytical chemistry: The considerable differences between nominal and actual concentrations, both within and between the laboratories, are questioned by many reviewers. In phase 1 some laboratories had difficulty maintaining water concentrations of 4-tert-pentylphenol (see p 13-14, i.e. laboratory 3 nominal 600 μg/L measured 23.2 μg/L, laboratory 4 nominal 320 μg/L, measured 1.8 μg/L) (JP). There is also a high variability in the mean measured concentrations among laboratories for 4-tert-octylphenol (draft report Phase 2, pg 11, ex. Laboratories 2 and 8, Table 2) and DHT (draft report Phase 2, pg 13, Table 3) (PK, AN). The highest nominal concentration of DHT was 1000 ng/L and the mean measured concentration was 8.7 ng/L. Another laboratory failed to adequately create solutions (or

was unable to accurately measure DHT) with a mean measured concentration of 93.1 ng/L. (see pg 13 of FSTD Phase 2.) (PK, AN, JP).

It was noted that not every laboratory have tested all chemicals with all fish species. The repeatability over time within the same laboratory cannot be evaluated because no data is available (AN).

(ii) VTG: The use of sub-adult fish resulted in extreme variations particularly of the vtg endpoint (inter-laboratory variation of NOECs of factor >10, intra-laboratory variations of e.g. controls with coefficient variance > 100%). Therefore, it should be considered, whether the analysis of vtg in not-sex developed fish should be down-graded as a non-core endpoint in this method (RL).

Some participating laboratories were not able to find a vtg induction or an effect on sex ratio for the positive-control in Phase 1. It therefore remains unclear if there is a problem with the positive control depending on the fish species used (PK). Similarly, Laboratory 3 seemed not to be very sensitive in picking up the sex ratio changes caused by 4-tert-pentylphenol and prochloraz. LOECs were > 600 µg/L and >434 µg/L, respectively, while other laboratories had LOECs in the 1-6 µg/L range (JP).

(iii) Sex ratio: Phase 1 data does not allow for analysis of inter-laboratory variation and in phase 2, positive controls were not used to assess inter-laboratory variation for vtg and sex ratio in more detail (PK, JP).

There were considerable variation in sex ratio in 3-spined stickleback between laboratory 6 (no significant differences) and laboratory 8 (significantly more males) after exposure to DHT (PK).

Pg 26-27-28 FSTD Phase 2. Most laboratories scored fish for sex as male, female, intersex or undifferentiated. There was concern that some laboratories scored otherwise, and categorised fish as "Not males" (laboratory 3) or 'Not females' (laboratory 5). Laboratory 4 also scored fish as "not intersex" and "not undifferentiated" and similarly, laboratory 9 "not males," and "not females". It was questioned whether this lack of consistency in categorising fish would interfere with the comparisons among laboratories and subsequent interpretation of the results (JP).

The statistical evaluation of the sex ratio data showed that the change from genetic male or female fish to the phenotypic opposite sex is statistically more powerful than just sex ratio (see conclusion by statistician). This should be considered for the guideline development (RL).

(iv) Hatching rate and survival: Draft report Phase 2, pg 17, Table 7; in zebra fish exposed to 4-tert-octylphenol, hatching rate and survival are comparable between laboratories 1 and 4. The inter-laboratory variability is high in hatching rate and survival for medaka (laboratories 4, 5 and 9) and stickleback (laboratories 6 and 8). In zebra fish exposed to DHT, results are comparable (laboratories 1 and 3), as well as in medaka (laboratories 5 and 9) and stickleback (laboratories 6 and 8). Hatching rates and survival for 4-tert-pentylphenol between laboratories 9 and 10 are comparable. Results for Flutamide and 17β-estradiol cannot be discussed as they were generated by one laboratory only (laboratory 6) (AN).

Overall, the inter-laboratory variability as a measure of reproducibility was estimated to be high for VTG and sex ratio measurements and, in some cases, for hatching rate and survival. This high variability might be due to the high variation in mean measured test concentrations. The degree to which the biological variability is affecting test method reproducibility was not discussed, this need to be included (AN).

- It was also pointed out that the strain or population of medaka and zebra fish should be described as the sensitivity for environmental disrupting chemicals is different in each strain even if they are from the same species affecting the inter-laboratory reproducibility (TK).

