Guidance on Assessment and Application of Adverse Outcome Pathways (AOPs) Relevant to the Endocrine System

Technical Report No. 128

EUROPEAN CENTRE FOR ECOTOXICOLOGY AND TOXICOLOGY OF CHEMICALS

Guidance on Assessment and Application of Adverse Outcome Pathways (AOPs) Relevant to the Endocrine System

Technical Report No. 128

Brussels, December 2016

ISSN-2079-1526-128 (online)

[Updated 20 December 2016]

ECETOC Technical Report No. 128

European Centre for Ecotoxicology and Toxicology of Chemicals
2 Avenue E. Van Nieuwenhuyse (Bte 8), B-1160 Brussels, Belgium.

Guidance on Assessment and Application of Adverse Outcome Pathways (AOPs) Relevant to the Endocrine System

Contents

SUMMARY

Various European chemical regulations, e.g. Regulation (EC) No. 1907/2006 ('REACH') and Regulation (EC) No. 1107/2009 (plant protection products) only support the marketing and use of chemical products on the basis that they do not induce endocrine disruption in humans and/or non-target species. Therefore, there is keen interest in the development of new tools to help identify and regulate chemicals that may affect endocrine systems.

Adverse outcome pathways (AOPs) have the potential to be important tools for the identification and regulation of endocrine disrupters as they can aid in the determination of the mechanistic basis for adverse outcomes observed in *in vivo* (eco)toxicological studies; therefore they can help identify if an observed adverse outcome can be plausibly linked to an endocrine mechanism, which is a key requirement for identification of a chemical as an endocrine disrupter according to the widely accepted WHO/IPCS (2002) definition: "*An endocrine disrupter is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations.*" They could also potentially be used to help predict the potential for an adverse outcome *in vivo* based on the results of *in vitro* mechanistic data.

If AOPs are to be used to identify and/or predict endocrine disrupting properties, it must be ensured that they are sufficiently robust and fit for purpose. To this end, this Technical Report provides guidance on identifying the basic requirements of a defined AOP, and how to establish the minimum scientific standards that allow the use of AOPs in different contexts, such as hazard identification, read-across and risk assessment. These requirements are described as 'Key Elements in Assessment of AOP Utility' and each is discussed separately in detail in Sections 2.2 – 2.10. In addition, a case study is presented to provide examples of how these Key Elements may be considered collectively when evaluating an AOP.

1. INTRODUCTION

1.1 Background

The **adverse outcome pathway (AOP)** is a concept that provides a framework for organising knowledge about the progression of events across scales of biological organisation that lead to adverse effects. AOPs link existing knowledge in a linear fashion along a series of one or more causally connected **key events (KE)** between two anchors — a **molecular initiating event (MIE)** and **an adverse outcome (AO)**. The AO occurs at a level of biological organisation relevant to risk assessment. The linkage amongst the events is described by **key event relationships (KERs)** that describe the causal relationships.

There is considerable global interest in Adverse Outcome Pathways (AOPs), with a number of high-profile international activities. One such activity is the formation and curation of an AOP-Wiki (https://aopwiki.org/wiki/index.php/Main_Page), which is intended to be an interactive, collaborative space where all relevant stakeholders (including industry, academia and regulatory agencies) can contribute to the development and refinement of AOPs.

There are two main research areas for AOP development: (1) Exploratory Research: AOPs developed largely for academic purposes of establishing biological understanding and identification of data gaps; and (2) Defined Pathways: AOPs developed for utilisation in screening, prioritisation, and possibly decision-making within a regulatory context. This Technical Report provides guidance on delineating the basic requirements of a defined AOP, described in this document as 'Key Elements in Assessment of AOP Utility', and how to establish the minimum scientific standards that allow the use of AOPs to either predict an AO or to explain a known AOP.

The guidance is focussed on AOPs relating to the endocrine system, because amongst the AOPs currently being developed, those relevant to the area of endocrine disruption are particularly prevalent. This prevalence is likely attributable to the fact that the AOP concept, which links mechanism and adverse effects fits well with the most widely accepted definition of an endocrine disrupter: "*An endocrine disrupter is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations.*" (WHO/IPCS, 2002). This definition requires three things: (1) Relevant 'endocrine mechanism(s)'; (2) Relevant adverse effect(s); and (3) a plausible link between mechanism(s) and adverse effect(s).

1.2 Terms of Reference

The Task Force used the following terms of reference in compiling this Technical Report:

- Evaluate the current state of the science with respect to AOPs relevant to the endocrine system, how they are currently being developed, assessed and used and how they may be used in the future.
- Identify the key elements that must be considered when evaluating the utility of an AOP for different, defined purposes.
- Provide guidance on how to establish and achieve the minimum scientific standards to ensure that the requirements of each of the defined key elements are met.

2. KEY ELEMENTS IN ASSESSMENT OF AOP UTILITY

2.1 Determination of Key Elements

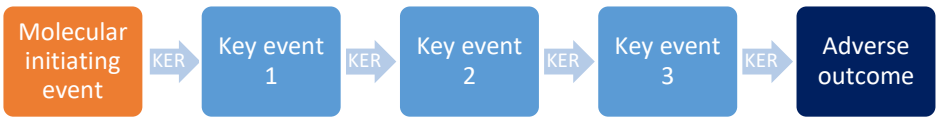
The key elements in assessment of AOP utility described in this section were determined following discussions within the Task Force and by reference to available individual AOPs, particularly those on the AOP wiki, and noting in which areas the AOPs were robust and in which areas the Task Force felt more detail was needed. Following this review, the Task Force collectively identified the nine key elements described in Sections 2.2 to 2.10 as those most important for consideration when developing and/or utilising an AOP. It should be noted that these key elements are those considered by the Task Force to be the most generally applicable across all AOPs and that for any given AOP there will be other important elements to be considered. It should also be noted that each of these key elements is not intended to carry equal weight; with the relative importance varying on a case-by-case basis.

2.2 Biological Plausibility

2.2.1 Overview

Biological plausibility is a critical determinant in both the development and the application of an adverse outcome pathway, because for an AOP to be used in any context it needs to be biologically plausible. In other words, the proposed causal association between the molecular initiating event (MIE) and the AO, linked by key events (illustrated conceptually in Figure 1), needs to be consistent with existing scientific knowledge. As AOPs are applied, the plausibility that the AOP is relevant to the chemical under evaluation needs to be demonstrated.

Figure 1: General AOP concept



KER = Key event relationship.

2.2.2 AOP development: Is the AOP in general biologically plausible?

In an AOP, the key events (KE) are measurements of a biological state, such as increased or decreased production of a hormone from a specific cell type. Key events can therefore be verified experimentally depending on the availability of appropriate test systems and the accuracy and precision of the methods used. Key event relationships (KER) refer to the causal associations which link a KE with events both upstream and downstream. For example, the KE of 'decreased testosterone production from Leydig cells' may have an

upstream KE of 'reduced serum luteinising hormone levels' and a downstream KE of 'decreased serum testosterone'. The relationships between these KEs are therefore inferences based on the most up-to-date understanding of the biological system being described. Therefore, although it may be possible to measure a KE, the KERs cannot be directly measured in a test system. As an AOP is being developed, it is important to ensure that the AOP in general and the key event relationships specifically, are biologically plausible. Once an AOP is proposed, further development and testing of the AOP need to include scrutiny of its overall biological plausibility. This involves considering the other focus areas discussed in this document and documenting alternative KERs and explanations of the observed effects.

2.2.3 AOP application: Is the AOP biologically plausible for the chemical under evaluation?

For an AOP to be used in any context, biological plausibility remains an important consideration. For example, in the context of priority setting, biological plausibility means that further data generation must address gaps in the proposed AOP and KER where alternative explanations to the proposed AOP have been identified, but not excluded. This is particularly important when a KE is not very specific in nature (reduced growth or weight of endocrine organs) and can have multiple causes. The same conditions apply when AOPs are used in a predictive way for risk assessment (i.e. using mechanistic data to predict an apical endpoint in the absence of empirical data) or to explain observed adverse effects in either experimental animals or exposed populations (i.e. where apical endpoint data is supplemented by mechanistic information to describe aetiology of the effect).

One critical consideration in determining whether there is a biologically plausible cause-and-effect relationship between a chemical exposure and endocrine-related adverse effects is potency. This indicates the importance of a good understanding of the dose-response for any effects seen and consideration of relevant exposure levels. Furthermore, as the building blocks for the AOP will come from different experiments, there must be consistency in the changes seen in these experiments and the doses that elicit the response. Since potency is such an important consideration in determining the biological plausibility of a cause-and-effect relationship between chemical exposure and endocrine-related adverse effects, some understanding of dose-response is needed to demonstrate that an adverse effect seen *in vivo* is related to an endocrine mode of action. For the fully quantitative AOPs used in a risk assessment context, detailed dose-response data for at least one model chemical are needed to demonstrate the level of perturbation required at different steps along the pathway to trigger the adverse outcome.

The same conditions apply when AOPs are used in a predictive way for risk assessment i.e. the endpoint traditionally used in the RA is not available but can be deduced from an earlier KE that has been measured.

When an AOP is applied to a group of (structurally related) chemicals, both the substances fitting the proposed AOP and those that do not fit should be considered, and both differences and similarities between substances discussed.

2.3 Essentiality

2.3.1 Overview

The molecular initiating event (MIE) is the initial interaction between a chemical and a biomolecule/biosystem of the AOP while key events (KEs) are events that can be observed across different levels of biological organisation (e.g. cellular, organ) which could result in an adverse outcome (Ankley et al, 2010; Allen et al, 2014). The evaluation of the essentiality of an AOP should lead to evidence that MIE and KEs contribute to the pathway. It is expected that some specific studies would investigate whether the blocking of MIE/KEs events would lead to blocking of the adverse effects. Indirect evidence could also be sufficient to prove essentiality; modification of expected modulating factors would impact KE and the related expected KE response or the Adverse Outcome (OECD, 2015a).

2.3.2 Examples

For existing AOPs focusing on endocrine active chemicals, direct evidence showing that blocking of the MIE lead to blocking the AO is not available. For most existing AOPs focusing on endocrine active chemicals, the essentiality of the MIE is often weak, but the strength of evidence increases for subsequent KEs and when approaching the AO (see Table 1).

Indirect evidence can nevertheless be sufficient to prove essentiality, for example, AOP 42 (from AOP wiki): Xenobiotic induced inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes in mammals. Specifically, the essentiality of the MIE for inducing subsequent KE is proven; cessation of xenobiotic exposure results in a return to normal levels of synthesis and circulatory hormones (e.g. Cooper et al, 1983). Other studies demonstrated essentiality for subsequent KEs (decreased serum thyroid hormones, decreased brain thyroid hormones, decreased cognitive function) by exposure/recovery experiments. This leads to a weight of evidence (WOE) that can be considered sufficient to fully demonstrate essentiality.

Factors such as the specificity of the association of adverse effect and initiating event are also closely linked to essentiality. As example, the existence of multiple MOAs may contribute to/synergise with an AOP and question the essentiality of the AOP. For AOP 18 (from AOP wiki): PPARα activation leading to impaired fertility, both PPARα-dependent and -independent mechanisms are plausible mechanism for fertility impairment. Ward et al (1998) observed that PPAR alpha-null mice exposed to DEHP developed toxic lesions in the liver, the kidney and testis but wild type mice developed also toxic lesions (less severe) in kidney and testis. This study clearly questioned the essentiality of PPARα activation for developing effects on fertility. Further studies are needed to elucidate the role of other MOAs (e.g. other PPAR isoforms).

