**COUNCIL OF
THE EUROPEAN UNION**

**Brussels, 7 March 2014
(OR. en)**

**7447/14
ADD 1**

**MI 252
CHIMIE 12
ENV 247
COMPET 165
ENT 80**

COVER NOTE

From:	European Commission
date of receipt:	5 March 2014
To:	General Secretariat of the Council
Subject:	ANNEX to COMMISSION REGULATION (EU) No …/.. of XXX amending, for the purpose of its adaptation to technical progress, Regulation (EC) No 440/2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)

Delegations will find attached document D032186/01 ANNEX.

––––––––––––––––––

Encl.:D032186/01 ANNEX

EN

ANNEX

The Annex to Regulation (EC) No 440/2008 is amended as follows:

Chapters B.53, B.54, B.55, B.56, B.57 and B.58 are inserted:

"B.53. Developmental Neurotoxicity Study

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 426 (2007). In Copenhagen in June 1995, an OECD Working Group on Reproduction and Developmental Toxicity discussed the need to update existing OECD test guidelines for reproduction and developmental toxicity, and the development of new guidelines for endpoints not yet covered (1). The working group recommended that a test guideline for developmental neurotoxicity should be written based on a US EPA guideline, which has since been revised (2). In June 1996, a second consultation meeting was held in Copenhagen to provide the Secretariat with guidance on the outline of a new test guideline on developmental neurotoxicity, including the major elements, e.g., details concerning choice of animal species, dosing period, testing period, endpoints to be assessed, and criteria for evaluating results. A US neurotoxicity risk assessment guideline was published in 1998 (3). An OECD Expert Consultation Meeting and an ILSI Risk Science Institute Workshop were held back-to-back in October 2000 and an expert consultation meeting was held in Tokyo 2005. These meetings were held to discuss the scientific and technical issues related to the current test guideline and the recommendations from the meetings (4)(5)(6)(7) were considered in the development of this test method. Additional information on the conduct, interpretation and terminology used for this test method can be found in OECD Guidance Documents No.43 on "Reproductive Toxicity Testing and Assessment" (8) and No.20 on "Neurotoxicity Testing"(9).

INITIAL CONSIDERATIONS

2. A number of chemicals is known to produce developmental neurotoxic effects in humans and other species (10)(11)(12)(13). Determination of the potential for developmental neurotoxicity may be needed to assess and evaluate the toxic characteristics of a chemical. Developmental neurotoxicity studies are designed to provide data, including dose-response characterisations, on the potential functional and morphological effects on the developing nervous system of the offspring that may arise from exposure in utero and during early life.

3. A developmental neurotoxicity study can be conducted as a separate study, incorporated

into a reproductive toxicity and/or adult neurotoxicity study (*e.g.,* test methods B.34 (14), B.35 (15), B.43 (16)), or added onto a prenatal developmental toxicity study (*e.g.,* test method B.31 (17)). When the developmental neurotoxicity study is incorporated within or attached to another study, it is imperative to preserve the integrity of both study types. All testing should comply with applicable legislation or government and institutional guidelines for the use of laboratory animals in research (*e.g.,* 18).

4. The testing laboratory should consider all available information on the test chemical prior to conducting the study. Such information will include the identity and chemical structure of the chemical; its physico-chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the chemical; toxicological data on structurally related chemicals; and the anticipated use(s) of the chemical. This information is necessary to satisfy all concerned that the test is relevant for the protection of human health, and will help in the selection of an appropriate starting dose.

PRINCIPLE OF THE TEST

5. The test chemical is administered to animals during gestation and lactation. Dams are tested to assess effects in pregnant and lactating females and may also provide comparative information (dams versus offspring). Offspring are randomly selected from within litters for neurotoxicity evaluation. The evaluation consists of observations to detect gross neurologic and behavioural abnormalities, including the assessment of physical development, behavioural ontogeny, motor activity, motor and sensory function, and learning and memory; and the evaluation of brain weights and neuropathology during postnatal development and adulthood.

6. When the test method is conducted as a separate study, additional available animals in each group could be used for specific neurobehavioral, neuropathological, neurochemical or electrophysiological procedures that may supplement the data obtained from the examinations recommended by this test method (16)(19)(20)(21). The supplemental procedures can be particularly useful when empirical observation, anticipated effects, or mechanism/mode-of-action indicate a specific type of neurotoxicity. These supplemental procedures may be used in the dams as well as in the pups. In addition, *ex vivo* or *in vitro* procedures may also be used, as long as these procedures do not alter the integrity of the *in vivo* procedures.

PREPARATIONS FOR THE TEST

Selection of animal species

7. The preferred test species is the rat; other species can be used when appropriate. Note, however, the gestational and postnatal days specified in this test method are specific to commonly used strains of rats, and comparable days should be selected if a different species or unusual strain is used. The use of another species should be justified based on toxicological, pharmacokinetic, and/or other data. Justification should include availability of species-specific postnatal neurobehavioral and neuropathological assessments. If there was an earlier test that raised concerns, the species/strain that raised a concern should be considered. Because of the differing performance attributes

of different rat strains, there should be evidence that the strain selected for use has adequate fecundity and responsiveness. The reliability and sensitivity of other species to detect developmental neurotoxicity should be documented.

Housing and feeding conditions

8. The temperature in the experimental animal room should be $22\pm3^{\circ}C$. Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. It is also possible to reverse the light cycle prior to mating and for the duration of the study, in order to perform the assessments of functional and behavioural endpoints during the dark period (under red light), *i.e.,* during the time the animals are normally active (22). Any changes in the light-dark cycle should include adequate acclimation time to allow animals to adapt to the new cycle. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The type of food and water should be reported and both should be analysed for contaminants.

9. Animals may be housed individually or be caged in small groups of the same sex. Mating procedures should be carried out in cages suitable for the purpose. After evidence of copulation or no later than day 15 of pregnancy, mated animals should be caged separately in delivery or maternity cages. Cages should be arranged in such a way that possible effects due to cage placement are minimised. Mated females should be provided with appropriate and defined nesting materials when parturition is near. It is well known that inappropriate handling or stress during pregnancy can result in adverse outcomes, including prenatal loss and altered foetal and postnatal development. To guard against foetal loss from factors which are not treatment-related, animals should be carefully handled during pregnancy, and stress from outside factors such as excessive outside noise should be avoided.

Preparation of the animals

10. Healthy animals should be used, which have been acclimated to laboratory conditions and have not been subjected to previous experimental procedures, unless the study is incorporated in another study (see paragraph 3). The test animals should be characterised as to species, strain, source, sex, weight and age. Each animal should be assigned and marked with a unique identification number. The animals of all test groups should, as nearly as practicable, be of uniform weight and age, and should be within the normal range of the species and strain under study. Young adult nulliparous female animals should be used at each dose level. Siblings should not be mated, and care should be taken to ensure this. Gestation Day (GD) 0 is the day on which a vaginal plug and/or sperm are observed. Adequate acclimation time (*e.g.,* 2-3 days) should be allowed when purchasing time-pregnant animals from a supplier. Mated females should be assigned in an unbiased way to the control and treatment groups, and as far as possible, they should be evenly distributed among the groups (*e.g.,* a stratified random procedure is recommended to provide even distribution among all groups, such as that based on body weight). Females inseminated by the same male should be equalised across groups.

PROCEDURE

Number and sex of animals

11. Each test and control group should contain a sufficient number of pregnant females to be exposed to the test chemical to ensure that an adequate number of offspring are produced for neurotoxicity evaluation. A total of 20 litters are recommended at each dose level. Replicate and staggered-group dosing designs are allowed if total numbers of litters per group are achieved, and appropriate statistical models are used to account for replicates.

12. On or before postnatal day (PND) 4 (day of delivery is PND 0), the size of each litter should be adjusted by eliminating extra pups by random selection to yield a uniform litter size for all litters (23). The litter size should not exceed the average litter size for the strain of rodents used (8-12). The litter should have, as nearly as possible, equal numbers of male and female pups. Selective elimination of pups, *e.g.*, based upon body weight, is not appropriate. After standardisation of litters (culling) and prior to further testing of functional endpoints, individual pups that are scheduled for pre-weaning or post-weaning testing should be identified uniquely, using any suitable humane method for pup identification (*e.g., 24*).

Assignment of animals for functional and behavioural tests, brain weights, and neuropathological evaluations

13. The test method allows various approaches with respect to the assignment of animals exposed *in utero* and through lactation to functional and behavioural tests, sexual maturation, brain weight determination, and neuropathological evaluation (25). Other tests of neurobehavioral function (*e.g.,* social behaviour), neurochemistry or neuropathology can be added on a case-by-case basis, as long as the integrity of the original required tests are not compromised.

14. Pups are selected from each dose group and assigned for endpoint assessments on or after PND 4. Selection of pups should be performed so that to the extent possible both sexes from each litter in each dose group are equally represented in all tests. For motor activity testing the same pair of male and female pups should be tested at all pre-weaning ages (see paragraph 35). For all other tests the same or separate pairs of male and female animals may be assigned to different behavioural tests. Different pups may need to be assigned to weanling versus adult tests of cognitive function in order to avoid confounding the effects of age and prior training on these measurements (26)(27). At weaning (PND 21), pups not selected for testing can be disposed of humanely. Any alterations in pup assignments should be reported. The statistical unit of measure should be the litter (or dam) and not the pup.

15. There are different ways to assign pups to the pre-weaning and post-weaning examinations, cognitive tests, pathological examinations, etc, (see Figure 1 for general design and Appendix 1 for examples of assignment). Recommended minimum numbers of animals in each dose group for pre-weaning and post-weaning examinations are as follows:

Clinical observations and bodyweight	All animals
Detailed clinical observations	20/sex (1/sex/litter)
Brain weight (post fixation) PND 11-22	10/sex (1/litter)
Brain weight (unfixed) ~ PND 70	10/sex (1/litter)
Neuropathology (immersion or perfusion fixation) PND 11-22	10/sex (1/litter)
Neuropathology (perfusion fixation) PND ~70	10/sex (1/litter)
Sexual maturation	20/sex (1/sex/litter)
Other developmental landmarks (optional)	All animals
Behavioural ontogeny	20/sex (1/sex/litter)
Motor activity	20/sex (1/sex/litter)
Motor and sensory function	20/sex (1/sex/litter)
Learning and memory	10/sex[a] (1/litter)

[a] Depending on the sensitivity of cognitive function tests, investigation of a large higher number of animals should be considered *e.g.,* up to 1 male and 1 female per litter (for animal assignments see Appendix 1) (further guidance on sample size is provided in the OECD Guidance Document 43 (8)).

Dosage

16. At least three dose levels and a concurrent control should be used. The dose levels should be spaced to produce a gradation of toxic effects. Unless limited by the physico-chemical nature or biological properties of the chemical, the highest dose level should be chosen with the aim to induce some maternal toxicity (*e.g.,* clinical signs, decreased body weight gain (not more than 10%) and/or evidence of dose-limiting toxicity in a target organ). The high dose may be limited to 1000 mg/kg/day body weight, with some exceptions. For example, expected human exposure may indicate the need for a higher dose level to be used. Alternatively, pilot studies or preliminary range-finding studies should be performed to determine the highest dosage to be used which should produce a minimal degree of maternal toxicity. If the test chemical has been shown to be developmentally toxic either in a standard developmental toxicity study or in a pilot study, the highest dose level should be the maximum dose which will not induce excessive offspring toxicity, or *in utero* or neonatal death or malformations, sufficient to preclude a meaningful evaluation of neurotoxicity. The lowest dose level should aim to not produce any evidence of either maternal or developmental toxicity including neurotoxicity. A descending sequence of dose levels should be selected with a view to demonstrating any dose-related response and a No-Observed-Adverse Effect Level (NOAEL), or doses near the limit of detection that would allow the determination of a benchmark dose. Two- to four-fold intervals are frequently optimal for setting the descending dose levels, and the addition of a fourth dose group is often preferable to using very large intervals (*e.g.,* more than a factor of 10) between dosages.

17. Dose levels should be selected taking into account all existing toxicity data as well as additional information on metabolism and toxicokinetics of the test chemical or related materials. This information may also assist in demonstrating the adequacy of the dosing regimen. Direct dosing of pups should be considered based on exposure and pharmacokinetic information (28)(29). Careful consideration of benefits and disadvantages should be made prior to conducting direct dosing studies (30).

18. The concurrent control group should be a sham-treated control group or a vehicle-

control group if a vehicle is used in administering the test chemical. All animals should normally be administered the same volume of either test chemical or vehicle on a body weight basis. If a vehicle or other additive is used to facilitate dosing, consideration should be given to the following characteristics: effects on the absorption, distribution, metabolism, or retention of the test chemical; effects on the chemical properties of the test chemical which may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals. The vehicle should not cause effects that could interfere with the interpretation of the study neither should it be neurobehaviourally toxic nor have effects on reproduction or development. For novel vehicles , a sham-treated control group should be included in addition to a vehicle control group. Animals in the control group(s) should be handled in an identical manner to test group animals.

Administration of doses

19. The test chemical or vehicle should be administered by the route most relevant to potential human exposure, and based on available metabolism and distribution information in the test animals. The route of administration will generally be oral (*e.g.,* gavage, dietary, via drinking water), but other routes (*e.g.,* dermal, inhalation) may be used depending on the characteristics and anticipated or known human exposure routes (further guidance is provided in the Guidance Document 43 (8)). Justification should be provided for the route of administration chosen. The test chemical should be administered at approximately the same time every day.

20. The dose administered to each animal should normally be based on the most recent individual body weight determination. However, caution should be exercised when adjusting the doses during the last third of pregnancy. If excess toxicity is noted in the treated dams, those animals should be humanely killed.

21. The test chemical or vehicle should, as a minimum, be administered daily to mated females from the time of implantation (GD 6) throughout lactation (PND 21), so that the pups are exposed to the test chemical during pre- and postnatal neurological development. The age at which dosing starts, and the duration and frequency of dosing, may be adjusted if evidence supports an experimental design more relevant to human exposures. Dosing durations should be adjusted for other species to ensure exposure during all early periods of brain development (*i.e.,* equivalent to prenatal and early postnatal human brain growth). Dosing may begin from the initiation of pregnancy (GD 0) although consideration should be given to the potential of the test chemical to cause pre-implantation loss. Administration beginning at GD 6 would avoid this risk, but the developmental stages between GD 0 and 6 would not be treated. When a laboratory purchases time-mated animals, it is impractical to begin dosing at GD 0, and thus GD 6 would be a good starting day. The testing laboratory should set the dosing regimen according to relevant information about the effects of the test chemical, prior experience, and logistical considerations; this may include extension of dosing past weaning. Dosing should not occur on the day of parturition in those animals which have not completely delivered their offspring. In general, it is assumed that exposure of the pups will occur through the maternal milk; however, direct dosing of pups should be considered in those cases where there is a lack of evidence of continued exposure to offspring. Evidence of continuous exposure can be retrieved from *e.g.,* pharmacokinetic information, offspring toxicity or changes in bio-markers (28).

OBSERVATIONS

Observations on dams

22. All dams should be carefully observed at least once daily with respect to their health condition, including morbidity and mortality.

23. During the treatment and observation periods, more detailed clinical observations should be conducted periodically (at least twice during the gestational dosing period and twice during the lactational dosing period) using at least ten dams per dose level. The animals should be observed outside the home cage by trained technicians who are unaware of the animals' treatment, using standardised procedures to minimise animal stress and observer bias, and maximise inter-observer reliability. Where possible, it is advisable that the observations in a given study be made by the same technician.

24. The presence of observed signs should be recorded. Whenever feasible, the magnitude of the observed signs should also be recorded. Clinical observations should include, but not be limited to, changes in skin, fur, eyes, mucous membranes, occurrence of secretions, and autonomic activity (*e.g.,* lacrimation, piloerection, pupil size, unusual respiratory pattern and/or mouth breathing, and any unusual signs of urination or defecation).

25. Any unusual responses with respect to body position, activity level (*e.g.,* decreased or increased exploration of the standard area) and co-ordination of movement should also be noted. Changes in gait, (*e.g.,* waddling, ataxia), posture (*e.g.,* hunched-back) and reactivity to handling, placing or other environmental stimuli, as well as the presence of clonic or tonic movements, convulsions, tremors, stereotypies (*e.g.,* excessive grooming, unusual head movements, repetitive circling), bizarre behaviour (*e.g.,* biting or excessive licking, self-mutilation, walking backwards, vocalisation), or aggression should be recorded.

26. Signs of toxicity should be recorded, including the day of onset, time of day, degree, and duration.

27. Animals should be weighed at the time of dosing at least weekly throughout the study, on or near the day of delivery, and on PND 21 (weaning). For gavage studies dams should be weighed at least twice weekly. Doses should be adjusted at the time of each body weight determination, as appropriate. Food consumption should be measured weekly at a minimum during gestation and lactation. Water consumption should be measured at least weekly if exposure is via the water supply.

Observations on offspring

28. All offspring should be carefully observed at least daily for signs of toxicity and for morbidity and mortality.

29. During the treatment and observation periods, more detailed clinical observations of the offspring should be conducted. The offspring (at least one pup/sex/litter) should be observed by trained technicians who are unaware of the animals' treatment, using standardised procedures to minimise bias and maximise inter-observer reliability. Where

possible, it is advisable that the observations are made by the same technician. At a minimum, the endpoints described in paragraphs 24 and 25 should be monitored as appropriate for the developmental stage being observed.

30. All signs of toxicity in the offspring should be recorded, including the day of onset, time of day, degree, and duration.

Physical and developmental landmarks

31. Changes in pre-weaning landmarks of development (*e.g.,* pinna unfolding, eye opening, incisor eruption) are highly correlated with body weight (30)(31). Body weight may be the best indicator of physical development. Measurement of developmental landmarks is, therefore, recommended only when there is prior evidence that these endpoints will provide additional information. Timing for the assessment of these parameters is indicated in Table 1. Depending on the anticipated effects, and the results of the initial measurements, it may be advisable to add additional time points or to perform the measurements in other developmental stages.

32. It is advisable to use post-coital age instead of postnatal age when assessing physical development (33). If pups are tested on the day of weaning, it is recommended that this testing be carried out prior to actual weaning to avoid a confounding effect by the stress associated with weaning. In addition, any post-weaning testing of pups should not occur during the two days after weaning.

Table 1: Timing of the assessment of physical and developmental landmarks, and functional/behavioural endpoints[a].

Age Periods / Endpoints	Pre-weaning[b]	Adolescence[b]	Young adults[b]
Physical and developmental landmarks			
Body weight and Clinical Observations	weekly[c]	at least every two weeks	at least every two weeks
Brain weight	PND 22[d]		at termination
Neuropathology	PND 22[d]		at termination
Sexual maturation	--	as appropriate	--
Other developmental landmarks[e]	as appropriate	--	--
Functional/behavioural endpoints			
Behavioural ontogeny	At least two measures		
Motor activity (including habituation)	1–3 times (f)	--	once

8

Motor and sensory function	--	once	once
Learning and memory	--	once	once

[a] This table presents the minimum number of times when measurements should be performed. Depending on the anticipated effects, and the results of the initial measurements, it may be advisable to add additional time points (*e.g.,* aged animals) or to perform the measurements in other developmental stages.

[b] It is recommended that pups not be tested during the two days after weaning (see paragraph 32). Recommended ages for adolescent testing are: learning and memory = PND 25±2; motor and sensory function = PND 25±2. Recommended ages for testing young adults is PND 60-70.

[c] Body weights should be measured at least twice weekly when directly dosing pups for adjustment of doses at a time of rapid body weight gain.

[d] Brain weights and neuropathology may be assessed at some earlier time (*e.g.,* PND 11), if appropriate (see paragraph 39).

[e] Other developmental landmarks in addition to the body weight (*e.g.,* eye opening) should be recorded when appropriate (see paragraph 31).

[f] See paragraph 35.

33. Live pups should be counted and sexed *e.g.,* by visual inspection or measurement of anogenital distance (34)(35), and each pup within a litter should be weighed individually at birth or soon thereafter, at least weekly throughout lactation, and at least once every two weeks thereafter. When sexual maturation is evaluated, the age and body weight of the animal when vaginal patency (36) or preputial separation (37) occurs should be determined for at least one male and one female per litter.

Behavioural ontogeny

34. Ontogeny of selected behaviours should be measured in at least one pup/sex/litter during the appropriate age period, with the same pups being used on all test days for all behaviours assessed. The measurement days should be spaced evenly over that period to define either the normal or treatment-related change in ontogeny of that behaviour (38). The following are some examples of behaviours for which their ontogeny could be assessed: righting reflex, negative geotaxis and motor activity (38)(39)(40).

Motor activity

35. Motor activity should be monitored (41)(42)(43)(44)(45) during the pre-weaning and adult age periods. For testing at the time of weaning, see paragraph 32. The test session should be long enough to demonstrate intra-session habituation for non-treated controls. Use of motor activity to assess behavioural ontogeny is strongly recommended. If used as a test of behavioural ontogeny, then testing should utilise the same animals for all pre-weaning test sessions. Testing should be frequent enough to assess the ontogeny of intra-session habituation (44). This may require three or more time periods prior to, and including the day of weaning (*e.g.,* PND 13, 17, 21). Testing of the same animals, or littermates, should also occur at an adult age close to study termination (*e.g.,* PND 60-70). Testing on additional days may be done as necessary. Motor activity should be monitored by an automated activity recording apparatus which should be capable of detecting both increases and decreases in activity, (*i.e.,* baseline activity as measured by the device should not be so low as to preclude detection of decreases, nor so high as to

9

preclude detection of increases in activity). Each device should be tested by standard procedures to ensure, to the extent possible, reliability of operation across devices and across days. To the extent possible, treatment groups should be balanced across devices. Each animal should be tested individually. Treatment groups should be counter-balanced across test times to avoid confounding by circadian rhythms of activity. Efforts should be made to ensure that variations in the test conditions are minimal and are not systematically related to treatment. Among the variables that can affect many measures of behaviour, including motor activity, are sound level, size and shape of the test cage, temperature, relative humidity, light conditions, odours, use of home cage or novel test cage and environmental distractions.

Motor and sensory function

36. Motor and sensory function should be examined in detail at least once for the adolescent period and once during the young adult period (*e.g.,* PND 60-70). For testing at the time of weaning, see paragraph 32. Sufficient testing should be conducted to ensure an adequate quantitative sampling of sensory modalities (*e.g.,* somato-sensory, vestibular) and motor functions (*e.g.,* strength, coordination). A few examples of tests for motor and sensory function are extensor thrust response (46), righting reflex (47)(48), auditory startle habituation (40)(49)(50)(51)(52)(53)(54), and evoked potentials (55).

Learning and memory tests

37. A test of associative learning and memory should be conducted post-weaning (e.g., 25±2 days) and for young adults (PND 60 and older). For testing at the time of weaning, see paragraph 32. The same or separate test(s) may be used at these two stages of development. Some flexibility is allowed in the choice of test(s) for learning and memory in weanling and adult rats. However, the test(s) should be designed so as to fulfil two criteria. First, learning should be assessed either as a change across several repeated learning trials or sessions, or, in tests involving a single trial, with reference to a condition that controls for non-associative effects of the training experience. Second, the test(s) should include some measure of memory (short-term or long-term) in addition to original learning (acquisition), but this measure of memory cannot be reported in the absence of a measure of acquisition obtained from the same test. If the test(s) of learning and memory reveal(s) an effect of the test chemical, additional tests to rule out alternative interpretations based on alterations in sensory, motivational, and/or motor capacities may be considered. In addition to the above two criteria, it is recommended that the test of learning and memory be chosen on the basis of its demonstrated sensitivity to the class of chemical under investigation, if such information is available in the literature. In the absence of such information, examples of tests that could be made to meet the above criteria include: passive avoidance (43)(56)(57), delayed-matching-to-position for the adult rat (58) and for the infant rat (59), olfactory conditioning (43)(60), Morris water maze (61)(62)(63), Biel or Cincinnati maze (64)(65), radial arm maze (66), T-maze (43), and acquisition and retention of schedule-controlled behaviour (26)(67)(68). Additional tests are described in the literature for weanling (26)(27) and adult rats (19)(20).

Post-mortem examination

38. Maternal animals can be euthanised after weaning of the offspring.

39. Neuropathological evaluation of the offspring will be conducted using tissues from animals humanely killed at PND 22 or at an earlier time point between PND 11 and PND 22, as well as at study termination. For offspring killed through PND 22, brain tissues should be evaluated; for animals killed at termination, both central nervous system (CNS) tissues and peripheral nervous system (PNS) tissues should be evaluated. Animals killed on PND 22 or earlier may be fixed either by immersion or perfusion. Animals killed at study termination should be fixed by perfusion. All aspects of the preparation of tissue samples, from the perfusion of animals, through the dissection of tissue samples, tissue processing, and staining of slides should employ a counterbalanced design such that each batch contains representative samples from each dose group. Additional guidance on neuropathology can be found in OECD Guidance Document No. 20 (9), see also (103).

Processing of tissue samples

40. All gross abnormalities apparent at the time of necropsy should be noted. Tissue samples taken should represent all major regions of the nervous system. The tissue samples should be retained in an appropriate fixative and processed according to standardised published histological protocols (69)(70)(71) (103). Paraffin embedding is acceptable for tissues of the CNS and PNS, but the use of osmium in post-fixation, together with epoxy embedding, may be appropriate when a higher degree of resolution is required (*e.g.,* for peripheral nerves when a peripheral neuropathy is suspected and/or for morphometric analysis of peripheral nerves). Brain tissue collected for morphometric analysis should be embedded in appropriate media at all dose levels at the same time in order to avoid shrinkage artifacts that may be associated with prolonged storage in fixative (6).

Neuropathological examination

41. The purposes of the qualitative examination are:

i) to identify regions within the nervous system exhibiting evidence of neuropathological alterations;

ii) to identify types of neuropathological alterations resulting from exposure to the test chemical; and

iii) to determine the range of severity of the neuropathological alterations.

Representative histological sections from the tissue samples should be examined microscopically by an appropriately trained pathologist for evidence of neuropathological alterations. All neuropathologic alterations should be assigned a subjective grade indicating severity. A hematoxylin and eosin stain may be sufficient for evaluating brain sections from animals humanely killed at PND 22, or earlier. However, a myelin stain (*e.g.,* luxol fast blue / cresyl violet) and a silver stain (*e.g.,* Bielschowsky's or Bodians stains) are recommended for sections of CNS and PNS tissues from animals killed at study termination. Subject to the professional judgement of the pathologist and the kind of alterations observed, other stains may be considered appropriate to identify and characterise particular types of alterations (*e.g.,* glial fibrillary acidic protein (GFAP) or lectin histochemistry to assess glial and microglial

alterations (72), fluoro-jade to detect necrosis (73)(74), or silver stains specific for neural degeneration (75)).

42. Morphometric (quantitative) evaluation should be performed as these data may assist in the detection of a treatment-related effect and are valuable in the interpretation of treatment-related differences in brain weight or morphology (76)(77). Nervous tissue should be sampled and prepared to enable morphometric evaluation. Morphometric evaluations may include *e.g.*, linear or areal measurements of specific brain regions (78). Linear or areal measurements require the use of homologous sections carefully selected based on reliable microscopic landmarks (6). Stereology may be used to identify treatment-related effects on parameters such as volume or cell number for specific neuroanatomic regions (79)(80)(81)(82)(83)(84).

43. The brains should be examined for any evidence of treatment-related neuropathological alterations and adequate samples should be taken from all major brain regions (*e.g.*, olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, midbrain (tectum, tegmentum, and cerebral peduncles), pons, medulla oblongata, cerebellum) to ensure a thorough examination. It is important that sections for all animals are taken in the same plane. In adults humanely killed at study termination, representative sections of the spinal cord and the PNS should be sampled. The areas examined should include the eye with optic nerve and retina, the spinal cord at the cervical and lumbar swellings, the dorsal and ventral root fibers, the proximal sciatic nerve, the proximal tibial nerve (at the knee), and the tibial nerve calf muscle branches. The spinal cord and peripheral nerve sections should include both cross or transverse and longitudinal sections.

44. Neuropathological evaluation should include an examination for indications of developmental damage to the nervous system (6)(85)(86)(87)(88)(89), in addition to the cellular alterations (*e.g.*, neuronal vacuolation, degeneration, necrosis) and tissue changes (*e.g.*, gliosis, leukocytic infiltration, cystic formation). In this regard, it is important that treatment-related effects be distinguished from normal developmental events known to occur at a developmental stage corresponding to the time of sacrifice (90). Examples of significant alterations indicative of developmental insult include, but are not restricted to:

- alterations in the gross size or shape of the olfactory bulbs, cerebrum or cerebellum;

- alterations in the relative size of various brain regions, including decreases or increases in the size of regions resulting from the loss or persistence of normally transient populations of cells or axonal projections (*e.g.*, external germinal layer of cerebellum, corpus callosum);

- alterations in proliferation, migration, and differentiation, as indicated by areas of excessive apoptosis or necrosis, clusters or dispersed populations of ectopic, disoriented or malformed neurons or alterations in the relative size of various layers of cortical structures;

- alterations in patterns of myelination, including an overall size reduction or altered staining of myelinated structures;

- evidence of hydrocephalus, in particular enlargement of the ventricles, stenosis of

the cerebral aqueduct and thinning of the cerebral hemispheres.

Analysis of the dose-response relationship of neuropathological alterations

45. The following stepwise procedure is recommended for the qualitative and quantitative neuropathological analyses. First, sections from the high dose group are compared with those of the control group. If no evidence of neuropathological alterations is found in animals of the high dose group, no further analysis is required. If evidence of neuropathological alterations is found in the high dose group, then animals from the intermediate and low dose groups are examined. If the high dose group is terminated due to death or other confounding toxicity, the high and intermediate dose groups should be analysed for neuropathological alterations. If there is any indication of neurotoxicity in lower dose groups, neuropathological analysis should be performed in those groups. If any treatment-related neuropathological alterations are found in the qualitative or quantitative examination, the dose-dependence of the incidence, frequency and severity grade of the lesions or of the morphometric alterations should be determined, based on an evaluation of all animals from all dose groups. All regions of the brain that exhibit any evidence of neuropathologic alteration should be included in this evaluation. For each type of lesion, the characteristics used to define each severity grade should be described, indicating the features used to differentiate each grade. The frequency of each type of lesion and its severity grade should be recorded and a statistical analysis should be performed to evaluate the nature of a dose-response relationships. The use of coded slides is recommended (91).

DATA AND REPORTING

Data

46. Data should be reported individually and summarised in tabular form, showing for each test group the types of change and the number of dams, offspring by sex, and litters displaying each type of change. If direct postnatal exposure of the offspring has been performed, the route, duration and period of exposure should be reported.

Evaluation and interpretation of results

47. A developmental neurotoxicity study will provide information on the effects of repeated exposure to a chemical during *in utero* and early postnatal development. Since emphasis is placed on both general toxicity and developmental neurotoxicity endpoints, the results of the study will allow for the discrimination between neurodevelopmental effects occurring in the absence of general maternal toxicity, and those which are only expressed at levels that are also toxic to the maternal animal. Due to the complex interrelationships among study design, statistical analysis, and biological significance of the data, adequate interpretation of developmental neurotoxicity data will involve expert judgment (107)(109). The interpretation of test results should use a weight-of-evidence-approach (20)(92)(93)(94). Patterns of behavioural or morphological findings, if present, as well as evidence of dose-response should be discussed. Data from all studies relevant to the evaluation of developmental neurotoxicity, including human epidemiological studies or case reports, and experimental animal studies (*e.g.*, toxicokinetic data, structure-activity information, data from other toxicity studies)

should be included in this characterisation. This includes the relationship between the doses of the test chemical and the presence or absence, incidence, and extent of any neurotoxic effect for each sex (20)(95).

48. Evaluation of data should include a discussion of both the biological and statistical significance. Statistical analysis should be viewed as a tool that guides rather than determines the interpretation of data. Lack of statistical significance should not be the sole rationale for concluding a lack of treatment related effect, just as statistical significance should not be the sole justification for concluding a treatment-related effect. To guard against possible false-negative findings and the inherent difficulties in "proving a negative," available positive and historical control data should be discussed, especially when there are no treatment-related effects (102)(106). The probability of false positives should be discussed in light of the total statistical evaluation of the data (96). The evaluation should include the relationship, if any, between observed neuropathological and behavioural alterations.

49. All results should be analysed using statistical models appropriate to the experimental design (108). The choice of a parametric or a nonparametric analysis should be justified by considering factors such as the nature of the data (transformed or not) and their distribution, as well as the relative robustness of the statistical analysis selected. The purpose and design of the study should guide the choice of statistical analyses to minimise Type I (false positive) and Type II (false negative) errors (96)(97)(104)(105). Developmental studies using multiparous species where multiple pups per litter are tested should include the litter in the statistical model to guard against an inflated Type I error rates (98)(99)(100)(101). The statistical unit of measure should be the litter and not the pup. Experiments should be designed such that littermates are not treated as independent observations. Any endpoint repeatedly measured in the same subject should be analysed using statistical models that account for the non-independence of those measures.

Test report

50. The test report should include the following information:

Test chemical:

- physical nature and, where relevant, physiochemical properties;
- identification data, including source;
- purity of the preparation, and known and/or anticipated impurities.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water or physiological saline solution.

Test animals:

- species and strain used, and a justification if other than the rat;
- supplier of test animals;
- number, age at start, and sex of animals;

- source, housing conditions, diet, water, etc.;

- individual weights of animals at the start of the test.

