
ECETOC

ECETOC Document No. 39

**Screening and Testing Methods for
Ecotoxicological Effects of
Potential Endocrine Disrupters:
Response to the EDSTAC
Recommendations and a Proposed
Alternative Approach**

January 1999

**Screening and Testing Methods for
Ecotoxicological Effects of
Potential Endocrine Disrupters:
Response to the EDSTAC
Recommendations and a Proposed
Alternative Approach**

ECETOC Document No. 39

January 1999

Screening and Testing Methods for Ecotoxicological Effects of Potential Endocrine Disruptors: Response to the EDSTAC Recommendations and a Proposed Alternative Approach

CONTENTS

SUMMARY	1
INTRODUCTION	3
1. CONCEPTUAL FRAMEWORK.....	5
2. TIER 1 SCREENS	8
2.1 FROG METAMORPHOSIS ASSAY	9
2.1.1 Description of proposed assay.....	9
2.1.2 Advantages/disadvantages of the method.....	9
2.1.3 Alternative approaches	10
2.1.4 Conclusions.....	11
2.2 FISH GONADAL RECRUDESCENCE ASSAY	12
2.2.1 Description of proposed assay.....	12
2.2.2 Advantages/disadvantages of the method.....	12
2.2.3 Alternative approaches	14
2.2.4 Conclusion	15
3. TIER 2 TESTS.....	17
3.1 AVIAN REPRODUCTION TESTING.....	17
3.1.1 Description of proposed assay.....	17
3.1.2 Advantages/disadvantages of the method.....	17
3.1.3 Alternative approaches	18
3.1.4 Conclusion	19
3.2 FISH FULL LIFE-CYCLE TEST	19
3.2.1 Description of the proposed assay.....	19
3.2.2 Advantages/disadvantages of the method.....	20
3.2.3 Alternative approaches	22
3.2.4 Conclusion	23
3.3 MYSID LIFE-CYCLE TEST	24
3.3.1 Description of the proposed assay.....	24

3.3.2 Advantages/disadvantages of the method	24
3.3.3 Research requirements.....	25
3.3.4 Conclusion	27
3.4. AMPHIBIAN DEVELOPMENT AND REPRODUCTION.....	27
3.4.1 Description of the assay.....	27
3.4.2 Alternative approach	27
3.4.3 Conclusion	28
4. STANDARDISATION, VALIDATION, METHODS DEVELOPMENT AND RESEARCH.....	29
4.1 EDSTAC-PROPOSED APPROACH	29
4.2 ECETOC RECOMMENDATIONS.....	29
4.2.1 Selection of test compounds.....	29
4.2.2 Inter-species comparisons.....	30
4.2.3 Selection of reference chemicals.....	30
4.2.4 OECD activities	30
4.2.5 Conclusions.....	31
BIBLIOGRAPHY	32
AUTHORS OF THE REVIEW DOCUMENT	35
MEMBERS OF THE SCIENTIFIC COMMITTEE.....	36

SUMMARY

The scientific rationale for, and the practicalities of implementing, the EDSTAC recommendations for endocrine disrupter (ED) screening and testing with ecotoxicological (non-mammalian wildlife) species are evaluated. Overall, the major points of weakness in EDSTAC include [1] the unnecessary duplication of mammalian and wildlife assays for androgen, oestrogen and thyroid activity; and [2] the lack of sufficient representation of environmental exposure considerations when moving through the screening and testing tiers for wildlife species. In view of these significant weaknesses, an alternative approach for wildlife has been developed under the ECETOC Screening and Testing Strategy for Environmental Endocrine Modulators (ESTEEM). At the Tier 1 screening level, the ESTEEM framework combines an *in vivo* fish assay for androgens and oestrogens (and their antagonists), supplementing the suite of mammalian assays. The ESTEEM proposal also includes two subsequent testing tiers (as opposed to the one testing tier recommended by EDSTAC) which gives appropriate emphasis to environmental exposure considerations and also to the data provided on ED activity by *in vitro* screens and the extensive battery of tier 1 mammalian and fish screening assays.

In principle, ECETOC agrees with the EDSTAC recommendation for tier 1 screen with an egg-laying species (namely, fish) due to the distinct reproductive roles of androgens and oestrogens in oviparous species versus mammals. However, the proposed EDSTAC tier 1 screening assays for wildlife, namely the frog metamorphosis assay and fish gonadal recrudescence assay are not supported by the available information. The tier 1 screening frog metamorphosis assay, proposed by EDSTAC to specifically study (anti-)thyroid activity, duplicates aspects of the mammalian *in vivo* screening assays within EDSTAC tier 1. Due to this unnecessary duplication, ECETOC recommends that the frog metamorphosis assay should not be used for tier 1 screening and should only be considered for tier 2 testing, if concern is raised both by a positive response in the mammalian thyroid screen and there is potential for environmental exposure. Extended amphibian development tests (ESTEEM tier 3) may be carried out if results in the amphibian larval metamorphosis test (ESTEEM tier 2) are positive, again with due consideration of environmental exposure potential.

The EDSTAC recommended fish gonadal recrudescence assay is still in a very early conceptual phase, there is no indication of the sensitivity to EDs and it is considered to have significant scientific and technical flaws for routine screening. To some extent, EDSTAC recognises this significant problem in stating that no validated methods exist at present for the screening for endocrine disruption in fish. Currently, several international research institutes are actively developing alternative fish screening methods for practical application using the existing OECD fish species. For example, ongoing research funded by the European chemical industry is now evaluating vitellogenin induction and other biochemical endpoints in juvenile fathead minnows exposed to test substances (androgens, oestrogens and aromatase inhibitors) for less than 21 days. Other complementary approaches are

being explored in North America and Japan, using fathead minnow and medaka, respectively. Importantly, a recent OECD Fish Consultation Meeting (London, October 1998) agreed that the results of these international ED research efforts should be discussed at a special OECD meeting scheduled for autumn 1999.

For higher tier testing, the fish full life-cycle (FFLC or fish multi-generation) protocol is agreed to represent a reasonable approach for evaluating developmental and reproductive effects in fish exposed to EDs. Given the existing knowledge of reproductive and developmental effects of many chemicals in fish, and the practical need to conduct partial life-cycle tests before moving to the very challenging FFLC protocol, ECETOC proposes that shorter-term tests such as the fish early life stage and partial life-cycle tests are a more pragmatic means of detecting effects to a similar extent as can be determined in a FFLC test protocol. In view of these scientific and practical considerations, it is therefore more cost-effective to consider an alternative tiered approach to EDSTAC, first incorporating extended early life-stage and partial life-cycle tests with fish (ESTEEM tier 2) before moving on to the 'gold standard' FFLC (ESTEEM tier 3). Again, data on ED activity from all previous assays, together with environmental exposure data should be employed in considering the possible need to move beyond ESTEEM tier 2.

For higher tier avian reproduction testing, within both the EDSTAC and ESTEEM approaches, modifications and validation of the existing protocols may result in appropriate test methods for this purpose. ECETOC considers that the current OECD activity towards revision of the avian reproduction test guideline 206 (using Japanese quail) is the optimal way to identify a cost-effective avian reproduction protocol. Further discussion is required, however, to identify any new endpoints that may be considered useful for detecting ED activity in birds.

For aquatic and terrestrial invertebrate species, there is limited knowledge of the role of androgens, oestrogens and thyroid hormones (or related hormones such as ecdysteroids) in their reproductive and developmental systems. While EDSTAC recommends a tier 2 test for mysid reproduction, the scientific basis for such a recommendation is weak since there is very limited evidence for crustacean species being affected by substances acting on the relevant steroid hormonal systems. ECETOC therefore suggests that decisions on appropriate invertebrate species and higher tier testing procedures are deferred until the report and recommendations are available from the international SETAC expert workshop 'Endocrine Disruption in Invertebrates: Endocrinology, Testing and Assessment' (EDIETA) held on 12 - 15 December 1998.

INTRODUCTION

The European Centre for Ecotoxicology and Toxicology (ECETOC), is a scientific, non-commercial association financed by the European chemical industry. In 1995 ECETOC formed the Environmental Oestrogens Task Force *inter alia* to identify the best ways to detect and assess endocrine disrupting (ED) chemicals (ECETOC, 1996) and to assist in the development of reliable methods of human and wildlife ED risk assessment. For the latter environmental aspects, the Task Force subsequently established a Wildlife Working Group (WG1E).

In 1996, the United States Congress mandated the EPA to establish the Endocrine Disrupter Screening and Testing Advisory Committee (EDSTAC), with multi-stakeholder inputs from academia, government, industry and various public interest groups. The US Congressional mandate specifically required reporting by August 1998 of the EDSTAC recommendations for a screening and testing strategy applicable to human and environmental risk assessment processes (EDSTAC, 1998; EPA, 1998). In turn, the EDSTAC recommendations are required by Congress to be implemented by August 1999 with a further report to Congress on the progress made by August 2000. In view of the significance for industry of the EDSTAC recommendations for wildlife screening and testing, the ECETOC WG1E has reviewed the published report from the perspective of scientific rationale and practical feasibility. In particular, ECETOC has commented on the important areas of *in vivo* testing procedures for wildlife (non-mammalian) and the testing strategies (for EDSTAC tier 1 screening and tier 2 testing) in the overall context of a tiered testing scheme and its role in the environmental risk assessment of EDs (EDSTAC Chapter 5 and Appendix Q).

The *in vitro* assays recommended for screening and prioritisation are mainly in the domain of screening for mechanistic actions in human cell lines or organs in order to obtain valuable information on the intrinsic mode of action. These *in vitro* test systems are not addressed by this ECETOC report, since it has been widely agreed at several international workshops that the assessment of effects in wildlife species should focus on *in vivo* studies (Tattersfield *et al*, 1997; Ankley *et al*, 1998).