The review panel agrees that this criterion has been partially met

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.

All reviewers were satisfied that this criterion has been fully met, i.e. the reference chemicals tested are all representative, they are all known from published literature to have oestrogenic or androgenic effects on fish. In addition, all data were analysed in a blinded fashion.

The review panel agrees that this criterion has been 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.

The general consensus was that this criterion has been fully met.

It was felt that the published toxicity testing data on species of interest are available. The existing published methods appear to be similar or comparable to the FSDT with regard to the variability of the results. There is no existing OECD Test Guideline to which the FSDT may be compared (especially relative to vtg measurements). The FSDT method can be adapted to other species, provided fish are sexually differentiated and homologous vtg antibodies are available (AN).

The peer review panel agrees that this criterion has been fully met

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

All peer reviewers were satisfied that this criterion has been partially met.

There is no indication that the tests were conducted in accordance with the principles of GLP. Similarly, it is not clear whether the participating laboratories were GLP-certified. It was felt however that if the GLP principles translates 'sound scientific principles' the criteria was met. Additional comments included:

- The protocol is in principle an enhancement of OECD Guideline No. 210 "Fish Early Life Stage Toxicity Test". The same GLP conditions therefore apply. The Test Guideline for FSDT should specify that the assay is to be carried out under GLP principles (AN).
- For Laboratory 1, survival of zebra fish in the DHT test was poor (51.5 %) in FSTD Phase 2, page 18. There was concern that this did not pass GLP and the criteria for a successful test (JP).
- Concern was expressed that some of the presented data questioned the "sound scientific principles". This includes for example the, at times, substantial differences between nominal and actual chemical water concentrations (especially DHT) which make interpretation of data difficult (PK).

All reviewers agreed that although no tests were formally certified as GLP studies, most participating laboratories appear to have conducted their studies according to GLP principles.

The review panel agrees that this criterion has been met provided GLP was translated to 'sound scientific principles'.

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

The general consensus was that all key data were available and that this criterion has been partially met.

- The majority of peer reviewers commented on the fact that data on the negative control substance ammonia are still missing (PK, AN, RL, JP). Phase 2 document therefore needs the addition of backup data to support the information for the Ammonia tests including figures (missing from Phase 2, pg 42 and onwards).

- Three reviewers noted that hatching rate and survival data for fry is lacking for laboratory 2 (Phase 2) (TK, PK, JP) and data on embryonic development, abnormal appearance and behaviour appear to be missing all together (PK). These data may be helpful to discriminate between general toxicity and mechanism of action endpoints, in particular at higher test concentrations, and therefore needs adding.

- Two reviewers (PK, JP) call attention to the fact that the draft Phase 2 report appeared not to be complete (i.e. sections on performance of endpoints, summary, discussion etc) and note that several headings with incomplete paragraphs are present. This needs to be addressed accordingly.

<u>The review panel agree that this criterion has been partially met.</u>

CONCLUSIONS:
In general, there was a good agreement and often consensus between the reviewers on most of the criteria. The following conclusions can be obtained:
Criterion 1: Considered fully met - the rationale is clearly set out.
Criterion 2: Considered at least partially met - fully met for the sex ratio endpoint but for the vtg endpoint the relevance to population level effects is considered less clear. Defined histological criteria for sex ratio determination currently lacking and therefore requested.
Criterion 3: Considered at least partially met - but it was clear that more detailed information is required in particular on sexing and staging fish gonads (see also comment above).
Criterion 4: Considered partially met - but both inter- and intra-laboratory variability remains to be established. Some significant issues related to both inter- and intra- laboratory variability with regards to the analytical chemistry, vtg and sex ratio endpoints still need to be addressed.
Criterion 5: Considered fully met - the reference chemicals tested are all representative and data analysed in a blinded fashion.
Criterion 6: Considered fully met - FSDT is not a substitute test but a supplement test to the existing set of tests, all relevant data is considered available.
Criterion 7: Considered partially met - if however the principles of GLP were translated to sound scientific principles then the criterion is fully met.
Criterion 8: Considered partially met - but more data supporting the validity of FSDT need to be obtained, specifically data on the negative control substance ammonia in addition to data on hatching and survival in laboratory 2 (phase 2).