Test conditions:

- rationale for dose level selection;

- rationale for dosing route and time period;

- specifications of the doses administered, including details of the vehicle, volume and physical form of the material administered;

- details of test chemical formulation/diet preparation, achieved concentration, stability and homogeneity of the preparation;

- method used for unique identification of dams and offspring;

- a detailed description of the randomisation procedure(s) used to assign dams to treatment groups, to select pups for culling, and to assign pups to test groups;

- details of the administration of the test chemical;

- conversion from diet/drinking water or inhalation test chemical concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable;

- environmental conditions;

- details of food and water (*e.g.,* tap, distilled) quality;

- dates of study start and end.

Observations and test procedures:

- a detailed description of the procedures used to standardise observations and procedures as well as operational definitions for scoring observations;

- a list of all test procedures used, and justification for their use;

- details of the behavioural/functional, pathological, neurochemical or electrophysiological procedures used, including information and details on automated devices;

- procedures for calibrating and ensuring the equivalence of devices and the balancing of treatment groups in testing procedures;

- a short justification explaining any decisions involving professional judgement.

Results (individual and summary, including mean and variance when appropriate):

- the number of animals at the start of the study and the number at the end of the study;

- the number of animals and litters used for each test method;

- identification number of each animal and the litter from which it came;

- litter size and mean weight at birth by sex;

- body weight and body weight change data, including terminal body weight for dams and offspring;

- food consumption data, and water consumption data if appropriate (*e.g.,* if test chemical is administered via water);

- toxic response data by sex and dose level, including signs of toxicity or mortality, including time and cause of death, if appropriate;

- nature, severity, duration, day of onset, time of day, and subsequent course of the detailed clinical observations;

- score on each developmental landmark (weight, sexual maturation and behavioural ontogeny) at each observation time;

- a detailed description of all behavioural, functional, neuropathological, neurochemical, electrophysiological findings by sex, including both increases and decreases from controls;

- necropsy findings;

- brain weights;

- any diagnoses derived from neurological signs and lesions, including naturally-occurring diseases or conditions;

- images of exemplar findings;

- low-power images to assess homology of sections used for morphometry;

- absorption and metabolism data, including complementary data from a separate toxicokinetic study, if available;

- statistical treatment of results, including statistical models used to analyse the data, and the results, regardless of whether they were significant or not;

- list of study personnel, including professional training.

Discussion of results:

- dose response information, by sex and group;

- relationship of any other toxic effects to a conclusion about the neurotoxic potential of the test chemical, by sex and group;

- impact of any toxicokinetic information on the conclusions;

- similarities of effects to any known neurotoxicants;

- data supporting the reliability and sensitivity of the test method (*i.e.,* positive and historical control data);

- relationships, if any, between neuropathological and functional effects;

- NOAEL or benchmark dose for dams and offspring, by sex and group.

Conclusions:

- a discussion of the overall interpretation of the data based on the results, including a conclusion of whether or not the test chemical caused developmental neurotoxicity and the NOAEL.

LITERATURE

(1) OECD (1995). Draft Report of the OECD *Ad Hoc* Working Group on Reproduction and Developmental Toxicity. Copenhagen, Denmark, 13-14 June 1995.

(2) US EPA (1998). U.S. Environmental Protection Agency Health Effects Test Guidelines. OPPTS 870.6300. Developmental Neurotoxicity Study. US EPA 712-C-98-239. Available: [http://www.epa.gov/opptsfrs/OPPTS_Harmonized/870_Health_Effects_ Test_Guidelines/Series/].

(3) US EPA (1998). Guidelines for Neurotoxicity Risk Assessment. US EPA 630/R-95/001F. Available: [http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?PrintVersion=True&d eid=12479].

(4) Cory-Slechta, D.A., Crofton, K.M., Foran, J.A., Ross, J.F., Sheets, L.P., Weiss, B., Mileson, B. (2001). Methods to identify and characterize developmental neurotoxicity for human health risk assessment: I. Behavioral effects. Environ. Health Perspect., 109:79-91.

(5) Dorman, D.C., Allen, S.L., Byczkowski, J.Z., Claudio, L., Fisher, J.E. Jr., Fisher, J.W., Harry, G.J., Li, A.A., Makris, S.L., Padilla, S., Sultatos, L.G., Mileson, B.E. (2001). Methods to identify and characterize developmental neurotoxicity for human health risk assessment: III. Pharmacokinetic and pharmacodynamic considerations. Environ. Health Perspect., 109:101-111.

(6) Garman, R.H., Fix,A.S., Jortner, B.S., Jensen, K.F., Hardisty, J.F., Claudio, L., Ferenc, S. (2001). Methods to identify and characterize developmental neurotoxicity for human health risk assessment: II. Neuropathology. Environ. Health Perspect., 109:93-100.

(7) OECD (2003). Report of the OECD Expert Consultation Meeting on Developmental Neurotoxicity Testing. Washington D.C., US, 23-25 October 2000.

(8) OECD (2008). OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 43. Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment. Environment Directorate, OECD, Paris. July 2008 Available: [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=en v/jm/mono(2008)16&doclanguage=en].

(9) OECD (2003). OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 20. Guidance Document for Neurotoxicity Testing. Environment Directorate, OECD, Paris, September

2003. Available: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html].

(10) Kimmel, C.A., Rees, D.C., Francis, E.Z. (1990) Qualitative and quantitative comparability of human and animal developmental neurotoxicity. Neurotoxicol. Teratol., 12: 173-292.

(11) Spencer, P.S., Schaumburg, H.H., Ludolph, A.C. (2000) *Experimental and Clinical Neurotoxicology, 2nd Edition,* ISBN 0195084772, Oxford University Press, New York.

(12) Mendola, P., Selevan, S.G., Gutter, S., Rice, D. (2002) Environmental factors associated with a spectrum of neurodevelopmental deficits. Ment. Retard. Dev. Disabil. Res. Rev. 8:188-197.

(13) Slikker, W.B., Chang, L.W. (1998) *Handbook of Developmental Neurotoxicology, 1st Edition*, ISBN 0126488606, Academic Press, New York.

(14) Chapter B.34 of this Annex, One-generation reproduction toxicity study.

(15) Chapter B.35 of this Annex, Two-generation reproduction toxicity study.

(16) Chapter B.43 of this Annex, Neurotoxicity Study in Rodents.

(17) Chapter B.31 of this Annex, Prenatal developmental toxicity study.

(18) Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33

(19) WHO (1986) *Principles and Methods for the Assessment of Neurotoxicity Associated with Exposure to Chemicals*, (Environmental Health Criteria 60), Albany, New York: World Health Organization Publications Center, USA. Available: [http://www.inchem.org/documents/ehc/ehc/ehc060.htm].

(20) WHO (2001) *Neurotoxicity Risk Assessment for Human Health: Principles and Approaches,* (Environmental Health Criteria 223), World Health Organization Publications, Geneva. Available: [http://www.intox.org/databank/documents/supplem/supp/ehc223.htm].

(21) Chang, L.W., Slikker, W. (1995) *Neurotoxicology: Approaches and Methods, 1st Edition*, ISBN 012168055X, Academic Press, New York.

(22) De Cabo, C., Viveros, M.P. (1997) Effects of neonatal naltrexone on neurological and somatic development in rats of both genders. Neurotoxicol. Teratol., 19:499-509.

(23) Agnish, N.D., Keller, K.A. (1997) The rationale for culling of rodent litters. Fundam. Appl. Toxicol., 38:2-6.

(24) Avery, D.L., Spyker, J.M. (1977) Foot tattoo of neonatal mice. Lab. Animal Sci., 27:110-112.

(25) Wier, P.J., Guerriero, F.J., Walker, R.F. (1989) Implementation of a primary screen for developmental neurotoxicity. Fundam. Appl. Toxicol., 13:118-136.

(26) Spear, N.E., Campbell, B.A. (1979) *Ontogeny of Learning and Memory*. ISBN 0470268492, Erlbaum Associates, New Jersey.

(27) Krasnegor, N.A., Blass, E.M., Hofer, M.A., Smotherman, W. (1987) *Perinatal Development: A Psychobiological Perspective.* Academic Press, Orlando.

(28) Zoetis, T., Walls, I. (2003) *Principles and Practices for Direct Dosing of Pre-Weaning Mammals in Toxicity Testing and Research.* ILSI Press, Washington, DC.

(29) Moser, V., Walls, I., Zoetis, T. (2005) Direct dosing of preweaning rodents in toxicity testing and research: Deliberations of an ILSI RSI expert working group. Int. J. Toxicol., 24:87-94.

(30) Conolly, R.B., Beck, B.D., Goodman, J.I. (1999) Stimulating research to improve the scientific basis of risk assessment. Toxicol. Sci., 49: 1-4.

(31) ICH (1993) ICH Harmonised Tripartite Guideline: Detection of Toxicity to Reproduction for Medical Products (S5A). International Conference on Harmonisation of Technical Requirements for Registration of Phamaceuticals for Human Use.

(32) Lochry, E.A. (1987) Concurrent use of behavioral/functional testing in existing reproductive and developmental toxicity screens: Practical considerations. J. Am. Coll. Toxicol., 6:433-439.

(33) Tachibana, T., Narita, H., Ogawa, T., Tanimura, T. (1998) Using postnatal age to determine test dates leads to misinterpretation when treatments alter gestation length, results from a collaborative behavioral teratology study in Japan. Neurotoxicol. Teratol., 20:449-457.

(34) Gallavan, R.H. Jr., Holson, J.F., Stump, D.G., Knapp, J.F., Reynolds, V.L. (1999) Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. Reprod. Toxicol., 13:383-390.

(35) Gray, L.E. Jr., Ostby, J., Furr, J., Price, M., Veeramachaneni, D.N., Parks, L. (2000) Perinatal exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. Toxicol. Sci., 58:350-365.

(36) Adams, J., Buelke-Sam, J., Kimmel, C.A., Nelson, C.J., Reiter, L.W., Sobotka, T.J., Tilson, H.A., Nelson, B.K. (1985) Collaborative behavioral teratology study: Protocol design and testing procedure. Neurobehav. Toxicol. Teratol., 7:579-586.

(37) Korenbrot, C.C., Huhtaniemi, I.T., Weiner, R.W. (1977) Preputial separation as an external sign of pubertal development in the male rat. Biol. Reprod., 17:298-303.

(38) Spear, L.P. (1990) Neurobehavioral assessment during the early postnatal period. Neurotoxicol. Teratol., 12:489-95.

(39) Altman, J., Sudarshan, K. (1975) Postnatal development of locomotion in the laboratory rat. Anim. Behav., 23:896-920.

(40) Adams, J. (1986) Methods in Behavioral Teratology. In: *Handbook of Behavioral Teratology.* Riley, E.P., Vorhees, C.V. (eds.) Plenum Press, New York, pp. 67-100.

(41) Reiter, L.W., MacPhail, R.C. (1979) Motor activity: A survey of methods with potential use in toxicity testing. Neurobehav. Toxicol., 1:53-66.

(42) Robbins, T.W. (1977) A critique of the methods available for the measurement of spontaneous motor activity, *Handbook of Psychopharmacology*, Vol. 7, Iverson, L.L., Iverson, D.S., Snyder, S.H., (eds.) Plenum Press, New York, pp. 37-82.

(43) Crofton, K.M., Peele, D.B., Stanton, M.E. (1993) Developmental neurotoxicity following neonatal exposure to 3,3'-iminodipropionitrile in the rat. Neurotoxicol. Teratol., 15:117-129.

(44) Ruppert, P.H., Dean, K.F., Reiter, L.W. (1985) Development of locomotor activity of rat pups in figure-eight mazes. Dev. Psychobiol., 18:247-260.

(45) Crofton, K.M., Howard, J.L., Moser, V.C., Gill, M.W., Reiter, L.W., Tilson, H.A., MacPhail, R.C. (1991) Interlaboratory comparison of motor activity experiments: Implications for neurotoxicological assessments. Neurotoxicol. Teratol., 13:599-609.

(46) Ross, J. F., Handley, D. E., Fix, A. S., Lawhorn, G. T., Carr, G. J. (1997) Quantification of the hind-limb extensor thrust response in rats. Neurotoxicol. Teratol., 19:1997. 405-411.

(47) Handley, D.E., Ross, J.F., Carr, G.J. (1998) A force plate system for measuring low-magnitude reaction forces in small laboratory animals. Physiol. Behav., 64:661-669.

(48) Edwards, P.M., Parker, V.H. (1977) A simple, sensitive, and objective method for early assessment of acrylamide neuropathy in rats. Toxicol. Appl. Pharmacol., 40:589-591.

(49) Davis, M. (1984) The mammalian startle response. In: *Neural Mechanisms of Startle Behavior*, Eaton, R.C. (ed), Plenum Press, New York, pp. 287-351

(50) Koch, M. (1999) The neurobiology of startle. Prog. Neurobiol., 59:107-128.

(51) Crofton, K.M. (1992) Reflex modification and the assessment of sensory dysfunction. In *Target Organ Toxicology Series: Neurotoxicology*, Tilson, H., Mitchell, C. (eds). Raven Press, New York, pp. 181-211.

(52) Crofton, K.M., Sheets, L.P. (1989) Evaluation of sensory system function using reflex modification of the startle response. J. Am. Coll. Toxicol., 8:199-211.

(53) Crofton, K.M, Lassiter, T.L, Rebert, C.S. (1994) Solvent-induced ototoxicity in rats: An atypical selective mid-frequency hearing deficit. Hear. Res., 80:25-30.

(54) Ison, J.R. (1984) Reflex modification as an objective test for sensory processing following toxicant exposure. Neurobehav. Toxicol. Teratol., 6:437–445.

(55) Mattsson, J.L., Boyes, W.K., Ross, J.F. (1992) Incorporating evoked potentials into neurotoxicity test schemes. In: *Target Organ Toxicology Series: Neurotoxicity*, Tilson, H., Mitchell, C., (eds.), Raven Press, New York. pp. 125-145.

(56) Peele, D.B., Allison, S.D., Crofton, K.M. (1990) Learning and memory deficits in rats following exposure to 3,3'-iminopropionitrile. Toxicol. Appl. Pharmacol., 105:321-332.

(57) Bammer, G. (1982) Pharmacological investigations of neurotransmitter involvement in passive avoidance responding: A review and some new results. Neurosci. Behav. Rev., 6:247-296.

(58) Bushnell, P.J. (1988) Effects of delay, intertrial interval, delay behavior and trimethyltin on spatial delayed response in rats. Neurotoxicol. Teratol., 10:237-244.

(59) Green, R.J., Stanton, M.E. (1989) Differential ontogeny of working memory and reference memory in the rat. Behav. Neurosci., 103:98-105.

(60) Kucharski, D., Spear, N.E. (1984) Conditioning of aversion to an odor paired with peripheral shock in the developing rat. Develop. Psychobiol., 17:465-479.

(61) Morris, R. (1984) Developments of a water-maze procedure for studying spatial learning in the rat. J. Neurosci. Methods, 11:47-60.

(62) Brandeis, R., Brandys, Y., Yehuda, S. (1989) The use of the Morris water maze in the study of memory and learning. Int. J. Neurosci., 48:29-69.

(63) D'Hooge, R., De Deyn, P.P. (2001) Applications of the Morris water maze in the study of learning and memory. Brain Res. Rev, 36:60-90.

(64) Vorhees, C.V. (1987) Maze learning in rats: A comparison of performance in two water mazes in progeny prenatally exposed to different doses of phenytoin. Neurotoxicol. Teratol., 9:235-241.

(65) Vorhees, C.V. (1997) Methods for detecting long-term CNS dysfunction after prenatal exposure to neurotoxins. Drug Chem. Toxicol., 20:387-399.

(66) Akaike, M., Tanaka, K., Goto, M., Sakaguchi, T. (1988) Impaired Biel and Radial arm maze learning in rats with methyl-nitrosurea induced microcephaly. Neurotoxicol. Teratol., 10:327-332.

(67) Cory-Slechta, D.A., Weiss, B., Cox, C. (1983) Delayed behavioral toxicity of lead with increasing exposure concentration. Toxicol. Appl. Pharmacol., 71:342-352.

(68) Campbell, B.A., Haroutunian, V. (1981) Effects of age on long-term memory: Retention of fixed interval responding. J. Gerontol., 36:338–341.

(69) Fix, A.S, Garman, R.H. (2000) Practical aspects of neuropathology: A technical guide for working with the nervous system. Toxicol. Pathol., 28: 122-131.

(70) Prophet, E.B., Mills, B., Arrington, J.B., Sobin, L.H. (1994) *Laboratory Methods in Histotechnology*, American Registry of Pathology, Washington, DC, pp. 84-107.

(71) Bancroft, J.D., Gamble, M. (2002) *Theory and Practice of Histological Techniques*, 5th edition, Churchill Livingstone, London.

(72) Fix, A.S., Ross, J.F., Stitzel, S.R., Switzer, R.C. (1996) Integrated evaluation of central nervous system lesions: stains for neurons, astrocytes, and microglia reveal the spatial and temporal features of MK-801-induced neuronal necrosis in the rat cerebral cortex. Toxicol. Pathol., 24: 291-304.

(73) Schmued, L.C., Hopkins, K.J. (2000) Fluoro-Jade B: A high affinity tracer for the localization of neuronal degeneration. Brain Res., 874:123-130.

(74) Krinke, G.J., Classen, W., Vidotto, N., Suter, E., Wurmlin, C.H. (2001) Detecting necrotic neurons with fluoro-jade stain. Exp. Toxic. Pathol., 53:365-372.
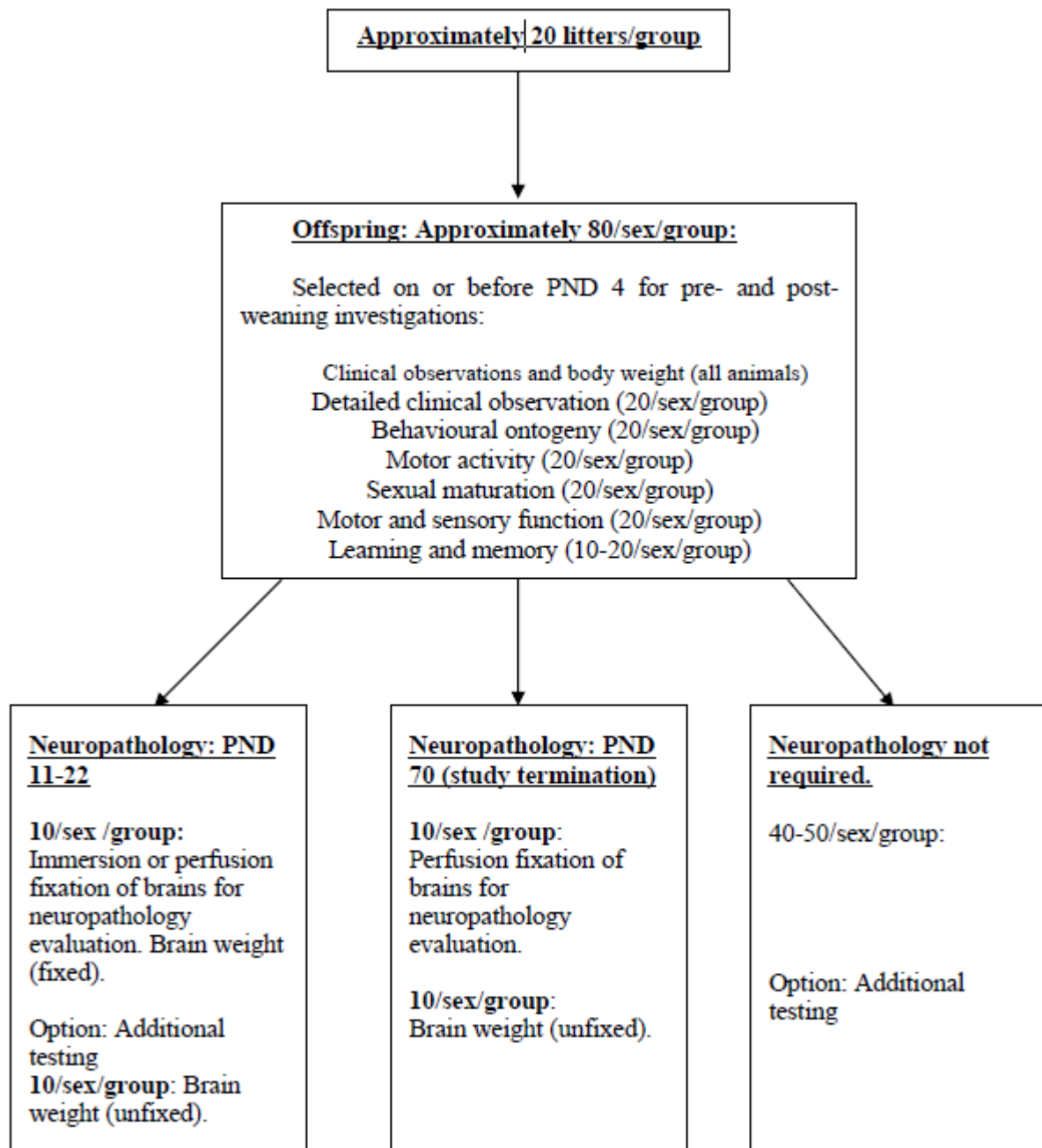
(75) De Olmos, I.S., Beltramino, C.A., and de Olmos de Lorenzo, S. (1994) Use of an amino-cupric-silver technique for the detection of early and semiacute neuronal degeneration caused by neurotoxicants, hypoxia and physical trauma. Neurotoxicol. Teratol., 16, 545-561.

(76) De Groot, D.M.G., Bos-Kuijpers, M.H.M., Kaufmann, W.S.H., Lammers, J.H.C.M., O'Callaghan, J.P., Pakkenberg, B., Pelgrim, M.T.M., Waalkens-Berendsen, I.D.H., Waanders, M.M., Gundersen, H.J. (2005a) Regulatory developmental neurotoxicity testing: A model study focusing on conventional neuropathology endpoints and other perspectives. Environ. Toxicol. Pharmacol., 19:745-755.

(77) De Groot, D.M.G., Hartgring, S., van de Horst, L., Moerkens, M., Otto, M., Bos-Kuijpers, M.H.M., Kaufmann, W.S.H., Lammers, J.H.C.M., O'Callaghan, J.P., Waalkens-Berendsen, I.D.H., Pakkenberg, B., Gundersen, H.J. (2005b) 2D and 3D assessment of neuropathology in rat brain after prenatal exposure to methylazoxymethanol, a model for developmental neurotoxicity. Reprod. Toxicol., 20:417-432.

(78) Rodier, P.M., Gramann, W.J. (1979) Morphologic effects of interference with cell proliferation in the early fetal period. Neurobehav. Toxicol., 1:129–135.

(79) Howard, C.V., Reed, M.G. (1998) *Unbiased Stereology: Three-Dimensional Measurement in Microscopy*, Springer-Verlag, New York.

(80) Hyman, B.T., Gomez-Isla, T., Irizarry, M.C. (1998) Stereology: A practical primer for neuropathology. J. Neuropathol. Exp. Neurol., 57: 305-310.

(81) Korbo, L., Andersen, B.B., Ladefoged, O., Møller, A. (1993) Total numbers of various cell types in rat cerebellar cortex estimated using an unbiased stereological method. Brain Res., 609: 262-268.

(82) Schmitz, C. (1997) Towards more readily comprehensible procedures in disector stereology. J. Neurocytol., 26:707-710.

(83) West, M.J. (1999) Stereological methods for estimating the total number of neurons and synapses: Issues of precision and bias. Trends Neurosci., 22:51-61.

(84) Schmitz, C., Hof, P.R. (2005) Design-based stereology in neuroscience. Neuroscience, 130: 813–831.

(85) Gavin, C.E., Kates, B., Gerken, L.A., Rodier, P.M. (1994) Patterns of growth deficiency in rats exposed *in utero* to undernutrition, ethanol, or the neuroteratogen methylazoxymethanol (MAM). Teratology, 49:113-121.

(86) Ohno, M., Aotani, H., Shimada, M. (1995) Glial responses to hypoxic/ischemic encephalopathy in neonatal rat cerebrum. Develop. Brain Res., 84:294-298.

(87) Jensen KF, Catalano SM. (1998) Brain morphogenesis and developmental neurotoxicology. In: *Handbook of Developmental Neurotoxicology*, Slikker, Jr. W., Chang, L.W. (eds) Academic Press, New York, pp. 3-41.

(88) Ikonomidou, C., Bosch, F., Miksa, M., Bittigau, P., Vöckler, J., Dikranian, K., Tenkova, T.I., Stefovska, V., Turski, L., Olney, J.W. (1999) Blockade of NMDA receptors and apoptotic neurodegeneration in the developing brain. Science, 283:70-74.

(89) Ikonomidou, C., Bittigau, P., Ishimaru, M.J., Wozniak, D.F., Koch, C., Genz, K., Price, M.T., Sefovska, V., Hörster, F., Tenkova, T., Dikranian, K., Olney, J.W. (2000) Ethanol-induced apoptotic degeneration and fetal alcohol syndrome. Science, 287:1056–1060.

(90) Friede, R. L. (1989) *Developmental Neuropathology*. Second edition. Springer-Verlag, Berlin.

(91) House, D.E., Berman, E., Seeley, J.C., Simmons, J.E. (1992) Comparison of open and blind histopathologic evaluation of hepatic lesions. Toxicol. Let., 63:127-133.

(92) Tilson, H.A., MacPhail, R.C., Crofton, K.M. (1996) Setting exposure standards: a decision process. Environ. Health Perspect., 104:401-405.

(93) US EPA (2005) Guidelines for Carcinogen Risk Assessment. US EPA NCEA-F-0644A.

(94) US EPA (1996) Guidelines for Reproductive Toxicity Risk Assessment, Federal Register 61(212): 56274-56322.

(95) Danish Environmental Protection Agency (1995) *Neurotoxicology*. Review of Definitions, Methodology, and Criteria. Miljøprojekt nr. 282. Ladefoged, O., Lam, H.R., Østergaard, G., Nielsen, E., Arlien-Søborg, P.

(96) Muller, K.E., Barton, C.N., Benignus, V.A. (1984). Recommendations for appropriate statistical practice in toxicologic experiments. Neurotoxicology, 5:113-126.

(97) Gad, S.C. (1989) Principles of screening in toxicology with special emphasis on applications to Neurotoxicology. J. Am. Coll. Toxicol., 8:21-27.

(98) Abby, H., Howard, E. (1973) Statistical procedures in developmental studies on a species with multiple offspring. Dev. Psychobiol., 6:329-335.

(99) Haseman, J.K., Hogan, M.D. (1975) Selection of the experimental unit in teratology studies. Teratology, 12:165-172.

(100)   Holson, R.R., Pearce, B. (1992) Principles and pitfalls in the analysis of prenatal treatment effects in multiparous species. Neurotoxicol. Teratol., 14: 221-228.

(101)   Nelson, C.J., Felton, R.P., Kimmel, C.A., Buelke-Sam, J., Adams, J. (1985) Collaborative Behavioral Teratology Study: Statistical approach. Neurobehav. Toxicol. Teratol., 7:587-90.

(102)   Crofton, K.M., Makris, S.L., Sette, W.F., Mendez, E., Raffaele, K.C. (2004) A qualitative retrospective analysis of positive control data in developmental neurotoxicity studies. Neurotoxicol. Teratol., 26:345-352.

(103)   Bolon, B., Garman, R., Jensen, K., Krinke, G., Stuart, B., and an *ad hoc* working group of the STP Scientific and Regulatory Policy Committee. (2006) A 'best practices' approach to neuropathological assessment in developmental neurotoxicity testing – for today. Toxicol. Pathol. 34:296-313.

(104)   Tamura, R.N., Buelke-Sam, J. (1992) The use of repeated measures analysis in developmental toxicology studies. Neurotoxicol. Teratol., 14(3):205-210.

(105)   Tukey, J.W., Ciminera, J.L., Heyse, J.F. (1985) Testing the statistical certainty of a response to increasing doses of a drug. Biometrics, 41:295-301.

(106)   Crofton, K.M., Foss, J.A., Haas, U., Jensen, K., Levin, E.D., and Parker, S.P. (2008) Undertaking positive control studies as part of developmental neurotoxicity testing: report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology*, 30(4):266-287.

(107)   Raffaele, K.C., Fisher, E., Hancock, S., Hazelden, K., and Sobrian, S.K. (2008) Determining normal variability in a developmental neurotoxicity test: report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology,* 30(4):288-325.

(108)   Holson, R.R., Freshwater, L., Maurissen, J.P.J., Moser, V.C., and Phang, W. (2008) Statistical issues and techniques appropriate for developmental neurotoxicity testing: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology*, 30(4):326-348.

(109)   Tyl, R.W., Crofton, K.M., Moretto, A., Moser, V.C., Sheets, L.P., and Sobotka, T.J.   (2008) Identification and interpretation of developmental neurotoxicity effects: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints   *Neurotoxicology and Teratology*, 30(4):349-381.

Figure 1. General testing scheme for functional/behavioural tests, neuropathology evaluation, and brain weights. This diagram is based on the description in paragraphs 13-15 (PND=postnatal day). Examples of animal assignment are given in Appendix 1.

```
┌─────────────────────────────────┐
│ Approximately 20 litters/group  │
└─────────────────────────────────┘
                 │
                 ▼
┌──────────────────────────────────────────────────┐
│ Offspring: Approximately 80/sex/group:            │
│                                                    │
│   Selected on or before PND 4 for pre- and post-   │
│ weaning investigations:                            │
│                                                    │
│   Clinical observations and body weight (all animals) │
│   Detailed clinical observation (20/sex/group)     │
│   Behavioural ontogeny (20/sex/group)              │
│   Motor activity (20/sex/group)                    │
│   Sexual maturation (20/sex/group)                 │
│   Motor and sensory function (20/sex/group)        │
│   Learning and memory (10-20/sex/group)            │
└──────────────────────────────────────────────────┘
```

Neuropathology: PND 11-22	Neuropathology: PND 70 (study termination)	Neuropathology not required.
10/sex /group: Immersion or perfusion fixation of brains for neuropathology evaluation. Brain weight (fixed).	10/sex /group: Perfusion fixation of brains for neuropathology evaluation.	40-50/sex/group:
Option: Additional testing 10/sex/group: Brain weight (unfixed).	10/sex/group: Brain weight (unfixed).	Option: Additional testing

Appendix 1

1. Examples of possible assignments are described and tabulated below. These examples are provided to illustrate that assignment of study animals to various testing paradigms can be accomplished in a number of different ways.

Example 1

2. One set of 20 pups/sex/dose level (i.e., 1 male and 1 female per litter) is used for pre-weaning testing of behavioural ontogeny. Out of these animals, 10 pups/sex/dose level (i.e., 1 male or 1 female per litter) are humanely killed at PND 22. The brains are removed, weighed and processed for histopathologic evaluation. In addition, brain weight data are collected using unfixed brains from the remaining 10 males and 10 females per dose level.

3. Another set of 20 animals/sex/dose level (i.e., 1 male and 1 female per litter) is used for post-weaning functional/behavioral tests (detailed clinical observations, motor activity, auditory startle and cognitive function testing in adolescents) and assessing age of sexual maturation. Of these animals, 10 animals/sex/dose level (i.e., 1 male or 1 female per litter), are anesthetised and fixed via perfusion at study termination (approximately PND 70). After additional fixation in situ, the brain is removed and processed for neuropathological evaluation.

4. For cognitive function testing in young adults (e.g., PND 60-70), a third set of 20 pups/sex/dose level is used (i.e., 1 male and 1 female per litter). Of these animals, 10 animals/sex/group (1 male or 1 female per litter) are killed at study termination and the brain is removed and weighed.

5. The remaining 20 animals/sex/group are reserved for possible additional tests.

Table 1

Pup no.[a]		No. of pups assigned to test	Examination / Test
m	f		
1	5	20 m + 20 f	Behavioural ontogeny
		10 m + 10 f	PND 22 brain weight/neuropathology/ morphometry
		10 m + 10 f	PND 22 brain weight

28

2	6	20 m + 20 f	Detailed clinical observations
		20 m + 20 f	Motor activity
		20 m + 20 f	Sexual maturation
		20 m + 20 f	Motor and sensory function
		20 m + 20 f	Learning and memory (PND 25)
		10 m + 10 f	Young adult brain weight/neuropathology/ morphometry ~PND 70
3	7	20 m + 20 f	Learning and memory (young adults)
		10 m + 10 f	Young adult brain weight ~ PND 70
4	8	- -	Reserve animals for replacements or additional tests

[a] For this example, litters are culled to 4 males + 4 females; male pups are numbered 1 through 4, female pups 5 through 8.

Example 2

6. One set of 20 pups/sex/dose level (*i.e.,* 1 male and 1 female per litter) is used for pre-weaning testing of behavioural ontogeny. Out of these animals, 10 pups/sex/dose level (1 male or 1 female per litter), are humanely killed at PND 11. The brains are removed, weighed and processed for histopathologic evaluation.

7. Another set of 20 animals/sex/dose level (1 male and 1 female per litter) is used for post-weaning examinations (detailed clinical observations, motor activity, assessing age of sexual maturation and motor and sensory function). Of these animals, 10 animals/sex/dose level (*i.e.,* 1 male or 1 female per litter) are anesthetised and fixed via perfusion at study termination (approximately PND 70). After additional fixation *in situ*, the brain is removed, weighed and processed for neuropathological evaluation.

8. For cognitive function testing in adolescents and young adults, 10 pups/sex/dose level are used (*i.e.,* 1 male or 1 female per litter). Different animals are used for testing for cognitive function tests at PND 23 and young adults. At termination, the 10 animals/sex/group tested as adults are killed, the brain is removed and weighed.