ECETOC fully agrees with the principal concept of a tiered ED evaluation programme, in which prioritisation and initial assessment is performed on the basis of short-term, less-complex screening and testing protocols. In contrast to the EDSTAC recommendations, however, ECETOC proposes that greater emphasis is placed on the triggers for the conduct of higher tier, long-term and complex tests since these have not been adequately described in the EDSTAC report. In particular, the important aspects of environmental exposures to chemicals need to be given greater transparency than is apparent in the EDSTAC recommendations. It is such scientific concern that has led to the development of an alternative approach in the form of ECETOC's Screening and Testing Strategy for Environmental Endocrine Modulators (ESTEEM).

This report addresses each of the EDSTAC recommendations for *in vivo* ecotoxicological (wildlife) assays for screening and testing for EDs, with each assay being briefly described and the advantages and disadvantages discussed. Where appropriate, ECETOC has described alternative approaches according to both scientific, practical and ethical (animal testing) considerations.

1. CONCEPTUAL FRAMEWORK

The fundamental EDSTAC concept of using a tiered screening and testing programme, in which prioritisation and initial assessment is performed on the basis of short-term, less-complex assays is fully supported. Focusing on the key elements of ecotoxicological relevance, the strategy proposed by EDSTAC is summarised in Figure 1. In view of concerns on the ecotoxicological assays suggested in EDSTAC tiers 1 and 2, together with the limited number of tiers and uncertainty over the triggers for conducting the more complex tests, WG1E has developed an alternative approach. ESTEEM aims to help address these concerns (summarised in Figure 2).

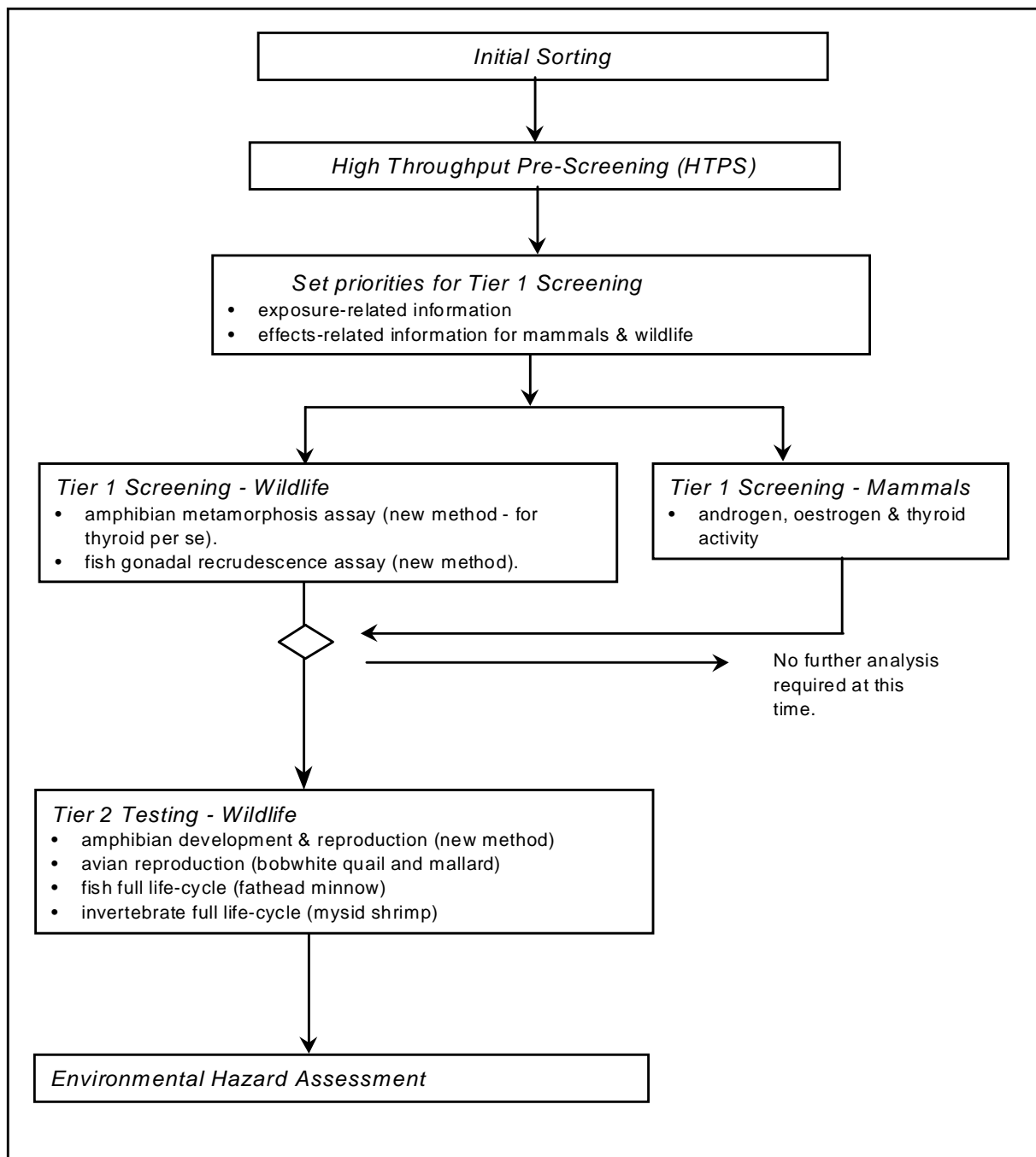
Key elements of the ESTEEM approach are:

- to include greater consideration of environmental exposure information when deciding which types (e.g. avian, fish or invertebrate) of long-term ED tests may be required at the higher tiers (Tillitt *et al*, 1998);
- to rationalise the EDSTAC recommendation to move to technically challenging multi-generation tests (e.g. fish full life-cycle) directly after the tier 1 screening, without first recognising the value of partial life-cycle tests.

The various ecotoxicological screening and testing protocols recommended by EDSTAC are discussed in greater detail in the subsequent sections of this document. In these discussions, a distinction is made with regard to the broad state of knowledge for a given protocol and the degree of standardisation currently available. The following concepts have proved useful in discussions to date:

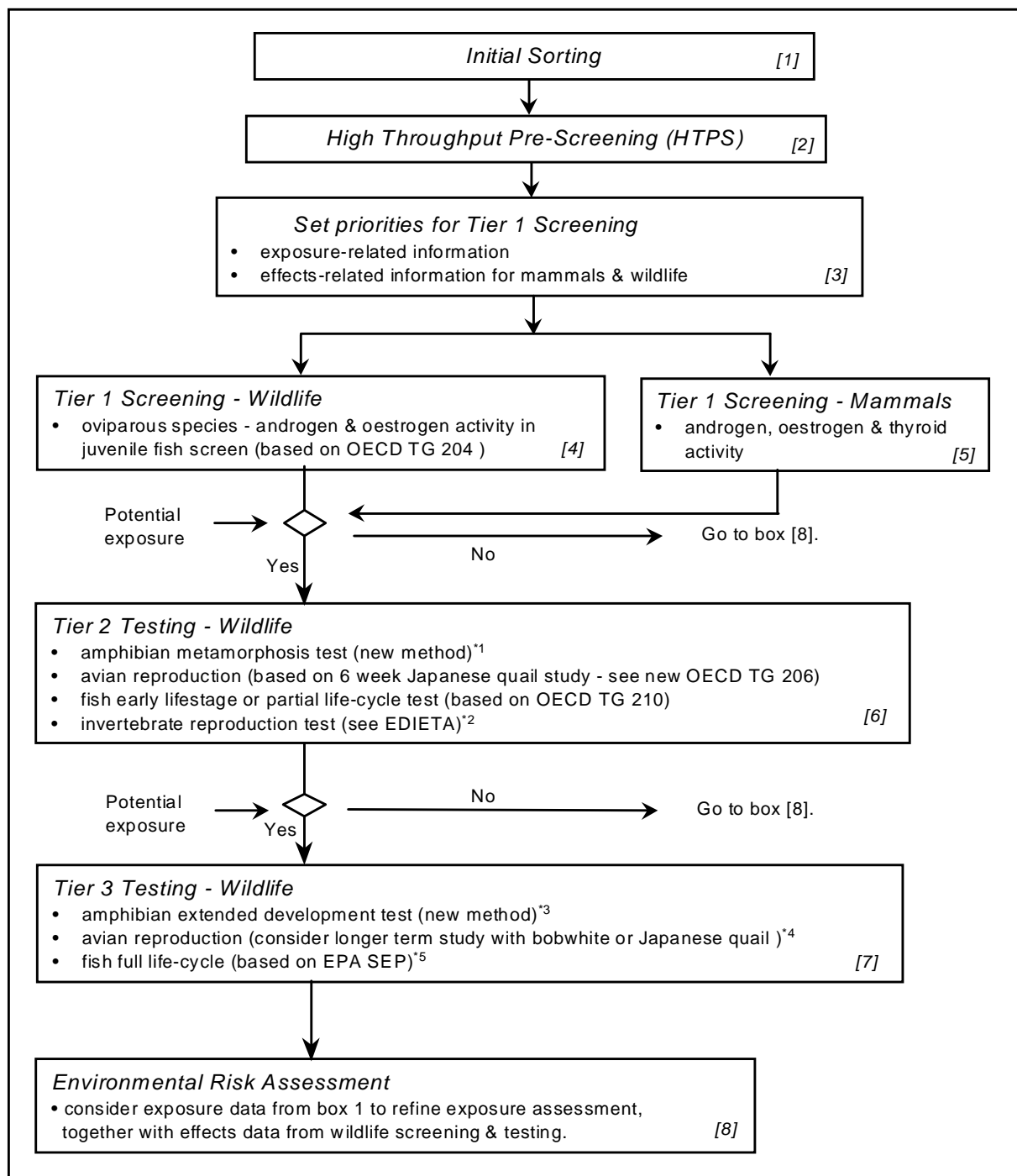
1. conceptual protocol - a novel protocol which has not yet been actually conducted and is not yet reported in the peer-reviewed scientific literature. No validation data are therefore available for the protocol at this preliminary stage;
2. developmental protocol - a protocol which has been published in the peer-reviewed scientific literature and which forms the basis for further research (e.g. using a broader range of reference ED substances) towards the development of a reliable method. No validation data are normally available at this interim stage;
3. validated protocol - a protocol that has been demonstrated to be reliable and reproducible as a result of appropriate intra- and inter-laboratory comparisons (e.g. ring- testing). Protocols need to achieve this status before they can be reliably employed for regulatory purposes, for example, within the OECD test guideline programme.

