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

24-Aug-2011

English - Or. English

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

Cancels & replaces the same document of 31 May 2010

Series on Testing and Assessment

No. 127

PEER REVIEW REPORT OF THE VALIDATION OF THE 21-DAY ANDROGENISED FEMALE STICKLEBACK SCREENING ASSAY

(As revised in 2011)

JT03306114

ENV/JM/MONO(2010)18
Unclassified

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 127

PEER REVIEW REPORT OF THE VALIDATION OF THE 21-DAY ANDROGENISED FEMALE STICKLEBACK SCREENING ASSAY

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

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

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2010

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 and 2009)*

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 (2009)*

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, revised 2009)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006, revised 2009)*

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)*

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 on the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox (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. 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 *Explanatory Background Document to the OECD Draft Test Guideline on in vitro Skin Irritation Testing (2010)*

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 *Draft 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)*

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 31 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 is available electronically, at no charge.

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

or contact:

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

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document contains the peer review report of the validation of the 21-day Androgenised Female Stickleback Screening Assay, preceded by a statement from the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up to the peer-review. The validation report of the assay forms the basis for the peer-review and is also published in the OECD Series on Testing and Assessment.

In June 2009, the Working Group of National Coordinators of the Test Guidelines Programme (WNT) was requested by the Secretariat to nominate candidate peer-reviewers. The peer review was managed by an independent contractor identified by the United Kingdom and funded by the OECD Secretariat. The peer-review was carried out between July and September 2009. In September 2009, a report was agreed by the peer-review panel. In December 2009, the Validation Management Group for Ecotoxicity Testing took note of the peer-review report and formulated their views, described further in the WNT agreement.

The peer-review report was endorsed by the WNT at its meeting held on 23-25 March 2010. The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 5 May 2010.

The report *Ammonia as an Endocrine, Negative Substance in the Androgenised Female Stickleback Assay* was prepared, as a follow-up to Agreement of the WNT on the Follow-up to the Peer-Review Report, to demonstrate that ammonia has no androgenic or anti-androgenic activity when tested with the AFSS. It was added at the end of the document No. 127, which was already published in 2010.

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

Table of contents

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

The Peer Review Report of the 21-day Androgenised Female Stickleback Screening Assay was submitted for endorsement to the Working Group of National Coordinators of the Test Guidelines Programme (WNT) at its March 2010 meeting.

Following the conclusions of the report, and taking note of the views below from the Validation Management Group for Ecotoxicity Testing (VMG-eco) on technical issues raised in the Peer Review Report, i.e.:

- Issues of lack of assessment of the reproducibility of the spiggin ELISA: the lead laboratory responded that although it is true that spiggin measurements were performed in a single laboratory in the ring-test referenced in the validation report, previous work by Allen *et al*, 2008, available to peer-reviewers but not taken into consideration, had already evaluated the reproducibility of the spiggin measurements in an intercalibration exercise undertaken in three laboratories. The VMG-eco was satisfied with the response and did not recommend further work,

- Issue of high solvent (methanol) concentration: the lead laboratory responded that although the concentration of solvent was high, it did not affect the spiggin concentration compared with the water control. The VMG-eco was satisfied with the response and did not suggest an alternative solvent,

- Issue of inappropriate use of dihydrotestosterone (DHT) as an androgenic control substance: the lead laboratory corrected an erroneous comment implying that DHT could be aromatised an thus not a good choice of androgenic substance of reference. In previous experiments in stickleback, the lead laboratory never observed vitellogenin induction following DHT exposure. Other androgenic substances had already been considered or used by the lead laboratory (e.g. methyltestosterone) and none were as good as DHT for various reasons. The VMG-eco was satisfied with the response provided,

- Issue of absence of coding of the substances: the lead laboratory indicated that despite the lack of coding, all measurements were analysed blind by the technician who was unaware of the exposure level of samples analysed. The VMG-eco was satisfied with the response provided and did not consider the issue as an impediment to the validation of the assay,

- Issue of lack of commercial availability of reagents: the lead laboratory indicated that the reagents can be obtained free of charge from the lead laboratory and that a commercial kit is in preparation by at least one European company, and others may follow. The VMG-eco did not see this as a real issue,

- Issue of absence of use of a negative substance: the VMG-eco recommended that further consideration be given to possible candidate negative substances,

The WNT agreed that:

i) a negative substance should be identified and tested in one laboratory;

ii) the added value of the Androgenised Female Stickleback Screening Assay should be examined by comparing its sensitivity to detect antiandrogenic activity with that of the fathead minnow and medaka in TG 229 and TG 230 and with the Hershberger assay (TG 441); and

iii) the Androgenised Female Stickleback Screening Assay should initially be developed as a Guidance Document and its role in the Test Guidelines Programme be further considered based on what is found in (ii) above.

PEER REVIEW REPORT FOR THE VALIDAITON OF THE 21-DAY ANDROGENISED FEMALE STICKLEBACK ENDOCRINE SCREENING ASSAY

Background:

In 1996, the Organisation for Economic Co-operation and Development (OECD) established a special activity on Endocrine Disrupter Testing and Assessment with the objectives of:

- providing information and co-ordinating activities
- developing new and revise existing Test Guidelines to detect endocrine disrupters
- harmonising hazard and risk characterisation approaches

This activity was launched at the request of the member countries and the Business and Industry Advisory Committee to the OECD (BIAC) to ensure that testing and assessment approaches for endocrine disrupters would not substantially differ among countries.

In this context, the UK submitted in 2008 a standard project submission form of a 21-day androgenised female stickleback endocrine screening assay. The proposal intended to complement the newly adopted Test Guideline 230 (the 21-day fish endocrine screening assay), which lacks sensitivity in terms of detecting anti-androgenic xenobiotics. The validation work of this assay is under the overview of the Validation Management Group-Eco (VMG-eco).

The purpose of this work, managed by the UK, was to administer an independent peer review of a report summarising the OECD validation of the 21-day Androgenised Female Stickleback Endocrine Screening Assay (AFSS). This document contains the peer review report of the validation of the AFSS.

The peer review Process.

A peer review panel was formed in July 2009 to provide review of the validation process for the AFSS. 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 peer review was managed by Christina Lye (contractor). The members of the peer review panel and their affiliations are as follows:

- Robert Hoke, Ph.D., DuPont, Haskell Global Centres for Health and Environmental Sciences, Delaware, USA
- Henrik Holbech, PhD, Institute of Biology, University of Southern Denmark, Odense, Denmark.
- Dalma Martinovic, Assistant Prof, University of St Thomas, MN, United States
- Wilfred Sanchez, PhD. INERIS (Institut National de l'Environnement Industriel et des Risques), Verneuil en Halatte, France
- John Sumpter, Distinguished. Prof., Head of Institute for the Environment, Brunel University, Uxbridge, UK

The reviewers were asked to consider whether (a) the 8 specific criteria that are set out in the 'Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, GD34' developed by OECD are met, partially met or not met, and (b) 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.

The principles and criteria for test method validation are as follows:

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.

The peer-reviewers submitted separate reports. The comments to each question have been sorted, compiled and summarised below. The original unedited comments from the peer reviewers related to the eight criteria above are provided in Annex 1.

The peer review package, provided by the United Kingdom, includes the draft validation report, a number of relevant scientific articles and a preliminary draft Test Guideline. A list of the documents that were used by the peer reviewers is provided in Annex 2.

The peer review panel was requested to report their views concerning the validation process of the 21-day female stickleback screening test. This report will be further submitted to the Working Group of National Coordinators (WNT). On the basis of the peer review report, 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 WNT in their decisions to the next step of this project.

Summary of comments for each criterion:

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

All reviewers apart from one agree that the rationale is clearly set out. According to 4 out of 5 peer reviewers (HH, DM, WS and JS); the proposed AFSS test meets all the following criteria:

- The reduction in spiggin is a direct anti-androgenic mode of action. Thus, it is considered that the AFSS has the potential to reliably and quantitatively detect antiandrogen chemicals (HH, JS).

- The current fish test screened guidelines (TG 229 and TG 230) does not cover anti-androgenic chemicals (HH, DM, WS, JS). Therefore, the AFSS will provide a called for supplement test to existing test.

- The only existing test for anti-androgens, the Hershberger Test Bioassay in Rats, has significant drawbacks as documented (WS, JS). The AFSS has the potential to be a considerable improvement on the Hershberger Test. In this context it was suggested that a presentation of tests using mammals could be added in the 'background & objectives' to complete the rationale for the AFSS test (WS).

- It is clear that anti-androgenic chemicals have been detected in significant amounts in the aquatic environment worldwide (HH, WS, JS). Thus, there is a regulatory purpose and well defined need for the test (HH, DM).

A different view was expressed by the remaining reviewer (BH):

- The reviewer agrees that there is clear statement of the rationale for the AFSS in the validation report. However, the reviewer does not agree with the concept that the AFSS is the only method of addressing the scientific question at issue, or the perceived regulatory purpose and need for a test for androgenicity/anti-androgenicity.

- The reviewer believes that there has been much research recently that suggests there may be *in vitro* assays available to address these mechanisms. In addition, the existing 21-day fish screen already in draft OECD test guideline form could be slightly modified to address the androgenicity/anti-androgenicity endpoints. This view was corroborated by another reviewer (DM).

- The reviewer BH also believes that the evaluation of these effects can be done based on a much simpler endpoint (e.g. tubercule expression) as opposed to the more complicated organism dissection and ELISA methodology required for spiggin measurement in the AFSS.

- The reviewer did not feel that the larger issues of how the proposed assay meshes within the overall framework of OECD's other fish testing guidelines has been adequately addressed. Nor was it clear to the reviewer that the promulgation of a separate fish assay for essentially a single endocrine endpoint is in keeping with the existing animal welfare guidance.

4 out of 5 peer reviewers agree that this criterion is met. One peer reviewer believes it is not met for the reasons outlined above.

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 fully met. The mechanism behind spiggin reduction is clearly defined and is specifically related to androgen receptor interference. Although spiggin is not relevant to other fish species than stickleback, the androgen receptor antagonists affecting spiggin are highly relevant to other fish species (HH).

One peer reviewer (DM) expressed some concern regarding the choice of dihydro-testosterone (DHT) as androgenising agent, on the basis of DHT and/or its metabolite's demonstrated ability to exert estrogenic activity via conversion to weak estrogenic steroids. The use of the male vitellogenin data to eliminate the doubts over the performance of the assay that may stem from this phenomenon was therefore suggested.

The review panel agrees that this criterion is 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. However, there was also a consensus that the draft protocol provided is not as clear as it could be, and in some details is in disagreement with existing OECD aquatic effect test guidelines.

Many views were submitted from the reviewers, these are listed below:

- Serious concern was raised over animal sources (e.g. what are the recommended sources of the animals for this test, are there any culture facilities, what are the culture requirements etc.) (DM). This reviewer emphasised that without proper animal sources this test will be very difficult to conduct and the results will be highly variable if different entities are collecting wild fish from different environments. In addition, the guidance on biomass loading and fish numbers required for testing could be better clarified (BH).