9. The remaining 20 animals/sex/group not selected for testing are killed and discarded at

weaning.

Table 2.

| Pup no.[a] | | No. of pups assigned to test | Examination / Test |
m	f		
1	5	20 m + 20 f	Behavioural ontogeny
		10 m + 10 f	PND 11 brain weight/ neuropathology/ morphometry
2	6	20 m + 20 f	Detailed clinical observations
		20 m + 20 f	Motor activity
		20 m + 20 f	Sexual maturation
		20 m + 20 f	Motor and sensory function
		10 m + 10 f	Young adult brain weight/neuropathology/ morphometry ~PND 70
3	7	10 m + 10 f [b]	Learning and memory (PND 23)
3	7	10 m + 10 f [b]	Learning and memory (young adults)
			Young adult brain weight
4	8	- -	Animals killed and discarded PND 21.

[a] For this example, litters are culled to 4 males + 4 females; male pups are numbered 1 through 4, female pups 5 through 8.

[b] Different pups are used for cognitive tests at PND 23 and in young adults (e.g., even /odd litters from total of 20).

Example 3

10. One set 20 pups/sex/dose level (*i.e.,* 1 male and 1 female per litter) is used for brain weight and neuropathology assessment at PND 11. Out of these animals, 10 pups/sex/dose level (*i.e.* 1 male or 1 female per litter) are humanely killed at PND 11 and brains are removed, weighed and processed for histopathologic evaluation. In addition, brain weight data are collected using unfixed brains from the remaining 10 males and 10 females per dose level.

11. Another set of of 20 animals/sex/dose level (*i.e.,* 1 male and 1 female per litter) are used for behavioural ontogeny (motor activity), post-weaning examinations (motor activity and assessing age of sexual maturation), and cognitive function testing in adolescents.

12. Another set of 20 animals/sex/dose level (*i.e.,* 1 male and 1 female per litter) is used for motor and sensory function tests (auditory startle) and detailed clinical observations. Of these animals, 10 animals/sex/dose level (*i.e.,* 1 male or 1 female per litter) are anesthetised and fixed via perfusion at study termination (approximately PND 70). After additional fixation *in situ*, the brain is removed, weighed and processed for neuropathological evaluation.

13. Another set of 20 pups/sex/dose level are used for cognitive function testing in young adults (*i.e.,* 1 male and 1 female per litter). Of these, 10 animals/sex/group (*i.e.,* 1 male or 1 female per litter) are killed at termination, the brain removed and weighed.

Table 3.

Pup no.[a]		No. of pups assigned to test	Examination / Test
m	f		
1	5	10 m + 10 f	PND 11 brain weight/ neuropathology/ morphometry
		10 m + 10 f	PND 11 brain weight
2	6	20 m + 20 f	Behavioural ontogeny (motor activity)
		20 m + 20 f	Motor activity
		20 m + 20 f	Sexual maturation
		20 m + 20 f	Learning and memory (PND 27)
3	7	20 m + 20 f	Auditory startle (adolescents and young adults)
		20 m + 20 f	Detailed clinical observations
		10 m + 10 f	Young adult brain weight/neuropathology/ morphometry ~PND 70
4	8	20 m + 20 f	Learning and memory (young adults)
		10 m + 10 f	Young adult brain weight

[a] For this example, litters are culled to 4 males + 4 females; male pups are numbered 1

31

through 4, female pups 5 through 8.

Appendix 2

Definitions

Chemical: A substance or a mixture

Test chemical: Any substance or mixture tested using this test method

B.54. Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 440 (2007). The OECD initiated a high-priority activity in 1998 to revise existing guidelines and to develop new guidelines for the screening and testing of potential endocrine disrupters (1). One element of the activity was to develop a test guideline for the rodent Uterotrophic Bioassay. The rodent Uterotrophic Bioassay then underwent an extensive validation programme including the compilation of a detailed background document (2)(3) and the conduct of extensive intra- and interlaboratory studies to show the relevance and reproducibility of the bioassay with a potent reference oestrogen, weak oestrogen receptor agonists, a strong oestrogen receptor antagonist, and a negative reference chemical (4)(5)(6)(7)(8)(9). This test method B.54 is the outcome of the experience gained during the validation test programme and the results obtained thereby with oestrogenic agonists.

2. The Uterotrophic Bioassay is a short-term screening test that originated in the 1930s (27)(28) and was first standardised for screening by an expert committee in 1962 (32)(35). It is based on the increase in uterine weight or uterotrophic response (for review, see 29). It evaluates the ability of a chemical to elicit biological activities consistent with agonists or antagonists of natural oestrogens (e.g. 17ß-estradiol), however, its use for antagonist detection is much less common than for agonists. The uterus responds to oestrogens in two ways. An initial response is an increase in weight due to water imbibition. This response is followed by a weight gain due to tissue growth (30). The uterus responses in rats and mice qualitatively are comparable.

3. This bioassay serves as an *in vivo* screening assay and its application should be seen in the context of the "OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals" (Appendix 2). In this Conceptual Framework the Uterotrophic Bioassay is contained in Level 3 as an *in vivo* assay providing data about a single endocrine mechanism, i.e. oestrogenicity.

4. The Uterotrophic Bioassay is intended to be included in a battery of *in vitro* and *in vivo* tests to identify chemicals with potential to interact with the endocrine system, ultimately leading to risk assessments for human health or the environment. The OECD validation programme used both strong and weak oestrogen agonists to evaluate the performance of the assay to identify oestrogenic chemicals (4)(5)(6)(7)(8). Thereby the sensitivity of the test procedure for oestrogen agonists was well demonstrated besides a good intra- and interlaboratory reproducibility.

5. With regard to negative compounds, only one "negative" reference chemical already reported negative by uterotrophic assay as well as *in vitro* receptor binding and receptor assays was included in the validation programme, but additional test data, not related to the OECD validation programme, have been evaluated, giving further support to the specificity of the Uterotrophic Bioassay for the screening of oestrogen agonists (16).

INITIAL CONSIDERATIONS AND LIMITATIONS

6. Oestrogen agonists and antagonists act as ligands for oestrogen receptors α and β and may activate or inhibit, respectively, the transcriptional action of the receptors. This may have the potential to lead to adverse health hazards, including reproductive and developmental effects. Therefore, the need exists to rapidly assess and evaluate a chemical as a possible oestrogen agonist or antagonist. While informative, the affinity of a ligand for an oestrogen receptor or transcriptional activation of reporter genes *in vitro* is only one of several determinants of possible hazard. Other determinants can include metabolic activation and deactivation upon entering the body, distribution to target tissues, and clearance from the body, depending at least in part on the route of administration and the chemical being tested. This leads to the need to screen the possible activity of a chemical *in vivo* under relevant conditions, unless the chemical's characteristics regarding Absorption – Distribution – Metabolism – Elimination (ADME) already provide appropriate information. Uterine tissues respond with rapid and vigorous growth to stimulation by oestrogens, particularly in laboratory rodents, where the oestrous cycle lasts approximately 4 days. Rodent species, particularly the rat, are also widely used in toxicity studies for hazard characterisation. Therefore, the rodent uterus is an appropriate target organ for the *in vivo* screening of oestrogen agonists and antagonists.

7. This test method is based on those protocols employed in the OECD validation study which have been shown to be reliable and repeatable in intra- and interlaboratory studies (5)(7). Currently two methods, namely the ovariectomised adult female method (ovx-adult method) and the immature non-ovariectomised method (immature method) are available. It was shown in the OECD validation test programme that both methods have comparable sensitivity and reproducibility. However, the immature, as it has an intact hypothalamic-pituitary-gonadal (HPG) axis, is somewhat less specific but covers a larger scope of investigation than the ovariectomized animal because it can respond to chemicals that interact with the HPG axis rather than just the oestrogen receptor. The HGP axis of the rat is functional at about 15 days of age. Prior to that, puberty cannot be accelerated with treatments like GnRH. As the females begin to reach puberty, prior to vaginal opening, the female will have several silent cycles that do not result in vaginal opening or ovulation, but there are some hormonal fluctuations. If a chemical stimulates the HPG axis directly or indirectly, precocious puberty, early ovulation and accelerated vaginal opening result. Not only chemicals that act on the HPG axis do this but some diets with higher metabolisable energy levels than others will stimulate growth and accelerate vaginal opening without being oestrogenic. Such chemicals would not induce an uterotrophic response in OVX adult animals as their HPG axis does not work.

8. For animal welfare reasons preference should be given to the method using immature rats, avoiding surgical pre-treatment of the animals and avoiding also a possible non-use of those animals which indicate any evidence entering oestrous (see paragraph 30).

9. The uterotrophic response is not entirely of oestrogenic origin, i.e. chemicals other than agonists or antagonists of oestrogens may also provide a response. For example, relatively high doses of progesterone, testosterone, or various synthetic progestins may all lead to a stimulative response (30). Any response may be analysed histologically for keratinisation and cornification of the vagina (30). Irrespective of the possible origin of

the response, a positive outcome of an Uterotrophic Bioassay should normally initiate actions for further clarification. Additional evidence of oestrogenicity could come from *in vitro* assays, such as the ER binding assays and transcriptional activation assays, or from other *in vivo* assays such as the female pubertal assay.

10. Taking into account that the Uterotrophic Bioassay serves as an *in vivo* screening assay, the validation approach taken served both animal welfare considerations and a tiered testing strategy. To this end, effort was directed at rigorously validating reproducibility and sensitivity for oestrogenicity - the main concern for many chemicals-, while little effort was directed at the antioestrogenicity component of the assay. Only one antioestrogen with strong activity was tested since the number of chemicals with a clear antioestrogenic profile (not obscured by some oestrogenic activity) is very limited. Thus this test method is dedicated to the oestrogenic protocol, while the protocol describing the antagonist mode of the assay is included in a Guidance Document (37). The reproducibility and sensitivity of the assay for chemicals with purely anti-oestrogenic activity will be more clearly defined later on, after the test procedure has been in routine use for some time and more chemicals with this modality of action are identified.

11. It is acknowledged that all animal based procedures will conform to local standards of animal care; the descriptions of care and treatment set forth below are minimal performance standards, and will be superseded by local regulations such as Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes (38). Further guidance of the humane treatment of animals is given by the OECD (25).

12. As with all assays using live animals, it is essential to ensure that the data are truly necessary prior to the start of the assay. For example, two conditions where the data may be required are:

    - high exposure potential (Level 1 of the Conceptual Framework, Appendix 2) or indications for oestrogenicity (Level 2) to investigate whether such effects may occur *in vivo*;

    - effects indicating oestrogenicity in Level 4 or 5 *in vivo* tests to substantiate that the effects were related to an oestrogenic mechanism that cannot be elucidated using an *in vitro* test.

13. Definitions used in this test method are given in Appendix 1.

PRINCIPLE OF THE TEST

14. The Uterotrophic Bioassay relies for its sensitivity on an animal test system in which the hypothalamic-pituitary-ovarian axis is not functional, leading to low endogenous levels of circulating oestrogen. This will ensure a low baseline uterine weight and a maximum range of response to administered oestrogens. Two oestrogen sensitive states in the female rodent meet this requirement:

    i) immature females after weaning and prior to puberty, and

    ii) young adult females after ovariectomy with adequate time for uterine tissues

to regress.

15. The test chemical is administered daily by oral gavage or subcutaneous injection. Graduated test chemical doses are administered to a minimum of two treatment groups (see paragraph 33 for guidance) of experimental animals using one dose level per group and an administration period of three consecutive days for immature method and a minimum administration period of three consecutive days for ovx-adult method. The animals are necropsied approximately 24 hours after the last dose. For oestrogen agonists, the mean uterine weight of the treated animal groups relative to the vehicle group is assessed for a statistically significant increase. A statistically significant increase in the mean uterine weight of a test group indicates a positive response in this bioassay.

DESCRIPTION OF THE METHOD

Selection of animal species

16. Commonly used laboratory rodent strains may be used. As an example, Sprague-Dawley and Wistar strains of rats were used during the validation. Strains with uteri known or suspected to be less responsive should not be used. The laboratory should demonstrate the sensitivity of the strain used as described in paragraphs 26 and 27.

17. The rat and mouse have been routinely used in the Uterotrophic Bioassay since the 1930s. The OECD validation studies were only performed with rats based on an understanding that both species are expected to be equivalent and therefore one species should be enough for the world-wide validation in order to save resources and animals. The rat is the species of choice in most reproductive and developmental toxicity studies. Taking into consideration that a vast historical database exists for mice and thus to broaden the scope of the Uterotrophic Bioassay test method in rodents to the use of mice as test species, a limited follow-up validation study was carried out in mice (16). A bridging approach with a limited number of test chemicals, participating laboratories and without coded sample testing has been selected in keeping with the original intent to save resources and animals. This bridging validation study shows for the Uterotrophic Bioassay in young adult ovariectomized mice that, qualitatively and quantitatively, the data obtained in rats and mice correspond well with each other. Where the Uterotrophic Bioassay result may be preliminary to a long-term study, this allows animals from the same strain and source to be used in both studies. The bridging approach was limited to the OVX mice and the report does not provide a robust data set to validate the immature model, thus the immature model for mice is not considered under the scope of the current test method.

18. Thus, in some cases mice may be used instead of rats. A rationale should be given for this species, based on toxicological, pharmacokinetic, and/or other criteria. Modifications of the protocol may be necessary for mice. For example, the food consumption of mice on a body weight basis is higher than that of rats and therefore the phyto-oestrogen content in food should be lower for mice than for rats (9)(20)(22).

Housing and feeding conditions

19. All procedures should conform with local standards of laboratory animal care. These descriptions of care and treatment are minimum standards and will be superseded by local regulations such as Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes (38) . The temperature in the experimental animal room should be 22°C (with an approximate range ± 3°C). The relative humidity should be a minimum of 30% and preferably should not exceed a maximum 70%, other than during room cleaning. The aim should be relative humidity of 50-60%. Lighting should be artificial. The daily lighting sequence should be 12 hours light, 12 hours dark.

20. Laboratory diet and drinking water should be provided *ad libitum*. Young adult animals may be housed individually or be caged in groups of up to three animals. Due to the young age of the immature animals, social group housing is recommended.

21. High levels of phyto-oestrogens in laboratory diets have been known to increase uterine weights in rodents to a degree enough as to interfere with the Uterotrophic Bioassay (13)(14)(15). High levels of phyto-oestrogens and of metabolisable energy in laboratory diets may also result in early puberty, if immature animals are used. The presence of phyto-oestrogens results primarily from the inclusion of soy and alfalfa products in the laboratory diets and concentrations of phyto-oestrogens have been shown to vary from batch-to-batch of standard laboratory diets (23). Body weight is an important variable, as the quantity of food consumed is related to body weight. Therefore, the actual phyto-oestrogen dose consumed from the same diet may vary among species and by age (9). For immature female rats, food consumption on a body weight basis may be approximately double that of ovariectomised young adult females. For young adult mice, food consumption on a body weight basis may be approximately quadruple that of ovariectomised young adult female rats.

22. Uterotrophic Bioassay results (9)(17)(18)(19), however, show that limited quantities of dietary phyto-oestrogens are acceptable and do not reduce the sensitivity of the bioassay. As a guide, dietary levels of phyto-oestrogens should not exceed 350 μg of genistein equivalents/gram of laboratory diet for immature female Sprague Dawley and Wistar rats (6)(9). Such diets should also be appropriate when testing in young adult ovariectomised rats because food consumption on a body weight basis is less in young adult as compared to immature animals. If adult ovariectomised mice or more phyto-oestrogen-sensitive rats are to be used, proportional reduction in dietary phyto-oestrogen levels must be considered (20). In addition, the differences in available metabolic energy from different diets may lead to time shifts for the onset of puberty (21)(22).

23. Prior to the study, careful selection is required of a diet without an elevated level of phyto-oestrogens (for guidance see (6)(9)) or metabolisable energy, that can confound the results (15)(17)(19)(22)(36). Ensuring the proper performance of the test system used by the laboratory as specified in paragraphs 26 and 27 is an important check on both of these factors. As a safeguard consistent with good laboratory practice (GLP) representative sampling of each batch of diet administered during the study should be conducted for possible analysis of phyto-oestrogen content (e.g. in the case of high uterine control weight relative to historic controls or an inadequate response to the reference oestrogen, 17 alpha ethinyl estradiol). Aliquots should be analysed as part of the study or frozen at -20°C or in such a way as to prevent the sample from decomposing

prior to analysis.

24. Some bedding materials may contain naturally occurring oestrogenic or antioestrogenic chemicals (e.g. corn cob is known to affects the cyclicity of rats and appears to be antioestrogenic). The selected bedding material should contain a minimum level of phyto-oestrogens.

Preparation of animals

25. Experimental animals without evidence of any disease or physical abnormalities are randomly assigned to the control and treatment groups. Cages should be arranged in such a way that possible effects due to cage placement are minimised. The animals should be identified uniquely. Preferably, immature animals should be caged with dams or foster dams until weaning during acclimatisation. The acclimatisation period prior to the start of the study should be about 5 days for young adult animals and for the immature animals delivered with dams or foster dams. If immature animals are obtained as weanlings without dams a shorter duration of the acclimatisation period may become necessary as dosing should start immediately after weaning (see paragraph 29).

PROCEDURE

Verification of Laboratory Proficiency

26. Two different options can be used to verify laboratory proficiency:

    - Periodic verification, relying on an initial baseline positive control study (see paragraph 27). At least every 6 months and each time there is a change that may influence the performance of the assay (e.g. a new formulation of diet, change in personnel performing dissections, change in animal strain or supplier, etc.), the responsiveness of the test system (animal model) should be verified using an appropriate dose (based on the baseline positive control study described in paragraph 27) of a reference oestrogen: 17α-ethinyl estradiol (CAS No. 57-63-6) (EE).

    - Use of concurrent controls, by including a group administered with an appropriate dose of reference oestrogen in each assay.

    If the system does not respond as expected, the experimental conditions should be examined and modified accordingly. It is recommended that the dose of reference oestrogen to be used in either approach be approximately the ED70 to 80.

27. **Baseline Positive Control Study** – Before a laboratory conducts a study under this test method for the first time, laboratory proficiency should be demonstrated by testing the responsiveness of the animal model, by establishing the dose response of a reference oestrogen: 17α-ethinyl estradiol (CAS No. 57-63-6) (EE) with a minimum of four doses. The uterine weight response will be compared to established historical data (see reference (5)). If this baseline positive control study does not yield the anticipated results the experimental conditions should be examined and modified.

Number and condition of animals

28. Each treated and control group should include at least 6 animals (for both immature and ovx-adult method protocols).

Age of immature animals

29. For the Uterotrophic Bioassay with immature animals the day of birth must be specified. Dosing should begin early enough to ensure that, at the end of test chemicall administration, the physiological rise of endogenous oestrogens associated with puberty has not yet taken place. On the other hand, there is evidence that very young animals may be less sensitive. For defining the optimal age each laboratory should take its own background data on maturation into consideration.

As a general guide, dosing in rats may begin immediately after early weaning on postnatal day 18 (with the day of birth being postnatal day 0). Dosing in rats preferably should be completed on postnatal day 21 but in any case prior to postnatal day 25, because, after this age, the hypothalamic-pituitary-ovarian axis becomes functional and endogenous oestrogen levels may begin to rise with a concomitant increase in baseline uterine weight means and an increase in the group standard deviations (2)(3)(10)(11)(12).

Procedure for ovariectomy

30. For the ovariectomised female rat and mouse (treatment and control groups), ovariectomy should occur between 6 and 8 weeks of age. For rats, a minimum of 14 days should elapse between ovariectomy and the first day of administration in order to allow the uterus to regress to a minimum, stable baseline. For mice, at least 7 days should elapse between ovariectomy and the first day of administration. As small amounts of ovarian tissue are sufficient to produce significant circulating levels of oestrogens (3), the animals should be tested prior to use by observing epithelial cells swabbed from the vagina on at least five consecutive days (e.g., days 10-14 after ovariectomy for rats). If the animals indicate any evidence entering oestrous, the animals should not be used. Further, at necropsy, the ovarian stubs should be examined for any evidence that ovarian tissue is present. If so, the animal should not be used in the calculations (3).

31. The ovariectomy procedure begins with the animal in ventral recumbency after the animal has been properly anesthetised. The incision opening the dorso-lateral abdominal wall should be approximately 1 cm lengthways at the mid-point between the costal inferior border and the iliac crest, and a few millimetres lateral to the lateral margin of the lumbar muscle. The ovary should be removed from the abdominal cavity onto an aseptic field. The ovary should be disconnected at the junction of the oviduct and the uterine body. After confirming that no massive bleeding is occurring, the abdominal wall should be closed by a suture and the skin closed by autoclips or appropriate suture. The ligation points are shown schematically in Figure 1. Appropriate post-operative analgesia should be used as recommended by a veterinarian experienced in rodent care.

Body weight

32. In the ovx-adult method, body weight and uterine weight are not correlated because

uterine weight is affected by hormones like oestrogens but not by the growth factors that regulate body size. On the contrary, body weight is related to uterine weight in the immature model, while it is maturing (34). Thus, at the commencement of the study the weight variation of animals used, in the immature model, should be minimal and not exceed ± 20 % of the mean weight. This means that the litter size should be standardised by the breeder, to ensure that offspring of different mother animals will be fed approximately the same. Animals should be assigned to groups (both control and treatment) by randomised weight distribution, so that mean body weight of each group is not statistically different from any other group. Consideration should be given to avoid assignment of littermates to the same treatment group as far as practicable without increasing the number of litters to be used for the investigation.

Dosage

33. In order to establish whether a test chemical can have oestrogenic action *in vivo*, two dose groups and a control are normally sufficient and this design is therefore preferred for animal welfare reasons. If the purpose is either to obtain a dose-response curve or to extrapolate to lower doses, at least 3 dose groups are needed. If information beyond identification of oestrogenic activity (such as an estimate of potency) is required, a different dosing regimen should be considered. Except for treatment with the test chemical, animals in the control group should be handled in an identical manner to the test group subjects. If a vehicle is used in administering the test chemical, the control group should receive the same amount of vehicle used with the treated groups (or highest volume used with the test groups if different among groups).

34. The objective in the case of the Uterotrophic Bioassay is to select doses that ensure animal survival and that are without significant toxicity or distress to the animals after three consecutive days of chemical administration up to a maximum dose of 1000 mg/kg/d. All dose levels should be proposed and selected taking into account any existing toxicity and (toxico-) kinetic data available for the test chemical or related materials. The highest dose level should first take into consideration the LD50 and/or acute toxicity information in order to avoid death, severe suffering or distress in the animals (24)(25)(26). The highest dose should represent the maximum tolerated dose (MTD); a study conducted at a dose level that induced a positive uterotrophic response would be accepted too. As a screen, large intervals (e.g. one half log units corresponding to a dose progression of 3.2 or even up to one log units) between dosages are generally acceptable. If there are no suitable data available, a range finding study may be performed to aid the determination of the doses to be used.

35. Alternatively, if the oestrogenic potency of an agonist can be estimated by *in vitro* (or *in silico*) data, these may be taken into consideration for dose selection. For example, the amount of the test chemical that would produce uterotrophic responses equivalent to the reference agonist (ethinyl estradiol) is estimated by its relative *in vitro* potencies to ethinyl estradiol. The highest test dose would be given by multiplying this equivalent dose by an appropriate factor e.g. 10 or 100.

Considerations for range finding

36. If necessary, a preliminary range finding study can be carried out with few animals. In

this respect, OECD Guidance Document n°19 (25) may be used defining clinical signs indicative of toxicity or distress to the animals. If feasible within this range finding study after three days of administration, the uteri may be excised and weighed approximately 24-hours after the last dose. These data could then be used to assist the main study design (select an acceptable maximum and lower doses and recommend the number of dose groups).

Administration of doses

37. The test chemical is administered by oral gavage or subcutaneous injection. Animal welfare considerations as well as toxicological aspects like the relevance to the human route of exposure to the chemical (e.g. oral gavage to model ingestion, subcutaneous injection to model inhalation or dermal adsorption), the physical/chemical properties of the test material and especially existing toxicological information and data on metabolism and kinetics (e.g. need to avoid first pass metabolism, better efficiency via a particular route) have to be taken into account when choosing the route of administration.

38. It is recommended that, wherever possible, the use of an aqueous solution/suspension be considered first. But as most oestrogen ligands or their metabolic precursors tend to be hydrophobic, the most common approach is to use a solution/suspension in oil (e.g. corn, peanut, sesame or olive oil). However, these oils have different caloric and fat content, thus the vehicle might affect total metabolisable energy (ME) intake, thereby potentially altering measured endpoints such as the uterine weight especially in the immature method (33). Thus, prior to the study, any vehicle to be used should be tested against controls without vehicles. Test chemicals can be dissolved in a minimal amount of 95% ethanol or other appropriate solvents and diluted to final working concentrations in the test vehicle. The toxic characteristics of the solvent must be known, and should be tested in a separate solvent-only control group. If the test chemical is considered stable, gentle heating and vigorous mechanical action can be used to assist in dissolving the test chemical. The stability of the test chemical in the vehicle should be determined. If the test chemical is stable for the duration of the study, then one starting aliquot of the test chemical may be prepared, and the specified dosage dilutions prepared daily.

39. Dosage timing will depend of the model used (refer to paragraph 29 for the immature model and to paragraph 30 for ovx-adult model). Immature female rats are dosed with the test chemical daily for three consecutive days. A three-day treatment is also recommended for ovariectomised female rats but longer exposures are acceptable and may improve the detection of weakly active chemicals. With ovariectomised female mice, an application duration of 3 days should be sufficient without a significant advantage by an extension of up to seven days for strong oestrogen agonists, however, this relation was not demonstrated for weak oestrogens in the validation study (16) thus dosage should be extended up to 7 consecutive days in ovx-adult mice. The dose should be given at similar times each day. They should be adjusted as necessary to maintain a constant dose level in terms of animal body weight (e.g., mg of test chemical per kg of body weight per day). Regarding the test volume, its variability, on a body weight basis, should be minimised by adjusting the concentration of the dosing solution to ensure a constant volume on a body weight basis at all dose levels and for any route of administration.

40. When the test chemical is administered by gavage, this should be done in a single daily dose to the animals using a stomach tube or a suitable intubation cannula. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. Local animal care guidelines should be followed, but the volume should not exceed 5 ml/kg body weight, except in the case of aqueous solutions where 10 ml/kg body weight may be used.

41. When the test chemical is administered by subcutaneous injection, this should be done in a single daily dose. Doses should be administered to the dorsoscapular or lumbar regions via sterile needle (e.g. 23- or 25-gauge) and a tuberculin syringe. Shaving the injection site is optional. Any losses, leakage at the injection site or incomplete dosing should be recorded. The total volume injected per rat per day should not exceed 5 ml/kg body weight, divided into 2 injection sites, except in the case of aqueous solutions where 10 ml/kg body weight may be used.

Observations

General and clinical observations

42. General clinical observations should be made at least once a day and more frequently when signs of toxicity are observed. Observations should be carried out preferably at the same time(s) each day and considering the period of anticipated peak effects after dosing. All animals are to be observed for mortality, morbidity and general clinical signs such as changes in behaviour, skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity (e.g. lacrimation, piloerection, pupil size, unusual respiratory pattern).

Body weight and food consumption

43. All animals should be weighed daily to the nearest 0.1 g, starting just prior to initiation of treatment i.e., when the animals are allocated into groups. As an optional measurement, the amount of food consumed during the treatment period may be measured per cage by weighing the feeders. The food consumption results should be expressed in grammes per rat per day.

Dissection and measurement of uterus weight

44. Twenty-four hours after the last treatment, the rats will be humanely killed. Ideally, the necropsy order will be randomised across groups to avoid progression directly up or down dose groups that could subtly affect the data. The bioassay objective is to measure both the wet and blotted uterus weights. The wet weight includes the uterus and the luminal fluid contents. The blotted weight is measured after the luminal contents of the uterus have been expressed and removed.

45. Before dissection the vagina will be examined for opening status in immature animals. The dissection procedure begins by opening the abdominal wall starting at the pubic symphysis. Then, uterine horn and ovaries, if present, are detached from the dorsal abdominal wall. The urinary bladder and ureters are removed from the ventral and lateral side of uterus and vagina. Fibrous adhesion between the rectum and the vagina is

detached until the junction of vaginal orifice and perineal skin can be identified. The uterus and vagina are detached from the body by incising the vaginal wall just above the junction between perineal skin as shown in Figure 2. The uterus should be detached from the body wall by gently cutting the uterine mesentery at the point of its attachment along the full length of the dorsolateral aspect of each uterine horn. Once removed from the body, uterine handling should be sufficiently rapid to avoid desiccation of the tissues. Loss of weight due to desiccation becomes more important with small tissues such as the uterus (23). If ovaries are present, the ovaries are removed at the oviduct avoiding loss of luminal fluid from the uterine horn. If the animal has been ovariectomised, the stubs should be examined for the presence of any ovarian tissue. Excess fat and connective tissue should be trimmed away. The vagina is removed from the uterus just below the cervix so that the cervix remains with the uterine body as shown in Figure 2.

46. Each uterus should be transferred to a uniquely marked and weighed container (e.g. a petri-dish or plastic weight boat) with continuing care to avoid desiccation before weighing (e.g. filter paper slightly dampened with saline may be placed in the container). The uterus with luminal fluid will be weighed to the nearest 0.1 mg (wet uterine weight).

47. Each uterus will then be individually processed to remove the luminal fluid. Both uterine horns will be pierced or cut longitudinally. The uterus will be placed on lightly moistened filter paper (e.g. Whatman No. 3) and gently pressed with a second piece of lightly moistened filter paper to completely remove the luminal fluid. The uterus without the luminal contents will be weighed to the nearest 0.1 mg (blotted uterine weight).

48. The uterus weight at termination can be used to ensure that the appropriate age in the immature intact rat was not exceeded, however, the historical data of the rat strain used by the laboratory are decisive in this respect (see paragraph 56 for interpretation of the results).

Optional investigations

49. After weighing, the uterus may be fixed in 10% neutral buffered formalin to be examined histopathologically after Haematoxylin & Eosin (HE)-staining. The vagina may be investigated accordingly (see paragraph 9). In addition, morphometric measurement of endometrial epithelium may be done for quantitative comparison.

DATA AND REPORTING

Data

50. Study data should include:
    - the number of animals at the start of the assay,
    - the number and identity of animals found dead during the assay or killed for

humane reasons and the date and time of any death or humane kill,

- the number and identity of animals showing signs of toxicity, and a description of the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, and

- the number and identity of animals showing any lesions and a description of the type of lesions.

51. Individual animal data should be recorded for the body weights, the wet uterine weight, and the blotted uterine weight. One-tailed statistical analyses for agonists should be used to determine whether the administration of a test chemical resulted in a statistically significant ($p < 0.05$) increase in the uterine weight. Appropriate statistical analyses should be carried out to test for treatment related changes in blotted and wet uterine weight. For example, the data may be evaluated by an analysis of covariance (ANCOVA) approach with body weight at necropsy as the co-variable. A variance-stabilising logarithmic transformation may be carried out on the uterine data prior to the data analysis. Dunnett and Hsu's test are appropriate for making pair wise comparisons of each dosed group to vehicle controls and to calculate the confidence intervals. Studentised residual plots can be used to detect possible outliers and to assess homogeneity of variances. These procedures were applied in the OECD validation programme using the PROC GLM in the Statistical Analysis System (SAS Institute, Cary, NC), version 8 (6)(7).

52. A final report shall include:

Testing facility:

- Responsible personnel and their study responsibilities
- Data from the Baseline Positive Control Test and periodic positive control data (see paragraphs 26 and 27)

Test chemical:

- Characterisation of test chemicals
- Physical nature and where relevant physicochemical properties
- Method and frequency of preparation of dilutions
- Any data generated on stability
- Any analyses of dosing solutions

Vehicle:

- Characterisation of test vehicle (nature, supplier and lot)
- Justification of choice of vehicle (if other than water)

Test animals:

- Species and strain and justification for their choice
- Supplier and specific supplier facility

- Age on supply with birth date
- If immature animals, whether or not supplied with dam or foster dam and date of weaning
- Details of animal acclimatisation procedure
- Number of animals per cage
- Detail and method of individual animal and group identification

Assay Conditions:
- Details of randomisation process (i.e., method used)
- Rationale for dose selection
- Details of test chemical formulation, its achieved concentrations, stability and homogeneity
- Details of test chemical administration and rationale for the choice of exposure route
- Diet (name, type, supplier, content, and, if known, phyto-oestrogen levels)
- Water source (e.g., tap water or filtered water) and supply (by tubing from a large container, in bottles, etc.)
- Bedding (name, type, supplier, content)
- Record of caging conditions, lighting interval, room temperature and humidity, room cleaning
- Detailed description of necropsy and uterine weighing procedures
- Description of statistical procedures

Results

   For individual animals:
- All daily individual body weights (from allocation into groups through necropsy) (to the nearest 0.1 g)
- Age of each animal (in days counting day of birth as day 0) when administration of test chemical begins
- Date and time of each dose administration
- Calculated volume and dosage administered and observations of any dosage losses during or after administration
- Daily record of status of animal, including relevant symptoms and observations
- Suspected cause of death (if found during study in moribund state or dead)
- Date and time of humane killing with time interval to last dosing
- Wet uterine weight (to the nearest 0.1 mg) and any observations of luminal fluid losses during dissection and preparation for weighing
- Blotted uterine weight (to the nearest 0.1 mg)

For each group of animals:

- Mean daily body weights (to the nearest 0.1 g) and standard deviations (from allocation into groups through necropsy)

- Mean wet uterine weights and mean blotted uterine weights (to the nearest 0.1 mg) and standard deviations

- If measured, daily food consumption (calculated as grams of food consumed per animal)

- The results of statistical analyses comparing both the wet and blotted uterine weights of treated groups relative to the same measures in the vehicle control groups.