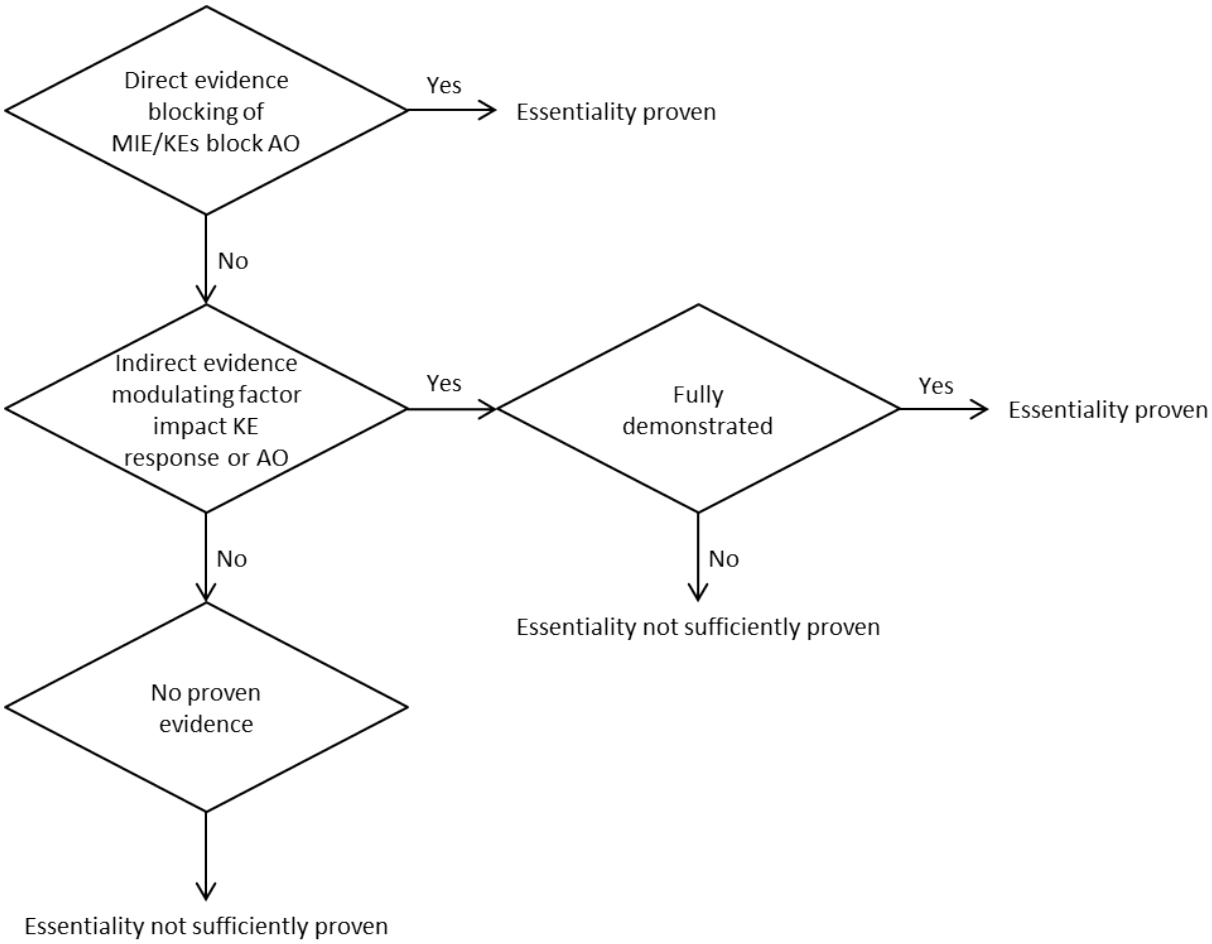
Table 1: Evaluation of essentiality for several example AOP taken from the AOP wiki

AOP	Essentiality as reported in the AOP wiki	Direct evidence	Indirect evidence	No proven evidence	Comments
AOP 7: PPARγ activation leading to impaired fertility in adult female rodents	Weak to strong	-	*Weak evidence for MIE* PPARγ activation was found to indirectly alter the expression of aromatase. *Demonstrated for KEs* The observed decrease insteroidogenesis and subsequent decrease in testosterone levels are well established as precursors to anatomical changes in the developing male reproductive tract.	-	Essentiality not proven.
AOP 18: PPARα activation leading to impaired fertility upon *utero* exposure in rodent males	Weak to moderate	-	*Weak evidence for MIE* Decreased testosterone levels in PPARα(−/−) null control mice. Increases in gene expression of PPARα followed by decreases the expression of genes which are associated with steroidogenesis. *Demonstrated for KEs* The observed decrease insteroidogenesis and subsequent decrease in testosterone levels is well established as precursors to anatomical changes in the developing male reproductive tract.	-	Mainly based on phthalates, potentially multiple modes of action. Essentiality not proven.
AOP 42: Xenobiotic induced inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes in mammals	Moderate to strong	-	*Demonstrated for MIE* (inhibition of thyroperoxidase) Cessation of exposure results in return to normal levels of synthesis and circulatory hormones. *Demonstrated for KEs* (decreased serum TH, decreased brain TH, decreased cognitive function) by exposure / recovery experiments *Partly demonstrated for KEs* (altered regulation of TH responsive genes, altered neuroanatomy, altered neurophysiology).	-	Essentiality proven.

2.3.3 Application comments

Components of evidence for essentiality evaluation are summarised in Figure 2. An AOP with weak evidence to support the KEs and especially MIE as essential should not be relied upon for a hazard/risk assessment, an IATA or read-across. This can lead to misuse or misguided reliance on this AOP. Furthermore, when essentiality is not sufficiently proven, improvement of the AOP is required before any regulatory use. Indirect evidence that the MIE modulates the KE response and that a KE produces the AO is needed.

Figure 2: Components of evidence for essentiality evaluation



2.4 Empirical Evidence

2.4.1 Overview

Empirical evidence refers to the quality and strength of the database that underpins the AOP. In judging the empirical evidence, it is important to assess the quality and quantity of the supporting data and the concordance of observations in MIE, KE, and AO, including thresholds and dose-response.

This is a key consideration for all applications, but the relative strength may vary depending on application. For example, good quality *in vitro* data describing a plausible MIE could be sufficient to decide whether a chemical is a priority for a further more detailed evaluation. However, for risk assessment purposes greater scrutiny should be placed on the evidence across the whole AOP, including quantitative (dose-response) and temporal considerations.

2.4.2 Data quality

The quality of data should be critically evaluated before it is accepted to form part of the weight of evidence for the development or application of an AOP. One accepted method of judging (eco)toxicological data quality is Klimisch scoring (Klimisch et al, 1997). The concept has been expanded for ecotoxicological data with the addition of detailed evaluation sheets (Moermond et al, 2016) and in quantitative weight of evidence, where no studies are excluded, but the weight of a study in the weight of evidence evaluation changes (Van der Kraak et al, 2014).

Although some approaches place a lot of emphasis on compliance with Good Laboratory Practice (GLP) and adherence to OECD test guidelines, mechanistic data used to develop and apply AOPs may not be generated to GLP or to standard test guidelines. Therefore, the fundamentals of reliability should be observed, not just test guideline or GLP compliance. Common elements in data quality assessment are use of appropriate controls, meeting guideline validity criteria, adequate description of test systems, exposure confirmation and conditions and statistical evaluation. If one of these factors cannot be confirmed then the study will carry less weight, and depending on the nature of the deficiency may be disregarded altogether. Note that reliability and relevance of a study are not the same: while a study can be fully reliable (i.e. well conducted and reported) it is not necessarily relevant, because it may not address the precise question of relationship between MIE, KE and AO (Ruden et al, in press).

To develop a robust AOP not only should studies be used which support the proposed AOP, but also those that apparently contradict (or do not support) it and the differences explained or uncertainties highlighted. Some key areas where consistency needs to be demonstrated are described below:

2.4.3 Temporal concordance

There should be a clear relationship between time of MIE, subsequent KEs and AO. Some MIEs can result in a KE or an AO within a short timeframe for example reproductive and mechanistic endpoints in fish short term reproduction assays (e.g. Ankley and Villeneuve 2015). Whilst for other effects there may be a significant latency between the MIE and the AO. For example, developmental processes where adverse outcomes become evident only after development is completed or thyroid effects, where longer-term "depletion" of thyroid hormone or its precursors is required to produce an effect, while short term disturbance may have no adverse impact, depending on the timing of exposure. In either case, the temporal concordance between the exposure and the effect should be predictable and consistent.

2.4.4 Thresholds and dose / concentration response

For each AOP It should be determined if the relationship between the events is still plausible on a quantitative scale, i.e. that effect thresholds and dose-response of upstream events are comparable to those of downstream events. Compensatory mechanisms should be considered in this interpretation, as should confounding factors such as age, size and physiological state of organisms, but the overall picture should be consistent. This represents a particular challenge for judging the quality of an *in vitro* only dataset and selecting the most reliable data, since many *in vitro* test methods for endocrine activity do not provide consistent dose response data and can show significant inter-laboratory variation in concentrations required to elicit a response.

For example, in the OECD validation study for the H295R steroidogenesis assay, there was good consistency amongst laboratories in terms of fold changes in testosterone synthesis caused by different androgen agonists and antagonists. However, the lowest observed effect concentrations (LOECs) varied by several orders of magnitude amongst laboratories (Hecker et al, 2011). This level of variability observed in a well-defined test system demonstrates the challenges that exist in demonstrating consistency in dose-response relationships both within an assay and amongst assays.

Another example of challenges associated with determining the weight of the empirical evidence is from the OECD validation (OECD Series on Testing & Assessment No. 226; OECD 2015b) of the *in vitro* oestrogen receptor binding assay for the weak binder 4-n-heptylphenol (heptylphenol). While most of the 23 substances tested were identified by all labs as either binder, non-binder or equivocal, in the case of heptylphenol, four laboratories classified the test item as a 'binder', one as 'equivocal' and one as a 'non-binder'. This was due to very high concentrations tested (up to 10^{-3} mol/L, i.e. 192 mg/L) and also associated solubility issues. Therefore, depending on which laboratory conducts the test up to which concentration, the sequence from MIE to AO could be quantitatively convincing, or not.

The dose metrics for the various tests in the OECD toolbox vary from µM in test solutions for *in vitro* tests, mg/kg diet or body weight, or µg/L in aqueous media. To properly link MIE, KEs, KERs and AOs, these units should be made comparable e.g. by consideration of ADME and bioconcentration. Internal concentrations in the target tissue (e.g. after aquatic exposure measured in fish or estimated from known bioconcentration values,) should at least roughly correspond to the concentrations that elicit the corresponding response in *in*

vivo studies. There is no agreed quantitative measure, when reasonable correspondence can be established, but if the exposure concentrations needed to produce effects in several tests differ by orders of magnitude, the differences should be explained or considered not coherent in the WoE.

2.4.5 Benchmarking to known 'positives'

Proper benchmarking should be a prerequisite for developing any (eco)toxicological assay as a part of an AOP. Difficulties in finding an appropriate positive control can also result from undetected variables or confounding factors in the assay and can be an indication that improvements are still needed. This should be considered in the weight that results from such a test have and the applicability of the AOP in which they are used.

2.4.6 Application comments

An understanding of the empirical evidence, including quantitative concordance is of critical importance when using AOPs for risk assessment, hazard identification and read-across. It is also important, but to a lesser extent, for priority setting.

A misjudgement of the empirical evidence could lead to misidentification of a chemical as an endocrine disrupter.

2.5 Analytical Validation

2.5.1 Overview

In toxicity testing, it has long been recognised that there is a need to have confidence in test methods used to evaluate biological activity. Within the AOP construct, the same is true – there is a need to have confidence in the assays/test methods used to measure responses indicative of Key Events (KEs). The Scientific Confidence Framework (SCF) for high throughput assays and prediction models based on the proposal of Cox et al (2014), and extended by Patlewicz et al (2015) to AOPs, is composed of three elements, (1) Analytical Validation, (2) Qualification and (3) Utilisation. In the most general sense, Analytical Validation within the context of an AOP is simply documentation of the performance of each assay or test method used for measuring a response that provides knowledge of a substance's activity for a specific KE within the AOP.

Documentation of analytical validation for KEs in an AOP should follow the principles and procedures described in OECD's 2005 Guidance Document 34 (Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14) and ICCVAM's 1997 report on Validation and Regulatory Acceptance of Toxicological Test Methods (https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf).

For AOPs, emphasis is placed on documentation of test method relevance and reliability for the intended purpose – i.e. provide data on the biological activity of a substance with respect to a given KE. The confidence

in a given test method is dependent upon the extent to which the documentation cited covers the elements listed below (modified from ICCVAM, 1994):

- The scientific rationale for the test method, including a clear statement of its proposed use, including the relationship of the test method's endpoint(s) to the biological effect of interest.
- The availability of a detailed protocol for the test method, to include, but not necessarily limited to, test performance criteria (e.g. positive and negative control responses), a description of how data will be analysed, a description of the known limitations of the test, the degree to which biological variability affects test reproducibility, etc.
- A description of the classes of materials that the test can and cannot accurately assess.
- The description of within-test variability and the reproducibility of the test within and among laboratories.
- Description of performance, such as sensitivity (the proportion of active substances that are correctly identified by the new test), specificity (the proportion of inactive substances that are correctly identified) as well as the reference chemicals (both known positive and known negative agents) representative of the types of substances to which the test method has been applied.
- The availability of data supporting the assessment of the validity of the test method.
- Results of independent scientific review.

For endocrine AOPs, in particular those dealing with oestrogen, androgen and thyroid pathways, a number of test methods and assays with endpoints relevant to specific KEs have been validated by the OECD and US EPA and incorporated into regulatory testing guidelines. See for example OCSPP Harmonised Test Guidelines, Series 890 - Endocrine Disruptor Screening Program Test Guidelines (US EPA, 2009)[1] and OECD's Test Guidelines Specifically Developed or Updated for the Screening or Testing of Chemicals for Endocrine Disruption[2]. When referring to these methods for use in measuring a KE in an AOP, analytical validation would consist of citing the validation documents published by US EPA or OECD.

