EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

**Institute for Health and Consumer Protection**
**European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)**

**ECVAM**
**SCIENTIFIC**
**ADVISORY**
**COMMITTEE**
**(ESAC)**

# ESAC Working Group Peer Review Consensus Report

## on the validity of the zebrafish embryo toxicity test for acute aquatic toxicity testing

# TABLE OF CONTENTS

# ESAC Working Group

This report was prepared by the "ESAC Working Group ZFET" (ESAC WG), charged with conducting a detailed scientific peer review of

(a) prospective study, coordinated by ECVAM on behalf of the OECD, concerning the transferability and reproducibility of the Zebrafish Embryo Toxicity Test (ZFET) for acute aquatic toxicity testing and

(b) a retrospective analysis of the correlation of fish embryo toxicity data versus acute fish toxicity data in view of determining whether the Fish Embryo Toxicity test (FET), including the ZFET variant, could be used as an alternative, i.e. to substitute, the acute fish toxicity test. This latter analysis has been compiled by Proctor & Gamble (S. Belanger and colleagues) and was submitted to ECVAM for evaluation and ESAC peer review in summer 2012.

The ESAC WG was set up by written procedure in August 2012. Basis for the scientific review was the ECVAM request for ESAC advice on the validity (reliability and relevance) of the zebrafish embryo toxicity test for acute aquatic toxicity testing (ESAC request 2012-03).

The ESAC WG conducted the peer review from October 2012 to February 2013. This report was endorsed by the ESAC WG on 18. February 2013 and represents the consensus view of the ESAC WG.

This ESAC WG peer review consensus report was endorsed by the ESAC on 15. March 2013.

<u>The ESAC WG had the following members:</u>

- Dr. Andrea Seiler, Chair (ESAC member, Chair of the WG)
- Dr. Neil Carmichael (ESAC member)
- Prof. Lucio G. Costa (ESAC member)
- Prof. Kristin Schirmer (ESAC member)

<u>EURL ECVAM:</u>

- Dr. Claudius Griesinger *(EURL ECVAM Coordinator for ESAC Peer Reviews and ECVAM recommendations)*
- Dr. Marlies Halder *(Project Leader (Z)FET Study)*

## NOTE ON THIS REPORTING TEMPLATE

The template follows the ECVAM modular approach and allows at the same time for the description of the analysis and conclusions concerning more specific questions. The template was approved by the ESAC through written procedure on 29 October 2010.

The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies).

Depending on the study type and the objective of the study, not all sections may be applicable. However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

- **Explanatory notes to the paragraph titles (in green)** provide guidance on the type of information / analysis expected under each section. Depending on the purpose and scope of the study to be reviewed, some of the aspects mentioned in the explanatory notes may not be applicable or only be applicable to some extent. Moreover, the explanatory notes are not intended to represent an exhaustive list of possible issues to be addressed under the respective heading, but are thought to provide some guidance with respect to the considerations typically expected.

## ABBREVIATIONS USED IN THE DOCUMENT

- **AFT** Acute Fish Toxicity
- **CSG** Chemical Selection Group
- **CV** Coefficient of variation
- **3,4-DCA** 3,4-dichloroaniline
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **ESAC** ECVAM Scientific Advisory Committee
- **ESAC WG** ESAC Working Group
- **FET** Fish Embryo Toxicity Test
- **GLP** Good Laboratory Practice
- **SOP** Standard Operating Procedure (used here as equivalent to 'protocol')
- **VMG** Validation Management Group
- **VSR** Validation Study Report
- **WLR** Within-laboratory reproducibility
- **ZFET** Zebrafish Embryo Toxicity Test

# Executive summary

Following a request from ECVAM to ESAC for peer review of and scientific advice on an ECVAM-coordinated validation study concerning the use of fish embryos as alternative to juvenile and adult fish acute lethality test, an ESAC Working Group (ESAC WG) was set up by ESAC. The ESAC WG was charged with conducting a detailed scientific peer review of two studies which had addressed (i) prospectively, the within- and between-laboratory reproducibility of the zebrafish embryo toxicity test (ZFET) and (ii) retrospectively, the ability of the fish embryo toxicity test (FET) to predict acute fish toxicity based on peer-reviewed and otherwise available data.

The ESAC WG met in person at ECVAM on Nov.20/21 2012 and communicated further by email and teleconferences in December 2012 and January/February 2013. The ESAC WG reviewed the following documents:

1. Two reports resulting from the prospective OECD validation study conducted in two distinct phases and coordinated by ECVAM, "Validation report (Phases 1a and 1b) for the zebrafish embryo toxicity test (comprising two parts)" from August 25/2011 and "Validation report (Phases 2a and 2b ) for the zebrafish embryo toxicity test" from August 10/2012, as well as all the related annexes;
2. The retrospective analysis provided by Belanger et al. entitled, "An update to the fish embryo toxicity-acute fish toxicity relationship and prospects for support of the use of the FET as an animal alternative" from February 22/2012, which is a document developed for the February, 2012 meeting of the OECD ad hoc expert group on the fish embryo test.

The goal of the prospective study was to evaluate the transferability and within- as well as between-laboratory reproducibility of the zebrafish embryo toxicity test (ZFET), for which an OECD draft guideline (status May 2008) and a draft standard operating procedure (SOP) already existed. The draft SOP was provided by the lead laboratory, which has significant experience in the ZFET and was instrumental in developing the first draft OECD guideline. Prior to the start of the prospective study, the validation management group (VMG) reviewed the SOP and incorporated the concerns expressed by the OECD *ad hoc* expert group FET (a group established by OECD to develop the new OECD guideline "Fish embryo test"), e.g. the number of embryos per test concentration and control was increased from 10 to 20, the extension of the test to 96 h to provide a time period in which the embryo was no longer protected by the chorion, and an acceptance criterion was set for the fertility rate. Moreover, during Phase 1a the pre-saturation of exposure vessels and daily renewal of test concentrations so as to promote establishment of chemical equilibrium during exposure was introduced into the SOP. In Phase 1a of the prospective study, the SOP was transferred to seven participating laboratories, all testing one chemical (3,4-dichloroaniline; 3,4-DCA), which served as a positive control throughout. The transfer of the SOP was successful, with promising within- and between-laboratory reproducibility. In addition, a test concentration for the positive control was derived for which a minimum of 30% lethality over the 96 h can be expected (in contrast to the 10% lethality initially required in the OECD draft guideline, which is difficult to distinguish statistically from a lack of lethality). Then in Phase 1b, six laboratories were involved in testing six additional chemicals, which were selected based on a wide range of toxicity. Overall, the ZFET was successfully transferred to the participating laboratories with rather good within- and between-laboratory reproducibility for five of the six chemicals (CV<30%). The lower reproducibility of the sixth chemical was attributed to its high volatility and indeed, significantly lower test concentrations were analytically confirmed in one laboratory in one run.

The goal of the second phase of the prospective study was to increase confidence in the very good within- and between-laboratory reproducibility by testing 13 additional chemicals. In Phase 2a, newly joining laboratories went through a training phase using 3,4-DCA as in Phase 1a before. Then, nine laboratories, trained either in Phase 1a or Phase 2a, were involved in testing the 13 additional chemicals. These chemicals were selected based on rational criteria, which included, aside from a wide range of toxicity, different physico-chemical properties and modes of toxic action. As well, a coefficient of variation (CV) below 30% was agreed upon as an acceptable within- and between-laboratory reproducibility. The testing results confirmed that the ZFET SOP could be transferred successfully to new laboratories and that the within- and between-laboratory reproducibility is indeed generally below 30%. A higher CV for three chemicals was attributed to very steep concentration-response curves. Moreover, analytical measurements were performed in two laboratories, demonstrating that (i) quantitative chemical analysis can be performed despite the small testing volumes, and that (ii) chemicals with a combination of low water solubility, high potential to be biotransformed, and volatility, comprise the most challenging to test.

Overall, the ESAC WG considered the scientific work presented for the prospective study of very high quality. The rational for the testing design and chemical selection was well described and the results thoroughly evaluated. The conclusions drawn are very well justified and supported by the data. A minor weakness in the report was the inconsistent assignment of toxicity classes to the chosen chemicals, likely due to the utilisation of different sources of information.