This report will 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 finalise the validation activity which links to the development of a new OECD test guideline.

Annex I:

OECD validation of the Fish Sexual Development Test

CRITERIA		PEER-REVIEWERS comments on whether criteria is met, partially met or not met				
		Tohru Kobayashi				
A	The rationale for the test method should be available	Fully met	It is no problem.			
B	The relationship between the test method's endpoint and the biological phenomenon of interest should be described	Partially met	Sex ratio endpoint: What is the histological criteria for female, male, intersex or undifferentiated determined? (such as ovarian cavity, intratesticular efferent duct, seminiferous tubule (lobule)?) Defined criteria are required. Does intersex have male or female function?			
C	A detailed protocol for the test method should be available	Partially met	The protocol is no problem, mostly. However, it appears that the following explanation is lacked: In medaka, genetic sex is judged by sex-specific DNA markers, such as *dmy* etc. To determine the genotypic sex, the method of extraction of genome from medaka, PCR condition and primers should be described.			
D	The intra-, and inter-laboratory reproducibility of the test method should be demonstrated	Partially met	Overall: It appears that it is not easy to compare the results in inter-lab. I guessed one of the reasons is from the culture condition in each lab. However, big problem may be fish tested. Throughout the DRAFT, no the strain (or population) of fish (zebrafish, medaka) tested is described. Generally, the sensitivity for environmental disrupting chemicals is different in each strain, even if they are same species. For example, in medaka, South population is more sensitive for estrogens as compared with North population. It means that feminization is occurred easily in genetic males in South population of medaka as compared with North population. I think that at least the strain name (or population name) of tested fish should be described. If not so, it is difficult to compare and judge the results between inter-laboratories.			
E	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	Fully met	Reference chemicals are no problem.			
F	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	Fully met	Comparable data should be available.			
G	Ideally, all data	Partially met	It seems that the data is agreed with GLP in mind. But, it appears			

	supporting the validity of a test method should have been obtained in accordance with the principles of GLP		that some data is not (at survival ratio).
H	All data supporting the assessment of the validity of the test method should be available for expert review	Partially met	Hatching rate and survival is lacked in LAB 2 (Phase 2). For comparison of the results in inter-lab, it is easy to evaluate that the data for analytical chemistry combines to test-data in the Table and Figure.

OECD validation of the Fish Sexual Development Test

CRITERIA		PEER-REVIEWERS comments on whether criteria is met, partially met or not met				
		Petra Kunz				
A	The rationale for the test method should be available	Fully met	The rationale for the test method is clearly met. Scientific basis, regulatory purpose and need for the test are described in the draft reports and the draft proposal.			
B	The relationship between the test method's endpoint and the biological phenomenon of interest should be described	Fully met	Met, provided further clarification added. The biological relevance of the endpoint sex ratio is clearly stated. For the endpoint VTG it may be helpful to more clearly emphasize its advantages and limitations (i.e. variation among laboratories). It may be helpful to have a section in the protocol that states more clearly on what to expect from the combined data/information gained by the endpoints VTG and sex ratio (...and the optional endpoints 2nd sex characteristics and gonad histopathology). Especially regarding the ability of the test system to differentiate between strong or weak (anti-)estrogenic (i.e. VTG induction in 4tert pentylphenol exposed females), (anti-)androgenic and aromatase-inhibition (i.e. VTG induction in prochloraz exposed males) and to point out the limitations of the test system.			
C	A detailed protocol for the test method should be available	Fully met	I believe that the protocol includes a full description of the test method. I think the annex documents - especially the statistical flowcharts for Vtg and Sex Ration - are very helpful. Due to the incorrect storage of some water samples in Phase 1 it may be helpful to add a recommendation on the storage of water samples to the draft proposal. In addition some remarks of the pros and cons of the different species that can be used in the test system may be helpful.			
D	The intra-, and inter-laboratory reproducibility of the test method should be demonstrated	Partially met	Regarding the reproducibility of the test (intra- and inter lab) I am a little bit concerned about the following points: - the considerable differences between nominal and actual concentrations (i.e. DHT and 4tert-pentylphenol) - The fact that some labs were not able to find a VTG induction or an effect on sex ratio for the positive-control in phase 1. As, to my knowledge a positive control was not used in phase 2, it remains unclear if there may be a problem with the positive control depending on the fish used. - Phase 1 data does not allow for analysis of interlab variation and in phase 2, positive controls were not used to assess interlab variation for VTG and Sex ratio in more detail. - Differences in sex ratio in 3-spined stickleback between lab 6 (no significant differences) and 8 (significantly more males) after exposure to DHT. Concerning the above mentioned points interlab variation should be examined throughout and if necessary, additional experiments should be conducted.			