- The current version relies on referring to previous documents which makes it difficult for the users to access/comprehend the full protocol (DM).

- The clarity of the protocol could be significantly improved from the material procedure summaries similar to those presented for the 21-day fish endocrine screening assay (DM).

- Several questions were raised regarding solvent choice and concentration (BH, DM). There was concern that the methanol concentration for DHT used in the protocol exceeded the OECD recommendation, and that methanol would not be the current solvent of choice. Thus, it is currently unclear to what solvent and at what concentrations will be recommended in the final version. On the other hand, it was also pointed out that although the solvent used was not ideal, it has apparently not affected the spiggin results, demonstrating the robustness of the endpoint (HH). The use of solvent in the assay is a direct result of the choice of DHT; therefore, it is possible that the substitution of DHT with another chemical would negate the need for a solvent in the assay (BH). This need to be considered.

- More explicit guidance on what ratios of androgen/anti-androgen should be used is requested as this ratio appears to have quite an effect on the spiggin production (DM).

- Authors should clearly state what statistical methods should be used to analyse the data (DM). Spiggin concentrations in different treatments have prominently unequal variances and thus a recommendation regarding this should be made. The multiple comparison methods should be recommended. Currently t-tests are used, but there is no rationale provided why more widely accepted post-hoc methods were not used. If t-tests are advised the authors should explicitly state whether the results should be adjusted for multiple comparisons. This seems to vary within the validation report, and thus need clarification.

The review panel agrees that this criterion has been at least partially met (3 reviewers - partially met, 2 reviewers – fully 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. There were a number of critical comments on this criterion.

- Inter-laboratory variability remains to be established (BH, WS, JS) as demonstrated by the fact that all of the spiggin analyses were conducted by a single laboratory (e.g. Laboratory 2).

- The demonstration of repeatability and reproducibility suffers because the exposures were not conducted in an identical manner (DM, WS, BH). Several confounding factors have been listed including various exposure conditions, especially for origin of the fish, fish size and exposure media.

- One reviewer (JS) highlighted the fact that all of the 4 participating laboratories involved in collating data were excellent laboratories, all with considerable experience with the test species 3-spined stickleback (*Gasterosteus aculeatus*). Thus, some concern was expressed over the inter-laboratory variability between other laboratories less experienced with this test species.

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 was partially met. Four androgenic chemicals were tested and these chemicals were representative of the type of substances likely to be tested in the test method. Notwithstanding this, some significant comments were made:

- None of the chemicals were tested 'blind'. The majority of reviewers (HH, BH, DM, and JS) agree that coding is preferable to avoid bias in the results due to prior knowledge of the expected responses. This should be considered.

- To fulfil this criterion, 3 out of 5 reviewers (BH, DM, and WS) suggested that additional chemicals representative of additional modes of action (estrogens, anti-estrogens, aromatase modulators, narcosis etc) should have been tested.

- The exposure to potassium permanganate (PP) suffered from multiple problems and in one reviewer's opinion (DM) these data are not sufficient to draw final conclusions regarding the effects of PP in this assay. The reviewer questioned the value of using PP as a negative control.

This opinion was corroborated by another reviewer (HH) who suggested that a new negative compound may have to be tested.

The review panel agrees that this criterion has been partially 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 at least partially met. Three reviewers stated that it had been partially met (HH, BH, JS) and one reviewer fully met (WS). The remaining reviewer (DM) thought 'not met' but that this is acceptable given that this requirement may relate only to the substitute test methods and the AFSS assay is not a substitute test but a supplement test to the existing tests. Thus, this reviewer suggested that it may be desirable to compare the results of the test with the existing data from complementary assays (YAS and 21-day fish endocrine screening assay). Other comments were:

- 3 reviewers (DM, WS, and JS) complained that all data supporting the validity of the test method was not obtained. No information was provided on the toxicity of other chemicals to the test species (JS).

- It would also be useful if authors could compare LOEC/LC50 values with those provided for other fish species, or with those of stickleback, even if different endpoints were used (DM). One reviewer (BH) suggested that, in particular, it would be useful to evaluate the comparative sensitivity of fathead minnow and stickleback in order to obtain an adequate assessment of the sensitivity of the test species/method compared to other relevant species/methods.

- Given that functional androgens in sticklebacks (11KA vs. 11KT in the majority of teleosts) differ from the other small fish models, it was requested that comments should be provided on the relevance of using stickleback as a model for other fish species (DM). A discussion on similarities and differences in androgen pathways and whether this may impact AFSS result was therefore suggested.

Overall, the peer review panel considers this criterion has been at least partially met (apart from one reviewer who thought reasons for 'not met' was acceptable)

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 agreed that if the principles of GLP were translated to 'sound scientific principles' this criterion was acceptable and met.

Therefore, 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 at least partially met (3 reviewers HH, DM, JS – fully met, 2 reviewers BH, WS – partially met).

According to one reviewer (WS), the test method is not sufficiently detailed to permit an independent laboratory to generate equivalent data. In addition, spiggin standard and primary antibody raised against spiggin preparation for the spiggin ELISA, is not commercially available. This should be considered.

The same reviewer also point out that two other methods have been described recently that measure spiggin and question how these measurements can be used in the AFSS.

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 – but there was a problem in acceptance of the rationale by one of the reviewer.

Criterion 2: Considered fully met – some concern raised regarding the choice of DHT as androgenising agent.

Criterion 3: Considered at least partially met – but it was clear that more detailed information is required.

Criterion 4: Considered partially met – but inter-laboratory variability remains to be established, the selective selection of participating laboratories need to be reflected upon, and some issues related to inter-laboratory variability still need to be addressed.

Criterion 5: Considered partially met – 'coding' is undoubtedly needed, and testing of other chemicals representative of additional modes of actions recommended. A new negative compound, replacing potassium permanganate (PP), may have to be tested?

Criterion 6: Considered at least partially met (one 'not met' but reasons acceptable). More data supporting the validity of AFSS need to be obtained.

Criterion 7: Considered partially met if the principles of GLP was translated to sound scientific principles - no major problem.

Criterion 8: Considered at least met – some concern (see also statement for criterion 3) regarding lack of sufficient details of test method.

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: Comments sent by each reviewer

Henrik Holbech

Table 1. Principles and criteria for test method validation

a) **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.

The Criterion is fully met.

The scientific basis is clear: Reduction in spiggin is a direct anti-androgenic mode of action. The need for the test is also well defined because the current fish test screening guidelines (TG 229 and TG 230) does not cover anti-androgenic chemicals and because anti-androgenic chemicals have been detected in significant amounts in the aquatic environment worldwide.

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

This should include a reference to scientific relevance of the effect(s) measured by the test method in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.

The Criterion is met.

The mechanism behind spiggin reduction is clearly defined and is specific related to androgen receptor interference. Although spiggin is not relevant to other fish species than stickleback, the androgen receptor antagonists affecting spiggin are highly relevant to all fish species. One must though bear in mind that toxic effects could influence spiggin concentrations!

c) **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.

The Criterion is fully met.

The draft test guideline and annexes provides all the information's needed.

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

Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.

The Criterion is partially met.

The intra- and inter laboratory validation comprised 12 full dose response exposures that took place in four different laboratories. The consistence of the spiggin endpoint is verified by these experiments. If the detection of other modes of action than antiandrogenic and androgenic will be included in the protocol, validation of the response to chemicals with these modes of action should be performed in a few laboratories. A new negative compound might have to be tested

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.**

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").

The Criterion is partially met.

Flutamide is the most widely used model anti-androgen. Was the chemicals blinded?

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.**

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.

The Criterion is at least partially met.

It is not a substitute test but a supplement test to the existing tests

g) **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.

The Criterion GLP is probably not directly met.

The impacts are though explained in details and if "the principles of GLP" are translated to "sound scientific principles" the criterion is acceptable and met.

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

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

The Criterion is met.

Robert Hoke

The reviewer would like to commend OECD and all parties involved in the development of the underlying science for the level of effort and attention to detail in the preparation of the documentation provided for the review process. There clearly has been a great deal of meaningful work conducted on various aspects of the basic biology of the three-spine stickleback as well as the utility of the spiggin endpoint for assessment of androgens/anti-androgens in the context of OECD's development of test methods for assessment of the potential endocrine activity of chemicals.

The research provided in the review package is clearly promising but the reviewer is concerned that it may not reflect the current science related to the assessment of the effects of androgens/anti-androgens in fish. In addition, the material provided for the review also does not address the larger issue of how the proposed assay meshes within the overall framework of OECD's other fish testing guidelines nor is it clear to the reviewer that the promulgation of a separate fish assay (and a fourth species for EDC assessments) for essentially a single endocrine endpoint, i.e., androgenicity/anti-androgenicity, is in keeping with the existing animal welfare guidance, particularly if the desired endpoint may be obtained by minor modifications to existing draft or test guidelines (e.g., the 21-day fish screen).

Existing animal welfare guidance would suggest that maximizing the utility of existing (including draft) test guidelines and species should be favoured over the introduction of a new species and test method, unless the new species and test method provides information unobtainable in another manner. Introduction of another species is also likely to lead to additional, unnecessary (from a scientific standpoint) country-specific requests for testing with a specific favoured species and therefore, multiple tests with different species on the same chemical, a situation not in keeping with the goal of test guideline harmonization. The purpose of the "endocrine screens" is merely to identify chemicals that are potentially "endocrine active" and not to provide information suitable for a particular country-specific or region-specific risk assessment.

The documentation reviewed and the proposed need for the AFSS is based on the hypothesis that there is no other existing method for evaluating the androgenicity/anti-androgenicity of a chemical that potentially acts via these endocrine mechanisms. The reviewer believes that there has been much research recently that suggests there in fact may be in vitro assays available to address these mechanisms but will focus on in vivo assays in these comments. My comments based on the 8 validation criteria to be addressed in the review are indicated below.

a) **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.

The reviewer agrees that there is a clear statement of the rationale for the AFSS, however, the reviewer does not agree with the concept that the AFSS is the only method of addressing the scientific question at issue or the perceived regulatory purpose and need for a test for androgenicity/anti-androgenicity. As indicated above, new developments in the area of in vitro assays may provide a means for assaying these endocrine mechanisms without the need for vertebrate testing. In addition, there are existing research and peer-reviewed literature that suggest the existing 21-d fish screen already in draft OECD test guideline form, and which has been extensively validated, could be slightly modified to address the androgenicity/anti-androgenicity endpoints. In a publication in 2004, Ankley et

al. (2004; Env. Sci. Technol. 38:6322-6327) presented the results of initial work designed to characterize the effects of flutamide, a model androgen antagonist, on endocrine function in the fathead minnow, one of the existing species utilized in the draft OECD 21-day fish screen. Ankley et al. demonstrated that 17-beta trenbolone (a model androgen that does not elicit the "mixed responses" sometimes seen with methyltestosterone) effectively masculinised reproductively mature female fathead minnows as evidenced by the development of "breeding tubercules" typically only seen in reproductively mature males. Female fathead minnows co-exposed to trenbolone and flutamide did not develop breeding tubercules.

