2.0 FETAX TEST METHOD PROTOCOL

2.1 Standard Detailed Protocol

Under the auspices of the ASTM, a comprehensive guideline for FETAX was published in 1991. The guideline was subsequently revised, and the updated version was published in 1998. The two versions of this guideline are designated as a "Standard Guide for Conducting the Frog Embryo Teratogenesis Assay—*Xenopus* (FETAX)," Annual Book of ASTM Standards, Designation E1439–91 and E1439–98, respectively. The most recent guideline expands the information procedures outlined in the 1991 ASTM Guideline. The two versions are provided in **Appendix 10** and **11**, respectively. The ASTM FETAX Guideline includes information on the following topics:

- terminology,
- summary of the guideline,
- significance and use of the assay,
- safety precautions,
- apparatus,
- water for culturing *Xenopus* adults,
- preparation of FETAX solution water,
- test material,
- test organisms,
- procedure,
- analytical methodology,
- test acceptability,
- documentation,
- key words, and
- references.

The appendices to the ASTM FETAX Guideline include a list of alternative species, additional endpoints, and alternative exposure scenarios. The 1998 ASTM FETAX Guideline also includes appendices on concentration steps for range-finding tests, microsome isolation reagents, and nicotinamide adenine dinucleotide phosphate (NADPH)-generating system components. The procedures presented in the ASTM FETAX Guideline (1991, 1998) are considered to be applicable to all chemicals individually or in formulations, commercial products, or mixtures. In addition, the 1998 ASTM FETAX Guideline allows, with appropriate modification, the use of FETAX for conducting tests on surface and ground waters, solid phase samples such as soils and sediments, and whole bulk soils and sediments.

A brief description of the ASTM FETAX Guideline (1991, 1998) follows.

2.1.1 Materials, Equipment, and Supplies

Adults should be kept in an animal room isolated from extraneous light that might interfere with a consistent 12-hour photoperiod. Adults can be maintained in large aquaria or in fiberglass or stainless steel raceways at densities of four to six animals per 1800 cm^2 of water surface area. The sides of the tanks should be opaque and at least 30 cm high. The water depth should be between 7 and 14 cm. Water temperature for adults should be 23 ± 3°C. Two types of breeding aquaria are described in detail (ASTM, 1991; 1998). For conducting FETAX, a constant temperature room or a suitable incubator for embryos is required, although a fixed photoperiod is unnecessary. The incubator must be capable of maintaining a temperature of 24 ± 2°C. Covered 60-mm glass Petri dishes should be used as test chambers, except that disposable 55-mm polystyrene Petri dishes should be used if a substantial amount of the test substance binds to glass, but not to polystyrene, or when metabolic activation is incorporated.

Equipment and facilities that contact stock solutions, test solutions, or water in which embryos will be placed should not contain substances that can be leached or dissolved by aqueous solutions in amounts that would adversely affect embryonic growth or development. Additionally, items that contact stock solutions or test solutions should be chosen to minimize sorption of most test materials from water. Glass, Type 316 stainless steel, nylon, and fluorocarbon plastic should be used whenever possible to minimize dissolution, leaching, and sorption. Rigid plastics may be used for holding, acclimation, and in the water supply system, but they should be soaked for a week before use.

FETAX Solution, stock solutions, or test solutions should not contact brass, copper, lead, galvanized metal, or natural rubber before or during the test. Items made of neoprene rubber or other materials not mentioned above should not be used unless it has been shown that their use will not adversely affect either survival or growth of the embryos and larvae of the test species.

A binocular dissection microscope capable of magnifications up to 30x is required to count and evaluate embryos for malformations. A simple darkroom enlarger is used to enlarge embryo images two to three times for head to tail length measurements. It is also possible to measure embryo length through the use of a map measurer or an ocular micrometer. However, the process is greatly facilitated by using a digitizer interfaced to a microcomputer. The microcomputer is also used in data analysis.

Before FETAX is conducted in new test facilities, it is recommended that a "non-toxicant" test be conducted, in which all test chambers contain FETAX solution with no added test material. The embryos should grow, develop, and survive in numbers consistent with an acceptable test. The magnitude of the chamber-to-chamber variation should be evaluated.

2.1.2 Detailed Procedures for FETAX

As recommended in the ASTM FETAX Guideline (1991, 1998), the following information should be known about the test material before a test is conducted:

- identities and concentrations of major ingredients and major impurities,
- solubility and stability in water,
- estimate of toxicity to humans, and
- recommended safe-handling procedures.

An acceptable clutch of eggs has the capability of developing into Developmental Stage 46 tadpoles with less than 10% gross abnormalities and less than 10% mortality. In practice, 95% normal, live embryos should be obtained routinely. Recognition of high quality eggs is based on the following (J. Bantle, personal communication):

- The eggs must be normally pigmented on the top surface;
- The pigment must be even in coloration and not mottled;
- They cannot have been laid in strings (see Bantle et al., 1998);
- Less than 30% of the eggs should exhibit abnormal pigmentation when first laid;
- Greater than 70% should rotate such that the animal (dark) pole is facing up in the dish;
- Fertilization and normal cleavage rates must be in excess of 70%;

The process of dejellying with 2% cysteine is critical if the developing embryos are not to be damaged by the treatment. Damaged embryos often look normal but soon undergo abnormal cleavage. Treatment with cysteine must only progress until the embryos roll with just a slight amount of stickiness. A method of quantifying this step has not yet been developed. Additionally, excess treatment does not also show immediately as a change in morphology. To learn this process, embryos must be dejellied, the normal-looking ones chosen, and then allowed to grow to Developmental Stage 46 to see if mortality and malformation rates are acceptable. Eggs from the same batch may be subjected to different lengths of dejelly time in order to assess the effects of time on the process.