E	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	Fully met	The selection of the reference chemicals is fine and it represents the substances the test is designed for.
F	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	Fully met	The already existing literature on tests with similar endpoints with the species of concern was taken into account.
G	Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP	Partially met- no proof of GLP (however if GLP principles translates 'sound scientific principles' – criteria met)	I did not find any statements which of the labs were GLP-labs or if the labs were working in accordance with GLP. But I think it can be assumed that the experiments were conducted according to the principles of GLP. However, some of the presented data question the "sound scientific principles" for example the sometimes big differences between nominal and actual chemical water concentrations (especially DHT). This makes it difficult to interpret the data.
H	All data supporting the assessment of the validity of the test method should be available for expert review	Partially met	Partly met. Data on negative control are missing. Remarks/Question: 1. The draft report of phase 2 does not seem to be complete (i.e. performance of endpoints, summary...) 2. After phase 1 it was suggested that positive controls should always be used. Was this done by some labs for example for 4tert-octylphenol (positive control estradiol)?? If so, it would be nice if this data could be added, especially with regard to the problems with the positive controls in phase I. 3. Data on observations like embryonic development, abnormal appearance or behaviour seem to be missing. These data may be helpful to discriminate between general toxicity and MOA endpoints, especially at higher test concentrations.

Peer review comment for OECD FSDT validation by Reinhard Laenge

CRITERIA		PEER-REVIEWERS comments on whether criteria is met, partially met or not met
		Reinhard Laenge
A	The rationale for the test method should be available	Fully met
B	The relationship between the test method's endpoint and the biological phenomenon of interest should be described	Fully met, provided the fact that the method consists of a mechanistic endpoint (Vtg), which is a marker endpoint not necessary quantitatively correlated to a toxicological effect, and a toxicological endpoint (sex ratio) which is environmentally relevant. However, since reproduction as a toxicological endpoint is not implied to be addressed by this method, the test method can only be applied in a case-specific context with additional information.
C	A detailed protocol for the test method should be available	Fully met. The considerations of the statistician regarding the statistical procedures should be taken into account for the description in the guideline.
D	The intra-, and inter-laboratory reproducibility of the test method should be demonstrated	Partly met. The use of sub-adult fish results in extreme variations particularly of the Vtg endpoint (interlab variation of NOECs of factor >10, intralab variations of e.g. controls with coeff. var. > 100%). It should be considered, whether the analysis of Vtg in not sex developed fish should be down-graded as a non-core endpoint in this method. The statistical evaluation of the sex ratio data showed that the change from genetic male or female fish to the phenotypic opposite sex is statistically more powerful than just sex ratio (see conclusion by statistician). This should be considered for the guideline development.
E	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	Fully met
F	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	Fully met
G	Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP	Partly met. The validation reports do not mention the quality assurance systems in place in the different laboratories, and reviewers were not able to review any raw data. However, if GLP principles translates to 'sound scientific principles' the reviewer thinks that criteria are met.
H	All data supporting the assessment of the validity of the test method should be available for expert review	Partly met. Data on negative control are missing.