Expression of male secondary sexual characteristics, such as breeding tubercules, in fish is controlled by the AR and the evaluation of the secondary sexual characteristics in fathead minnows in already an important endpoint in the draft OECD 21-d fish screen in conjunction with measurement of vitellogenin. The development of male breeding tubercules in female fathead minnows clearly demonstrates the androgenic activity of trenbolone while the lack of these same tubercules in females co-exposed to both trenbolone and flutamide (or hydroxyflutamide) demonstrates the anti-androgenic activity of flutamide. Therefore, fathead minnows, in addition to stickleback, can be used to evaluate the androgenic/anti-androgenic activity of a compound. Additional support for this observation was provided in a subsequent study by Martinovic et al. (2008; Environ. Toxicol. Chem. 27(2):478-488) where exposures and co-exposures of reproductively mature fathead minnows to 17-beta trenbolone and vinclozolin demonstrated the androgenic activity of trenbolone and the anti-androgenic activity of vinclozolin.

In addition to demonstrating that a fish species (i.e., the fathead minnow) already included in an OECD draft endocrine test guideline can be used to evaluate the potential androgenic/anti-androgenic effects of chemicals, the studies identified above suggest that the evaluation of these effects could be done based on a much simpler endpoint (i.e., tubercule expression) as opposed to the more complicated organism dissection and ELISA methodology required for spiggin measurements in the AFSS. The expression of breeding tubercules is also already included in the draft OECD 21-d fish screen as part of the assessment of secondary sexual characteristics so is a relatively familiar endpoint.

The reviewer's assessment of this endpoint is that it is not met.

b) **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 testmethod in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.

The relationship of the spiggin endpoint to the androgenicity/anti/androgenicity of a chemical has been adequately demonstrated in the reviewer's opinion based on the research results contained in the review documentation.

The reviewer's assessment of this endpoint is that it is met.

c) **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.

The draft protocol provided as part of the review documentation is not clear as clearly written as it could be and in some details in disagreement with existing OECD aquatic effects test guidelines. As examples, the use of solvents needs to be clarified (including appropriate solvent concentrations). This is an issue of much discussion in the aquatic toxicology community at the moment. Although the reviewer generally believes the use of solvents should be allowed in most aquatic toxicity tests, there may be reason to limit their use and/or their concentrations for endocrine testing purposes beyond the restrictions already in place for typical aquatic testing. The guidance on biomass loading and fish numbers required for testing could also be better clarified. The test guideline recommends all female fish for testing and describes a somewhat convoluted mechanism for attaining female-only lab populations but then indicates that if female-only populations are not possible or if fish cannot be sexed before testing that additional fish should be utilized for testing…this does not seem to be in keeping with animal welfare guidance.

The guidance to set test concentrations based on the MTC or 10 mg/L, which ever is lower, also seems anomalous 10 mg/L is the highest test concentration required in An OECD fish early life stage study surely endocrine effects testing should be conducted at test concentrations below those required by normal aquatic testing schemes (otherwise, traditional testing would provide endpoints that in conjunction with the normal assessment factors would account for the risks associated with chemicals causing "endocrine effects") and perhaps only after consideration/generation of more traditional acute and chronic test endpoints. As an aside the typical use of fish species other than stickleback for the generation of traditional aquatic toxicity endpoints is a powerful argument both scientifically and from the standpoint of animal welfare for the alignment of fish test species across test methods, e.g., the fathead minnow is the only species, based on existing studies, that appears capable of addressing all traditional aquatic toxicity endpoints for fish (acute through fish full life cycle) as well as all relevant endocrine endpoints for fish.

Several disconcerting issues were raised later in the draft TG or in the appendices, i.e., page 9 of the draft TG, paragraph 45, suggests that production of spiggin in females can also be affected by general toxicity and non-endocrine toxic modes of action such as nephrotoxicity this would seem to

indicate that spiggin production is not necessarily a definitive assessment of androgenicity/anti-androgenicity and that more research may be needed to determine how endocrine effects on spiggin production can be separated from non-endocrine effects on spiggin production. Another potential issue arose during the review of Appendix 2 where experimental conditions are presented for the AFSS. The indicated photoperiod is 12 light:12 dark which is at odds with the 16L:8D maintained in most testing laboratories for organism cultures and that is specified in most (all) OECD aquatic test guidelines for fish and invertebrates. The need for a different photoperiod regime is likely to present difficulties in many laboratories and aquatic organism culturing facilities.

The reviewer's assessment of this endpoint is that it is partially met.

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

Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.

The following text was taken from the February version of the document supplied for the review of the AFSS. The reviewer commends the authors of this document for identifying these issues, but while they may be strengths in terms of indicating the association between the concentration of an anti-androgen in water and spiggin levels in female stickleback kidneys, they are clearly weaknesses in terms of the assessment of the intra- and inter-laboratory variability of the AFSS itself. In addition, the fact that all of the spiggin analyses were conducted by a single laboratory does not permit evaluation of the inter-laboratory evaluation of these data and the fact there have been no studies conducted using the methodology outlined in the draft TG does not permit evaluation of the laboratory variability associated with the TG as written.

"In many ways, the results of this analysis were surprisingly uniform even between experiments because several confounding factors had the potential to adversely affect the data interpretations. Some of these factors are listed below:

- The origin of the fish used for the exposures was not standardised, ranging from populations fully adapted to freshwater to fully adjusted to marine environment.
- The size of the fish used was not uniform, their weight ranging from 0.37 to 4.27g.
- The methodology employing in water extraction for analytical verification of compounds changed over the period of experimentations.
- The analytical instruments (GC-MS for FN, LC-MS for VZ, FL, LN) employed for anti-androgen concentrations were also different over the 7-year period.
- The standard curve employed in the spiggin assay was altered after the first set of data were analysed."

The reviewer's assessment of this endpoint is that it is partially met.

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.**

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").

Comparison of the data supplied in the review documentation with the validation conducted for the draft OECD 21-d fish screen would suggest that an insufficient number/range of

compounds have been tested in the AFSS. In addition, the compounds evaluated in the AFSS were not evaluated "under code" in a blind manner.

The reviewer's assessment of this endpoint is that it is partially 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.**

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.

The reviewer had some difficulty interpreting the exact intent of this question. If I am to interpret the question as "has there been an adequate assessment of the sensitivity of the test species/method compared to other relevant species/methods" I would have to say that I believe the answer is no. Given the reviewer's previous discussion of the fathead minnow it would be useful to evaluate the comparative sensitivity of the two species for the assessment of the endpoints at issue as well as more traditional acute and chronic endpoints.

The reviewer's assessment of this endpoint is that it is partially met.

g) **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.

The reviewer believes that essentially none of the data included for the review were collected under strict GLP with the associated audits, etc. However, much of the data appears to have been collected "in the spirit of GLP", meaning that there was appropriate data recording, notebooks, data transparency, etc., and that it would be possible for the investigators involved to provide this documentation for independent review if requested. Therefore, the reviewer believes that full compliance with GLP is unnecessary and unlikely to occur in most academic research laboratories.

The reviewer's assessment of this endpoint is that it is partially met.

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

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

The reviewer's response to the previous criterion is also applicable to this criterion. To the reviewer's knowledge, the draft test method is not "in the public domain", nor should it be until it is indeed ready for widespread adoption. The relevant data for evaluation were provided as part of the review process.

The reviewer's assessment of this endpoint is that it is partially met.

Dalma Martinović-Weigelt

GD34 - Criteria 1/8

a) **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.

This criterion was fully met.

Scientific basis, regulatory purpose and need for the test

The authors clearly presented scientific basis of the anti-androgen assay and supported it by referencing a number of manuscripts (published by their group, as well as by the others) in the peer-reviewed literature.

The authors clearly identified regulatory purpose of the assay; initially the assay was developed to fulfil OECD Test Guidelines for development of the assays that enable the detection of endocrine active substances in the aquatic environment. They also provided a concise overview of the stickleback tests in relation to the OECD guidelines (history of the assay development, validation etc.).Finally the authors stated that the need for this assay was identified in response to the OECD conclusion (OECD report, 2006) that the 21-day fish screen was not adequate for detection of anti-androgens.

GD34 - Criteria 2/8

b) 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.

This criterion was fully met, but I do have some concerns regarding the choice of DHT as androgenising agent (please see the last paragraph). Below I am providing a review of the endpoints addressed in the androgenised female stickleback screen (AFSS). It is my understanding that the stickleback breeding test was eliminated as a candidate test for detection of anti-androgens because the results of the intercalibration exercises (especially the second ring of testing) were not satisfactory due to the fact that two out of three laboratories did not detect anti-androgenic activity via spiggin reduction.

The authors clearly demonstrated the causative/correlative relationship between the spiggin and androgens, thus indicating that this endpoint has a clear biological relevance to the evaluation of the anti-androgenic activity of the chemicals.

Spiggin is normally produced by the kidney of male sticklebacks in response to circulating endogenous androgens. Spiggin is needed for nestbuilding. Spiggin is normally present at very low to undetectable levels in the kidney of the females. However, the female kidney is capable of synthesizing and secreting spiggin into the urinary tract in response to exogenous androgen exposure. Because spiggin

production is an androgen inducible/regulated it is a suitable endpoint for detection of both androgenic and anti-androgenic activity. The authors demonstrated that anti-androgens also reduce spiggin production in sexually mature males, but that the responses can be quite variable due to the natural variation in the reproductive status, sensitivity etc. In order to address this, the authors suggested that the females be androgenised (and thus the spiggin production induced in a fairly uniform manner) with DHT (an unaromatizable androgen) and concurrently exposed to an anti-androgen. The co-exposure to DHT and chemical of interest should result in the reduction of spiggin production if tested chemicals have anti-androgenic effects. The authors have successfully demonstrated that this in fact happens when they tested a series of known anti-androgens. The literature review in combination with the data presented suggests that the detection of anti-androgenic chemicals is possible via the measurement of spiggin level in androgenised female fish. The basic mechanism of action behind this phenomenon has been elucidated in some detail, but the full details of the pathway are yet to be elucidated (identification of subtype of receptors, the most active forms of androgens (11-KA, 11KT vs. T etc).

In addition to above authors speculate that this assay may be of use for detection of androgens, estrogens, and potentially aromatase inhibitors. While reasonable, this claim is not fully supported by the data.