To perform and document analytical validation for KEs in an AOP, the following steps are recommended:

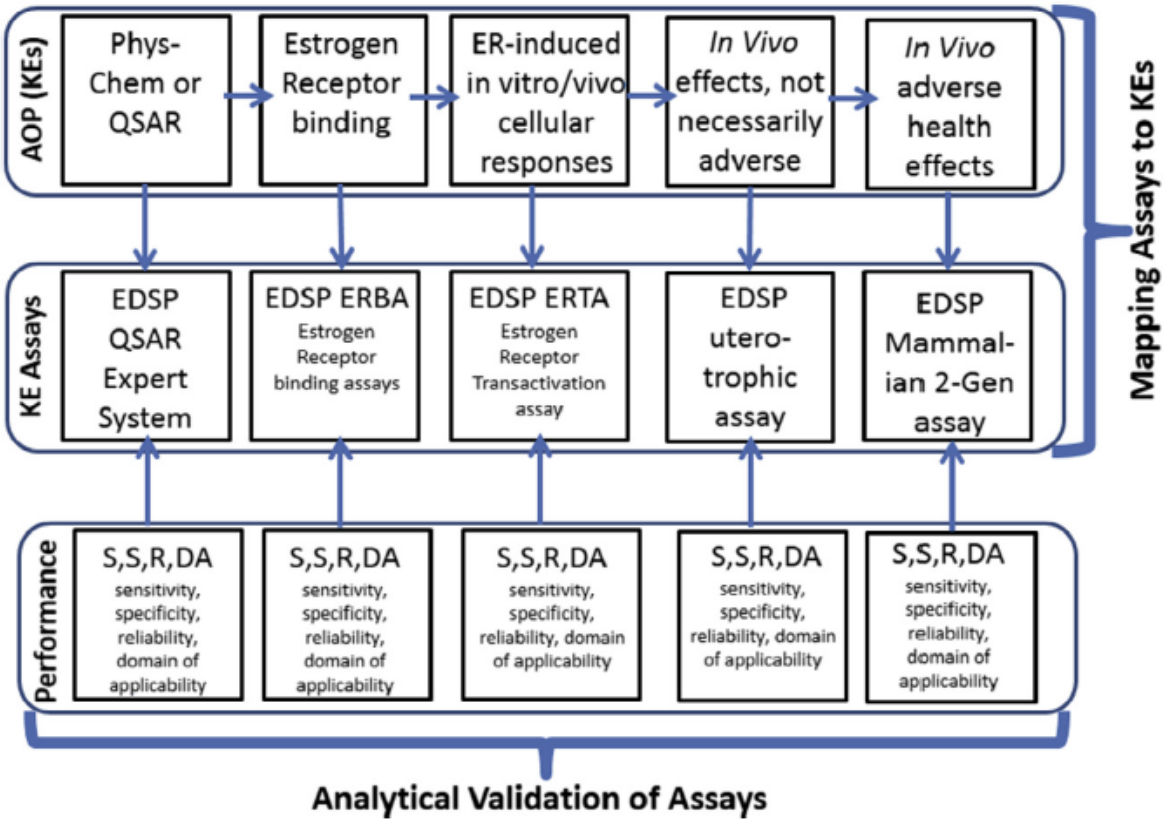
- **Step 1.** Develop the AOP.
- **Step 2.** Map existing, specific assays to each of the KEs within the AOP.
- **Step 3.** Document analytical validation of each assay. For example, specify the biological basis and document of assay performance characteristics (reliability, sensitivity, specificity and domain of applicability).

An example of Analytical Validation of an oestrogen pathway AOP is presented in Patlewicz et al, 2015. The example in Figure 3 is similar, except this example is for an oestrogen pathway AOP in fish. [Note this example is for illustration purposes only, and is not meant to be a full representation of the AOP.] An example of documenting analytical validation for the Fish Short-Term Reproduction Assay is presented as Appendix A.

[1] https://www.epa.gov/test-guidelines-pesticides-and-toxic-substances/series-890-endocrine-disruptor-screening-program

[2] http://www.oecd.org/env/ehs/testing/oecdworkrelatedtoendocrinedisrupters.htm

Figure 3: An example of mapping assays to the AOP and characterising assay performance. *Reproduced from Patlewicz* et al*, 2015*



In cases where a KE in an AOP is mapped to a new or novel method, it is incumbent upon the AOP developer to provide sufficient documentation and citations to enable independent verification of the test method performance (sensitivity, specificity, reliability and domain of applicability).

2.5.2 Conclusions

Having confidence in the scientific methods used to evaluate biological activity for each KE is essential for supporting applications of the AOP for specific uses. As demonstrated by Patlewicz et al (2015), and summarised here, documenting Analytical Validation as part of incorporating the Scientific Confidence Framework into AOPs will help to communicate broadly to potential users of an AOP the scientific strengths and limitations associated with specific test methods that measure the nature and degree of response that a substance elicits which is indicative of its activity for a given KE.

2.6 Exposures

2.6.1 Overview

Since AOPs are, by definition, non-specific regarding chemical identity, the AOP description does not, as of yet, specifically require any information on external or internal exposure aspects, though it can be included on a volunatary basis (Groh et al, 2015).

However, if AOPs are to be used as supporting tools to answer questions in the regulatory field of screening, prioritisation, hazard and risk assessment of chemicals it is necessary to consider exposure conditions and toxicokinetic aspects.

Adsorption, distribution, metabolism and elimination (ADME) of the substance has to be integrated in the assessment process to understand the relation between internal and external concentrations of chemicals *in vivo* and *in vitro*. This is especially crucial for the understanding of the relevance of effects observed in *in vitro* assays, which are commonly used to assess the MIE, for the *in vivo* situation. The *in vitro* to *in vivo* extrapolation (IVIVE) should represent the actual dose in the *in vitro* test system rather than the applied dose. It is therefore important for extrapolations to be based on the free concentration of the test chemical which is available to interact with the biological systems (Groothuis et al, 2015).

2.6.2 Absorption

"Practical application of AOPs in chemical-based risk assessment will require extrapolation of an *in vitro* concentration expected to trigger an MIE to an *in vivo* biologically-effective target tissue dose, which can then be used to estimate a regulatory-relevant external dose (i.e. using reverse toxicokinetics). This extrapolation cannot be made without considering exposure, as well as absorption, distribution, metabolism, and excretion (ADME) properties of a chemical. The most active chemical in an *in vitro* assay may not induce *in vivo* toxicity if concentrations necessary to trigger an MIE are unlikely to be attained due to limited exposure or ADME-mediated processes" (Phillips et al, 2016). Phillips et al (2016), developed a conceptual workflow to refine *in vitro* results by factoring in exposure potential and ADME behaviour, which can then be used to predict *in vivo* MIEs that would initiate an AOP. Philips et al (2016) demonstrate, using the example of chemicals shown to inhibit acetylcholinesterase *in vitro*, that this conceptual framework can be used to (de-)prioritise substances for further assessment/testing. Phillips et al (2016) acknowledged that predicting the likelihood of a substance being metabolised is challenging and propose the use of computational programs to predict metabolism, such as Meteor Nexus[3], if no experimental data are available.

The workflow developed by Phillips et al (2016) primarily addresses qualitative aspects of exposure to identify high priority chemicals for further quantitative analysis. To make such a workflow useable for regulatory purposes, quantitative rates of uptake metabolism and elimination will need to be included.

[3] https://www.lhasalimited.org/products/meteor-nexus.htm

Stadickna-Michalak et al (2014) developed a toxicokinetic model to predict concentrations of chemicals in cells which might facilitate the *in vitro* to *in vivo* toxicity extrapolation. To better understand the relationship between external and internal concentrations of chemicals in fish, they measured and modelled toxicokinetics in cultured fish cell lines of rainbow trout (*Onchorhynchus mykiss*). They found a correlation between the internal effect concentration in fish gills and the fish cell line with log K_{ow}. It was concluded that it might be possible to predict effects on fish based on internal effect concentrations in fish cells, but more data have to be generated. Unfortunately, the project does not take into account metabolism of the chemicals tested. This should be considered in future work to refine the results.

The importance of organism-internal distribution for species sensitivity has also to be taken into account. Nyman et al (2014) showed that compared to *Gammaridae*, the snail *Lymnea stagnalis* was much less sensitive to some neurotoxic pesticides, despite accumulating significantly higher amounts of these chemicals on a whole-body basis. This was explained by the observation that accumulation in the snails was largely restricted to the gastrointestinal complex, while much lower amounts were detected in the nervous system, which is the target toxicity site of these compounds (Nyman et al, 2014).

2.6.3 Metabolism

It is also important to identify relevant metabolites and determine their potential effects. A substance that, for example, undergoes a rapid first pass metabolism in the liver, may not reach the target organ via systemic circulation. On the other hand, metabolism may lead to systemically available metabolites that drive a toxicological effect.

Metabolism should be considered even if a substance tests negative in an *in vitro* assay because it is still possible that the metabolite reaches the molecular target. Conversely, metabolism should also be investigated if an effect occurs in an *in vitro* system: This might be due to the fact that the system has no metabolic capacity and in contrast to the organism, no ability for detoxification. Without this check an *in vitro* false positive result might be overlooked.

An understanding of the nature and the kinetics of the formation, distribution and excretion of relevant metabolites and their possible interactions with the target tissues is therefore crucial to determine the relevance of an effect observed *in vitro* and to the design of relevant studies also taking into consideration metabolites. Metabolic processes or the induction of metabolising enzyme(s) also influence the metabolism of hormones and it is important to understand the kinetics of both processes to assess whether adverse effects can occur at relevant exposure concentrations.

2.6.4 External exposure

Exposure concentrations and conditions in *in vitro* assays should consider the availability of the tested substance to the target cells. Chemicals may differentially and non-specifically bind to medium constituents such as serum protein and lipids, well plate plastic and cells. They may also evaporate, degrade or be metabolised over the exposure period at different rates. Studies have shown that these processes may significantly alter the bioavailable and biologically effective dose of test chemicals in *in vitro* assays. This

subsequently hampers the interpretation of *in vitro* data to predict and compare the true potency of test substances (Groothuis et al, 2015).

External exposure factors e.g. pH, temperature, composition of exposure medium, can alter the speciation of chemicals and with this, bioavailability as well as bioaccumulation potential. For example, exposure medium composition affects physico-chemical properties and toxicity of silver nanoparticles and silver ions (Groh et al, 2014). Medium pH is also known to affect the speciation and bioconcentration potential of ionisable compounds.

The uptake rate into cells may also be affected by the presence of uptake and efflux transporters. *In vivo*, blood-brain barrier epithelia, gut epithelia and hepatocytes, for example, contain uptake and efflux transporters that actively transport specific chemicals across the cell membrane. Cells lacking specific transporters, as a number of cell lines do, will be poor surrogates for *in vivo* toxicity regulated by these transporters (Groothuis et al, 2015).

Exposure routes (e.g. oral dosage, waterborne exposure) may also have a large influence on subsequent toxicokinetic processes, in the organisms. This has to be taken into consideration when extrapolating laboratory results to the real exposure situation in the environment.

Although human external exposure data may be available in some situations (e.g. where the chemical of interest is present in foods or other consumer products with well-defined consumer use patterns), for a number of environmental chemicals exposure data are often lacking. Furthermore, information on internal exposure of humans or environmental species is rarely available. These data are critical in enabling either priority setting or risk assessment using AOPs.

Additionally, when using *in vivo* hazard or mode of action data it is important to consider the exposure level compared to actually occurring exposure of humans and the environment. *In vivo* studies using very high levels of exposure can potentially lead to a saturation of metabolic pathways and lead to a shift to another pathway that is interacting with hormone metabolism.

2.6.5 Quantitative relationships

To use the AOP methodology for regulatory purposes such as risk assessment, quantitative AOPs will need to be established and thresholds for MIE to trigger the downstream KEs or AOs need to be defined. As a way forward for ecotoxicological assessments Groh et al (2015) recommend to first define the approximate concentration ranges where a particular AOP is likely to be activated. This concentration range could then be compared to the environmentally relevant concentrations to which most of the organisms in the wild are likely to be exposed, taking into account exposure routes, toxicokinetic considerations and cumulative exposure to mixtures of similarly acting toxicants. This comparison would allow identification of the AOPs most likely to operate in a particular case. In addition, it may also be used to prioritise the AOPs for which more detailed quantitative definitions should be developed first. KE thresholds would need to be defined with regard not only to the dose of the chemical, but also to the duration of exposure necessary to trigger a particular AOP. These exact same principles could also be applied to human health assessments with minor adaptations such

as considering human-specific exposure scenarios, e.g., occupational exposures and residues in processed foodstuffs.

The usability of the AOPs depends on the regulatory question being asked and the context of the regulatory decision determines what level of uncertainty (e.g. amount of quantitative data) is acceptable (Willet et al, 2014).

2.7 Quantitative Understanding and Predictive Modelling

2.7.1 Overview

AOPs are laid out in a linear fashion, progressing from the MIE to the AO. However, it cannot be assumed that demonstration of an MIE will always lead to the AO as a significant number of factors are required for a quantitative understanding of the KEs and KERs. These factors include the intrinsic potency of a chemical to initiate the MIE and biological redundancy / adaptation processes that can affect the threshold for the toxicological response. As many of the tools used to assess MIE/KEs are *in vitro* assay systems, a thorough understanding of how these relate to what may happen *in vivo* must also be assessed using reliable models e.g. IVIVE. In their article on the need for and use of IVIVE models, Gülden and Seibert (2005) identified two 'key problems' with the use of *in vitro* assays for hazard assessment:

- The toxicodynamic problem - The endpoints of toxic action detectable *in vitro* are less complex and, importantly, mostly different from those assessed *in vivo*.
- The toxicokinetic problem - Toxic concentrations determined *in vitro* are not equivalent to toxic doses or concentrations *in vivo*. This is due to important differences in biokinetics and bioavailability of chemicals *in vitro* and *in vivo* systems.

