4. General test considerations

4.1 Concentration setting

Test concentration setting is an important part of the study design. Effective choice of concentrations ensures a successful test, minimizing the need for repeat runs, with consequent benefits in terms of animal welfare, time and cost. Key to the strategy employed is the purpose of the test in terms of quantitative endpoint determination (e.g. NOEC, LC_{50}) or qualitative hazard determination (e.g. endocrine screens to demonstrate or exclude *in vivo* endocrine activity). For quantitative tests, concentration selection is optimized to achieve the point estimate. However, for qualitative hazard identification, test concentration selection is driven by the need to test in the "concentration space" most likely to observe the hazard whilst not confounding the interpretation (e.g. inducing other unwanted effects that may result from systemic toxicity). This difference is acknowledged in the existing test guidelines by recommendations for test concentration spacing factors. For quantitative tests, the spacing factor typically should not exceed 2.2 to 3.2 (OECD TGs 203 and 210, respectively). The scientific rational, justifying such an approach for test concentration spacing, is given in Doudoroff et al. (1951) and was further supported by the work of Sprague (1969). For the fish endocrine screens, a factor not exceeding 10 is recommended (OECD TGs 229 and 230), i.e. a larger concentration span.

An exception to the strategies employed above is the recently published guidance on the use of the threshold approach for fish acute toxicity testing (OECD 2010). Essentially, the approach uses a limit test at a single threshold concentration determined by the results of *Daphnia* and algae tests. If no mortality is observed in the limit test, the fish acute value can be expressed as greater than the threshold value. However, if mortality is observed, a full concentration response test is triggered. This is discussed further in chapter XX.

Maximum test concentrations may be specified by the regulatory body. For example, for OECD tests it is 100 mg/L for acute tests (e.g. OECD TG 203) and 10 mg/L for chronic and endocrine tests (e.g. OECD TGs 210 and 229). However, other regulatory bodies may specify other maximum test concentrations, for example the maximum test concentration is 1000 mg/L for industrial chemicals (OPPTS 850.1075, U.S. EPA 1996)

Range finding tests, to inform definitive test concentration selection, are an important tool. However, the need for such confirmatory data should be weighed against the existing data on the test item or related substances. Existing data may negate the need for range-finding, since the effect range can be reasonably predicted. For example, effect levels can often be reliably predicted for formulated plant protection or biocidal products from existing data on the component active ingredient(s) that drive the overall toxicity. Mammalian and non-mammalian data for pharmaceutical studies may be useful, as could be read across from similar compounds (ECETOC 2007) for setting test concentrations. However, where such data are not available, it is often in the interests of animal reduction to run a well-designed range-finding test to avoid the need to repeat definitive tests that fail to capture the relevant endpoint(s).

Best practise for range-finding experiments is difficult to describe, as it is substance- and existing information-specific. Ideally, fish used in range-finding experiments should be as similar to the definitive test organisms as possible in terms of size and age (preferably from the same batch or source). For fish acute toxicity tests, typically three concentrations (no control) with a wide spacing factor of 10, when practical, and three fish per treatment offers sufficient information to set bounds around the approximate position of the LC_{50}. The rational for this type of approach is given in Hutchinson et al. (2003). For chronic tests, range-finding is more difficult because of the longer duration, multiple endpoints and, due to reduced replication, differences in statistical power compared to definitive test. However, in general it is not necessary to range-find to the full duration of the definitive test, but only sufficient time to assess the relevant parameters. For example, for the fish early life-stage test (OECD TG 210), a duration of 14 days (ca. 9 days post-hatch in fathead minnow (*Pimephales promelas*) or 10 – 11 days post-hatch in zebrafish (*Danio rerio*)) usually allows for the fish to have grown sufficiently for estimation of growth effects. Again, a large spacing factor of 10 is useful. In general, range-finding should also be conducted in similar conditions to the definitive test in relation to exposure (static, semi-static, flow-through conditions). This is particularly important for unstable, volatile and high octanol-water partition coefficient substances. Large differences in exposure between range-finding and definitive tests may occur due to differences in test systems and fish loading resulting in range-finding that is not predictive.