Figure 1. Summary of EDSTAC Screening & Testing Scheme for Wildlife.



Notes: ◇ decision point (EDSTAC criteria for decision making undefined).

Figure 2. Draft ECETOC Screening & Testing Strategy for Environmental Endocrine Modulators (ESTEEM).



- Notes:
- 1 - only required if mammalian screens show evidence of thyroid disruption.
 - 2 - addressed at OECD-SETAC workshop (EDIETA) 12 - 15 December 1998.
 - 3 - only required if active in amphibian larval metamorphosis test at Tier 2.
 - 4 - only required if positive in avian reproduction test at Tier 2.
 - 5 - only required if active in fish early lifestage test at Tier 2.
- ◇ decision point.

2. TIER 1 SCREENS

2.1 FROG METAMORPHOSIS ASSAY

2.1.1 Description of proposed assay

The frog metamorphosis assay is presented in Chapter 5, Section III of the EDSTAC report. The main objective of the inclusion of this *in vivo* assay into the EDSTAC tier 1 screening battery appears to be to detect activity of thyroid active chemicals *per se*, rather than to represent amphibians as a taxonomic group. Thyroid hormones play an important role in vertebrate developmental and reproductive endocrinology and ECETOC agrees with the principle that identifying potential effects of chemicals on the thyroid system forms a key part of an ED screening and testing strategy.

The EDSTAC report outlines a protocol in which the larvae (tadpoles) of the African clawed frog *Xenopus laevis* (age 50 - 64 days) are exposed to the test substance for 14 days. The main endpoint observed is the tail resorption rate. It is suggested that the measurement is made using computer-aided video image analysis. Importantly, EDSTAC appears to be recommending a novel conceptual protocol. ECETOC is unaware of peer-reviewed publications on the metamorphosis assay, although a related approach has been applied to embryo development in the same species (Fort and Stover, 1997). Apparently aware of such uncertainties, EDSTAC (Chapter 5, Section VII) recognises that more research must be conducted to determine whether the *Xenopus* metamorphosis assay can be adequately developed and validated for a potentially useful role within the ED screening and testing programme. Further specific comments are summarised below.

2.1.2 Advantages/disadvantages of the method

The frog metamorphosis assay is at the conceptual stage of development. Significant research is needed using a range of thyroid agonists and antagonists in order to evaluate the scientific value of the assay prior to embarkation on the necessary validation of the protocol. The description of the assay provided within the EDSTAC report is abbreviated and does not give important details of experimental design (e.g. protocol design, numbers of animals required, sensitivity to thyroid disrupters, frequency of measurements).

The advantages of the system are:

- *Xenopus laevis* is a well established laboratory organism;
- the reported ease of handling the test organisms.

The disadvantages are:

- this assay is conceptual and has yet to be reported in the peer-reviewed scientific literature;
- while the mechanisms for tail resorption are well understood, there is a critical lack of development and validation of this assay;
- evaluation of tail resorption using computer-aided image analysis is technically complex;
- the interpretation of the rate of tail resorption requires a reference database on the variability of this parameter under controlled conditions; ECETOC is unaware of such a database;
- developmental toxicants other than thyroid-hormone disrupters may also affect gross morphological parameters such as tail resorption, hence measurement of solely a morphological endpoint is unlikely to be adequate to detect specifically thyroid agonists and antagonists.
- thyroid disruption may be detected in mammalian assays described in the EDSTAC tier 1, hence the adoption of this proposal would lead to an unnecessary increase in animal testing.

2.1.3 Alternative approaches

The objective of tier 1 screening is to detect chemical substances or mixtures capable of interacting with androgen, oestrogen, or thyroid hormone systems (EDSTAC Chapter 5, Section I). In terms of wildlife screening, an exhaustive assessment of a variety of different species is not the scope of tier 1 screening tests. Available information does not suggest that a specific mode of action or a fundamentally different system of hormonal function is affected in amphibians compared to other vertebrates (e.g. mammals, birds and fish). Hence, an alternative approach is to recommend that mammalian testing be used at the tier I screening level (Figure 1) for thyroid effect. If thyroid activity is detected in the mammalian tier 1 screen, and if the environmental fate predictions suggest significant exposure to aquatic organisms, then a suitably-validated frog metamorphosis assay may be considered for use at ESTEEM tier 2 (Figure 2).

The physiological processes described below suggest that a number of endpoints may be indicative for thyroid effects. However, these are not specific to amphibians, but represented in many vertebrate taxa. For example, Gorbman (1983) states that all vertebrates form and secrete comparable hormones from their thyroid gland, namely thyroxine (T4) and triiodothyronine (T3). Many organs, physiological systems, and metabolic processes are also influenced by these thyroid hormones. Morphological effects of thyroid hormones include metamorphosis in both amphibians and non-amphibian vertebrates (e.g. marine flatfish), post-embryonic maturation, developmental differentiation and growth. Physiological and metabolic effects are seen in the nervous system and

neuromusculature, digestive tract, kidney and gill, oxygen consumption, body-temperature regulation and nutrient metabolism.

Mechanisms of thyroid hormone action vary between species and among tissues. Two types of intracellular binding receptor proteins, one in the mitochondria and one in nuclei have been identified. Recent evidence suggests that a true mitochondrial T3 receptor has not been demonstrated. Nuclear receptors consist of variants that can be classed as α - and β -variants (McNabb, 1992). The frog thyroid receptors (TRs) are highly homologous to TRs of other vertebrates (Kanamori & Brown, 1996). In all species studied to date, α -TR is expressed early in development and is ubiquitous in its expression. The respective roles of the α - and β -TRs in mammals remains to be determined.

Studies in populations of Great Lakes salmon have demonstrated that fish can be affected by thyroid-disrupting chemicals (Leatherland, 1992; Kime 1998). These observations in wild fish suggest that there is also an important opportunity to detect thyroid disruption in laboratory fish populations. Subject to further research, this may allow the incorporation of thyroid relevant endpoints in ESTEEM tier 2 and 3 fish tests as an alternative to amphibians. Such a possibility would be attractive in terms of minimising the overall numbers of animals required for ED testing.

A number of reviews (Brown *et al*, 1995; Hayes and Wu, 1995; Kanamori and Brown, 1996) describe the role of thyroid hormones (TH) on amphibian metamorphosis and the effect of exogenous thyroid hormones and thyroid hormone synthesis inhibitors. It is widely recognised that premature metamorphosis in tadpoles can be stimulated by the addition of TH to the aquarium water, with one of the first signs of metamorphosis being the growth of limbs. In the absence of TH, tadpole limb buds still form but these structures do not progress beyond the bud stage. During normal development, TH levels rise resulting in growth and differentiation of the limbs. Tail resorption is the last of the gross morphological changes occurring during metamorphosis, when TH levels are highest and occurring between 24 and 48 h following the addition of exogenous TH. Metamorphosis is reported to be probably initiated by binding of TH to α -TR. However, β -TR is presumed to play a later role in developmental programmes such as tail resorption. Notably, steroids such as corticoids can modulate TH activity during amphibian metamorphosis (enhancement of tail resorption) (Brown *et al* 1995; Hayes and Wu, 1995; Kanamori and Brown, 1996).

2.1.4 Conclusions

It is clear from the above discussion that potential effects on the thyroid system should be detected in principle by mammalian tier 1 screening, as recommended in the EDSTAC testing strategy (EDSTAC report, Chapter 5) and currently being evaluated by the OECD-Endocrine Disruptors Testing and Assessment working group (OECD, 1998). A specific tier 1 screen for thyroid activity *per se* in

amphibians is not scientifically warranted for solely hazard identification and simply duplicates data which can be obtained from other screening assays. Such a duplication of screening endpoints is difficult to justify in terms of the use of additional animals.

In summary, the decision to include a frog metamorphosis assay in ESTEEM tier 2 (Figure 2) should be based on a positive result from the mammalian thyroid screen, and only if significant exposure of aquatic ecosystems is predicted. The development of a comparative database on the sensitivity of fish and amphibians to thyroid disruptors is also a priority for research. For example, if such a database suggested that fish larval development (e.g. in an early life-stage test) was also affected by thyroid agonists and antagonists, then amphibian testing might no longer be required at ESTEEM tier 2 and this would fit with the acknowledged need, from both ethical and economic perspectives, to reduce animal testing.