Test substance exposure is continuous throughout the test. For each dose group, two dishes each containing 25 embryos and 10 mL of test solution are used. For the control group, four dishes of 25 embryos each are used. However, studies that employ 55-mm polystyrene Petri dishes rather than 60-mm glass Petri dishes use 20 embryos per dish (Bantle et al., 1998). Both versions of the ASTM FETAX Guideline (1991, 1998) state that embryos must be randomly assigned to test dishes, but the 1998 version includes a revision to make an exception to random assignment when a forced air incubator is used to eliminate the occurrence of hot or cold locations. A

temperature of 24 ± 2°C must be maintained throughout the 96-hour test duration. Temperatures higher than 26°C cause malformation, whereas low temperatures prevent the controls from reaching Developmental Stage 46 within 96 hours. If 90% of the embryos in the control dishes have not reached Developmental Stage 46 by 96 hours, the test may be extended by three hours. Deviations from this standard exposure time must be reported as deviating from standard FETAX conditions. The pH of the stock and test solutions must be between 6.5 and 9.0, with 7.7 considered optimal. The pH of a control dish and the pH of the highest test concentration should be measured at the beginning of the test, and subsequently at 24-hour intervals.

Since early *Xenopus* embryos have limited ability to metabolize xenobiotics, particularly in regard to cytochrome P-450 activity, the incorporation of metabolic activation into the standard protocol is necessary when FETAX is used to evaluate developmental toxicity/teratogenicity for human health hazard assessment. The MAS is composed of rat liver microsomes and a NADPH-generating system. The rat liver microsomes may be obtained from an Aroclor 1254-treated male rat. Aroclor 1254 is a broad-spectrum cytochrome P-450-inducing agent and liver microsomes from such rats are appropriate in the majority of experimental studies. Rats exposed to isoniazid or uninduced microsomes may be used in those cases where Aroclor 1254-induction is known to repress specific P-450 isozymes. The nature of the test material may suggest the most appropriate inducing system to use. In cases where limited data are available concerning test substance biotransformation, the ASTM FETAX Guideline (1991, 1998) proposes that a set of Aroclor 1254- and isoniazid-induced rat liver microsomes mixed in equivalent activity ratios be used. However, D. Fort (personal communication) has concluded that a mixture of -napthoflavone- (or 3-methylcholanthrene-), phenobarbital-, and isoniazid-induced microsomes is the most effective source for an MAS. The P-450 activities of each lot of prepared microsomes will vary. Therefore, the P-450 activity of each lot must be determined and a standard amount added to each dish. It is important to include an MAS-only (microsomes and generator system without test material) negative control. Microsomal protein can slow growth and development at concentrations greater than 60 µg/mL. Reduced nicotinamide adenine dinucleotide, which is required for microsomal activity, can also cause abnormal development and its concentration must be kept low. Additional research may be needed to establish the most appropriate criteria for using the different MAS proposed and the optimal conditions for each. The use of an

exogenous MAS in FETAX may not result in the same effects that would be expected to occur if *Xenopus* embryos were P-450 metabolically competent.

Following range-finding tests to identify the appropriate doses to test (see **Section 2.1.3**), three replicate definitive tests are performed. Each of the three definitive tests is conducted using embryos from a different male/female pair of *X. laevis*. If FETAX is being used for human health developmental hazard assessment, definitive tests should be conducted with and without metabolic activation. At a minimum, five concentrations for each endpoint are used. However, additional concentrations between the EC16 and EC84 are highly recommended to ensure obtaining accurate 96-hour LC_{50} and EC_{50} values. The same test material concentrations must be used for each replicate definitive test. The experiments, with and without metabolic activation, should yield acceptable 96-hour LC_{50} and EC_{50} values. If they do not, the tests should be repeated. Prior testing suggests that intra-test variability should yield a coefficient of variation that is less than 100%. In some cases where test variability is extremely high, it may be necessary to determine whether the test material is rapidly degrading, salting out, or volatilizing out of solution.

As defined by the ASTM FETAX Guideline (1991, 1998), a FETAX study should be considered unacceptable if one or more of the following occurs:

- Embryos from more than one mating pair were used in the same test or in replicate tests;
- Hardware cloth or metal mesh was used as a support in the breeding aquarium;
- In the negative controls, either the mean survival is less than 90% or the mean malformation in embryos is greater than 10%, or both;
- Ninety percent of the FETAX-solution-only controls do not reach Developmental Stage 46 by the end of 96 to 99 hours;

