SECRETARIAT PAPER:
APPLICATION OF GUIDANCE DOCUMENT 34 CRITERIA TO THE VALIDATION OF THE 21-DAY FISH ENDOCRINE SCREENING ASSAY

This document was prepared at the initiative of the Secretariat and is aiming to a transparent presentation of the current status of the validation of the 21-day Fish Screening Assay to the Peer-Review panel, as of April 2007. A first draft of this document was prepared in February 2006 by the Secretariat in consultation with the co-Chairs of the Validation Management Group for ecotoxicity testing (VMG-eco), and presented to the Working Group of National Coordinators of the Test Guidelines Programme (WNT) in May 2006. The WNT stated that the paper was a clear presentation of the validation status of the assay and suggested that such paper be presented in the peer-review package to facilitate peer-reviewers' understanding of the validation status of the assay.

For each criterion set out in the OECD Guidance Document 34 adopted in June 2005, some answers and/or rationale are provided, based on the discussions of the Validation Management Group for Ecotoxicity Testing (VMG-eco) and on the information contained in the validation reports. Further explanation on the current status of discussion at the VMG level is sometimes provided to give context for some of the open questions.

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.

Scientific basis

Aquatic contaminants with androgenic, oestrogenic and aromatase inhibiting properties have been demonstrated to affect the reproductive fitness of fish, including disruption of sexual development and reproduction (cessation of spawning, infertility, intersex gonads, etc.) (Sumpter, 1998; Vos et al., 2000). The *in vivo* fish screening assay is specifically intended to detect chemicals that disturb normal androgenic, estrogenic and aromatase activity in fish of both sexes over a 21 day period. Two endocrine specific endpoints: vitellogenin level and secondary sex characteristics are measured at the end of the exposure period.

In the assay, adult males and females in a reproductive status are exposed together in test vessels. Their adult and reproductive status enables a clear differentiation of each sex, and thus a sex-related analysis of each endpoint.

Regulatory purpose and need for the test

The regulatory need for this assay stems from the absence of existing OECD Test Guidelines which enable the detection of endocrine active substances in the aquatic environment. The endocrine system regulation and the metabolism in oviparous aquatic organisms is distinct from mammalians, although similarities also exist, such that responses to endocrine active substances in mammals can not fully predict those in aquatic vertebrates, in nature or in intensity. Data show that fish should not be considered as simply a rat in the water, hence the need for a validated fish screening assay for (anti-)androgenic, oestrogenic and aromatase inhibiting compounds (Ankley et al., 1998; OECD 1998; 2000).

The assay, the decision for its conduct, and the interpretation of its outcome, are intended to be appraised/ weighed in conjunction with other sources of information (e.g. physico-chemical properties, *in vitro* data, results from mammalian assays, scientific literature, non test data, etc.) for determining the need or not for further investigations, including further testing. The assay measures responses which can not, at this stage of knowledge, be considered predictive or not predictive of adverse effects (Hutchinson et al., 2006). The assay is primarily intended for the *in vivo* screening of substances for aquatic vertebrates. Further testing would normally be expected for further hazard characterisation and risk assessment.

The assay may also be used for the testing of effluents from water treatment plants.

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.

Vitellogenin

Vitellogenin is normally produced by the liver of female oviparous vertebrates in response to circulating endogenous oestrogen. It is a precursor of egg yolk proteins and, once produced in the liver, travels in the bloodstream to the ovary, where it is taken up and modified by developing eggs. Vitellogenin is normally undetectable in the plasma of immature and male oviparous animals because they lack circulating oestrogen; however, their liver is capable of synthesizing and secreting vitellogenin into the blood in response to exogenous oestrogen stimulation.

- Thus, the detection of estrogenic chemicals is possible via the measurement of vitellogenin level in male fish, and it has been abundantly documented in the scientific peer-reviewed literature (Sumpter and Jobling 1996). The biological relevance of the response evaluated is established.

- Furthermore, a reduction in the circulating level of oestrogen in females, for instance through the inhibition of the aromatase, causes a decrease in the vitellogenin level, which thus usefully serves for the detection of aromatase inhibiting substances (Ankley et al., 2002; Zerulla et al., 1998). The biological relevance of the response evaluated is established.