CRITERIA		PEER-REVIEWERS comments on whether criteria is met, partially met or not met				
		Aristocle Ndayibagira				
A	The rationale for the test method should be available	Fully met	The usefulness of the test to screen for endocrine disrupting chemicals (EDCs) for regulatory purposes was, however, not discussed in the rationale. In draft report (DR) *Phase 2 document, pg 5, paragraph 6*: Reference was not provided for the following statement: "the production of VTG is controlled by interaction of oestrogenic hormones with the oestrogen receptor". In *DR Phase 2 document, pg 5, paragraph 7*: Reference was not provided for the following statement: "…skewed sex ratio in a fish population exposed to EDCs impact its sustainability".				
B	The relationship between the test method's endpoint and the biological phenomenon of interest should be described	Partially met	The biological relevance of the responses evaluated (VTG induction and sex ratio change) is established. Several studies are available in the scientific literature to document these responses. *Suggested additional reference*. The following review paper describes the relationship between endpoints of interest and the effects of known estrogens/anti-androgens and androgens/anti-estrogens: *Sholz S. and Klüver N. 2009. Effects of endocrine disrupters on sexual, gonadal development in fish. Sex Dev 3 (2-3):136-151.* The presence of positive results in control (*VTG: DR phase 2, pgs 34, 47, 48, 51*) should be considered as false positive. A better understanding of the false positive results is needed and it is necessary to specify what should be the acceptable rate of false positive/negative results.				
C	A detailed protocol for the test method should be available	Meet provided further clarification added	The test method is described in the Draft Proposal (DP Version of 1 July 2010). - There is no specification of how sex determination is done during gonad histological examination. Guidance (including photographs and micrographs) on how to recognize female, male, intersex, testis-ova or undifferentiated gonads should be included. - Reference to scientific literatures was insufficient in DP. (Might include references from DR phases 1 and 2). - The test acceptance criteria (DP, *pg 3*) indicate that the concentrations of the test substance in solution must be maintained within ±20% of the mean measured values. The protocol left unaddressed the difficulties in maintaining measured test concentrations within the acceptable range (DR phases 1 and phase 2). - Need to clarify whether concentrations higher than 10% of the acute adult LC_{50} or 10 mg/L are limit concentrations (DP, pg 6). Should these concentrations be used in limit test for regulatory purposes? - Guidance for sexing and staging fish gonads should be				

<table>
<tr><td></td><td></td><td></td><td>provided. The limit test conditions (concentrations) have to be specified.</td></tr>
<tr><td>D</td><td>The intra-, and inter-laboratory reproducibility of the test method should be demonstrated</td><td>Partially met</td><td>Not every laboratory has tested all chemicals with all species. The repeatability over time within the same laboratory cannot be evaluated because no data is available
Analytical chemistry
There is a high variability in the mean measured concentrations among laboratories for 4-tert-octylphenol (DR Phase 2, pg 11, ex. Labs 2&8, Table 2) and dihydrotestosterone (DR Phase 2, pg 13, Table 3: -A possible explanation is that the glassware used may have adsorbed the test chemicals.
Hatching rate and survival
DR Phase 2, pg 17, Table 7; In zebrafish exposed to 4-tert-octylphenol, hatching rate and survival are comparable between Lab 1 and Lab 4. The inter-laboratory variability is high in hatching rate and survival for medaka (Lab 4, 5 and 9) and stickleback (Lab 6 and 8). In zebra fish exposed to DHT, results are comparable (Lab 1 and 3), as well as in medaka (Lab 5 and 9) and stickleback (Lab 6 and 8). Hatching rates and survival for 4-tert-pentylphenol between Lab 9 and 10 are comparable. Results for Flutamide and 17β-estradiol cannot be discussed as they were generated by one laboratory only (Lab 6).
Vitellogenin and Sex ratios
In DR Phase 2: The overviews of NOEC/LOEC for inter-laboratory VTG and sex ratio measurements are reported (Table 12 & 13). The estimated coefficients of inter-laboratory variation of VTG and sex ratio measurements ranged between 7 and 114% (Table below). Note that the NOEC of DHT for Japanese medaka sex ratio in Lab 5 is <51.0 ng/L (not 704.8 ng/L, pg 59).
Table of coefficients of inter-laboratory variation (based on the results from DR Phase 2, Table 12 & 13).
<table>
<tr><th rowspan="2">Exposure chemical</th><th rowspan="2">Species</th><th rowspan="2">Lab</th><th colspan="2">Coefficient of variation</th></tr>
<tr><th>VTG</th><th>Sex ratio</th></tr>
<tr><td>4-tert-octylphenol</td><td>Zebra fish</td><td>1,2,4</td><td>NOEC: 30%
LOEC: 52%</td><td>NOEC: 42%
LOEC: 32%</td></tr>
<tr><td>4-tert-octylphenol</td><td>Japanese medaka</td><td>4, 5, 9</td><td>NOEC: 80%
LOEC: 124%</td><td>NOEC: 41%
LOEC: 89%</td></tr>
<tr><td>4-tert-octylphenol</td><td>Three spined stickleback</td><td>6,8</td><td>NOEC: 33%
LOEC: 31%</td><td>NOEC: 31%
LOEC: 31%</td></tr>
<tr><td>DHT</td><td>Zebra fish</td><td>1,2,3</td><td>NOEC: 90%
LOEC: 90%</td><td>NOEC: 90%
LOEC: 90%</td></tr>
<tr><td>DHT</td><td>Japanese</td><td>5,9</td><td>NOEC:</td><td>NOEC:</td></tr>
</table>
</td></tr>
</table>