- Dilution water was used in the test, and it did not allow embryonic growth at the same rate as FETAX solution;
- The deionized or distilled water does not conform to the Type I ASTM standard;
- A water, FETAX Solution, an MAS control (where an MAS is used), or solvent control was not included in the test;
- The concentration of solvent was not the same in all treatments, except for a dilution-water or FETAX-solution control;
- Identification of the Developmental Stage of the embryos was performed using a reference other than Nieuwkoop and Faber (1975);
- The test was started either with less than Developmental Stage 8 blastulae or with greater than Developmental Stage 11 gastrulae;
- All Petri dishes (or other containers) were not physically identical throughout the test.
- Petri dishes were not randomly assigned to their positions in a non-forced air incubator.
- The embryos were not randomly assigned to the Petri dishes;
- Required data concerning mortality, malformation, and growth were not collected;
- The pH of the test solution was less than 6.5 or greater than 9.0 in the control or highest test concentration;
- Dead embryos were not removed after each 24-hour (±2 hour) interval;

- There was consistent deviation from the temperature limits (a short-term deviation of more than ±2°C might be inconsequential); or

- The reference toxicant produced significant variability (± 2 standard deviation units from the historical mean values) compared to historical data plotted on a control chart.

2.1.3 Dose-Selection Procedures—Range-Finding Test

The ASTM FETAX Guideline (1991, 1998) describes the range-finding tests and the concentration selection procedure. Range-finding tests should be used whenever possible to identify the best approximation of the 96-hour LC_{50} and EC_{50} for definitive testing. Concentration selection is a multistep process that depends on the nature of the test material and the results of the first range-finding test. The first range-finding test consists of a series of at least seven concentrations that differ by a factor of ten. If FETAX is being used for human developmental hazard assessment, range-finding tests should be conducted with and without metabolic activation.

A second range-finding test series is performed using a sliding scale of concentrations provided in the ASTM FETAX Guideline (1991, 1998). The concentration values range from 0.001 to 100; in steps of 0.0005 between 0.001 and 0.1, in steps of 0.05 between 0.1 and 1, in steps of 0.5 between 1 and 10, and in steps of 5 between 10 and 100. Using the sliding scale, the value closest to the 96-hour LC_{50} (for tests conducted with and without metabolic activation) should be identified and then three values immediately below and three values immediately above the estimated LC_{50} should be chosen. The same method should be used to estimate concentrations surrounding the 96-hour EC_{50}. In addition, the 96-hour LC_5, LC_{16}, LC_{84}, and LC_{95} and the EC_5, EC_{16}, EC_{84}, and EC_{95} may be calculated. By determining these values, the concentrations to be tested in the definitive tests are established and the slopes of the concentration-response curves are taken into consideration. Growth inhibition data are not collected from range-finding tests. For some test materials, it may be necessary to use the results of the first definitive experiment as another range-finding test and to adjust the test concentrations accordingly.

2.1.4 Endpoints Measured

The three endpoints measured are mortality, malformations, and embryonic growth.

Mortality: Dead embryos must be removed when solutions are changed at the end of each 24-hour period during the 96-hour test. If dead embryos are not removed, microbial growth can occur that might kill live embryos. Death at 24 hours (Developmental Stage 27) is ascertained by the extent of skin pigmentation, structural integrity, and irritability of the embryo. At 48 hours (Developmental Stage 35), 72 hours (Developmental Stage 42), and 96 hours (Developmental Stage 46), the lack of a heartbeat is an unambiguous sign of death. Based on the mortality data obtained over a range of dose levels, the LC_{50} value is calculated (ASTM, 1991; 1998).

Malformations: Malformations must be recorded at the end of the 96-hour treatment period. The Atlas of Abnormalities (Bantle et al., 1998) should be used in scoring malformations. The number of malformations in each category should be reported in standard format for ease of comparison. Based on malformation data obtained over a range of dose levels, the EC_{50} value is calculated (ASTM, 1991; 1998).

Generally, the two point estimates for mortality and malformations are then used to calculate a TI, which is equal to the LC_{50} divided by the EC_{50} (Bantle et al., 1989; ASTM, 1991; 1998).

Embryonic Growth: The ability of a material to inhibit embryonic growth is often the most sensitive indicator of developmental toxicity (ASTM, 1991; 1998). Head to tail length data (growth) must be collected at the end of each test. If the embryo is curved or kinked, then the measurement follows the contour of the embryo. Measurement should be made after the embryos are fixed in 3% formalin. Using length data, the MCIG is determined by statistically comparing the mean head to tail length of the treated embryos at each dose group to that of the embryos in the control group (ASTM, 1991; 1998).

2.1.5 Duration of Exposure

The ASTM FETAX Guideline (1991, 1998) specifies that *X. laevis* embryos are exposed for 96 hours to the test material. However, if 90% of the embryos in the control dishes have not reached Developmental Stage 46 by this time, the test may be extended by three hours to attain this developmental stage.

2.1.6 Known Limits of Use

As presented in **Section 1.2.4**, FETAX is considered to be applicable to most chemicals and mixtures. Testing of water-insoluble materials would be limited by the highest concentration that can be achieved using an appropriate organic solvent. The test method is incompatible with substances (or concentrations of substances) that alter the pH, hardness, alkalinity, and conductivity of the FETAX solution beyond the acceptable range indicated by the ASTM FETAX Guideline (1991, 1998).