2.2 FISH GONADAL RECRUDESCENCE ASSAY

2.2.1 Description of proposed assay

The recommendation for an *in vivo* screening assay based on fish gonadal recrudescence is outlined in Chapter 5, pages 5-29 of the EDSTAC report. In this conceptual approach, it is proposed that mature fathead minnows (*Pimephales promelas*) of both sexes be held under unfavourable conditions for reproductive activity (short day length and low temperatures; 'winter phase') and then be subjected to an increasing day length and temperature regime ('summer phase') to stimulate gonadal maturation and the development of secondary sexual characteristics. The proposed assay includes assessment of gamete quality, fecundity, gross morphological examination of the gonads and the measurement of relative gonad weight (gonadal somatic index; GSI). Additionally, EDSTAC recommends the measurement of plasma vitellogenin since this is a widely-used biomarker of ED exposure in fish (Ankley *et al*, 1998; Panter *et al*, 1998).

2.2.2 Advantages/disadvantages of the method

The disadvantages are:

- the fish gonadal recrudescence assay is proposed as a screening test by EDSTAC. There is, however, no published fathead minnow protocol for this conceptual assay and hence the database underlying this approach is minimal;
- unlike other approaches, which utilise existing FIFRA and OECD test guidelines (e.g. OECD test guidelines 204 or 210) (OECD, 1993) as a basis for the enhancing existing protocols for ED substances (Tyler *et al*, in press), the fish gonadal recrudescence assay has no obvious link to an existing regulatory protocol. To ECETOC's knowledge, there are no published data using the

- gonadal recrudescence approach with OECD test species (e.g. fathead minnow, medaka or zebrafish);
- unlike the standardised FIFRA and OECD short-term tests with fish (e.g. OECD test guideline 204), the method requires a variable environmental regime, experience in gross pathology for gonad analysis and an extensive practical knowledge of the reproductive cycle laboratory fish populations which will need to be specifically cultured for this purpose;
 - the practical use of endpoints such as gamete quality and fecundity in such a screening assay is fraught with difficulty, due to the fact that commonly-used laboratory species (e.g. fathead minnow, medaka and zebrafish) have widely variable fecundity between breeding pairs, even within control populations;
 - the term recrudescence indicates that reproductively-mature fish will be used (namely, those fish that have already spawned once before the onset of the simulated 'winter phase') and then must undergo another period of gonadal growth before spawning again;
 - the fathead minnow is the recommended species for this conceptual protocol, although other species of fish may be used. The need to have sufficient synchronously spawning fish to conduct such screening assays would cause extensive practical difficulties;
 - this approach would necessitate that breeding pairs of fathead minnow are minutely observed and detailed records kept of spawning periodicity. On account of the natural variability of gonadal growth and spawning periodicity between breeding pairs it will be in practice very difficult synchronously to expose sufficient numbers of fish at the correct stage of development. This natural variability in mature fish will also hamper the development of a standardised screening protocol as interpretation of the results of the gonadal recrudescence assay will have to take into account the variability in spawning periodicity;
 - the common laboratory species recommended by OECD such as *Danio rerio* (zebrafish), *Pimephales promelas* (fathead minnow) or *Oryzias latipes* (medaka) have different sexual behaviour and requirements for sexual reproduction. For example, zebrafish have been reported to display a normal juvenile hermaphrodite stage (Takahashi, 1977). More information is needed to enable selection of the most appropriate range of species for conducting this test;
 - the sexes in fathead minnow are usually very distinct. As in other fish species, however, secondary sexual characteristics in males develop differently as dominant (territorial) males appear during group culturing, while others show less obvious secondary sexual characteristics (Pyron and Beiting, 1989). This behavioural plasticity may have a significant influence on the parameters recommended for assessment by EDSTAC;

- the sexing of medaka (*Oryzias latipes*) is also not difficult. For this species, however, it is not known whether the laboratory conditions under which the fish are cultured during the phase prior to gonadal maturation may influence the reproductive parameters which have been recommended by EDSTAC for evaluation;
- in zebrafish, in contrast to fathead minnow or medaka, the sex of the fish is difficult to determine when they are not in a reproductive stage. On the other hand, zebrafish can produce eggs even at sub-optimal conditions and it is therefore difficult to maintain this species in a non-reproductive stage (Laale, 1977). Furthermore, the reproductive assays with zebrafish may incur severe problems after the fish have been reproducing for some time, due to significant increases in the frequencies of developmental abnormalities (Laale, 1977; Piron, 1978; Newsome and Piron, 1982; Rundle 1996);
- without prior experience, or appropriate standardisation and validation, there is little evidence to demonstrate that a fish gonadal recrudescence assay could be robust, practical, or sensitive to EDs, hence this aspect of the EDSTAC recommendations appears to be weak.

2.2.3 Alternative approaches

It has been argued that the gonadal recrudescence assay is the preferred screening method for fish, since the period of gonadal growth would probably be the most sensitive period to endocrine disruption. However, there is evidence that other periods of development could also be sensitive to endocrine active substances. Importantly, positive features of a screening assay include: sensitivity and specificity to EDs, potential for screening many substances and practicality. By analogy with the widely-used immature rodent uterotrophic assay, the use of juvenile fish as the basis of a screening assay has many advantages and should be given serious consideration as an alternative to the EDSTAC proposal for the adult fish gonadal recrudescence assay. For example, it is easy to obtain supplies of juvenile fish at the same stage of development, hence it is practical to expose sufficiently large numbers of juvenile fish to the test substances. Also, the use of juvenile fish avoids the potential problem associated with adult fish reproductive and territorial behaviour (e.g. where the presence of sexually-dominant males (or females) may lead to behavioural and pheromone modulation of other fish in the laboratory test population) (Pyron and Beitinger, 1989). Once the baseline endocrinology has been established for a given species and stage of development, it is possible to use sensitive and specific endpoints (e.g. vitellogenin induction) to screen for ED activity. Using this approach, a number of research groups have shown that juvenile fish are very sensitive to substances such as oestrogens (Purdom *et al*, 1994; Panter *et al*, 1998; Tyler *et al*, in press).

To further these ideas, an extensive research programme is currently being sponsored by the European chemical industry's Endocrine Modulators Steering Group (CEFIC-EMSG), based on an

ECETOC research proposal entitled "*Research on the effects of endocrine modulators in aquatic organisms: an assessment of different testing approaches*". This programme includes research towards the development of a short-term screening assay with fathead minnows (EMSG project A), in which biochemical parameters such as vitellogenin and endogenous steroid hormone levels are being studied for their applicability as biomarkers of ED activity (Campbell and Hutchinson, 1998, Länge *et al*, 1998; Panter *et al*, 1998).

The aim of this project is to develop the basic science necessary for evaluating short-term *in vivo* effects of ED chemicals in fish in the context of existing OECD test guidelines 204 and 215 (OECD, 1993). Such information is essential for the potential future development of a cost-effective, rapid *in vivo* fish screening assay. The primary endpoint being evaluated for oestrogen activity is vitellogenin induction, however, on-going research is also examining the assessment of other relevant endocrine endpoints in fathead minnows (e.g. sex hormone levels).

Although the parameters identified above for further research are yet to be investigated for their relevance and practicality in a short-term screening assay, there appears to be several advantages in the EMSG's juvenile fish approach over the EDSTAC recommendation to develop a novel fish gonadal recrudescence assay. Using the analogous concept of the immature rat uterotrophic assay, these advantages in juvenile fish include the ready availability of appropriate test organisms, the use of biomarkers which are specifically involved in fish reproductive physiology (e.g. vitellogenin and endogenous sex hormones levels) and the ability to link this approach into the existing philosophy for several OECD fish test guidelines. Importantly, ECETOC notes that while vitellogenin has been demonstrated to be a sensitive biomarker for oestrogen exposure, further research is required on the natural variations in this and other biomarkers (e.g. plasma hormone levels) in order to establish the baseline of normal values. Certainly, the development of the juvenile fish assay and other new methods for ED screening in fish and the subsequent optimisation of existing regulatory guidelines will require extensive validation and standardisation work. This approach was extensively discussed at a recent OECD Fish Consultation Meeting held in London in October 1998, with the agreement to evaluate the EMSG results with those from other groups (e.g. Japan, North America) at a second OECD Fish Consultation Meeting provisionally scheduled for autumn 1999 (OECD, 1999).

2.2.4 Conclusion

A screening assay based on the exposure of juvenile fish, and incorporating relevant endocrine endpoints, would seem a more practical and robust approach to the screening of compounds for endocrine activity than the gonadal recrudescence assay. Conceptually, this approach parallels the 'gold-standard' immature rat uterotrophic assay which has proved so useful for detecting oestrogenic activity in mammals (EDSTAC, 1998). However, the current lack of data with any of the conceptual

assays proposed to date requires that all current proposals for the development of an *in vivo* fish screening assay should be examined with equal rigour; ECETOC believes that it will become clear in the coming months, once experimental work is completed, which approach should go forward for further validation. If the gonadal recrudescence assay could be shown to be workable with selected US and OECD fish species (eg fathead minnow), if the practical difficulties could be minimised, and if a reliable protocol is developed then validated, then the EDSTAC recommendation may lead to a useful screening assay. ECETOC also considers, however, that a scientifically more robust and more practical assay should be sought for screening purposes. On-going EMSG-funded research suggests that the juvenile fish assay is a good candidate for this role.

3. TIER 2 TESTS

3.1 AVIAN REPRODUCTION TESTING

3.1.1 Description of proposed assay

For substances active in the tier 1 screens, avian reproduction testing is recommended by EDSTAC at Tier 2 (see Chapter 5, pages 5-61 and Appendix Q; EDSTAC 1998). Specifically, EDSTAC recommends conducting *in vivo* reproduction tests with two avian species, northern bobwhite quail (*Colinus virginianus*) and mallard duck (*Anas platyrhynchos*), using existing protocols with possible modifications to enhance the ability to detect endocrine-related effects. The existing avian reproduction protocols are well known and it is estimated that they have been used approximately 300 - 500 times. The methods are standardised and regarded as reproducible, hence, protocols for both northern bobwhite quail and mallard duck have *de facto* acceptance by the scientific community as being valid. EDSTAC's recommendations for suggested modifications to the current one-generation avian reproduction protocol include:

- extending the current study design to two generations;
- measuring circulating sex steroid and thyroid hormones;
- monitoring a suite of morphological and histological parameters, functional skills of offspring and reproductive capability of offspring.