The retrospective analysis provided by Belanger et al. (2012) comprised an exhaustive update of the correlation of the fish embryo toxicity test and the acute fish toxicity test. It takes into account previously published comparisons (e.g., the review by Lammer et al. (2009) on the ZFET) but adds any information available until the completion of the report (Feb/2012): peer-reviewed and open literature for the ZFET and fish embryo test (FET) data obtained for other species of fish; and data made available by a number of groups prior to peer-reviewed publication. In fact, the final database for evaluating the relationship between the FET and the acute fish toxicity consisted of 985 FET studies on 229 compounds (dominated by ZFET) and of 1531 fish acute toxicity studies on 151 compounds (dominated by rainbow trout, fathead minnow, bluegill sunfish), presenting more than double the size of the information presented in Lammer et al. (2009). The analysis was largely based on orthogonal regression analysis, which takes uncertainty of both the FET and the fish acute toxicity data into account. The analysis showed that (i) the ZFET basically responds to chemical exposure like a fish in the acute fish toxicity test – the regression lines were statistically indistinguishable, meaning that the ZFET data can be taken directly to establish the acute toxicity to fish; (ii) even though data are sparse for FETs with other species, data available thus far look similarly promising; (iii) the very good correlations hold true despite chemicals having a wide range of physico-chemical properties, toxicities, modes of toxic action, and substance classes, thus revealing a broad applicability.

The ESAC WG considered also this study of very high quality and the results supportive of the application of the ZFET as alternative to the acute fish toxicity test. Embryos of other species may be considered as alternatives as well although more studies should be conducted to confirm this. The presented analysis overall was very robust, partly because of the high number of data points; in general, no class of chemical revealed an exceptional difference to the predictivity of the method.

# 1. Study objective and design

## 1.1 Analysis of the clarity of the definition of the study objective

*NOTE: (a) please summarise briefly the study objective and (b) provide an appraisal as to whether the study objective is clearly and comprehensibly defined in the VSR?*

***(a) ESAC WG summary of the study objective as outlined in the VSR***

This evaluation considers two different, but related studies. The first focussed on the zebrafish embryo toxicity test (ZFET) and the other on fish embryo toxicity test (FET), including ZFET, as an alternative to the acute fish toxicity test outlined in OECD 203. Here and elsewhere in this report, the term alternative is used in the broad sense of the Three Rs (replacement, reduction, refinement).

The first was a prospective study, similar in scope to other "traditional" validation studies. The aim of the prospective study was to determine whether the ZFET was transferable to different laboratories, and to assess the within- and between-laboratory reproducibility of the test. After the development of Standard Operating Procedure (SOP), and a first test (which included three runs) with a positive control (3,4-dichloroaniline; 3,4-DCA) (Phase 1a), six chemicals were tested in six laboratories (Phase 1b). After results were independently reviewed and the SOP was slightly modified, Phase 2a started. This consisted in the testing of the positive control 3,4-DCA in four new laboratories and the testing of thirteen new chemicals in nine laboratories (Phase 2b). It should be noted that in both Phases not all chemicals were tested in all laboratories.

The aim of the retrospective assessment was different, in that it consisted in a review of published and unpublished data on ZFET (including those generated by the prospective study) as well as other FET results together with available data on acute fish toxicity, to assess the suitability of such tests to serve as an alternative to the acute fish toxicity test outlined in OECD 203.

***(b) Appraisal of clarity of study objective as outlined in the VSR***

The WG judged that the objectives of the study were clearly articulated. The study design of the prospective and the retrospective part of the validation study were well thought through and planned accordingly.

*The prospective study* was divided into two phases: The aim of Phase 1 was to evaluate the transferability, and the within- and between-laboratory reproducibility of the ZFET test with seven chemicals. The aim of Phase 2 was to further evaluate the within- and between-laboratory reproducibility of the ZFET with an additional thirteen chemicals covering specific areas of use (chemicals, pharmaceuticals, pesticides, biocides), a wide range of toxicity, physico-chemical characteristics, and various modes of action.

The objective of *the retrospective study* was to assess the suitability of the FET as an alternative model by analysing the correlation between data from FET and data from acute fish toxicity tests (AFT) for as many chemicals as possible. A total of 985 FET studies (229 chemicals) and 1531 acute fish toxicity studies (151 chemicals in common with the FET) were analysed by FET-AFT regressions to understand the potential relationships or biases based on physico-chemical properties, species choices, duration of exposure, chemical classes, chemical functional uses, and modes of action. This study was not a traditional validation study (i.e. validating a specific method) but rather an analysis of existing data to see how well the FET predicts, based on correlations, acute fish toxicity (measured according to OECD 203).

## 1.2 Quality of the background provided concerning the purpose of the test method

*NOTE: What is the overall purpose of the test method: a) scientific use, b) regulatory application?*

***(a) Analysis of the scientific rationale provided in the VSR***

*NOTE: Is the scientific rationale for the test method AND (consequently) for conducting the study clearly explained? How does the test method contribute to (a) the scientific understanding of the specified health/environmental effect or aspects of it? (b) the prediction of the specified health/environmental effect or aspects of it? Does the VSR make sufficient reference to the relevant body of scientific literature?*

The scientific rationale for both studies was presented in the documents which were provided to the WG. The basic questions are quite straightforward: 1) can the ZFET be implemented in different laboratories and what are the within- and between-laboratory coefficients of variation? 2) Can the FET in general predict acute fish toxicity, and thus serve as an alternative to the standard acute fish toxicity test (OECD 203)?

The zebrafish embryo toxicity test is designed to determine the lethal effects of chemicals on embryonic stages of fish and constitutes a potential alternative test method to the acute toxicity tests with juvenile and adult fish, i.e., the OECD Test Guideline 203 (OECD TG 203, 1992), thus providing a reduction in fish usage. Zebrafish embryos are individually exposed in, e.g., 24-well microtiter plates or crystallisation dishes. The test is initiated immediately after fertilisation and is continued for 96 hours. Lethal effects, assessed by four apical observations (coagulation of the embryo, non-detachment of the tail, non-formation of somites, and non-detection of heart beat), are determined and used to determine acute toxicity and to calculate the LC50 value. In addition, non-hatch is recorded. The test method is based on using a minimum of five test concentrations as well as appropriate negative and positive controls. Each chemical is tested with 20 embryos per test concentration and controls. All these aspects (part of the prospective study) are scientifically sound.

Question two was adequately addressed in the document “An Update to the Fish Embryo Toxicity-Acute Fish Toxicity Relationship and Prospects for Support of the Use of the FET as an Animal Alternative” (Belanger et al. 2012), and in other scientific publications (e.g. Lammer et al., 2009; Knöbel et al. 2012). The data analysed in the two latter studies (Lammer et al., 2009; Knöbel et al. 2012) were already included in the Belanger et al. (2012) study. This thorough evaluation of existing data demonstrates that the zebrafish embryo toxicity test correlates well with acute fish toxicity.

***(b) Analysis of the regulatory rationale provided in the VSR***

*NOTE: Is a regulatory rationale specified, i.e. a specific application of the test method for purposes of generating data with respect to regulatory requirements as specified in legislation or internationally agreed guidelines etc.? If so, how does the study and its objective and design relate to this regulatory rationale? Are the relevant regulatory documents appropriately referenced?*

The relevant regulatory documents were appropriately referenced in the VSR. The main issue is that a number of legislations require acute fish toxicity tests, and the use of ZFET (or FET) would significantly lower the number of fish used for such toxicity testing. The fish embryo toxicity test is designed to determine the lethal effects of chemicals on embryonic stages of fish and constitutes a potential alternative test method to the acute toxicity tests with juvenile and adult fish, thus providing a reduction in fish usage.

The acute fish toxicity test is a mandatory component in the environmental safety assessment of industrial chemicals, agrochemicals, biocides, pharmaceuticals, feed stuff etc. In the European Union, Directive 2010/63/EU on the protection of animals used for scientific purposes (EU, 2010) demands

that tests on vertebrate animals are reduced, refined or replaced whenever possible. In addition, the legislation on chemicals (REACH) foresees that animal studies are only to be conducted if the data gaps can not be adequately addressed by other means (e.g. *in vitro* studies, etc.). In Germany, the Fish Egg Toxicity test (DIN 2001) was validated and replaced the 48 h acute fish test for routine whole effluent testing in 2005. Recently, a modified international version of the fish egg toxicity test was published (ISO 2007).

## 1.3 Appraisal of the appropriateness of the study design

*NOTE: This includes an analysis of the number of laboratories involved in the study, the organisation of study management, of statistical analysis and may include more technical aspects such as (a) a brief appraisal of the nature and number of test items used (details however to be provided in section 6, test items), retesting in case of unqualified tests and others.*

Since the two studies were quite different with regard to objectives, design and approach, two separate appraisals by the WG are provided below.

### 1.3.1 The prospective study

The prospective study was divided into two phases. The aim of Phase 1 was to evaluate the transferability, and the within- and between-laboratory reproducibility of the zebrafish FET (ZFET) with six chemicals. The aim of Phase 2 was to further evaluate the within- and between-laboratory reproducibility of the ZFET with an additional thirteen chemicals covering specific areas of use (chemicals, pharmaceuticals, pesticides, biocides), a wide range of toxicity, and various modes of action.

Based also on specific questions indicated in the EURL ECVAM request for ESAC advice, the WG addressed a number of different issues which are indicated below.