			<table><tr><td></td><td>medaka</td><td></td><td>41% LOEC: 102%</td><td>42% LOEC: 42%</td></tr><tr><td>DHT</td><td>Three spined stickleback</td><td>6,8</td><td>NOEC: 114% LOEC: 114%</td><td>NOEC: 114% LOEC: 114%</td></tr><tr><td>4-tert-pentylphenol</td><td>Japanese medaka</td><td>9,10</td><td>NOEC: 11% LOEC: 7%</td><td>NOEC: 83% LOEC: 77%</td></tr></table> *Overall*, the inter-laboratory variability as a measure of reproducibility was estimated to be high for VTG and sex ratio measurements and, in some cases, for hatching rate and survival. This high variability might be due to the high variation in mean measured test concentrations. The degree to which the biological variability is affecting test method reproducibility was not discussed
E	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	Fully met	The reference chemicals tested are representative. They are all known, from published literature, to have estrogenic or androgenic effects in fish. All data were analysed in a blinded fashion (*DR Phase 1, pg 35*).
F	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	Fully met	Published toxicity testing data on species of interest are available (refer to scientific literatures). The existing published methods appear to be similar or comparable to the essay with regard to the variability of the results. There is no existing OECD Test Guideline to which the essay may be compared (especially relative to VTG measurements). The method can be adapted to other species, provided fish are sexually differentiated and homologous VTG antibodies are available.
G	Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP	Partially met	There is no indication that the tests were conducted in accordance with the principles of GLP. It is not clear whether the participating laboratories were GLP-certified. *Overall*, since the protocol is in principle an enhancement of OECD Guideline No. 210 "Fish Early Life Stage Toxicity Test", the same GLP conditions apply. The Test Guideline should specify that the essay is to be carried out under GLP principles.
H	All data supporting the assessment of the validity of the test method should be available for expert review	Partially met	Lab 2 data (phase 2) are still to come

OECD validation of the Fish Sexual Development Test

CRITERIA		PEER-REVIEWERS comments on whether criteria is met, partially met or not met				
		Joanne Parrott				
A	The rationale for the test method should be available	Fully met	This criterion is satisfied in the introduction.			
B	The relationship between the test method's endpoint and the biological phenomenon of interest should be described	Partially met	This criterion is fully met for the endpoint of sex ratio. This is clearly linked to impact on the population. For the endpoint of vitellogenin, the relevance to population-level effects is less clear, as stated in the document. For secondary sex characteristics, this appears to be achievable in the medaka only (via anal fin rays). In 60 dph zebra fish, I anticipate this will not be a useful endpoint. Thus – the 3 test endpoints can be done for medaka – but only 2 endpoints can be done for zebra fish. Does one egg in testes = INTERSEX?? I am unclear for the sex ratio endpoint if one egg in one section of testes would cause the male to be classified as intersex (or ovo-testis). There need to be clear criteria, e.g. 3 or 5 eggs per zebra fish, one egg per medaka testis in order for classifications of intersex to be defined and consistent across labs.			
C	A detailed protocol for the test method should be available	Fully met	I feel the protocol is fully described. I like the full description for the histology tissue, embedding and sectioning.			
D	The intra-, and inter-laboratory reproducibility of the test method should be demonstrated	Partially met	In FSDT Phase 1 – some labs had difficulty maintaining water concentrations of 4 tert pentylphenol (see pg 13-14) Lab 3 nominal 600 ug/L measured 23.2, lab 4 nominal 320 ug/L, measured 1.8 ug/L. Lab 3 seemed to be not very sensitive in picking up the sex ratio changes caused by 4-tert pentylphenol and prochloraz. LOECS were > 600 and >434, respectively, while other labs had LOECs in the 1-6 ug/L range. This table0-4 also compared LOECs and NOECs using both nominal and measured concentrations (table 0-3 for Vitellogenin results does this also). I am not sure if this should be done. Why not compare all measured concns? (see page 38 FSDT Phase 1) Conclusion 1 – Phase 1 - page 42 – variability of endpoints is high. Yes – agree with this. Did it improve in Phase 2? Overall – phase 1 generated no data in inter-lab variability (mentioned on page 43 Phase 1 recommendations). I am a little concerned that some labs could not achieve very high exposure concentrations for the DHT. The highest nominal concentration was 1,000 ng/L and the mean measured			