3.1.2 Advantages/disadvantages of the method

A definitive test on endocrine effects in birds may be considered appropriate when a relevant exposure of birds is expected. The reproductive function, although generally similarly regulated as in other vertebrates, may be affected by different endpoints such as egg-laying or different hormonal imprinting of males and females. Importantly, the 'default sex' for birds is the male, whereas in mammals the 'default sex' is female (Campbell and Hutchinson, 1998).

Disadvantages

- the extension of the northern bobwhite quail or mallard duck test protocol to a two-generation study makes these tests even longer and more costly. At present, the OECD is progressing the revised test guideline 206 for the Japanese quail (*Coturnix coturnix japonica*), which has a much shorter generation time than either the northern bobwhite quail or mallard duck (OECD, 1997);

- the modifications mentioned in the EDSTAC document for extending the current EPA protocol include endpoints such as plasma-steroid levels, neurobehavioural tests, organ weights, and histopathology. However, these modifications were not specifically related to endocrine parameters and appear to reflect an interest in exploratory research in topics of growth, development, reproductive endocrinology, and behaviour rather than towards their practical use in a regulatory programme aimed at environmental risk assessment;
- EDSTAC describes a number of new avian reproduction endpoints in Appendix Q and states that these were developed by the OECD Avian Reproduction Working Group (EDSTAC 1998). ECETOC members of this OECD Working Group do not consider that this statement is correct and believe that OECD has not discussed the endpoints in Appendix Q.
- the suggested modifications represent non-standard and non-validated methods that are potentially of varying reliability or relevance for detecting biological effects;
- an extensive validation programme with known ED compounds (including agonists and antagonists for androgens and oestrogens) is needed to identify relevancy of parameters in Appendix Q to endocrine toxicity testing in birds;
- extensive control data in the various avian species of potential interest are essential in order to allow reliable interpretation of the many parameters in EDSTAC Appendix Q;
- the sensitivity of each parameter to detect adverse biological effects in birds is not well characterised. Also, the specificity of the various endocrine, behavioural and morphological parameters are unknown among bird species or across chemical groups;
- natural variation in blood-hormone concentrations in most avian species is unknown, thus comparisons of experimentally-derived laboratory data to baseline normal values would not be possible without further research;
- biological significance of experimentally-derived effects in birds may be difficult to define. Thus, a large research effort is needed to provide a scientific context in which to interpret the regulatory meaning of new endpoints proposed for the avian testing tier;
- the use of two avian species does not seem to give relevant additional information. The only reason for using mallard duck is the demonstrated ability to show egg-shell thinning after exposure to contaminants (Lundholm, 1997), however, this has been observed in only a very small percentage of test compounds.

3.1.3 Alternative approaches

The revised OECD Test Guideline 206 is currently based on a one-generation study for Japanese quail, although the aim is to adapt the guideline to both Japanese and northern bobwhite quail. Reproductive parameters such as egg production and egg quality, fertility, embryo survival, hatching success, and growth and development are part of the current draft; gonad histology is at present not included as a requirement. Histology is mentioned by EDSTAC as a possible further endpoint; subject to research, ECETOC supports this flexible approach. In addition, gonadal gross morphology and accessory organ development of offspring may be a valuable addition to the current draft to enable evaluation of the potential effects on the endocrine system of birds (Tattersfield *et al*, 1997).

Other endpoints, such as behavioural tests or steroid levels may be included in the test guideline at a later stage, but not without fundamental research and validation efforts. ECETOC proposes that for both scientific and practical purposes, there is a useful role in ESTEEM tier 2 for the Japanese quail assay (duration ca. 6 weeks), with the longer-term avian studies being more appropriate for ESTEEM tier 3 (Figure 2). Further discussion on this issue is expected at an international workshop on avian reproduction testing scheduled for 1999.

3.1.4 Conclusion

The modification of existing chronic bird test guidelines has relevant merits. However, the potential scientific benefits which may arise from adding new biological endpoints and from growing out a second generation of birds are as yet unknown. It will be useful to consider the relative cost to information gain of the suggested modifications to the avian reproduction protocol. Particularly, the testing of two different species does not seem to have adequate benefits for the additional efforts, and the optimisation of the currently drafted chronic OECD guideline on Japanese quail seems to have the highest potential for a cost-effective test methodology. ECETOC therefore considers that the revised OECD avian reproduction test guideline should be sufficient to detect any significant reproductive effects in ESTEEM tier 2, with triggering to a longer term tier 3 avian reproduction study based on exposure assessment and significant reproductive effects in ESTEEM tier 2 (Figure 2).

3.2 FISH FULL LIFE-CYCLE TEST

3.2.1 Description of the proposed assay

The EDSTAC report (Chapter 5, pages 5-62 and Appendix Q) recommends the use of a full life-cycle test on the freshwater fathead minnow as a definitive tier 2 test for ED effects in fish (Figure 1). Additionally, it is suggested that other species such as the sheepshead minnow should be used when estuarine or marine environments are expected to be exposed to the test compound. In the fish full life-cycle (FFLC) test as standardised by EPA, fathead minnows are exposed to the test compound

from egg (F_0 generation) to early development of F_1 generation offspring, with a test duration of approximately 9 - 10 months. At maturation of the F_1 generation, breeding pairs or groups are formed to promote spawning (USEPA, 1986). The endpoints analysed in the existing FFLC include spawning frequency, number of eggs produced, fertility, viability of embryos, hatching success, growth and development. The EDSTAC recommendations are to add other parameters such as sexual differentiation, sex ratio, gonadal weight and histopathology, sperm motility and egg maturation, and plasma vitellogenin and steroid hormone levels. Behavioural tests are also suggested.

3.2.2 Advantages/disadvantages of the method

A major disadvantage of FFLC tests is their technical complexity. In fact, due to these problems and high cost, FFLC tests are not routinely conducted for chemicals with widespread outdoor exposure and therefore ECETOC does not support the EDSTAC view given in Chapter 5, pages 5-60 of the final report. However, FFLC tests have been employed on occasions where there is significant concern about selected substances and EDs on long-term fish health (Länge *et al*, 1997).

The fathead minnow is considered by ECETOC to be a suitable test species for FFLC due to the large number of chronic tests performed with this species and hence widespread experience in many laboratories which culture fathead minnows. There also exists a large reference database of reproductive parameters (e.g. egg numbers per spawning) in control groups of fathead minnow which is a considerable strength above other non-OECD fish species. The fathead minnow is also considered to be an adequate surrogate for other fish species and the additional use of the sheepshead minnow (where exposure to a particular chemical substance or mixture is predominantly estuarine or marine) or other species appears unwarranted. However, justification for use of alternative fish species could be made, where a specific mode of action is expected to which a particular part of the life-cycle might be particularly susceptible. ECETOC welcomes EDSTAC's performance-based approach to species selection as further research results become available.

EDSTAC proposes that the FFLC test would need to be conducted following a trigger from either the sorting of initial data or from tier 1 screening (Figure 1) (EDSTAC 1998). This jump to such a high-level-definitive test is considered somewhat extreme given the scientific and technical problems, long time frame and associated costs involved in FFLC tests. EDSTAC also appear to have overlooked the practical necessity to conduct an early life-stage test with the same fish species as a prior range-finding study for a FFLC test on a given substance. ECETOC therefore recommends incorporation of an intermediate fish testing tier (see ESTEEM Figure 2) which may include a flexible evaluation of developmental or reproductive effects based upon early life-stage (ELS) or partial life-cycle (PLC) studies, respectively. The main reasons behind ECETOC's recommendation of a three-tiered approach are:

- experience of life-cycle tests with the fathead minnow indicates that shorter tests (e.g. ELS) may detect effects for many organic chemicals to a similar extent as can be determined in FFLC tests;
- the ELS test with fathead minnows is likely to cover the most-sensitive stage regarding sex determination and early development. Further research is essential to explore this issue;
- fertility and ovulation can be tested in a PLC test using mature fish.

The number of substances tested could be greatly increased taking into account the duration and technical challenges of the studies and the capacities of laboratories with experience in FFLC tests. ECETOC estimates that probably ten times more compounds can be tested world wide at ESTEEM tier 2 (Figure 2) compared with EDSTAC tier 2 (Figure 1) (assumed 200 compounds as opposed to 20 per year).

For established ED compounds such as ethinylestradiol, WG1E concluded that there is only limited evidence that ED effects occur in the F_1 generation, which were not already apparent in the F_0 generation (see Länge *et al.*, 1997 and 1998). The conclusion of this FFLC study with a potent xenoestrogen indicates that an extended ELS test has significant potential for detecting developmental effects due to ED compounds. Clearly, more data need to be made available to explore this possibility further for fish ELS and PLC tests on a variety of ED chemicals.

Further validation of the predictability of ELS and PLC studies for full life-cycle effects is required and forms part of the ongoing research programme of the European chemical industry's Endocrine Modulators Steering Group (CEFIC-EMSG), based upon an ECETOC research proposal entitled "*Research on the effects of endocrine modulators in aquatic organisms: an assessment of different testing approaches*". Other groups should also be encouraged to generate data for such validation (see below). ECETOC WG1E welcome the additional endpoints recommended for validation and inclusion at the higher-tier testing levels of ESTEEM (Figure 2).