2.1.7 Nature of the Responses Assessed

In FETAX, the primary endpoints assessed are mortality, malfomations, and growth inhibition (ASTM, 1991; 1998; Finch, 1994) (see **Section 2.1.4**). Mortality is an easily observable endpoint. Growth inhibition, as measured by a significant decrease in the head to tail length, is also easily measured. Malfomations in *Xenopus* can be difficult to identify (see **Section 6.6.2**)

2.1.8 Appropriate Vehicle, Negative, and Positive Controls

As specified by the most recent ASTM FETAX Guideline (1998), a stock solution should be prepared anytime the test substance can not be directly added to the test vessel. Test substances administered using a stock solution should be prepared in such a manner as to ensure that the embryos are exposed to a homogeneous mixture. The concentration and stability of the test substance in a stock solution should be determined before testing. Stock solutions should be

prepared daily unless analytical data indicate the solution is stable with time. If the test material is subject to photolysis, the stock solution should be shielded from light.

The preferred solvent for this assay is FETAX Solution; ingredients are provided in the ASTM FETAX Guideline (1991, 1998). The minimum necessary amount of a strong acid or base may be used in the preparation of an aqueous stock solution, but this might appreciably affect the pH of test solutions. Use of a more soluble form of the test material, such as chloride or sulfate salts of organic amines, sodium or potassium salts of phenols or organic acids, and chloride or nitrate salts of metals, might affect the pH more than the use of a minimum necessary amount of a strong acid or base. Prior to testing, all available chemical and physical data on the test substance should be obtained and considered prior to making decisions on pH adjustments.

If a solvent other than FETAX Solution is used, its concentration in test solutions must be demonstrated to not adversely affect *Xenopus* embryo growth and survival. Because of its low toxicity, low volatility, and high ability to dissolve many organic chemicals, triethylene glycol is often a good organic solvent for preparing stock solutions. Other water-miscible organic solvents such as dimethyl sulfoxide and acetone also may be used. Ethanol is not recommended because of its potential teratogenicity. Methanol has high toxicity in FETAX. Acetone might stimulate the growth of microorganisms and is quite volatile. Organic solvents should be reagent-grade six or better. A surfactant should not be used in the preparation of a stock solution because it might affect the form and toxicity of the test material in the test solutions.

If a solvent other than dilution-water or FETAX Solution is used, at least one solvent control test group, using solvent from the same batch used to make the stock solution, must be included in the test. A dilution-water or FETAX Solution control should also be included in the test. If no solvent other than dilution-water or FETAX Solution is used, then a dilution-water or FETAX Solution control must be included in the test. The concentration of solvent must be the same in the solvent control and in all test solutions.

For studies conducted without metabolic activation, 6-aminonicotinamide (6-AN; purity > 99%) is proposed in the ASTM FETAX Guideline (1991; 1998) as the positive or reference toxicant,

as this substance presents a mortality and malformation database convenient for reference purposes. The 1998 ASTM FETAX Guideline provides reference values for 6-AN for the 96-hour LC_{50} (2.23 mg/mL) and the 96-hour EC_{50} (0.005 mg/mL), which yield a TI of 446. The MCIG should be ~1.15 mg/mL. However, based on the excessive variability in results obtained for 6-AN in the Phase I Validation Study, the investigators concluded that a replacement reference substance should be identified (Bantle et al., 1994a). A replacement for 6-AN has not yet been identified. A concurrent positive control for studies conducted without metabolic activation is not recommended in the ASTM FETAX Guideline (1991, 1998). Rather, at least quarterly, concentration-response experiments must be performed and the results of these tests compared with historical tests to judge the laboratory quality of FETAX data. Only those biological responses related to mortality and malformations are considered in this analysis; growth inhibition need not be evaluated for 6-AN. NICEATM suggests that the appropriateness of a reference positive control, as opposed to a concurrent positive control, for FETAX studies conducted without metabolic activation should be critically evaluated.

The recommended concurrent bioactivation positive control for studies conducted with metabolic activation is cyclophosphamide (CP) at a concentration of 4 mg/mL. The MAS-only control and the CP-only control should result in less than 10% mortality and malformations. With metabolic activation, bioactivated CP should kill 100% of the embryos within 96 hours. The ASTM FETAX Guideline (1991, 1998) states that a control is needed also to demonstrate that the cytochrome P-450 system is responsible for the observed bioactivation. For this control, a small amount of dithionite may be added directly to the microsomes followed by bubbling carbon monoxide through the microsomal protein at a steady rate for three minutes to inactivate the cytochrome P-450.

NICEATM concluded that the appropriateness of using CP at a concentration that results in 100% mortality should be critically evaluated. A response of this magnitude limits a statistical consideration of historical data. Also, as the TI is considered a primary measure of teratogenic potential, it may be more informative if a concentration of CP is used that allows for an assessment of malformations, as well as mortality.

2.1.9 Acceptable Range of Negative and Positive Control Responses

For negative or solvent controls, the percentage of malformed embryos must not exceed 10%, while mean survival must be greater than 90% (ASTM, 1998).

For 6-AN and CP, the two positive control chemicals recommended in the ASTM FETAX Guideline (1991, 1998), no specific acceptable range of values was provided. However, the ASTM FETAX Guideline (1991, 1998) states that the reference toxicant test must produce data within two standard deviations of the historical mean values. No information is provided in the ASTM FETAX Guideline (1991; 1998) on the number of studies required to generate appropriate historical data or the time period over which such data should be retrospectively assessed. When conducting studies with metabolic activation, the MAS-only control and the CP-only control should result in less than 10% mortality and malformations. With metabolic activation, bioactivated CP should kill 100% of the embryos within 96 hours. No other information is provided in the ASTM FETAX Guideline (1991, 1998).