- The results of statistical analysis comparing the total body weight and the body weight gain of treated groups relative to the same measures in the vehicle control groups.

53. Summary of the important guidance facts of the test method

	Rat	*Mice*
Animals		
Strain	*Commonly used laboratory rodent strain*	
Number of animals	***A minimum of 6 animals per dose group***	
Number of groups	*A minimum of 2 test groups (see paragraph 33 for guidance) and a negative control group* *For guidance on positive control groups see paragraphs 26 and 27*	
Housing and feeding conditions		
T° in animal room	*22°C ± 3°C*	
Relative humidity	*50-60% and not below 30% or above 70%*	
Daily lighting sequence	*12 hours light, 12 hours dark*	
Diet and drinking water	*Ad libitum*	
Housing	*Individually or in groups of up to three animals (social group housing is recommended for immature animals)*	
Diet and bedding	*Low level of phyto-oestrogens recommended in diet and bedding*	
Protocol		
Method	*Immature non-ovariectomised method (the preferred one).* *Ovariectomised adult female method*	*Ovariectomised adult female method*
Age of dosing for immature animals	*PND 18 at the earliest. Dosing should be completed prior to PND 25*	*Not relevant under the scope of the current test method.*
Age of ovariectomy	*Between 6 and 8 weeks of age.*	
Age of dosing for ovariectomised animals	*A minimum of 14 days should elapse between ovariectomy and the 1st day of administration.*	*A minimum of 7 days should elapse between ovariectomy and the 1st day of administration.*
Body weight	*Body weight variation should be minimal and not exceed ± 20% of the mean weight.*	
Dosing		
Route of administration	*Oral gavage or subcutaneous injection*	
Frequency of administration	*Single daily dose*	
Volume amount for gavage and injection	*≤ 5ml/kg body weight (or up to 10 ml/kg body weight in case of aqueous solutions) (in 2 injection sites for subcutaneous route)*	
Duration of	*3 consecutive days for immature model*	*7 consecutive days for the OVX model*

| administration	Minimum of 3 consecutive days for the OVX model	
Time of necropsy	Approximately 24 hours after the last dose	
Results		
Positive response	Statistically significant increase of the mean uterus weight (wet and/or blotted)	
Reference oestrogen	17α-ethinyl estradiol	

GUIDANCE FOR THE INTERPRETATION AND ACCEPTANCE OF THE RESULTS

54. In general, a test for oestrogenicity should be considered positive if there is a statistically significant increase in uterine weight ($p < 0.05$) at least at the high dose level as compared to the solvent control group. A positive result is further supported by the demonstration of a biologically plausible relationship between the dose and the magnitude of the response, bearing in mind that overlapping oestrogenic and antioestrogenic activities of the test chemical may affect the shape of the dose-response curve.

55. Care must be taken in order not to exceed the maximum tolerated dose to allow a meaningful interpretation of the data. Reduction of body weight, clinical signs and other findings should be thoroughly assessed in this respect.

56. An important consideration for the acceptance of the data from the Uterotrophic Bioassay is the uterine weights of the vehicle control group. High control values may compromise the responsiveness of the bioassay and the ability to detect very weak oestrogen agonists. Literature reviews and the data generated during the validation of the Uterotrophic Bioassay suggest that instances of high control means do occur spontaneously, particularly in immature animals (2)(3)(6)(9). As the uterine weight of immature rats depends on many variables like strain or body weight, no definitive upper limit for the uterine weight can be given. As a guide, if blotted uterine weights in immature control rats are comprised between 40 and 45 mg, results should be considered as suspicious and uterine weights above 45 mg may lead to rerun the test. However, this needs to be considered on a case by case basis (3)(6)(8). When testing in adult rats incomplete ovariectomy will leave ovarian tissue that can produce endogenous oestrogen and retard the regression of the uterine weight.

57. Blotted vehicle control uterine weights less than 0.09% of body weight for immature female rats and less than 0.04% for ovariectomised young adult females appear to yield acceptable results [see Table 31 (2)]. If the control uterine weights are greater than these numbers, various factors should be scrutinised including the age of the animals, proper ovariectomy, dietary phyto-oestrogens, and so on, and a negative assay result (no indication for oestrogenic activity) should be used with caution.

58. Historical data for vehicle control groups should be maintained in the laboratory. Historical data for responses to positive reference oestrogens, such as 17α-ethinyl estradiol, should also be maintained in the laboratory. Laboratories may also test the

response to known weak oestrogen agonists. All these data can be compared to available data (2)(3)(4)(5)(6)(7)(8) to ensure that the laboratory's methods yield sufficient sensitivity.

59. The blotted uterine weights showed less variability in the course of the OECD validation study than the wet uterine weights (6)(7). However, a significant response in either measure would indicate that the test chemical is positive for oestrogenic activity.

60. The uterotrophic response is not entirely of oestrogenic origin, however, a positive result of the Uterotrophic Bioassay should generally be interpreted as evidence for oestrogenic potential in vivo, and should normally initiate actions for further clarification (see paragraph 9 and the "OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals", Annex 2).

Figure 1: Schematic diagram showing the surgical removal of the ovaries

Mesometrium, vasculature
and fat pad not shown
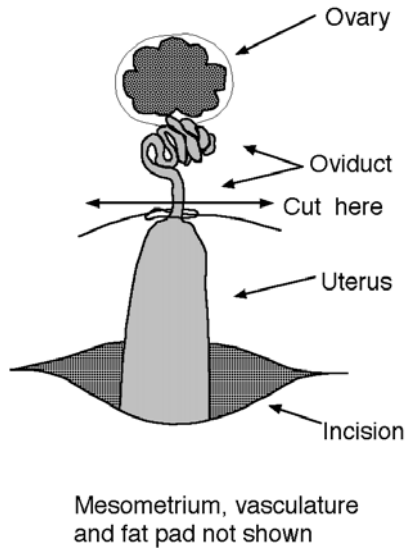
The procedure begins by opening dorso-lateral abdominal wall at the mid-point between the costal inferior border and the iliac crest, and a few millimetres lateral to the lateral margin of the lumbar muscle. Within the abdominal cavity, the ovaries should be located. On an aseptic field, the ovaries are then physically removed from the abdominal cavity, a ligature placed between the ovary and uterus to control bleeding, and the ovary detached by incision above the ligature at the junction of the oviduct and each uterine horn. After confirming that no significant bleeding persists, the abdominal wall should be closed by suture, and the skin closed, e.g., by autoclips or suture. The animals should be allowed to recover and the uterus weight to regress for a minimum of 14 days before use.

Figure 2: The removal and preparation of the uterine tissues for weight measurement.

UTERINE
WEIGHT

Disconnection line at necropsy
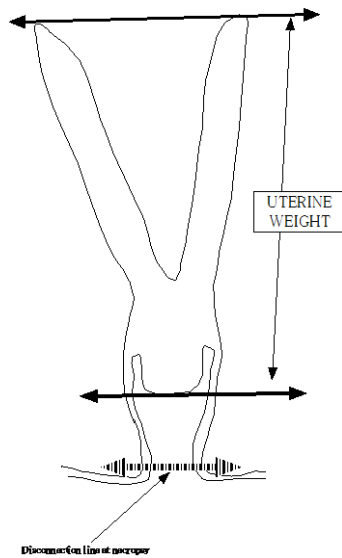
The procedure begins by opening the abdominal wall at the pubic symphysis. Then, each ovary, if present and uterine horn is detached from the dorsal abdominal wall. Urinary bladder and ureters are removed from the ventral and lateral side of uterus and vagina. Fibrous adhesion between the rectum and the vagina are detached until the junction of vaginal orifice and perineal skin can be identified. The uterus and vagina are detached from the body by incising the vaginal wall just above the junction between perineal skin as shown in the figure. The uterus should be detached from the body wall by gently cutting the uterine mesentery at the point of its attachment along the full length of the dorsolateral aspect of each uterine horn. After removal from the body, the excess fat and connective tissue is trimmed away. If ovaries are present, the ovaries are removed at the oviduct avoiding loss of luminal fluid from the uterine horn. If the animal has been ovarectomised, the stubs should be examined for the presence of any ovarian tissue. The vagina is removed from the uterus just below the cervix so that the cervix remains with the uterine body as shown in the figure. The uterus can then be weighed.

Appendix 1

DEFINITIONS

Antioestrogenicity is the capability of a chemical to suppress the action of estradiol 17ß in a mammalian organism.

Chemical means a substance or a mixture.

Date of birth is postnatal day 0.

Dosage is a general term comprising of dose, its frequency and the duration of dosing.

Dose is the amount of test chemical administered. For the Uterotrophic Bioassay, the dose is expressed as weight of test chemical per unit body weight of test animal per day (e.g. mg/kg body weight/day).

Maximum Tolerable Dose (MTD) is the highest amount of a chemical that, when introduced into the body does not kill test animals (denoted by LD_0) (IUPAC, 1993)

Oestrogenicity is the capability of a chemical to act like estradiol 17ß in a mammalian organism.

Postnatal day X is the Xth day of life after the day of birth.

Sensitivity is the proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method.

Specificity is the proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method.

Test chemical means any substance or mixture tested using this test method.

Uterotrophic is a term used to describe a positive influence on the growth of uterine tissues.

Validation is a scientific process designed to characterise the operational requirements and limitations of a test method and to demonstrate its reliability and relevance for a particular purpose.

Note: Document prepared by the Secretariat of the Test Guidelines Programme based on the agreement reached at the 6th Meeting of the EDTA Task Force

OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals

Level	Description	Examples
Level 1 Sorting & prioritization based upon existing information		- physical & chemical properties, e.g., MW, reactivity, volatility, biodegradability, - human & environmental exposure, e.g., production volume, release, use patterns - hazard, e.g., available toxicological data
Level 2 *In vitro* assays providing mechanistic data	- ER, AR, TR receptor binding affinity - Transcriptional activation - Aromatase and steroidogenesis *in vitro* - Aryl hydrocarbon receptor recognition/binding - QSARs	-High Through Put Prescreens - Thyroid function - Fish hepatocyte VTG assay - Others (as appropriate)
Level 3 *In vivo* assays providing data about single endocrine Mechanisms and effects	- Uterotrophic assay (estrogenic related) - Hershberger assay (androgenic related) - Non -receptor mediated hormone function - Others (e.g. thyroid)	- Fish VTG (vitellogenin) assay (estrogenic related)
Level 4 *In vivo* assays providing data about multiple endocrine Mechanisms and effects	- enhanced OECD 407 (endpoints based on endocrine mechanisms) - male and female pubertal assays - adult intact male assay	- Fish gonadal histopathology assay - Frog metamorphosis assay
Level 5 *In vivo* assays providing data on effects from endocrine & other mechanisms	- 1-generation assay (TG415 enhanced)[1] - 2-generation assay (TG416 enhanced)[1] - reproductive screening test (TG421 enhanced)[1] - combined 28 day/reproduction screening test (TG 422 enhanced)[1] [1] Potential enhancements will be considered by VMG mamm	- Partial and full life cycle assays in fish, birds, amphibians & invertebrates (developmental and reproduction)

NOTES TO THE FRAMEWORK

Note 1: Entering at all levels and exiting at all levels is possible and depends upon the nature of existing information needs for hazard and risk assessment purposes

Note 2: In level 5,ecotoxicology should include endpoints that indicate mechanisms of adverse effects, and potential population damage

Note 3: When a multimodal model covers several of the single endpoint assays, that model would replace the use of those single endpoint assays

Note 4: The assessment of each chemical should be based on a case by case basis, taking into account all available information, bearing in mind the function of the framework levels.

Note 5: The framework should not be considered as all inclusive at the present time. At levels 3,4 and 5 it includes assays that are either available or for which validation is under way. With respect to the latter, these are provisionally included. Once developed and validated, they will be formally added to the framework.

Note 6: Level 5 should not be considered as including definitive tests only. Tests included at that level are considered to contribute to general hazard and risk assessment.

LITERATURE

(1) OECD (1998). Report of the First Meeting of the OECD Endocrine Disrupter Testing and Assessment (EDTA) Task Force, 10th-11th March 1998, ENV/MC/CHEM/RA(98)5.

(2) OECD (2003). Detailed Background Review of the Uterotrophic Bioassay: Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Bioassay. OECD Environmental Health and Safety Publication Series on Testing and Assessment No. 38. ENV/JM/MONO(2003)1.

(3) Owens JW, Ashby J. (2002). Critical Review and Evaluation of the Uterotrophic Bioassay for the Identification of Possible Estrogen Agonists and Antagonists: In Support of the Validation of the OECD Uterotrophic Protocols for the Laboratory Rodent. Crit. Rev. Toxicol. 32:445-520.

(4) OECD (2006). OECD Report of the Initial Work Towards the Validation of the Rodent Uterotrophic Assay - Phase 1. OECD Environmental Health and Safety Publication Series on Testing and Assessment No. 65. ENV/JM/MONO(2006)33.

(5) Kanno, J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W. (2001). The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase 1. Environ Health Perspect. 109:785-94.

(6) OECD (2006). OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2 - Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories. OECD Environmental Health and Safety Publication Series on Testing and Assessment No. 66. ENV/JM/MONO(2006)34.

(7) Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. (2003). The OECD program to validate the rat uterotrophic bioassay: Phase Two - Dose Response Studies. Environ. Health Persp. 111:1530-1549

(8) Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. (2003). The OECD program to validate the rat uterotrophic bioassay: Phase Two – Coded Single Dose Studies. Environ. Health Persp. 111:1550-1558.

(9) Owens W, Ashby J, Odum J, Onyon L. (2003). The OECD program to validate the rat uterotrophic bioassay: Phase Two – Dietary phytoestrogen analyses. Environ. Health Persp. 111:1559-1567.

(10) Ogasawara Y, Okamoto S, Kitamura Y, Matsumoto K. (1983). Proliferative pattern of uterine cells from birth to adulthood in intact,

neonatally castrated, and/or adrenalectomized mice assayed by incorporation of [I^{125}]iododeoxyuridine. Endocrinology 113:582-587.

(11) Branham WS, Sheehan DM, Zehr DR, Ridlon E, Nelson CJ. (1985). The postnatal ontogeny of rat uterine glands and age-related effects of 17β-estradiol. Endocrinology 117:2229-2237.

(12) Schlumpf M, Berger L, Cotton B, Conscience-Egli M, Durrer S, Fleischmann I, Haller V, Maerkel K, Lichtensteiger W. (2001). Estrogen active UV screens. SÖFW-J. 127:10-15.

(13) Zarrow MX, Lazo-Wasem EA, Shoger RL. (1953). Estrogenic activity in a commercial animal ration. Science 118:650-651.

(14) Drane HM, Patterson DSP, Roberts BA , Saba N. (1975). The chance discovery of oestrogenic activity in laboratory rat cake. Fd. Cosmet. Toxicol. 13:425-427.

(15) Boettger-Tong H, Murphy L, Chiappetta C, Kirkland JL, Goodwin B, Adlercreutz H, Stancel GM, Makela S. (1998). A case of a laboratory animal feed with high estrogenic activity and its impact on *in vivo* responses to exogenously administered estrogens. Environ. Health Perspec. 106:369-373.

(16) OECD (2007). Additional data supporting the Test Guideline on the Uterotrophic Bioassay in rodents. OECD Environmental Health and Safety Publication Series on Testing and Assessment No. 67.

(17) Degen GH, Janning P, Diel P, Bolt HM. (2002). Estrogenic isoflavones in rodent diets. Toxicol. Lett. 128:145-157.

(18) Wade MG, Lee A, McMahon A, Cooke G, Curran I. (2003). The influence of dietary isoflavone on the uterotrophic response in juvenile rats. Food Chem. Toxicol. 41:1517-1525.

(19) Yamasaki K, Sawaki M, Noda S, Wada T, Hara T, Takatsuki M. (2002). Immature uterotrophic assay of estrogenic compounds in rats given different phytoestrogen content diets and the ovarian changes in the immature rat uterotrophic of estrogenic compounds with ICI 182,780 or antide. Arch. Toxicol. 76:613-620.

(20) Thigpen JE, Haseman JK, Saunders HE, Setchell KDR, Grant MF, Forsythe D. (2003). Dietary phytoestrogens accelerate the time of vaginal opening in immature CD-1 mice. Comp. Med. 53:477-485.

(21) Ashby J, Tinwell H, Odum J, Kimber I, Brooks AN, Pate I, Boyle CC. (2000). Diet and the aetiology of temporal advances in human and rodent sexual development. J. Appl. Toxicol. 20:343-347.

(22) Thigpen JE, Lockear J, Haseman J, Saunders HE, Caviness G, Grant MF, Forsythe DB. (2002). Dietary factors affecting uterine weights of

immature CD-1 mice used in uterotrophic bioassays. Cancer Detect. Prev. 26:381-393.

(23) Thigpen JE, Li L-A, Richter CB, Lebetkin EH, Jameson CW. (1987). The mouse bioassay for the detection of estrogenic activity in rodent diets: I. A standardized method for conducting the mouse bioassay. Lab. Anim. Sci. 37:596-601.

(24) OECD (2008). Acute oral toxicity – up-and-down procedure. OECD Guideline for the testing of chemicals No 425.

(25) OECD (2000). Guidance document on the recognition, assessment and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No 19. ENV/JM/MONO(2000)7.

(26) OECD (2001). Guidance document on acute oral toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment No 24. ENV/JM/MONO(2001)4.

(27) Bulbring, E., and Burn, J.H. (1935). The estimation of oestrin and of male hormone in oily solution. J. Physiol. 85: 320 - 333.

(28) Dorfman, R.I., Gallagher, T.F. and Koch, F.C (1936). The nature of the estrogenic substance in human male urine and bull testis. Endocrinology 19: 33 - 41.

(29) Reel, J.R., Lamb IV, J.C. and Neal, B.H. (1996). Survey and assessment of mammalian estrogen biological assays for hazard characterization. Fundam. Appl. Toxicol. 34: 288 - 305.

(30) Jones, R.C. and Edgren, R.A. (1973). The effects of various steroid on the vaginal histology in the rat. Fertil. Steril. 24: 284 – 291.

(31) OECD (1982). Organization for Economic Co-operation and Development - Principles of Good Laboratory Practice, ISBN 92-64-12367-9, Paris.

(32) Dorfman R.I. (1962). Methods in Hormone Research, Vol. II, Part IV: Standard Methods Adopted by Official Organization. New York, Academic Press.

(33) Thigpen J. E. et al. (2004). Selecting the appropriate rodent diet for endocrine disruptor research and testing studies. ILAR J 45(4): 401-416.

(34) Gray L.E. and Ostby J. (1998). Effects of pesticides and toxic substances on behavioral and morphological reproductive development: endocrine versus non-endocrine mechanism. Toxicol Ind Health. 14 (1-2): 159-184.

(35) Booth AN, Bickoff EM and Kohler GO. (1960). Estrogen-like activity in vegetable oils and mill by-products. Science 131:1807-1808.

(36) Kato H, Iwata T, Katsu Y, Watanabe H, Ohta Y, Iguchi T (2004). Evaluation of estrogenic activity in diets for experimental animals using in vitro assay. J. Agric Food Chem. 52, 1410-1414.

(37) OECD (2007). Guidance Document on the Uterotrophic Bioassay Procedure to Test for Antioestrogenicity. Series on Testing and Assessment. No. 71.

(38) Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33.

B.55. Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti)Androgenic Properties

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 441 (2009). The OECD initiated a high-priority activity in 1998 to revise existing guidelines and to develop new guidelines for the screening and testing of potential endocrine disrupters (1). One element of the activity was to develop a test guideline for the rat Hershberger Bioassay. After several decades of use by the pharmaceutical industry, this assay was first standardised by an official expert committee in 1962 as a screening tool for androgenic chemicals (2). In 2001-2007, the rat Hershberger Bioassay has undergone an extensive validation programme including the generation of a Background Review Document (23), compilation of a detailed methods paper (3), development of a dissection guide (21) and the conduct of extensive intra- and interlaboratory studies to show the reliability and reproducibility of the bioassay. These validation studies were conducted with a potent reference androgen (testosterone propionate (TP)), two potent synthetic androgens (trenbolone acetate and methyl testosterone), a potent antiandrogenic pharmaceutical (flutamide), a potent inhibitor of the synthesis (finasteride) of the natural androgen (dihydrotestosterone-DHT), several weakly antiandrogenic pesticides (linuron, vinclozolin, procymidone, p,p' DDE), a potent 5α reductase inhibitor (finasteride) and two known negative chemicals (dinitrophenol and nonylphenol) (4) (5) (6) (7) (8). This test method is the outcome of the long historical experience with the bioassay and the experience gained during the validation test programme and the results obtained therein.

2. The Hershberger Bioassay is a short-term *in vivo* screening test using accessory tissues of the male reproductive tract. The assay originated in the 1930s and was modified in the 1940s to include androgen-responsive muscles in the male reproductive tract (2) (9-15). In the 1960s, over 700 possible androgens were evaluated using a standardised version of the protocol (2) (14), and use of the assay for both androgens and antiandrogens was considered a standard method in the 1960s (2) (15). The current bioassay is based on the changes in weight of five androgen-dependent tissues in the castrate-peripubertal male rat. It evaluates the ability of a chemical to elicit biological activities consistent with androgen agonists, antagonists or 5α-reductase inhibitors. The five target androgen-dependent tissues included in this test method are the ventral prostate (VP), seminal vesicle (SV) (plus fluids and coagulating glands), levator ani-bulbocavernosus (LABC) muscle, paired Cowper's glands (COW) and the glans penis (GP). In the castrate-peripubertal male rat, these five tissues all respond to androgens with an increase in absolute weight. When these same tissues are stimulated to increase in weight by administration of a potent reference androgen, these five tissues all respond to antiandrogens with a decrease in absolute weight. The primary model for the Hershberger bioassay has been the surgically castrated peripubertal male, which was validated in Phases 1, 2 and 3 of the Hershberger validation programme.

3. The Hershberger bioassay serves as a mechanistic *in vivo* screening assay for androgen

agonists, androgen antagonists and 5α-reductase inhibitors and its application should be seen in the context of the "OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals" (Appendix 2). In this Conceptual Framework the Hershberger Bioassay is contained in Level 3 as an *in vivo* assay providing data about a single endocrine mechanism, i.e. (anti)androgenicity. It is intended to be included in a battery of *in vitro* and *in vivo* tests to identify chemicals with potential to interact with the endocrine system, ultimately leading to hazard and risk assessments for human health or the environment.

4. Due to animal welfare concerns with the castration procedure, the intact (uncastrated) stimulated weanling male was sought as an alternative model for the Hershberger Bioassay to avoid the castration step. The stimulated weanling test method was validated (24); however, in the validation studies, the weanling version of the Hershberger Bioassay did not appear to be able to consistently detect effects on androgen-dependent organ weights from weak anti-androgens at the doses tested. Therefore, it was not included in this test method. However, recognising that its use may provide not only animal welfare benefits but also may provide information on other modes of action, it is available in OECD Guidance Document 115 (25).

INITIAL CONSIDERATIONS AND LIMITATIONS

5. Androgen agonists and antagonists act as ligands for the androgen receptor and may activate or inhibit, respectively, gene transcription controlled by the receptor. In addition, some chemicals inhibit the conversion of testosterone to the more potent natural androgen dihydrotestosterone in some androgen target tissues (5α-reductase inhibitors). Such chemicals have the potential to lead to adverse health hazards, including reproductive and developmental effects. Therefore, the regulatory need exists to rapidly assess and evaluate a chemical as a possible androgen agonist or antagonist or 5α-reductase inhibitor. While informative, the affinity of a ligand for an androgen receptor as measured by receptor binding or transcriptional activation of reporter genes *in vitro* is not the only determinant of possible hazard. Other determinants include metabolic activation and deactivation upon entering the body, chemical distribution to target tissues, and clearance from the body. This leads to the need to screen the possible activity of a chemical *in vivo* under relevant conditions and exposure. *In vivo* evaluation is less critical if the chemical's characteristics regarding Absorption – Distribution – Metabolism – Elimination (ADME) are known. Androgen-dependent tissues respond with rapid and vigorous growth to stimulation by androgens, particularly in castrate-peripubertal male rats. Rodent species, particularly the rat, are also widely used in toxicity studies for hazard characterisation. Therefore, the assay version, using the castrated peripubertal rat and the five target tissues in this assay, is appropriate for the *in vivo* screening of androgen agonists and antagonists and 5α-reductase inhibitors.

6. This test method is based on those protocols employed in the OECD validation study which have been shown to be reliable and reproducible in intra- and inter-laboratory studies (4) (5)

(6) (7) (8). Both androgen and antiandrogen procedures are presented in this test method.

7. Although there was some variation in the dose of TP used to detect antiandrogens in the OECD Hershberger Bioassay validation programme by the different laboratories (0.2 versus 0.4 mg/kg/d, subcutaneous injection) there was little difference between these two protocol variations in the ability to detect weak or strong antiandrogenic activity. However, it is clear that the dose of TP should not be too high to block the effects of weak androgen receptor (AR) antagonists or so low that the androgenic tissues display little growth response even without antiandrogen coadministration.

8. The growth response of the individual androgen-dependent tissues is not entirely of androgenic origin, i.e. chemicals other than androgen agonists can alter the weight of certain tissues. However, the growth response of several tissues concomitantly substantiates a more androgen-specific mechanism. For example, high doses of potent oestrogens can increase the weight of the seminal vesicles; however, the other androgen-dependent tissues in the assay do not respond in a similar manner. Antiandrogenic chemicals can act either as androgen receptor antagonists or 5α-reductase inhibitors. 5α-reductase inhibitors have a variable effect, because the conversion to more potent dihydrotestosterone varies by tissue. Antiandrogens that inhibit 5α−reductase, like finasteride, have more pronounced effects in the ventral prostate than other tissues as compared to a potent AR antagonist, like flutamide. This difference in tissue response can be used to differentiate between AR mediated and 5α−reductase mediated modes of action. In addition, the androgen receptor is evolutionarily related to that of other steroid hormones, and some other hormones, when administered at high, supraphysiological dosage levels, can bind and antagonise the growth-promoting effects of TP (13). Further, it also is plausible that enhanced steroid metabolism and a consequent lowering of serum testosterone could reduce androgen-dependent tissue growth. Therefore, any positive outcome in the Hershberger Bioassay should normally be evaluated using a weight of evidence approach, including *in vitro* assays, such as the AR and oestrogen receptor (ER) binding assays and corresponding transcriptional activation assays, or from other *in vivo* assays that examine similar androgen target tissues such as the male pubertal assay, 15-day intact adult male assay, or 28-day or 90-day repeat dose studies.

9. Experience indicates that xenobiotic androgens are rarer than xenobiotic antiandrogens. The expectation then is that the Hershberger bioassay will be used most often for the screening of antiandrogens. However, the procedure to test for androgens could, nevertheless, be recommended for steroidal or steroid-like chemicals or for chemicals for which an indication of possible androgenic effects was derived from methods contained in Level 1 or 2 of the conceptual framework (Appendix 2). Similarly, adverse effects associated with (anti)androgenic profiles may be observed in Level 5 assays, leading to the need to assess whether a chemical operates by an endocrine mode of action.

10. It is acknowledged that all animal-based procedures should conform to local standards of animal care; the descriptions of care and treatment set forth below are minimal performance standards, and will be superseded by local regulations such as Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals

used for scientific purposes (26). Further guidance of the humane treatment of animals is given by the OECD (17).

11. As in any bioassay using experimental animals, careful considerations should be given to the necessity to carry out this study. Basically there may be two reasons for such a decision:

   - high exposure potential (Level 1 of the Conceptual Framework) or indications for (anti)androgenicity in *in vitro* assays (Level 2) supporting investigations whether such effects may occur *in vivo;*

   - effects consistent with (anti)androgenicity in Level 4 or 5 *in vivo* tests supporting investigations of the specific mode of action, e.g., to determine whether the effects were due to an (anti)androgenic mechanism.

12. Definitions used in this test method are given in Appendix 1.

PRINCIPLE OF THE TEST

13. The Hershberger Bioassay achieves its sensitivity by using males with minimal endogenous androgen production. This is achieved through the use of castrated males provided an adequate time after castration for the target tissues to regress to a minimal and uniform baseline weight is allowed. Thus, when screening of potential androgenic activity, there are low endogenous levels of circulating androgens, the hypothalamic - pituitary - gonad axis is rendered unable to compensate via feedback mechanisms, the ability of the tissues to respond is maximised, and the starting tissue weight variability is minimised. When screening of potential anti-androgenic activity, a more consistent tissue weight gain can be achieved when the tissues are stimulated by a reference androgen. As a result, the Hershberger Bioassay requires only 6 animals per dose group whereas other assays with intact pubertal or adult males suggest using 15 males per dose group.

14. Castration of peripubertal male rats should be done in an appropriate manner using approved anaesthetics and aseptic technique. Analgesics should be administered on the first few days following surgery to eliminate post-surgical discomfort. Castration enhances the precision of the assay to detect weak androgens and antiandrogens by eliminating compensatory endocrine feed-back mechanisms present in the intact animal that can attenuate the effects of administered androgens and antiandrogens and by eliminating the large inter-individual variability in serum testosterone levels. Hence, castration reduces the numbers of animals required to screen for these endocrine activities.

15. When screening for potential androgenic activity, the test chemical is administered daily by oral gavage or subcutaneous (sc) injection for a period of ten consecutive days. Test chemicals are administered to a minimum of two treatment groups of experimental animals using one dose level per group. The animals are necropsied approximately 24 hours after the last dose. A statistically significant increase in two or more target organ weights of the test chemical groups compared to the vehicle control group indicates that the test chemical is positive for potential androgenic activity (See paragraph 60). Androgens, like trenbolone that

cannot be 5α−reduced have more pronounced effects on the LABC and GP versus TP, but all tissues should display increased growth.

16. When screening for potential antiandrogenic activity, the test chemical is administered daily by oral gavage or subcutaneous injection for a period of ten consecutive days in concert with daily TP doses (0.2 or 0.4 mg/kg/d) by sc injection. It was determined in the validation programme that either 0.2 or 0.4 mg/kg/d of TP could be used as both were effective in the detection of antiandrogens and, therefore, only one dose should be selected for use in the assay. Graduated test chemical doses are administered to a minimum of three treatment groups of experimental animals using one dose level per group. The animals are necropsied approximately 24 hours after the last dose. A statistically significant decrease in two or more target organ weights of the test chemical plus TP groups compared to the TP only control group indicates that the test chemical is positive for potential antiandrogenic activity (See paragraph 61).

DESCRIPTION OF THE METHOD

Selection of species and strain

17. The rat has been routinely used in the Hershberger Bioassay since the 1930s. Although it is biologically plausible that both the rat and mouse would display similar responses, based upon 70 years of experience with the rat model, the rat is the species of choice for the Hershberger Bioassay. In addition, since Hershberger Bioassay data may be preliminary to a long-term multigenerational study, this allows animals from the same species, strain and source to be used in both studies.

18. This protocol allows laboratories to select the strain of rat to be used in the assay which should generally be that used historically by the participating laboratory. Commonly used laboratory rat strains may be used; however, strains that mature significantly later than 42 days of age should not be used since castration of these males at 42 days of age could preclude measurement of glans penis weights, which can only be done after the prepuce is separated from the penile shaft. Thus, strains derived from the Fisher 344 rat should not be used, except in rare cases. The Fisher 344 rat has a different timing of sexual development compared with other more commonly used strains such as Sprague Dawley or Wistar strains (16). If such a strain is to be used, the laboratory should castrate them at a slightly older age and be able to demonstrate the sensitivity of the strain used. The rationale for the choice of rat strain should be clearly stated by the laboratory. Where the screening assay may be preliminary to a repeated dose oral study, a reproductive and developmental study, or a long-term study, preferably animals from the same strain and source should be used in all studies.

Housing and feeding conditions

19. All procedures should conform to all local standards of laboratory animal care. These descriptions of care and treatment are minimum standards and will be superseded by more stringent local regulations, such as Directive 2010/63/EU of the European Parliament and of

the Council of 22 September 2010 on the protection of animals used for scientific purposes (26). The temperature in the experimental animal room should be 22°C (with an approximate range ± 3°C). The relative humidity should be a minimum of 30% and preferably should not exceed a maximum 70%, other than during room cleaning. The aim should be relative humidity of 50-60%. Lighting should be artificial. The daily lighting sequence should be 12 hours light, 12 hours dark.

20. Group housing is preferable to isolation because of the young age of the animals and the fact that rats are social animals. Housing of two or three animals per cage avoids crowding and associated stress that may interfere with the hormonal control of the development of the sex accessory tissue. Cages should be thoroughly cleaned to remove possible contaminants and arranged in such a way that possible effects due to cage placement are minimised. Cages of a proper size (~2000 square centimetres) will prevent overcrowding.

21. Each animal should be identified individually (e.g., ear mark or tag) using a humane method. The method of identification should be recorded.

22. Laboratory diet and drinking water should be provided *ad libitum*. Laboratories executing the Hershberger Bioassay should use the laboratory diet normally used in their chemical testing work. In the validation studies of the Bioassay, no effects or variability were observed that were attributable to the diet. The diet used will be recorded and a sample of the laboratory diet should be retained for possible future analysis.

Performance Criteria for androgen-dependent organ weights

23. During the validation study, there was no evidence that a decrease in body weight affected increases or decreases in the growth of tissue weights for target tissues (i.e., that should be weighted in this study).