One of my concerns is the choice of DHT as an androgenising agent. It has been demonstrated that DHT and/or its metabolites can exert estrogenic activity via conversion to weak estrogenic steroids (Ishikawa et al., 2006, Journal of Steroid Biochemistry and Molecular Biology, Vol. 98, No. 2-3, pp. 133-138). Please elaborate whether this is a concern given the DHT concentration that is recommended for the testing, given that fish are used (what does metabolism of DHT looks like in fish?). Please use the male vitellogenin data you collected in the past to support your explanation and to address the doubts over AFSS performance that may stem from this phenomenon.

GD34 - Criteria 3/8

c) **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.

This criterion was partially met, but it is not acceptable as presented.

The current description of protocol is not fully acceptable. However, it is my opinion that it can be mended if some of the suggested changes are incorporated. Under the points below I also comment on some of the science aspects of the assay. These comments do not seem to fit into other parts of the review and I felt that they needed to be addressed.

After carefully examining the description of the anti-androgen assay I have come to the conclusion that while majority of the protocol is described in the current version it frequently relies on referring to previous document (AE1150, 2003; http://randd.defra.gov.uk). Similarly the 2003 documents refer to other publications for the detailed methods etc. In my opinion, this makes it quite difficult for the users to access/comprehend the full protocol.

Furthermore, the 2003 protocol contains the information that is not relevant to the most current protocol (e.g. includes males, MT exposure etc.). In general the clarity of the presentation needs to be improved.

Finally, there are several recommendations that in my opinion should have been incorporated in the last version of the protocol, but are not currently included. Below are my detailed comments.

Protocol summary

The clarity of the protocol could be greatly improved from the material/procedure summaries similar to those presented for 21d fish assay (for ideas please see Table 1. in http://www.epa.gov/endo/pubs/att-f_fish_assay_protocol.pdf).)

Chemicals

Androgen: anti-androgen ratios

It would be very useful if the authors could provide more explicit guidance on what ratios of androgen: antiandrogen should be used, and how these should be estimated. This ratio seems to have quite an effect on the spiggin production as discussed in the report. It would also be beneficial to examine how these ratios could be estimated based on the YAS data (activity equivalents etc). I realize that this would not address antiandrogenic activity via "non-AR antagonism" MOAs, but it may be useful for some of the tested chemicals. It appears that the authors have a very nice data set that deals with some of this, and therefore such recommendations should be possible. I am concerned that unless this is specified in some detail it leaves a lot of room for mistakes and inadequate assessment of effects might be made by the inexperienced users.

Solvent choice and concentration

It appears that the solvent (methanol) concentration for DHT that was used in the most recent version of the protocol exceeded the OECD recommendation. What is the recommended methanol concentration?

Furthermore the authors noted that methanol would not be their current solvent of choice. Please state explicitly what solvent you do recommend, at what concentration and provide the rationale for the change in solvent choice (if any is planned) in the introduction.

Animal sources

I have multiple concerns here. Please address the following questions: What are the recommended sources of the animals for this test? Are there any culture facilities? What are the culture requirements? If the wild fish are to be used please make minimal requirements for conducting the test and identify the criteria for the control fish (mortality, age, size, basal spiggin concentrations, basal vitellogenin concentrations in males? etc.). Conducting exposures with field-caught impaired animals could lead to inconclusive results and/or erroneous test outcomes.

Endpoint measurement

Please make the clear recommendation as to whether VTG measurement should be made concurrently with spiggin measurements.

Data analyses recommendation

Authors should state clearly what statistical methods should be used to analyse the data.

Spiggin concentrations in different treatments have prominently unequal variances and thus a recommendation regarding this should be made; in the report the authors used regular ANOVA as well as ANOVA that uses Welch method to accommodate for unequal variances.

The multiple comparison methods should be recommended. Currently t-tests are used, but there is no rationale provided why more widely accepted post-hoc methods were not used. If t-tests are advised the authors should explicitly state whether the results should be adjusted for multiple comparisons (this seems to vary within the report and/or was not clear).

Criteria for acceptable test performance

While these may be provided elsewhere it is important to incorporate this information in the current protocol. As of now this is not adequately addressed.

GD34 - Criteria 4/8

d) **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.

Partially met, but acceptable

For the androgenised female stickleback screen there is a fairly large dataset that addresses the repeatability and reproducibility issues. However the demonstration of repeatability and reproducibility suffers because the exposures were not conducted in identical manner. This resulted in increased variability of the measurement. As of now it is not fully clear what the main contributors of the variability are (ELISA assay, the variability of the delivered chemical concentrations or inherent biological variability). It would be very helpful if the authors reported commented on coefficients of variation for spiggin in control and treated animals. The users should be provided with some guidance on what the acceptable coefficients of variation are. In general, I do believe that the overall analysis of the data demonstrates that the proposed assay was able to consistently detect anti-androgenic activity (qualitatively), but the large variation in estimated LOEC values is of some concern (e.g. Cefas Weymouth flutamide data).

GD34 - Criteria 5/8

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.**

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”).This criterion was partially met. For acceptability comment please see below.

Representativity

All of the tested chemicals (with the exception of the potassium permanganate) are anti-androgenic chemicals. If the representativity of chemicals is described in terms of the modes of action they represent then the available information is not acceptable. To fulfil this criteria additional chemicals representative of additional modes of action (estrogens, antiestrogens, aromatase modulators, and narcosis) would have to be tested. If these were tested please incorporate the summary of the tests in this document. If the representativity of chemicals is described in terms of diversity of chemical structure/use then it is met.

From the scientific perspective, testing additional chemicals with diverse modes of action would be extremely valuable (for example chemicals for which antiandrogenic effects may be overshadowed by other MOAs – prochloraz is a known AR antagonist, CYP19 and CYP 17 inhibitor, and inducer of other P450s). For example, in case of prochloraz it would be important to see if the putative increase in circulating T due to the lack of conversion to E2, can be overshadowed by antiandrogenic activity and to describe dose-response for it.

The exposure to potassium permanganate suffered from multiple problems and in my opinion these data are not sufficient to draw the final conclusions regarding the effects of PP in this assay. The authors themselves recognized the multiple weaknesses of this experiment (inability to consistently deliver DHT and thus lack of appropriate controls for PP/DHT treated fish, increase in spiggin production in PP/DHT treated animals and toxicity/mortality issues). Given the toxicity of the PP I agree with the authors and question the value of using PP as a negative control. Thus I am not sure that this experiment is worth repeating. Regardless, there is a need for an exposure to a negative control chemical.

Coding

"The purpose of coded substance is to avoid bias in the results due to prior knowledge of the expected responses. Guidance Document 34 says that it is "preferable" but it is not a requirement". The coding issue was not addressed in the current version of the protocol and it should be recommended in the future version.

GD34 - Criteria 6/8

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.**

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.

This criterion was not met, but this is acceptable.

Not met, but acceptable given that this requirement may relate only to the substitute test methods. It is my understanding that this assay may be used to complement YAS findings and "21-DAY FISH ENDOCRINE SCREENING ASSAY". Thus it may be desirable to compare the results of AFSS with the existing data for those toxicity tests.

Authors should provide a summary of data available for comparison, and if these are not available then they should explicitly state this. It would be useful if authors could compare LOEC/LC 50 values with those provided for other fish species or with those for stickleback even if different endpoints were used. This would allow for better evaluation of the performance of the AFSS test and for comparison of its sensitivity/utility.

Furthermore given that functional androgens in sticklebacks (11KA vs. 11KT in majority of the teleosts) differ from the other small fish models please provide some comments on the relevance of using stickleback as a model for other fish species. Please discuss similarities and differences in androgen pathways and whether this may impact AFSS results.

GD34 - Criteria 7/8

g) **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.

This criterion was partially met. It is acceptable, but needs to be clarified in the final version.

It appears that the laboratories followed GLP practices. However, this was not explicitly stated and it remains unclear whether the laboratories were "GLP-certified".

GD34 - Criteria 8/8

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

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

This criterion was fully met. It is acceptable.

Authors published/prepared manuscripts for submission to scientific journals. Furthermore, the past reports and assay descriptions are publicly available. Furthermore, there are explicit plans for the current report review to be available to the public.

Wilfried Sanchez

Introduction

The 21-day androgenised female stickleback endocrine screening test is presented as a complementary test to the OECD test guidelines (TG 229 and TG 230) while it allows to assess anti-androgenic activity of chemicals. For this purpose, the selected model fish species is the three-spined stickleback (Gasterosteus aculeatus L.) and the major measured end-point is the spiggin synthesis in kidney of female sticklebacks co-exposed to the tested chemical and model androgen DHT.

In light to the bibliography, all laboratories involved in this project (i.e. Cefas Burnham, Cefas Weymouth, CEH Lancaster, Bergen University) have a solid experience on stickleback ecotoxicology especially on the effects of endocrine disrupting chemicals. Based on this experience, the report addresses a retrospective validation of a large dataset produced over a period of 7 years. Moreover, a test investigating the requirement to employ methanol as solvent carrier and the suitability of potassium permanganate as a negative control is presented. This report is accompanied to a draft test guideline and any scientific publications from partner laboratories.

a) **The rationale for the test method should be available**

The rationale for the test method is available. In the "Background and objectives" section (11-12), the report describes a limit of TG229 and TG230: lack of ability to detect antiandrogens. Moreover, several scientific publications are cited to demonstrate that antiandrogens are an important class of endocrine disrupting compounds frequently encountered in water ecosystems. In this section, complementary data addressing the effects of antiandrogens on organisms could be added. Indeed, except the work of Jobling et al (2009) that suggests the contribution of antiandrogens in sexual disruption of wild fish, no other data are provided.

In the "General discussion" section (151-154), the Hershberger's test (i.e. alternative test for the detection of antiandrogens) and more rapid alternative are described and the relevance of stickleback test is discussed. Presentation of test using mammals could be added in the "Background and objectives" section to complete the rationale for the test.

Moreover, in a recent paper, Berg et al. (2009) described a first quantitative spiggin ELISA based on recombinant spiggin. This point and associated perspectives are not presented in the document. Also, it appears as very important to integrate this data in the discussion of test while the assay used in this comparison study is only a semi-quantitative study similarly to the test developed by Sanchez et al. (2008).

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

The relationship between spiggin measurement and antiandrogenic activity of chemicals is partially described in the report. In the milestones section (14), the mechanism of spiggin synthesis and the role of this protein are described using appropriate scientific references. However, the molecular mechanism of spiggin and the involvement of androgen receptor can be described here. These data are completed by relevant information on molecular effects of antiandrogens (93) and on mechanisms of action of endocrine disruptor chemical (156-159).

c) **A detailed protocol for the test method should be available**

A draft test guideline is available in an attached document. This document describes clearly animal species, husbandry conditions, what is measured, how it is measured, data analysis and decision criteria for data evaluation.

Several weaknesses are noticed in this attached document:

- Section 17: in this section, fish manipulations for sex determination were described as easy.However, in the comparison study, we show that male is introduced in groups due to mistakes in sex determination. In this section, it could be more interesting to describe how many fish are necessary to obtain a sufficient number of females in accordance with the required statistical power.