2.7.2 Examples

In mammalian toxicology, *in vitro* to *in vivo* extrapolation (IVIVE) and physiologically-based pharmacokinetic (PBPK) modelling techniques are being developed at in increasingly rapid pace and are proposed for use in several areas of toxicology (e.g. Yoon et al, 2012; Bale et al, 2013) with recent examples suggesting how they can be used in the area of endocrine disruption:

- Becker et al (2015) have used IVIVE as part of a proof on concept exercise building an exposure : activity profiling method for interpreting high-throughput screening data for oestrogenic activity.
- Campbell et al (2015) have used PBPK and IVIVE approaches to determine margins of safety for methyl-, propyl- and butylparaben by comparing the effective concentrations from an *in vitro* assay of oestrogenicity to calculated human internal doses.
- Silva et al (2015) have used IVIVE to compare data from endocrine- (and neurotoxicity- and developmental-) relevant *in vitro* assays conducted as part of US EPA's ToxCast program with data generated using the same chemicals *in vivo*.

For ecotoxicology, the development of IVIVE models has been slow with only a few examples developed so far e.g. Stadnicka-Michalak et al (2014). However, no specific applications to endocrine disruption have yet to be employed. Further, adversity needs to be developed beyond the individual since the protection goal is at the population level. Therefore, it is likely that the relevance of an AOP needs to be assessed using suitable modelling techniques that can assess population level impacts. Although such tools are widely available for many taxa their use in a regulatory context is limited (Schmolke et al, 2010). Further, for AOP application mechanistic models will undoubtedly be required to form the linkage between the mode of action and the adverse effect (Ankley et al, 2009). Where these linkages have been investigated, for example vitellogenin and sex steroids in small fish models relatively simple models with toxicological input variables from short-term screening data (fecundity) have been used (Miller et al, 2007; Ankley et al, 2008). Such approaches are limited by the poor analytical power of the mechanistic endpoints and lack of realism in the modelling approaches to account for true population dynamics (e.g. density dependence) in order to accurately predict the adverse effect at the population level. Therefore, developing appropriate tools is a high priority to enable quantitative applications for ecotoxicology.

2.7.3 Application Comments

An understanding of the quantitative aspects of an AOP and use of reliable IVIVE/population relevance tools are important for hazard assessment and absolutely critical for their use in risk assessment and therefore also of critical importance in an integrated testing and assessment (IATA) approach. Failing to consider quantitative aspects could lead to misidentification of a chemical as an endocrine disrupter.

For priority setting and read-across, a thorough understanding of the quantitative aspects of an AOP is not necessarily required. However, a basic understanding of these factors, particularly potency, will considerably improve priority setting by discriminating chemicals of higher priority (high potency) from those of lower priority (low potency). An understanding of how potency relates to human or environmental exposure is critical for priority setting, since where there is a low probability of exposure and effects concentrations overlapping (i.e. for low activity chemicals associated with low exposures) there is a low priority for further testing (Dent et al, 2015). An understanding of potency will also help with read-across approaches, particularly when reading-across from a potent chemical that has been used as an exemplar to demonstrate an AOP to a less potent chemical.

2.8 Taxonomic Applicability / Relevance / Species Concordance

2.8.1 Overview

Mammals

Most toxicological guideline testing, as well as mechanistic data, is based on and is available for mammalian systems. As such, chemicals are tested in a variety of species (depending on the specific chemical regulation), including rat, mouse, rabbit, dog and guinea pig. Many physiological systems and receptors are well conserved during evolution in the various mammalian species which allows the extrapolation of effects and mechanisms

to humans. The default assumption is therefore that any finding in mammalian test species is relevant to humans. Thus, it is assumed that AOPs based on the mammalian test systems mentioned above are also existing in and relevant for humans.

In general, this is a correct approach and forms the basis for hazard and risk assessment, however, it is important to recognise potentially strain- or species-specific responses. Two prominent examples for this are the tissue-site specific induction of tumours resulting in adenomas in the thyroid of rodents and sex/strain specific induction of tumours resulting in Leydig cell tumours in rodents, which are a pronounced effect in some rat strains (i.e. in Fischer rats) while almost absent in others.

Further, there can be differences between mammalian species regarding ADME processes, so that there can be quantitative differences in AOPs between species.

Due to their taxonomic relatedness, mammalian data are potentially also relevant for other vertebrates such as fish, birds, amphibians and reptiles. However, differences between these vertebrates and mammals regarding ADME processes will likely lead to greater quantitative differences in AOPs. However, for priority setting, based on activity (yes/no), this need not be a limitation.

Other vertebrates

The second most studied vertebrate group is fish. Receptor-mediated oestrogenicity and androgenicity as well as aromatase inhibition as observed in mammalian systems seem to translate reasonably well to fish; in the sense that both organism groups show typical and rather specific *in vivo* responses. In fact, there is some evidence for biosynthesis inhibition that fish are a better model than mammals (Ankley and Gray 2013). However, for other MoAs the degree of similarity is less pronounced and/or specific endpoints absent (e.g. low sensitivity for receptor-mediated anti-androgenicity in fish).

Birds have a different genetic sex determination system than mammals: males are the homozygote sex (ZZ), while females are heterozygotes (ZW) and thus the egg determines a chick's gender. This obviously has consequences for transferability of mammalian findings regarding hormonal influences on sexual differentiation to birds.

Amphibian metamorphosis is governed by thyroid hormones; so the same hormone regulates a different process than in mammals. Some sexual steroid hormones do have the same function in these vertebrate groups.

Overall it can be stated that there would be more confidence in an AOP when it is observed in multiple vertebrate species, while it is also acknowledged that some AOPs can be specific for certain taxa.

Invertebrates

The relevance of mammalian data for invertebrates is much lower, due to dissimilarities in hormonal systems; e.g. arthropods have different hormone systems (ecdysteroids, juvenile hormones) which are absent in vertebrates (Gunnarsson et al, 2008). Even where certain hormone systems are shared there is evidence for functional differences (e.g. the molluscs; Scott 2012 and 2013).

For invertebrates there are few mechanistic *in silico*, *in vitro* and *in vivo* assays (e.g. for insects, see Weltje, 2013) since traditionally invertebrate testing has been more focussed on capturing apical endpoints. So adverse outcomes in invertebrates are well described, while the underlying mechanisms are often poorly understood.

Finally, it should be considered that the general protection goal for ecotoxicological risk assessments is the population (for human health it is the individual), but in some regulatory contexts (e.g. EFSA Aquatic GD, 2013) vertebrates follow stricter rules than invertebrates. In extrapolating AOPs across species, other aspects than taxonomic relatedness should also be considered, e.g. reproductive strategies such as uni- or multivoltinism and r- and K-strategists, which can have a large influence on compensation for stress at the population level.

2.9 Test System Relevance

2.9.1 Overview

A key motivation for developing AOPs is to assist in the organisation of toxicological knowledge to facilitate the application of mechanistic information to regulatory decision making. In this context, a KE is defined as an observable change that is necessary (but not necessarily sufficient) for the progression towards a specific AO. A KE must be measurable, and therefore by definition will often be characterised by experimental observations and mechanistic data such as changes in gene expression, protein expression, alterations in morphology or a physiological dysfunction. Therefore, confidence in the test systems that inform an AOP greatly impacts the confidence in the predictive outcome and the relevance of the outcome to informing the regulatory decision. In section 2.6, factors that inform confidence in the test methods used to measure responses indicative of KEs (biological and analytical) are discussed. In this section, the test system factors that underpin the confidence in their relevance to the specific regulatory question are discussed. Considerations important in the context of test system relevance fall into two broad categories: 1) compatibility of the test system with the chemical of interest, and 2) appropriateness of extrapolating the test system response to the ecological or biological system in question.

2.9.2 Factors to consider

2.9.2.1 Chemical domain

AOPs are not intended to be chemical specific, however for application purposes they will be applied in a chemical specific manner. Chemical properties are an important determinant of compatibility with a given test system or biological model. Therefore, the limitations of a given test method to accommodate the physical and chemical properties must be considered. This is important to consider in the context of test system relevance because it informs if the properties of the chemical of interest will limit the usefulness of the AOP to inform a regulatory question. For example, are there solubility properties of the substance that limit their compatibility with the system? Does the chemical need to be modified or manipulated in some manner to be

used in the test system (e.g. how relevant are exposures that are facilitated by use of solvents in a test system to a “real world” situation?).

2.9.2.2 Biological domain test systems

It is recognised that most biological responses mediated by chemical exposures will occur in a systems context, meaning interaction and crosstalk with other pathways are important to the observed whole animal outcome. However, AOPs are likely to be characterised primarily by *in vitro* or *ex-vivo* assays. Therefore if the test system outcome is to be used to inform a regulatory application, the systems relevance of the assays that are mapped to the various events that define an AOP needs to be considered. This means considering if there are any limitations imposed by the reduced biological complexity of the test system that may influence interpretation of the outcome. This needs to be reflected upon in both a chemical —and regulatory question-specific manner. For example, does the test system adequately reflect the in vivo situation (e.g. bioavailability or toxicokinetic differences), does it represent the complexity of the whole organism (e.g. loss of critical cell interactions), are their critical feedback loops missing that influence outcome interpretation (e.g. surgically modified animals), does the system accommodate bioactivation and is this important? For ecotoxicology questions, it is also important to consider the appropriateness of the test system outcome to predict population-relevant effects.

2.9.2.3 Species applicability domain

By their nature, test systems are typically limited in their immediate species relevance. However, further test system development or experimentation can often support relevance across a broader species domain. The relevance of the test system to the species of focus for the regulatory application and level of confidence in this knowledge base needs to be considered. Is relevance to that species assumed, or is the relevance based on data? What uncertainty exists in the understanding of species relevance that should be recognised in the regulatory decision? Can it be further informed by supplementing with additional data?

2.9.2.4 Predictive domain

This factor considers the biological bounds of interpretation of the assay outcome, and the level of confidence in this knowledge base. For example, how well has the assay been demonstrated to correlate with an *in vivo* outcome? What KEs in the AOP does the assay map to and is this endpoint specific (receptor binding) or is it representative of a more generic response (e.g. oxidative stress). Is the effect an appropriate measurement for regulatory purposes (e.g. is it adverse in all cases, does it reflect a population level effect)? Can a test system stand-alone in its predictive power for the key event, or does it need to be integrated with outcomes from other test systems to support a confident prediction?

2.9.3 Application considerations

AOPs are fundamentally intended to be a simplified description of a biological response pathway built from biologically simplified test systems. Therefore, limitations and informed assumptions will be necessary for

AOPs to be useful in most regulatory applications. The intent here is to emphasise that test system limitations need to be acknowledged and the assumptions for extrapolating the test system outcomes to the regulatory decision understood. Test system relevance factors can also be useful for guiding improvements in the description of the assays and test systems that map to the KEs in an AOP.

In general, understanding the chemical domain of the test systems is essential for understanding if the outcome of an assay is meaningful in a chemical specific context. Therefore, the chemical domain factor is seen as a minimal requirement for all chemical specific regulatory applications. For the other factors, the better defined the limitations of the test system are, the more they can be considered in the interpretation of outcomes resulting in greater confidence in the utilisation of the test system outcome for regulatory decision making. Understanding the predictive domain becomes more relevant for regulatory applications of a predictive nature (hazard identification and IATA); however, integration of a battery of assays may improve overall predictive confidence for a number of low confidence assays. Uncertainty in species domain applicability can result in false positive and negative predictions. The limitations and uncertainties in the approaches taken should be characterised in all cases of AOP application.

Table 2 below summarises which relevance consideration need to be considered in different AOP application scenarios.

Table 2: Table summarising which relevance considerations need to be considered in different AOP application scenarios. X *= needs to be considered.*

	Chemical domain	Species applicability domain	Biological domain	Predictive domain
Priority Setting	X			
Hazard ID	X	X	X	X
Risk Assessment	X	X	X	X
IATA	X	X	X	X
Read-across	X	X		

2.10 Other Factors

2.10.1 Overview

It is well known that adverse effects can manifest in components of the endocrine system (e.g. the gonads and the thyroid) and/or endocrine system-dependent biological functions / processes (e.g. reproduction and development) as a secondary consequence of other factors (Hutchinson et al, 2009; Everds et al, 2013). In any case, these other factors are non-endocrine toxicities; these can be very general (e.g. general systemic toxicity) or very specific (e.g. hepatic effects leading to perturbations of circulating hormone levels). Identification of these types of effects is very important as they may lead to the misidentification of a chemical as an endocrine disrupter (WHO/IPCS [2002] definition: "*An endocrine disrupter is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or [sub]populations.*").

It should also be noted that in some cases, a KE or AO may manifest as a result of an effect (or multiple effects) on the endocrine system different from that specified in the adverse outcome pathway (AOP) being applied. Understanding of this is important as it will influence the design of any necessary follow-up studies.