A. Selection of chemicals

A total of six chemicals were selected for Phase 1b and an additional thirteen chemicals were selected for Phase 2b. In addition, 3,4-DCA was utilised as a positive control for both Phase 1a and Phase 2a. For Phase 1, no detailed selection criteria were included in the VSR, and it was only stated that "Chemicals were selected based on the recommendations of the ad hoc Expert Group (see Minutes of the meeting in May 2008). The University of Heidelberg purchased and distributed 3,4-DCA to the laboratories, whereas ECVAM purchased, aliquoted and distributed the six chemicals tested in Phase 1b". Upon a request from the WG, the VMG provided additional details, and indicated that a preliminary list of 27 chemicals and a list of 8 chemicals presented by Procter & Gamble (P&G) were considered. In the planning of Phase 1, six chemicals were chosen to cover the range of fish toxicity from non-toxic to very toxic. Two chemicals (sodium chloride, triclosan) were from the list of 27; three (dibutylmaleate, 6-methyl-5-hepten-2-one, and 2,3,6-trimethylphenol) were from the P&G list, while ethanol was chosen as analytical measurements were readily available. The WG believes that the choices are fully acceptable.

The criteria utilised for selection of chemicals for Phase 2 of the study are well documented (see Annex II, Phase 2 report), and the WG judged the criteria for selecting the chemical very appropriate for Phase 2 of the study. A "chemical selection group" (established by the OECD ad hoc expert group Fish Embryo Test) was charged of selecting thirteen chemicals from the previously defined list of 27 compounds. Criteria for selecting the chemicals for Phase 2 included the area of use (industrial chemicals, pharmaceuticals, pesticides, biocides); the range of fish toxicity (non-toxic, moderately toxic, toxic and very toxic); the availability of an analytical quantification method; the availability of fish data and, if possible, FET data; and the commercial availability of the chemicals. To some extent the following parameters were also considered: the chemical classes; the water solubility (use of

solvent); and the mode of action (either known or as predicted by other available tools such as the OECD QSAR tool box [OASIS] or USEPA ECOSAR).

Overall, 20 chemicals (including the positive control) were tested in eleven participating laboratories covering specific areas of use, a wide range of toxicity and various modes of action. The WG considered the number of test items and the number of independent runs/laboratory per chemical (N=3) sufficient to draw conclusions on the reproducibility. The distribution of chemicals ensured that each chemical was tested in at least 4 (Phase 1) or at least 3 (Phase 2) laboratories.

B. General study design

The WG group judged the design of the study as appropriate for evaluating transferability, within-laboratory reproducibility, and between-laboratory reproducibility. Assessing the predictive capacity of the ZFET was not an objective of the study.

Transferability was done in two steps in Phase 1. First, the SOP was transferred using one chemical to see if any amendments to the SOP were needed (the chemical was 3,4-DCA, which was used throughout as a positive control). Subsequently, six additional chemicals were tested by the participating laboratories.

For each phase acceptance criteria were well defined in the SOPs. For a run to be considered qualified, the following criteria were applied: (1) The fertility rate of the parent generation should be ≥70%; (2) The dissolved oxygen concentration should be ≥80 % of the air saturation value at the beginning of the test; (3) The water temperature should be maintained at 26 ± 1 °C in test chambers at any time during the test; (4) Overall survival of embryos in the negative external control (and, where relevant, in the solvent control) should be ≥90% until the end of exposure; (5) Exposure to the positive control (e.g., 4.0 mg/L 3,4-DCA) should result in a minimum mortality of 30 % at the end of the exposure; and (6) Controls and test solutions must be renewed on a daily basis.

The WG noted that the latest version of the SOP indicates that “If more than one dead embryo is observed in the negative internal control, the plate might be rejected” (this was not indicated in the SOP for Phase 1). This criterion was phrased explicitly in the OECD draft guideline (July 2012). The WG discussed whether the term “might” ought to be changed to a stronger one (e.g. shall or must). The December 2012 version of the draft OECD guidelines on FET (that went into commenting) stated that “If more then 1 dead embryo is observed in the internal plate control, the plate is rejected, thus reducing the number of concentrations used to derive the LC50”. Thus, the WG considers that its concern has been already addressed.

Prior to statistical analysis, the data underwent a quality check by the coordinator and by an independent person, who checked whether complete information was provided and whether the runs met the acceptance criteria as described in the SOP. In Phase 1b the laboratories provided data of 81 runs to the coordinator using the updated reporting template (see Annex I of Phase 1 report). For Phase 2b the laboratories provided data of 153 runs. The results (also of failed experiments) were reported using the reporting template. Data not taken into account for a given analysis were correctly marked in the report. In this second phase, with one exception (methylmercury (II) chloride, which was tested by three labs), each chemical was tested in four out of nine laboratories that participated in Phase 2 (so not each chemical was tested by all labs). See Tables 1 (from VSR, Phase 1, p. 35) and 2 (from VSR, Phase 2, p. 33) for the distribution of chemicals among laboratories. Though the design involved the testing of chemicals by only a subset of all laboratories (due to the fact the laboratories did not receive any funding and could thus only test the number of chemicals for which they had resources available), the results are very good, which speaks for the robustness of the ZFET.

Table 1. Design of chemical distribution for Phase 1b (as provided in VSR Phase 1, p. 35)

| | | Laboratories* | | | | | |
|---|---|---|---|---|---|---|---|
| **Toxicity** | **Chemicals** | **A** | **B** | **C** | **D** | **F** | **G** |
| Very toxic to fish *(LC50 < 1 mg/L)* | Triclosan | | X | X | | X | X |
| Toxic to fish *(LC50 from 1 to 10 mg/L)* | Dibutyl maleate | X | | X | X | X | X |
| Moderately toxic to fish *(LC50 from 10 to 100 mg/L)* | 2,3,6-Trimethylphenol | X | | X | X | X | X |
| | 6-Methyl-5-heptene-2-one | | X | X | | X | X |
| Non-toxic to fish *(LC50 > 100 mg/L)* | Sodium chloride | | X | X | | X | X |
| | Ethanol | X | | X | X | X | X |

*: Laboratory E did not participate in Phase 1b

In Phase 2, all chemicals (not coded) were tested at five different concentrations in three independent runs in four laboratories (except for methylmercury (II) chloride which was assessed in three laboratories) with appropriate controls. Stock solutions and test concentrations were analytically confirmed for carbamazepine, prochloraz, 1-octanol, copper (II) sulfate pentahydrate and tetradecyl sulfate sodium salt (see Table 2).

Table 2. Design of chemical distribution for Phase 2b (as provided in VSR Phase 2, p. 33)

| | | Laboratories | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Chemicals** | **Fish Toxicity** | **B** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** |
| Methylmercury (II) chloride[1, 2] | +++ | | X | | X | | | | X | |
| Copper (II) sulfate pentahydrate [1] | +++ | | | | X | X[3] | X | X | | |
| 4,6-Dinitro-*o*-cresol | +++ | | | X | X | | X | | | X |
| 2,4-Dinitrophenol | +++ | | X | | X | | | X | | X |
| Merquat 100 | ++ | | | | X | | X | | X | X |
| Luviquat HM 552 | ++ | | | | X | | X | X | | X |
| Tetradecyl sulfate sodium salt [1] | ++ | X | | | X | X[3] | X | | | |
| Malathion | ++ | X | | | X | | | | X | X |
| Prochloraz | ++ | | | X[3] | X | | X | X | | |
| 1-Octanol | + | X | | | X | X[3] | X | | | |
| Carbamazepine | + | | X | X[3] | X | | | | | X |
| Dimethyl sulfoxide | - | | | | X | | X | | X | X |
| Triethylene glycol | - | X | | X | X | | | X | | |

1) Methylmercury (II) chloride; copper (II) sulfate pentahydrate and tetradecyl sulfate sodium salt are listed according to the fish toxicity of their soluble form.
2) Tested only in three laboratories; due to safety reasons no further laboratory could test the chemical.
3) Analytical measurements of stock solutions and test concentrations were carried out. Reports on the analytics are available in Annexes IIIa and IIIb.

### 1.3.2. The retrospective study

The analysis of the correlation between FET and AFT (provided by the document authored by Belanger et al. 2012) was extremely wide-ranging and thorough. Data inclusions and exclusions were described in detail. Sound rational explanations were always given.

The report of Belanger et al. (2012, p. 11) indicates that "*the data for FET were derived from numerous sources including studies from Lammer et al. (2009) and Braunbeck and Lammer (2005); studies conducted under the oversight of the OECD Validation Management Group to assess the transferability, and within- and between-laboratory variability of the FET (the prospective study); information from the peer reviewed literature that were identified through hand searches and via literature search engines (Google Scholar, SCOPUS/SciVerse, PubMed); data made available by laboratories engaged in FET method development including those of Th. Braunbeck (University of Heidelberg), S. Scholz (UFZ-Helmholtz Centre, Leipzig), and K. Schirmer (EAWAG); and data call-ins made at SETAC-Europe (Milan, Italy) and SETAC-North America (Boston, USA) Annual Meetings, in May and November 2011, respectively. Targeted e-mails to distribution lists of the SETAC Global Advisory Committee on Animal Alternatives in Environmental Science and the ILSI-HESI Project Committee on Animal Alterative Needs in Environmental Risk Assessment (approximately 100 contacts) were sent seeking additional information*" (Belanger et al. 2012, p. 11).