- Section 17: fish can be sexed using DNA techniques. However, I think that these techniques are not applicable in this context due to cost, time and required technicity of the assay.

- Section 40: R^2 is not a valuable indicator to guarantee adequate prediction of concentrations. Indeed, R^2 describes data dispersion but a shift of standard curve can be observed with a great R^2 value. Hence, it is necessary to compare the four parameters describing standard curve using ANOVA test to be sure that predicted values of spiggin are in the same area.

- Section 41: LOD and LOQ definitions are valid only if standard curve is linear. However, if standard curve is a classical four parameter curve, LOD is determined when binding ratio is 90% and LOQ is determined when binding ratio is 80%.

- Annex 4: several data on spiggin standard production and titration or on primary antibody production are lacking. In this condition, this protocol does not permit to generate equivalent data in an independent laboratory.

d) **The intra-and inter-laboratory reproducibility of the test method should be demonstrated**

The intra- and inter-laboratory reproducibility is not demonstrated. In this study, data from a large dataset are analysed. These data are produced by 4 laboratories using various exposure conditions (annex 1). However, as indicated in "Statistical analysis of all existing data to date – Background information on the datasets" (79), analysis were not performed by each partner after exposure. For example, all spiggin and DHT measurement were performed at Lab 2. In this condition, intra and inter-laboratory reproducibility cannot be evaluated. For this purpose, it is important that each partner involved in test makes all measurements require by the test.

However, in this comparison study, partner laboratories used various exposure conditions especially for origin of the fish, fish size and exposure media. In this context, test variability due to these confounding factors is low. Moreover, in this version of AFSS, reproductive status of sticklebacks had not effect on test result while female fish do not synthesised spiggin in normal conditions. Hence, AFSS appears as weak sensitive to confounding biological factors.

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**

To assess the performance of the test method, 4 chemicals were used. The first one is flutamide, a model antiandrogen. The other is fenitrothion, linuron and vinclozolin. All of them are agricultural chemicals known to possess antiandrogenic activity.

To complete this work, it could be interesting to evaluate the response of other endocrine disruptor compounds having androgenic or estrogenic activity. As indicated in the "General discussion" section (157), the EDEN project evaluated response of E1, E2, EE2 and NP. Unfortunately, these data are not presented in the report. Moreover, in the "Description of the stickleback anti-androgen assays – the androgenised female stickleback screen" section (37), it is speculate that the test is able to detect androgenic activity by spiggin induction in the negative control group and by higher spiggin content in groups treated by DHT. Hence, it could be useful to test an androgenic chemical such as trenbolone to validate this hypothesis.

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**

This point is fully met. However, we can notice that few relevant toxicity data are available in the three-spined stickleback.

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

This point is not addressed in the report. However, several data used in this report have been previously peer-reviewed. Hence, we can consider that this point is partially met.

h) **All data supporting the assessment of the validity of the test method should be available for expert review**

This point is partially met. The data supporting the validity of the test method are organized and easily accessible. However, the test method is not sufficiently detailed to permit an independent laboratory to generate equivalent data. Spiggin ELISA test uses semi-quantitative spiggin standard and primary antibody raised against spiggin preparation. These two components are not commercially available and can be provided by the Cefas (Partner 2 of the test). What are the conditions to obtain these components? Is Cefas able to provide sufficient component quantity around the world to perform AFSS test? Unfortunately, no data is provided to produce independently these components and in this condition, it is not possible to obtain similar data especially due to the utilisation of a semi-quantitative standard.

Recently, two other methods have been described to measure spiggin. How these methods can be used in the AFSS test?

Summary

Criteria A: fully met

Criteria B: fully met

Criteria C: fully met

Criteria D: partially met

Criteria E: partially met

Criteria F: fully met

Criteria G: partially met

Criteria H: partially met

John Sumpter

My opinions on each of the eight criteria for test method validation are as follows:

a) **The rationale for the test should be available.**

It is clear that the only existing test for anti-androgens, the Hershberger Test, has significant drawbacks, as is documented. There is a clear need for a better test, especially one appropriate to environmental issues. The proposed test meets all these criteria.

Conclusion: Met

b) **The relationship between the test method's end points(s) and the (biological) phenomenon of interest should be described.**

The androgen-specific nature of spiggin production is universally accepted. It is logical that anti-androgens will inherit spiggin synthesis in a dose-dependent manner. Hence, the biological relevance of the test is very well established.

Conclusion: Met

c) **A detailed protocol for the test method should be available.**

In the documentation I reviewed, there was no "detailed protocol". However, it is clear that one could be produced (and perhaps has been); the report I reviewed certainly contained most, if not all, the information needed for a detailed protocol.

Conclusion: Partially-met

d) **The intra and inter-laboratory reproducibility of the test method should be demonstrated.**

Four labs contributed to the data in the report. A lot (perhaps half) of that report was taken up with intra- and inter-laboratory comparisons of data. The analysis supplied was both thorough and comprehensive, and believable. However, all the labs involved were excellent labs, all with considerable experience with the test species. It remains to be seen if the labs with less experience and/or expertise can produce such reproducible data. Further, one specific aspect of the test, the measurement of spiggin, was done in only one of the labs (for all partners), and hence it is unknown how much variability there would be if other laboratories conducted the spiggin analyses.

Conclusion: Partially-met.

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.**

Four anti-androgenic chemicals were tested by one or more of the laboratories. These chemicals are definitely representative of the type of substance likely to be tested. All four demonstrated anti-androgenic activity. However, none were tested "blind", as ideally they should have been.

Conclusion: Partially-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**.

It is somewhat unclear exactly what is required here! The toxicity (e.g. LC50 values) of the tested chemicals to fish was reported, and this information was used to ensure that toxic concentrations were not used. No information was provided on the toxicity of other chemicals to the test species if indeed it was required).

Conclusion: Either met or partially-met, depending on what was required.

g) **Ideally all data supporting the validity of the test method should have been obtained with the principals of GLP.**

Although all the tests appear to have been conducted very completely (Hence the good data obtained), it is unlikely any of the work was done to GLP. In fact, GLP is not mentioned once throughout the report as it should have been). However, I doubt this criteria is relevant to the acceptability (or otherwise) of the test proposed.

Conclusion: Partially-met.

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

The applicant should be congratulated on supplying a lot of data, and analyses of that data. I checked some of it, and support the applicant's conclusions. Of course more data can always be provided (no data from the ELISA of spiggin was provided, for example), but what was here was all the key data, and I was convinced of its quality.

Conclusion: Met

Overall conclusion:

This was a high quality report. I was convinced by the data. I consider that the 21-day androgenised female stickleback endocrine screening assay has the potential to reliably and quantitatively detect anti-androgenic chemicals. The assay has the potential to be a considerable improvement on the existing Hershberger Test. It is up to the OPECD to take into account non-scientific reasons for accepting or rejecting the proposed test: the science supporting it is, in my opinion, very strong.

ANNEX II - Peer Review Package for the Androgenised Female Stickleback Screening Assay

	Peer Review Package for the Androgenised Female Stickleback Screening Assay
1	Validation report of the 21-day androgenised female stickleback endocrine screening assay. Katsiadaki, Cefas Weymouth Laboratory, Weymouth, Dorset, UK, July 2009.
2	Intercalibration exercise using a stickleback endocrine disrupter screening assay. Allen et al (2008). Toxicology & Chemistry, Vol 27, No 2, pp 404-412.
3	Use of the three-spine stickleback (*Gasterosteus aculeatus*) as a sensitive in vivo test for detection of environmental antiandrogens. (Katsiadaki et al, 2007, Environmental Health Perspective, Vol. 114, supplement 1, April 2006)
4	Detection of the anti-androgenic effect of endocrine disrupting environmental contaminants using in vivo and in vitro assays in the three-spined stickleback (Jolly et al, 2009)
5	Development of a stickleback kidney cell culture assay for the screening of androgenic and anti-androgenic endocrine disrupters (Jolly et al, 2006)
6	Detection of environmental androgens: a novel method based on enzyme-linked immunosorbent assay of spiggin, the stickleback (*Gasterosteus aculeatus*) glue protein (Katsiadaki et al, 2002a, Toxicology & Chemistry, Vol. 21, No. 9, pp. 1946-1954)
7	The potential of the three-spined stickleback (*Gasterosteus aculeatus L.*) as a combined biomarker for oestrogens and androgens in European waters (Katsiadaki et al, 2002b, Marine Environmental Research, Vol. 54, pp. 725-728).
8	The model anti-androgen flutamide suppresses the expression of typical male stickleback reproductive behavior (Sebire et al, 2008, Aquatic Toxicology, Vol. 90, pp. 37-47)
	Preliminary Draft Test Guideline: Androgenised Female Stickleback Screen (AFSS)

AMMONIA AS AN ENDOCRINE, NEGATIVE SUBSTANCE IN THE ANDROGENISED FEMALE STICKLEBACK ASSAY

This document presents *the testing* of ammoniac with the Androgenised Female Stickleback Screen (AFSS), which was performed to demonstrate that ammonia has no androgenic or anti-androgenic activity when tested with the AFSS. It was developed by the United Kingdom (Ioanna Katsiadaki and Marion Sebire).

The peer review report of the AFSS was published in 2010 as No. 127 in the Series on Testing and Assessment, together with the agreement of the Working Group of National Coordinators of the Test Guidelines Programme (WNT) on the follow-up of the peer review. One of the follow-up actions agreed by the WNT was that a negative substance should be identified and tested in one laboratory.

The results of the test were presented at the Fish Drafting Group meeting held in Tokyo on 9-10 February 2011. The Validation Management Group for Ecotoxicity testing and the Fish Drafting Group were requested to send comments on the document after the Tokyo meeting. The Secretariat proposed to attach this report to the document N°127.

1. INTRODUCTION

Background and Objectives

1. Cefas has developed over the past 10 years a sensitive and specific test for the detection of androgenic xenobiotics using the 3-spined stickleback (*Gasterosteus aculeatus*) and a single biomarker, the androgen-regulated glue protein spiggin. The test is particularly useful in a regulatory context for the detection of androgen antagonists in view of:

a) The lack of clear response from OECD TG229 and TG230 during the validation phase of both screens (see OECD VMG-eco report, 2006).

b) The concern stemming from the high anti-androgenic activity detected in the environment (Johnson *et al*, 2006; Urbatzka *et al*, 2007; Tollefsen *et al*, 2007; Hill *et al*, 2009).

2. For this reasons the United Kingdom submitted a standard protocol submission form (SPSF) requesting the retrospective validation of a large dataset produced by Cefas and collaborators (over an 8 year period), using this test. In the test, female sticklebacks are simultaneously treated with a modest concentration (5µg/L) of the non aromatisable androgen dihydrotestosterone, DHT, and the test compound (Katsiadaki *et al*, 2006; Katsiadaki *et al*, 2007). Any inhibition or reduction of spiggin induction by DHT is attributed to the anti-androgenic activity of the test compound, when compared to the spiggin induction in female sticklebacks exposed to the test compound alone (without DHT; test compound control). The test is called the androgenised female stickleback screen (AFSS).