2.10.2 Examples

2.10.2.1 General systemic toxicity

General systemic toxicity (generally measured as effects on survival, body weight and/or the presence of clinical signs of toxicity) is known to have a profound effect on endocrine organs / processes and so the observation of a KE/AO of an endocrine AOP in the presence of such toxicity must be carefully scrutinised. Examples of this type of effect are:

- Mating behaviour and fertility / fecundity are known to be adversely affected by general systemic toxicity (Terry et al, 2005).
- Increases in post-implantation loss, abortions and certain types of developmental variations and malformations are known to occur in mammalian species as a result of marked maternal general systemic toxicity during gestation (Khera, 1985).
- Pubertal development (vaginal opening / preputial separation) in rodents is bodyweight-dependent. Therefore, reduced bodyweight gain, often of the magnitude required to demonstrate achievement of a maximum tolerated dose (MTD), can lead to delays in pubertal development (Marty et al, 2003).
- General systemic toxicity can lead to a developmental delay in amphibians, including in the appearance of the thyroid gland. Therefore, it is critical that a developmental stage-matching approach is used when comparing thyroid histology between control and treated groups (OPPTS 890.1100 - US EPA, 2009).
- In the presence of general systemic toxicity in mammals, is it common to see weight changes (decreases) and histopathological changes (e.g. atrophy) in gonads.
- Non-specific effects on fecundity, development, vitellogenin synthesis, expression of secondary sexual characteristics and gonadal histopathology have been observed in fish and amphibian models (Wheeler and Coady 2016).

2.10.2.2 Specific non-endocrine toxicities / modes of action

Other specific non-endocrine toxicities can result in secondary effects on the endocrine system. In many cases, these specific non-endocrine toxicities can be considered the ‘lead’ or ‘primary’ toxicity of the chemical under evaluation and often occur at lower doses / concentrations than any apparent effect on the endocrine system. In some cases, the toxicities are related to the intended mode-of-action (e.g. pharmaceutical or pesticidal) of the chemical. Examples of this type of effect are:

- Hepatic effects leading to perturbations of circulating hormone levels.
- In rodents, induction of hepatic UDPglucuronyltransferase (UDPGT) activity can result in perturbations of circulating T_3/T_4 and thyroid stimulating hormone (TSH) levels and subsequent adverse effects, including tumours, in the thyroid (Dellarco et al, 2006).

- In fish, hepatotoxicity can lead to a reduction in vitellogenin (VTG) expression, with subsequent effects on egg production.
- Inhibitors of angiogenesis are known to have effects on mammalian oestrus cycling and fertility (Fraser and Lunn, 2000).
- Osteo-renal syndrome (renal osteodystrophy) results from hyperparathyroidism secondary to hyperphosphatemia combined with hypocalcaemia, both of which are due to decreased excretion of phosphate by the damaged kidney.

2.10.2.3 Other (or multiple) endocrine toxicities

In some cases, a KE or AO may manifest as a result of an effect (or multiple effects) on the endocrine system different from that specified in the adverse outcome pathway (AOP) being applied. Understanding of this is important as it will influence the design of follow-up studies, if necessary. For example, AOP 23 (from AOP wiki; androgen receptor agonism leading to reproductive dysfunction) states that a reduction in plasma 17β-oestradiol is attributable to androgen receptor agonism.

Table 3: Summary of AOP 23 (from AOP wiki) androgen receptor agonism leading to reproductive dysfunction

Molecular Initiating Event	Support for Essentiality
Androgen receptor, Agonism	Strong

Key Events	Support for Essentiality
Testosterone synthesis by ovarian theca cells, Reduction	Moderate
17β-oestradiol synthesis by ovarian granulosa cells, Reduction	Moderate
Plasma 17β-oestradiol concentrations, Reduction	Strong
Transcription and translation of vitellogenin in liver, Reduction	Moderate
Cumulative fecundity and spawning, Reduction	Moderate
Plasma vitellogenin concentrations, Reduction	Strong
Vitellogenin uptake into oocytes and oocyte growth/development, Reduction	Weak
Gonadotropins, circulating concentrations, Reduction	Moderate

Adverse Outcome
Population trajectory, Decrease

The same KE (reduction in plasma 17β-oestradiol) and likely the remainder of the AOP could also be observed as a result of inhibition of aromatase (CYP19). Understanding this is important for designing follow-up studies, as well as for priority selection/read-across (i.e. ensuring the right screening *in vitro* assay is employed).

2.10.3 Application comments

For the reasons detailed above, an understanding of the potential impacts of other factors is of critical importance when using AOPs for hazard identification, risk assessment and read-across. Failing to consider such effects could lead to misidentification of a chemical as an endocrine disrupter.

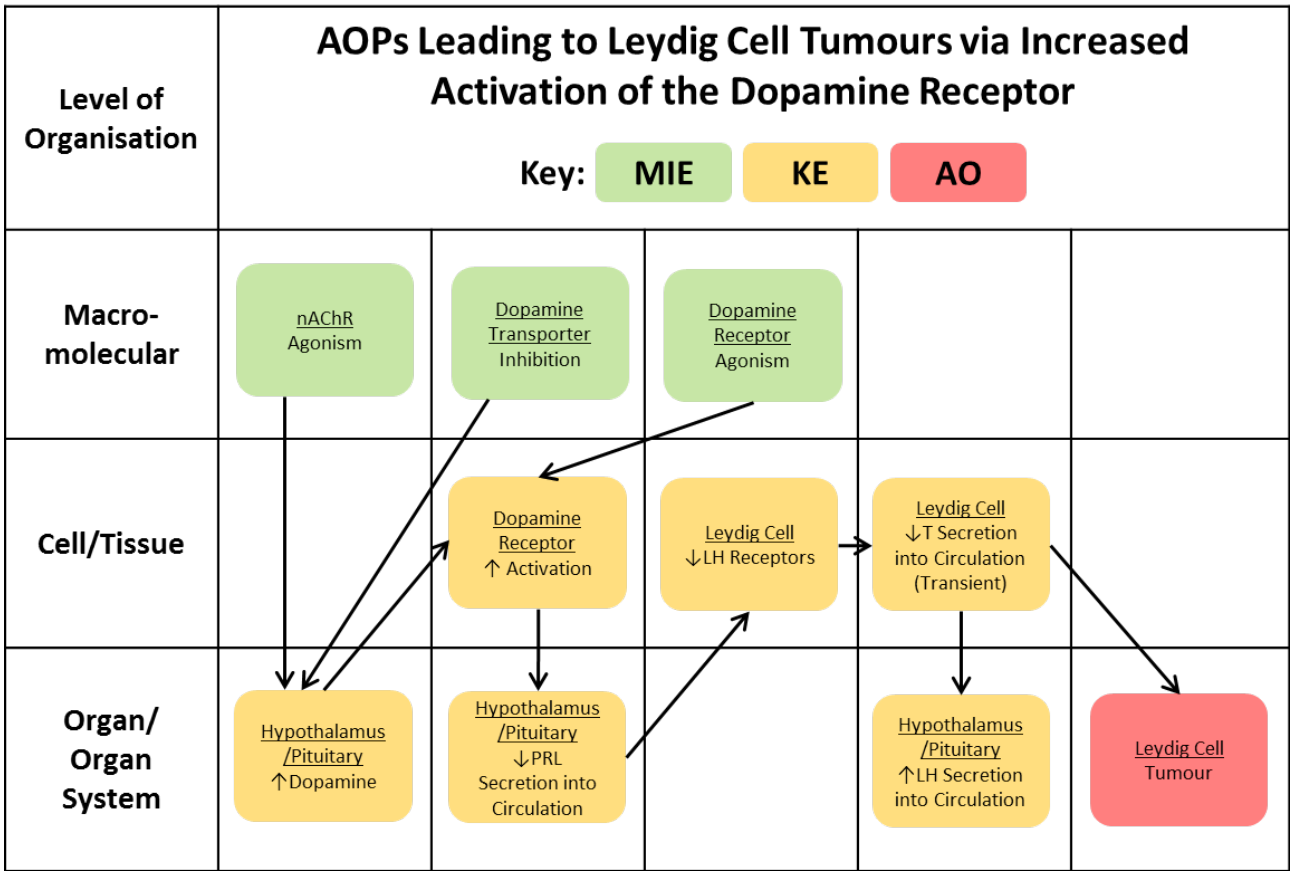
For priority setting (e.g. screening for effects against defined targets in *in vitro* assays), the potential impact of other factors is not as important. However, where priority setting has been triggered based on concern from related chemical(s), it is important to ensure that the correct tools (*in vitro* assays measuring effects against the correct endpoints) are used.

3. CASE STUDY: ADVERSE OUTCOME PATHWAYS (AOPS) LEADING TO LEYDIG CELL TUMOURS VIA INCREASED ACTIVATION OF THE DOPAMINE RECEPTOR

3.1 Introduction

Figure 4 describes a series of biological and biochemical events leading to an increased incidence of Leydig cell tumours, a tumour of the testis. Although these events could be described as three separate AOPs with three distinct Molecular Initiating Events (MIEs), they can be sensibly considered together as a collective of AOPs leading to Leydig cell tumours via increased activation of the dopamine receptor as all three converge at the same shared Key Event (KE) of 'increased activation of the dopamine receptor' with the same downstream KEs culminating in the same AO of 'Leydig Cell Tumour'.

Figure 4: AOPs Leading to Leydig Cell Tumours via Increased Activation of the Dopamine Receptor



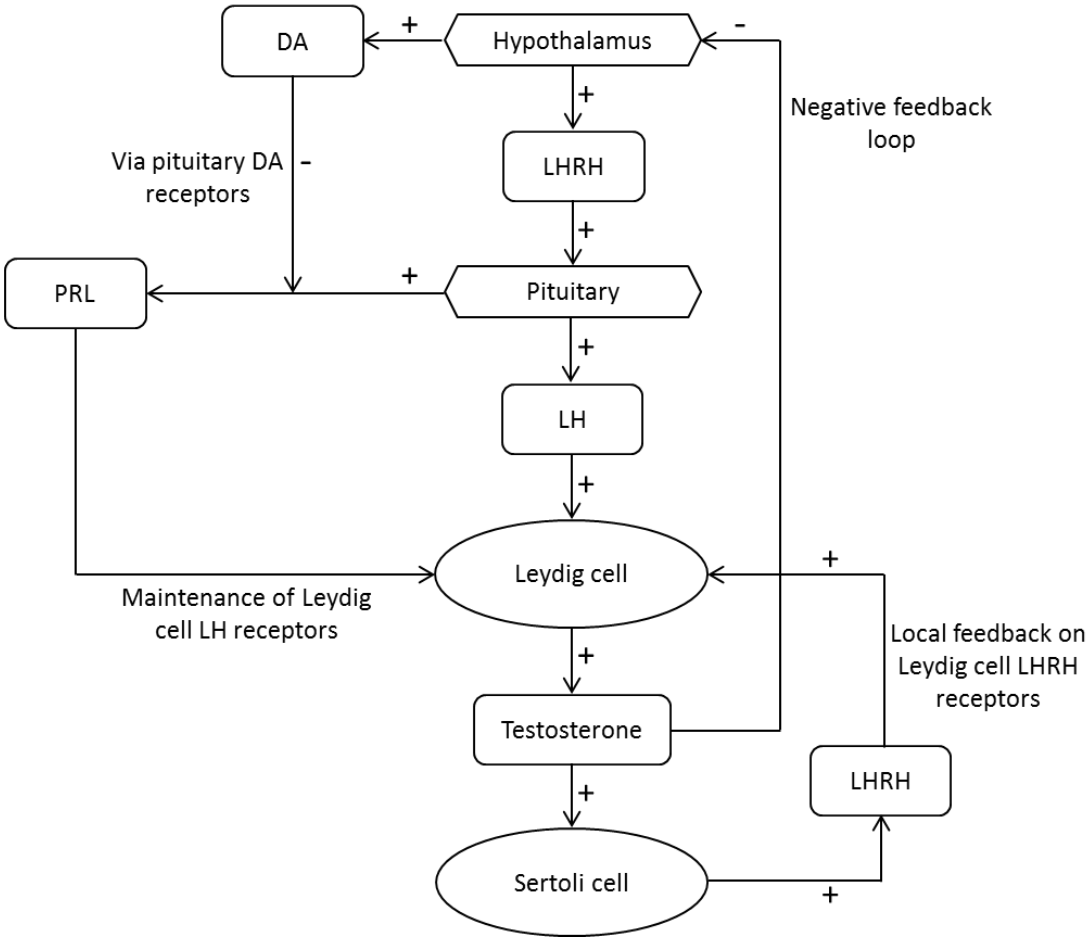
MIE = Molecular Initiating Event, **KE** = Key Event, **AO** = Adverse Outcome, **nAChR** = Nicotinic Acetylcholine Receptor, **PRL** = Prolactin, **LH** = Luteinising Hormone, **T** = Testosterone

3.2 Biological Plausibility

3.2.1 Biological Plausibility of the AOPs

The biological plausibility of these AOPs is well established. These AOPs are fully consistent with the current understanding of the underlying biology. The relevant underlying biology is described as the hypothalamic-pituitary-gonadal (HPG) axis and is well characterised and understood. In males, the HPG axis describes the series of chemical and biological interactions involved in regulation of testosterone biosynthesis and is described in Figure 5.

Figure 5: The Hypothalamic-Pituitary-Gonadal (HPG) Axis. *Adapted from Prentice and Meikle, 1995*



DA = Dopamine, **PRL** = Prolactin, **LH** = Luteinising Hormone, **LHRH** = Luteinising Hormone Releasing Hormone (also known as Gonadotropin Releasing Hormone [**GnRH**]).

