Testing Strategies for Endocrine Disruption in the Aquatic Environment

Daniel B. Pickford

AstraZeneca Global Safety Health and Environment

Thank you, Mr. Chairman. Firstly, I have to give a slight correction. I am not really an expert on fish testing, but maybe more of an expert on amphibian testing. I hope to cover some aspects on both of these topics today.

Firstly, I should point out that as we are working within the general definition of an endocrine disrupter set out in Weybridge in 1996: "That which is causing adverse effects in an intact organism." This places an emphasis on *in vivo* testing as we have heard very well from the previous two speakers.

However, this should be backed up with mechanistic information from *in vitro* data, and this can be used in prioritization, and I think we have already heard very elegant description of that process.

Finally, depending on the type of test we are doing, this may include in wildlife species as well as in mammals, use of reproductive, developmental, or behavioral endpoints to detect endocrine disrupters.

Before I show you some brief strategies of ecological risk assessment for aquatic vertebrates, maybe I just need to give a quick introduction of why we need these other tests in addition to mammalian tests.

Fish are not just rats in water. We need to consider that while there is obviously evolutionary conservation of the endocrine system in vertebrates, there are differences in both the roles of hormones and their regulatory interactions. For example, prolactin is principally involved in lactation in mammals whereas in fish it is involved in osmoregulation and in amphibians it is a key regulator of metamorphosis.

Furthermore, taking Amphibia as an example, there are different regulator interactions. So in Anuran amphibians — that is frogs and toads — there is an interaction between the thyroid axis and sex steroids. This may have implications for the utility of vitellogenin as an estrogenic biomarker in amphibian tests: amphibian larvae do not respond to estrogen with vitellogenin synthesis until metamorphic climax, after there has been a cross talk between thyroid receptor and estrogen receptor.

Furthermore, there may be differences in metabolism between mammals and aquatic lower invertebrates, and this may have an effect on the relative importance of systemic toxicity vs. endocrine toxicity. An example of that would be methoxychlor, which we have just heard about. Hydroxylated metabolites are responsible for its estrogenic effects *in vivo*. However, in fish, where there seems to be a limited ability to metabolize this compound, we see a truncated estrogenic response, where at concentrations which would not be toxic in mammals we see systemic toxicity in fish.

There may also be differences in exposure route or the influence of exposure route between oral exposure in mammals versus dermal or gill exposure in fish, and this may influence the bioavailability of compounds, which are endocrine disrupters potentially. This may be further complicated by effects of first-pass metabolism.

Finally, there may be species differences in the affinity of steroid receptors for xenobiotics, and a good example is vinclozolin metabolites, which seem to have different affinities for the androgen receptor in fish and rats.

Now, I would like to provide a brief overview of an industry perspective on how we might go about testing for endocrine disrupters in the process of ecological risk assessment. This is only my interpretation, do not hold me to it. This has been proposed by ECETOC, the European Center for Ecotoxicology and Toxicology of Chemicals, and this was published last year by Tom Hutchinson *et al*.

In this program, as we have already heard, the idea is to use SAR data and *in vitro* data for priority setting. Maybe in aquatic species as we are talking about, we might also incorporate information from mammalian testing that has already been done. Obviously, this is all combined with information on pattern of use and fate data to determine realistic potential for exposure.

Then, we would enter a tiered testing system of *in vivo* tests. To use the fish situation as an example, ECETOC are actually promoting a three-tier system that differs slightly from maybe the U.S. approach or even the OECD approach, but I think really it is compatible. As we have already heard from Dr. Koëter, we have got a Tier 1 screening type test, a short-term test that is really to provide mechanistically specific information on endocrine activities.

So is a compound estrogenic or androgenic in fish? If we can confirm that in this screening test, we may then go onto Tier 2 and Tier 3, which respectively give increased levels of effects characterization using more apical endpoints. At the end of all this we can incorporate exposure data and our hazard data for the ecological risk assessment.

We have already had an introduction to what those methods might be in fish. The juvenile fish screening assay should be for detecting estrogens, or that is one of the aims, and the idea is that we may need an oviparous model, a lower vertebrate egg-layer to test for estrogens as well as a mammal, to reduce false negatives.

One of the candidate tests for this is a rainbow trout assay. A lot of work on this has been done by a colleague of mine at Brixham, Karen Thorpe. It is an adaptation of existing OECD test guidelines, and it involves a 14 day aqueous exposure at 15°C. The endpoints, as we heard, are survival, growth, hepato-somatic index, and plasma vitellogenin as an estrogenic biomarker. As we have already heard, this is a concept which we could apply to other OECD test species.