24. Among the different strains of rat used successfully in the validation programme, androgen-dependent organ weights are larger in the heavier rat strains than in the lighter strains. Therefore, the Hershberger Bioassay performance criteria do not include absolute expected organ weights for positive and negative controls.

25. Because the Coefficient of Variation (CV) for a tissue has an inverse relationship with statistical power, the Hershberger Bioassay performance criteria are based on maximum CV values for each tissue (Table 1). The CVs are derived from the OECD validation studies. In the case of negative outcomes, laboratories should examine the CVs from the control group and the high dose treatment group to determine if the maximum CV performance criteria have been exceeded.

26. The study should be repeated when: 1) three or more of the ten possible individual CVs in the control and high dose treatment groups exceed the maximums designated for agonist and antagonist studies in Table 1 and 2) at least two target tissues were marginally insignificant, i.e., ρ values between 0.05 and 0.10.

Table 1: Maximum allowable CVs Determined for the Target Sex Accessory Tissues for the castrate model in the OECD Validation Studies[1].

Tissue	Antiandrogenic effects	Androgenic effects
Seminal vesicles	40%	40%
Ventral prostate	40%	45%
LABC	20%	30%
Cowper's glands	35%	55%
Glans penis	17%	22%

PROCEDURE

Regulatory compliance and laboratory verification

27. Unlike the Uterotrophic assay (Chapter B.54 of this Annex), a demonstration of laboratory competence prior to the initiation of the study is not necessary for the Hershberger assay because concurrent positive (Testosterone Propionate and Flutamide) and negative controls are run as an integral part of the assay.

Number and condition of animals

28. Each treated and control group should include a minimum of 6 animals. This applies to both the androgenic and antiandrogenic protocols.

Castration

29. There should be an initial acclimatisation period of several days after receipt of the animals to ensure that the animals are healthy and thriving. Since animals castrated before 42 days of age or postnatal day (pnd) 42 may not display preputial separation, animals should be castrated on pnd 42 or thereafter, not before. The animals are castrated under anaesthesia by placing an incision in the scrotum and removing both testes and epididymides with ligation of blood vessels and seminal ducts. After confirming that no bleeding is occurring, the scrotum should be closed with suture or autoclips. Animals should be treated with analgesics for the first few days after surgery to alleviate any post-surgical discomfort. If castrated animals are purchased from an animal supplier, the age of animals and stage of sexual maturity should be assured by the supplier.

---

[1] The threshold CV for a given tissue was identified from a graph of CV values - arranged from smallest sequentially to largest - for all means from all experiments in the validation exercise using a specific model (agonist or antagonist). The threshold CV was read from the point at which the increments between to the next highest CVs in the series are dramatically larger than the preceding few CVs- the "breakpoint". It should be noted that although this analysis identified relatively reliable "breakpoints" for the antagonist model of the assay, CV curves for the agonist assay showed a more uniform increase making identification of a threshold CV by this method somewhat arbitrary.

Acclimatisation after castration

30. The animals should continue acclimation to the laboratory conditions to allow for the regression in the target tissue weights for a minimum of 7 days following castration. Animals should be observed daily, and any animals with evidence of disease or physical abnormalities should be removed. Thus, treatment with initiation of dosing (on study) may commence as early as pnd 49 days of age, but not later than pnd 60. Age at necropsy should not be greater than pnd 70. This flexibility allows a laboratory to schedule the experimental work efficiently.

Body weight and group randomisation

31. Differences in individual body weights are a source of variability in tissue weights both within and among groups of animals. Increasing tissue weight variability results in an increased coefficient of variation (CV) and decreases the statistical power of the assay (sometimes referred to as assay sensitivity). Therefore, variations in body weight should be both experimentally and statistically controlled.

32. Experimental control involves producing small variations in body weight within and among the study groups. First, unusually small or large animals should be avoided and not placed in the study cohort. At study commencement the weight variation of animals used should not exceed ± 20 % of the mean weight (e.g. 175g ± 35g for castrated peripubertal rats). Second, animals should be assigned to groups (both control and treatment) by randomised weight distribution, so that mean body weight of each group is not statistically different from any other group. The block randomisation procedure used should be recorded.

33. Because toxicity may decrease the body weight of treated groups relative to the control group, the body weight on the first day of test chemical administration could be used as the statistical covariate, not the body weight at necropsy.

Dosage

34. In order to establish whether a test chemical can have androgenic action *in vivo*, two dose groups of the test chemical plus positive and vehicle (negative) controls (See paragraph 43) are normally sufficient, and this design is therefore preferred for animal welfare reasons. If the purpose is either to obtain a dose-response curve or to extrapolate to lower doses, at least 3 dose groups are needed. If information beyond identification of androgenic activity (such as an estimate of potency) is required, a different dosing regimen should be considered. To test for antiandrogens, the test chemical is administered together with a reference androgen agonist. A minimum of 3 test groups with different doses of the test chemical and a positive and a negative control (See paragraph 44) should be used. Except for treatment with the test chemical, animals in the control group should be handled in an identical manner to the test group subjects. If a vehicle is used in administering the test chemical, the control group should receive the vehicle in the highest volume used with the test groups.

35. All dose levels should be proposed and selected taking into account any existing toxicity and

(toxico-) kinetic data available for the test chemical or related materials. The highest dose level should first take into consideration the LD_{50} and/or acute toxicity information in order to avoid death, severe suffering or distress in the animals (17) (18) (19) (20) and, second, take into consideration available information on the doses used in subchronic and chronic studies. In general, the highest dose should not cause a reduction in the final body weight of the animals greater than 10% of control weight. The highest dose should be either 1) the highest dose that ensures animal survival and that is without significant toxicity or distress to the animals after 10 consecutive days of administration up to a maximal dose of 1000 mg/kg/day (See paragraph 36) or 2) a dose inducing (anti)androgenic effects, whichever is lower. As a screen, large intervals, e.g., one half log units (corresponding to a dose progression of 3.2) or even one log units, between dosages are acceptable. If there are no suitable data available, a range finding study (See paragraph 37) may be performed to aid the determination of the doses to be used.

Limit dose level

36. If a test at the limit dose of 1000 mg/kg body weight/day and a lower dose using the procedures described for this study fails to produce a statistically significant change in reproductive organ weights, then additional dose levels may be considered unnecessary. The limit dose applies except when human exposure data indicate the need for a higher dose level to be used.

Considerations for range finding

37. If necessary, a preliminary range finding study can be carried out with a few animals to select the appropriate dose groups [using methods for acute toxicity testing (Chapters B.1 bis, B.1 tris of this Annex (27), OECD TG 425 (19))]. The objective in the case of the Hershberger Bioassay is to select doses that ensure animal survival and that are without significant toxicity or distress to the animals after ten consecutive days of chemical administration up to a limit dose of 1000 mg/kg/d as noted in paragraphs 35 and 36. In this respect an OECD Guidance Document (17) may be used defining clinical signs indicative of toxicity or distress to the animals. If feasible within this range finding study after ten days of administration, the target tissues may be excised and weighed approximately 24-hours after the last dose is administered. These data could then be used to assist the selection of the doses in the main study.

Reference chemicals and vehicle

38. The reference androgen agonist should be Testosterone Propionate (TP), CAS No 57-82-5. The reference TP dosage may be either 0.2 mg/kg-bw/d or 0.4 mg/kg-bw/d. The reference androgen antagonist should be Flutamide (FT), CAS No 1311-84-7. The reference FT dosage should be 3 mg/kg-bw/d, and the FT should be co-administered with the reference TP dosage.

39. It is recommended that, wherever possible, the use of an aqueous solution/suspension be considered first. However, since many androgen ligands or their metabolic precursors tend to

be hydrophobic, the most common approach is to use a solution/suspension in oil (e.g. corn, peanut, sesame or olive oil). Test chemicals can be dissolved in a minimal amount of 95% ethanol or other appropriate solvents and diluted to final working concentrations in the test vehicle. The toxic characteristics of the solvent should be known, and should be tested in a separate solvent-only control group. If the test chemical is considered stable, gentle heating and vigorous mechanical action can be used to assist in dissolving the test chemical. The stability of the test chemical in the vehicle should be determined. If the test chemical is stable for the duration of the study, then one starting aliquot of the test chemical may be prepared, and the specified dosage dilutions prepared daily using care to avoid contamination and spoilage of the samples.

Administration of doses

40. TP should be administered by subcutaneous injection, and FT by oral gavage.

41. The test chemical is administered by oral gavage or subcutaneous injection. Animal welfare considerations and the physical/chemical properties of the test chemical need to be taken into account when choosing the route of administration. In addition, toxicological aspects like the relevance to the human route of exposure to the chemical (e.g. oral gavage to model ingestion, subcutaneous injection to model inhalation or dermal adsorption) and existing toxicological information and data on metabolism and kinetics (e.g. need to avoid first pass metabolism, better efficiency via a particular route) should be taken into account before extensive, long-term testing is initiated if positive results are obtained by injection.

42. The animals should be dosed in the same manner and time sequence for ten consecutive days at approximately 24 hour intervals. The dosage level should be adjusted daily based on the concurrent daily measures of body weight. The volume of dose and time that it is administered should be recorded on each day of exposure. Care should be taken in order not to exceed the maximum dose described in paragraph 35 to allow a meaningful interpretation of the data. Reduction of body weight, clinical signs, and other findings should be thoroughly assessed in this respect. For oral gavage, a stomach tube or a suitable intubation cannula should be used. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. Local animal care guidelines should be followed, but the volume should not exceed 5 ml/kg body weight, except in the case of aqueous solutions where 10 ml/kg body weight may be used. For subcutaneous injections, doses should be administered to the dorsoscapular and or lumbar regions via sterile needle (e.g. 23- or 25-gauge) and a tuberculin syringe. Shaving the injection site is optional. Any losses, leakage at the injection site or incomplete dosing should be recorded. The total volume injected per rat per day should not exceed 0.5 ml/kg body weight.

Specific procedures for androgen agonists

43. For the test for androgen agonists, the vehicle is the negative control, and the TP-treated group is the positive control. Biological activity consistent with androgen agonists is tested by administering a test chemical to treatment groups at the selected doses for 10 consecutive days. The weights of the five sex accessory tissues from the test chemical groups are

compared to the vehicle group for statistically significant increases in weight.

Specific procedures for androgen antagonists and 5α-reductase inhibitors

44. For the test for androgen antagonists and 5α-reductase inhibitors, the TP-treated group is the negative control, and the group coadministered with reference doses of TP and FT is the positive control. Biological activity consistent with androgen antagonists and 5α-reductase inhibitors is tested by administering a reference dose of TP and administering the test chemical for 10 consecutive days. The weights of the five sex accessory tissues from the TP plus test chemical groups are compared to the reference TP-only group for statistically significant decreases in weights.

OBSERVATIONS

Clinical observations

45. General clinical observations should be made at least once a day and more frequently when signs of toxicity are observed. Observations should be carried out preferably at the same time(s) each day and considering the period of anticipated peak effects after dosing. All animals should be observed for mortality, morbidity and general clinical signs such as changes in behaviour, skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity (e.g. lacrimation, piloerection, pupil size, unusual respiratory pattern).

46. Any animal found dead should be removed and disposed of without further data analysis. Any mortality of animals prior to necropsy should be included in the study record together with any apparent reasons for mortality. Any moribund animals should be humanely terminated. Any moribund and subsequently euthanised animals should be included in the study record with apparent reasons for morbidity.

Body weight and food consumption

47. All animals should be weighed daily to the nearest 0.1 g, starting just prior to initiation of treatment i.e., when the animals are allocated into groups. As an optional measurement, the amount of food consumed during the treatment period may be measured per cage by weighing the feeders. The food consumption results should be expressed in grams per rat per day.

Dissection and measurement of tissue and organ weights

48. Approximately 24 hours after the last administration of the test chemical, the rats should be euthanised and exsanguinated according to the normal procedures of the conducting laboratory, and necropsy carried out. The method of humane killing should be recorded in the laboratory report.

49. Ideally, the necropsy order should be randomised across groups to avoid progression directly up or down dose groups that could affect the data. Any finding at necropsy, i.e., pathological changes/visible lesions should be noted and reported.

50. The five androgen-dependent tissues (VP, SV, LABC, COW, GP) should be weighted. These tissues should be excised, carefully trimmed of excess adhering tissue and fat, and their fresh (unfixed) weights determined. Each tissue should be handled with particular care to avoid the loss of fluids and to avoid desiccation, which may introduce significant errors and variability by decreasing the recorded weights. Several of the tissues may be very small or difficult to dissect, and this will introduce variability. Therefore, it is important that persons carrying out the dissection of the sex accessory tissues are familiar with standard dissection procedures for these tissues. A standard operating procedure (SOP) manual for dissection is available from the OECD (21). Careful training according to the SOP guide will minimise a potential source of variation in the study. Ideally the same prosector should be responsible for the dissection of a given tissue to eliminate inter-individual differences in tissue processing. If this is not possible, the necropsy should be designed such that each prosector dissects a given tissue from all treatment groups as opposed to one individual dissecting all tissues from a control group, while someone else is responsible for the treated groups. Each sex accessory tissues should be weighed without blotting to the nearest 0.1 mg, and the weights recorded for each animal.

51. Several of the tissues may be very small or difficult to dissect, and this will introduce variability. Previous work has indicated a range of coefficient of variations (CVs) that appears to differ based upon the proficiency of the laboratory. In a few cases, large differences in the absolute weights of the tissues such as the VP and COWS have been observed within a particular laboratory.

52. Liver, paired kidney, and paired adrenal weights are optional measurements. Again, tissues should be trimmed free of any adhering fascia and fat. The liver should be weighed and recorded to the nearest 0.1 g and the paired kidneys and paired adrenals should be weighed and recorded to the nearest 0.1 mg. The liver, kidney and adrenals are not only influenced by androgens; they also provide useful indices of systemic toxicity.

53. Measurement of serum luteinising hormone (LH), follicular stimulating hormone (FSH) and testosterone (T) is optional. Serum T levels are useful to determine if the test chemical induces liver metabolism of testosterone, lowering serum levels. Without the T data, such an effect might appear to be via an antiandrogenic mechanism. LH levels provides information about the ability of an antiandrogen to not only reduce organ weights, but also to affect hypothalamic-pituitary function, which in long term studies can induce testis tumors. FSH is an important hormone for spermatogenesis. Serum T4 and T3 also are optional measures that would provide useful supplemental information about the ability to disrupt thyroid hormone homeostasis. If hormone measurements are to be made, the rats should be anesthetised prior to necropsy and blood taken by cardiac puncture, and the method of anaesthesia should be chosen with care so that it does not affect hormone measurement. The method of serum preparation, the source of radioimmunoassay or other measurement kits, the analytical procedures, and the results should be recorded. LH levels should be reported as

ng per ml of serum, and T should also be reported as ng per ml of serum.

54. The dissection of the tissues is described as follows with a detailed dissection guide with photographs published as supplementary materials as part of the validation programme (21). A dissection video is also available from the Korea Food and Drug Administration web page (22).

- With the ventral surface of the animal upwards, determine if the prepuce of the penis has separated from the glans penis. If so, then retract the prepuce and remove the glans penis, weigh (nearest 0.1 mg), and record the weight;

- Open the abdominal skin and wall, exposing the viscera. If the optional organs are weighed, remove and weigh liver to nearest 0.1 g, remove the stomach and intestines, remove and weigh the paired kidneys and paired adrenals to the nearest 0.1 mg. This dissection exposes the bladder and begins the dissection of the target male accessory tissues.

- To dissect the VP, separate bladder from the ventral muscle layer by cutting connective tissue along the midline. Displace the bladder anteriorly towards the seminal vesicles (SV), revealing the left and right lobes of the ventral prostate (covered by a layer of fat). Carefully tease the fat from the right and left lobes of the VP. Gently displace the VP right lobe from the urethra and dissect the lobe from the urethra. While still holding the VP right lobe, gently displace the VP left lobe from the urethra and then dissect; weigh to nearest 0.1 mg and record the weight.

- To dissect the SVCG, displace the bladder caudally, exposing the vas deferens and right and left lobes of the seminal vesicles plus coagulating glands (SVCG). Prevent leakage of fluid by clamping a haemostat at the base of the SVCGs, where the vas deferens joins the urethra. Carefully dissect the SVCGs, with the haemostat in place trim fat and adnexa away, place in a tared weigh-boat, remove the haemostat, and weigh to the nearest 0.1 mg and record the weight.

- To dissect the levator ani plus bulbocavernosus muscles (LABC), the muscles and the base of the penis are exposed. The LA muscles wrap around the colon, while the anterior LA and BC muscles are attached to the penile bulbs. The skin and adnexa from the perianal region extending from the base of the penis to the anterior end of the anus are removed. The BC muscles are gradually dissected from the penile bulb and tissues. The colon is cut in two and, the full LABC can be dissected and removed. The LABC should be trimmed of fat and adnexa, weighed to the nearest 0.1 mg, and record the weight.

- After the LABC has been removed, the round Cowper's or bulbourethral glands (COW) are visible at the base of, and slightly dorsal to, the penile bulbs. Careful dissection is required to avoid nicking the thin capsule in order to prevent fluid leakage. Weigh the paired COW to the nearest 0.1 mg, and record the weight.

- In addition, if fluid is lost from any gland during the necropsy and dissection, this

should be recorded.

55. If the evaluation of each chemical requires necropsy of more animals than is reasonable for a single day, the study start may be staggered on two consecutive days, resulting in the staggering of the necropsy and the related work over two days. If staggered in this manner, one-half of the animals per treatment group should be used per day.

56. Carcasses should be disposed of in an appropriate manner following necropsy.

REPORTING

Data

57. Data should be reported individually (i.e. body weight, accessory sex tissue weights, optional measurements and other responses and observations) and for each group of animals (means and standard deviations of all measurement taken). The data should be summarised in tabular form. The data should show the number of animals at the start of the test, the number of animals found dead during the test or found showing signs of toxicity, a description of the signs of toxicity observed, including time of onset, duration and severity.

58. A final report should include:

Testing facility

- Name of facility, location
- Study director and other personnel and their study responsibilities
- Dates the study began and ended, i.e., first day of test chemical administration and last day of necropsy, respectively.

Test chemical

- Source, lot/batch number, identity, purity, full address of the supplier and characterisation of the test chemical(s)
- Physical nature and, where relevant, physicochemical properties;
- Storage conditions and the method and frequency of dilution preparation
- Any data generated on stability
- Any analyses of dosing solutions/suspensions.

Vehicle

- Characterization of the vehicle (identity, supplier and lot #)
- Justification of the vehicle choice (if other than water)

Test animals and animal husbandry procedures

- Species/strain used and rationale for choice

- Source or supplier of animals, including full address

- Number and age of animals supplied

- Housing conditions (temperature, lighting, and so on)

- Diet (name, type, supplier, lot number, content and if known, phytooestrogens levels)

- Bedding (name, type, supplier, content)

- Caging conditions and number of animals per cage;

Assay Conditions

- Age at castration and duration of acclimatisation after castration;

- Individual weights of animals at the start of the study (to nearest 0.1 g);

- Randomisation process and a record of the assignment to vehicle, reference, test chemical groups, and cages

- Mean and standard deviation of the body weights for each group for each weigh day throughout the study;

- Rationale for dose selection

- Route of administration of test chemical and rationale for the choice of exposure route

- If an assay for antiandrogenicity, the TP treatment (dose and volume),

- Test chemical treatment (dose and volume),

- Time of dosing

- Necropsy procedures, including means of exsanguinations and any anesthesia

- If serum analyses are performed, details of the method should be supplied. For example, if RIA is used, the RIA procedure, source of RIA kits, kit expiration dates, procedure for scintillation counting, and standardisation should be reported.

Results

- Daily observations for each animal during dosing, including:
    - Body weights (to the nearest 0.1 g),
    - Clinical signs (if any),
    - Any measurement or notes of food consumption.

- Necropsy observations for each animal, including:
    - Date of necropsy,
    - Animal treatment group,
    - Animal ID,

- Prosector,

- Time of day necropsy and dissection are performed,

- Animal age,

- Final body weight at necropsy, noting any statistically significant increase or decrease,

- Order of animal exsanguination and dissection at necropsy,

- Weights of the five target androgen dependent tissues:

  - Ventral prostate (to the nearest 0.1 mg)

  - Seminal vesicles plus coagulating glands, including fluid (paired, to nearest 0.1 mg)

  - Levator ani plus bulbocavernosus muscle complex (to nearest 0.1 mg)

  - Cowper's glands (fresh weight – paired, to nearest 0.1 mg).

  - Glans penis (fresh weight to nearest 0.1 mg)

- Weights of optional tissues, if performed:

  - Liver (to nearest 0.1 g)

  - Kidney (paired, to nearest 0.1 mg)

  - Adrenal (paired, to nearest 0.1 mg)

- General remarks and comments

- Analyses of serum hormones, if performed.

  - - Serum LH (optional – ng per ml of serum), and

  - - Serum T (optional – ng per ml of serum)

  - General remarks and comments

Data summarisation

Data should be summarised in tabular form containing the sample size for each group, the mean of the value, and the standard error of the mean or the standard deviation. Tables should include necropsy body weights, body weight changes from the beginning of dosing until necropsy, target accessory sex tissues weights, and any optional organ weights.

Discussion of the results

Analysis of results

59. Necropsy body and organ weights should be statistically analysed for characteristics such as homogeneity of variance with appropriate data transformations as needed. Treatment groups should be compared to a control group using techniques such as ANOVA followed by

pairwise comparisons (e.g. Dunnett's one tailed test) and the criterion for statistical difference, for example, p ≤ 0.05. Those groups attaining statistical significance should be identified. However, "relative organ" weights should be avoided due to the invalid statistical assumptions underlying this data manipulation.

60. For androgen agonism, the control should be the vehicle-only test group. The mode of action characteristics of a test chemical can lead to different relative responses amongst the tissues, for example trenbolone, which cannot be 5 alpha-reduced, has more pronounced effects on the LABC and GP than does TP. A statistically significant increase (p≤ 0.05) in any two or more of the five target androgen-dependent tissue weights (VP, LABC, GP, CG and SVCG) should be considered a positive androgen agonist result, and all the target tissues should display some degree of increased growth. Combined evaluation of all accessory sex organs (ASO) tissue responses could be achieved using appropriate multivariate data analysis. This could improve the analysis, especially in cases where only a single tissue gives a statistically significant response.

61. For androgen antagonism, the control should be the reference androgen (testosterone propionate only) test group. The mode of action characteristics of a test chemical can lead to different relative responses amongst the tissues, for example 5 alpha α-reductase inhibitors, like finasteride, have more pronounced effects on the ventral prostate than other tissues as compared to potent AR antagonists, like flutamide. A statistically significant reduction (p ≤ 0.05) in any two or more of the five target androgen-dependent tissue weights (VP, LABC, GP, CG and SVCG) relative to TP treatment alone should be considered a positive androgen antagonist result and all the target tissues should display some degree of reduced growth. Combined evaluation of all ASO tissue responses could be achieved using appropriate multivariate data analysis. This could improve the analysis, especially in cases where only a single tissue gives a statistically significant response.

62. Data should be summarised in tabular form containing the mean, standard error of the mean (standard deviation would also be acceptable) and sample size for each group. Individual data tables should also be included. The individual values, mean, SE (SD) and CV values for the control data should be examined to determine if they meet acceptable criteria for consistency with expected historical values. CVs that exceed CV values listed in Table 1 (see paragraphs 25 and 26) for each organ weight should determine if there are errors in data recording or entry or if the laboratory has not yet mastered accurate dissection of the androgen-dependent tissues and further training/practice is warranted. Generally, CVs (the standard deviation divided by the mean organ weight) are reproducible from lab to lab and study to study. Data presented should include at least: ventral prostate, seminal vesicle, levator ani plus bulbocavernosus, Cowper's glands, glans penis, liver, and body weights and body weight change from the beginning of dosing until necropsy. Data also may be presented after covariance adjustment for body weight, but this should not replace presentation of the unadjusted data. In addition, if preputial separation (PPS) does not occur in any of the groups, the incidence of PPS should be recorded and statistically compared to the control group using Fisher Exact test.

63. When verifying the computer data entries with the original data sheets for accuracy, organ

weight values that are not biologically plausible or vary by more than three standard deviations from that treatment group means should be carefully scrutinised and may need to be discarded, likely being recording errors.

64. Comparison of study results with OECD CV values (in Table 1) is often an important step in interpretation as to the validity of the study results. Historical data for vehicle control groups should be maintained in the laboratory. Historical data for responses to positive reference chemicals, such as TP and FT, should also be maintained in the laboratory. Laboratories may also periodically test the response to known weak androgen agonists and antagonists and maintain these data. These data can be compared to available OECD data to ensure that the laboratory's methods yield sufficient statistical precision and power.

Appendix 1

DEFINITIONS

Androgenic is a term used to describe a positive influence on the growth of androgen-dependent tissues

Antiandrogenic is the capability of a chemical to suppress the action of TP in a mammalian organism.

Chemical means a substance or a mixture.

Date of birth is postnatal day 0.

Dose is the amount of test chemical administered. For the Hershberger Bioassay, the dose is expressed as weight of test chemical per unit body weight of test animal per day (e.g. mg/kg body weight/day).

Dosage is a general term comprising of dose, its frequency and the duration of dosing.

Moribund is a term used to describe an animal in a dying state, i.e., near the point of death.

Postnatal day X is the Xth day of life after the day of birth.

Sensitivity is the capability of a test method to correctly identify chemicals having the property that is being tested for.

Specificity is the capability of a test method to correctly identify chemicals not having the property that is being tested for.

Test chemical means any substance or mixture tested using this test method.

Validation is a scientific process designed to characterize the operational requirements and limitations of a test method and to demonstrate its reliability and relevance for a particular purpose.

Appendix 2

Note: Document prepared by the Secretariat of the Test Guidelines Programme based on the agreement reached at the 6th Meeting of the EDTA Task Force

OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals

Level 1
Sorting & prioritization based upon existing information

- physical & chemical properties, e.g., MW, reactivity, volatility, biodegradability,
- human & environmental exposure, e.g., production volume, release, use patterns
- hazard, e.g., available toxicological data

Level 2
In vitro assays providing mechanistic data

- ER, AR, TR receptor binding affinity
- Transcriptional activation
- Aromatase and steroidogenesis *in vitro*
- Aryl hydrocarbon receptor recognition/binding
- QSARs
- High Through Put Prescreens
- Thyroid function
- Fish hepatocyte VTG assay
- Others (as appropriate)

Level 3
In vivo assays providing data about single endocrine Mechanisms and effects

- Uterotrophic assay (estrogenic related)
- Hershberger assay (androgenic related)
- Non -receptor mediated hormone function
- Others (e.g. thyroid)
- Fish VTG (vitellogenin) assay (estrogenic related)

Level 4
In vivo assays providing data about multiple endocrine Mechanisms and effects

- enhanced OECD 407 (endpoints based on endocrine mechanisms)
- male and female pubertal assays
- adult intact male assay
- Fish gonadal histopathology assay
- Frog metamorphosis assay

Level 5
In vivo assays providing data on effects from endocrine & other mechanisms

- 1-generation assay (TG415 enhanced)[1]
- 2-generation assay (TG416 enhanced)[1]
- reproductive screening test (TG421 enhanced)[1]
- combined 28 day/reproduction screening test (TG 422 enhanced)[1]

[1] Potential enhancements will be considered by VMG mamm

- Partial and full life cycle assays in fish, birds, amphibians & invertebrates (developmental and reproduction)

VMG   mamm: Validation   Management   Group   on   Mammalian   Testing   and   Assessment

NOTES TO THE FRAMEWORK

Note 1: Entering at all levels and exiting at all levels is possible and depends upon the nature of existing information needs for hazard and risk assessment purposes

Note 2: In level 5, ecotoxicology should include endpoints that indicate mechanisms of adverse effects, and potential population damage

Note 3: When a multimodal model covers several of the single endpoint assays, that model would replace the use of those single endpoint assays

Note 4: The assessment of each chemical should be based on a case by case basis, taking into account all available information, bearing in mind the function of the framework levels.

Note 5: The framework should not be considered as all inclusive at the present time. At levels 3,4 and 5 it includes assays that are either available or for which validation is under way. With respect to the latter, these are provisionally included. Once developed and validated, they will be formally added to the framework.

Note 6: Level 5 should not be considered as including definitive tests only. Tests included at that level are considered to contribute to general hazard and risk assessment.

LITERATURE

(1) OECD (1998). Report of the First Meeting of the OECD Endocrine Disrupter Testing and Assessment (EDTA) Task Force, 10th-11th March 1998, ENV/MC/CHEM/RA(98)5.

(2) Dorfman RI (1962). Standard methods adopted by official organization. Academic Press, NY.

(3) Gray LE Jr, Furr J and Ostby JS (2005). Hershberger assay to investigate the effects of endocrine disrupting compounds with androgenic and antiandrogenic activity in castrate-immature male rats. In: *Current Protocols in Toxicology* 16.9.1-16.9.15. J Wiley and Sons Inc.

(4) OECD (2006). Final OECD report of the initial work towards the validation of the rat Hershberger assay. Phase 1. Androgenic response to testosterone propionate and anti-androgenic effects of flutamide. Environmental Health and Safety, Monograph Series on Testing and Assessment N° 62. ENV/JM/MONO(2006)30.

(5) OECD (2008). Report of the OECD Validation of the Rat Hershberger Bioassay: Phase 2: Testing of Androgen Agonists, Androgen Antagonists and a 5α-Reductase Inhibitor in Dose Response Studies by Multiple Laboratories. Environmental Health and Safety, Monograph Series on Testing and Assessment N° 86. ENV/JM/MONO(2008)3.

(6) OECD (2007). Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol. Environmental Health and Safety, Monograph Series on Testing and Assessment N° 73. ENV/JM/MONO(2007)20.

(7) Owens, W, Zeiger E, Walker M, Ashby J, Onyon L, Gray, Jr, LE (2006). The OECD programme to validate the rat Hershberger bioassay to screen compounds for in vivo androgen and antiandrogen responses. Phase 1: Use of a potent agonist and a potent antagonist to test the standardized protocol. Env. Health Persp. 114:1265-1269.

(8) Owens W, Gray LE, Zeiger E, Walker M, Yamasaki K, Ashby J, Jacob E (2007). The OECD program to validate the rat Hershberger bioassay to screen compounds for in vivo androgen and antiandrogen responses: phase 2 dose-response studies. Environ Health Perspect. 115(5):671-8.

(9) Korenchevsky V (1932). The assay of testicular hormone preparations. Biochem J 26:413-422.

(10) Korenchevsky V, Dennison M, Schalit R (1932). The response of castrated male rats to the injection of the testicular hormone. Biochem J 26:1306-1314.

(11) Eisenberg E, Gordan GS (1950). The levator ani muscle of the rat as an index of myotrophic activity of steroidal hormones. J Pharmacol Exp Therap 99:38-44.

(12) Eisenberg E, Gordan GS, Elliott HW (1949). Testosterone and tissue respiration of the castrate male rat with a possible test for mytrophic activity. Endocrinology 45:113-119.

(13) Hershberger L, Shipley E, Meyer R (1953). Myotrophic activity of 19-nortestosterone and other steroids determined by modified levator ani muscle method. Proc Soc Exp Biol Med 83:175-180.

(14) Hilgar AG, Vollmer EP (1964). Endocrine bioassay data: Androgenic and myogenic. Washington DC: United States Public Health Service.

(15) Dorfman RI (1969). Androgens and anabolic agents. In: Methods in Hormone Research, volume IIA. (Dorfman RI, ed.) New York:Academic Press, 151-220.

(16) Massaro EJ (2002). Handbook of Neurotoxicology, volume I. New York: Humana Press, p 38.

(17) OECD (2000). Guidance document on the recognition, assessment and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No 19. ENV/JM/MONO(2000)7.

(18) OECD (1982). Organization for Economic Co-operation and Development - Principles of Good Laboratory Practice, ISBN 92-64-12367-9, Paris.

(19) OECD (2008). Acute oral toxicity – up-and-down procedure. OECD Guideline for the testing of chemicals No 425.

(20) OECD (2001). Guidance document on acute oral toxicity . Environmental Health and Safety Monograph Series on Testing and Assessment No 24. ENV/JM/MONO(2001)4.

(21) Supplemental materials for Owens et al. (2006). The OECD programme to validate the rat Hershberger bioassay to screen compounds for in vivo androgen and antiandrogen responses. Phase 1: Use of a potent agonist and a potent antagonist to test the standardized protocol. Env. Health Persp. 114:1265-1269. See, section II, The dissection guidance provided to the laboratories: http://www.ehponline.org/docs/2006/8751/suppl.pdf.

(22) Korea Food and Drug Administration. Visual reference guide on Hershberger assay procedure, including a dissection video. http://rndmoa.kfda.go.kr/endocrine/reference/education_fr.html .

(23) OECD (2008). Background Review Document on the Rodent Hershberger Bioassay. Environmental Health and Safety Monograph Series on Testing and Assessment No 90. ENV/JM/MONO(2008)17.

(24) OECD (2008). Draft Validation report of the Intact, Stimulated, Weanling Male Rat Version of the Hershberger Bioassay.

(25) OECD (2009). Guidance Document on the Weanling Hershberger Bioassay in rats: A shortterm screening assay for (anti)androgenic properties. Series on Testing and Assessment, Number 115.

(26) Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33.