3.2.2 Chemical-Specific Plausibility of the AOPs

In addition to being biologically plausible, it is important that AOPs can be shown to be plausible when actually applied to specific chemicals. Some examples are listed below. These examples have been collectively referred to as 'dopamine agonists' in that they all result in increased dopaminergic activity, i.e. they all converge at the same shared KE of 'increased activation of the dopamine receptor' but may differ with respect to the MIE/KEs upstream of this shared KE.

Direct dopamine receptor agonism

Mesulergine: Specifically designed as a dopamine D2 receptor agonist for the treatment of hyperprolactinaemia. Resulted in an increased incidence of Leydig cell tumours (LCTs) after chronic administration to male Wistar rats (Prentice et al, 1992).

Norprolac (Quinagolide): Specifically designed as a dopamine D2 receptor agonist for the treatment of hyperprolactinaemia. Resulted in an increased incidence of LCTs after chronic administration to male Sprague Dawley rats (Roberts et al, 1993).

Inhibition of dopamine transport

Oxonilic acid: A quinolone antibiotic that inhibits bacterial DNA gyrase. Resulted in an increase in LCTs after chronic administration to male Wistar rats. Subsequent mode of action work showed that this was a result of the ability of oxolinic acid to inhibit dopamine transporters in the hypothalamus, resulting in decreased re-uptake of dopamine. This results in increased dopamine within the hypothalamic-hypophyseal portal system, resulting in increased activation of dopamine receptors in the anterior pituitary (Yamada et al, 1994; Garcia de Mateos-Verchere et al, 1998; Casarett et al, 2007).

Nicotinic acetylcholine receptor (nAChR) agonism

Sulfoxaflor: A plant protection product active ingredient with insecticidal properties mediated through its agonism of insect nicotinic acetylcholine receptors (nAChRs). Resulted in increased size and bilateral incidence of LCTs after chronic administration to Fischer 344 rats. Subsequent mode of action work showed that this was a result of the ability of sulfoxaflor to act as an agonist of hypothalamic nAChRs (Rasoulpour et al, 2014). Hypothalamic nAChRs, such as α4β2 and α4α6β2, play a key regulatory role in dopamine release from dopaminergic neurons (Maskos, 2010). Increased activation of central nAChRs results in increased dopamine secretion into the hypothalamic–hypophyseal portal system, resulting in increased activation of dopamine receptors in the anterior pituitary in a similar fashion as described above for oxolinic acid.

3.3 Essentiality

The essentiality of the molecular initiating events (MIE) and all the key events (KEs) has been well demonstrated for all the chemicals referred to under ***Chemical-Specific Plausibility of the AOPs*** using relevant, well conducted and robust experiments.

3.4 Empirical Evidence

As mentioned above, strong empirical evidence, including temporal and dose/concentration concordance, has been generated for these AOPs using a range of chemicals in relevant, well conducted and robust experiments.

In their assessment of sulfoxaflor, Rasoulpour et al (2014) provide an excellent example of how temporal and dose concordance can be considered (see Table 4).

Table 4: Temporal and Dose Concordance of Key Events for Sulfoxaflor-Induced Rat LCTs.
Adapted from Rasoulpour et al*, 2014*

Temporal →

Dose ↓

Dose (ppm in diet)	KE1	KE2	KE3	KE4	KE5	KE6
	↑Dopamine release via nAChR agonism	↓Serum prolactin levels **2-8 weeks**	↓Luteinzing hormone receptors in Leydig cells **4-8 weeks**	↓Serum testosterone **2-8 weeks**	↑Serum Luteinizing hormone **2-8 weeks**	↑Size/ bilateral incidence of LCTs **>1 year**
25		-	-	-	-	-
100		-	-	-	-	-
400				+[b]		+ (size)
500	+[a]	+ (4 weeks) - (2, 8 weeks)	+ (4 weeks) - (2, 8 weeks)	-	+ (4 weeks) - (2, 8 weeks)	+ (size and bilateral incidence)

+ = effect present, **-** = effect absent, **blank cell** = no data, **a** = data from microdialysis experiment at 400 µM sulfoxaflor to target 40 µM in plasma (approximately equivalent to internal dose following administration of 500 ppm in the diet), **b** =inferred from a delay in balanopreputial separation in male offspring in a 2-generation reproductive toxicity study.

3.5 Analytical Validation

Although no national or international test guidelines are available for assessment of the MIEs or any of the KEs, robust and reliable *in vitro* assays to assess effects on nAChR and dopamine receptors and the dopamine transporter are widely available. In addition, reliable assays (e.g. radioimmunoassays and electroimmunoassays) have been developed by numerous different laboratories for routine assessment of circulating levels of all the hormones involved in this AOP (prolactin [PRL], testosterone [T], luteinising

hormone [LH]). Incidences of LCTs would typically be observed in chronic rat studies conducted according to internationally-recognised test guidelines.

3.6 Exposures

As with all AOPs, an understanding of factors such as exposure and absorption, distribution, metabolism and excretion are critical in application. As mentioned in ***Analytical Validation*** above the MIEs described here are can be readily assessed using *in vitro* assays. Whilst useful for identifying if a chemical has intrinsic activity, both the potency of the chemical and the relevance of the concentrations tested to those achieved *in vivo* must be considered. In the absence of *in vivo* kinetic data, there are a number of readily available *in vitro*/*in silico* physiologically based pharmacokinetic (PBPK) modelling techniques that can be used to inform *in vitro* to *in vivo* extrapolations (IVIVE).

3.7 Quantitative Understanding and Predictive Modelling

Whilst no predictive models have yet been established for these AOPs, the available data for the example compounds described in ***Chemical-Specific Plausibility of the AOPs*** allow for a strong quantitative understanding of the key Event Relationships (KERs). For these example compounds, the mechanistic data that described most of the KEs were generated in response to the observation of LCTs at the conclusion of a chronic study; however, these same data could be used, for example, to help guide prioritisation of untested chemicals. The first step in such as assessment could be to assess the potential of a chemical with respect to the MIEs, considering PBPK and IVIVE where possible. Next, those compounds that show potential to activate the MIE could be tested in short term *in vivo* studies, again considering 'real world' exposures to assess potential for effects on PRL and LH. Based on the available data, absence of effect on PRL and LH in a short term *in vivo* study would be a reliable indicator that intrinsic activity observed *in vitro* was not expressed *in vivo*.

3.8 Taxonomic Applicability/Species Concordance

These AOPs are considered to be specific to the male rat (and have been demonstrated for a range of strains: Wistar, Sprague-Dawley and Fischer-344) and not relevant to humans owing to multiple qualitative and quantitative differences between the species (Prentice and Meikle, 1995; Clegg et al, 1997; Foster, 2007; Rasoulpour et al, 2014).

Literature data suggests that there may also be fundamental differences between rats and mice in terms of Leydig cell tumourigenesis. For example, Murakami et al (1995) showed a species difference in the response of the Leydig cell to increased LH levels as a result of treatment with procymidone and suggested this as a mechanistic rationale as to why chronic administration of procymidone resulted in increased incidences of LCTs in rats but not mice.

3.9 Test System Relevance

These AOPs have been demonstrated for a range of unrelated molecules and so the chemical domain of the AOPs in general and specific assays (e.g. *in vitro* nAChR or dopamine receptor agonism or dopamine transporter inhibition) can be considered broad. Owing to the species specificity of these AOPs as discussed in ***Taxonomic Applicability / Species Concordance***, care should be taken to ensure the MIEs/KEs are demonstrated using the appropriate test species (i.e. male rat). Where applicable, the strain of rat used to demonstrate the KEs should match the strain of rat used for the evaluation of carcinogenicity.

3.10 Other Factors

Increases in LCTs in male rats have been observed to occur as a result of other MoAs/AOPs including mutagenicity, androgen receptor antagonism, oestrogen receptor (ant)agonism, 5α-reductase inhibition, aromatase inhibition, reduced testosterone biosynthesis, increased testosterone metabolism and LHRH/GnRH agonism. These other MoAs/AOPs have been reviewed extensively (Prentice and Meikle, 1995; Clegg et al, 1997; Foster, 2007). Therefore, it is important that these other potential causes are considered; especially because some of these alternatives are considered to be relevant to humans, whereas others are not, making this critical information in terms of AOP application for human health hazard / risk assessment (see ***Taxonomic Applicability / Species Concordance***).

In their assessment of sulfoxaflor, Rasoulpour et al (2014) provide an excellent example of how other modes of action/AOPs potentially operating can be ruled out based on a holistic evaluation of all the available data (see Table 5).

Table 5: Evaluation of Alternative Modes of Action/AOPs for Sulfoxaflor-induced Rat LCTs. *Adapted from Rasoulpour et al, 2014*

Alternative AOP	Example Chemical(s)	Strength of association	Consistency of association	Specificity of association	Dose concordance	Temporal concordance	Coherence and plausibility
Mutagenicity	Cadmium	- Not genotoxic	-	-	-	-	- (both)
Androgen receptor (AR) antagonism	Vinclozolin Flutamide	- No effect on AR *in vitro*	-	-	-	-	- (both)
Oestrogen receptor (ER) (ant)agonism	Diethylstilb-estrol	- No effect on ER *in vitro*	-	-	-	-	- (both)
5α-reductase inhibition	Finasteride	- No effect on 5α-reductase gene expression in testes	-	-	-	-	- (both)
Aromatase inhibition	Formestane Letrozole	- No effect on aromatase *in vitro*	-	-	-	-	- (both)
↓Testosterone biosynthesis	Cimetidine	- No effect on steroidogenic genes	-	-	-	-	- (both)
↑Testosterone metabolism	Triazoles	- No increased testosterone metabolism	-	-	-	-	- (both)
LHRH (GnRH) agonism	Buserelin	-	- No evidence from apical endpoints in multiple *in vivo* studies	-	-	-	- (both)
Dopamine agonism/ enhancement	Mesulergine Norprolac Oxolinic acid	+ Moderate	+ Moderate	+ Moderate	+ Moderate	± Weak	+ (both)

+ = attribute present, **-** = attribute absent, **±** = equivocal

4. CONCLUSIONS AND RECOMMENDATIONS

AOPs can be used in many different contexts, and each use of an AOP may require different considerations. It is critical that AOPs are only used to predict AOs where sufficient data exists to allow this. Similarly, AOPs should only be used to explain the aetiology of observed AOs when there is sufficient evidence on the mode of action (MoA) documented in an AOP. Currently, most AOPs should be utilised for developing the knowledge base that can be used in predictive eco/toxicology screening. Inferring the definitive manifestation of AOs from screening based on molecular-initiating events is typically not supported by the level of understanding of a given AOP.

Inappropriate use of AOPs, especially as they refer to endocrine disruption, could result in misleading classification/categorisation of substances as having endocrine disrupting properties without adequate data. Currently, AOPs should be used to strengthen the knowledge of correlative and causative linkages, qualitative and quantitative key event relationships, and understanding of current data gaps in order to be employed appropriately as part of first-level screening.

It must be understood that not all KEs will be equally applicable for all scenarios in which an AOP could be applied. For example, when using an AOP for hazard identification, i.e. determining if a particular AOP is potentially applicable for a specific chemical or group of chemicals, then definitive data describing exposure are less important at that stage. However, understanding of the likelihood of exposure occurring would be important. For priority setting, exposure data can be useful to put the *in vitro* mode of action data into context and determine the need for further work, whilst for risk assessment, a quantitative understanding of exposure is of critical importance to enable a safety decision to be made. Other elements, such as biological plausibility and essentiality, are important for all potential applications. Comments on the applicability of particular Key Elements in different AOP application scenarios are captured in Table 6.