For the endocrine screens, the purpose of range-finding is to ensure that the definitive test exposes fish to a suitable range of concentrations maximising the likelihood of observing the effect. The important distinction being that achieving a NOEC is not the purpose of the screening test, but rather to inform decision making for further testing investigations. Therefore, test concentration selection becomes a trade-off between testing sufficiently high to find the effect (if present) whilst not confounding the results by inducing systemic toxicity. To this end, the Maximum Tolerable Concentration (MTC), as defined by Hutchinson et al. (2009), is cited in the test guidelines OECD TGs 229 and 230. The MTC is the highest concentration that does not lead to a reduction in survival, feeding, normal behaviour and normal morphology and colour. Therefore, professional judgement is required to analyse all the available data to determine if the MTC can be estimated without range-finding. Here it is important to note that predicting the MTC from acute lethality data is problematic, since the screens are effectively long-term exposures (21 days) assessing sublethal effects. Therefore, chronic data are important in determining a suitable MTC. Furthermore, when range-finding is required, it is recommended that durations longer than acute tests be used. The test should assess indicators of systemic toxicity (mortality and symptoms of toxicity). These can then be used to estimate the MTC to be used as the highest definitive test concentration.

In summary, the purpose of the test needs to be considered in order to decide on an appropriate concentration range. In line with animal welfare considerations, where possible, existing information alone should be used to determine definitive test concentrations. However, it is often justifiable to conduct a suitable range-finding experiment to inform test concentration selection (leading to a higher test success rate and overall reduction in animal numbers).

4.2 Preparation of test solutions, including solvent-free methods

The preparation of test solutions is an important part of the experimental design. The physicochemical properties of the test item can make testing in aqueous media difficult. The Guidance document on aquatic toxicity testing of difficult substances and mixtures (OECD 2000) provides useful guidance for such substances. This includes recommended organic solvents and solvent-free preparation methods. The preference is always to present the test item in the form it is most likely to occur in the environment, which for single chemicals is generally as the dissolved form, although preparations (pesticide formulations and other mixtures) may be tested as homogenous emulsions or suspensions. However, the physicochemical properties of the substance and necessary delivery options (e.g. flow-through systems) often mean it is not possible or practical to simply dissolve the substance in the test media. The objective is to achieve a biological (toxicological) response with testing up to the practical solubility limit in test media. It should be acknowledged that it is not always possible to achieve a biological response below the practical solubility limit and so a limit test at this concentration is often the best approach.

The most common practise to aid dissolution is the use of a solvent. However, this can be problematic in itself as it may potentially alter the bioavailability of the test substance and/or influence the test system (additional carbon source leading to microbial growth). As required by the individual test guidelines, a solvent control group should always be included with as many replicates as the water control group. Further, a recent review observed evidence that some low concentrations of solvents may affect the reproduction of certain fish species, and also impact biomarkers of endocrine disruption (Hutchinson et al. 2006). Therefore, where ever practically feasible, the use of solvents should be avoided.

Guidance for methods of solvent-free preparation has been described (Rufli et al. 1998, OECD 2000) and include generator columns, coating of stock solution vessels, sonication, and large volume (typically dilute) saturated aqueous stock solutions. Note, some of the non-solvent methods may result in the formation of micelles of the test substance, and this should be considered in test stock preparation and chemical analysis. However, it must be acknowledged that there are limitations to these methods, notwithstanding the considerable increase in time and cost associated with implementation in commercial and necessarily high throughput ecotoxicology laboratories. Difficulties also arise depending on the duration of the test and the supply of stock solutions, for example, to a flow-through delivery system. With these durations and at concentrations required, it may be difficult to ensure the stock concentration is maintained over the duration of the test. Ultimately, if not controlled, this may lead to unacceptable variability in test solution concentrations. Since maintaining acceptable variability in exposure solutions is a validity criterion (e.g. ± 20 % of the mean measured values; OECD TG 210), this has major implications for a laboratory's ability to conduct a valid study.

Therefore, there is a place for the use of solvents; however, it is sensible considering the issues described above to minimize the solvent concentration as is recommended in the current fish endocrine screening assays (OECD TGs 229 and 230). Hutchinson et al. (2006) recommended where solvent use is necessary, in reproduction studies with aquatic organisms, the

maximum solvent concentration should not exceed 20 µL/L of dilution water (0.002 %). This recommendation is a good target maximum concentration, although ultimately this may depend on the physciochemical properties of the test substance, not least its solubility in solvents. However, where potential solvent effects are suspected, the potential influence of the solvent on the test results should be discussed (e.g. enhanced growth in the solvent controls).