(27) The following chapters of this Annex:
   B.1 bis, Acute oral toxicity – fixed dose procedure
   B.1 tris, Acute oral toxicity – acute toxic class method

B.56 Extended One-Generation Reproductive Toxicity Study

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 443 (2012). It is based on the International Life Science Institute (ILSI)-Health and Environmental Sciences Institute (HESI), Agricultural Chemical Safety Assessment (ACSA) Technical Committee proposal for a life stage F_1 extended one generation reproductive study as published in Cooper et al., 2006 (1). Several improvements and clarifications have been made to the study design to provide flexibility and to stress the importance of starting with existing knowledge, while using in-life observations to guide and tailor the testing. This test method provides a detailed description of the operational conduct of an Extended One-Generation Reproductive Toxicity Study. The test method describes three cohorts of F_1 animals:

    -   Cohort 1: assesses reproductive/developmental endpoints; this cohort may be extended to include an F_2 generation.

    -   Cohort 2: assesses the potential impact of chemical exposure on the developing nervous system.

    -   Cohort 3: assesses the potential impact of chemical exposure on the developing immune system.

2. Decisions on whether to assess the second generation and to omit the developmental neurotoxicity cohort and/or developmental immunotoxicity cohort should reflect existing knowledge for the chemical being evaluated, as well as the needs of various regulatory authorities. The purpose of the test method is to provide details on how the study can be conducted and to address how each cohort should be evaluated.

3. Procedure for the decision on the internal triggering for producing a 2nd generation is described in OECD Guidance Document 117 (39) for those regulatory authorities using internal triggers.
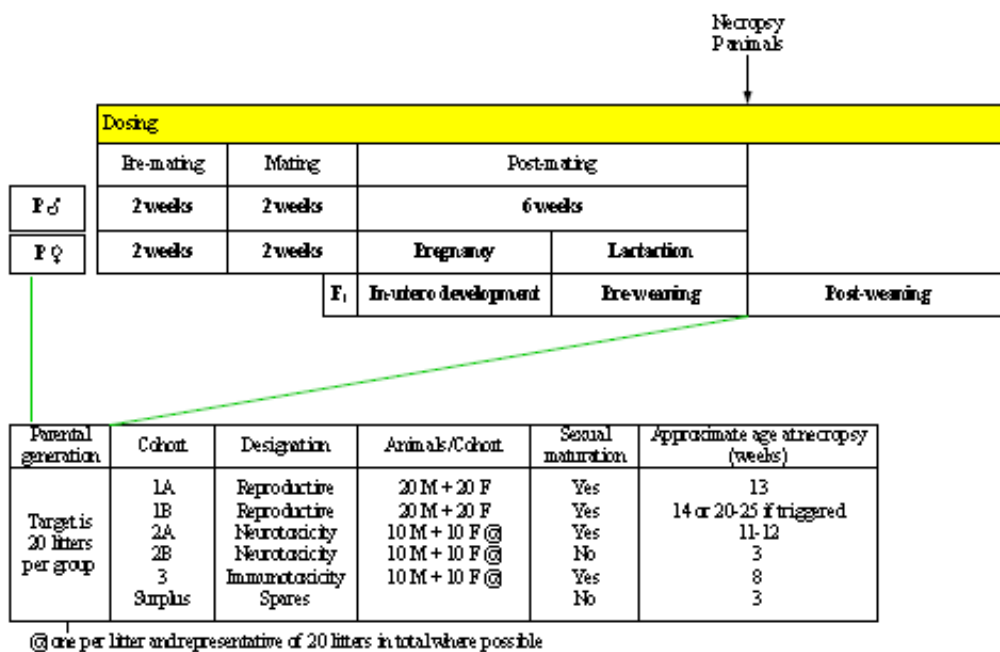

INITIAL CONSIDERATIONS AND OBJECTIVES

4. The main objective of the Extended One-Generation Reproductive Toxicity Study is to evaluate specific life stages not covered by other types of toxicity studies and test for effects that may occur as a result of pre- and postnatal chemical exposure. For reproductive endpoints, it is envisaged that, as a first step and when available, information from repeat-dose studies (including screening reproductive toxicity studies, *e.g.* OECD TG 422 (32)), or short term endocrine disrupter screening assays, (*e.g.* Uterotrophic assay - test method B.54 (36); and Hershberger assay – test method B.55 (37)) is used to detect effects on reproductive organs for males and females. This might include spermatogenesis (testicular histopathology) for males and oestrous cycles, follicle counts/oocyte maturation and ovarian integrity (histopathology) for females. The Extended One-Generation Reproductive Toxicity Study then serves as a test for reproductive endpoints that require the interaction of males with females, females with conceptus, and females with offspring and the F_1 generation until after sexual maturity

(see OECD Guidance Document 151 supporting this test method (40)).

5. The test method is designed to provide an evaluation of the pre- and postnatal effects of chemicals on development as well as a thorough evaluation of systemic toxicity in pregnant and lactating females and young and adult offspring. Detailed examination of key developmental endpoints, such as offspring viability, neonatal health, developmental status at birth, and physical and functional development until adulthood, is expected to identify specific target organs in the offspring. In addition, the study will provide and/or confirm information about the effects of a test chemical on the integrity and performance of the adult male and female reproductive systems. Specifically, but not exclusively, the following parameters are considered: gonadal function, the oestrous cycle, epididymal sperm maturation, mating behaviour, conception, pregnancy, parturition, and lactation. Furthermore, the information obtained from the developmental neurotoxicity and developmental immunotoxicity assessments will characterise potential effects in those systems. The data derived from these tests should allow the determination of No-Observed Adverse Effect Levels (NOAELs), Lowest Observed Adverse Effect Levels (LOAELs) and/or benchmark doses for the various endpoints and/or be used to characterise effects detected in previous repeat-dose studies and/or serve as a guide for subsequent testing.

6. A schematic drawing of the protocol is presented in Figure 1. The test chemical is administered continuously in graduated doses to several groups of sexually mature males and females. This parental (P) generation is dosed for a defined pre-mating period (selected based on the available information for the test chemical; but for a minimum of two weeks) and a two-week mating period. P males are further treated at least until weaning of the F_1. They should be treated for a minimum of 10 weeks. They may be treated for longer if there is a need to clarify effects on reproduction. Treatment of the P females is continued during pregnancy and lactation until termination after the weaning of their litters (*i.e.* 8-10 weeks of treatment). The F_1 offspring receive further treatment with the test chemical from weaning to adulthood. If a second generation is assessed (see OECD Guidance Document 117 (39)), the F_1 offspring will be maintained on treatment until weaning of the F_2, or until termination of the study.

7. Clinical observations and pathology examinations are performed on all animals for signs of toxicity, with special emphasis on the integrity and performance of the male and female reproductive systems and the health, growth, development and function of the offspring. At weaning, selected offspring are assigned to specific subgroups (cohorts 1-3, see paragraphs 33 and 34 and Figure 1) for further investigations, including sexual maturation, reproductive organ integrity and function, neurological and behavioural endpoints, and immune functions.

8. In conducting the study, the guiding principles and considerations outlined in the OECD Guidance Document n° 19 on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations (34) should be followed.

9. When a sufficient number of studies are available to ascertain the impact of this new study design, the test method will be reviewed and if necessary revised in light of experience gained.

Figure 1: Scheme of the Extended One-Generation Reproductive Toxicity Study

Parental generation	Cohort	Designation	Animals/Cohort	Sexual maturation	Approximate age at necropsy (weeks)
	1A	Reproductive	20 M + 20 F	Yes	13
	1B	Reproductive	20 M + 20 F	Yes	14 or 20-25 if triggered
Target is 20 litters per group	2A	Neurotoxicity	10 M + 10 F @	Yes	11-12
	2B	Neurotoxicity	10 M + 10 F @	No	3
	3	Immunotoxicity	10 M + 10 F @	Yes	8
	Surplus	Spares		No	3

@ one per litter and representative of 20 litters in total where possible

DESCRIPTION OF THE METHOD/PREPARATIONS FOR THE TEST

Animals

Selection of animal species and strain

10. The choice of species for the reproductive toxicity test should be carefully considered in light of all available information. However, because of the extent of background data and the comparability to general toxicity tests, the rat is normally the preferred species, and criteria and recommendations given in this test method refer to this species. If another species is used, justification should be given and appropriate modifications to the protocol will be necessary. Strains with low fecundity or a well-known high incidence of spontaneous developmental defects should not be used.

Age, body weight and inclusion criteria

11. Healthy parental animals, which have not been subjected to previous experimental procedures, should be used. Both males and females should be studied and the females should be nulliparous and non-pregnant. The P animals should be sexually mature, of similar weight (within sex) at initiation of dosing, similar age (approximately 90 days) at mating, and representative of the species and strain under study. Animals should be acclimated for at least 5 days after arrival. The animals are randomly assigned to the control and treatment groups, in a manner, which results in comparable mean body weight values among the groups (*i.e.* ± 20% of the mean).

87

Housing and feeding conditions

12. The temperature in the experimental animal room should be 22 °C (± 3°). Relative humidity should be between 30-70%, with an ideal range of 50-60%. Artificial lighting should be set at 12 hours light, 12 hours dark. Conventional laboratory diets may be used with an unlimited supply of drinking water. Careful attention should be given to diet phytoestrogen content, as a high level of phytoestrogen in the diet might affect some reproductive endpoints. Standardised, open-formula diets in which estrogenic chemicals have been reduced are recommended (2) (30). The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this method. Content, homogeneity and stability of the test chemical in the diets should be verified. The feed and drinking water should be regularly analysed for contaminants. Samples of each batch of the diet used during the study should be retained under appropriate conditions (*e.g.* frozen at -20°C), until finalisation of the report, in case the results necessitate a further analysis of diet ingredients.

13. Animals should be caged in small groups of the same sex and treatment group. They may be housed individually to avoid possible injuries (*e.g.* males after the mating period). Mating procedures should be carried out in suitable cages. After evidence of copulation, females that are presumed to be pregnant are housed separately in parturition or maternity cages where they are provided with appropriate and defined nesting materials. Litters are housed with their mothers until weaning. F_1 animals should be housed in small groups of the same sex and treatment group from weaning to termination. If scientifically justified, animals can be housed individually. The level of phytoestrogens contained in the selected bedding material should be minimal.

Number and identification of animals

14. Normally, each test and control group should contain a sufficient number of mating pairs to yield at least 20 pregnant females per dose group. The objective is to produce enough pregnancies to ensure a meaningful evaluation of the potential of the chemical to affect fertility, pregnancy and maternal behaviour of the P generation and growth and development of the F_1 offspring, from conception to maturity. Failure to achieve the desired number of pregnant animals does not necessarily invalidate the study and should be evaluated on a case-by-case basis, considering a possible causal relationship to the test chemical.

15. Each P animal is assigned a unique identification number before dosing starts. If laboratory historical data suggest that a significant proportion of females may not show regular (4 or 5-day) oestrous cycles, then an assessment of oestrous cycles before start of treatment is advised. Alternatively, the group size may be increased to ensure that at least 20 females in each group would have regular (4 or 5-day) oestrous cycles at start of treatment. All F_1 offspring are uniquely identified when neonates are first examined on postnatal day (PND) 0 or 1. Records indicating the litter of origin should be maintained for all F_1 animals, and F_2 animals where applicable, throughout the study.

Test chemical

Available information on the test chemical

16. The review of existing information is important for decisions on the route of administration, the choice of the vehicle, the selection of animal species, the selection of dosages and potential modifications of the dosing schedule. Therefore, all the relevant available information on the test chemical, *i.e.* physico-chemical, toxicokinetics (including species-specific metabolism), toxicodynamic properties, structure-activity relationships (SARs), *in vitro* metabolic processes, results of previous toxicity studies and relevant information on structural analogues should be taken into consideration in planning the Extended One-Generation Reproductive Toxicity Study. Preliminary information on absorption, distribution, metabolism and elimination (ADME) and bioaccumulation may be derived from chemical structure, physico-chemical data, extent of plasma protein binding or toxicokinetic (TK) studies, while results from toxicity studies give additional information, *e.g.* on NOAEL, metabolism or induction of metabolism.

Consideration of toxicokinetic data

17. Although not required, TK data from previously conducted dose range-finding or other studies are extremely useful in the planning of the study design, selection of dose levels and interpretation of results. Of particular utility are data which: 1) verify exposure of developing foetuses and pups to the test chemical (or relevant metabolites), 2) provide an estimate of internal dosimetry, and 3) evaluate for potential dose-dependent saturation of kinetic processes. Additional TK data, such as metabolite profiles, concentration-time courses, etc. should also be considered, if they are available. Supplemental TK data may also be collected during the main study, provided that it does not interfere with the collection and interpretation of the main study endpoints.

As a general guide, the following TK data set would be useful in planning the Extended One-Generation Reproductive Toxicity Study:

- Late pregnancy (*e.g.* Gestation Day 20) - maternal blood and foetal blood
- Mid-lactation (PND 10) - maternal blood, pup blood and/or milk
- Early post-weaning (*e.g.* PND 28) - weanling blood samples.

Flexibility should be employed in determining the specific analytes (e.g. parent chemical and/or metabolites) and sampling scheme. For example, the number and timing of sample collection on a given sampling day will be dependent upon route of exposure and prior knowledge of TK properties in non-pregnant animals. For dietary studies, sampling at a single consistent time on each of these days is sufficient, whereas gavage dosing may warrant additional sampling times to obtain a better estimate of the range of internal doses. However, it is not necessary to generate a full concentration time-course on any of the sampling days. If necessary, blood can be pooled by sex within litters for fetal and neonatal analyses.

Route of administration

18. Selection of the route should take into consideration the route(s) most relevant for human exposure. Although the protocol is designed for administration of the test chemical through the diet, it can be modified for administration by other routes (drinking water, gavage, inhalation, dermal), depending on the characteristics of the

chemical and the information required.

Choice of the vehicle

19. Where necessary, the test chemical is dissolved or suspended in a suitable vehicle. It is recommended that, where possible, the use of an aqueous solution/suspension is considered first, followed by consideration of a solution/suspension in oil (*e.g.* corn oil). For vehicles other than water, the toxic characteristics of the vehicle should be known. Use of vehicles with potential intrinsic toxicity should be avoided (*e.g.* acetone, DMSO). The stability of the test chemical in the vehicle should be determined. Considerations should be given to the following characteristics if a vehicle or other additive is used to facilitate dosing: effects on the absorption, distribution, metabolism, or retention of the test chemical; effects on the chemical properties of the test chemical that may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals.

Dose selection

20. Normally, the study should include at least three dose levels and a concurrent control. When selecting appropriate dose levels, the investigator should consider all available information, including the dosing information from previous studies, TK data from pregnant or non-pregnant animals, the extent of lactational transfer, and estimates of human exposure. If TK data are available which indicate dose-dependent saturation of TK processes, care should be taken to avoid high dose levels which clearly exhibit saturation, provided of course, that human exposures are expected to be well below the point of saturation. In such cases, the highest dose level should be at, or just slightly above the inflection point for transition to nonlinear TK behaviour.

21. In the absence of relevant TK data, the dose levels should be based on toxic effects, unless limited by the physical/chemical nature of the test chemical. If dose levels are based on toxicity, the highest dose should be chosen with the aim to induce some systemic toxicity, but not death or severe suffering of the animals.

22. A descending sequence of dose levels should be selected in order to demonstrate any dose-related effect and to establish NOAELs or doses near the limit of detection that would allow for derivation of a benchmark dose for the most sensitive endpoint(s). To avoid large dose spacing between NOAELs and LOAELs, two- or four-fold intervals are frequently optimal. The addition of a fourth test group is often preferable to using a very large interval (*e.g.* more than a factor of 10) between doses.

23. Except for treatment with the test chemical, animals in the control group are handled in an identical manner to the test group subjects. This group should be untreated or sham-treated or a vehicle-control group if a vehicle is used in administering the test chemical. If a vehicle is used, the control group should receive the vehicle in the highest volume used.

Limit test

24. If there is no evidence of toxicity at a dose of at least 1000 mg/kg body weight/day in repeat-dose studies, or if toxicity would not be expected based upon data from

structurally- and/or metabolically-related chemicals, indicating similarity in the *in vivo*/*in vitro* metabolic properties, a study using several dose levels may not be necessary. In such cases, the Extended One-Generation Reproductive Toxicity Study could be conducted using a control group and a single dose of at least 1000 mg/kg body weight/day. However, should evidence for reproductive or developmental toxicity be found at this limit dose, further studies at lower dose levels will be required to identify a NOAEL. These limit test considerations apply only when human exposure does not indicate the need for a higher dose level.

PROCEDURES

Exposure of offspring

25. Dietary exposure is the preferred method of administration. If gavage studies are performed, it should be noted that the pups will normally only receive test chemical indirectly through the milk, until direct dosing commences for them at weaning. In diet or drinking water studies, the pups will additionally receive test chemical directly when they commence eating for themselves during the last week of the lactation period. Modifications to the study design should be considered when excretion of the test chemical in milk is poor and where there is lack of evidence for a continuous exposure of the offspring. In these cases, direct dosing of pups during the lactation period should be considered based on available TK information, offspring toxicity or changes in bio-markers (3) (4). Careful consideration of benefits and disadvantages should be made prior to conducting direct-dosing studies on nursing pups (5).

Dosing schedule and administration of doses

26. Some information on oestrous cycles, male and female reproductive tract histopathology and testicular/epididymal sperm analysis may be available from previous repeat-dose toxicity studies of adequate duration. The duration of the pre-mating treatment in the Extended One-Generation Reproductive Toxicity Study is therefore aimed at the detection of effects on functional changes that may interfere with mating behaviour and fertilisation. The pre-mating treatment should be sufficiently long to achieve steady-state exposure conditions in P males and females. A 2-week pre-mating treatment for both sexes is considered adequate in most cases. For females, this covers 3-4 complete oestrous cycles and should be sufficient to detect any adverse effects on cyclicity. For males, this is equivalent to the time required for epididymal transit of maturing spermatozoa and should allow the detection of post-testicular effects on sperm (during the final stages of spermiation and epididymal sperm maturation) at mating. At the time of termination, when testicular and epididymal histopathology and analysis of sperm parameters are scheduled, the P and F_1 males, will have been exposed for at least one entire spermatogenic process ((6) (7) (8) (9) and OECD Guidance Document 151 (40)).

27. Pre-mating exposure scenarios for males could be adapted if testicular toxicity (impairment of spermatogenesis) or effects on sperm integrity and function have been clearly identified in previous studies. Similarly, for females, known effects of the test chemical on the oestrous cycle and thus sexual receptivity, may justify different pre-mating exposure scenarios. In special cases it may be acceptable that treatment of the P females is initiated only after a sperm-positive smear has been obtained (see OECD

Guidance Document 151 (40)).

28. Once the pre-mating dosing period is established, the animals should be treated with the test chemical continuously on a 7-days/week basis until necropsy. All animals should be dosed by the same method. Dosing should continue during the 2-week mating period and, for P females, throughout gestation and lactation up to the day of termination after weaning. Males should be treated in the same manner until termination at the time when the F_1 animals are weaned. For necropsy, priority should be given to females which should be necropsied on the same/similar day of lactation. Necropsy of males can be spread over a larger number of days, depending on laboratory facilities. Unless already initiated during the lactation period, direct dosing of the selected F_1 males and females should begin at weaning and continue until scheduled necropsy, depending on cohort assignment.

29. For chemicals administered via the diet or drinking water, it is important to ensure that the quantities of the test chemical involved do not interfere with normal nutrition or water balance. When the test chemical is administered in the diet, either a constant dietary concentration (ppm) or a constant dose level in terms of the body weight of the animal may be employed; the option chosen should be specified.

30. When the test chemical is administered by gavage, the volume of liquid administered at one time should not normally exceed 1 ml/100 g body weight (0.4 ml/100 g body weight is the maximum for oil, *e.g.* corn oil). Except for irritant or corrosive chemicals, which will normally reveal exacerbated effects with higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure a constant volume at all dose levels. The treatment should be given at similar times each day. The dose to each animal should normally be based on the most recent individual body weight determination and adjusted at least weekly in adult males and adult non-pregnant females, and every two days in pregnant females and F_1 animals when administered prior to weaning and during the 2 weeks following weaning. If TK data indicate a low placental transfer of the test chemical, the gavage dose during the last week of pregnancy may have to be adjusted to prevent administration of an excessively toxic dose to the dam. Females should not be treated by gavage, or any other route of treatment where the animal needs to be handled, on the day of parturition; omission of test chemical administration on that day is preferable to a disturbance of the birth process.

Mating

31. Each P female should be placed with a single, randomly selected, unrelated male from the same dose group (1:1 pairing) until evidence of copulation is observed or 2 weeks have elapsed. If there are insufficient males, for example due to male death before pairing, then male(s) which have already mated may be paired (1:1) with a second female(s) such that all females are paired. Day 0 of pregnancy is defined as the day on which mating evidence is confirmed (a vaginal plug or sperm are found). Animals should be separated as soon as possible after evidence of copulation is observed. If mating has not occurred after 2 weeks, the animals should be separated without further opportunity for mating. Mating pairs should be clearly identified in the data.

Litter size

32. On day 4 after birth, the size of each litter may be adjusted by eliminating extra pups by random selection to yield, as nearly as possible, five males and five females per litter. Selective elimination of pups, *e.g.* based upon body weight, is not appropriate. Whenever the number of male or female pups prevents having five of each sex per litter, partial adjustment (for example, six males and four females) is acceptable.

Selection of pups for post-weaning studies (see Figure 1)

33. At weaning (around PND 21) pups from all available litters up to 20 per dose and control group are selected for further examinations and maintained until sexual maturation (unless earlier testing is required). Pups are selected randomly, with the exception that obvious runts (animals with a body weight more than two standard deviations below the mean pup weight of the respective litter) should not be included, as they are unlikely to be representative of the treatment group.

On PND 21, the selected F_1 pups are randomly assigned to one of three cohorts of animals, as follows:

  - Cohort 1 (1A and 1B) = Reproductive/developmental toxicity testing

  - Cohort 2 (2A and 2B) = Developmental neurotoxicity testing

  - Cohort 3 = Developmental immunotoxicity testing

Cohort 1A: One male and one female/litter/group (20/sex/group): priority selection for primary assessment of effects upon reproductive systems and of general toxicity.

Cohort 1B: One male and one female/litter/group (20/sex/group): priority selection for follow-up assessment of reproductive performance by mating F1 animals, when assessed (see OECD Guidance Document 117 (39)), and for obtaining additional histopathology data in cases of suspected reproductive or endocrine toxicants, or when results from cohort 1A are equivocal.

Cohort 2A: Total of 20 pups per group (10 males and 10 females per group; one male or one female per litter) assigned for neurobehavioral testing followed by neurohistopathology assessment as adults.

Cohort 2B: Total of 20 pups per group (10 males and 10 females per group; one male or one female per litter) assigned for neurohistopathology assessment at weaning (PND 21 or PND 22). If there are insufficient numbers of animals, preference should be given to assign animals to Cohort 2A.

Cohort 3: Total of 20 pups per group (10 males and 10 females per group; one per litter, where possible). Additional pups may be required from the control group to act as positive control animals in the T-cell dependant antibody response assay (TDAR) at PND 56 ± 3.

34. Should there be an insufficient number of pups in a litter to serve all cohorts, the cohort 1 takes precedence, as it can be extended to produce an F_2 generation. Additional pups may be assigned to any of the cohorts in case of specific concern, *e.g.* if a chemical is suspected to be a neurotoxicant, immunotoxicant or reproductive toxicant. These pups may be used for examinations at different timepoints or for the evaluation of

supplementary endpoints. Pups not assigned to cohorts will be submitted to clinical biochemistry (paragraph 55) and gross necropsy (paragraph 68).

Second mating of the P animals

35. A second mating is not normally recommended for the P animals, as it comes at the expense of losing important information on the number of implantation sites (and thus post-implantation and peri-natal loss data, indicators of a possible teratogenic potential) for the first litter. The need to verify or elucidate an effect in exposed females would be served better by extending the study to include a mating of the F_1 generation. However, a second mating of the P males with untreated females is always an option to clarify equivocal findings or for further characterisation of effects on fertility observed in the first mating.

IN-LIFE OBSERVATIONS

Clinical observations

36. For the P and the selected F_1 animals, a general clinical observation is made once a day. In the case of gavage dosing, the timing of clinical observations should be prior to and post dosing (for possible signs of toxicity associated with peak plasma concentration). Pertinent behavioural changes, signs of difficult or prolonged parturition and all signs of toxicity are recorded. Twice daily, during the weekend once daily, all animals are observed for severe toxicity, morbidity and mortality.

37. In addition, a more detailed examination of all P and F_1 animals (after weaning) is conducted on a weekly basis and could conveniently be performed on an occasion when the animal is weighed, which would minimise handling stress. Observations should be carefully conducted and recorded using scoring systems that have been defined by the testing laboratory. Efforts should be made to ensure that variations in the test conditions are minimal. Signs noted should include, but not be limited to, changes in skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity (*e.g.* lacrimation, piloerection, pupil size, unusual respiratory pattern). Changes in gait, posture, response to handling, as well as the presence of clonic or tonic movements, stereotypy (*e.g.* excessive grooming, repetitive circling) or bizarre behaviour (*e.g.* self-mutilation, walking backwards) should also be recorded.

Body weight and food/water consumption

38. P animals are weighed on the first day of dosing and at least weekly thereafter. In addition, P females are weighed during lactation on the same days as the weighing of the pups in their litters (see paragraph 44). All F_1 animals are weighed individually at weaning (PND 21) and at least weekly thereafter. Body weight is also recorded on the day when they attain puberty (completion of preputial separation or vaginal patency). All animals are weighed at sacrifice.

39. During the study, food and water consumption (in the case of test chemical administration in the drinking water) are recorded at least weekly on the same days as animal body weights (except during cohabitation). The food consumption of each cage

of F$_1$ animals is recorded weekly commencing with selection to a respective cohort.

Oestrous cycles

40. Preliminary information of test chemical-related effects on the oestrous cycle may already be available from previous repeat-dose toxicity studies, and may be used in designing a test chemical-specific protocol for the Extended One-Generation Reproductive Toxicity Study. Normally the assessment of oestrous cyclicity (by vaginal cytology) will start at the beginning of the treatment period and continue until confirmation of mating or the end of the 2-week mating period. If females have been screened for normal oestrous cycles before treatment, then it is useful to continue smearing as treatment starts, but if there is concern about non-specific effects at the start of treatment (such as an initial marked reduction in food consumption) then animals may be allowed to adapt to treatment for up to two weeks before the start of the 2-week smearing period leading into pairing. If the female treatment period is extended in this way (*i.e.* to a 4-week pre-mating treatment) then consideration should be made to purchasing animals younger and to extending the period of male treatment before pairing. When obtaining vaginal/cervical cells, care should be taken to avoid disturbance of mucosa and subsequently, the induction of pseudopregnancy (10) (11).

41. Vaginal smears should be examined daily for all F$_1$ females in cohort 1A, after the onset of vaginal patency, until the first cornified smear is recorded, in order to determine the time interval between these two events. Oestrous cycles for all F$_1$ females in cohort 1A should also be monitored for a period of two weeks, commencing around PND 75. In addition, should mating of the F$_1$ generation be necessary, the vaginal cytology in cohort 1B will be followed from the time of pairing until mating evidence is detected.

Mating and pregnancy

42. In addition to the standard endpoints (*e.g.* body weight, food consumption, clinical observations including mortality/morbidity checks), the dates of pairing, the date of insemination and the date of parturition are recorded and the precoital interval (pairing to insemination) and the duration of pregnancy (insemination to parturition) are calculated. The P females should be examined carefully at the time of expected parturition for any signs of dystocia. Any abnormalities in nesting behaviour or nursing performance should be recorded.

43. The day on which parturition occurs is lactation day 0 (LD 0) for the dam and postnatal day 0 (PND 0) for the offspring. Alternatively, all comparisons may also be based on post-coital time to eliminate confounding of postnatal development data, by differences in the duration of pregnancy; however, timing relative to parturition should also be recorded. This is especially important when the test chemical exerts an influence on the duration of pregnancy.

Offspring parameters

44. Each litter should be examined as soon as possible after parturition (PND 0 or 1) to establish the number and sex of pups, stillbirths, live births, and the presence of gross anomalies (externally visible abnormalities, including cleft palate; subcutaneous haemorrhages; abnormal skin colour or texture; presence of umbilical cord; lack of milk

in stomach; presence of dried secretions). In addition, the first clinical examination of the neonates should include a qualitative assessment of body temperature, state of activity and reaction to handling. Pups found dead on PND 0 or at a later time should be examined for possible defects and cause of death. Live pups are counted and weighed individually on PND 0 or PND 1, and regularly thereafter, *e.g.* at least on PND 4, 7, 14, and 21. Clinical examinations, as applicable for the age of the animals, should be repeated when the offspring are weighed, or more often if case-specific findings have been made at birth. Signs noted could include, but may not be limited to, external abnormalities, changes in skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity. Changes in gait, posture, response to handling, as well as the presence of clonic or tonic movements, stereotypy or bizarre behaviour, should also be recorded.

45. The anogenital distance (AGD) of each pup should be measured on at least one occasion from PND 0 through PND 4. Pup body weight should be collected on the day the AGD is measured and the AGD should be normalised to a measure of pup size, preferably the cube root of body weight (12). The presence of nipples/areolae in male pups should be checked on PND 12 or 13.

46. All selected F_1 animals are evaluated daily for balano-preputial separation or vaginal patency for male/female respectively commencing before the expected day for achievement of these endpoints to detect if sexual maturation occurs early. Any abnormalities of genital organs, such as persistent vaginal thread, hypospadia or cleft penis, should be noted. Sexual maturity of F_1 animals is compared to physical development by determining age and body weight at balano-preputial separation or vaginal opening for male/female respectively (13).

Assessment of potential developmental neurotoxicity (cohorts 2A and 2B)

47. Ten male and 10 female cohort 2A animals and 10 male and 10 female cohort 2B animals, from each treatment group (for each cohort: 1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) should be used for neurotoxicity assessments. Cohort 2A animals should be subjected to auditory startle, functional observational battery, motor activity (see paragraphs 48-50), and neuropathology assessments (see paragraphs 74-75). Efforts should be made to ensure that variations in all test conditions are minimal and are not systematically related to treatment. Among the variables that can affect behaviour are sound level (e.g. intermittent noise), temperature, humidity, lighting, odors, time of day, and environmental distractions. Results of the neurotoxicity assays should be interpreted in relation to appropriate historical control reference ranges. Cohort 2B animals should be used for neuropathology assessment on PND 21 or PND 22 (see paragraphs 74-75).

48. An auditory startle test should be performed on PND 24 (±1 day) using animals in cohort 2A. The day of testing should be counterbalanced across treated and control groups. Each session consists of 50 trials. In performing the auditory startle test, the mean response amplitude on each block of 10 trials (5 blocks of 10 trials) should be determined, with test conditions optimised to produce intra-session habituation. These procedures should be consistent with test method B.53 (35).

49. At an appropriate time between PND 63 and PND 75, the cohort 2A animals are

subjected to a functional observational battery and an automated test of motor activity. These procedures should be consistent with test methods B.43 (33) and B.53 (35). The functional observational battery includes a thorough description of the subject's appearance, behaviour and functional integrity. This is assessed through observations in the home cage, after removal to a standard arena for observation (open field) where the animal is moving freely, and through manipulative tests. Testing should proceed from the least to the most interactive. A list of measures is presented in Appendix 1. All animals should be observed carefully by trained observers who are unaware of the animals' treatment status, using standardised procedures to minimise observer variability. Where possible, it is advisable that the same observer evaluates the animals in a given test. If this is not possible, some demonstration of inter-observer reliability is required. For each parameter in the behavioural testing battery, explicit operationally defined scales and scoring criteria are to be used. If possible, objective quantitative measures should be developed for observational endpoints, which involve subjective ranking. For motor activity, each animal is tested individually. The test session should be long enough to demonstrate intra-session habituation for controls. Motor activity should be monitored by an automated activity recording apparatus which should be capable of detecting both increases and decreases in activity, (*i.e.* baseline activity as measured by the device should not be so low as to preclude detection of decreases, nor so high as to preclude detection of increases in activity). Each device should be tested by standard procedures to ensure, to the extent possible, reliability of operation across devices and across days. To the extent possible, treatment groups should be balanced across devices. Treatment groups should be counter-balanced across test times to avoid confounding by circadian rhythms of activity.

50. If existing information indicates the need for other functional testing (*e.g.* sensory, social, cognitive), these should be integrated without compromising the integrity of the other evaluations conducted in the study. If this testing is performed in the same animals as used for standard auditory startle, functional observational battery and motor activity testing, different tests should be scheduled to minimise the risk of compromising the integrity of these tests. Supplemental procedures may be particularly useful when empirical observation, anticipated effects, or mechanistic/mode-of-action indicate a specific type of neurotoxicity.

Assessment of potential developmental immunotoxicity (cohort 3)

51. At PND 56 (±3 days), 10 male and 10 female cohort 3 animals from each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) should be used in a T-cell dependant antibody response assay, *i.e.* the primary IgM antibody response to a T-cell dependent antigen, such as Sheep Red Blood Cells (SRBC) or Keyhole Limpet Hemocyanin (KLH), consistent with current immunotoxicity testing procedures (14) (15). The response may be evaluated by counting specific plaque-forming cells (PFC) in the spleen or by determining the titer of SRBC- or KLH-specific IgM antibody in the serum by ELISA, at the peak of the response. Responses typically peak four (PFC response) or five (ELISA) days after intravenous immunisation. If the primary antibody response is assayed by counting plaque-forming cells, it is permissible to evaluate subgroups of animals on separate days, provided that: subgroup immunisation and sacrifice are timed so that PFCs are counted at the peak of the response; that subgroups contain an equal number of male and

female offspring from all dose groups, including controls; and that subgroups are evaluated at approximately the same postnatal age. Exposure to the test chemical will continue until the day before collecting spleens for the PFC response or serum for the ELISA assay.