3. A very comprehensive validation report on the AFSS was submitted to the OECD and was peer reviewed in 2009. At the December 2009 meeting of the Validation Management Group for ecotoxicity testing (VMG-eco), it was decided that complementary studies were needed on an endocrine negative substance, before the test could be fully validated and adopted as a guidance document. Testing of at least one negative substance would be useful to give an estimation of false positives outcomes generated by these tests. Following the VMG-eco meeting, two substances were proposed as first candidates to consider: n-octanol and ammonia. n-Octanol was already tested in the validation of OECD TGs 229 and 230 and was considered to be a "challenging" negative substance at that time (OECD and CEFIC reports are publicly available with the results of the various experiments performed). The main two challenges

with the use of n-octanol were the difficulty to maintain test concentrations (a very high renewal rate was needed) and high microbial growth in the system, requiring frequent change of test vessels.

4. In respect to the use of ammonia as a negative substance, the experience from the US EPA experimental programme was more positive, besides the issue of concentration setting (due to the very steep slope of the concentration-response curve for acute toxicity) and its dependence on pH and temperature conditions. Therefore, following the VMG-eco teleconference in March 2010, it was decided to use ammonia as a model negative substance.

5. ***The objective of this work was to complement the existing dataset for AFSS evaluating the effect of a negative substance, in this case ammonia. The performance of the test with the negative substance is critical for its adoption as an OECD guidance document for in vivo confirmation of suspected anti-androgenic activity of chemicals.***

Selection of concentration range for ammonia exposure

6. Since ammonia is a highly toxic compound, a thorough search was conducted into available data for fish toxicity before selecting the test concentrations. Toxicity of ammonia and ammonium-containing substances is highly dependent on the pH and also on temperature. Toxicity due to ammonium salts increases with increasing pH (US EPA, 1999), because the fraction of unionised ammonia increases according to the following equation (Emerson, 1975): fraction unionised = $1/(10^{pKa-pH}+1)$. Moreover, the pKa is reciprocally related to temperature. The percentage of total ammonia present as NH_3 in aqueous solutions at 20°C is 0.039% at pH 6 and 3.82% at pH 8. The concentration of unionised ammonia will be lower at higher ionic strengths of very hard fresh water (such as the laboratory own freshwater supply) or salt water environments.

7. The effect of temperature on the toxicity of ammonia and ammonia-containing substances was demonstrated in studies conducted by Thurston and Russo (1983) where acute toxicity decreased as temperature increased over the range 12-19°C or studies by Thurston *et al* (1983), where toxicity decreased as temperature increased over a range of 12-22°C. The toxicity of free ammonia to fish increased 1.5-5 fold as the temperature decreases from *ca*. 25°C to *ca*. 5°C.

8. The range of free ammonia concentrations for testing as a negative substance suggested by the VMG-Eco in March 2010 was 0.06, 0.13, 0.25 and 0.5 NH_3 mg/L (equivalent to 1.94, 3.88, 7.75 and 15.5 total ammonia (TA) (NH3+NH4)-N mg/L at pH 7.7 and 27°C). Adjusting this concentration range for the laboratory system (pH 8 and 17°C) meant that the range of total ammonia (referring thereafter as TA-N/L) should be 1.95-16.2mg/L. However, based on a literature search, it was estimated that the highest concentration proposed (0.5 mgNH_3 mg/L) was likely to exceed the level of severity of this test and result in considerable mortalities in the stickleback. Some data that have influenced the decision are provided below.

9. Table 1 shows the mean LC50 values for ammonia in fish species found in Canadian waters (Environment Canada, 2001). Most of the acute tests were conducted in laboratories where concentrations were maintained at a constant level, and after 48–96 hours mortality did not change. As table 1 displays for the majority of the species studied the LC50 was not far from the highest proposed concentration suggesting that 0.5 mgNH_3 mg/L had a clear potential for toxic effects.

10. The US EPA report on ammonia concentrations related to water quality criteria (2009), detailed examples for acute and chronic toxicity of ammonia in aquatic animals. The LC_{50} of ammonia in the three-spined stickleback at pH 7.5 was 143.9 mg TA-N/L at 15°C and 78.7 mg TA-N/L at 23.3°C, equivalent at *ca.* 1.23mg free ammonia (NH_3)/L (Hazel *et al*, 1971). Similarly, the LC_{50} in guppies (*Poecilia reticulata*) was 75.65mg TA-N/L at pH 7.45 and at 25°C i.e. *ca.* 1.19 mg NH_3/L.

11. Hence, in order to avoid toxic effects and comply with the United Kingdom home office licence on animal testing, it was decided to proceed with a lower range of ammonia concentrations, adopting the highest ammonia concentration to be at approximately 5 times lower than the LC_{50} measured by Hazel *et al* (1971). So the range of concentration tested was 0.06 mg/L (Low), 0.13mg/L (Medium) and 0.25 mg/L (High), all referring to total ammonia (mg TA-N/L) under pH 8 and 17°C. In terms of free ammonia (NH_3), the concentrations in the experimental tanks were 0.06, 0.13 and 0.25mg/L respectively.

Table 1: Mean LC50s for un-ionized ammonia in Canadian fish species (Environment Canada, 2001)

Common name	Species name	LC_{50}[1] (mg NH_3/L)	No. of studies	Minimum LC_{50} (mg NH_3/L)	Maximum LC_{50} (mg NH_3/L)
White perch	*Morone americana*	0.279	2	0.150	0.520
Mountain whitefish	*Prosopium williamsoni*	0.289	3	0.143	0.473
Chinook salmon	*Oncorhynchus tshawytscha*	0.442	3	0.399	0.476
Rainbow trout	*Oncorhynchus mykiss*	0.481	112	0.158	1.090
Pumpkinseed	*Lepomis gibbosus*	0.489	4	0.140	0.860
Coho salmon	*Oncorhynchus kisutch*	0.520	8	0.272	0.880
Cutthroat trout	*Oncorhynchus clarki*	0.642	4	0.520	0.800
Brown trout	*Salmo trutta*	0.657	3	0.597	0.701
Mountain sucker	*Catostomus platyrhynchus*	0.685	3	0.668	0.819
Walleye	*Stizostedion vitreum*	0.706	4	0.510	1.100
Golden shiner	*Notemigonus crysoleucas*	0.720	1		
Golden trout	*Oncorhynchus aguabonita*	0.755	1		
Brook trout	*Salvelinus fontinalis*	1.005	2	0.962	1.050
Smallmouth bass	*Micropterus dolomieu*	1.105	4	0.690	1.780
Largemouth bass	*Micropterus salmoides*	1.304	2	1.000	1.700
Fathead minnow	*Pimephales promelas*	1.344	45	0.240	3.440
White sucker	*Catostomus commersoni*	1.349	7	0.760	2.220
Mottled sculpin	*Cottus bairdi*	1.390	1		
Bluegill	*Lepomis macrochirus*	1.406	15	0.260	2.970
Spotfin shiner	*Cyprinella spiloptera*	1.479	3	1.200	1.620
Channel catfish	*Ictalurus punctatus*	1.707	14	0.500	4.200
Stoneroller	*Comostoma anonalum*	1.720	1		
Green sunfish	*Lepomis cyanellus*	1.860	6	0.590	2.110

[1] LC50 is the geometric mean when more than one study result is reported.

2. MATERIALS AND METHODS

12. The experiment followed the protocol design described in the AFSS draft guidance document.

Animals

13. The experimental fish were obtained from the wild when few months old (Houghton Springs, Dorset). They were maintained in the Cefas facilities under photoperiod that simulated the natural day length for 8 months. Two weeks prior to the test the fish were moved to summer photoperiod 18 hours light: 6 hours dark (18L: 6D) to facilitate the onset of sexual dimorphism). On day 0, 84 female sticklebacks were selected from the stock population and moved to the experimental tanks, under neutral photoperiod 12 hours light 12 hours dark (12L:12D) and maintained at 17°C (± 1ºC) throughout the experiment. There was no acclimation period as the test compounds need to be established in the tanks before the fish moved in. The fish were fed with frozen bloodworms once only every 48 hours (a slight deviation from the AFFS protocol as overfeeding could be a source of increased ammonia levels-and hence toxicity-present in the tanks).

14. At initiation of the experiment on day 0, five females from the stock population were sampled for obtaining a baseline on morphometric data and spiggin kidney levels. Their mean wet weight at the beginning of the test was 1.62 ± 0.1 g.

Exposure

15. The chemicals were added to the flow through system into the experimental tanks six days prior to the transfer of the fish to ensure saturation on the aquaria and fine tuning of the system. There were 7 experimental groups, each group comprising of 2 replicates with 6 females in each (Table 2). The AFSS Guidance Document recommends the use of 5 fish per experimental tank (10 per treatment level). In the testing of ammonia, this number was increased to 6 per experimental tank (2 per treatment) to ensure statistical robustness of the data due to potential mortalities as a result of ammonia toxicity.

16. Mixing vessels were employed to deliver the desired chemical concentrations to the tanks via gravity. The delivery of chemicals (or water and solvent controls) to the mixing vessels were achieved

with the use of accurately calibrated peristaltic pumps. The dilution water was delivered by means of accurately calibrated rotameters. The flow rate expected for each replicate was 65 ± 5 ml/min (i.e. ca. 12 changes of water/day). Fresh working solutions of the chemicals were prepared every two days.

Table 2: Treatment allocation for the AFSS using ammonia as a negative substance

Label	Treatment			No. of replicates	No. fish/tank	No. fish/ treatment
	DHT	Ammonia				
		as Total ammonia (TA-N) [a]	as Free ammonia (NH_3)			
WC	Water control, no chemical or solvent			2	6	12
SC	Methanol at 83µl/L			2	6	12
DHT Control	5µg/L	-	-	2	6	12
NH_3 Control	-	Max 12.75 mg/L	0.25 mg/L	2	6	12
NH_3 Low	5µg/L	Max 3 mg/L	0.06 mg/L	2	6	12
NH_3 Medium	5µg/L	Max 6.6 mg/L	0.13mg/L	2	6	12
NH_3 High	5µg/L	Max 12.75 mg/L	0.25 mg/L	2	6	12
					Total	84

[a] adjusted depending on pH and temperature to keep the free ammonia concentration stable

Preparation of the chemicals

17. DHT (Sigma Chemicals, Pool, UK, Cat No: A8380), was used at 5µg/L in the experimental tanks as described in the AFSS. The solvent vehicle was methanol at a final concentration in the aquaria of *ca.* 83 µg/L, which is below the maximum solvent concentration 100µg/L as recommended by the OECD. In respect to the preparation of ammonia solutions, the VMG-eco agreed on a mixture of ammonium chloride

and ammonium hydroxide in ratio of 45:1 to achieve the desired stock solution concentration under neutral pH. No detailed information on solution preparation or analysis was provided at the time.