In summary, where ever possible, the use of a solvent should be avoided and alternative preparation techniques be employed. In cases where a solvent is required the concentration should be minimized, as far as practically possible.

4.3 Acclimation/culture maintenance/pre-treatment

The quality of test organisms is key to the successful conduct of fish tests. To ensure quality and confirm that the test organisms have adapted to laboratory conditions, the fish acute test (OECD TG 203) recommends a minimum acclimation period and batch selection criteria based on mortality and signs of disease during the acclimation period. In general, this is a practical method to ensure suitability for testing. For the longer-term studies, acclimation may not be possible (e.g. the fish early life-stage test starting with newly fertilised eggs). Here the preferred option is for embryos to be derived from in-house cultures of breeding fish where quality controls, such as not using a breeding group's first spawn (typically low viability) and disease status can be assured. However, it is acknowledged that this is not always possible. For example, it is not always practical to culture in-house all the required fish species (e.g. bluegill sunfish (*Lepomis macrochirus*)). Therefore, external suppliers such as commercial fish breeders are often necessary sources of test organisms. It is recommended that organisms supplied in this manner are accompanied by documentation from the supplier outlining basic information on source, occurrence of any treatments, time of collection (particularly for embryos) and any other pertinent information. This provides some safeguard against poor practise and establishes time lines for approximate time post-fertilisation, as required by the fish early life-stage test guideline (OECD TG 210).

For tests requiring reproductively active fish (OECD 229 and 230), it is advisable to have a prolonged acclimatisation period. The time to a particular developmental stage of fish will differ, because of variability in certain parameters during their culture (e.g. feeding, temperature, density, etc.). Therefore, if it is not possible to culture the animals entirely in-house, a prolonged acclimation period will ensure they are fully adapted to laboratory conditions and are more likely to be at the appropriate developmental stage required in a test (e.g. actively spawning).

For some species, commercial sources may not be available, in which case the field collection of animals is required. In these cases, characterisation of the organisms and the site from which they are collected should be undertaken. Characterisation should include an assessment of the contamination history of the collection site, evidence that the animals are derived from a viable population (i.e. reproducing) and their parasite load. Once in the laboratory, acclimation of the population to laboratory conditions should include mortality, disease and stress

assessment. Ideally, if the fish are to be used for endocrine screening (e.g. the androgenised stickleback assay), successful reproduction under culture/acclimation conditions is preferable before use in a test.

In summary, it is preferable for test organisms to be cultured in the testing laboratory. However, for certain test types and species this is not always practical. Information should be supplied with the batch of fish concerning their history. Strain should be included (e.g. OECD TG 210) where feasible; however, this may not always be possible for certain species. For the endocrine screening methods, longer acclimation periods may be necessary to ensure the fish are at the required developmental age/stage. Field collected species should undergo a full characterisation and acclimation assessment.

4.4 Species selection

Species selection considers a number of different factors including, size, ease of maintenance in the laboratory, convenience for testing, relevant economic, biological or ecological factors, known sensitivity, pre-existing data, animal welfare, availability of test methods for subsequent tests that may be triggered, as well as national or regional preferences. There are also practical considerations, such the availability of cultured, as opposed, to field-collected organisms (see section 4.3 for considerations on acclimation/culture maintenance/pre-treatment). However, field-collected animals may be more appropriate for site-/situation-specific risk assessment questions. It is not always possible to meet all of these requirements within one test. However, species selection should always consider these factors, so further testing with additional species is less likely to be required.

In terms of acute toxicity, rainbow trout (*Oncorhynchus mykiss*) is considered to be amongst the most sensitive species. Dyer et al. (1997) reviewed sensitivity differences between tropical, temperate and coldwater species and found the latter consistently more sensitive for a diverse set of chemicals. This has also been established using acute toxicity Interspecies Correlation Estimation, a program developed by the USEPA, which indicate rainbow trout as more sensitive than fathead minnow, but often only at a factor of 2 to 3 (Dyer et al. 2006); see also chapter 5 on fish welfare in this document). Similarly, Lammer et al. (2009) summarized acute inter-species toxicity comparisons for zebrafish, medaka (*Oryzias latipes*), bluegill, fathead minnow and rainbow trout to 30 - 80 chemicals, depending on the species pair. Results were highly similar to those of Dyer et al. (1997, 2006). However, for fish early life-stage tests, smaller warm-water species (e.g. fathead minnow, medaka and zebrafish) are preferred, due to the shorter duration of the test, compared to the rainbow trout study (ca. 30 days *versus* ca. 90 days).