2.1.10 Data Collection

As described in the ASTM FETAX Guideline (1991, 1998), data are collected on the incidence of embryos that have died during the 96-hour culture period; the head to tail length, a measure of growth, among the surviving embryos at the end of the 96-hour culture period (see **Section 2.1.4**), and on the incidence and type of malformations present among the surviving embryos at the end of the 96-hour culture period. Malformations are scored using a binocular dissection microscope capable of magnifications up to 30x. The standard FETAX scoring form (ASTM, 1991; 1998) includes the following categories to be scored during an assessment of malformations:

- severe,
- stunted,
- gut,
- edema (multiple, cardiac, abdominal, facial, cephalic, optic),
- axial malformations (tail, notochord, fin),
- face,
- eye,
- brain,

- hemorrhage,
- cardiac,
- blisters,
- other (specify)

2.1.11 Data Storage Media

Original data are collected on FETAX-specific forms and maintained in study books. Example forms are provided in the ASTM FETAX Guideline (1991, 1998). Data are then generally entered into computerized spreadsheets for manipulation and analysis.

2.1.12 Measures of Variability

In FETAX, as described in the ASTM FETAX Guideline (1991, 1998), each test substance concentration involves the use of two replicate dishes, while each control treatment group involves the use of four replicate dishes. Each plastic or glass Petri dish contains 20 or 25 embryos, respectively. Sterile plastic Petri dishes are used with MAS to reduce the possibility of bacterial contamination. To evaluate the teratogenicity of a test material, three replicate definitive tests are performed. Each of the three definitive tests is conducted using embryos from a different male/female pair of *X. laevis.* If FETAX is being used for human health hazard assessment, definitive tests are conducted with and without metabolic activation. The ASTM FETAX Guideline (1991, 1998) specifies that the geometric mean for the 96-hour LC_{50}, the 96-hour EC_{50}, the TI, and the MCIG, as well as their 95% confidence limits be calculated using the data from the three replicate definitive tests and provided in the study report. **Section 2.1.13** describes the statistical methods used to calculate the 96-hour LC_{50}, the 96-hour EC_{50}, and the MCIG.

Intra- and inter-laboratory variation in FETAX has been evaluated using the four different measurements—LC_{50}, EC_{50}, TI, and the MCIG—that can be obtained from each experiment. In some studies, the types of malformations present in the embryos were considered also. Reproducibility and reliability of each FETAX endpoint were evaluated by calculating coefficients of variation (CV [%]), and comparing the CVs for each measure across laboratories. Additionally, a statistical approach described in ASTM E691—92 (ASTM, 1992) (**Appendix 12**), a guide for evaluating inter- and intra-laboratory variability, was used.

Historical negative and positive control data can be used to evaluate variability in performance within a laboratory across time. The ASTM FETAX Guideline (1991, 1998) states that, at least quarterly, concentration-response experiments must be performed for the positive control (without metabolic activation) and the results of these tests compared with historical tests to judge the laboratory quality of FETAX data. The reference toxicant test must produce data within two standard deviations of the historical mean values.

2.1.13 Statistical and Non-Statistical Methods

As described in the ASTM FETAX Guideline (1991, 1998), if the test contains a dilution-water or a FETAX-Solution control and a solvent control, the mortality, malformation, and growth inhibition of these treatment groups should be compared using a two-tailed student's t-test. If a statistically significant difference in mortality, malformation, or growth inhibition is detected between the two controls, only the solvent control may be used as the basis for comparison in the calculation of results.

For the range-finding and definitive tests, probit analysis, trimmed Spearman-Karber analysis, or the two-point graphical method are used to estimate the LC_{50} and EC_{50} values. The graphical method is used only when regular statistical analyses fail to generate useful data. Generally, probit analysis is used when the data meet normal distribution and homogeneity of variance assumptions, and the trimmed Spearman-Karber test is used when the data fail to meet these assumptions. However, range-finding tests may bypass the homogeneity of variance requirements. Data sets that are marginal in terms of concentration-response information should not be analyzed by probit analysis as it may skew the data (D. Fort, personal communication). Spearman-Karber should be used when in doubt or to confirm the results of probit analysis.

The TI, the ratio of the LC_{50} to the EC_{50}, is calculated for each test, and then the mean of the three tests determined. The MCIG is determined using a student's t-test for grouped observations, with significance at the $p = 0.05$ level.

The decision criteria described in the ASTM FETAX Guideline (1991, 1998) are based on non-statistical methods (see **Section 2.1.14**).

2.1.14 Decision Criteria

The decision criteria for FETAX are described in the ASTM FETAX Guideline (1991, 1998). The assay provides concentration-response data for mortality, malformations, and growth inhibition. These data can be compared with similar data on a molar basis using other pure test materials or using standard amounts of environmental samples to yield a relative ranking of toxicity. A test substance is considered to be a developmental toxicant when it causes any deficit in an embryo, especially at concentrations lower than those required to induce adult toxicity. In comparison, a teratogen causes some observable abnormality in embryonic development. Three separate FETAX decision criteria (i.e., TI, growth inhibition, and severity of malformations) are used to identify teratogens. Any single decision criterion is considered sufficient to identify a potential teratogenic hazard (ASTM, 1991; 1998).