- Finally, the decrease of vitellogenin production in females was also noted following exposure to non-aromatisable androgen (correlative relationship); however, the underlying mechanism is not well understood or documented yet and therefore the biological relevance is not established.

Several measurement methods have been successfully developed and standardised for routine use. This is the case of species-specific ELISA methods, using immunochemistry for the quantification of vitellogenin produced in small blood or liver samples collected from individual fish (Parks et al, 1999; Holbech et al, 2001; Nozaka et al, 2004). Blood of fathead minnow and zebrafish and liver of medaka are sampled for VTG measurement. In medaka, there is a good correlation between VTG measured from blood and from liver (Tatarazako et al, 2004).

Secondary sex characteristics

Secondary sex characteristics in male fish of certain species are externally visible, quantifiable and responsive to circulating levels of endogenous androgens; this is the case for the fathead minnow and the medaka in particular. Similarly, females maintain the capacity to develop male secondary sex characteristics, when they are exposed to androgenic substances in water. Several studies are available in the scientific literature to document this type of response in fathead minnow (Ankley et al, 2003; Panter et al, 2004) and medaka (Seki et al, 2004) in particular, but also in other species.

In the fathead minnow, the main indicator of exogenous androgenic exposure used is the number of nuptial tubercles located on the snout of the female fish. In the medaka, the number of papillary processes constitutes the main marker of exogenous exposure to androgenic compounds in female fish.

Important note:

The gonado-somatic index and the gonad histopathology endpoints were initially included in the assay. After a thorough evaluation of the inter-laboratory reproducibility with the test substances, it appears that *i*) the gonado-somatic index is not consistently responsive to strong test substances under the conditions of this assay, and *ii*) the gonad histopathology endpoint needs further work to demonstrate reproducibility of findings. The use of the fecundity endpoint was also considered, and together with the gonad histopathology, the Validation Management Group concluded that (extract from the VMG-eco 4 Summary Record, December 2005):

- vitellogenin measurement was validated for the detection of oestrogenic activity and aromatase inhibition;

- secondary sex characteristics were validated for the detection of androgenic activity;

- gonad histopathology was not yet validated, especially with regard to the lack of reproducibility of the current findings, therefore more work is needed using the methodology and guidance laid out in Phase 1B;

- fecundity was not yet validated and its non-specific role and interpretation in an endocrine screening assay context were questioned by several participants.

The VMG-eco also recommended that:

- The development of an OECD Test Guideline on a fish screening assay for endocrine active substances based on Phase 1B using VTG and SSC as core endpoints (with advice on species specific characteristics). Based on current knowledge, the primary objective is to detect estrogenic and androgenic agonists and aromatase inhibiting substances. The design of the TG would be such as to facilitate the inclusion of additional validated endpoints in the future.

- Simultaneously, to develop guidance documents addressing measurement, test design, techniques and interpretation of fecundity and gonad histopathology in the OECD fish species. The documents would also include information on the limitations and advantages of each of the four endpoints for detecting a range of endocrine activities.

- Subject to the provision of data generated according to the OECD validation principles (as per Guidance Document 34), to integrate fecundity and gonad histopathology into a revised OECD Test Guideline for fish screening for endocrine active substances (agonists and antagonists).

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.

A detailed protocol, including test materials needed: fish (age, weight and length ranges), appropriate test vessels and other acceptable test conditions the standard operating procedures for blood/liver sampling and measurement of vitellogenin and for counting the secondary sex characteristics is available. Acceptable ELISA methods and their specifications for each fish species have also been made available in the protocol. General acceptability criteria for the test, similar to existing OECD Test guidelines are available (e.g., mortality in controls, dissolved oxygen level, pH, temperature, etc.).

A preliminary draft OECD Test Guideline was prepared as recommended by the WNT to give a general sense of the level of details provided.