However, longer tests with rainbow trout may be sensible, when there are historical data on a class of compounds for which it advantageous to read across endpoints in the same species. For other tests species, preference may be driven by the endpoint of concern. For instance, rainbow trout is preferred in the fish juvenile growth test (OECD TG 215), since relative growth in the exponential phase is greater than in other species making differences easier to

detect. For endocrine-specific testing, there are clear advantages of using the same species throughout general toxicity testing, endocrine screening and definitive endocrine testing. Such an approach could reduce the need for range-finding between test levels, i.e. general toxicity tests (e.g. fish early life-stage tests) could be used to set the MTC (see section 4.1 – concentration setting) for the fish endocrine screens (OECD TGs 229 and 230), all of which would inform concentration setting for definitive endocrine tests (tests currently under consideration are the fish sexual development test (FSDT), the fish full life-cycle (FFLC) and fish multi-generation tests).

Principles for selection of test species for chronic fish testing (e.g. OECD TG 210) would appear to be applicable to FFLC testing (USEPA 1986, Benoit 1981). There should also be a consideration of whether the desired/necessary endpoints can be measured, or at least easily be measured in the species chosen. For example, if secondary sexual characters are a critical endpoint in either of endocrine screens (OECD TGs 229 and 230), the species selected for testing should be medaka or fathead minnow rather than zebrafish. However, for other endpoints, fecundity measurements (fractional *versus* continuous spawners), egg or larvae success and body size, making blood sampling easier, may also be considerations. Similarly, animal reductions needs may be addressed by choosing species for optimal for a particular endpoint. For example, the fish sexual development test with a species that can be genetically sexed (e.g. medaka), as opposed to one that cannot, would allow for the use of fewer animals.

In summary, there are a number of factors driving species selection, and it is not always possible to satisfy these within one single test species. There is also value in keeping some flexibility in test species choice, for example to allow for freshwater/estuarine testing or the use of a fish test species in which genetic sex markers may have been recently developed, etc. Coherent principles for long-term studies, but flexibility in test species should be considered. However, there are clear advantages (where possible) in terms of animal reduction for using the same species at higher testing tiers.

4.5 Chemical analysis

An appropriate validated test method should be available before the initiation of the definitive test. The analytical method should cover the anticipated test concentration range. The validation should include an assessment of the recovery from test media (i.e. from spiked samples) and determination of the limit of quantification. The purpose of the analysis is to confirm exposure, and, as a minimum, analytical samples should be taken at the beginning and at the end of the exposure period from all treatments and control(s). Where appropriate, for longer-term studies, samples should be taken at weekly intervals. Additional samples should be taken from the test system or stock solutions, at the discretion of the study director, to investigate the impact of any failures to the test system (failure of dosing systems etc.). In general, it is recommended to take analytical samples from all replicates at the start and weekly intervals thereafter (OECD TG 229). However, for well-understood compounds, it may be scientifically justifiable to measure concentrations in fewer replicates at every sampling interval.

Replicates should be alternated, unless otherwise stated in the test guideline (e.g. OECD TG 229); so samples are not taken from only one replicate throughout the study. Samples should be taken at the approximate midpoint of the test vessel. At each sampling interval, it can be useful to take two samples; one for analysis and one for storage as a back-up. When poorly soluble materials are tested, the samples should be centrifuged or filtered prior to analysis and the supernatant analysed to determine the concentration of the test substance in solution, as this is presumed to be that which is biologically available. It should be noted that some of the non-solvent methods may result in the formation of micelles of the test substance and centrifugation is highly recommended prior to chemical analysis.