18. After contacting the Fish Sexual Development Test (FSDT) lead laboratory in Denmark (DHI), that was testing ammonia as an endocrine negative using the zebrafish during the same period, some useful information was obtained in respect to the preparation of the ammonia stock solution.

19. A stock solution at 72g TA-N/L was prepared every 5 days using ammonium chloride (*ca*. 206 g in 1L of distilled water) and ammonium hydroxide (*ca*. 5ml in 1L). Every two days, fresh working solutions were prepared and their concentrations were adjusted according to the pH and the temperature in the aquaria. For achieving 0.25mg NH_3/L in the tanks, the range of working solution concentrations used was 5.2-12.75mg TA-N/L, for 0.13mg NH_3/L it was 2.7-6.6mg TA-N/L and for 0.06mg NH_3/L it was 1.3-3mg TA-N/L.

Termination

20. Three days prior to the planned termination, increased mortalities in the aquaria with the highest ammonia concentration were observed. Anticipating that the mortalities were going to continue in this treatment level it was decided to terminate the experiment on day 18 instead of day 21. Hence the duration of the ammonia exposure was 18 instead of 21 days, which is a slight deviation from the AFSS protocol. This was not considered a major issue since induction of spiggin by DHT at 5µg/L is evident as early as 7 days and very prominent after 14 days of exposure (see figure 5, Katsiadaki *et al*, 2002). Therefore, there was no doubt that the slightly shorter duration of exposure has not affected the *in vivo* responses.

21. On day 18, the fish were terminally anaesthetised by MS-222 and placed in liquid nitrogen. They were stored at -80ºC until dissection. The kidneys were dissected out and placed in 200µl of a strong urea buffer for digestion prior to spiggin analysis according Katsiadaki *et al* (2002).

Chemical analysis

22. At least once a week two batches of water samples were collected from each tank; one was used for DHT analysis and the other for total ammonia determination. For DHT, 1 ml was collected from all the tanks (mixing vessels and experimental tanks) into an eppendorf and stored at -20ºC until radio-immunoassay (RIA) analysis. The general procedure has been established for the measurement of several steroids in the same laboratory (Scott *et al*, 1994) and requires no water extraction that can lead to test compound losses. Only 10µl were added to 90µl RIA buffer (0.5M phosphate buffer containing 0.2% bovine serum albumin, 0.8% sodium chloride, 0.03% EDTA and 0.01% sodium azide) for the analysis.

23. For ammonia, 5ml were collected from all the tanks (mixing vessels and experimental tanks), and at the same time the pH and the temperature were recorded. Total ammonia was measured by adding the 5ml of aquaria water collected into an ammonium cuvette test (Hach Lange Ltd), containing all necessary reagents. The cuvette was shaken for 30 seconds then left for 15 minutes and then inserted into a portable spectrophotometer (DR2800, Hach Lange Ltd). The ammonia concentration range that can be measured by this method is 0.015-2mg TA-N/L.

RESULTS AND DISCUSSION

GENERAL

24. Upon dissection, the sex of each fish was identified by visual inspection of the gonads. No male fish were present in the experimental population as intended. At the end of the study, the mean wet weight of the fish was 1.58 g and was within ± 30% of the mean wet weight from the sub-sample of the stock.

25. The mortalities were below 10% in each treatment group with the exception of the NH_3 high concentrations (NH_3 Control and NH_3 High) where 25 and 33.3 % mortality, respectively were observed (Table 3). Besides best efforts to balance the VMG-Eco recommendations on ammonia concentration range and the reported toxicity, this was only partially successful. The data are in line with the results of the FSDT on ammonia using zebrafish. In the FSDT, the ammonia concentration range suggested by VMG-Eco was proved very toxic resulting in 100% larval mortality at 0.5mg/L NH_3/L (pH 7.7 and 27°C) whilst only 32 % of the larvae survived the exposure at 0.25mg NH_3/L (pH 7.7 and 27°C; Mette Albrektsen, DHI, personal communication).

Table 3: Stickleback mortalities during the ammonia exposure

	No of surviving fish	**No mortalities**	**% Mortality**
Water Control	11	1	8.33
Solvent Control	11	1	8.33
NH_3 Control	8	4	33.33
DHT Control	11	1	8.33

NH_3 Low	11	1	8.33
NH_3 Medium	11	1	8.33
NH_3 High	9	3	25.00

Chemical analysis

26. Table 4 displays in detail the measurements of both chemicals, temperature and pH data during the exposure. The ammonia concentration was constantly kept within the ±20% of the mean measured concentration and close to the nominal concentration at all times for all treatment levels. The minor amounts of free ammonia recorded in the tanks where no ammonia was added were all below the quantification limit of the assay and are within the expected range for fish aquaria.

27. The DHT concentrations were kept very close to the nominals. Although in a few occasions the measured concentrations were just outside ±20% of the mean measured concentration (all measurements were within ±30% of the mean measured concentration). Most importantly the mean values of the three ammonia treatments levels had highly comparable mean values for DHT.

Table 4: Measured concentrations of DHT and free ammonia (NH_3) along with pH and temperature records during the ammonia exposure.

Nominal concentration	Sample	DHT (in µg/L)	NH_3 (in mg/L)	Temperature(in °C)	pH
Water Control	1	<LOQ	0.003	16.6	8.3
	2	<LOQ	0.001	17.8	8.2
	3	<LOQ	0.001	17.4	8.2
	4	<LOQ	0.001		
	Mean		**0.002** [a]	**17.3**	**8.3**
Solvent Control (Methanol)	1	<LOQ	0.001	16.4	8.3
	2	<LOQ	0.000	17.6	7.9
	3	<LOQ	0.000	17.2	7.9
	4	<LOQ	0.000		
	Mean		**0.000** [a]	**17.1**	**8.0**
NH_3 Control (0.25 mg/L)	1	<LOQ	0.233	16.8	8.3
	2	<LOQ	0.269	17.7	8.0
	3	<LOQ	0.215	17.5	7.8
	4	<LOQ	0.258		
	Mean		**0.244**	**17.3**	**8.0**
DHT Control (5 µg/L)	1	3.79	0.001	16.6	8.3
	2	3.67	0.000	17.9	7.9
	3	4.80	0.001	17.6	7.9
	4	4.78	0.000		
	Mean	**4.26**	**0.001** [a]	**17.3**	**8.0**

NH_3 Low DHT at 5 µg/L + NH_3 at 0.06 mg/L	1	3.01	0.033	16.6	8.2
	2	3.10	0.048	17.7	7.9
	3	4.49	0.048	17.3	7.9
	4	5.26	0.052		
	Mean	**3.96**	**0.045**	**17.2**	**8.0**
NH_3 Medium DHT at 5µg/L + NH_3 at 0.13 mg/L	1	2.60	0.079	16.8	8.3
	2	4.96	0.087	18.0	7.9
	3	3.11	0.106	17.6	7.9
	4	4.70	0.106		
	Mean	**3.84**	**0.094**	**17.4**	**8.0**
NH_3 Low DHT at 5 µg/L + NH_3 at 0.25 mg/L	1	3.10	0.176	16.7	8.2
	2	3.42	0.236	17.8	8.0
	3	2.96	0.208	17.7	7.9
	4	4.93	0.215		
	Mean	**3.60**	**0.209**	**17.4**	**8.0**

Highlighted lines are outliers from the mean treatment concentration ± 20%.

[a] : recorded data were below the quantification limit of the assay.

Spiggin

28. The responses of fish to the treatments by means of kidney spiggin content are displayed in Figure 1. Spiggin data were logarithmically transformed before analysis. One-way ANOVA of ln(spiggin) showed that exposure to ammonia had no effect on spiggin induction by DHT. By comparing the spiggin values between groups where DHT was present (DHT control, DHT+ NH_3 Low, DHT+ NH_3 Medium, DHT + NH_3 High) and groups where DHT was not present (day 0 fish, water control, solvent control, NH_3 Control), a massively significant difference between non DHT-treated groups and DHT-treated groups, and the complete lack of significance between ammonia treatments within the groups ($p = 0.5251$; see Annex 1) were found.

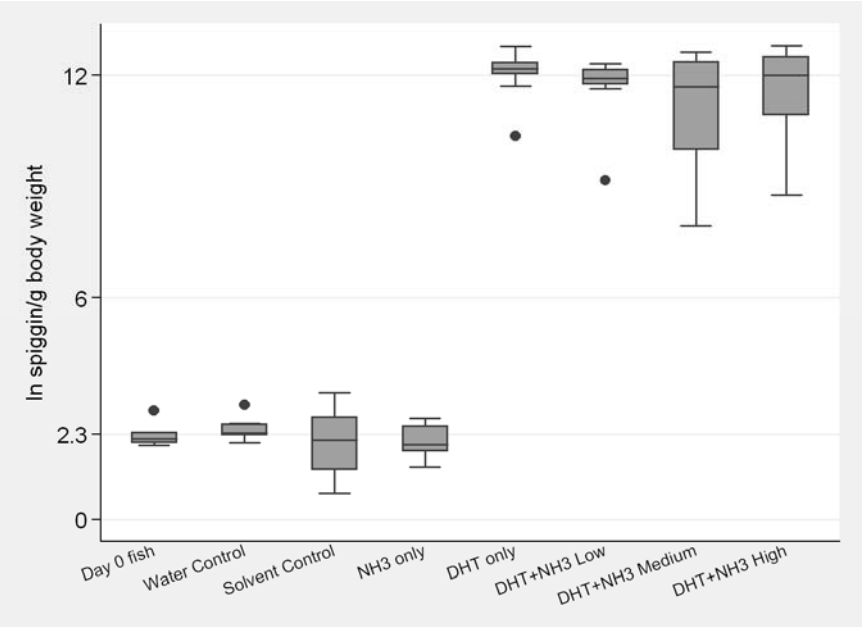

Figure 1. Spiggin responses of female sticklebacks after exposure to DHT alone or in combination with free ammonia (NH_3)

29. In this AFSS, it is hugely important that the DHT measured concentrations in the DHT control and the three treatment levels of test compound are highly comparable. If for example DHT is lower in the high test compound concentration in comparison to the DHT control, the test might produce a false positive result (lower spiggin induction, hence an anti-androgenic effect). Equally, if DHT is higher in the

high test compound concentration in comparison to the DHT control, the test might produce a false negative result (masking the antiandrogenic effect of the test compound by producing spiggin values that are comparable with the DHT controls). Although this was not the case in the testing of ammonia as a negative substance, as chemical analysis confirmed (table 4), further analysis were conducted to demonstrate the principle.

30. Spiggin levels were adjusted upwards when the measured DHT was further from the nominal using the equation 'ln(spiggin) = 8.2 + 0.5 DHT' from the AFSS validation report using the established linear relationship between the measured ln(spiggin) response and the actual DHT concentration. Figure 2 shows each ln(spiggin) adjusted to the nominal DHT at 5µg/L (pink boxes). An ANOVA confirmed that the differences between NH_3 treatment and spiggin levels were not statistically significant (p=0.4339; see Annex 1).