The Belanger et al. (2012, p. 11) report also provides details on the source of acute fish toxicity data. "*Acute fish toxicity data were obtained from peer-reviewed literature, government reports, the USEPA ECOTOX database (http://cfpub.epa.gov/ecotox/), the USEPA ASTER database (via Chris Russom, USEPA) (http://www.epa.gov/med/Prods_Pubs/aster.htm), USEPA ECOSAR (version 1.1, release 17 August 2011), the Pesticide Action Network database (http://www.pesticideinfo.org/Index.html), the eChemPortal (http://www.echemportal.org/echemportal/substancesearch/page.action?pageID=0) of OECD, and UBA ETOX (http://webetox.uba.de/webETOX/index.do;jsessionid=98FB01DFE86E14B210BE486882273D84?language=en). On a few occasions, specific data were made available through interested industry participants or from MSDS. Use of MSDS information was limited to occasions where all other avenues for obtaining fish data were exhausted and the entries were sufficiently detailed with appropriate information*" (Belanger et al. 2012, p. 11).

In the Belanger et al. (2012) report, *all information was collected in a spreadsheet, and details of the studies were entered into the database as individual line records, so that contributions of independent tests on toxicities of each chemical could be assessed for contributions to variability in regression estimates. To facilitate data collection and maintenance, a data template was distributed to all contacts.* Inclusion/exclusion criteria were clearly marked in the EXCEL sheet.

Overall, the WG judged the design of the database as appropriate to allow conclusions on the suitability of the FET as an alternative model for acute fish toxicity testing. Indeed, the database covered: a large chemical domain, specific areas of use (chemicals, pharmaceuticals, pesticides, biocides), a wide range of toxicity, various modes of action, chemicals with different physico-chemical properties, data from different species of fish (OECD 203-based acute fish toxicity data).

### 1.4 Appropriateness of the statistical evaluation

*NOTE: Are the statistical methods used for evaluating the study data appropriate. Is there a sufficient justification for the use of the methods chosen? Was the statistician independent from the test method submitter/developer?*

The WG judged the chosen statistical methods for evaluating data of both the prospective and retrospective study to be appropriate. For the prospective study, LC50 values were determined for 48 and 96 h exposure, by logistic regression while confidence intervals were calculated using the delta method (Phase 1a). For Phases 1b, 2a and 2b, LC50 values were calculated by logistic regression (two parameter logistic function or, in some problematic cases, three parameter logistic regression). Confidence intervals were calculated using the profile likelihood method. The reports state that LC50 values were calculated for 48 h and 96 h following the recommendations of the OECD Guidance 54 in the statistical analysis of ecotoxicity data (OECD, 2006).

Also for the retrospective study the statistical approach was very well explained and justified. Orthogonal regression was used to fit the linear relationship between the two experimental methods, thus adjusting for measurement errors. By applying this approach, both the variability between the FET and fish acute toxicity from different species and between acute toxicity data of different fish species could be tested.

# 2. Collection of existing data

*NOTE: (Pre)validation studies typically make use of existing data, e.g. either as reference data (prospective studies) OR as reference data and testing data as well (retrospective study).*

## 2.1 Existing data used as reference data

*Which data sources were used for compiling reference data associated with the test chemicals?*

For the prospective study (all Phases) reference data pertained to the values of acute toxicity of chemicals in fish, which were used to compare values of LC50 obtained in the ZFET with those obtained in fish of different species and ages (juvenile and adult). The source of data for the acute fish toxicity values was the open literature and the OECD QSAR toolbox (Version 2.0). The range of LC50 value (min-max) was reported, and the geometric mean was calculated and utilised for comparison with the mean LC50 value obtained in the ZFET assay (VSR, Phase 1).

## 2.2 Existing data used as testing data

*Point 2.2 only concerns retrospective validation studies or modular studies that used existing and newly generated data to assess the performance of an assay. Which data sources were used to collect existing testing data?*

For the retrospective study, data for FET and AFT were retrieved from available sources, as indicate in section 1.3.2 of this report (p. 14). Additional details can be found in the report by Belanger et al. (2012, p. 11-13).

## 2.3 Search strategy for retrieving existing data

*NOTE: Please describe and evaluate how the search for existing data described was planned, organised and executed? In particular: has a **search strategy** been described and consistently applied?*

A clear strategy was utilised for retrieving existing data, and this is described in detail in the Belanger et al. (2012) report (p. 11-15). The WG did not notice any additional specific search strategy mentioned in the VSR beyond what is already indicated in Items 2.1 and 2.2 above. It is the opinion of the WG that the search strategy was appropriately described and fully acceptable.

## 2.4 Selection criteria applied to existing data

*NOTE: Have consistent evaluation/decision criteria been pre-defined and applied in order to select the data and has the selection of data been explained in a transparent manner?*

This issue would apply to 1) the selection of acute fish toxicity data, utilised to provide a "categorisation" of chemicals (e.g. very toxic, toxic etc.) and for a comparison with LC50 values obtained in the ZFET study (see Table 20 of Phase 2 report); 2) the selection of chemicals.

With regard to chemicals, for Phase 1b these were selected based on the recommendations of the ad hoc Expert Group (see Minutes of the meeting in May 2008). No specific criterion is described. For Phase 2b, a "chemical selection group" (CSG) was established, and the criteria for selection of chemicals is described in more detail (VSR, Phase 2, Annex II, page 8/67) as follows: *"The CSG discussed during several teleconference calls the criteria to be applied for selection of chemicals. The CSG agreed with the VMG to present an extended list of 20 chemicals to the OECD FET ad hoc expert*

*group for review and approval. This list should be complementary to the chemicals tested in Phase 1 and would be used to define the final list of 13 chemicals to be tested in Phase 2. It was agreed to prioritise the list of chemicals taking into account: a) the area of use (industrial chemicals, pharmaceuticals, pesticides, biocides); b) the range of fish toxicity (non-toxic, moderately toxic, toxic and very toxic); c) the availability of a quantification analytical method; d) the availability of fish data and, if possible, FET data; e) the commercial availability of the chemicals. The following issues were also considered: chemical classes; water solubility (use of solvent); the mode of actions (either known or as predicted by other available tools such as the OECD QSAR tool box [OASIS] or USEPA ECOSAR)."*

With regard to the selection of data (both FET and AFT data) utilised in the retrospective study, a detailed analysis has been provide in the report by Belanger et al. (2012, p. 11-15). In addition to all sources, this document indicates the criteria used in data selection. For example, with regard to chemicals, consideration was given to physico-chemical characteristics, QSAR information, and functional domain. The latter aspect is shown in Table 3 below.

Table 3. Categories of functional use for chemicals tested in the Fish Embryo Test and the Acute Fish Toxicity test (Number of chemicals for each functional domain; adapted from Belanger et al. 2012, p. 16).

| | FET | AFT |
|---|---|---|
| Biocide | 10 | 5 |
| Flame retardant | 1 | 1 |
| Food Additive/Vitamin | 4 | 2 |
| Hair Dye | 1 | 1 |
| Industrial Organic | 125 | 77 |
| Inorganic | 2 | 2 |
| Metal | 7 | 7 |
| Natural/Botanical | 4 | 1 |
| Organometal | 1 | 1 |
| Perfume | 1 | 1 |
| Pesticide | 27 | 23 |
| Petrochemical | 1 | 1 |
| Pharmaceutical | 22 | 8 |
| Polymer | 5 | 2 |
| Surfactant | 19 | 19 |
| ***TOTAL*** | ***229*** | ***151*** |

# 3. Quality aspects relating to data generated during the study

## 3.1 Quality assurance systems used when generating the data

*NOTE: Have quality assurance systems such as GLP (Good Laboratory Practice) or GCCP (Good Cell Culture Practice) been followed when generating the data?*

From the Validation Study Reports (Phase 1 and Phase 2) it was unclear whether quality assurance systems such as GLP (Good Laboratory Practice) had been followed when generating the data in the prospective study. However, for generating the data, the trial plans for Phase 1 and 2 contained a statement on quality assurance. The participating laboratories had to document their quality assurance system. Quality assurance systems were followed when analysing and reporting the data. The quality check was performed by ECVAM staff and consisted of a check with respect to consistency of reported results (e.g. completeness, fulfilment of test criteria etc.). The LC50 values were calculated by a statistician: in Phase 1a, data analysis was carried out by the IHCP statistician, while in Phase 1b and in Phase 2 data analysis was carried out by a statistician from P&G (with approval of OECD).