Ideally samples should be analysed immediately. However, often this is not practical and the samples must be stored (e.g. refrigerated or frozen depending on the test item) until they can be analysed. If stored, storage stability should be confirmed (i.e. prepared storage stability spikes in test media). Where solid phase or liquid extraction of the sample is required, this can be conducted before storage, as this will often enhance the stability of the sample. The back-up sample can be analysed, if necessary, to confirm any results outside expectation. The backup sample can be especially useful for flow-through studies, where the dynamic nature of test system means it would otherwise be difficult to investigate erroneous results. For flow-through studies, it is advisable to conduct pre-exposure analyses to ensure the test system is in equilibrium and test concentrations are approximately in the expected range, before adding test organisms. In conjunction with the chemical analysis, prior to the start of any flow-through study, the dosing system to be used should also be checked to confirm correct delivery of the test solutions.

In summary, suitable validated test methods should be available before initiation of the definitive test. Sampling should be at the beginning, end and at regular intervals during the exposure. However, the various regulatory authorities may have different analytical requirements.

4.6 Water and diet quality

Water and diet should be of sufficient quality to support normal test organism growth and development. This can be demonstrated by the culture of fish in the medium used for the test. The ability to culture test organisms through a life-cycle provides definitive evidence for appropriate water quality and culture conditions. However, for certain test types (e.g. OECD TG 203), this level of evidence is not required, as long as acclimation and test validity criteria pertaining to the biological quality of the organisms are met. Further, confirmation can be provided by the periodic chemical analysis of water (and sometimes food) for substances that may be present at levels considered to be toxic. The ASTM (2002) guidance can be consulted as a point of reference for this determination. There are also generally applicable criteria for water quality parameters in a recent OECD test guideline (OECD TG 229). Similarly, the nutritional value of the diet should be considered.

The clear advantage of commercial (formulated pellet or flake foods) is that nutritional quality is known. However, it can often be useful to provide additional sources such as frozen adult or brine shrimp (*Artemia* sp.) nauplii or other live food (*Daphnia*, *Chironomus* spp., etc.), particularly where fish reproduction (in culture or testing) is important. For fish endocrine tests (screening and definitive), the presence of potentially endocrine-active contaminants or food components (e.g. high phytoestrogen content) should be avoided.

In summary, there are also generally applicable criteria for water quality parameters in the ASTM guidance document (ASTM 2002) and recent OECD test guidelines (e.g. OECD TG 229). It may also be important to consider the nutritional value and presence of contaminants in the diet.

4.7 Test acceptability criteria

The current test acceptance or validity criteria, specified in the fish test guidelines, include control mortality, dissolved oxygen concentrations and water temperature. Longer-term studies also include variability around analytical measurements and biological criteria (fertilisation success or spawning activity). These are important criteria and should be used in the assessment of data quality and the decision over whether to repeat a test. However, there should be some latitude and professional judgement to assess the likely impact of deviations from these requirements. Typically, up to 10 % mortality (or 1 out of 7 in the fish acute OECD TG 203) or species-dependent hatch and post-hatch mortality in early life-stage test (e.g. OECD TG 210) is allowed in the control group(s).

The level of control performance may affect the power of the test (see Chapter 3 on statistical considerations). In general, this should be considered an important criterion, since it pertains to the quality of the test organisms and factors of the test system that may cause significant stress sufficient to impact the reliability of the results determined.

The criterion related to dissolved oxygen ensures suitable conditions for the fish. However, for longer-term studies, judgement should be made as to the duration and magnitude of observations below the dissolved oxygen criterion of 60 % air saturation. For instance, the rapid instigation of aeration to bring the levels above guideline requirements may be acceptable, if the duration and magnitude were unlikely to have adversely impacted the results of the test.

Similarly, the requirements for water temperature between vessels at any one time (± 1.5 °C) can be challenging to meet, and again judgement can be applied to assess the potential impact.

Criteria based on analytical variability can be difficult to meet, particularly for flow-through studies which are more prone to occasional errors or drift of dosing systems and for “difficult” test substances. However, justifications for such exceptions should be fully described in the report.

Flexibility over biological criteria may be difficult, since they exist to ensure quality of the test organisms, but also to ensure there are sufficient individuals available at the end of a test

for the determination of certain endpoints (e.g. growth). However, it should be acknowledged these criteria are often challenging to meet in certain species (e.g. fertilisation success in trout and larval survival to the free-feeding stage in zebrafish).