Remaining issues:

Acceptability criteria:

At present, acceptability criteria for the endpoints (e.g. vitellogenin level in control males and females for each species have not yet been defined. This constitutes the next step before the Test Guideline and will be addressed by fish experts, as it is necessary to ensure that the controls remain within a commonly accepted baseline. To that end, existing information from Phase 1A, Phase 1B and Phase 2 and experience in the laboratories may serve as a basis.

Replication:

The replication was still an open question at the VMG-eco meeting in January 2007. Simulations performed, based on vitellogenin data from fathead minnow and zebrafish resulted in the following power estimation under the maximum variance scenario representing a worse case scenario (see Table 1). Based on the properties of baseline VTG data (normality vs. non-normal distribution, monotonous vs. non-monotonous dose-response, variance homogeneity vs. heterogeneity), various statistical tests were simulated using scenarios where sample size and replication were allowed to vary.

The Phase 1B of the validation study was performed with 2 replicates of 5 males and 5 females in each replicate for all three fish species. For the detection of increased VTG (oestrogen-like response, effect size of 100% increase), such test design places the power of the assay in a range of 67-96% (shaded cells), probably around 80%, which would appear to be acceptable. For the detection of decreased VTG (aromatase inhibition-like response, effect size of 50% decrease), the same test design (i.e. with 2 replicates per treatment level) seems to yield low power. A test design with 4 replicates per treatment level seems to increase power (range 54-88%).

The Phase 2 of the validation study (negative substance testing) followed Phase 1B design for the medaka and the zebrafish. For the fathead minnow, Phase 2 studies were performed with 4 replicates per treatment level; each replicate containing 2 males and 4 females. The reason for re-arranging the test design for the fathead minnows is a biological one: male territoriality required a decrease of the number of males per tank while keeping a similar number of males per treatment level.

Overall the data presented at he VMG-eco meetings (Phase 1B data in Dec. 2005 and Phase 2 data in Jan.2007) suggested that the optimal design for power properties is to use 4 replicates per treatment with 2 to 5 fish/sex per replicate (Green, personal communication). The following Table 1 set out the power properties under the Maximum Variance Scenario.

MAXIMUM VARIANCE SCENARIO	**Square root allocation rule** (double replicates in the control)	**Detection of 100% increase of vitellogenin (e.g. estrogen induction-type of response)** Number of animals/sex/replicate		**Detection of 50% decrease of vitellogenin (e.g. aromatase-type of response)** Number of animals/sex/replicate	
		2	**8**	**2**	**8**
2 replicates	No	67	96	26	56
	Yes	85	100	41	75
3 replicates	No	91	100	42	80
	Yes	94	100	51	88
4 replicates	No	96	100	54	88
	Yes	99	100	62	96

Table 1: Power to detect a significant effect at the highest dose tested, using Jonckheere-Terpstra or Williams tests (simulations are based on zebrafish and fathead minnow VTG data).

Statistics

As indicated earlier, based on the properties of baseline VTG data (normality vs. non-normal distribution, monotonous vs. non-monotonous dose-response, variance homogeneity vs. heterogeneity), various statistical tests were simulated using scenarios where sample size and replication were allowed to vary. One general outcome was that for 2 replicates per treatment, the Mann-Whitney test has no power and should not be used. For test designs using 2 to 4 replicates per treatment, with 2 to 5 fish per replicate, Dunnett, Jonckheere, Williams and Dunn tests should be preferred. As a result, a flowchart to facilitate the selection of appropriate statistical test was prepared and can be found in the report of Phase 2 of the validation.

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.

Intra-laboratory variability as a measure of repeatability

Assessment of the repeatability of the fish screening assay within the same laboratory over time is limited, due to the limited number of times this specific protocol has been applied. However, a number of baseline studies on control VTG levels address this important aspect (Panter at al., 2000; Hutchinson et al., 2006; Eiden et al., in press).

Also, an initial comparison over time of intra-laboratory results from laboratories having participated in both Phase 1A and Phase 1B, is proposed in Table 2a and Table 2b. It is uncertain how this can be analysed further.