Table 6: Summary of the Key Elements of AOP Utility

<table>
<tr><th></th><th>Priority setting</th><th>Hazard identification</th><th>Read-across</th><th>IATA*</th><th>Risk Assessment</th></tr>
<tr><td>Biological plausibility</td><td colspan="5">AOP consistent with current understanding of the underlying biology, including consideration of potency. Data should be sufficient to link the MIE to the AO via KEs for at least one chemical.</td></tr>
<tr><td>Essentiality</td><td colspan="5">1) Direct evidence showing that blocking of the MIE leads to blocking the AO.
or 2) indirect evidence that the MIE modulates the KE responses and that the KE produce AO.</td></tr>
<tr><td rowspan="2">Other factors</td><td colspan="2">Consider what other AOs could result from the observed MIEs.</td><td></td><td colspan="2">Consider what other AOPs could result in the observed AO.</td></tr>
<tr><td></td><td colspan="4">Understand how AOP can be influenced by factors not characterised as part of this AOP (e.g. general toxicity). Which AOPs share common KEs and KERs and how do they interact?</td></tr>
<tr><td>Empirical evidence</td><td colspan="5">Determine the quantity and quality of the data/information underlying the AOP.</td></tr>
<tr><td rowspan="2">Quantitative understanding/ predictive modelling</td><td colspan="2">Some understanding of quantitative aspects required.</td><td colspan="3">Quantitative aspects of AOP need to be well characterised.</td></tr>
<tr><td colspan="2">IVIVE desirable to put level of concern into context</td><td>IVIVE desirable as part of weight of evidence</td><td colspan="2">Robust IVIVE necessary</td></tr>
<tr><td>Test system validation</td><td colspan="5">Confidence needed in test methods used to evaluate AOP. Where non-validated methods are used documentation should be sufficient to demonstrate validity and reproducibility of the data used.</td></tr>
<tr><td>Taxonomic applicability/ relevance/ species concordance</td><td>Understanding of taxonomic differences</td><td colspan="4">Robust data describing qualitative and quantitative differences across different taxa, species, sex, etc.</td></tr>
<tr><td>Exposures</td><td colspan="5">Critical in all areas but now data are generated/used may vary depending on application.</td></tr>
<tr><td>Test system relevance</td><td colspan="5">What are the domains of applicability of the test systems and how do they relate to the AOP?</td></tr>
</table>

* IATA = Integrated Approach to Testing and Assessment

ABBREVIATIONS

ADME	Absorption, distribution, metabolism, and excretion
AO	Adverse outcome
AOP	Adverse outcome pathway
DEHP	Diethylhexyl phthalate
GLP	Good laboratory practice
GnRH	Gonadotropin releasing hormone
HPG	Hypothalamic-pituitary-gonadal
IATA	Integrated approaches to testing and assessment
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IVIVE	*In vitro* to *in vivo* extrapolation
KE	Key event
KER	Key event relationship
LCT	Leydig cell tumour
LH	Luteinising hormone
LHRH	Luteinising hormone releasing hormone
LOEC	Lowest observed effect concentrations
Log Kow	Octanol/water partition coefficient
MIE	Molecular initiating event
MoA	Mode of action
MTD	Maximum tolerated dose
nAChR	Nicotinic Acetylcholine Receptor
OCSPP	Office of Chemical Safety and Pollution Prevention
OECD	Organisation for Economic Co-operation and Development
PBPK	Physiologically based pharmacokinetic
PPAR	Peroxisome proliferator-activated receptor
PRL	Prolactin
RA	Risk Assessment
SCF	Scientific Confidence Framework

T	Testosterone
TH	Thyroid hormone
TSH	Thyroid stimulating hormone
UDPGT	Uridine diphospho-glucuronyltransferase
US EPA	United States Environmental Protection Agency
VTG	Vitellogenin
WOE	Weight of evidence
ZW	Heterozygote
ZZ	Homozygote

BIBLIOGRAPHY

Allen TE, Goodman JM, Gutsell S, Russell PJ. 2014. Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27(12):2100-2112.

Ankley GT, Gray LE. 2013. Cross-species conservation of endocrine pathways: A critical analysis of tier 1 fish and rat screening assays with 12 model chemicals. Environ Toxicol Chem 32: 1084–1087.

Ankley GT, Miller DH, Jensen KM, Villeneuve DL, Martinović D. 2008. Relationship of plasma sex steroid concentrations in female fathead minnows to reproductive success and population status. Aquatic Toxicol 88(1):69-74.

Ankley GT, Bencic DC, Breen MS, Collette TW, Conolly RB, Denslow ND, Edwards SW, Ekman DR, Garcia-Reyero N, Jensen KM, Lazorchak JM, Martinović D, Miller DH, Perkins EJ, Orlando EF, Villeneuve DL, Wang RL, Watanabe KH. 2009. Endocrine disrupting chemicals in fish: Developing exposure indicators and predictive models of effects based on mechanism of action. Aquatic Toxicol 92(3):168-178.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730-741.

Ankley GT, Villeneuve DL. 2015. Temporal Changes in Biological Responses and Uncertainty in Assessing Risks of Endocrine-Disrupting Chemicals: Insights from Intensive Time-Course Studies with Fish. Toxicol Sci 144: 259-275.

Bale AS, Kenyon E, Flynn TJ, Lipscomb JC, Mendrick DL, Hartung T, Patton GW. 2013. Correlating *in vitro* data to *in vivo* findings for risk assessment. ALTEX 31(1/14):79-90.

Becker RA, Friedman KP, Simon TW, Marty MS, Patlewicz G, Rowlands JC. 2015. An exposure:activity profiling method for interpreting high-throughput screening data for estrogenic activity—Proof of concept. Regul Toxicol Pharmacol 71(3):398-408.

Campbell JL, Yoon M, Clewell HJ. 2015. A case study on quantitative *in vitro* to *in vivo* extrapolation for environmental esters: Methyl-, propyl- and butylparaben. Toxicology 332:67-76.

Casarett LJ, Doull J, Klaassen CD. 2007. Casarett and Doull's toxicology: The basic science of poisons. New York: McGraw-Hill Professional.

Clegg ED, Cook JC, Chapin RE, Foster PMD, Daston GP. 1997. Leydig cell hyperplasia and adenoma formation: Mechanisms and relevance to humans. Reprod Toxicol 11(1):107-121.

Cooper DS, Kieffer JD, Halpern R, Saxe V, Mover H, Maloof F, Ridgway EC. 1983. Propylthiouracil (PTU) pharmacology in the rat II. Effects of PTU on thyroid function. Endocrinology 113(3): 921-928.

Cox LA, Popken D, Marty MS, Rowlands JC, Patlewicz G, Goyak KO, Becker RA. 2014. Developing scientific confidence in HTS-derived prediction models: Lessons learned from an endocrine case study. Regul Toxicol Pharmacol 69(3):443-450.

Dellarco VL, McGregor D, Berry SC, Cohen SM, Boobis AR. 2006. Thiazopyr and thyroid disruption: case study within the context of the 2006 IPCS Human Relevance Framework for analysis of a cancer mode of action. Crit Rev Toxicol 36(10):793-801.

Dent MP, Carmichael PL, Jones KC, Martin FL. 2015. Towards a non-animal risk assessment for anti-androgenic effects in humans. Environ Int 83:94-106.

EFSA. 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA J. 11, 3290.

Everds NE, Snyder PW, Bailey KL, Bolon B, Creasy DM, Foley GL, Rosol TJ, Sellers T. 2013. Interpreting stress responses during routine toxicity studies: a review of the biology, impact and assessment. Toxicol Pathol 41(4):560-614.

Foster PMD. 2007. Induction of Leydig cell tumors by xenobiotics. In Payne AH, Hardy MP, eds, Contemporary Endocrinology: The Leydig Cell in Health and Disease. Humana Press Inc, Totowa, NJ, USA, pp 383-392.

Fraser HM, Lunn SF. 2000. Angiogenesis and its control in the female reproductive system. Br Med Bull 56(3):787-797.

Garcia de Mateos-Verchere J, Vaugoies J-M, Naudin B, Costentin N. 1998. Behavioural and neurochemical evidence that the antimicrobial agent oxolinic acid is a dopamine uptake inhibitor. Euro Neuropsychopharmacol 8:255-259.

Groh KJ, Dalkvist T, Piccapietra F, Behra R, Suter MJ-F, Schirmer K. 2014. Critical influence of chloride ions on silver ion-mediated acute toxicity. Nanotoxicology 14:1-11.

Groh KJ, Carvalho RN, Chipman JK, Denslow ND, Halder M, Murphy CA, Roelofs D, Rolaki A, Schirmer K, Watanabe KH. 2015. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere 120:764–777.

Groothuis FA, Heringa MB, Nicol B, Hermens JL, Blaauboer BJ, Kramer NI. 2015. Dose metric considerations in *in vitro* assays to improve quantitative *in vitro–in vivo* dose extrapolations. Toxicology 332:30-40.

Gülden M, Seibert H. 2005. *In vitro-in vivo* extrapolation of toxic potencies for hazard and risk assessment – problems and new developments. ALTEX 22(Special Issue 2):218-225.

Gunnarsson L-M, Jauhiainen A, Kristiansson E, Nerman O, Larsson DGJ. 2008. Evolutionary conservation of human drug targets in organisms used for environmental risk assessments. Environ Sci Technol 42(15):5807-5813.

Hecker M, Hollert, Cooper R, Vinggaard AM, Akahori Y, Murphy M, Nellemann C, Higley E, Newsted J, Laskey J, Buckalew A, Grund S, Maletz S, Giesy, J, Timm G. 2011. The OECD validation program of the H295R steroidogenesis assay: phase 3. Final inter-laboratory validation study. Environ Sci Pollut Res Int 18(3):503-515.

Hutchinson TH, Bögi C, Winter MJ, Owens JW. 2009. Benefits of the maximum tolerated dose (MTD) and maximum tolerated concentration (MTC) concept in aquatic toxicology. Aquat Toxicol 91(3):197-202.

ICCVAM. 1994. http://iccvam.niehs.nih.gov Validation of Alternative Methods.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf

Khera KS. 1985. Maternal toxicity: a possible etiological factor *in embryo*-fetal deaths and fetal malformations of rodent-rabbit species. Teratology 31(1):129-153.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25(1):1-5.

Marty MS, Johnson KA, Carney EW. 2003. Effect of feed restriction on Hershberger and pubertal male assay endpoints. Birth Defects Res B Dev Reprod Toxicol 68(4):363-374.

Maskos U. 2010. Role of endogenous acetylcholine in the control of the dopaminergic system via nicotinic receptors. J Neurochem 114(3): 641-646.

Miller DH, Jensen KM, Villeneuve DL, Kahl MD, Makynen EA, Durhan EJ, Ankley GT. 2007. Linkage of biochemical responses to population-level effects: a case study with vitellogenin in the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 26(3):521-527.

Moermond CT, Kase R, Korkaric M, Ågerstrand M. 2016. CRED: Criteria for reporting and evaluating ecotoxicity data. Environ Toxicol Chem 35(5):1297-1309.

Murakami M, Hosokawa S, Yamada T, Harawaka M, Ito M, Koyama Y, Kimura J, Yoshitake A, Yamada H. 1995. Species-specific mechanism in rat Leydig cell tumorigenesis by procymidone. Toxicol Appl Pharmacol 131(2):244-252.

Nyman A-M, Schirmer K, Ashauer R. 2014. Importance of Toxicokinetics for Interspecies Variation in Sensitivity to Chemicals. Environ Sci Technol 48(10):5946–5954.

OECD. 2005. OECD Series on Testing and Assessment Number 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organisation for Economic Co-operation and Development, Paris.
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14

OECD. 2015a. Report of the workshop on a framework for the development and use of integrated approaches to testing and assessment. Series on Testing and Assessment No. 215, ENV/JM/MONO(2015)22. Organisation for Economic Co-operation and Development, Paris, France.
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)22&doclanguage=en

OECD. 2015b. Series on Testing & Assessment No. 226. Integrated summary report: Validation of Two Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrERa). Organisation for Economic Co-operation and Development, Paris, France.

Patlewicz G, Simon TW, Rowlands JC, Budinsky RA, Becker RA. 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul Toxicol Pharmacol 71(3):463-477.

Phillips MB, Leonard JA, Grulke CM, Chang DT, Edwards SW, Brooks R, Goldsmith MR, El-Masri H, Tan YM. 2016. A Workflow to Investigate Exposure and Pharmacokinetic Influences on High-Throughput *in Vitro* Chemical Screening Based on Adverse Outcome Pathways. Environ Health Perspect 124(1): 53-60.

Prentice DE, Siegel RA, Donatsch P, Qureshi S, Ettline RA. 1992. Mesulergine induced Leydig cell tumours, a syndrome involving the pituitary-testicular axis of the rat. Arch Toxicol Suppl 15:197-204.

Prentice DE, Meikle AW. 1995. A review of drug-induced Leydig cell hyperplasia and neoplasia in the rat and some comparisons with man. Human Exp Toxicol 14:562-572.

Rasoulpour RJ, Terry C, LeBaron M, Stebbins K, Ellis-Hutchings RG, Billington R. 2014. Mode-of-action and human relevance framework analysis for rat Leydig cell tumors associated with sulfoxaflor. Crit Rev Toxicol 44(S2):25-44.

Roberts SA, Robison RL, Van Ryzin RJ, Stoll RE. 1993. Increased incidence of Leydig cell tumours in male rats with the dopamine agonist SDZ205-502. Toxicologist 13:292.

Ruden C, Adams J, Ågerstrand M, Brock TCM, Poulsen V, Schlekat CE, Wheeler JR and Henry TR. In press. Assessing the relevance of ecotoxicological studies for regulatory decision-making. Integrated Environmental Assessment and Management.

Schmolke A, Thorbek P, DeAngelis D, Grimm V. 2010. Ecological models supporting environmental decision making: a strategy for the future. Trends Ecol Evol 25(8): 479-486.

Scott, A.P., 2012. Do mollusks use vertebrate sex steroids as reproductive hormones? Part I: Critical appraisal of the evidence for the presence, biosynthesis and uptake of steroids. Steroids 77, 1450-1468.

Scott, A.P., 2013. Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical review of the evidence that steroids have biological effects. Steroids 78, 268-281.

Silva M, Pham N, Lewis C, Iyer S, Kwok E, Solomon G, Zeise L. 2015. A Comparison of ToxCast Test results with *in vivo* and other *in vitro* endpoints for neuro, endocrine, and developmental toxicities: A case study using endosulfan and methidathion. Birth Defects Res (Part B) 104:71-89.

Stadnicka-Michalak J, Tanneberger K, Schirmer K, Ashauer R. 2014. Measured and modeled toxicokinetics in cultured fish cells and application to *in vitro - in vivo* toxicity extrapolation. Plos One 9(3):e92303. doi: http://dx.doi.org/10.1371/journal.pone.0092303

Terry KK, Chatman LA, Foley GL, Kadyszewski E, Fleeman TL, Hurtt ME, Chapin RE. 2005. Effects of feed restriction on fertility in female rats. *Birth Defects Res B Dev* Reprod Toxicol 74(5):431-441.