Duration of the study also enhances the challenge to meet the acceptance criteria. It should also be noted that when studies have numerous test acceptance criteria, even though test acceptability criteria are individually reasonable, the probability of random variation causing failure to meet at least one can be relatively high. Flexibility in interpretation of small deviations from acceptance criteria in these studies is recommended.

Minor statistical deviations from performance criteria should not be used to reject scientifically sound studies. For example, suppose the criteria for egg fertility of eggs is 95 %, but an individual study achieves 92 %. This is only a minor deviation, because over the course of time it is discovered that highly experienced laboratories can achieve fertility between 90 and 95 % and, in such a case, allowances should be made for studies that fall below the criterion to be acceptable.

Acceptance criteria should be taken as a holistic view and not just a tick-box exercise.

In summary, professional and scientific judgement should be applied to test acceptance criteria. Particularly, as a consequence can be repeat tests with consequent increases in animal use. However, all exceptions should be justified and the potential impact assessed and reported.

References

ASTM (2002) Conducting acute toxicity tests with fishes, macroinvertebrates and amphibians. Standard E729-96. Am. Soc. Test. Mat., 100 Barr Harbor Drive, West Conshohocken, Pennsylvania 19428.

Benoit, D.A. (1981) User's guide for conducting life cycle chronic toxicity tests with fathead minnows (*Pimephales promelas*). Environmental Research Laboratory – Duluth, Duluth, MN. EPA-600/8-81-001

Dyer, S. D., S. E. Belanger, and G. J. Carr. 1997. An initial evaluation of the use of Euro/North American fish species for tropical effects assessments. Chemosphere 35(11):2767-2781.

Duodoroff, P., Anderson, B.G., Burdick, G.E., Galtsoff, P.S., Hart, W.B., Patrick, R., Strong, E.R., Surber, E.W., van Horn, W.M. (1951) Bioassay methods for the evaluation of acute toxicity of industrial wastes to fish. Sewage Industr. Wastes 23: 1380-1397.

Dyer, S.D., Versteeg, D.J., Belanger, S.E., Chaney, J.G., Mayer F. L. (2006) Interspecies correlation estimates (ICE) predict protective environmental concentrations. Environ. Sci. Technol. 40: 3102-3111.

ECETOC (2007) TR 102 - Intelligent Testing Strategies in Ecotoxicology: Mode of Action Approach for Specifically Acting Chemicals. December 2007.

Hutchinson, T.H., Shillabeer, N., Winter, M.J., Pickford, D.B. (2006) Acute and chronic effects of carrier solvents in aquatic organisms: A critical review. Aquat. Toxicol. 76: 69-92.

Hutchinson, T.H., Bögi, C., Winter, M.J., Owens, J.W. (2009) Benefits of the maximum tolerated dose (MTD) and maximum tolerated concentration (MTC) concept in aquatic toxicology. Aquat. Toxicol. 91: 197-202.

Lammer, E., Carr, G.J., Wendler, K., Rawlings, J.M., Belanger, S.E., Braunbeck, T. (2009) Is the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test. Comp. Biochem. Physiol. 149C: 196-209.

OECD (2000) Guidance document on aquatic toxicity testing of difficult substances and mixtures. Organisation for Economic Cooperation and Development, Paris, 53 pp.

OECD (2010) Short guidance on the threshold approach acute fish toxicity. Organisation for Economic Cooperation and Development, Paris.

Rufli, H., Fisk, P.R., Girling, A.R., King, J.M.H., Länge, R., Lejeune, X., Stelter, N., Stevens, C., Suteau, P., Tapp. J., Thus, J., Versteeg, D.J., Niessen, H.J. (1998) Aquatic toxicity testing of sparingly soluble, volatile and unstable substances and interpretation of use of data. Ecotoxicology and Environmental Safety 39: 72-77.

Sprague, J.B. (1969) Measurement of pollutant toxicity to fish I. Bioassay methods for acute toxicity. Water Res. 3: 793-821.

US EPA (1986) Hazard evaluation division standard evaluation procedure – Fish life-cycle toxicity tests. EPA 540/9-86-137. United States Environmental Protection Agency, Washington DC.

US EPA (1996) OPPTS 850.1075 – Fish Acute Toxicity Test, Freshwater and Marine. EPA 712-C-96-118. United States Environmental Protection Agency, Washington DC.