			concentration was 8.7 ng/L. Another lab was also quite poor at creating solutions (or at measuring DHT) with a mean measured concentration of 93.1 ng/L. (Please see pg 13 of FSTD Phase 2.) Pg 26-27-28 FSTD Phase 2. Most labs scored fish for sex as male, female, intersex or undifferentiated. I am concerned that some labs scored otherwise, and categorized fish as "Not males" Lab 3 or 'Not females' Lab 5. Lab 4 also scored fish as "not intersex" and "not undifferentiated". Lab 9 "not males," and "not females". Does this lack of consistency in the categorizations interfere with the comparisons among labs, and interpretation of the results? FSTD Phase 2 doc…Pgs 34-35. Is the discussion complete? There are several headings with incomplete paragraphs, it appears.
E	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	Fully met	Reference chemical selection is fine. Phase 1 - Weak and strong estrogens tested, as well as an aromatase inhibitor. Phase 2 - Weak estrogens and strong un-aromatizable androgen tested.
F	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	Fully met	The performance criteria are not compared to relevant toxicity testing data – as this test is for subtle effects, changes in sex and estrogenic biomarkers.
G	Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP	Partially met- no proof of GLP (however if GLP principles translates 'sound scientific principles' – criteria met)	There is no absolute statement that the labs are GLP accredited, and probably several of them are not, as they are true research laboratories. However, it is felt that the data were gathered with good GLP in mind, with sufficient replicate exposures, and repeated measures of chemical concentrations in aquaria etc. For Lab 1, survival of zebra fish in the DHT test was poor (51.5 %) in FSTD Phase 2, page 18. Did this pass GLP and the criteria for a successful test?
H	All data supporting the assessment of the validity of the test method should be available for expert review	Partially met	Data not shown for the negative substance ammonia. FSTD Phase 2 document needs the addition of backup data to support the information for the Ammonia tests. Ammonia figures are missing from FSTD Phase 2, pg 42 and onwards. Pg 17-18 FSDT Phase 2 – It appears some data are still to come…Hatching success and survival for fry for Lab 2.

A**NNEX II**

	Peer Review Package for the Fish Sexual Development Test
1	Draft report of Phase 1 of the validation of the Fish Sexual Development Test: S:\Applic\EHS\PROJECTS\TG\Activities\500 Special Activities\Fish activities\Fish Sexual Development Test\Phase 1\Report\DRAFT REPORT OF PHASE 1 OF THE VALIDATION OF THE FISH SEXUAL DEVELOPMENT TEST.doc
2	Draft report of Phase 2 of the validation of the Fish Sexual Development Test: S:\Applic\EHS\PROJECTS\TG\Activities\500 Special Activities\Fish activities\Fish Sexual Development Test\Phase 2\Report\DRAFT REPORT OF PHASE 2 OF THE VALIDATION OF THE FISH SEXUAL DEVELOPMENT TEST.doc
3	Preliminary draft Test Guideline on the Fish Sexual Development Test: S:\Applic\EHS\PROJECTS\TG\Activities\500 Special Activities\Fish activities\Fish Sexual Development Test\Draft TG\FSDT draft proposal July 2010_rev_AG_HH.doc