US EPA. 2009. OPPTS 890.1100: Amphibian metamorphosis assay. US EPA October 2009.

Van Der Kraak GJ, Hosmer AJ, Hanson ML, Kloas W, Solomon KR. 2014. Effects of atrazine in fish, amphibians, and reptiles: an analysis based on quantitative weight of evidence. Crit Rev Toxicol 44(Suppl 5):1-66.

Ward JM, Peters JM, Perella CM, Gonzalez FJ. 1998. Receptor and nonreceptor-mediated organ-specific toxicity of di(2-ethylhexyl)phthalate (DEHP) in peroxisome proliferator-activated receptor alpha-null mice. Toxicol Pathol 26(2):240-246.

Weltje L. 2013. Techniques for Measuring Endocrine Disruption in Insects. In Matthiessen P, ed, Endocrine Disrupter Risk Assessment: Testing and Prediction Methods. John Wiley & Sons, Inc.

Wheeler JR and Coady K. 2016. Are all chemicals endocrine disrupters? Integr Enviro Assess Manage 12(2):397–406.

WHO/IPCS. 2002. Global assessment of the state-of-the-science of endocrine disruptors. http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/ (DoA: 18 May 2014).

Willett C, Caverly Rae J, Goyak KO, Minsavage G, Westmoreland C, Andersen M, Avigan M, Duché D, Harris G, Hartung T, Jaeschke H, Kleensang A, Landesmann B, Martos S, Matevia M, Toole C, Rowan A, Schultz T, Seed J, Senior J, Shah I, Subramanian K, Vinken M, Watkins P. 2014. Building Shared Experience to Advance Practical Application of Pathway-Based Toxicology: Liver Toxicity Mode-of-Action. ALTEX 31(4/14):500-519.

Yamada T, Nakamura J, Murakami M, Okuno Y, Hosokawa S, Matsuo M, Yamada H. 1994. The correlation of serum Luteinizing hormone levels with induction of Leydig cell tumours in rats by oxolinic acid. Toxicol Appl Pharmacol 129:146-154.

Yoon M, Campbell JL, Andersen ME, Clewell HJ. 2012. Quantitative *in vitro* to *in vivo* extrapolation of cell-based toxicity assay results. Crit Rev Toxicol 42(8):633-652.

APPENDIX A: AN EXAMPLE OF DOCUMENTING – ANALYTICAL VALIDATION OF THE FISH SHORT-TERM REPRODUCTION ASSAY

Endpoints evaluated:

Survival, Behaviour, Fecundity, Fertilisation Success, Hatchability (including larval appearance and survival, Appearance and Secondary Sex Characteristics, Gonad Histology, Biochemical Endpoints (collection and analysis of blood for quantification of sex steroids and vitellogenin.

Documentation of test method performance:

1. Validation of the Fish Short-Term Reproduction Assay: Integrated Summary Report (U.S. Environmental Protection Agency, Endocrine Disruptor Screening Program, Washington, DC December 15, 2007) http://www.epa.gov/endo/pubs/fish_assay_isr.pdf
2. OECD Revised Draft Report Phase 1B of the Validation of the 21-day Fish Assay for the Detection of Endocrine Active Substances (2005).

   http://www.epa.gov/endo/pubs/edmvac/revised_draft_oecd_fish_report_phase1b_disclaimer.pdf
3. Peer Review Results for the Fish Short-term Reproduction Assay (January 2008)

   http://www.epa.gov/endo/pubs/fishassay_peer_review_013008.pdf
4. EPA Response to Peer Review Comments (undated).

   http://www.epa.gov/endo/pubs/fish_peer_review_response.pdf
5. Endocrine Disruptor Screening Program Test Guidelines - OPPTS 890.1350: Fish Short-Term Reproduction Assay [EPA 740-C-09-007] (October 2009)

   http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0007

Conclusions reached: Sensitivity, specificity, reliability and domain of applicability:

- "Reproducibility of the assay: The reproducibility of the fish short-term reproduction assay, for screening purposes, has been broadly demonstrated using a number of representative endocrine-active chemicals across geographically diverse laboratories. In the inter-laboratory study with the optimised protocol, five chemicals were tested which generally demonstrate the reproducibility of the assay. In addition, a number of studies conducted prior to the inter-laboratory study have also shown reproducibility of the optimised assay."
- "The fish short-term reproduction assay as presented is intended to serve in a screening capacity to provide an indication of potential endocrine activity, not to confirm any specific mechanism, mode of action, or adverse effect. Therefore, any significant effect in one or more of the core endpoints of this assay (fecundity, histopathology, GSI, sex steroid measurements, vitellogenin, and secondary sex characteristics) should be considered a positive response in the Fish Short-term Reproduction assay, and supports further testing of the compound in the Tier 2 assays of the EDSP."

- "Specific Endpoint Limitations
  - Vitellogenin as endpoint: 1.) requires large blood volume relative to amount routinely able to collect; 2.) dependent in contract laboratories on available commercial kits which have no consistent validation or quality assurance standards which necessitates additional calibration in the individual laboratory.
  - Fecundity as endpoint: 1.) can be influenced by many chemical and non-chemical factors, 2.) requires pre-exposure monitoring to ensure suitable compliance.
  - Histopathology as endpoint: 1.) requires additional time and services of a qualified pathologist.
  - Secondary Sex Characteristics as endpoint: 1.) sensitivity and specificity are limited to certain modes of action; 2.) not all are as quantitative as some other endpoints; 3.) some alterations to physiology that manifest in morphological changes may not appear in the short duration of the fish assay.
  - Sex Steroids as an endpoint: 1.) requires radio-immuno assay for quantification which may be challenging for some commercial laboratories."
- "Specific Endpoint Strengths
  - Vitellogenin as an endpoint: 1) primarily controlled through oestrogen interaction with the oestrogen receptor, and hence is directly related to a mechanism of concern (Korte et al, 2000); 2) clear, unambiguous induction in male fish is well established as a response to oestrogen receptor agonists (Brodeur et al, 2006; Harries et al, 2000; Korte et al, 2000); 3) VTG may also respond secondarily to androgenic compounds through suppression of natural androgens and subsequent reduction in endogenous oestrogens; this mechanism may also manifest when fish are exposed to a steroidogenesis (e.g. aromatase) inhibitor due to the impaired ability to adequately produce endogenous oestrogens; 4) well-established endpoint and increasing commercial availability of ELISA kits that are specific to fathead minnow vitellogenin (Jensen et al, 2006).
  - Fecundity as an endpoint: 1) can be collected non-invasively with minimal effort and does not require additional animal use; 2) fertility data can be collected easily at the same time egg counts are made with minimal effort or time necessary, and they do not require technical expertise, allowing excellent transferability and inter-laboratory comparisons; 3) fecundity as an apical endpoint, when combined with gonadal histopathology, provides a good indicator of reproductive health of the fish as impaired fecundity is an adverse effect with regulatory importance whether it is due to endocrine-mediated activity or another mechanism of action (MOA); 4) fertility provides an indication of male reproductive function (sperm quality); 5) reduced fecundity has been the most consistently observed finding after exposure to diverse endocrine active substances, including all of the primary modes of action the assay is designed to detect: (anti-)oestrogens, (anti-)androgens, and modulators of steroidogenesis.
  - Gonad Histopathology as an endpoint: 1) sensitive indicator of endocrine dysfunction; 2) provides a direct evaluation of the reproductive organs of interest; 3) histopathologic changes express the integration of several molecular, cellular, and physiologic processes, 4) provides insight on the potential reproductive impacts of chemical disruption; 5) may decrease ambiguity when fish are exposed to chemicals with unknown modes of action, reconcile unexpected results from other endpoints and hence, may reduce the likelihood that assays must be run multiple times in such

instances; 6) ability to assess the general health of test populations, and the ability to identify causes of morbidity and mortality not associated with test compounds or reproductive endocrine activity, thus histopathological analysis can also help to reduce the number of false negative conclusions.

- Secondary Sex Characteristics as an endpoint: 1) biologically relevant, unique, robust and reproducible; 2) male secondary sex characteristics provide indicative androgenic / antiandrogenic effects that may not be observed with other endpoints; 3) inter-laboratory comparisons of secondary sex characteristics as endpoints have been relatively reproducible; 4) sensitive endpoint for androgen agonists which cause clear, unambiguous changes in secondary sex characteristics in females in the assay and trigger further testing.
- Sex Steroids as an endpoint: 1) provide additional supportive information that an endocrine mediated as opposed to non-endocrine mediated mode of action is occurring, which is especially valuable when decreased fecundity is also observed; 2) provide important insights into the specific mode of action."

References:

Brodeur JC, Woodburn KB, Zheng F, Bartels MJ, Kiecka GM. 2000. Plasma sampling and freezing procedures influence vitellogenin measurements by enzyme-linked immunoassay in the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 25(2):337-348.

Harries JE, Runnalls T, Hill EM Harris CA, Maddix S, Sumpter JP, Tyler CR. 2000. Development of a reproductive performance test for endocrine disrupting chemicals using pair-breeding fathead minnows (*Pimephales promelas*). Environ Sci Technol 34:3003-3011.

Jensen KM, Ankely GT. 2006. Evaluation of a commercial kit for measuring vitellogenin in the fathead minnow (*Pimephales promelas*). Ecotox. Environ. Safety 64(2):101-105.

Korte JJ, Kahl MD, Jensen KM, Pasha MS, Parks LG, LeBlanc GA, Ankley GT. 2000. Fathead minnow vitellogenin: Complementary DNA sequence and message RNA and protein expression after 17β-estradiol treatment. Environ Toxicol Chem 19:972-981.

MEMBERS OF THE TASK FORCE

W. Al-Husainy	Shell NL - The Hague
R. Becker	American Chemistry Council USA - Washington
M. Dent	Unilever UK - Bedford
S. Duhayon	Total B - Seneffe
U. Ensenbach	Clariant D - Sulzbach
I. Fegert	BASF D - Ludwigshafen
R. Green (Chairman)	Dow AgroSciences UK - Abingdon
S. Jacobi	Albemarle B - Louvain-La-Neuve
M. Léonard	L'Oréal F - Aulnay-sous-Bois
C. Palermo	ExxonMobil B - Machelen
K. Paul Friedman	Bayer USA - Research Triangle Park, NC
L. Weltje	BASF D - Limburgerhof
A. Weyers	Bayer D - Monheim
J. Wheeler	Dow AgroSciences UK - Abingdon
M. Galay Burgos *	ECETOC B - Brussels
A. Poole **	ECETOC B - Brussels

* Left ECETOC July 2016

** Took over Task Force management July 2016

MEMBERS OF THE SCIENTIFIC COMMITTEE

(Peer Review Committee)

B. van Ravenzwaay (Chairman) *	BASF
Senior Vice President - Experimental Toxicology	D - Ludwigshafen
R. Bars	Bayer CropScience
Team Leader, Toxicology Research	F - Sophia Antipolis
P. Boogaard	Shell Health
Senior Toxicologist	NL -The Hague
A. Flückiger	F. Hoffmann - La Roche
Chief Occupational Health Officer	CH - Basel
H. Greim	Technical University München
Institute of Toxicology and Environmental Hygiene	D – München
H. Hollnagel	Dow Europe
Regulatory Toxicologist	CH – Horgen
F. Lewis	Syngenta
Global Platform Lead	UK - Bracknell
G. Malinverno	Solvay
Global Governmental and Regulatory Affairs	I - Milano
L. Maltby	University of Sheffield
Professor of Environmental Biology	UK - Sheffield
S. Marshall	Unilever SEAC
Environmental Science Leader	UK - Bedford
M.L. Meisters	DuPont de Nemours
Manager Health and Environmental Sciences EMEA	B – Mechelen
M. Pemberton	Systox
Director	UK – Wilmslow
C. Rodriguez	Procter and Gamble
Principal Toxicologist, Corporate Central Product Safety	B - Strombeek-Bever

* Responsible for primary peer review

MEMBERS OF THE SCIENTIFIC COMMITTEE (cont'd)

D. Salvito **	RIFM
Vice President, Environmental Sciences	USA - Woodcliff Lake, NJ
J. Tolls	Henkel
Director Environmental Safety Assessment	D - Düsseldorf
S. van der Vies	VU University Medical Center
Professor of Biochemistry	NL - Amsterdam
C.J. van Leeuwen	KWR Watercycle Research Institute
Principal Scientist	NL - Nieuwegein
R. Zaleski	ExxonMobil
Exposure Sciences Section Head	USA - Annandale, NJ

** Resigned from the SC in June 2016

ECETOC PUBLISHED REPORTS

The full catalogue of ECETOC publications can be found on the ECETOC website:
http://www.ecetoc.org/publications

Responsible Editor:
Dr Alan Poole
ECETOC AISBL
Av. E. Van Nieuwenhuyse 2 (box. 8)
B-1160 Brussels, Belgium
VAT: BE 0418344469
www.ecetoc.org
D-2016-3001-247

Since 1978 ECETOC, an Industry-funded, scientific, not-for-profit think tank, strives to enhance the quality and reliability of science-based chemical risk assessment. Learn more at http://www.ecetoc.org/