The addition of new ED-relevant endpoints to study designs requires considerable development and validation before they can be considered useful for regulatory purposes. Subsequently, endpoints may best be targeted at the endocrine mechanisms of interest on a case-by-case basis, depending upon results from lower-tier assays. Ability to relate endpoints at the testing tiers 2 and 3 to those at the screening tier 1 would be beneficial and assist in predictive environmental risk assessments. Hence, the use of a short-term test procedure on fish, in which biochemical parameters such as vitellogenin and endogenous steroid hormone levels are proposed as specific biomarkers (see previous discussion of fish gonadal recrudescence assay) would be desirable in order to make such correlations on a clearer mechanistic basis.

Currently, all fish ELS, PLC and FLC tests include critical elements of growth and development, and thereby are vulnerable to thyroid-related effects (Kime, 1998). In the future, determination of such effects in fish may be sufficient to detect potential effects of thyroid-active compounds (with reference to the mammalian tier 1 data for thyroid activity) and potentially negate the need for duplicative testing with amphibia. This possibility should be reviewed as data become available; comparative research into this aspect is highly recommended.

Determination of genetic sex differentiation in fish may prove useful and would enable the sex ratio of phenotypically-similar offspring to be determined (Campbell and Hutchinson, 1998). Knowledge of the genotypic sex ratio relative to the phenotypic sex ratio would be useful but markers for genetic sex need to be developed for this purpose. A literature review on this aspect has been recently commissioned by the European chemical industry's Endocrine Modulators Steering Group, based upon an ECETOC research proposal entitled "*Research on the effects of endocrine modulators in aquatic organisms: an assessment of different testing approaches*" (ECETOC, 1998).

3.2.3 Alternative approaches

The research programme of the European chemical industry includes long-term studies in fathead minnows. This research involves validation of developmental (extended ELS) and reproduction (PLC) tests in respect of their ability to predict FFLC effects and their potential enhancement with certain additional, but as yet non-standardised endpoints (e.g. biochemical and histological parameters). This research programme is being undertaken in co-operation with other international research organisations and is planned to begin in 1999.

Development assay in fish - the first type of assay under investigation by the European chemical industry. In this approach, the standard procedure for fathead minnow ELS tests (e.g. OECD test guideline 210), where newly-fertilised eggs and the hatched fry are exposed to the test substance for a total duration of 4 weeks post-hatch. An additional feature of this approach will be an extended F_0 study period during which the fish will be held in clean water until maturity and egg laying. This extension will allow the evaluation of the effects on reproduction due to exposure to ED compounds in early developmental stages at modest additional costs.

Reproduction assay in fish - the second type of assay under investigation by the European chemical industry. As a complementary aim using a fish PLC exposure, mature breeding animals are exposed and the test period covers a mating phase of the adult F_0 generation, as well as the development and hatch of the F_1 generation.

The different objectives of these two study types can be summarised as follows:

- in the ELS exposure, primarily the effect of the test substance on the development of the fish embryo and fry is studied, augmented during the research phase with an evaluation of fecundity. The basis for this research is a currently-available standardised test procedure (OECD test guideline 210) with some modifications (e.g. gonad histopathology, biomarker analyses), together with existing research data (Tyler *et al.*, in press);
- in the PLC exposure, the reproductive performance, fertility of adult males and females, and mating behaviour can be studied, in addition to the effects covered by the ELS test. This extends markedly the information gained by the ELS study.

The two test methodologies should be used to give an indication of the sensitivity of the reproductive and the developmental endpoints in comparison. It has to be taken into account that in the fish PLC test, possible effects seen in the F₁ generation may be influenced by the exposure of the F₀ generation.

A literature study on the topic of sex determination and existing markers of genetic sex has been commissioned by CEFIC-EMSG. Some expert academics and laboratories already have experience with the development of such markers in specific species (e.g. those currently used for non-sacrificial analysis in salmonid aquaculture). The aim of the literature study is to evaluate the potential of sex-specific genetic markers for those species used in ecotoxicological testing. Depending on the results of this literature search, experimental research may be commissioned to develop sex-specific probes which could be utilised in testing procedures.

3.2.4 Conclusion

ECETOC supports the conduct of higher-tier tests utilising fathead minnows as the preferred test species for chronic studies of compounds active in tier 1, and which are predicted to enter aquatic ecosystems. However, the EDSTAC recommendation to default to the conduct of FFLC tests at tier 2 is not considered to be a realistic option and the ESTEEM approach is proposed to incorporate fish ELS and PLC tests at tier 2, with FFLC tests at tier 3 (Figure 2). Further data are required to validate the predictive ability of ELS and PLC exposures for FFLC tests. Additional endpoints will need to be incorporated into the protocols for these various tests but these endpoints should be validated and their usefulness assessed before regulatory implementation. Subsequent measurements may be targeted at different endpoints if a compound-specific mode of action is known (e.g. androgens, aromatase inhibitors, thyroid-active chemicals).

3.3 MYSID LIFE-CYCLE TEST

3.3.1 Description of the proposed assay

The EDSTAC document (Chapter 5, pages 5-64) refers to the standardised EPA protocol for the mysid life-cycle test. The mysid or opossum shrimp (*Americamysis bahia*, previously *Mysidopsis bahia*) test protocol involves a six-week chemical exposure period (flow-through conditions) and the endpoints currently evaluated include:

- length of time for the appearance of the first brood;
- sex determination (sex ratio);
- body length of males and females;
- cumulative number of young produced per female;
- effects on F₁ mysids (number of males and females, body length of males and females, and cumulative mortality).

3.3.2 Advantages/disadvantages of the method

Given the scope of EDSTAC, the scientific rationale for selecting an arthropod species to test for (anti-) androgenic or oestrogenic effects is unclear and its inclusion seems to be based simply on the existence of regulatory guidelines. Nevertheless, ECETOC supports the view that where aquatic invertebrates may be exposed to an ED compounds (as detected in the tier 1 screens) (Figures 1 and 2), then an alternative species to daphnids might be preferred in order to assess chronic effects of endocrine disruptors in crustacea since:

- daphnid populations maintained in laboratory conditions consists of parthenogenetic females that do not normally undergo a sexual reproductive cycle. Such a mechanism may be less representative for invertebrates as a whole and it can be hypothesised that sexual reproduction in invertebrates is more likely to be affected by ED chemicals is parthenogenesis.

The EDSTAC proposal of using mysids may be a valid option for a chronic invertebrate species in endocrine effects testing.

The advantages are:

- mysids have been widely used in regulatory ecotoxicology;
- they are a sexually dimorphic species;
- an extensive reference database exists on reproductive toxicity to this species.

The disadvantages are:

- relative to other invertebrate assays it is of long duration (6 weeks);
- requires large-scale, flow-through marine testing facilities (needing several thousands of litres of test solution per test);
- identification of the sexes is technically difficult;
- parental cannibalism upon offspring can compromise assessment of reproductive output.

Reproduction studies in other invertebrate species may prove to be more suitable for this test. Positive features of alternative invertebrates tests would include:

- the use of much easier culturing and testing procedures than those for the saltwater mysid;
- to aid landlocked laboratories, use of freshwater or synthetic sea water for culturing and chronic toxicity testing;
- static-renewal or modest flow-through testing facilities (requiring small volumes of test solutions).

However, currently there are few data on the influence of potential endocrine disruptors on development and reproduction in arthropods in general. Before an existing, probably extended, method can be suggested or a new test method be developed, further research is necessary (see below).

3.3.3 Research requirements

Mysids and daphnids, like many other arthropods, have a specific hormone regulation for growth and reproduction (e.g. the ecdysteroid system (Koolman, 1982)). This is markedly different from the androgen- and oestrogen-based regulation of reproduction in mammals and oviparous vertebrates. The statement of the EDSTAC document that oestrogen or androgen disrupters are likely to interfere with the ecdysteroid activity is considered to be speculative and does not reflect the other hormonal

mechanisms known to be important in arthropods. Additionally, juvenile hormones should be considered in insects and crustacea.

Although oestrogenic and androgenic steroids (e.g. 17β -oestradiol and testosterone, respectively) have been measured in *Daphnia magna* (Baldwin *et al*, 1995), their role in reproduction and sex determination is unclear. Baldwin *et al* (1995) also found that the model oestrogen diethylstilbestrol affected the reproduction and growth of *D. magna*. However, there are conflicting results of the effects of the natural and synthetic hormones oestradiol and ethinylestradiol on reproduction in *D. magna*. While Kopf *et al.* (1997) found a decrease in reproduction after exposure to oestradiol at a low $\mu\text{g/l}$ level, Schweinfurth *et al.* (1997) found no effect of ethinylestradiol exposure in *D. magna* up to several hundreds of $\mu\text{g/l}$.

In addition to daphnids, a number of other aquatic invertebrate species have been used to evaluate the development and reproductive effects of EDs. Recent work has shown inhibition of development and reproduction in marine copepods (*Tisbe battagliai*) by 20-hydroxyecdysone and DES (Hutchinson *et al.*, submitted). In contrast, life-cycle (21 day) studies with this species did not show significant effects following exposure at up to 100 $\mu\text{g/l}$ of either oestradiol, oestrone or ethinylestradiol (Hutchinson *et al.*, in press). Protocols for developmental and reproductive endpoints have also been produced for freshwater chironomids (ASTM, 1994) and have been applied to important classes of industrial chemicals (Brown *et al*, 1996). These and other methods have been summarised in ECETOC Document 31 (ECETOC, 1996).