## 3.2 Quality check of the generated data prior to analysis

*NOTE: Have the generated data been checked for quality including correct formatting (i.e. using agreed data reporting sheets/templates) prior to analysis? Has the quality check been performed by a staff member independent from the laboratory staff generating the data?*

The WG noted that the quality of reported data in both Phase 1 and Phase 2 of the prospective study was carefully checked. In Phase 1a (i.e. transferability of the SOP with use of the positive control 3,4-DCA) the laboratories reported the results back to the coordinator using the indicated template (see Annex I of Phase 1 report). Before statistical analysis, data were checked for completeness, and to assure that they met the acceptance criteria described in the SOP (VSR Phase 1, p. 29). For Phase 1b of the study, the laboratories provided data (for a total of 81 runs) using the appropriate template (see Annex I of Phase 1 report); all results underwent a quality check by the coordinator and by an independent person prior to statistical analysis (VSR Phase 1, p. 36). Quality check for Phase 2 was similar. Initially (Phase 2a) the laboratories provided the data to ECVAM using the indicated template (see Annex I of Phase 2 report). All data underwent a quality check essentially identical as in Phase 1a (VSR Phase 2, p. 31) before statistical analysis, which was performed by P&G. Finally, as part of Phase 2b, the laboratories provided data on 153 runs. During the quality check, 10/153 runs (6.5%) did not meet acceptance criteria for various reasons, and were disqualified (VSR, Phase 2, p. 34.). The laboratories repeated the disqualified runs and all of the repeated runs met the acceptance criteria.

The WG noticed that all data generated by the participating laboratories in all phases were carefully checked by a staff member who was independent from the laboratory staff generating the data. The WG is of the opinion that such quality checks were appropriately described in the VSRs and were sufficient to assure that all data submitted to analysis were of the highest quality and met all acceptance criteria set in advance of the study.

# 4. Quality of data used for the purpose of the study (existing and newly generated)

## 4.1 Overall quality of the evaluated testing data (newly generated or existing)

*NOTE: Please describe the quality of the testing data. This may concern prospective testing data generated during and for the purpose of the study as well as existing testing data used for the study in a retrospective manner.*

The testing data for both the prospective and retrospective studies are represented by the ZFET data (prospective study) and by the FET data (retrospective study). With regard to the prospective study, WG judged the data generated by the study (Phase 1 and Phase 2) to be reliable. A detailed SOP was generated before the start of the study, and further modified and improved throughout the study. An important aspect of the SOP present from the start, was exposure duration, which was set at 96 h to account for the presence of the chorion that could act as a barrier to chemical exposure of the embryos. The results obtained with two high MW polymers (Merquat 100 and Luviquat HM 552), which showed no toxicity at 48h and significant toxicity at 96h, fully support such indication of the SOP. Additional modifications were related to the number of embryos to be tested at each concentration (20, for better statistical reliability), and provisions for acceptance criteria regarding fertilisation rate, and internal negative control. The acceptance criteria for the positive control was set from Phase 1b onward, whereas rejection of the plate when mortality occurred in the internal negative control was set for Phase 2b. All these aspects, together with the planned experimental design and the careful quality check carried out before analysis of data, contribute to a good overall quality of the testing data in the prospective study.

Equally positive comments were provided by the WG with regard to the overall quality of the testing data from the retrospective study. Data for FET were obtained from various publications and data bases detailed in Section 2.2. All data were normalised for water hardness (set at 100 mg/L). Details of each individual FET value are indicated in the report by Belanger et al. (2012).

## 4.2 Quality of the reference data for evaluating reliability and relevance[1]

*NOTE: What is the quality of the **reference data** used? Is the quality of the reference data sufficient in view of the study objective or is the study objective difficult to attain due to quality/reliability issues of the reference data selected and used for the purpose of the study?*

For the prospective study, reference data are the acute fish toxicity data which were used for two purposes: 1) establishing the relative toxicity class of a chemical (e.g. very-toxic, toxic, etc.) and 2) Comparing the result of the ZFET study with acute fish toxicity data.

The first utilisation of this information was for the choice of a series of chemicals spanning all four classes of acute fish toxicity. The WG noted some issues related with these choices. The classification of chemicals into four different classes based on acute fish toxicity (indicated by -, +, ++, +++) appeared to be somewhat inconsistently applied throughout the report. Such inconsistency was likely due to the use of different sources for the indicated information, and did not in any way

[1] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

hamper the study. The second utilisation of these data was for comparing the ZFET and fish LC50 data (see Table 20 of the VSR, Phase 2). The WG noted some issues related to the fact that the mean of all available fish LC50 data was utilised for the comparison. While this may be appropriate for chemicals for which the variability of LC50 is low (e.g. prochloraz with a range of 0.53 to 0.68 mg/L), the approach may be more problematic with other compounds with a wide range of reported LC50 values (e.g. malathion with fish LC50 ranging from 0.003 to 25 mg/L; mean = 0.289 mg/L). Perhaps the median could have been used for such comparisons instead of the mean.

For the retrospective study, reference data are all the fish data that were collected from innumerable sources and databases (see Section 2, above), while testing data are the FET data, also deriving from several different sources. A search strategy was employed by Belanger et al. (2012), and quality checks were performed for both FET and acute fish toxicity data. All individual data for each LC50 are listed together with the source. The large Excel sheet contains a high quality data base, as the original data were carefully scrutinised. For example, if tests were done at concentrations beyond water solubility, results were not considered to be reliable.

### 4.3 Sufficiency of the evaluated reference and testing data in view of the study objective

*NOTE: Is, overall, the quality of all data (i.e. testing data and reference data) used sufficient in view of the stated objective of the study?*

The WG judged all testing and reference data for both the prospective and the retrospective study to be of sufficient quality in view of the stated objectives.

## 5. Test definition (Module 1)

### 5.1 Quality and completeness of the overall test definition

*NOTE: This includes an analysis of the quality of the description of (a) the test system, (b) the protocol and associated SOP, (c) the test acceptance criteria, (d) the prediction models, (e) the biological and/or mechanistic relevance of the test method for the target organ/species/system etc.*

According to the specified objectives of the study (prospective and retrospective parts), i.e. to assess the reliability (transferability, within- and between-laboratory reproducibility) and the relevance (predictive capacity) of the ZFET, the overall test definition was considered of very good quality. Important elements such as (a) the test system, (b) the protocol, (c) the test acceptance criteria, and (d) the biological relevance, were well defined and described in the validation report and in the supporting documents. A prediction model was not defined due to a 1:1 concordance between FET and AFT (96 h).

### 5.2 Quality and completeness of the documentation concerning (a) SOPs and (b) Prediction Model(s)

*NOTE: Are the SOPs sufficiently detailed and complete? Are the Prediction Models (e.g. used to analyse (a) reproducibility through concordance of predictions and (b) predictive capacity) sufficiently well explained to be applied in the correct manner?*

The WG considers the SOP used in the validation study during the different phases as sufficiently detailed and complete for standardised use. Several modifications of the ZFET protocol which were well documented in the VSR were performed before the start and during the study.

An overview of the modifications for each Phase, as stated in the Validation Reports, is summarised below:

*"Before the start of the study the lead laboratory provided a draft SOP (which was reviewed by the Validation Management Group (VMG) and discussed with the participating laboratories before the start of Phase 1a).*

*The SOP deviated from the latest version of the OECD draft guideline (status May 2008) in relation to the following points, taking into consideration the concerns expressed by the ad hoc Expert Group:*

- *Possibility that the chorion could act as a barrier to chemical exposure, therefore the exposure duration was extended beyond hatch (96h) with calculation of LC50 at 48h and 96h.*
- *Number of embryos per concentration and control were increased to 20 embryos instead of 10 embryos.*
- *Acceptance criteria were set for the fertilisation rate. (The fertility rate of the parent generation should be ≥70%)"* (VSR Phase 1, Part 1, P. 25-26).

Phase 1a -Single run with 3,4 DCA (**SOP_ZFET_OECD_V02.7**).

The results of this first experiment led to an amendment of the SOP and the trial plan.

*"Since the analytical measurements demonstrated a substantial loss (>20%) of the 3,4-DCA concentration the SOP was modified as follows:*

*a) test vessels and 24-well plates were pre-saturated with the respective test concentrations at least 24h before the start of the test, and*

*b) daily renewal of the test concentrations and controls was required in order to maintain the test concentration >80% during the exposure period, which corresponds to the semi-static method as defined in OECD TG 203 (OECD, 1992)"* (VSR Phase 1, Part 1, p. 29, item 15).

Phase 1a: Transferability: Three runs with 3,4 DCA (**SOP_ZFET-OECD_V02.8**).

*"The results of Phase 1a led to two amendments to the SOP:*

*a) A minimum microscopic magnification of 80x should be used for the detection of the heart beat.*

*b) The acceptance criteria for the positive control (3.4-DCA) was included: 4.0 mg/L 3,4-DCA should result in a minimum mortality of 30 %."* (VSR Phase 1, Part 1, p. 34, item 33).