The TI presents a relative ranking of hazard from nearly 1 to several thousand. The hazard becomes a concern when the mean TI value of three definitive tests is greater than 1.5 (ASTM, 1991; 1998). More recently, Fort et al. (2000a) used a decision criterion in which TI values greater than 1.5 indicate increasing teratogenic hazard, while TI values greater than 3.0 indicate concern. The mortality and malformation concentration-response curves should have similar slopes with acceptable confidence limits when compared to data from 6-AN reference experiments. The term "acceptable" is not defined in the guideline. The TI values of different test substances can be compared to generate relative potency rankings.

In terms of malformations, non-teratogens cause slight to moderate malformations at concentrations near the 96-hour LC_{50}. Teratogens generally cause moderate to severe malformations at these concentrations. Comparison can be made to the reference control 6-AN to identify what constitutes a severe malformation (ASTM, 1991; 1998). An Atlas of Abnormalities (Bantle et al., 1998) is available for judging the severity of malformation.

Growth inhibition is correlated with teratogenesis in FETAX. Teratogenic hazard becomes apparent when growth is significantly affected at concentrations below 30% of the 96-hour LC_{50}. When using this decision criterion, it is important to ensure that the test concentrations selected are adequate to define the MCIG.

Although the ASTM FETAX Guideline (1991, 1998) states that any single decision criterion is considered sufficient to identify a potential teratogenic hazard, Bantle et al. (1999) and Fort et al. (2000a) have also evaluated FETAX results based on multiple decision criteria. In the most recent multi-laboratory validation study (Bantle et al., 1999), each test chemical was judged to pose a developmental hazard when the TI and the MCIG/LC_{50} ratio both indicated hazard (i.e., the TI was value greater than 1.5 and the MCIG/LC_{50} ratio was less than 0.30), and definitely no hazard when both decision criteria fell into the non-hazard category (i.e., the TI value was less than or equal to 1.5 and the MCIG/LC_{50} ratio was greater than or equal to 0.30) (see **Section 7.2.5**). The teratogenic hazard was considered equivocal when one, but not both, of the two decision criteria were positive. In such cases, the types and severity of malformations were examined for guidance in assessing teratogenic hazard. However, due to the subjectivity of malformation identification, this approach was not made a permanent part of the decision criteria (Bantle et al., 1999). In the comparative FETAX - rat teratogenic study conducted by Fort et al. (2000a), a similar multiple criteria approach was used except that the TI decision criterion was based on a TI value greater than 3.0.

2.1.15 Test Report Information

As stated in the ASTM FETAX Guideline (1991, 1998), the test report for an acceptable FETAX study should include the following information either directly or by reference to existing publications:

- The name of the test substance, the name of the investigator(s), the location of laboratory, and the dates of initiation and termination of test;

- The source of test substance, its lot number, composition (identities and concentrations of major ingredients and major impurities), known chemical and physical properties, and the identity and concentration(s) of any solvent used;

- If a dilution water other than FETAX Solution is used, its chemical characteristics and a description of any pretreatment;

- A recent analyses of FETAX Solution and adult culture water;

- pH measurements of the control and of the highest test concentrations at the end of each 24-hour time period;

- Available data on sample hardness, alkalinity, conductivity, total organic carbon (TOC), concentration of dissolved oxygen, and metal content;

- The mortality, malformation rates, and the mean embryo length at 96 hours in the dilution-water, FETAX Solution, or solvent control;

- The mortality and malformation results obtained for the positive control. If a full concentration-response curve was performed, then the 96-hour LC_{50}, the 96-hour EC_{50}, and their confidence limits should be reported;

- The 96-hour LC_{50}, the 96-hour EC_{50}, the TI, and the MCIG for each test. Also, the geometric means of these values and their 95% confidence limits;

- Concentration-response data for mortality, malformation, and growth inhibition may be provided;

- A table for each test that lists the percent mortality, percent malformation, and the head to tail length at each concentration tested;

- The names of the statistical tests employed, the alpha-levels of the tests, and some measure of the variability of the hypothesis tested;

- The types, frequency, and severity of malformations; and

- Any deviations from a standard FETAX study (e.g., exposure periods exceeding 96 hours or pulse exposures, the use of static exposure techniques, not using FETAX Solution as the diluent).

Although this level of detail was specified in the ASTM FETAX Guideline (1991), such detail was not provided in any FETAX study report evaluated by NICEATM. The following were generally not included in FETAX study reports:

- a table for each test that listed the percent mortality, percent malformation, and the head to tail length at each concentration tested;

- quantitative information on the types, frequency, and severity of malformations detected;

- pH measurements of the control and of the highest test concentrations at the end of each 24-hour time period;

- quantitative information on mortality, malformation rates, and the mean embryo length at 96 hours in the dilution-water, FETAX Solution, or solvent control; and

- the mortality and malformation results obtained for the positive control.

2.2 Commonly Used Variations in the FETAX Standard Protocol and Rationale

The ASTM FETAX Guideline (1991, 1998) discusses other types of data that can be collected in FETAX, which increases its versatility. The types of data listed below have been collected in past studies (ASTM, 1991; 1998).