This is some of the data from Karen's validation of this test approach. Looking at exposure of juvenile rainbow trout to estradiol, nonylphenol, and methoxychlor. As we can see, firstly on the x-axes there are different scales, reflecting the different potencies of these chemicals, but the picture is the same in each one.

There is a nice concentration-response relationship, but it seems that we do not get much increased sensitivity by extending the assay from 14 days. You can see the response is quite similar for 14 and 21 days. The final thing to point out briefly, as I mentioned, is the effect of systemic toxicity vs. endocrine toxicity in fish models. Here we see a truncation of the estrogenic response in trout because of systemic toxicity of methoxychlor, as I mentioned earlier.

Moving up to Tier 2: we have some partial life-cycle tests, and there could well be two of these. One of them is an embryo-larval development test, which is again an enhanced existing OECD guideline. This includes endpoints of growth and development which are apical, plus whole body vitellogenin concentrations, and also possible effects on gonadal histology.

The reproduction test: there are a few candidates, one of which we are working on in Brixham. Grace Panter and Karen Thorpe are looking at 42 day pair breeding assay using fathead minnow. This incorporates a three-week pre-acclimatization period before exposure, followed by three weeks exposure. They are looking at gross morphology, secondary sex characteristics such as the fat pad in the male fathead minnow (which may be androgen sensitive), plasma VTG, and again fecundity and gamete quality.

Those tier 2 tests are possibly shorter-term tests, which could be used to give some level of effects characterization without going to the length and costs of a full lifecycle test. Obviously, a full lifecycle is the gold standard for ecological risk assessment of endocrine disrupters, and we have obviously got well-established guidelines for lifecycles.

But these can be optimized for endocrine disrupters by bolting on relevant endpoints such as plasma vitellogenin, and maybe even in alternative species, something like spiggin as an androgenic biomarker, and also gonadal histology.

The idea is that hopefully in this kind of test we can compare apical data, such as reproduction, with more specific biomarkers; this will improve the predictive value of the biomarkers. Then, hopefully beyond that, we can take these validated biomarkers and use them as linkage tools between lab and field studies so we get a lab-field iteration.

I will now briefly move onto amphibian testing. Activity in this area has increased a lot over the last couple of years. This is sort of marked by the OECD commission of the first expert consultation on endocrine disrupter testing in Amphibia, which was in Paris in April of this year.

I think I might say the arena of this will be twofold: primarily the development of a screen for thyroid activity, I will talk more about that in a minute, and the secondary objective — possibly further down the road — is development of a higher tier test similar to the sort of thing we have in fish.

This really follows on from the U.S.A. where the EDSTAC process recommended a tail absorption assay for thyroid activity in their Tier 1 screen, and opposite this idea, in Germany we have development of a *Xenopus* metamorphosis assay, again to look at thyroid activity. The XEMA test has been developed by Professor Kloas from Berlin, who is here, and is currently in the ring testing stage, funded by the German Environment Agency (UBA). Our lab in Brixham is also involved in this process.

To place these potential amphibian tests in their life-cycle context, typically anuran amphibians - frogs and toads - have a bi-phasic life history: we have eggs going through embryonic development, hatching as larvae which are fully aquatic, and these eventually metamorphose into juveniles. Typically, juveniles in the adult phase are more terrestrial. Gonad maturation then proceeds in the adult phase when we can get gamete production and start a new lifecycle.

In the current context, the most important aspect of this lifecycle is the pivotal role of thyroid hormone in triggering metamorphosis in frogs and toads. This really lends metamorphosis potentially a great deal of utility as a screen for thyroid activity since we can use a metamorphosis assay.

At the higher tiered testing level we may have the reproduction-development test, which could be quite possibly a partial reproduction test. Maybe we could start with adult exposure and look at gamete production and through to metamorphosis. But in an appropriate species, we can also extend this to a life-cycle test, but this would require thinking about the species we use. I will mention this briefly later.

To try and put these tests into the context of the strategy I outlined earlier for fish testing with the three tiers, where would the amphibian metamorphosis assay sit? Ideally, it should be a tier one screen for thyroid activity. I have put it here floating between the two, because I think in its current state we need to address the issue of specificity of this test: whether it can really be a specific test for detecting thyroid activity *per se*, or whether it is more of an effects characterization. I will hopefully explain why that is.