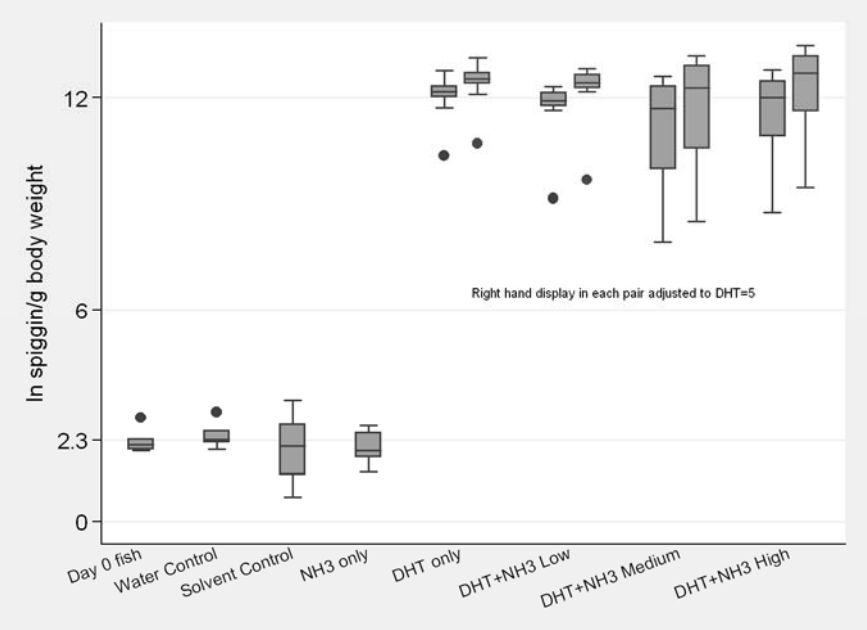

Figure 2: Spiggin responses of female sticklebacks after exposure to DHT alone or in combination with free ammonia (NH_3). Pink boxes represent the ln(spiggin) adjusted to the nominal DHT concentration of 5 µg/L.

31. Finally, in order to visualise the lack of effect of NH_3 exposure to the AFSS response, the spiggin results of the ammonia test were plotted over the spiggin results of all the AFSS validation data that included four known antiandrogens (Figure 3).

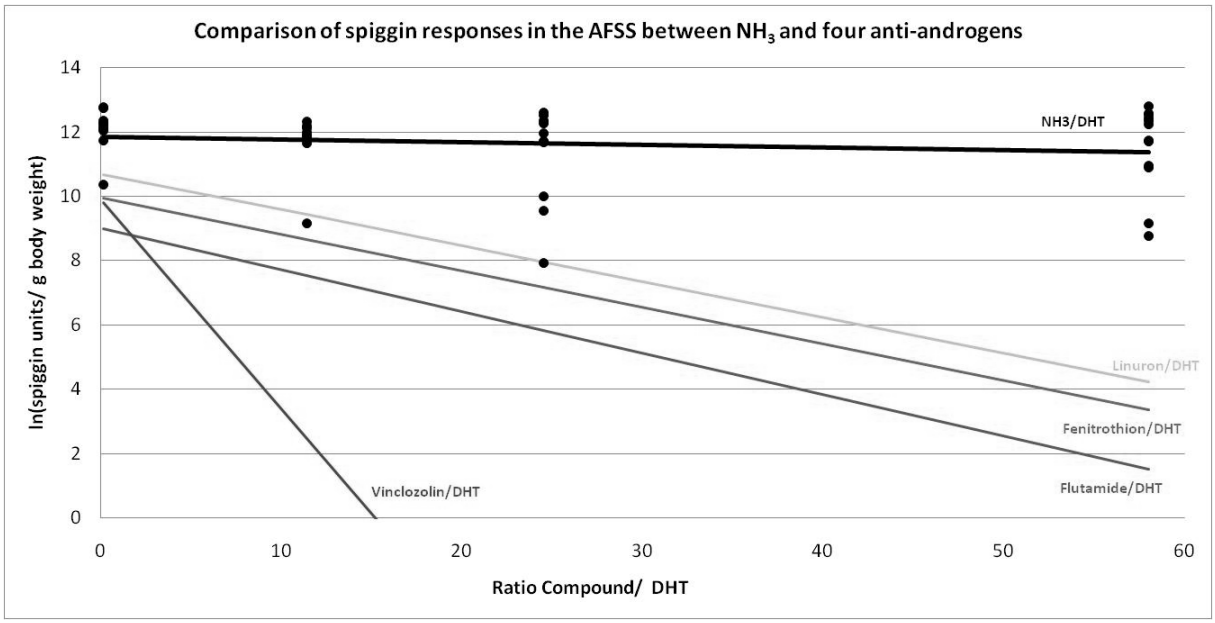

Figure 3: Scatter plot of the spiggin AFSS validation data using four antiandrogens and the ammonia data.

32. As figure 3 displays previous testing of all four antiandrogens (Flutamide, Vinclozolin, Fenitrothion and Linuron) produced negative slops in spiggin responses as the ratio between the test compound and DHT increased in the exposure tanks. This negative slope is indicative of antiandrogenic activity as it shows that spiggin induction by DHT is hampered by the test compound in a dose response manner. This was not the case for ammonia, demonstrating the lack of anti-androgenic activity; Ammonia did not induce spiggin in the female fish either when tested alone (NH_3 control, see figure 1 and 2), which denotes the lack of androgenic activity.

33. In conclusion, the results demonstrated that ammonia has no endocrine activity (androgenic or anti-androgenic) when tested in the AFSS, an expected result that completes the dataset presented in the AFFS validation report. This outcome provides further evidence on the robustness of AFSS for the screening of endocrine active compounds.

REFERENCES

Emerson, K., Lund, R. Thurston, R., Russo R. (1975), Aqueous ammonia equilibrium calculations: effect of pH and temperature, Journal of the Fisheries Research Board of Canada 32: 2379–2383.

Environment Canada (2001), Priority substances assessment report: ammonia in the aquatic environment, Minister of Public Works and Government Services Canada, Ottawa, Ontario.

Hill, E.M., Evans, K.L., Horwood, J., Rostkowski, P., Oladapo, F.O., Gibson, R., Shears, J.A., Tyler, C.R. (2010), Profiles and some initial identifications of (anti)androgenic compounds in fish exposed to wastewater treatment works effluents Environmental Science and Technology, 44 (3): 1137-1143.

Hazel, C.R., Thomsen W., Meith S.J. (1971), Sensitivity of striped bass and stickleback to ammonia in relation to temperature and salinity, California Fish and Game 57: 138-153.

Johnson, I., Hetheridge, M., Tyler, C.R. (2007), Assessment of (anti-) oestrogenic and (anti-) androgenic activities of final effluents from sewage treatment works, Science report-SC020118/SR, Environment Agency, Bristol, UK, page 61.

Katsiadaki, I., Scott, A.P., Hurst, M.R., Matthiessen, P., Mayer, I. (2002), Detection of environmental androgens: a novel method based on enzyme-linked immunosorbent assay of spiggin, the stickleback (*Gasterosteus aculeatus*) glue protein, Environmental Toxicology and Chemistry 21:1946-1954.

Katsiadaki, I., Morris, S., Squires, C., Hurst, M.R., James, J.D., Scott, A.P. (2006), A sensitive, *in vivo* test for the detection of environmental anti-androgens, using the three-spined stickleback (*Gasterosteus aculeatus*), Environmental Health Perspectives, 114 (suppl. 1):115-121.

Katsiadaki, I., Sanders, M.B., Sebire, M., Nagae, M., Soyano K., Scott, A.P. (2007), Three-spined stickleback: an emerging model in environmental endocrine disruption, Environmental Sciences 14 (5):263-283.

OECD (2006), Report of the validation of the 21d fish screening assay for the detection of endocrine substances (Phase 1B), Series on Testing and Assessment No. 61, ENV/JM/MONO(2006)29, OECD, Paris.

Scott, A.P., MacKenzie, D.S., Stacey, N.E. (1984), Endocrine changes during natural spawning in the white sucker, Catostomus commersoni, II. Steroid hormones. General and Comparative Endocrinology, 56(3):349-359.

Thurston, R.V., Russo, R.C. (1983), Acute toxicity of ammonia to rainbow trout, Transactions of the American Fisheries Society 112: 696-704.

Thurston, R.V., Russo, R.C., Phillips, G.R. (1983), Acute toxicity of ammonia to fathead minnows. Transactions of the American Fisheries Society 112: 705-711.

Tollefsen, K.E., Harman, C., Smith, A., Thomas, K.V. (2007), Estrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms, Marine Pollution Bulletin 54:277-283.

Urbatzka, R., van Cauwenberge, A., Maggioni, S., Viganò, L., Mandich, A., Benfenati, E., Lutz, I., Kloas, W. (2007), Androgenic and antiandrogenic activities in water and sediment samples from the river Lambro, Italy, detected by yeast androgen screen and chemical analyses, Chemosphere 67:1080-1087.

U.S. Environmental Protection Agency (USEPA) (1999), Update of ambient water quality criteria for ammonia — 1999, Criteria and Standards Division, Washington, D.C. (EPA-822-R-99-014).

ANNEX 1: STATISTICAL TABLES

Statistical analysis of data presented in figure 1

The anova for the raw (ln) results shows the massively significant difference between controls and DHT-treated groups, and the complete lack of significance between treatments within the groups.

. anova lnspiggin control ordtreat

Number of obs = 83 R-squared = 0.9660

Root MSE = .927842 Adj R-squared = 0.9628

Source	Partial SS	df	MS	F	Prob > F
Model	1833.52294	7	261.931848	304.26	0.0000
control	509.970352	1	509.970352	592.38	0.0000
ordtreat	4.4653551	6	.744225851	0.86	0.5251
Residual	64.5667389	75	.860889852		
Total	1898.08968	82	23.1474351		

Statistical analysis of data presented in figure 2

The DHT-treated groups were then adjusted to DHT=5, using the equation

gen adjlnspiggin = lnspiggin + .5*(5-dhtm) if dhmt>0

and an anova comparing just the DHT treated groups shows

. anova lnspiggin ordtreat if control==0

```
                   Number of obs =      45          R-squared     =  0.0638
                   Root MSE      = 1.14863   Adj R-squared = -0.0047

  Source |  Partial SS       df      MS             F          Prob > F
-----------+----------------------------------------------------------
   Model |   3.6888287     3  1.22960957       0.93     0.4339
         |
ordtreat |   3.6888287     3  1.22960957       0.93     0.4339
         |
Residual |  54.0937067    41   1.3193587
-----------+----------------------------------------------------------
   Total |  57.7825354    44  1.31323944
```