Phase 1b: Chemical testing (**SOP_ZFET_OECD_V02.9**).

Six laboratories trained in Phase 1a tested six chemicals in three independent runs following a slightly revised SOP (SOP_ZFET_OECD_V02.9).

The results of Phase 1b resulted in the following amendment to the SOP: A note on acceptance criteria for internal negative controls was added: *"If more than 1 dead embryo is observed in the internal negative control, the plate might be rejected."* (VSR Phase 2, Part 1, p. 32).

Phase 2a: Three runs with 3,4 DCA (**SOP_ZFET_OECD_V02.9**) (see Trial Plan for Phase 2a)

Phase 2b: Testing of thirteen chemicals used **SOP_ZFET_OECD_V02.10** (VSR Phase 2, Part 2, Annex VI)

# 6. Test materials

## 6.1 Sufficiency of the number of evaluated test items in view of the study objective

*NOTE: Is the **number** of test items (e.g. test chemicals) tested and analysed during the study sufficient in order to draw conclusions with respect to the objective of the study or are there shortcomings with regard to e.g. statistical sample size? If the number of items is considered insufficient, are there reasons for deviations and are these sufficiently explained and justified in the validation study report?*

Overall, the WG considers the number of test items sufficient to draw conclusions about the reliability (transferability within- and between-laboratory reproducibility) and the relevance (predictive capacity) of the ZFET.

In the prospective part of the study, 20 chemicals were tested covering specific areas of use (chemicals, pharmaceuticals, pesticides, biocides), a wide range of toxicity and various modes of action to evaluate the transferability and the within- and between-laboratory reproducibility of the ZFET (Tables 1 and 2).

For the retrospective part of the study, a total of 985 FET studies (229 chemicals) and 1531 acute fish toxicity studies (151 chemicals in common with the FET) were analysed by FET-acute fish toxicity (AFT) regressions to understand the potential relationships or biases based on physico-chemical properties, species choices, duration of exposure, chemical classes, chemical functional uses, and modes of action.

## 6.2 Representativeness of the test items with respect to applicability

*NOTE: Are the test items selected appropriate in view of understanding and describing the applicability and limitations of the test method. In case there is already information on applicability and limitations of the test method, consider to which extent the test items used during the study **map** applicability and limitations.*

The chemicals selected for the retrospective as well as prospective part of the study span a range of toxicity, chemical classes, mode of actions, areas of use (chemicals, pharmaceuticals, pesticides, biocides), molecular weights (<49 – >600) and log Kow (-2.49 – 7.5). Therefore, the selection of chemicals allowed gaining insight into the applicability domain of the FET (see also section 10 of the present report).

# 7. Within-laboratory reproducibility (Module 2)

### 7.1 Assessment of repeatability and within-laboratory reproducibility

*NOTE: How were repeatability and reproducibility assessed? Are the conclusions justified by the data as evaluated?*

Within laboratory reproducibility (WLR) was assessed by the analysis of the coefficient of variation (CV).

### 7.2 Conclusion on within-laboratory reproducibility

*NOTE: Are the conclusions on repeatability and within-laboratory reproducibility justified by the evaluated data?*

The WG confirms that the conclusions drawn are justified by the data. In agreement with the VMG the WG judges that the WLR of the ZFET is generally acceptable at 48h and 96h. Most compounds had intra-laboratory coefficients of variations below 30% (acceptance criterion set by the VMG). Plausible reasons were suggested for compounds with CVs above 30% (steep dose-response curves, physico-chemical properties).

Overall, the WLR (measured as coefficient of variation) was <30% for 14 compounds. At 48 h the WLR was >30% for 3 chemicals (6-methyl-5-hepten-2-one, 4,6-dinitro-o-cresol, tetradecylsulfate). For two substances (Merquat 100 and Luviquat HM 552) no CV could be calculated due to insufficient toxicity. At 96 h the WLR was >30% for 6 chemicals (triclosan, 6-methyl-5-hepten-2-one, 2,4-dinitrophneol, Merquat 100, tetradecylsulfate, malathion).

# 8. Transferability (Module 3)

### 8.1 Quality of design and analysis of the transfer phase

*NOTE: Was the transfer phase appropriately planned, e.g. transfer instructions, training, minimum requirements, training SOP (if appropriate). Were evaluation / decision criteria defining a successful transfer established beforehand and consistently applied during the analysis?*

In Phase 1a and Phase 2a of the prospective study the transferability of the ZFET to the participating laboratories was assessed using 3,4-DCA at six concentrations (0.5, 1, 2, 3.7, 4, and 8 mg/L plus negative control) as the test substance.

The transfer phases were appropriately planned as outlined in the plan for Phase 1a and Phase 2a. Evaluation/decision criteria defining a successful transfer were established beforehand. With regard to the WLR, the VMG agreed upon that coefficients of variation (CV) below 30% would be acceptable for demonstration of the transferability of the SOP using 3,4-DCA.

### 8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

*NOTE: Are the conclusions on transferability and "ease of transferability" (relating to robustness of the method) justified by the data generated? Have critical issues that may impact on transferability been identified?*

Phase 1a - Transferability: Three runs with 3,4 DCA

Considering the 5 labs with three qualified runs, the WLR was acceptable at 48h and 96h (CV<30%). The WG agrees with the conclusion of the VMG that the ZFET could be successfully transferred from the lead lab to the participating labs.

Critical issues that may have an impact on the transferability were identified and appropriately addressed. The VSR (Phase 1, Part 1, p. 29, item 15) states:

*"Since the analytical measurements demonstrated a substantial loss (>20%) of the 3,4-DCA concentration in the first experiment, the SOP was modified as follows:*

*a) test vessels and 24-well plates were pre-saturated with the respective test concentrations at least 24h before the start of the test, and*

*b) daily renewal of the test concentrations and controls was required in order to maintain the test concentration >80% during the exposure period, which corresponds to the semi-static method as defined in OECD TG 203 (OECD, 1992)."*

The VSR (Phase 1, Part 1, p. 34, item 33) further states:

*"The results of Phase 1a led to two amendments to the SOP:*

- *A minimum microscopic magnification of 80x should be used for the detection of the heart beat.*
- *The acceptance criteria for the positive control (3.4-DCA) was included: 4.0 mg/L 3,4-DCA should result in a minimum mortality of 30 %."*

The transferability of the protocol was further confirmed by testing 6 chemicals in Phase 1b.

Phase 2a - Training of new laboratories: Three runs with 3,4 DCA

The conclusions made by the VMG on the successful transferability of the ZFET to the new participating labs are justified by the generated data. Critical issues were not identified.

The transferability of the protocol was further confirmed by testing 13 chemicals in Phase 2b.

## 9. Between-laboratory reproducibility (Module 4)

### 9.1 Assessment of reproducibility in different laboratories

*NOTE: How was between-laboratory reproducibility assessed?*

In Phase 2b of the prospective study 9 laboratories tested a sub-selection from 13 compounds. Evaluation of the data is complicated by the fact that only one laboratory tested all the compounds, the other laboratories tested between 3-7 compounds from the 13. Each compound was tested 4 times except for methylmercury chloride which was tested three times. The target for acceptability

was an between-laboratory variability (CV) of less than 30%. This was exceeded for four compounds. These compounds were at the higher end of the toxicity spectrum and it is likely that the steepness of the dose-response curve for such compounds makes a reproducible LC50 more challenging.

### 9.2 Conclusion on reproducibility as assessed by the study

*NOTE: Are the conclusions on between-laboratory reproducibility justified by the data generated?*

Overall, the WG was of the opinion that between-laboratory reproducibility was acceptable.

## 10. Predictive capacity and overall relevance (Module 5)

### 10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

*NOTE: How was predictive capacity assessed? Where the reference data used in an appropriate manner to conclude on predictive capacity? Are the conclusions justified based on the data evaluated and in view of the test method's purpose?*

Overall, the WG agreed that the data are strong enough to justify the conclusion on the predictive capacity of the ZFET.

The prospective phases of the study were not designed to evaluate predictivity and this was not formally assessed. In Phase 1 (which included 7 chemicals, 1 in Phase 1a and 6 in Phase 1b), it was demonstrated that the ZFET LC50 values were within a factor of three of the literature-derived fish LC50 values. This is a reasonably good correspondence, given the variability intrinsic to the *in vivo* fish data.