Collecting data on pigmentation might be useful for measuring neural damage because it is thought that the size of the pigment patches is under control of the nervous system. Agents that affect these nerves cause smaller pigment patches and the overall color of the 96-hour larvae will pale. Comparison to the standard Atlas of Abnormalities (Bantle et al., 1998) and suitable controls must be made to determine abnormal pigmentation. Other causes of depigmentation are possible, including loss of melanin production. A concentration-response curve can be generated and an EC_{50} related to pigmentation can be determined. Scoring for pigmentation is considered to be subjective.

Collecting locomotion data is potentially useful in measuring specific neural or muscle damage since larvae with substantial cellular damage swim poorly, erratically, or not at all. The ability to swim properly should be determined by comparison to appropriate controls. A concentration-response curve can be generated and an EC_{50} related to locomotion can be determined. Scoring for locomotion is considered to be subjective.

The embryos hatch from the fertilization membrane between 18 and 30 hours. The number failing to hatch at 48 hours could be recorded. Delay or failure indicates a slowing of developmental processes. This is analogous to developmental staging the embryos at the end of the 96-hour time period except that it is much easier to score hatching. A concentration-response curve can be generated and an EC_{50} related to hatching can be determined.

In special circumstances, exposure periods exceeding 96 hours or pulse exposures, or both, may be performed. Studies conducted with longer exposures should be reported as deviating from the standard FETAX assay. In the static technique, the test substance is added at the beginning of the test and is not changed. It should be recognized that many test substances degrade in a short time

period. The static technique should only be used for test substances that are extremely stable and do not volatilize or sorb to the test dishes. The cost or the available amount of the test substance might also dictate that the static technique be used.

A toxicant-delivery system is used to continuously deliver toxicant and dilution water to the embryos in a flow-through system. Small glass containers with bottom screening are used to contain the embryos in a larger diluter apparatus. The flow-through technique is recommended for test substances that degrade quickly or are volatile. Every attempt should be made to use FETAX solution as the diluent. This variation in procedure must be reported as deviating from the standard FETAX assay.

2.3. Basis for Selection of FETAX

FETAX is proposed as a screen for human developmental hazards based on the conclusion of the developers that the assay is easy, rapid, reliable, inexpensive, and predictive of mammalian developmental hazards. FETAX is essentially an organogenesis test, and organogenesis is highly conserved across amphibians and mammals. The first 96 hours of embryonic development in *Xenopus* parallel many of the major processes of human organogenesis (ASTM, 1991; 1998).

2.4 Confidentiality of Information

Copies were obtained by NICEATM of all original data collected during the five FETAX validation studies (see **Section 7.0**). Original data was not sought by NICEATM for any other publication containing FETAX data or for any publication containing laboratory mammal or human data.

2.5 Basis for FETAX Decision Criteria

As specified by the ASTM FETAX Guideline (1991; 1998), three separate decision criteria (a TI value greater than 1.5, a MCIG/LC_{50} ratio less than 0.3, and severity of malformations) have been used to identify potential human teratogens. The ASTM FETAX Guideline (1991, 1998)

concludes that any single decision criterion is sufficient to identify a potential teratogenic hazard, and that these three decision criteria are based on empirical evidence resulting from over 100 materials tested (without metabolic activation) in FETAX.

More recently, Bantle et al. (1999) and Fort et al. (2000a) have evaluated study results based on multiple decision criteria. In the Phase III.3 Validation Study conducted by Bantle et al. (1999; see **Section 7.2.5**), each test substance was judged to have developmental hazard when both the TI value and the $MCIG/LC_{50}$ ratio indicated hazard (i.e., the TI >1.5 and the $MCIG/LC_{50}$ <0.30), and definitely not hazardous when both decision criteria fell into the non-hazard category (i.e., the TI 1.5 and the $MCIG/LC_{50}$ 0.30). The hazard was considered equivocal when any one of the two decision criteria suggested hazard. In such cases, the types and severity of malformations were examined for guidance in assessing teratogenic hazard. However, due to the subjectivity of malformation identification, this approach was not made a permanent part of the decision criteria. In the comparative FETAX - rat teratogenic study conducted by Fort et al. (2000a), a similar combined decision criteria approach was used except that the TI decision criterion was a TI value greater than 3.0. No further information for the basis of these criteria were provided.

2.6 Basis for Numbers of Replicates and Repeat Tests in FETAX

In FETAX, as defined by the ASTM FETAX Guideline (1991, 1998), one or more range-finding tests and three replicate definitive tests are performed on each test substance. Each of the three definitive tests is conducted using embryos from a different male/female pair of *X. laevis.* Each test consists of several different concentrations of the test substance with two replicate dishes at each test concentration and four replicate dishes for each control. Each plastic or glass Petri dish contains 20 or 25 embryos, respectively. The number of embryos per dish, the number of replicate dishes per test substance concentration, and the number of replicate tests per study were not based on a formal scientific analysis. Rather, selection was based on the best scientific judgement of the developers/users of the assay at the time the ASTM FETAX Guideline (1991, 1998) was prepared (J. Bantle and D. Fort, personal communication).

2.7 Validation Study Based Modifications to the Standard Protocol

The FETAX protocol used in the Phase I Validation Study (Bantle et al., 1994a) followed the 1991 ASTM FETAX Guideline. Based on the results obtained, several changes to the standard FETAX protocol were recommended by the investigators; including:

- increasing the acceptable malformation rate in FETAX Solution controls from 7% to 10%;
- distributing 25-mL volumes of the toxicant solution to 50-mL flasks prior to aliquoting into dishes; and
- potentially eliminating 6-AN as the positive reference control.