Follow-up assessment of potential reproductive toxicity (cohort 1B)

52. Cohort 1B animals can be maintained on treatment beyond PND 90 and bred to obtain a F_2 generation if necessary. Males and females of the same dose group should be cohabited (avoiding the pairing of siblings) for up to two weeks, beginning on or after PND 90, but not exceeding PND 120. Procedures should be similar to those for the P animals. However, based on a weight of evidence, it may suffice to terminate the litters on PND 4 rather than follow them to weaning or beyond.

TERMINAL OBSERVATIONS

Clinical biochemistry / Haematology

53. Systemic effects should be monitored in P animals. Fasted blood samples from a defined site are taken from ten randomly-selected P males and females per dose group at termination, stored under appropriate conditions and subjected to partial or full-scale haematology, clinical biochemistry, assay of T4 and TSH or other examinations suggested by the known effect profile of the test chemical (see OECD Guidance Document 151 (40)). The following haematological parameters should be examined: haematocrit, haemoglobin concentration, erythrocyte count, total and differential leukocyte count, platelet count and blood clotting time/potential. Investigations of plasma or serum should include: glucose, total cholesterol, urea, creatinine, total protein, albumin and at least two enzymes indicative of hepatocellular effects (such as alanine aminotranferase, aspartate aminotransferase, alkaline phosphatase, gamma glutamyl transpeptidase and sorbitol dehydrogenase). Measurements of additional enzymes and bile acids may provide useful information under certain circumstances. In addition, blood from all animals may be taken and stored for possible analysis at a later time to help clarify equivocal effects or to generate internal exposure data. If a second mating of P animals is not intended, the blood samples are obtained just prior to, or as part of, the procedure at scheduled sacrifice. In the case animals are retained, blood samples should be collected a few days before the animals are mated for the second time. Unless existing data from repeated-dose studies indicate that the parameter is not affected by the test chemical, urinalysis should be performed prior to termination and the following parameters evaluated: appearance, volume, osmolality or specific gravity, pH, protein, glucose, blood and blood cells, cell debris. Urine may also be collected to monitor excretion of test chemical and/or metabolite(s).

54. Systemic effects should also be monitored in F_1 animals. Fasted blood samples from a defined site are taken from ten randomly selected cohort 1A males and females per dose group at termination, stored under appropriate conditions and subjected to standard clinical biochemistry, including the assessment of serum levels for thyroid hormones (T4 and TSH), haematology (total and differential leukocyte plus erythrocyte counts) and urinalysis assessments.

55. The surplus pups at PND 4 are subject to gross necropsy and consideration given to measuring serum thyroid hormone (T4) concentrations. If necessary, neonatal (PND 4) blood can be pooled by litters for biochemical/thyroid hormone analyses. Blood is also collected for T4 and TSH analysis from weanlings subject to gross necropsy on PND 22 (F_1 pups not selected for cohorts).

Sperm parameters

56. Sperm parameters should be measured in all P generation males unless there is existing data to show that sperm parameters are unaffected in a 90-day study. Examination of sperm parameters should be performed in all cohort 1A males.

57. At termination, testis and epididymis weights are recorded for all P and F_1 (cohort 1A) males. At least one testis and one epididymis are reserved for histopathological examination. The remaining epididymis is used for enumeration of cauda epididymis sperm reserves (16) (17). In addition, sperm from the cauda epididymis (or vas deferens) is collected using methods that minimise damage for evaluation of sperm motility and morphology (18).

58. Sperm motility can either be evaluated immediately after sacrifice or recorded for later analysis. The percentage of progressively motile sperm could be determined either subjectively or objectively by computer-assisted motion analysis (19) (20) (21) (22) (23) (24). For the evaluation of sperm morphology, an epididymal (or vas deferens) sperm sample should be examined as fixed or wet preparations (25) and at least 200 spermatozoa per sample classified as either normal (both head and midpiece/tail appear normal) or abnormal. Examples of morphologic sperm abnormalities would include fusion, isolated heads, and misshapen heads and/or tails (26). Misshapen or large sperm heads may indicate defects in spermiation.

59. If sperm samples are frozen, smears fixed and images for sperm motility analysis recorded at the time of necropsy (27), subsequent analysis may be restricted to control and high-dose males. However, if treatment-related effects are observed, the lower dose groups should also be evaluated.

Gross necropsy

60. At the time of termination or premature death, all P and F_1 animals are necropsied and examined macroscopically for any structural abnormalities or pathological changes. Special attention should be paid to the organs of the reproductive system. Pups that are humanely killed in a moribund condition and dead pups should be recorded and, when not macerated, examined for possible defects and/or cause of death and preserved.

61. For adult P and F_1 females, a vaginal smear is examined on the day of necropsy to determine the stage of the oestrous cycle and allow correlation with histopathology in reproductive organs. The uteri of all P females (and F_1 females, if applicable) are examined for the presence and number of implantation sites, in a manner which does not compromise histopathological evaluation.

Organ weight and tissue preservation – P and F_1 adult animals

62. At the time of termination, body weights and wet weights of the organs listed below from all P animals and all F_1 adults, from relevant cohorts (as outlined below), are determined as soon as possible after dissection to avoid drying. These organs should then be preserved under appropriate conditions. Unless specified otherwise, paired organs can be weighed individually or combined, consistent with the typical practice of the performing laboratory.

   - Uterus (with oviducts and cervix), ovaries

   - Testes, epididymides (total and cauda for the samples used for sperm counts)

   - Prostate (dorsolateral and ventral parts combined). Care should be exercised when trimming the prostate complex to avoid puncture of the fluid filled seminal vesicles. In the event of a treatment-related effect on total prostate weight, the dorsolateral and ventral segments should be carefully dissected after fixation, and weighed separately.

   - Seminal vesicles with coagulating glands and their fluids (as one unit)

   - Brain, liver, kidneys, heart, spleen, thymus, pituitary, thyroid (post-fixation), adrenal glands and known target organs or tissues.

63. In addition to the organs listed above, samples of peripheral nerve, muscle, spinal cord, eye plus optic nerve, gastrointestinal tract, urinary bladder, lung, trachea (with thyroid and parathyroid attached), bone marrow, vas deferens (males), mammary gland (males and females) and vagina should be preserved under appropriate conditions.

64. Cohort 1A animals have all organs weighed and preserved for histopathology.

65. For the investigation of pre- and postnatally induced immunotoxic effects, 10 male and 10 female cohort 1A animals from each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) will be subject to the following at termination:

   - weighing of the lymph nodes associated with and distant from the route of exposure (in addition to the weight of the adrenal glands, the thymus and the spleen, already performed in all cohort 1A animals)

   - splenic lymphocyte subpopulation analysis (CD4+ and CD8+ T lymphocytes, B lymphocytes, and natural killer cells) using one half of the spleen, the other half of the spleen being preserved for histopathological evaluation,

   Analysis of splenic lymphocyte subpopulations in non-immunised (cohort 1A) animals will determine if exposure is related to a shift in the immunological steady state distribution of "helper" (CD4+) or cytotoxic (CD8+) thymus-derived lymphocytes or natural killer (NK) cells (rapid responses to neoplastic cells and pathogens).

66. Cohort 1B animals should have the following organs weighed and corresponding tissues processed to the block stage:

   - Vagina (not weighed)

   - Uterus with cervix

- Ovaries

- Testes (at least one)

- Epididymides

- Seminal vesicles and coagulating glands

- Prostate

- Pituitary

- Identified target organs

Histopathology in cohort 1B would be conducted if results from cohort 1A are equivocal or in cases of suspected reproductive or endocrine toxicants.

67. Cohorts 2A and 2B: Developmental neurotoxicity testing (PND 21 or PND 22 and adult offspring). Cohort 2A animals are terminated after behavioural testing, with brain weight recorded and full neurohistopathology for purposes of neurotoxicity assessment. Cohort 2B animals are terminated on PND 21 or PND 22, with brain weight recorded and microscopic examination of the brain for purposes of neurotoxicity assessment. Perfusion fixation is required for cohort 2A animals and optional for cohort 2B animals, as provided in test method B.53 (35).

Organ weight and tissue preservation – F_1 weanlings

68. The pups not selected for cohorts, including runts, are terminated after weaning, on PND 22, unless the results indicate the need for further in-life investigations. Terminated pups are subjected to gross necropsy including an assessment of the reproductive organs, as described in paragraphs 62 and 63. For up to 10 pups per sex per group, from as many litters as possible, brain, spleen, and thymus should be weighed and retained under appropriate conditions. In addition, mammary tissues for these male and female pups may be preserved for further microscopic analysis[2] (see OECD Guidance Document 151 (40)). Gross abnormalities and target tissues should be saved for possible histological examination.

Histopathology – P animals

69. Full histopathology of the organs listed in paragraphs 62 and 63 is performed for all high-dose and control P animals. Organs demonstrating treatment-related changes should also be examined in all animals at the lower dose groups to aid in determining a NOAEL. Additionally, reproductive organs of all animals suspected of reduced fertility, *e.g.* those that failed to mate, conceive, sire, or deliver healthy offspring, or for which oestrous cyclicity or sperm number, motility, or morphology were affected, and all gross lesions should be subjected to histopathological evaluation.

Histopathology – F_1 animals

---

[2] Research has shown the mammary gland, especially in early life mammary gland development, to be a sensitive endpoint for oestrogen action. It is recommended that endpoints involving pup mammary glands of both sexes be included in this test method, when validated.

Cohort 1 animals

70. Full histopathology of the organs listed in paragraphs 62 and 63 is performed for all high-dose and control adult cohort 1A animals. All litters should be represented by at least 1 pup per sex. Organs and tissues demonstrating treatment-related changes and all gross lesions should also be examined in all animals in the lower dose groups to aid in determining a NOAEL. For the evaluation of pre- and postnatally induced effects on lymphoid organs also the histopathology on the collected lymph nodes and bone marrow should be evaluated of 10 male and 10 female cohort 1A animals next to histopathological evaluation of the thymus, spleen, and the adrenal glands already performed in all 1A animals.

71. Reproductive and endocrine tissues from all cohort 1B animals, processed to the block stage as described in paragraph 66, should be examined for histopathology in cases of suspected reproductive or endocrine toxicants. Cohort 1B should also undergo histological examination if results from cohort 1A are equivocal.

72. Ovaries of adult females should contain primordial and growing follicles, as well as corpora lutea; therefore, a histopathological examination should be aimed at detecting a quantitative evaluation of primordial and small growing follicles, as well as corpora lutea, in F_1 females; the number of animals, ovarian section selection, and section sample size should be statistically appropriate for the evaluation procedure used. Follicular enumeration may first be conducted on control and high-dose animals, and in the event of an adverse effect in the latter, lower doses should be examined. Examination should include enumeration of the number of primordial follicles, which can be combined with small growing follicles, for comparison of treated and control ovaries (see OECD Guidance Document 151 (40)). Corpora lutea assessment should be conducted in parallel with oestrous cyclicity testing so that the stage of the cycle can be taken into account in the assessment. Oviduct, uterus and vagina are examined for appropriate organ-typic development.

73. Detailed testicular histopathology examinations are conducted on the F_1 males in order to identify treatment-related effects on testis differentiation and development and on spermatogenesis (38). When possible, sections of the rete testis should be examined. Caput, corpus, and cauda of the epididymis and the vas deferens are examined for appropriate organ-typic development, as well as for the parameters required for the P males.

Cohort 2 animals

74. Neurohistopathology is performed for all high-dose and control cohort 2A animals per sex following completion of neurobehavioral testing (after PND 75, but not to exceed PND 90). Brain histopathology is performed for all high-dose and control cohort 2B animals per sex on PND 21 or PND 22. Organs or tissues demonstrating treatment-related changes should also be examined for the animals in the lower dose groups to aid in determining a NOAEL. For cohort 2A and 2B animals, multiple sections are examined from the brain to allow examination of olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, mid-brain (thecum, tegmentum, and cerebral peduncles), brain-stem and cerebellum. For cohort 2A only, the eyes (retina and optic nerve) and samples of peripheral nerve, muscle and spinal cord are examined.

All neurohistological procedures should be consistent with test method B.53 (35).

75. Morphometric (quantitative) evaluations should be performed on representative areas of the brain (homologous sections carefully selected based on reliable microscopic landmarks) and may include linear and/or areal measurements of specific brain regions. At least three consecutive sections should be taken at each landmark (level) in order to select the most homologous and representative section for the specific brain area to be evaluated. The neuropathologist should exercise appropriate judgment as to whether sections prepared for measurement are homologous with others in the sample set and therefore suitable for inclusion, since linear measurements in particular may change over a relatively short distance (28). Non-homologous sections should not be used. While the objective is to sample all animals reserved for this purpose (10/sex/dose level), smaller numbers may still be adequate. However, samples from fewer than 6 animals/sex/dose level would generally not be considered sufficient for the purposes of this test method. Stereology may be used to identify treatment-related effects on parameters such as volume or cell number for specific neuroanatomic regions. All aspects of the preparation of tissue samples, from tissue fixation, through the dissection of tissue samples, tissue processing, and staining of slides, should employ a counterbalanced design, such that each batch contains representative samples from each dose group. When morphometric or stereological analyses are to be used, then brain tissue should be embedded in appropriate media at all dose levels at the same time in order to avoid shrinkage artifacts associated with prolonged storage in fixative.

REPORTING

Data

76. Data are reported individually and summarised in tabular form. Where appropriate, for each test group and each generation, the following should be reported: number of animals at the start of the test, number of animals found dead during the test or killed for humane reasons, time of any death or humane kill, number of fertile animals, number of pregnant females, number of females giving birth to a litter, and number of animals showing signs of toxicity. A description of the toxicity, including time of onset, duration, and severity should also be reported.

77. Numerical results should be evaluated by an appropriate, and accepted statistical method. The statistical methods should be selected as part of the study design and should appropriately address non-normal data (*e.g.* count data), censored data (*e.g.* limited observation time), non-independence (*e.g.* litter effects and repeated measures), and unequal variances. Generalised linear mixed models and dose-response models cover a broad class of analytical tools that may be appropriate for the data generated under this test method. The report should include sufficient information on the method of analysis and the computer program employed, so that an independent reviewer/statistician can evaluate/re-evaluate the analysis.

Evaluation of results

78. The findings should be evaluated in terms of the observed effects, including necropsy and microscopic findings. The evaluation includes the relationship, or lack thereof,

between the dose and the presence, incidence, and severity of abnormalities, including gross lesions. Target organs, fertility, clinical abnormalities, reproductive and litter performance, body weight changes, mortality and any other toxic and developmental effects should also be assessed. Special attention should be given to sex-specific changes. The physico-chemical properties of the test chemical, and when available, TK data, including placental transfer and milk excretion, should be taken into consideration when evaluating the test results.

Test report

79. The test report should include the following information obtained in the present study from P, F_1 animals and F_2 animals (where relevant):

Test chemical:

- All relevant available information on the chemical, toxicokinetic and toxicodynamic properties of the test chemical;

- Identification data;

- Purity;

Vehicle (if appropriate):

- Justification for choice of vehicle if other than water;

Test animals:

- Species/strain used;

- Number, age and sex of animals;

- Source, housing conditions, diet, nesting materials, etc.;

- Individual weights of animals at the start of the test;

- Vaginal smear data for P females before initiation of treatment (if data are collected at that time);

- P generation pairing records indicating male and female partner of a mating and mating success;

- Litter of origin records for adult F_1 generation animals;

Test conditions:

- Rationale for dose level selection;

- Details of test chemical formulation/diet preparation, achieved concentrations;

- Stability and homogeneity of the preparation in the vehicle or carrier (*e.g.* diet, drinking water), in the blood and/or milk under the conditions of use and storage between uses;

- Details of the administration of the test chemical;

- Conversion from diet/drinking water test chemical concentration (ppm) to the achieved dose (mg/kg body weight/day), if applicable;

- Details of food and water quality (including diet composition, if available);
- Detailed description of the randomisation procedures to select pups for culling and to assign pups to test groups;
- Environmental conditions;
- List of study personnel, including professional training;

Results (summary and individual data by sex and dose):

- Food consumption, water consumption if available, food efficiency (body weight gain per gram of food consumed, except for the period of cohabitation and during lactation), and test chemical consumption (for dietary/drinking water administration) for P and F_1 animals;
- Absorption data (if available);
- Body weight data for P animals;
- Body weight data for the selected F_1 animals postweaning;
- Time of death during the study or whether animals survived to termination;
- Nature, severity and duration of clinical observations (whether reversible or not);
- Haematology, urinalysis and clinical chemistry data including TSH and T4;
- Phenotypic analysis of spleen cells (T-, B-, NK-cells);
- Bone marrow cellularity;
- Toxic response data;
- Number of P and F_1 females with normal or abnormal oestrous cycle and cycle duration;
- Time to mating (precoital interval, the number of days between pairing and mating);
- Toxic or other effects on reproduction, including numbers and percentages of animals that accomplished mating, pregnancy, parturition and lactation, of males inducing pregnancy, of females with signs of dystocia/prolonged or difficult parturition;
- Duration of pregnancy and, if available, parturition;
- Numbers of implantations, litter size and percentage of male pups;
- Number and percent of post-implantation loss, live births and stillbirths;
- Litter weight and pup weight data (males, females and combined), the number of runts if determined;
- Number of pups with grossly visible abnormalities;
- Toxic or other effects on offspring, postnatal growth, viability, etc.;
- Data on physical landmarks in pups and other postnatal developmental data;
- Data on sexual maturation of F_1 animals;
- Data on functional observations in pups and adults, as applicable;

105

- Body weight at sacrifice and absolute and relative organ weight data for the P and adult F_1 animals;

- Necropsy findings;

- Detailed description of all histopathological findings;

- Total cauda epididymal sperm number, percent progressively motile sperm, percent morphologically normal sperm, and percent of sperm with each identified abnormality for P and F_1 males;

- Numbers and maturational stages of follicles contained in the ovaries of P and F_1 females, where applicable;

- Enumeration of corpora lutea in the ovaries of F_1 females;

- Statistical treatment of results, where appropriate;

Cohort 2 parameters:

- Detailed description of the procedures used to standardise observations and procedures as well as operational definitions for scoring observations;

- List of all test procedures used, and justification for their use;

- Details of the behavioural/functional, neuropathological and morphometric procedures used, including information and details on automated devices;

- Procedures for calibrating and ensuring the equivalence of devices and the balancing of treatment groups in testing procedures;

- Short justification explaining any decisions involving professional judgment;

- Detailed description of all behavioural/functional, neuropathological and morphometric findings by sex and dose group, including both increases and decreases from controls;

- Brain weight;

- Any diagnoses derived from neurological signs and lesions, including naturally-occurring diseases or conditions;

- Images of exemplar findings;

- Low-power images to assess homology of sections used for morphometry;

- Statistical treatment of results, including statistical models used to analyse the data, and the results, regardless of whether they were significant or not;

- Relationship of any other toxic effects to a conclusion about the neurotoxic potential of the test chemical, by sex and dose group;

- Impact of any toxicokinetic information on the conclusions;

- Data supporting the reliability and sensitivity of the test method (*i.e.* positive and historical control data);

- Relationships, if any, between neuropathological and functional effects;

- NOAEL or benchmark dose for dams and offspring, by sex and dose group;

- Discussion of the overall interpretation of the data based on the results, including

a conclusion of whether or not the chemical caused developmental neurotoxicity and the NOAEL;

Cohort 3 parameters:

- Serum IgM antibody titres (sensitisation to SRBC or KLH), or splenic IgM PFC units (sensitisation to SRBC);

- Performance of the TDAR method should be confirmed as part of the optimisation process by laboratory setting up the assay for the first time, and periodically (e.g. yearly) by all laboratories;

- Discussion of the overall interpretation of the data based on the results, including a conclusion of whether or not the chemical caused developmental immunotoxicity and the NOAEL;

Discussion of results

Conclusions, including NOAEL values for parental and offspring effects

All information not obtained during the study, but useful for the interpretation of the results (*e.g.* similarities of effects to any known neurotoxicants), should also be provided.

Interpretation of Results

80. An Extended One-Generation Reproductive Toxicity Study will provide information on the effects of repeated exposure to a chemical during all phases of the reproductive cycle, as necessary. In particular, the study provides information on the reproductive system, and on development, growth, survival, and functional endpoints of offspring up to PND 90.

81. Interpretation of the results of the study should take into account all available information on the chemical, including physico-chemical, TK and toxicodynamic properties, available relevant information on structural analogues, and results of previously-conducted toxicity studies with the test chemical (*e.g.* acute toxicity, toxicity after repeated application, mechanistic studies and studies assessing if there are substantial qualitative and quantitative species differences in *in vivo*/*in vitro* metabolic properties). Gross necropsy and organ weight results should be assessed in context with observations made in other repeat-dose studies, when feasible. Decreases in offspring growth might be considered in relationship to an influence of the test chemical on milk composition (29).

Cohort 2 (Developmental neurotoxicity)

82. Neurobehavioral and neuropathology results should be interpreted in the context of all findings, using a weight-of-evidence approach with expert judgment. Patterns of behavioural or morphological findings, if present, as well as evidence of dose-response should be discussed. The evaluation of developmental neurotoxicity, including human epidemiological studies or case reports, and experimental animal studies (e.g. toxicokinetic data, structure-activity information, data from other toxicity studies) should be included in this characterisation. Evaluation of data should include a

discussion of both the biological and statistical significance. The evaluation should include the relationship, if any, between observed neuropathological and behavioural alterations. For guidance on the interpretation of developmental neurotoxicity results, refer to test method B.53 (35) and Tyl et al., 2008 (31).

Cohort 3 (Developmental immunotoxicity)

83. Suppression or enhancement of immune function as assessed by TDAR (T-cell dependent antibody response), should be evaluated in the context of all observations made. Significance of the outcome of TDAR may be supported by other effects on immunologically-related indicators (*e.g.* bone marrow cellularity, weight and histopathology of lymphoid tissues, lymphocyte subset distribution). Effects established by TDAR may be less meaningful in case of other toxicities observed at lower exposure concentrations.

84. OECD Guidance Document 43 should be consulted for aid in the interpretation of reproduction and neurotoxicity results (26).

LITERATURE

(1) Cooper, R.L., J.C. Lamb, S.M. Barlow, K. Bentley, A.M. Brady, N. Doerr, D.L. Eisenbrandt, P.A. Fenner-Crisp, R.N. Hines, L.F.H. Irvine, C.A. Kimmel, H. Koeter, A.A. Li, S.L. Makris, L.P. Sheets, G.J.A. Speijers and K.E. Whitby (2006), "A Tiered Approach to Life Stages Testing for Agricultural Chemical Safety Assessment", *Critical Reviews in Toxicology*, 36, 69-98.

(2) Thigpen, J.E., K.D.R. Setchell, K.B. Ahlmark, J. Locklear, T. Spahr, G.F. Leviness, M.F. Goelz, J.K. Haseman, R.R. Newbold, and D.B. Forsythe (1999), "Phytoestrogen Content of Purified Open and Closed Formula Laboratory Animal Diets", *Lab. Anim. Sci.*, 49, 530- 536.

(3) Zoetis, T. and I. Walls (2003), Principles and Practices for Direct Dosing of Pre-Weaning Mammals in Toxicity Testing and Research, ILSI Press, Washington, DC.

(4) Moser, V.C., I. Walls and T. Zoetis (2005), "Direct Dosing of Preweaning Rodents in Toxicity Testing and Research: Deliberations of an ILSI RSI Expert Working Group", International Journal of Toxicology, 24, 87-94.

(5) Conolly, R.B., B.D. Beck, and J.I. Goodman (1999), "Stimulating Research to Improve the Scientific Basis of Risk Assessment", *Toxicological Sciences*, 49, 1-4.

(6) Ulbrich, B. and A.K. Palmer (1995), "Detection of Effects on Male Reproduction – a Literature Survey", *Journal of the American College of Toxicologists*, 14, 293-327.

(7) Mangelsdorf, I., J. Buschmann and B. Orthen (2003), "Some Aspects Relating to the Evaluation of the Effects of Chemicals on Male Fertility", *Regulatory Toxicology and Pharmacology*, 37, 356-369.

(8) Sakai, T., M. Takahashi, K. Mitsumori, K. Yasuhara, K. Kawashima, H. Mayahara and Y. Ohno (2000). "Collaborative work to evaluate toxicity on male reproductive organs by repeated dose studies in rats - overview of the studies", *Journal of Toxicological Sciences*, 25, 1-21.

(9) Creasy, D.M. (2003), "Evaluation of Testicular Toxicology: A Synopsis and Discussion of the Recommendations Proposed by the Society of Toxicologic Pathology", *Birth Defects Research*, Part B, 68, 408-415.

(10) Goldman, J.M., A.S. Murr, A.R. Buckalew, J.M. Ferrell and R.L. Cooper (2007), "The Rodent Estrous Cycle: Characterization of Vaginal Cytology and its Utility in Toxicological Studies", *Birth Defects Research*, Part B, 80 (2), 84-97.

(11) Sadleir, R.M.F.S. (1979), "Cycles and Seasons", in C.R. Auston and R.V. Short (eds.), *Reproduction in Mammals: I. Germ Cells and Fertilization*, Cambridge, New York.

(12) Gallavan, R.H. Jr, J.F. Holson, D.G. Stump, J.F. Knapp and V.L. Reynolds (1999), "Interpreting the Toxicologic Significance of Alterations in Anogenital Distance: Potential for Confounding Effects of Progeny Body Weights", *Reproductive Toxicology,* 13: 383-390.

(13) Korenbrot, C.C., I.T. Huhtaniemi and R.I. Weiner (1977), "Preputial Separation as an External Sign of Pubertal Development in the Male Rat", *Biological Reproduction*, 17, 298-303.

(14) Ladics, G.S. (2007), "Use of SRBC Antibody Responses for Immunotoxicity Testing", *Methods*, 41, 9-19.

(15) Gore, E.R., J. Gower, E. Kurali, J.L. Sui, J. Bynum, D. Ennulat and D.J. Herzyk (2004), "Primary Antibody Response to Keyhole Limpet Hemocyanin in Rat as a Model for Immunotoxicity Evaluation", *Toxicology*, 197, 23-35.

(16) Gray, L.E., J. Ostby, J. Ferrell, G. Rehnberg, R. Linder, R. Cooper, J. Goldman, V. Slott and J. Laskey (1989), "A Dose-Response Analysis of Methoxychlor-Induced Alterations of Reproductive Development and Function in the Rat", *Fundamental and Applied Toxicology*, 12, 92-108.

(17) Robb, G.W., R.P. Amann and G.J. Killian (1978), "Daily Sperm Production and Epididymal Sperm Reserves of Pubertal and Adult Rats", *Journal of Reproduction and Fertility*, 54, 103-107.

(18) Klinefelter, G.R., L.E. Jr Gray and J.D. Suarez (1991), "The Method of Sperm Collection Significantly Influences Sperm Motion Parameters Following Ethane Dimethanesulfonate Administration in the Rat". *Reproductive Toxicology*, 5, 39-44.

(19) Seed, J., R.E. Chapin, E.D. Clegg., L.A. Dostal, R.H. Foote, M.E. Hurtt, G.R. Klinefelter, S.L. Makris, S.D. Perreault, S. Schrader, D. Seyler, R. Sprando, K.A. Treinen, D.N. Veeramachaneni and L.D. Wise (1996), "Methods for Assessing Sperm Motility, Morphology, and Counts in the Rat, Rabbit, and Dog: a Consensus Report", *Reproductive Toxicology*, 10, 237- 244.

(20) Chapin, R.E., R.S. Filler, D. Gulati, J.J. Heindel, D.F. Katz, C.A. Mebus, F. Obasaju, S.D. Perreault, S.R. Russell and S. Schrader (1992), "Methods for Assessing Rat Sperm Motility", *Reproductive Toxicology*, 6, 267-273.

(21) Klinefelter, G.R., N.L. Roberts and J.D. Suarez (1992), "Direct Effects of Ethane Dimethanesulphonate on Epididymal Function in Adult Rats: an *In Vitro* Demonstration", *Journal of Andrology*, 13, 409-421.

(22) Slott, V.L., J.D. Suarez and S.D. Perreault (1991), "Rat Sperm Motility Analysis: Methodologic Considerations", *Reproductive Toxicology*, 5, 449-458.

(23) Slott, V.L., and S.D. Perreault (1993), "Computer-Assisted Sperm Analysis of Rodent Epididymal Sperm Motility Using the Hamilton-Thorn Motility Analyzer", *Methods in Toxicology, Part A*, Academic, Orlando, Florida. pp. 319-333.

(24) Toth, G.P., J.A. Stober, E.J. Read, H. Zenick and M.K. Smith (1989), "The Automated Analysis of Rat Sperm Motility Following Subchronic Epichlorhydrin Administration: Methodologic and Statistical Considerations", *Journal of Andrology*, 10, 401-415.

(25) Linder, R.E., L.F. Strader, V.L. Slott and J.D. Suarez (1992), "Endpoints of Spermatoxicity in the Rat After Short Duration Exposures to Fourteen Reproductive Toxicants", *Reproductive Toxicology*, 6, 491-505.

(26) OECD (2008), *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment*, Series on Testing and Assessment, No. 43, ENV/JM/MONO(2008)16, OECD, Paris.

(27) Working, P.K., M. Hurtt (1987), "Computerized Videomicrographic Analysis of Rat Sperm Motility", *Journal of Andrology*, 8, 330-337.

(28) Bolin, B., R. Garman, K. Jensen, G. Krinke, B. Stuart, and an ad Hoc Working Group of the STP Scientific and Regulatory Policy Committee (2006), "A 'Best Practices' Approach to Neuropathologic Assessment in Developmental Neurotoxicity Testing − for Today", *Toxicological Pathology*, 34, 296-313.

(29) Stütz, N., B. Bongiovanni, M. Rassetto, A. Ferri, A.M. Evangelista de Duffard, and R. Duffard (2006), "Detection of 2,4-dichlorophenoxyacetic Acid in Rat Milk of Dams Exposed During Lactation and Milk Analysis of their Major Components", *Food Chemicals Toxicology*, 44, 8-16.

(30) Thigpen, JE, K.D.R. Setchell, J.K. Haseman, H.E. Saunders, G.F. Caviness, G.E. Kissling, M.G. Grant and D.B. Forsythe (2007), "Variations in Phytoestrogen Content between Different Mill Dates of the Same Diet Produces Significant Differences in the Time of Vaginal Opening in CD-1 Mice and F344 Rats but not in CD Sprague Dawley Rats", *Environmental health perspectives,* 115(12), 1717-1726.

(31) Tyl, R.W., K. Crofton, A. Moretto, V. Moser, L.P. Sheets and T.J. Sobotka (2008), "Identification and Interpretation of Developmental Neurotoxicity Effects: a Report from the ILSI Research Foundation/Risk Science Institute Expert Working Group on

Neurodevelopmental Endpoints", *Neurotoxicology and Teratology*, 30: 349-381.

(32) OECD (1996), *Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test*, OECD Guideline for Testing of Chemicals, No. 422, OECD, Paris.

(33) Chapter B.43 of this Annex, Neurotoxicity Study in Rodents

(34) OECD (2000), *Guidance Document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations*, Series on Testing and Assessment, No. 19, ENV/JM/MONO(2000)7, OECD, Paris.

(35) Chapter B.53 of this Annex, Developmental Neurotoxicity Study

(36) Chapter B.54 of this Annex, Uterotrophic Bioassay in Rodents: A short-term Screening Test for Oestrogenic Properties

(37) Chapter B.55 of this Annex, Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti)Androgenic Properties

(38) OECD (2009), Guidance Document for Histologic Evalution of Endocrine and Reproductive Test in Rodents, Series on Testing and Assessment, No. 106, OECD, Paris.

(39) OECD (2011), Guidance Document on the Current Implementation of Internal Triggers in the Extended One Generation Reproductive Toxicity Study in the United States and Canada, Series on Testing and Assessment, No. 117, ENV/JM/MONO(2011)21, OECD, Paris.

(40) OECD (2013), Guidance Document supporting TG 443: Extended One Generation Reproductive Toxicity Study, Series on Testing and Assessment, No. 151, OECD, Paris.

Appendix 1

MEASURES AND OBSERVATIONS INCLUDED IN THE FUNCTIONAL OBSERVATIONAL BATTERY (COHORT 2A)

Home Cage & Open Field	**Manipulative**	**Physiologic**
Posture	Ease of removal	Temperature
Involuntary Clonic & Tonic	Ease of handling	Body weight
Palpebral Closure	Muscle Tone	Pupil response
Piloerection	Approach Response	Pupil size
Salivation	Touch Response	
Lacrimation	Auditory Response	
Vocalisations	Tail Pinch Response	
Rearing	Righting Response	
Gait Abnormalities	Landing Foot Splay	
Arousal	Forelimb Grip Strength	
Stereotypy	Hindlimb Grip Strength	
Bizarre Behaviour		
Stains		
Respiratory Abnormalities		

Appendix 2

DEFINITIONS

Chemical: A substance or a mixture.

Test Chemical: Any substance or mixture tested using this test method.