As part of the ongoing research programme of the European chemical industry, a review addressing the use of aquatic invertebrates to detect endocrine activity potential has been carried out in collaboration with the UK Environment Agency (Pinder and Pottinger, 1998). They concluded that the role of sex steroid hormones in development and reproduction of arthropods and the potential sites of action of endocrine disrupters have yet to be established.

There is clear evidence in other invertebrate species of hormonal disruption. For example, in gastropod molluscs, the biocide tributyl tin is suggested to inhibit the aromatisation of testosterone in female snails, thus affecting the oestradiol and testosterone levels in the tissue of these organisms (Bettin *et al*, 1996).

Following the completion of the Pinder and Pottinger review, an international SETAC-OECD expert workshop on *Endocrine Disruption in Invertebrates: Endocrinology, Testing and Assessment* (EDIETA) was held in the Netherlands from 12 - 15 December 1998. With a significant input from US EPA and other North American scientists, the workshop discussed key aspects of invertebrate endocrinology and physiology, together with laboratory test methods and the use of aquatic and

terrestrial invertebrates for environmental monitoring and assessment. Given that the output from this important workshop is key to reaching a science-based consensus about invertebrate testing for ED effects, ECETOC proposes that EDSTAC defer the decision on which invertebrate species to use in higher tier testing (Figures 1 and 2) until the workshop report is available during mid-1999 (Stahl, personal communication).

3.3.4 Conclusion

In arthropods such as mysids, juvenoid and ecdysteroid hormones are more important than vertebrate-type steroids in influencing sexual differentiation, growth and reproduction. ECETOC concludes that the EDSTAC proposal to employ the mysid life-cycle test is scientifically unjustifiable as an invertebrate assay of long-term ED activity, given the serious gaps in the basic knowledge of endocrine function in this species. Further activity in this area should await the report from the international SETAC-OECD expert workshop on *Endocrine Disruption in Invertebrates: Endocrinology, Testing and Assessment* (EDIETA) which was held in the Netherlands, 12 - 15 December 1998.

3.4. AMPHIBIAN DEVELOPMENT AND REPRODUCTION

3.4.1 Description of the assay

While EDSTAC recommends a multigeneration assay incorporating embryonic larval development and adult reproduction (Chapter 5, pages 5-66), no specific assay has been described.

3.4.2 Alternative approach

The merits and objectives of a specific test with amphibian species on endocrine disrupting effects is discussed in the context of the frog metamorphosis assay (see section 2.1). Subject to adequate research, the inclusion of such a test might be a valid option at ESTEEM tier 2 (Figure 2).

Since amphibians are considered to be especially sensitive to thyroid disrupters, if a compound is active in the ESTEEM tier 2 frog metamorphosis assay, and if exposure is likely to be significant, then the conduct of a longer term amphibian development assay may be justified (ESTEEM tier 3) (Figure 2). To be environmentally relevant on such a case-by-case basis, this type of test should consider an ecologically representative amphibian species (for the protection of wild amphibian populations), together with an assessment of exposure parameters. Such an amphibian development assay could be derived by extending the chemical exposure period from early embryo development until metamorphosis is complete.

3.4.3 Conclusion

The EDSTAC document states that the development and validation of an appropriate test procedure is needed. In contrast, ECETOC believes that the test in amphibians should be focussed on the effects of thyroid disruption (ESTEEM tier 2), while reproduction endpoints are not the main objective of amphibian assays.

4. STANDARDISATION, VALIDATION, METHODS DEVELOPMENT AND RESEARCH

4.1 EDSTAC-PROPOSED APPROACH

In Chapter 5 of the EDSTAC report, 13 criteria are proposed to determine the degree of validation required for the screening and definitive assays. An important step in the validation process is the development of a standard protocol and the performance on model compounds in a number of laboratories. Priorities for method development in the field of wildlife tests were stated to be:

- an avian androgenicity screening assay;
- an amphibian development and reproduction test;
- a reptile reproduction test.

It is not clear whether EDSTAC considers that the other proposed test methods are sufficiently developed and / or validated.

4.2 ECETOC RECOMMENDATIONS

Validation can be defined as the input control technique used to detect any data which are inaccurate, incomplete or unreasonable. What the validation process needs to deliver is a suite of robust screening protocols which can distinguish between (anti-) oestrogens, (anti-) androgens and thyroid-active compounds. The critical objective of the tier 1 screening assays is to detect chemicals which may be active *in vivo*. Such *in vivo* screens should ideally avoid false positives, but it is essential that the potential for false negatives is minimised.

4.2.1 Selection of test compounds

The compounds chosen for validation of the *in vitro* screening tests should be made with respect to their known activity *in vivo*. If compounds are selected for which we have no *in vivo* data (ideally definitive data from long-term studies such as the mammalian multi-generation or similar) then there is a high risk that the context of any attempted validation will be extremely limited.

The selection of test compounds for *in vitro* and *in vivo* screening test validation should also consider the relevant mechanisms of action; e.g. whether metabolism is required, whether the substance is transported in serum via binding globulins or how the presence of such binding globulins might affect the performance of *in vitro* assays.

Following the adequate validation of the tier 1 screens, a number of selected substances in each class of activity may be selected for validation of appropriate long-term tests for hazard evaluation (ESTEEM tier 2 and 3 tests) (Figure 2).

4.2.2 Inter-species comparisons

The validation of *in vivo* assays must provide consistency across the different animal species being studied. Compounds which have been selected for the validation of mammalian protocols should also be considered for the validation of protocols with egg laying species (e.g. fish and birds). The above cited aquatic effects research programme of the European chemical industry is an illustration of the approach. The compounds selected have been principally chosen for their published database from mammalian systems and the intention is to generate comparative data for a range of animal species looking at a range of diverse endocrine disrupter endpoints.

The development of assays with new species, for example amphibians, is more problematic since the baseline data for the endocrinology and other reproductive biology is far less developed than for either mammalian species or for those wildlife species which have been well studied (e.g. fish species used in aquaculture). This basic problem should be recognised in drawing together a timetable of validation for the wildlife protocols and particularly for the protocols on thyroid-active developments.

4.2.3 Selection of reference chemicals

Chemical characterisation process should consider:

- reference substances for each hormone endpoint (negative and positive controls);
- natural and man-made substances;
- composition and definition of reference substance purity;
- the verification of chemical stability in the test systems;
- reference samples should be coded prior to their distribution to testing laboratories and there should be a centralised distribution of test materials to ensure all laboratories use the same batch number, etc.

4.2.4 OECD activities

The issue of validation of endocrine disrupter assays also formed a major discussion point at the recent OECD-Endocrine Disrupters Testing and Assessment (EDTA) working group meetings (Paris and Washington, 1998). The collation of a database on reference chemicals was considered to be of

the highest priority at the OECD working group meeting and this should be recognised for the progress of any validation of the screens or tests proposed under EDSTAC.

4.2.5 Conclusions

In summary, the method development and validation process should be a joint activity between governmental regulators, international harmonisation organisations such as OECD and industry research groups, such as ECETOC and CEFIC-EMSG. Such positive collaboration will help ensure that best use is made of existing data, initiated research programmes and collaborative activities in order to avoid duplication and conflicting approaches.

BIBLIOGRAPHY

Ankley, G., Mihaich, E., Stahl, R., Tillitt, D., Colborn, T., McMaster, S., Miller, R., Bantle, J., Campbell, P., Denslow, N., Dickerson, R., Folmar, L., Fry, M., Giesy, J., Gray, L.E., Guiney, P., Hutchinson, T.H., Kennedy, S., Kramer, V., LeBlanc, G., Mayes, M., Nimrod, A., Patino, R., Peterson, R., Purdy, R., Ringer, R., Thomas, P., Touart, L., Van der Kraak, G. and Zacharewski, T. (1998). Overview of a workshop on screening methods for detecting potential (anti-) estrogenic/androgenic chemicals in wildlife. *Environ. Toxicol. Chem.* 17, 68-87.

ASTM (1994). Standard guide for conducting sediment toxicity tests with freshwater invertebrates. E1383-94. American Society for Testing and Materials, Philadelphia, USA.

Baldwin, W.S., Milam, D.L. and LeBlanc, G.A. (1995). Physiological and biochemical perturbation in *Daphnia magna* following exposure to the model environmental oestrogen diethylstilbestrol. *Environ. Toxicol. Chem.* 14, 945-952.

Bettin, C., Öhlmann, J. and Stroben, E. (1996). TBT-induced imposex in marine gastropods is mediated by an increasing androgen level. *Helgoländer Meeresunters* 50, 299-317.

Brown, D.D., Wang, Z., Kanamori, A., Eliceiri, B., Furlow, J.D., Schwartzman, R. (1995). Amphibian metamorphosis – a complex program of gene-expression changes controlled by the thyroid-hormone. *Recent Progress in Hormone Research*, 50, 309-315.

Brown, D., Thompson, R.S., Stewart, K.M., Croudace, C.P. and Gillings, E. (1996). The effect of phthalate ester plasticisers on the emergence of the midge *Chironomus riparius* from treated sediments. *Chemosphere* 32, 2177 - 2187.

Campbell, P.M. and Hutchinson, T.H. (1998). Wildlife and endocrine disrupters: Requirements for hazard identification. *Environ. Toxicol. Chem.* 17, 127-135.

ECETOC (1996). Environmental Oestrogens - A Compendium of Test Methods. ECETOC Document No 33. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC (1998). Endocrine Disrupter Newsletter number 1, July 1998, 4 pp. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

EDSTAC (1998). Endocrine Disruptor Screening and Testing Advisory Committee Final Report, August 1998.

EPA (1998). Endocrine Disruptor Screening Program: Proposed Statement of Policy. Federal Register, 28th December 1998, vol 63, no 248, pp 71542 - 71568.