As Dr. Koëter pointed out, *Xenopus* is probably the best option for this test. It is well established and the larval development of this species is well characterized. We need to think about the exposure window because this can have an impact on survival, sensitivity of the test and variability.

As an example of that, in the EDSTAC process the recommendation was a tail resorption assay. But if we look at amphibian larval development, this is an arbitrary graph, with increasing development the tadpole is growing, evident as increased tail length. Then, during metamorphic climax the tail is resorbed, and this is the basis for the tail resorption assay.

But this is happening in the context of very high concentrations of endogenous thyroid hormone, which may mean that this is not a very sensitive window because it would be hard to show agonist or antagonist effects. So it may be better to move exposure earlier when larvae are sensitive to thyroid hormone, but they are not secreting much themselves. This is the basis of the German protocol.

Briefly, I want to mention one thing that may be a spanner in the works. What we need to think about is that if we are looking at growth and development as the endpoint for this metamorphosis test, these are apical endpoints. They can integrate multiple influences on the thyroid axis.

We have known for a long time in amphibian endocrinology that estrogens can influence larval development. We have looked at in our laboratory in Brixham the effects of 17 α -ethinylestradiol on larval development in two species in *Xenopus laevis*. We exposed them to 17 α -ethinylestradiol at 10 to 1000 ng/l, and we found that it caused a concentration dependent increase in development time, measured as the number of days it took the larvae to complete metamorphosis. And there is a significant delay at 16 ng/l.

In contrast, we also looked in the same concentration range, the same exposure system, in *Rana temporaria*, a native frog. The endpoint we have used here is forelimb emergence: it is a nice clear developmental endpoint. In this case, we see at a higher dose again an inhibition of development. But at lower concentrations like at 16 ng/l, we actually had what appeared to be an acceleration of development although this may be confounded by solvent effects. Here we have the dilution water control, and there was an acceleration in solvent solvent control. These data flag up the possibility that we need to be careful about extrapolating from *Xenopus* to other native species.

To try and summarize some issues on the amphibian metamorphosis assay, we need to think about the possibility of false positives due to extra-thyroidal influences such as from estrogens or even from systemic toxicants. We have to consider also there are multiple points at which we can disrupt the thyroid axis. We can disrupt synthesis, transport, clearance, and action at the target site.

So we need maybe to look at adding biochemical and molecular biomarkers so we can define the mechanism underlying any apical effect. Then we can say this is influencing development, and we know because of this biomarker that it is causing a thyroid active effect as opposed to something else.

For that reason, at the moment I think that in the absence of these biomarkers, we need to discuss whether it is a screen or a test. That maybe semantics, but we need to think about it. Maybe also we should think about co-validation of this test and sharing compounds with development of potential thyroid endpoints in mammalian assays.

Briefly, I will just try and finish up with thinking about the amphibian reproduction-development test. Again, I have put it floating here because depending on whether it is a partial or a full lifecycle, it could be in either of the tiers that ECETOC are promoting.

This is really, I think, still a conceptual test. The exposure window and the endpoints that we use are going to be driven probably by information by lower tiers and whether we see an effect on metamorphosis or whether we see an effect maybe in another test of estrogenic activity.

Choice of species, we need to consider generation time because this can have implications for the test duration and the cost.

The exposure route, many amphibians are terrestrial in the adult phase, as I pointed out, and therefore, if we are looking at lifecycle, we need to think about how we continue exposure from the larval to the adult phase, and this may take some work.

I think we can probably agree that *Xenopus* — in the OECD meeting we agreed — that *Xenopus* are probably not ecologically representative for many parts of the world as a test species for a higher tier test. It is obviously ecologically representative in South Africa, but we need to look at whether we can use this as a surrogate for other native species. Given the data I already showed, I think that that requires some discussion. Obviously, because of that we need to validate appropriate models.

In the interest of time, I am going to skip this slide because it may be inaccurate, anyway. It is fair to say frogs and invertebrates are probably lagging behind other species.

In summary, I would like to point out, well, hopefully I have shown that with a multi-tiered testing approach in aquatic wildlife for ecotoxicology, it would be nice to combine lower tier tests, which provide mechanistic information through use of specific biomarkers, with higher tier tests that give us apical data from apical endpoints, which we may be able to use to predict effects at the population level.

I will just finish up with thanks to my colleagues Tom Hutchinson, Karen Thorpe, and Grace Panter working with fish work, and to Nadine Pounds who has helped with just about everything. Finally, thank you very much to the Ministry of Environment and the organizers for inviting me to Japan to talk at this very important symposium. Thank you.