In Phase 2, the predictive capacity was again not formally tested although a table was produced (Table 20, VSR, Phase 2, p. 53) showing fish acute toxicity data with the ranges of published values. Again, the performance relative to the mean values was reasonably good. However the limitations of this approach were noted by the WG, specifically: when the range of published values spans more than one order of magnitude, it is not evident that the mean value should be construed as the "correct" value. The fish LC50 value with a range of LC50s of less than an order of magnitude constitute less than half of the compounds tested; therefore, while the WG cautiously supports the view that these data are promising, no firm conclusions were, or could be, drawn from this comparison

The WG was concerned that use of the chemical classification values as an aid to evaluating predictivity was a source of confusion. As these classification values are bounded by cut-offs, it is important to ensure that they are used consistently. This has not been done in the VSR, Phase 2, and means that there are small inconsistencies between Tables 1 (p. 29) and 20 (p.53). It is debateable whether these regulatory classifications (with their arbitrary boundaries) add anything to the analysis of the data as the choice of "representative" LC50 values can lead to different classifications. Therefore, evaluating the data in terms of correctly predicting the toxicity "class" can be very misleading.

The retrospective study was designed specifically to look at predictivity and used orthogonal regression analysis which takes into account the variability in the reference data and the experimental data. In this way it was possible to look at the predictivity with 151 chemicals without needing to decide the "correct values" for the reference compounds. From this analysis with all compounds a correlation coefficient of 0.9 was derived with a slope close to 1 (95% C.I. 0.95-1.11) and an intercept close to the origin.

Similar comparisons were made between different species in the *in vivo* fish database and revealed differences in species sensitivity (rainbow trout most sensitive). The zebra fish embryo toxicity data fell within the variability of the fish inter-species comparisons.

Correlations were examined between classes of chemicals and, in general, no class of compound was revealed an exceptional difference to the predictivity of the method. In general, as might be expected, the correlation was less robust for classes of compounds with fewer examples.

### 10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

*NOTE: Are the conclusions reg. biological relevance and relevance in terms of making accurate predictions/measurements for the specific toxicity effect justified by the evaluated data?*

The WG was of the opinion that the data from the ZFET test were relevant for fish toxicity as the data were not distinguishable from the range of toxicity seen between different species of fish. Given that reference *in vivo* data span as much as 3 orders of magnitude for some chemicals and typically one order of magnitude for most chemicals, there appears to be no basis to distinguish ZFET data from fish data.

## 11. Applicability domain (Module 6)

### 11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

*NOTE: When considering the objective of the study, was the study designed in a way to ensure that conclusions on the **applicability domain, the limitations and exclusions** of the test method can be drawn?*

The WG concludes that the data obtained in the prospective as well as retrospective part of the study provides new information on the applicability and the possible limitations of the (Z)FET.

*Prospective part:*

Overall the prospective study did not reveal any specific limitations on the applicability domain by chemical class or mode of action. However two specific reservations were made, namely:

- The test needs to be run for 96 hours as standard practice to enable high molecular weight compounds to be accurately predicted.
- Analysis of test concentration needs to be performed to ensure that solubility, volatility, degradation and adsorption are fully accounted for.

*Retrospective part:*

Examination of 151 chemicals including testing of specific subgroups of chemicals, (pesticides, surfactants etc), did not reveal any clear exclusions from the applicability domain for the FET.

### 11.2 Quality of the description of applicability domain, limitations, exclusions

*NOTE: When considering the objective of the study and the data generated/analysed, have the applicability domain, the limitations and the exclusions of the method been sufficiently described?*

The study designs in both the prospective and retrospective Phases were intended to cover a broad range of chemistry from both the use category and the toxicity perspectives. As a result, there is no obvious gap in the applicability domain, even if some categories may have relatively few examples. It is reasonable to conclude on the basis of this data that the (Z)FET has a broad applicability domain with no identified exclusions, provided that it is run for 96 h and exposure concentrations are verified.

# 12. Performance standards (Module 7)

This section is not applicable in the context of this review project.

# 13. Readiness for standardised use

### 13.1 Assessment of the readiness for regulatory purposes

*NOTE: Is the test method ready for regulatory purposes? If yes, why? If no – what impediments currently exclude advancement of the method for regulatory purposes?*

The FET using zebrafish (i.e., the ZFET) for 96 h constitutes an alternative test method to the acute toxicity tests with juvenile and adult fish, i.e., the OECD Test Guideline 203 (OECD TG 203, 1992) and similar guidelines thus providing a reduction in fish usage. An OECD guideline for the ZFET is currently under development and it would be important to finalise this guideline, which is currently being circulated as draft guideline (updated December 2012), as soon as possible. Once the guideline is finalised, the ZFET would be ready for regulatory use. As the prospective and retrospective studies show, it has been thoroughly evaluated for a wide range of chemicals with different physico-chemical properties and modes of toxic action; as well, transferability has been shown with acceptable within- and between-laboratory variability (see p. 54 VSR, Phase 2). The new OECD guideline could make its way into many testing schemes, including industrial chemicals, agrochemicals, veterinary pharmaceuticals, biocides, effluent testing. With regard to testing effluents for acute toxicity, an ISO guideline (DIN EN ISO 15088, called the fish egg test) already exists. It has been the basis of the first draft for the FET OECD guideline submitted to OECD by the German Environmental Protection Agency (UBA) in 2006 and is accepted as replacement of the acute fish toxicity test for effluent testing in Germany since 2005.

The application of the FET in a regulatory context should be considered with regard to the OECD fish toxicity testing framework (OECD Fish Toxicity Testing Framework 2012), which suggests reduction of use of fish tests on several levels (e.g., use of invertebrate and algae/aquatic plant EC50 values, application of limit tests). On page 150, reference is given to the FET as an alternative to the fish acute toxicity test: "…There may, furthermore, be scope to use the draft Fish Embryo Test (OECD 2006a), although this test has not yet been fully evaluated."

Thus, with the thorough evaluation now provided by the prospective and retrospective studies in particular for the ZFET, specific areas of regulatory use, for which reference to the FET is already given, include:

- Classification and labelling:

see Guidance for implementation of REACH: "Guidance on information requirements and chemical safety assessment" Chapter R.7b, page 43, where it is stated that, in case a fish acute toxicity test

would be required (i.e., if data on aquatic invertebrates and plants/algae indicate EC50 values > 1 mg/L and a QSAR values for fish acute toxicity cannot be convincingly obtained):

*"…if alternative and adequate test methods are available for the acute fish toxicity they may be considered to be used instead for classification… E.g., a proposal to use the fish embryo test (FET) as an alternative to the acute fish toxicity test has been made and is currently under evaluation in the OECD Guideline program…".* [Note: reference is also given in the associated Figure R.7.8-3 on page 45]

- Chemical safety assessment of industrial chemicals (REACH), see Guidance for implementation of REACH: "Guidance on information requirements and chemical safety assessment" Chapter R.7b, page 49, where it is stated that:

*"If there is a need to generate new data on the toxicity in fish and an accepted alternative method is available instead of in vivo fish testing perform the alternative test… A possible alternative, the fish embryo toxicity test, is currently under evaluation in the OECD Guideline program…"* [Note: reference is also given in the associated Figure R.7.8-4 on page 53]

One issue requiring discussion in the context of regulatory use is the performance of the embryo test under GLP. It is thus far not common practice to run this test under GLP conditions; however, for regulatory use, compliance to GLP quality control should become standard.

### 13.2 Assessment of the readiness <u>for other uses</u>

*NOTE: Is the test method ready for other uses (e.g. screening purposes, testing to gain mechanistic insight, to generate supportive information for hazard/risk assessment).*

One obvious other use of the FET is for in-house product development. Basically, any application requiring knowledge on acute fish toxicity can be done using the fish embryo test.

### 13.3 Critical aspects impacting on standardised use

*Note: What are the factors that may impact on standardised use (in regulatory or non-regulatory settings)?*

The WG suggests to make verification of chemical concentrations mandatory. Although good correlations between the fish and FET data have been derived without verification of chemical concentrations, there is clear evidence that verification improves the reliability of the concentration-response data used to derive LC50 values (see also Knöbel et al., 2012; the data of this study were included in the retrospective analysis). Importantly, chemical verification also provides confidence to the test results with chemicals that do not yield toxicity for concentrations up to 100 mg/L (above which a chemical is considered non-toxic). Verification of chemical exposure concentrations is also requested in both the OECD test guideline 203 (fish acute toxicity) and the OECD draft version to the ZFET although both guidelines include statements that give the impression that chemical verification could be omitted ("…if evidence is available to demonstrate that the test substance has been satisfactorily maintained…").

The critical aspect of this is that for companies/agencies not yet regularly confirming chemical concentrations, additional costs need to be calculated for the chemical analysis (which also may need to be outsourced). Moreover, chemical concentration verification should clearly be stated as a definitive testing requirement for both the OECD 203 as well as the alternative test using fish embryos so as to not put the alternative test to a disadvantage.

Another critical issue is whether the FET is equally valid as an alternative to the OECD 203 independent of the fish species from which the embryos are used. Thus far, most research was performed with zebrafish. The prospective study was done focussing on zebrafish; likewise, the

retrospective study is largely dominated by that species. As well, the OECD draft guideline is focussed on zebrafish. Thus, the applicability of the ZFET as an alternative to fish acute test according to OECD 203 (regardless of the species suggested in this guideline) is clear. However, although evidence provided by the retrospective study also speaks in favour of using embryos of e.g., medaka (*Oryzias latipes*) and fathead minnow (*Pimephales promelas*), more research is needed to verify the applicability of other fish species for the FET.