The first recommendation was based on the larger than anticipated range in the incidence of malformations among control cultures in several laboratories. The purpose of the second recommendation was to potentially reduce intra-dish variability within a treatment group. The recommendation for potentially eliminating 6-AN as the positive control for studies conducted without metabolic activation was based on the extensive variability seen within and across laboratories for this test material. However, a possible replacement positive control for 6-AN has not yet been identified (J. Bantle, personal communication) and this chemical was still recommended as reference control in studies conducted without metabolic activation in the revised 1998 ASTM FETAX Guideline.

The FETAX protocol used in the subsequent validation studies (Phase II, Phase III.1, Phase III.2, and Phase III.3) incorporated the first two protocol changes recommended in Phase I. In Phase III.2 and Phase III.3, the validation protocol was modified to include an exogenous MAS and CP as the appropriate concurrent positive to demonstrate the suitability of the MAS for bioactivation. Also, in Phase III.2 and III.3, 20 embryos were used per dish rather than the 25 recommended by the ASTM FETAX Guideline (1991, 1998). This modification to the protocol was due to the use of slightly smaller plastic Petri dishes in studies incorporating an MAS. The

arithmetic mean rather than the geometric mean recommended by the ASTM FETAX Guideline (1991, 1998) was calculated for the 96-hour LC_{50}, the 96-hour EC_{50}, the TI, and the MCIG in these studies.

2.8 Section 2 Conclusions

The 1991 and the revised and expanded 1998 FETAX Guideline published by ASTM are detailed, comprehensive, and well-structured. Adequate information is provided on the necessary materials, equipment, and supplies; range-finding and definitive tests; endpoint (mortality, malformations, and embryonic growth) assessment; nature of the responses assessed; the duration of exposure; data collection and data storage media; measures of variability; statistical and non-statistical methods; test report information; commonly used protocol variations and rationale; the use of alternative species; and the basis for selection of FETAX.

Known limits of use for FETAX were not described, except that it was stated that the test method is incompatible with materials (or concentrations of materials) that alter the pH, hardness, alkalinity, and conductivity of the FETAX solution beyond the acceptable range indicated by the ASTM FETAX Guideline (1991, 1998). It would also be expected that the testing of water insoluble materials would be limited by the highest concentration that can be achieved using an appropriate organic solvent (and concentration) that does not alter embryonic growth or survival.

Appropriate vehicle, negative, and positive controls were described. The recommended positive controls were 6-AN for studies without metabolic activation and CP for studies with metabolic activation (ASTM, 1991; 1998). However, one conclusion of the Phase I Validation Study (Bantle et al., 1994a) was that 6-AN was not an appropriate positive control for studies without metabolic activation and that another chemical should be identified for this purpose. To date, a replacement for 6-AN has not been identified. The ASTM FETAX Guideline (1991, 1998) recommends that concentration-response experiments for 6-AN be performed at least quarterly and the results of these tests compared with historical tests to judge the laboratory quality of FETAX data. NICEATM concluded that the inclusion of a concurrent positive control in each study without metabolic activation should be considered.

Information on the acceptable range of negative control response for FETAX was provided in the ASTM FETAX Guideline (1998). It was also stated that the reference toxicant 6-AN test must produce data within two standard deviations of the historical mean values. However, no information was provided on the number of experiments required to generate appropriate historical data or the time period over which such data should be retrospectively assessed. For studies conducted with metabolic activation, the bioactivated CP should kill 100% of the embryos within 96 hours. A response of this magnitude limits the utility of historical control data and the use of a test concentration that would enable an analysis of both mortality and malformation data should be considered.

The ASTM FETAX Guideline (1991, 1998) specifies that the geometrical mean for the 96-hour LC_{50}, the 96-hour EC_{50}, the TI, and the MCIG, as well as their 95% confidence limits be calculated using the data from the three replicate definitive tests and provided in the study report. However, in all reports evaluated, the arithmetic mean only has been calculated, and 95% confidence limits were generally not provided.

The three decision criteria used to distinguish between a teratogen and a non-teratogen in FETAX are well described in the ASTM FETAX Guideline (1991, 1998). In the ASTM Guideline, it was stated that these three decision criteria are based on empirical evidence resulting from over 100 materials tested in FETAX, without metabolic activation. Data to support this statement were not provided. Recently, Bantle et al. (1999) and Fort et al. (2000a) have also evaluated study results based on multiple decision criteria. In their analysis, Fort et al. (2000a) increased the TI decision point value from 1.5 to 3.0. In addition, in both studies, the types and severity of malformations were examined for guidance in assessing teratogenic hazard. As judged by NICEATM, the use of multiple decision criteria rather than single decision criterion does not appear to improve the performance characteristics of FETAX against laboratory mammal or human data (see **Section 6.6**).

Selection of the number of embryos per dish (i.e., 20 or 25), the number of replicate dishes per test concentration (i.e., two), and the number of replicate tests per FETAX definitive study (i.e., three) were based on the best scientific judgement of the developers/users of the assay at the time

the ASTM FETAX Guideline (1991, 1998) was developed (J. Bantle and D. Fort, personal communication). It may be useful to conduct a formal analysis of the impact of different numbers of embryos per dish, dishes per test concentration, and replicate definitive tests on the performance of FETAX.