Fort, D.J. and Stover, E.I. (1997). Development of short-term, whole embryo assays to evaluate detrimental effects on amphibian limb development and metamorphosis. In *Environmental Toxicology and Risk Assessment: Modelling and Risk Assessment* 6th Edn (Dwyer, F.J., Doane, T.R. and Hinman, M.I., Eds.). ASTM STP 1317, pp 376 - 390. American Society for the Testing of Materials, Philadelphia.

Gorbmann, A., Dickhoff, W.W., Vigna, S.R., Clark, N.B. and Ralph, C.L. (1983). *Comparative Endocrinology*. John Wiley and Sons Inc., New York.

Hayes, T.B. and Wu, T.H. (1995). Role of corticosterone in anuran metamorphosis and potential role in stress-included metamorphosis. *Netherlands J. Zool.* 45 (1-2), 107-109.

Hutchinson, T.H., Pounds, N.A., Hampel, M. and Williams, T.D. (in press). Impact of ecdysteroids and oestrogens on developmental and reproductive parameters in the marine copepod *Tisbe battagliai*. *Sci. Tot. Environ.*

Hutchinson, T.H., Pounds, N.A., Hampel, M. and Williams, T.D. (submitted). Life-cycle effects of 20-hydroxyecdysone and diethylstilbestrol on the marine copepod *Tisbe battagliai*. *Environ. Toxicol. Chem.*

Kanamori, A. and Brown, D.D. (1996). The analysis of complex developmental programmes: Amphibian metamorphosis. *Genes to Cells*, 1 (5), 429-435.

Kime, D.E. (1998). *Endocrine disruption in fish*. Kluwer Academic Publishers, London. 396 pp..

Koolman, J. (1982). Ecdysone metabolism. *Insect Biochem.* 12, 225-250.

Kopf, W. (1997). Wirkung endokriner Stoffe in Biotests mit Wasserorganismen. In: Bayerisches Landesamt f. Wasserwirtschaft - Institut für Wasserforschung -, München. Hrsg. Stoffe mit endokriner Wirkung im Wasser. Münchener Beiträge zur Abwasser-, Fischerei- u. Flußbiologie. Band 50. München: R. Oldenbourg, 1997: 39-54.

Laale, H.W. (1977). The biology and use of zebrafish *Brachydanio rerio* in fisheries research. *J. Fish Biol.* 10, 121-173.

Länge, R., Campbell, P., Einarson, S., Holt, M., Hutchinson, T., McCahon, P., Scholz, N., Panter, G., Tattersfield, L. and van Egmond, R. (1998). Endocrine disrupters in wildlife organisms: an approach of the European chemical industry to hazard identification and testing methods, Proceedings of the 3rd IBC Conference on Endocrine Disrupters in the Environment, London, 14-15 May 1998.

Länge, R., Schweinfurth, H., Panter, G. and Croudace, C. (1997). Growth and reproduction of fathead minnow (*Pimephales promelas*) exposed to the synthetic steroid hormone ethinylestradiol in a life-cycle test. Proceedings of the 7th Annual Meeting of SETAC-Europe, Amsterdam 6-10 April 1997.

Leatherland, J.F. (1992). Endocrine and reproductive function in great lakes salmon. In Chemical-Induced Alterations in Sexual and Functional Development: The Human/Wildlife Connection (Eds, Colborn, T. and Clement, C.) 129-145. Advances in Modern Environmental Toxicology No 21.

Lundholm, C.E. (1997). DDE-induced eggshell thinning in birds: effects of p,p'-DDE on the calcium and prostaglandin metabolism of the eggshell gland. *Comp. Biol. Physiol.*, 118C, 113-128.

McNabb, F.M.C. (1992). Thyroid hormones, Prentice Hall, New Jersey.

Newsome, C.S. and Piron, R.D. (1982). Aetiology of skeletal deformities in the zebra danio fish (*Brachydanio rerio*). *J. Fish Biol.* 21, 231 - 237.

OECD (1993). Guidelines for Testing of Chemicals, Paris.

OECD (1997). Proposed draft OECD Guideline for testing of chemicals. Avian reproduction toxicity test in the Japanese quail (*Coturnix coturnix japonica*) and northern bobwhite (*Colinus virginianus*) OECD, Paris, Dec. 1997.

OECD (1999). Fish Expert Consultation Meeting, London, 28 - 29 October 1998. Minutes available from the Environmental Health and Safety Division, OECD, Paris

Rundle, J. (1996). Danio throwbacks. *Practical Fish Keeping*, June 1996 pp 44 - 46.

Panter, G.H., Thompson, R.S. and Sumpter, J.P. (1998). Adverse reproductive effects in male fathead minnows (*Pimephales promelas*) exposed to environmentally relevant concentrations of the natural oestrogens, oestradiol and oestrone. *Aquat. Toxicol.* 42, 243-253.

Pinder, L.C.V. and Pottinger, T.G. (1998). Endocrine function in aquatic invertebrates and evidence for disruption by environmental pollutants. Draft internal report, Centre for Ecology and Hydrology, Institute of Freshwater Ecology, UK.

Piron, R.D. (1978). Spontaneous skeletal deformities in zebra danio (*Brachydanio rerio*) bred for fish toxicity tests. *J. Fish Biol.* 13, 79 - 83.

Purdom, C.E., Hardiman, P.A., Bye, V.F., Eno, N.C., Tyler, C.R. and Sumpter, J.P. (1994). Estrogenic effects of effluents from sewage treatment works. *Chemistry and Ecology* 8, 275-285.

Pyron, M. and Beitinger, L.T. (1989). Behaviour of male fathead minnows in the presence of an additional male or female fathead minnow. *The Texas Journal of Science* 41, 151-154.

Schweinfurth, H., Länge, R., Miklantz, H. and Schauer, G. (1997). Umweltverhalten und aquatische Toxizität von Ethinylestradiol. In: Bayerisches Landesamt f. Wasserwirtschaft - Institut für Wasserforschung -, München. Hrsg. Stoffe mit endokriner Wirkung im Wasser. Münchener Beiträge zur Abwasser-, Fischerei- u. Flußbiologie. Band 50. München: R. Oldenbourg, 1997, 39-54.

Takahashi, H. (1977). Juvenile hermaphroditism in the zebrafish, *Brachydanio rerio*. *Bulletin of the Faculty of Fisheries Hokkaido University*, 28, 57-65.

Tattersfield, L., Matthiessen, P., Campbell, P.M., Grandy, N. and Länge, R. (eds), 1997. SETAC Europe/OECD/EC Expert Workshop on Endocrine Modulators and Wildlife: Assessment and Testing (EMWAT), Chapter 4, Veldhoven, the Netherlands 10-13 April 1997, SETAC Europe, Brussels 1997.

Tillitt, D.E., Solomon, K.R., Mihaich, E.M., Cobb, G., Touart, L. and Kubiak, T.J. (1998). Role of exposure assessment in characterizing risks of endocrine-disrupting substances to wildlife. In *Principles and Processes for Evaluating Endocrine Disruption in Wildlife* (Kendall, R., Dickerson, R., Giesy, J. and Suk, W., eds.), pp 39 - 68. SETAC Press, Pensacola, FL..

Tyler, C.R., van Aerle, R., Hutchinson, T.H., Maddix, S. and Tripp, H. (in press). An *in vivo* testing system for endocrine disrupters in fish early life-stages using induction of vitellogenin. *Environ. Toxicol. Chem.*

USEPA (1986). Fish life-cycle toxicity tests. Hazard Evaluation Division Standard Evaluation Procedure EPA 540/9-86-137. Office of Pesticide Programs, Washington.

AUTHORS OF THE REVIEW DOCUMENT

K. Brugger	DuPont USA – Wilmington
P. M. Campbell	Procter and Gamble CDN – Toronto
S. Einarson	Norsk Hydro N – Porsgrunn
T. H. Hutchinson	Zeneca UK – Brixham
R. Länge	Schering D – Berlin
P. McCahon	Rhône-Poulenc USA – Research Triangle Park
L. Tattersfield	Shell UK – London
R. Van Egmond	Unilever UK – Bebington
M. Holt	ECETOC B – Brussels

MEMBERS OF THE SCIENTIFIC COMMITTEE

(Peer Review Committee)

W. Tordoir (Chairman), Group Adviser, Environmental Health and Human Toxicology	Shell International NL - Den Haag
O. Bøckman, Scientific Adviser	Norsk Hydro N - Porsgrunn
C. Braun Occupational Toxicologist	Akzo Nobel NL - Arnhem
N. Carmichael, Toxicology Director Worldwide	Rhône-Poulenc F - Lyon
C. d'Hondt ^a , Head of Environmental Safety Department	Novartis CH - Basel
T. Feijtel ^a Section Manager	Procter and Gamble B - Brussels
B. Hildebrand, Director, Experimental Toxicology	BASF D - Ludwigshafen
J. Jackson, Senior Associate Medical Adviser	Monsanto B - Brussels
E. Löser, Head, Institute of Industrial Toxicology	Bayer D - Wuppertal
R. Millischer, Head, Industrial Toxicology Department	Elf Atochem F - Paris
G. Randall ^a , Director, Environmental Laboratory	Zeneca UK - Brixham
A. Sarrif, Associate Director Toxicology Affairs	DuPont D - Bad Homburg
J. Solbé ^a , Head of SEAC Environment	Unilever UK - Bebington
H-J. Wiegand, Head, Product Safety Department	Hüls AG D - Marl

^a Stewards responsible for primary peer review

Responsible Editor: FM Carpanini, ECETOC
Av E Van Nieuwenhuyse 4 (Bte 6)
B - 1160 Brussels, Belgium