It was also suggested that baseline VTG measures from control animals from within participating laboratories which have performed similar studies in the past years could be evaluated to assess repeatability over time. However, the test designs used may vary.

Inter-laboratory variability as a measure of reproducibility

Intra-laboratory and inter-laboratory variability was reported and compared for individual laboratories in Phase 1B report (Tables 71 to 80). As an example, for induced males, the coefficient of variation of VTG measurements generally ranged between 30 and 70%, with a few exceptions. The coefficient of variation for secondary sex characteristics was lower, between 15 and 40% in average. Importantly, the reproducibility for VTG and SSC need to be put in context with other endpoints successfully used for many years in OECD and other regulatory test guidelines. An initial analysis of this aspect has recently been undertaken by Hutchinson et al (2006) and the data show that VTG reproducibility data compare well with fish survival and growth endpoints.

Several factors come into play when assessing the overall variability and sources of error of measurements: variability in the biological material (same for any test using biological material), variability at levels close to the detection limit of the measured endpoint (low VTG levels in control males generate high CV), variability in the test concentrations from one vessel to the next (specific to aquatic environment), variability in the ELISA kit and reagents. Provided some of these factors are known and controlled (e.g. maintenance of test concentrations within a narrow range), the coefficient of variation can also be maintained within acceptable limits.

Control FEMALES	Medaka (ng/mg liver)		Fathead minnow (ng/ml blood)		Zebrafish (ng/ml blood)	
	Ph1A	Ph1B	Ph1A	Ph1B	Ph1A	Ph1B
CERI	1700 (350)	690 (450)	1.3E+7 (2.4E+6)	-	1.3E+5 (5.3E+4)	-
NIES	270 (290)	280 (220)	1.7E+7 (5.8E+6)	-	8.8E+4 (3.9E+4)	-
METO	1800 (230)	1600 (490)	-	-	6.9E+4 (4.3E+4)	-
BCS	-	-	5.2E+6 (1.9E+6)	7.7E+6 (5E+6)	7.8E+4 (8.3E+5)	-
100 ng E_2/l [nominal] MALES	Medaka (ng/mg liver)		Fathead minnow (ng/ml blood)		Zebrafish (ng/ml blood)	
	Ph1A	Ph1B	Ph1A	Ph1B	Ph1A	Ph1B
CERI	5500 (1900)	6300 (1300)	1.1E+8 (5.6E+7)	-	9.3E+4 (6.4E+4)	-
NIES	3300 (1500)	1100 (470)	3.4E+7 (1.3E+7)	-	2.5E+6 (6.2E+5)	-
METO	11000 (2700)	5900 (2000)	-	-	3.3E+4 (5E+4)	-
BCS	**-**	**-**	**5.7E+7 (1.5E+7)**	**1.4E+8 (4.6E+7)**	**1.7E+6 (3.5E+6)**	**-**

Tables 2a and 2b: Means (standard deviation) of VTG measurements from four laboratories who participated in both Phase 1A and Phase 1B:
2a: baseline VTG levels in control females
2b: VTG level in males exposed to the positive control used in both phases [17beta-estradiol].

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").

Representativity

It is currently understood that representativity of chemicals is meant in terms of the modes of action they represent, rather than the chemical class they represent. Therefore the possible modes of action investigated and/or considered are as follows:

Chemical	**Mode of action**	**Class of chemical**	**Outcome in 21d fish screening assay for endocrine active substances**
17beta-estradiol	Potent oestrogen	Natural hormone	tested
4-tert-pentylphenol	Weak oestrogen	Industrial chemical	tested
Fadrozole	Aromatase inhibitor	Pharmaceutical	tested
Prochloraz	Aromatase inhibitor	Pesticide, other uses	tested
Flutamide	Anti-androgen	Pharmaceutical	tested
Trenbolone	Androgen (non-aromatisable)	Steroid anabolic	tested
Potassium permanganate	Oxidizer (endocrine negative)	Disinfectant	tested
n-Octanol	Non-polar narcosis (endocrine negative)	Several, including solvents	tested
Vinclozolin	Anti-androgen	Pesticide	Not tested: too expensive in a flow-through system (pure form)
ZM 154-189	Anti-oestrogen	Pharmaceutical	Not tested: difficult to obtain*
Fenarimol	Aromatase inhibitor	Pesticide	Not tested: too expensive in flow-through system (pure form)

Methyltestosterone	Aromatisable androgen	Pharmaceutical	Not tested: not a good reference androgen because of the aromatisation

Table 3: Chemicals tested in Phase 1A, Phase 1B and Phase 2(*but shown to be active in a closely related static-renewal 21d fish protocol (Panter et al., 2000)).