### 13.4 Gap analysis

*NOTE: Identify, if appropriate, gaps in the study design and/or execution that impact on the stated study objective or the conclusions drawn.*

Inasmuch as the prospective study had to rely on voluntary participation of the testing labs, not all labs tested all the chemicals, leading to a complex matrix of test results. Nevertheless, statistics was appropriately done and the fact that more labs were involved can also be interpreted as strength in that the very good overall results were obtained despite this difficulty. For future validation studies, the WG feels that it would be important to provide funding and to establish legal contracts for round-robin tests.

Despite its thoroughness, the retrospective study did not consider volatility (logH) as a parameter (both for the spread in terms of applicability domain and for investigating its influence on fish-fish embryo relationships [as was done for solubility and hydrophobicity (focussing on logKow)]. Impact of test results due to volatility might be mitigated in the testing design used in the prospective study by the pre-saturation of plates and daily exposure solution renewal. Consideration is nevertheless important because neither plate sealers nor completely closed test vessels with a headspace can completely abolish losses due to evaporation (Schreiber et al., 2008; Knoebel, Scholz, Schirmer, personal communication).

## 14. Other considerations

*NOTE: Please address any other consideration you might have not covered by any of the other sections in this report template.*

The retrospective study shows that the zebrafish embryo basically acts like another fish. From this point of view, there is no need for a prediction model. The slope is within the 95% confidence interval of the 1:1 correspondence (line of unity) and the slope of 1 is also within the 95% confidence interval for the slope in the equation. In other words, one cannot statistically discern the result from the regression from the direct 1:1 correspondence. Thus, the results obtained from the ZFET can be taken without any conversions.

## 15. Conclusions on the study

*NOTE: This section should present (a) an overview over the study results and conclusions described on the basis of the validation study report(s) (subsection 14.1), (b) a discussion to which extent the conclusions drawn in the validation study report(s) are justified by the study results on their own (subsection 14.2) and (c) to which extent the conclusions are plausible with respect to other information (subsection 14.3).*

### 15.1 ESAC WG summary of the results and conclusions of the study

The WG considers both the prospective as well as the retrospective study as well designed and pursued. For the prospective study, the SOP used in the validation study during the different phases is sufficiently detailed and complete for standardised use. Several modifications, which were well documented, were performed during the study. For example, pre-saturation of plates and daily renewal of exposure medium helps to ensure more stable exposure concentrations. For the retrospective study, all information available at this point with regard to fish embryo tests and their comparisons to acute fish toxicity data have been included, revealing that the embryo overall acts like a fish in the acute toxicity study design and that this is valid for chemicals with a wide range of physico-chemical properties and different modes of toxic action.

### 15.2 Extent to which study conclusions are justified by the study results alone

Both the prospective and retrospective studies are very well documented and allow conclusions by their results alone. The conclusion of the prospective study is that the ZFET can be transferred to other laboratories and that the SOP is sufficiently detailed and the test procedure robust so as to obtain comparable test results. The conclusion from the retrospective study is that, indeed, fish embryo tests forecast juvenile or adult fish acute toxicity.

### 15.3 Extent to which conclusions are plausible in the context of existing information

The conclusions are plausible also in the context of existing information. In fact, the retrospective study is based on existing information and the prospective study supports the results of the data collection compiled in the retrospective study.

## 16. Recommendations

*Note: This section should provide recommendations on the test method (e.g. further work required, possible use) and its constituting elements (e.g. test system, prediction model, SOP).*

### 16.1 General recommendations

The WG strongly feels that analytical confirmation of exposure concentrations should be an explicit requirement, in fact, for both the ZFET as well as the OECD 203. The WG feels that considering measured exposure concentrations is particularly important for aqueous exposure assessments because it will be used to deduce environmentally safe concentrations.

The ZFET draft OECD guideline (version December 2012) points out this requirement by saying that, “A reliable analytical method for the quantification of the substance in the test solution with known and reported accuracy and limit of detection should be available” (page 1; §4) and that, “As a minimum, the concentration of the test substance should be measured in the highest and lowest test concentration, but preferably in all treatments, at the beginning and end of the test” (page 7; §35). Efforts were also made in this draft guideline to allow for a more stable exposure concentration by pre-saturating exposure plates and by daily medium renewal. On the other hand, these statements are weakened by the statement that, “if evidence is available to demonstrate that the concentration of the test substance has been satisfactorily maintained within 20 per cent of the nominal or measured initial concentration throughout the test, results can be based on nominal or measured initial values.” (page 7; §35) However, in order to demonstrate that the concentrations are “satisfactorily” maintained, one needs to measure the concentrations, so this is a circular argument.

Nevertheless, the WG appreciates that the statements made regarding chemical concentration verification for the ZFET are much stronger than in the OECD 203 testing guideline, where it is only stated that, “There must be evidence that the concentration of the test substance being tested has been satisfactorily maintained, and preferably should be at least 80 per cent of the nominal concentration throughout the test. If the deviation from the nominal concentration is greater than 20 per cent, results should be based on measured concentrations.” Thus, while chemical concentration verification is implied in this sentence, it is not explicitly stated as a requirement. It would be important to have the same level of quality assurance with regard to chemical concentration verification in both the FET and the acute fish toxicity test.

Based on the results for high molecular weight compounds (which elicited toxicity only if test duration was extended beyond hatch), the WG supports the strategy to run the ZFET for 96 hours.

According to the retrospective study, where the correlation of FET and fish acute toxicity is non-discernible from the line of unity, a prediction model is not required, and this is true independent of the fish species used in the acute exposure tests. One has to keep in mind, however, that while the FET studies are still dominated by zebrafish (*Danio rerio*), the retrospective study also shows that a similar level of correlation can be found for, e.g. embryos of fathead minnow (*Pimephales promelas*).

Finally, one issue requiring discussion in the context of regulatory use is the performance of the zebrafish embryo toxicity test under GLP. It is thus far not common practice to run this test under GLP conditions; however, for regulatory use, compliance to GLP quality control should become standard.

### 16.2 Specific recommendations

The WG recommends to keep 3,4-DCA as the positive control, since it has been used in the validation exercise. Otherwise, criteria should be given for the selection of benchmark chemicals or a list of benchmark chemicals used for positive control could be suggested.

The OECD draft guideline recommends the use of self-adhesive foil to cover the 24-well plates or vapour-dense lids provided with plates; while this is certainly useful, one has to consider that none of these measures really prevents loss due to evaporation of volatile compounds. Similarly, the use of glass instead of plastic “in case adsorption to polystyrene is suspected” may decrease losses due to adsorption but may not abolish them.

In the prospective study, certain procedures were implemented based on recommendations by the OECD ad-hoc working group. Specifically, these concerned the number of embryos per concentrations (20), an exposure time beyond hatch up to 96 h, pre-saturation of well plates and daily exposure medium renewal, and a coefficient of variation (CV) below 30%. Based on the data and reports provided, the WG fully supports these recommendations, which have now also been included in the OECD draft guideline for the ZFET.

## 17. References

Belanger SE, Rawlings JM, Carr GJ (2012) An update to the fish embryo toxicity-acute fish toxicity relationship and prospects for support of the use of the FET as an animal alternative. Document prepared for the 2012 OECD ad hoc Expert Meeting Group on the Fish Embryo Test, pp. 1- 156.

Braunbeck Th, Lammer E (2005) Fish Embryo Toxicity Assays. Detailed Review Paper for the GermanFederal Environmental Agency, UBA Contract Number 203 85 422, pp. 1-302.

EU (2010) Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animal used for scientific purposes. Official J. of the EU L276, 33-79.

Knoebel M, Busser FJM, Rico-Rico A, Kramer NI, Hermens JLM, Hafner C, Tanneberger K, Schirmer K, Scholz S (2012) Predicting adult fish acute lethality with the zebrafish embryo: Relevance of test duration, endpoints, compound properties, and exposure concentration analysis. Environmental Science & Technology 46(17): 9690-9700.

Lammer E, Carr GJ, Wendler K, Rawlings JM, Belanger SE, Braunbeck Th (2009) Is the fish embryo toxicitytest (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test? Comp. Biochem. Physiol. CToxicol. Pharmacol. 149(2): 196-209.

OECD (1992) Fish Acute Toxicity Test. OECD Guidelines for Testing Chemicals, No. 203, Adopted 17.07.1992, pp. 1-9.

OECD (2012) Fish Toxicity Testing Framework. OECD Series on testing and assessment, No. 171.

Schreiber R, Altenburger R, Paschke A, Küster E (2008) How to deal with lipophilic and volatile organic substances in microtiter plate assays. Environmental Toxicology and Chemistry 27(8): 1676-1682.