One comment from some member of the VMG-eco was that weakly active substances were not sufficiently represented in the validation studies. Other members of the VMG-eco argued that more data were available from closely related validation programmes (USEPA, 2005; MoE Japan, 2005). The VMG-eco agreed to perform a catch-up study (Phase 3) where additional substances, weakly active or challenging negative, selected from those closely related validation programme would be tested: 4-tert-octylphenol, androstenedione and pentachlorophenol. The WNT also agreed with the collection of additional data, but recommended not to delay the peer-review, as the main objectives of the validation, i.e. demonstration of the relevance and reproducibility of the assay, were covered. The results of Phase 3 will be brought to the attention of the peer-review panel when they become available.

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. Historically this has not been common practice in ecotoxicology, however, today it is widely recognised as an important aspect of test method validation (e.g. for effluent or sediment ecotoxicity testing protocols). In Phase 1A and Phase 1B, none of the substances tested and none of the samples of biological material was coded.

- in the case of vitellogenin, the biological samples are read by a spectrophotometer measuring the optical density, and there is limited chance to introduce a bias.

- for secondary sex characteristics, due to the fact that the counting is done by a technician who is potentially aware of the expected response, it is true that bias could be introduced in the results. However, one could argue that counting and reporting a number is different from giving a personal judgement on the increase/decrease or variation in parameters evaluated.

To overcome this, it was agreed that the catch-up validation study, Phase 3, will be conducted with coding. The results of this study will be reported in the course of 2007. Initial results have demonstrated that the assay performed well, i.e. endpoints responded following the expectations.

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.

It is uncertain whether this criterion applies to new test method, or exclusively to replacement methods, where the performance of the replacement method has to be evaluated against the existing method.

In the case of the 21-day fish screening assay, there is no "existing" OECD Test Guideline to which the assay may be compared. Some comparison may be done with published long term studies (life-cycle type studies), however, there are not many of these available in the public domain so far. In a Japanese initiative, 7 substances have been evaluated in fish (21-day VTG assays, partial life-cycle test not incl. reproduction, and life-cycle test). The Table 4 below provides a summary of key findings:

Chemical	Vitellogenin Assay	Partial Life Cycle (extended early life-stage)	Full Life Cycle
17β-Estradiol	↑ ♂ 14 & 21-d 22ng/L	↓ BL & BW 71 ng/L; ↑ testis-ova ♂ 24 ng/L; ↑ LSI ♀ 2.5 & ♂ 7.5 ng/L; ↑ VTG ♂ 0.9 & ♀ 24 ng/L	**F_0** ↓ fecundity 28 ng/L; ↓ egg fertilization 8.7 ng/L; ↑ testis-ova ♂ 28 ng/L; ↑ LSI ♂ & ♀ 28 ng/L; ↑ VTG ♂ 8.7 ng/L **F_1** ↑ body length 8.7 ng/L; no male or female VTG change at 8.7 ng/L
Ethinyl Estradiol	↑ ♂ 21-d 15 ng/L ↑ ♀ 21-d 61 ng/L	↑ GSI ♂ 93 ng/L; ↑ testis-ova ♂ 25 ng/L; ↑ LSI ♂ 25 ng/L; ↑ VTG ♂ 25 & ♀ 93 ng/L	**F_0** ↑ time to hatching 94 ng/L; ↑ testis-ova ♂ 29 ng/L; ↓ egg fertilization 9.3 ng/L; ↑ VTG ♂ 9.3 ng/L **F_1** ↑ VTG ♂ as low as 1 ng/L
Bisphenol A	↑ ♂ 14-d 72 µg/L and 21-d 334 µg/L	↑ hatching time 4,410 µg/L; ↑ testis-ova ♂ 890 µg/L; ↑ LSI ♀ 2,120 & ♂ 4,410 µg/L; ↑ VTG ♂ 470 & ♀ 2,120 µg/L	**F_0** ↑ mortality1,185 µg/L; ↑ LSI ♂ 1,185 ng/L **F_1** ↑ body weight & length 1,185 µg/L; ↑ testis-ova ♂ (2/10) 1,185 µg/L; ↑ VTG ♂ 1,185 µg/L
p,p'-DDE	↑ ♂ 14 & 21-d 54 µg/L	↑ time to hatching & mortality 32 µg/L; ↑ testis-ova ♂ 32 µg/L; erratic LSI; ↑ VTG ♂ 32 & ♀ 111 µg/L	In progress
o,p'-DDT	↑ ♂ 14-d 3.2 µg/L and 21-d 1.5 µg/L	Absolute ↑ mortality; ↓ body length 3.4 µg/L; ↓ GSI ♀ 1.7 µg/L; ↑ VTG ♀ 0.83 µg/L	**F_0** ↑ testis-ova ♂ (2/10) 0.5 µg/L; ↓ egg fertilization 0.5 µg/L; ↑ VTG ♂ 0.5 µg/L **F_1** ↑ increased time to hatching; ↑ testis-ova ♂ (2/18) 0.52 µg/L; no change in male VTG at any dose
Di-*n*-butylphthalate	No effect 21-d in either ♂ or ♀ up to 822 µg/L	↓ hatchability 850 µg/L; ↑ mortality 73 µg/L (***100% mortality at post-hatch day 8 850 µg/L***)	**F_0** ↑ testis-ova ♂ (2/8) 233 µg/L **F_1** ↑ hatchability 233 µg/L; ↑ time to hatching 75 µg/L; Erratic increases in VTG for both ♂ & ♀
4-Nonylphenol	↑ ♂ 21-d 22 µg/L	↓ body length 45 µg/L;	**F_0** ↓ hatchability 183 µg/L; ↑ mortality 18

	↑ ♀ 21-d 118 µg/L	↓ body weight 24 µg/L; ↑ testis-ova ♂ 11.6 µg/L; ↑ VTG ♂ & ♀ 11.6 µg/L	µg/L; ↑ GSI ♀ 8 µg/L, no VTG data in this generation **F_1** ↑ testis-ova ♂ (5/9) 233 µg/L; no VTG data
4-*t*-Octylphenol	↑ ♂ 21-d 190 µg/L ↑ ♀ 21-d 3,300 µg/L	24 µg/L dose unexplained ↑ mortality, ↓ body length & weight; ↓ GSI ♀ 94 µg/L; ↑ testis-ova ♂ 11.4 µg/L; ↑ VTG ♂ 11.4 µg/L & ♀ 48 µg/L	**F_0** ↓ fecundity 82 µg/L; ↑ testis-ova ♂ 30 µg/L; ↑ LSI ♂ 30 µg/L; ↑ VTG ♂ 10 µg/L & ♀ 30 µg/L **F_1** ↑ testis-ova ♂ 30 µg/L; ↑ VTG ♂ & ♀ 10 µg/L

↑: significant increase ($p<0.001$ or $p<0.05$): ↓: significant decrease ($p<0.001$ or $p<0.05$)

Table 4: VTG, partial and full life-cycle tests performed on identical substances (Ministry of Environment, Japan, 2005).

Other data sources may exist which could add to the comparison of the fish screening assay versus longer-term studies; however the term "performance" does not seem exactly appropriate.

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.

Most participating laboratories replied that they conducted the Phase 1A, Phase 1B and Phase 2 in compliance with GLP. However, not many 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.

Manuscripts are in preparation for submission to a scientific journal. Additionally, validation reports are already declassified (Phase 1A and Phase 1B) or will be declassified in the near future (Phase 2) by the Joint Meeting and thus publicly available.