B.57. H295R Steroidogenesis Assay

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 456 (2011). The OECD initiated a high-priority activity in 1998 to revise existing, and to develop new, test guidelines for the screening and testing of potential endocrine disrupting chemicals. The 2002 OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals comprises five levels, each level corresponding to a different level of biological complexity (1). The *in vitro* H295R *S*teroidogenesis Assay (H295R) described in this test method utilises a human adreno-carcinoma cell line (NCI-H295R cells) and constitutes a level 2 "*in vitro* assay, providing mechanistic data", to be used for screening and prioritisation purposes. Development and standardisation of the assay as a screen for chemical effects on steroidogenesis, specifically the production of 17β-oestradiol (E2) and testosterone (T), was carried out in a multi–step process. The H295R assay has been optimised and validated (2) (3) (4) (5).

2. The objective of the H295R Steroidogenesis Assay is to detect chemicals that affect production of E2 and T. The H295R assay is intended to identify xenobiotics that have as their target site(s) the endogenous components that comprise the intracellular biochemical pathway beginning with the sequence of reactions from cholesterol to the production of E2 and/or T. The H295R assay is not intended to identify chemicals that affect steroidogenesis due to effects on the hypothalamic-pituitary-gonadal (HPG) axis. The goal of the assay is to provide a YES/NO answer with regard to the potential of a chemical to induce or inhibit the production of T and E2; however, quantitative results may be obtained in some cases (see paragraphs 53 and 54). The results of the assay are expressed as relative changes in hormone production compared with the solvent controls (SCs). The assay does not aim to provide specific mechanistic information concerning the interaction of the test chemical with the endocrine system. Research has been conducted using the cell line to identify effects on specific enzymes and intermediate hormones such as progesterone (2).

3. Definitions and abbreviations used in this test method are described in the Appendix. A detailed protocol including instructions on how to prepare solutions, cultivate cells and perform various aspects of the test is available as Appendix I-III of the OECD document "Multi-Laboratory Validation of the H295R Steroidogenesis Assay to Identify Modulators of Testosterone and Estradiol Production" (4).

INITIAL CONSIDERATIONS AND LIMITATIONS

4. Five different enzymes catalysing six different reactions are involved in sex steroid hormone biosynthesis. Enzymatic conversion of cholesterol to pregnenolone by the cytochrome P450 (CYP) cholesterol side-chain cleavage enzyme (CYP11A) constitutes the initial step in a series of biochemical reactions that culminate in synthesis of steroid end-products. Depending upon the order of the next two reactions, the steroidogenic pathway splits into two paths, the Δ^5-hydroxysteroid pathway and Δ^4-ketosteroid pathway, which converge in the production of androstenedione (Figure 1).

115

5. Androstenedione is converted to testosterone (T) by 17β-hydroxysteroid dehydrogenase (17β-HSD). Testosterone is both an intermediate and end-hormone product. In the male, T can be converted to dihydrotestosterone (DHT) by 5α-reductase, which is found in the cellular membranes, nuclear envelope, and endoplasmic reticulum of target tissues of androgenic action such as prostate and seminal vesicles. DHT is significantly more potent as an androgen than T and is also considered an end-product hormone. The H295R assay does not measure DHT (see paragraph 10).

6. The enzyme in the steroidogenic pathway which converts androgenic chemicals into oestrogenic chemicals is aromatase (CYP19). CYP19 converts T into 17β-oestradiol (E2) and androstenedione into oestrone. E2 and T are considered end-product hormones of the steroidogenic pathway.

7. The specificity of the lyase activity of CYP17 differs for the intermediate substrates among species. In the human, the enzyme favours substrates of the Δ^5-hydroxysteroid pathway (pregnenolone), whereas substrates in the Δ^4-ketosteroid pathway (progesterone) are favoured in the rat (19). Such differences in the CYP17 lyase activity may explain some species-dependent differences in response to chemicals that alter steroidogenesis *in vivo* (6). The H295 cells have been shown to most closely reflect human adult adrenal enzyme expression and steroid production pattern (20), but are known to express enzymes for both the Δ^5-hydroxysteroid and Δ^4-ketosteroid pathways for androgen synthesis (7) (11) (13) (15).
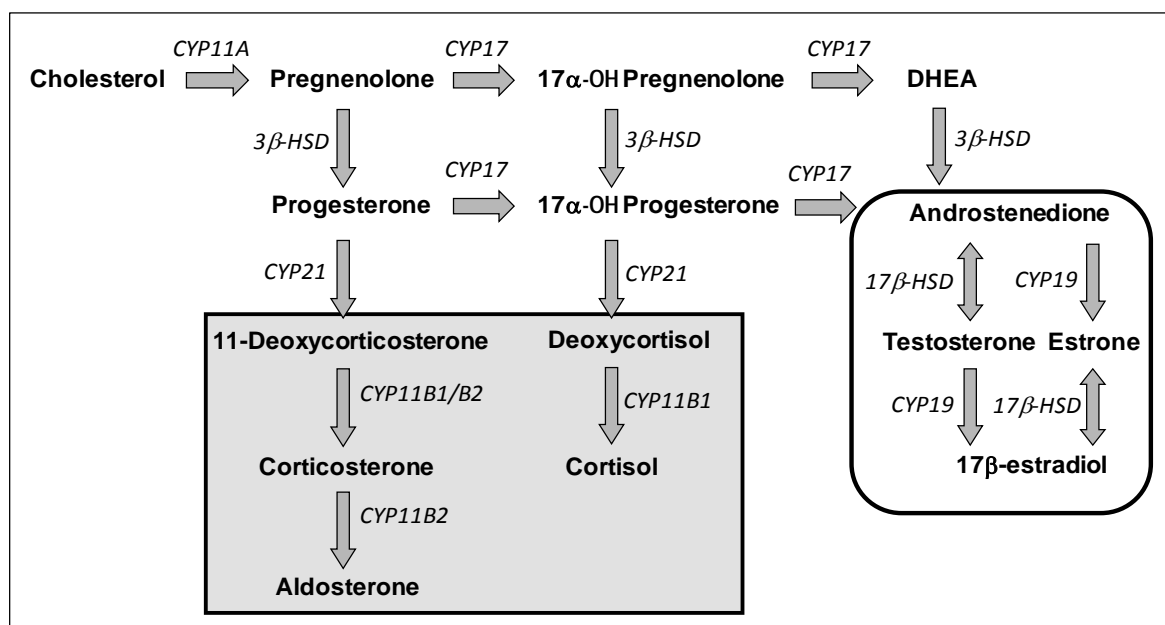

Figure 1: Steroidogenic pathway in H295R cells. Enzymes are in italics, hormones are bolded and arrows indicate the direction of synthesis. Gray background indicates corticosteroid pathways/products. Sex steroid pathways/products are circled. CYP = cytochrome P450; HSD = hydroxysteroid dehydrogenase; DHEA = dehydroepiandrosterone.

8. The human H295R adreno-carcinoma cell line is a useful *in vitro* model for the investigation of effects on steroid hormone synthesis (2) (7) (8) (9) (10). The H295R cell line expresses genes that encode for all the key enzymes for steroidogenesis noted

above (11) (15) (Figure 1). This is a unique property because *in vivo* expression of these genes is tissue and developmental stage-specific with typically no one tissue or one developmental stage expressing all of the genes involved in steroidogenesis (2). H295R cells have physiological characteristics of zonally undifferentiated human foetal adrenal cells (11). The cells represent a unique *in vitro* system in that they have the ability to produce all of the steroid hormones found in the adult adrenal cortex and the gonads, allowing testing for effects on both corticosteroid synthesis and the production of sex steroid hormones such as androgens and oestrogens, although the assay was validated only to detect T and E2. Changes recorded by the test system in the form of alteration in the production of T and E2 can be the result of a multitude of different interactions of test chemicals with steroidogenic functions that are expressed by the H295R cells. These include modulation of the expression, synthesis or function of enzymes involved in the production, transformation, or elimination of steroid hormones (12) (13) (14). Inhibition of hormone production can be due to direct competitive binding to an enzyme in the pathway, impact on co-factors such as NADPH (Nicotinamide Adenine Dinucleotide Phosphate) and cAMP (cyclic Adenosine Monophosphate), and/or increase in steroid metabolism or suppression of gene expression of certain enzymes in the steroidogenesis pathway. While inhibition can be a function of both direct or indirect processes involved with hormone production, induction is typically of an indirect nature, such as by affecting co-factors such as NADPH and cAMP (as in the case of forskolin), decreasing steroid metabolism (13), and or up-regulating steroidogenic gene expression.

9. The H295R assay has several advantages:

    -   It allows for the detection of both increases and decreases in the production of both T and E2;

    -   It permits the direct assessment of the potential impact of a chemical on cell viability/cytotoxicity. This is an important feature as it allows for the discrimination between effects that are due to cytotoxicity from those due to the direct interaction of chemicals with steroidogenic pathways, which is not possible in tissue explants systems that consist of multiple cell types of varying sensitivities and functionalities;

    -   It does not require the use of animals;

    -   The H295R cell line is commercially available.

10. The principle limitations of the assay are as follows:

    -   Its metabolic capability is unknown but probably quite limited; therefore, chemicals that need to be metabolically activated will probably be missed in this assay.

    -   Being derived from adrenal tissue, the H295R possesses the enzymes capable of producing the gluco-, and mineral-corticoids as well as the sex hormones; therefore, effects on the production of gluco-, and mineral corticoids could influence the levels of T and E2 observed in the assay.

    -   It does not measure DHT and, therefore, would not be expected to detect chemicals that inhibit 5α-reductase in which case the Hershberger assay (16) can be used.

- The H295R assay will not detect chemicals that interfere with steroidogenesis by affecting the hypothalamic-pituitary-gonadal axis (HPG) axis as this can only be studied in intact animals.

PRINCIPLE OF THE TEST

11. The purpose of the assay is the detection of chemicals that affect T and E2 production. T is also an intermediate in the pathway to produce E2. The assay can detect chemicals that typically inhibit or induce the enzymes of the steroidogenesis pathway.

12. The assay is usually performed under standard cell culture conditions in 24-well culture plates. Alternatively, other plate sizes can be used for conducting the assay; however, seeding and experimental conditions should be adjusted accordingly to maintain adherence to the performance criteria.

13. After an acclimation period of 24 h in multi-well plates, cells are exposed for 48 h to seven concentrations of the test chemical in at least triplicate. Solvent and a known inhibitor and inducer of hormone production are run at a fixed concentration as negative and positive controls. At the end of the exposure period, the medium is removed from each well. Cell viability in each well is analysed immediately after removal of medium. Concentrations of hormones in the medium can be measured using a variety of methods including commercially available hormone measurement kits and/or instrumental techniques such as liquid chromatography-mass spectrometry (LC-MS). Data are expressed as fold change relative to the solvent control and the Lowest-Observed-Effect-Concentration (LOEC). If the assay is negative, the highest concentration tested is reported as the No-Observed-Effect-Concentration (NOEC). Conclusions regarding the ability of a chemical to affect steroidogenesis should be based on at least two independent test runs. The first test run may function as a range finding run with subsequent adjustment of concentrations for runs 2 and 3, if applicable, if solubility or cytotoxicity problems are encountered or the activity of the chemical seems to be at the end of the range of concentrations tested.

CULTURE PROCEDURE

Cell Line

14. The NCI-H295R cells are commercially available from the American Type Culture Collections (ATCC) upon signing a Material Transfer Agreement (MTA)[3].

Introduction

15. Due to changes in the E2 producing capacity of the cells with increasing age/passages (2), cells should be cultured following a specific protocol before they are used and the number of passages since the cells were defrosted as well as the passage number at which the cells were frozen and placed in liquid nitrogen storage should be noted. The

---

[3] ATCC CRL-2128; ATCC, Manassas, VA, USA, [http://www.lgcstandards-atcc.org/].

first number indicates the actual cell passage number and the second number describes the passage number at which the cells were frozen and placed in storage. For example, cells that were frozen after passage five and defrosted and then were split three times (4 passages counting the freshly thawed cells as passage 1) after they were cultured again would be labelled passage 4.5. An example of a numbering scheme is illustrated in Appendix I of the validation report (4).

16. Stock medium is used as the base for the supplemented and freezing mediums. Supplemented medium is a necessary component for culturing cells. Freezing medium is specifically designed to allow for impact-free freezing of cells for long-term storage. Prior to use, Nu-serum (or a comparable serum of equal properties that has been demonstrated to produce data that meets the test performance and Quality Control (QC) requirements), which is a constituent of supplemented media, should be analysed for background T and E2 concentrations. The preparation of these solutions is described in Appendix II of the validation report (4).

17. After initiation of an H295R cell culture from an original ATCC batch, cells should be grown for five passages (*i.e.* the cells are split 4 times). Passage five cells are then frozen in liquid nitrogen for storage. Prior to freezing the cells, a sample of the previous passage four cells is run in a QC plate (See paragraph 36 and 37) to verify whether the basal production of hormones and the response to positive control chemicals meet the assay quality control criteria as defined in Table 5.

18. H295R cells need to be cultured, frozen and stored in liquid nitrogen to make sure that there are always cells of the appropriate passage/age available for culture and use. The maximum number of passages after taking a new[4] or frozen[5] batch of cells into culture that is acceptable for use in the H295R assay should not exceed 10. For example, acceptable passages for cultures of cells from a batch frozen at passage 5 would be 4.5 through 10.5. For cells started from these frozen batches, the procedure described in paragraph 19 should be followed. These cells should be cultured for at least four (4) additional passages (passage 4.5) prior to their use in testing.

Starting Cells from the Frozen Stock

19. The procedure for starting the cells from frozen stock is to be used when a new batch of cells is removed from liquid nitrogen storage for the purpose of culture and testing. Details for this procedure are set forth in Appendix III of the validation report (4). Cells are removed from liquid nitrogen storage, thawed rapidly, placed in supplemented medium in a centrifuge tube, centrifuged at room temperature, re-suspended in supplemented medium, and transferred to a culture flask. The medium should be changed the following day. The H295R cells are cultivated in an incubator at 37°C with 5% CO_2 in air atmosphere and the medium is renewed 2-3 times per week. When the cells are approximately 85-90% confluent, they should be split. Splitting of the cells is necessary to ensure the health and growth of the cells and to maintain cells for performing bioassays. The cells are rinsed three times with phosphate-buffered saline (PBS, without Ca^{2+} Mg^{2+}) and freed from the culture flask by the addition of an

---

[4] "New batch" refers to a fresh batch of cells received from ATCC.

[5] "Frozen batch" refers to cells that have been previously cultured and then frozen at a laboratory other than ATCC.

appropriate detachment enzyme, *e.g.* trypsin, in PBS (without Ca^{2+} Mg^{2+}). Immediately after the cells detach from the culture flask, the enzyme action should be stopped with the addition of supplemented medium at a ratio of 3X the volume used for the enzyme treatment. Cells are placed into a centrifuge tube, centrifuged at room temperature, the supernatant is removed and the pellet of cells is re-suspended in supplemented medium. The appropriate amount of cell solution is placed in the new culture flask. The amount of cell solution should be adjusted so that the cells are confluent within 5-7 days. The recommended sub-cultivation ratio is 1:3 to 1:4. The plate should be carefully labelled. The cells are now ready to be used in the assay and excess cells should be frozen in liquid nitrogen as described in paragraph 20.

Freezing H295R Cells (preparing cells for liquid nitrogen storage)

20. To prepare H295R cells for freezing, the procedure described above for splitting cells should be followed until the step for re-suspending the pellet of cells in the bottom of the centrifuge tube. Here, the pellet of cells is re-suspended in freezing medium. The solution is transferred to a cryogenic vial, labelled appropriately, and frozen at -80°C for 24 hours after which the cryogenic vial is transferred to liquid nitrogen for storage. Details for this procedure are set forth in Appendix III of the validation report (4).

Plating and Pre-incubation of Cells for Testing

21. The number of 24-well plates, prepared as outlined in paragraph 19, that will be needed depends on the number of chemicals to be tested and the confluency of the cells in the culture dishes. As a general rule, one culture flask (75 cm^2) of 80-90% confluent cells will supply sufficient cells for one to 1.5 (24-well) plates at a target density of 200,000 to 300,000 cells per ml of medium resulting in approximately 50-60% confluency in the wells at 24 hours (Figure 2). This is typically the optimal cell density for hormone production in the assay. At higher densities, T as well as E2 production patterns are altered. Before conducting the assay the first time, it is recommended that different seeding densities between 200 000 and 300 000 cells per ml be tested, and the density resulting in 50-60% confluency in the well at 24 hours be selected for further experiments.
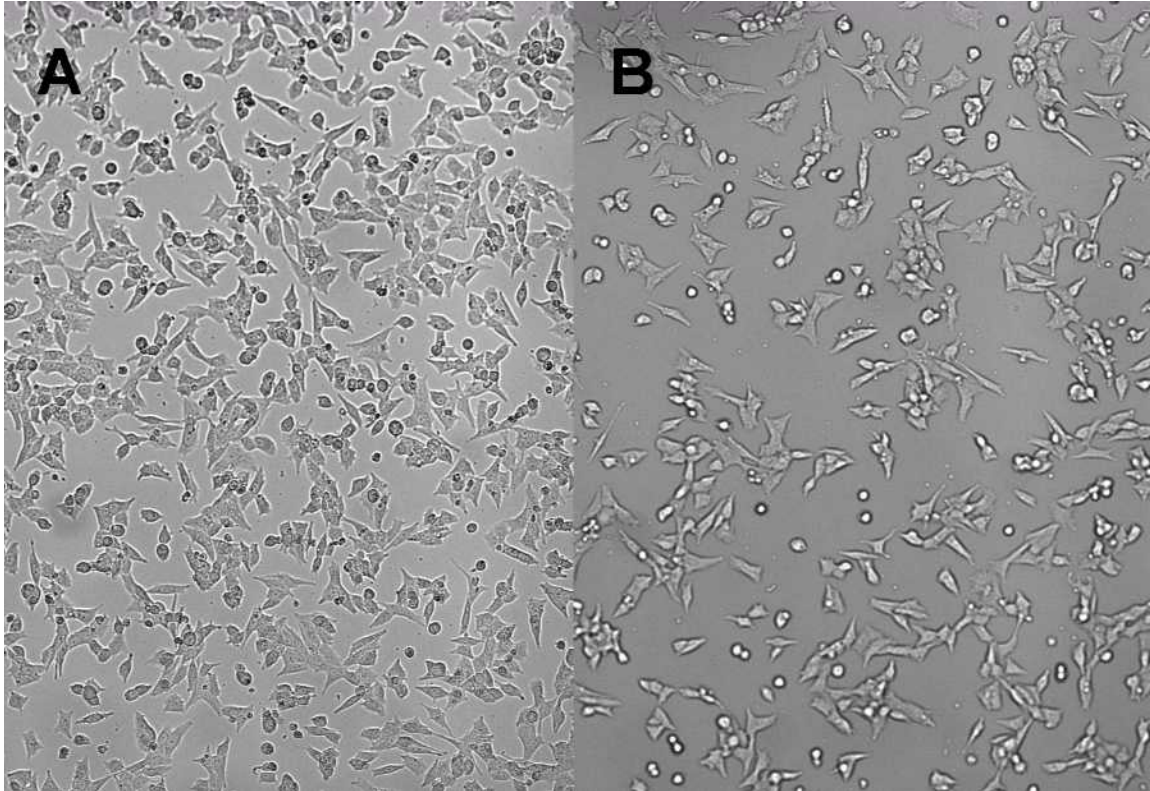
Figure 2: Photomicrograph of H295R cells at a seeding density of 50% in a 24 well culture plate at 24 hours taken at the edge (A) and centre (B) of a well.

22. The medium is pipetted off the culture flask, and the cells are rinsed 3 times with sterile PBS (without $Ca^{2+}Mg^{2+}$). An enzyme solution (in PBS) is added to detach the cells from the culture flask. Following an appropriate time for detachment of the cells, the enzyme action should be stopped with the addition of supplemented medium at a ratio of 3X the volume used for the enzyme treatment. Cells are placed into a centrifuge tube, centrifuged at room temperature, the supernatant is removed, and the pellet of cells is re-suspended in supplemented medium. The cell density is calculated using e.g. a haemocytometer or cell counter. The cell solution should be diluted to the desired plating density and thoroughly mixed to assure homogenous cell density. The cells should be plated with 1 ml of the cell solution/well and the plates and wells labelled. The seeded plates are incubated at 37°C under 5% CO_2 in air atmosphere for 24 hours to allow the cells to attach to the wells.

QUALITY CONTROL REQUIREMENTS

23. It is critical that exact volumes of solutions and samples are delivered into the wells during dosing because these volumes determine the concentrations used in the calculations of assay results.

24. Prior to the initiation of cell culture and any subsequent testing, each laboratory should demonstrate the sensitivity of its hormone measurement system (paragraphs 29-31).

25. If antibody-based hormone measurement assays are to be used, the chemicals to be

tested should be analysed for their potential to interfere with the measurement system used to quantify T and E2 as outlined in paragraph 32 prior to initiating testing.

26. DMSO is the recommended solvent for the assay. If an alternative solvent is utilised, the following should be determined:

   - The solubility of the test chemical, forskolin and prochloraz in the solvent; and

   - The cytotoxicity as a function of the concentration of solvent.

   It is recommended that the maximum allowable solvent concentration should not exceed a 10x dilution of the least cytotoxic concentration of the solvent.

27. Prior to conducting testing for the first time, the laboratory should conduct a qualifying experiment demonstrating that the laboratory is capable of maintaining and achieving appropriate cell culture and experimental conditions required for chemical testing as described in paragraphs 33-35.

28. When initiating testing using a new batch, a control plate should be run before using a new batch of cells to evaluate the performance of the cells as described in paragraphs 36 and 37.

Performance of the Hormone Measurement System

Method sensitivity, accuracy, precision and cross-reactivity with sample matrix

29. Each laboratory may use a hormone measurement system of its choice for the analysis of the production of T and E2 by H295R cells so long as it meets performance criteria, including the Limit of Quantification (LOQ). Nominally these are 100 pg/ml for T and 10 pg/ml for E2, which are based on the basal hormone levels observed in the validation studies. However, greater or lower levels may be appropriate depending upon the basal hormone levels achieved in the performing laboratory. Prior to initiation of QC plate and test runs, the laboratory should demonstrate that the hormone assay to be used can measure hormone concentrations in supplemented medium with sufficient accuracy and precision to meet the QC criteria specified in Tables 1 and 5 by analysing supplemented medium spiked with an internal hormone control. Supplemented medium should be spiked with at least three concentrations of each hormone (*e.g.* 100, 500 and 2500 pg/ml of T; 10, 50 and 250 pg/ml of E2; or the lowest possible concentrations based upon the detection limits of the chosen hormone measurement system can be used for the lowest spike concentrations for T and E2) and analysed. Measured hormone concentrations of non-extracted samples should be within 30% of nominal concentrations, and variation between replicate measurements of the same sample should not exceed 25% (see also Table 8 for additional QC criteria). If these QC criteria are fulfilled it is assumed that the selected hormone measurement assay is sufficiently accurate, precise and does not cross-react with components in the medium (sample matrix) such that a significant influence on the outcome of the assay would be expected. In this case, no extraction of samples prior to measurement of hormones is required.

30. In the case that the QC criteria in tables 1 and 8 are not fulfilled, a significant matrix effect may be occurring, and an experiment with extracted spiked medium should be conducted. An example of an extraction procedure is described in Appendix II of the

validation report (4). Measurements of the hormone concentrations in the extracted samples should be made in triplicate.[6] If it can be shown that after extraction the components of the medium do not interfere with the hormone detection method as defined by the QC criteria, all further experiments should be conducted using extracted samples. If the QC criteria cannot be met after extraction, the utilised hormone measurement system is not suitable for the purpose of the H295R Steroidogenesis Assay, and an alternative hormone detection method should be used.

Standard curve

31. The hormone concentrations of the solvent controls (SC) should be within the linear portion of the standard curve. Preferably, the SC values should fall close to the centre of the linear portion to ensure that induction and inhibition of hormone synthesis can be measured. Dilutions of medium (or extracts) to be measured are to be selected accordingly. The linear relationship is to be determined by a suitable statistical approach.

Chemical interference test

32. If antibody-based assays such as Enzyme-Linked Immunosorbent Assays (ELISAs) and Radio-Immuno Assays (RIAs) are going to be used to measure hormones, each chemical should be tested for potential interference with the hormone measurement system to be utilised prior to initiation of the actual testing of chemicals (Appendix III of the validation report (4)) because some chemicals can interfere with these tests (17). If interference occurs that is $\geq$ 20% of basal hormone production for T and/or E2 as determined by hormone analysis, the Chemical Hormone Assay Interference Test (such as described in Appendix III of the validation report (4) section 5.0) should be run on all test chemical stock solution dilutions to identify the threshold dose at which significant ($\geq$ 20%) interference occurs. If interference is less than 30%, results may be corrected for the interference. If interference exceeds 30%, the data are invalid and the data at these concentrations should be discarded. If significant interference of a test chemical with a hormone measurement system occurs at more than one non-cytotoxic concentration, a different hormone measurement system should be used. In order to avoid interference from contaminating chemicals it is recommended that hormones are extracted from the medium using suitable solvent, possible methods can be found in the validation report (4).

Table 1: Performance criteria for hormone measurement systems

Parameter	*Criterion*
Measurement Method Sensitivity	Limit of Quantification (LOQ) T: 100 pg/ml; E2: 10 pg/ml [a]
Hormone Extraction Efficiency (only when extraction is needed)	The average recovery rates (based on triplicate measures) for the spiked amounts of hormone should not deviate more than 30% from amount that was added.
Chemical Interference (only	No substantial ($\geq$30 % of basal hormone production of the

---

[6] Note: If extraction is required, three replicate measurements are made for each extract. Each sample will be extracted only once.

antibody based systems)	respective hormone) cross-reactivity with any of the hormones produced by the cells should occur b, c

[a] Note: Method measurement limits are based on the basal hormone production values provided in Table 5, and are performance based. If greater basal hormone production can be achieved the limit can be greater.

[b] Some T and E2 antibodies may cross-react with androstendione and oestrone, respectively, at a greater percentage. In such cases it is not possible to accurately determine effects on 17β-HSD. However, the data can still provide useful information regarding the effects on oestrogen or androgen production in general. In such cases data should be expressed as androgen/oestrogen responses rather than E2 and T.

[c] These include: cholesterol, pregnenolone, progesterone, 11-deoxycorticosterone, corticosterone, aldosterone, 17α-pregnenolone,17α-progesterone, deoxycortisol, cortisol, DHEA, androstenedione, oestrone.

Laboratory Proficiency Test

33. Before testing unknown chemicals, a laboratory should demonstrate that it is capable of achieving and maintaining appropriate cell culture and test conditions required for the successful conduct of the assay by running the laboratory proficiency test. As the performance of an assay is directly linked to the laboratory personnel conducting the assay, these procedures should be partly repeated if a change in laboratory personnel occurs.

34. This proficiency test will be conducted under the same conditions listed in paragraphs 38 through 40 by exposing cells to 7 increasing concentrations of strong, moderate and weak inducers and inhibitors as well as a negative chemical (see Table 2). Specifically, chemicals to be tested include the strong inducer forskolin (CAS no. 66575-29-9); the strong inhibitor prochloraz (CAS no. 67747-09-5); the moderate inducer atrazine (CAS no. 1912-24-9); the moderate inhibitor aminoglutethimide (CAS no. 125-84-8); the weak inducer (E2 production) and weak inhibitor (T production) bisphenol A (CAS no. 80-05-7); and the negative chemical human chorionic gonadotropin (HCG) (CAS no. 9002-61-3) as shown in Table 2. Separate plates are run for all chemicals using the format as shown in Table 6. One QC plate (Table 4, paragraphs 36-37) should be included with each daily run for the proficiency chemicals.

Table 2: Proficiency chemicals and exposure concentrations

Proficiency chemical	Test Concentrations [μM]
Prochloraz	0[a], 0.01, 0.03, 0.1, 0.3, 1, 3, 10
Forskolin	0[a], 0.03, 0.1, 0.3, 1, 3, 10, 30
Atrazine	0[a], 0.03, 0.1, 1, 3, 10, 30, 100
Aminoglutethimide	0[a], 0.03, 0.1, 1, 3, 10, 30, 100
Bisphenol A	0[a], 0.03, 0.1, 1, 3, 10, 30, 100
HCG	0[a], 0.03, 0.1, 1, 3, 10, 30, 100

[a]Solvent (DMSO) control (0), 1 μl DMSO/well

Exposure of H295R to proficiency chemicals should be conducted in 24 well plates during the laboratory proficiency test. Dosing is in μM for all test chemical doses. Doses should be administered in DMSO at 0.1% v/v per well. All test concentrations should be tested in triplicate wells (Table 6). Separate plates are run for each chemical. One QC plate is included with each daily run.

35. Cell viability and hormone analyses should be conducted as provided in paragraphs 42 through 46. The threshold value (lowest observed effect concentration, LOEC) and classification decision should be reported and compared with the values in Table 3. The data are considered acceptable if they meet the LOEC and decision classification in Table 3.

Table 3: Threshold values (LOECs) and decision classifications for Proficiency Chemicals

	CAS no.	LOEC [µM]		Decision Classification	
		T	*E2*	*T*	*E2*
Prochloraz	67747-09-5	≤0.1	≤1.0	+ [a] (Inhibition)	+ (Inhibition)
Forskolin	66575-29-9	≤10	≤0.1	+ (Induction)	+ (Induction)
Atrazine	1912-24-9	≤100	≤10	+ (Induction)	+ (Induction)
Aminoglutethimide	125-84-8	≤100	≤100	+ (Inhibition)	+ (Inhibition)
Bisphenol A	80-05-7	≤10	≤10	+ (Inhibition)	+ (Induction)
HCG	9002-61-3	n/a	n/a	Negative	Negative

[a] +, positive

n/a: not applicable as no changes should occur after exposure to non-cytotoxic concentrations of negative control.

Quality Control Plate

36. The quality control (QC) plate is used to verify the performance of the H295R cells under standard culture conditions, and to establish a historical database for hormone concentrations in solvent controls, positive and negative controls, as well as other QC measures over time.

    - H295R cell performance should be assessed using a QC plate for each new ATCC batch or after using a previously frozen stock of cells for the first time unless the laboratory proficiency test (paragraphs 32-34) has been run with that batch of cells.

    - A QC plate provides a complete assessment of the assay conditions (*e.g.* cell viability, solvent controls, negative and positive controls, as well as intra- and inter-assay variability) when testing chemicals and should be part of each test run.

37. The QC test is conducted in a 24-well plate and follows the same incubation, dosing, cell viability/cytotoxicity, hormone extraction and hormone analysis procedures described in paragraphs 38 through 46 for testing chemicals. The QC plate contains blanks, solvent controls, and two concentrations of a known inducer (forskolin, 1, 10 µM) and inhibitor (prochloraz, 0.1, 1 µM) of E2 and T synthesis. In addition, MeOH is used in select wells as a positive control for the viability/cytotoxicity assay. A detailed description of the plate layout is provided in Table 4. The criteria to be met on the QC plate are listed in Table 5. The minimum basal hormone production for T and E2 should be met in both the solvent control and blank wells.

Table 4: Quality control plate layout for testing performance of unexposed H295R cells and

cells exposed to known inhibitors (PRO = prochloraz) and stimulators (FOR = forskolin) of E2 and T production. After termination of the exposure experiment and removal of medium, a 70% methanol solution will be added to all MeOH wells to serve as a positive control for cytotoxicity (see cytotoxicity assay in Appendix III of the validation report (4))

	1	2	3	4	5	6
A	Blank [a]	Blank [a]	Blank [a]	Blank [a] (+ MeOH) [b]	Blank [a] (+ MeOH) [b]	Blank [a] (+ MeOH) [b]
B	DMSO [c] 1 µl	DMSO [c] 1 µl	DMSO [c] 1 µl	DMSO [c] 1 µl (+ MeOH) [b]	DMSO [c] 1 µl (+ MeOH) [b]	DMSO [c] 1 µl (+ MeOH) [b]
C	FOR 1 µM	FOR 1 µM	FOR 1 µM	PRO 0.1 µM	PRO 0.1 µM	PRO 0.1 µM
D	FOR 10 µM	FOR 10 µM	FOR 10 µM	PRO 1 µM	PRO 1 µM	PRO 1 µM

[a] Cells in Blank wells receive medium only (*i.e.* no solvent).
[b] Methanol (MeOH) will be added **after** the exposure is terminated and the medium is removed from these wells.
[c] DMSO solvent control (1 µl/well).

Table 5: Performance criteria for the Quality Control Plate

	T	E2
Basal Production of hormone in the solvent control (SC)	$\geq$ 5 times the LOQ	$\geq$ 2.5 times the LOQ
Induction (10 µM forskolin)	$\geq$ 1.5 times the SC	$\geq$ 7.5 times the SC
Inhibition (1µM prochloraz)	$\leq$ 0.5 times the SC	$\leq$ 0.5 times the SC

CHEMICAL EXPOSURE PROCEDURE

38. The pre-incubated cells are removed from the incubator (paragraph 21) and checked under a microscope to assure that they are in good condition (attachment, morphology) prior to dosing.

39. The cells are placed in a bio-safety cabinet and the supplemented medium removed and replaced with new supplemented medium (1 ml/well). DMSO is the preferred solvent for this test method. However, if there are reasons for using other solvents the scientific rationale should be described. Cells are exposed to the test chemical by adding 1 µl of the appropriate stock solution in DMSO (see Appendix II of the validation report (4)) per 1 ml supplemented medium (well volume). This results in a final concentration of 0.1% DMSO in the wells. To assure adequate mixing it is generally preferred that the

appropriate stock solution of the test chemical in DMSO is mixed with supplemented medium to yield the desired final concentration for each dose, and the mixture added to each well immediately after removal of old medium. If this option is used, the concentration of DMSO (0.1%) should remain consistent among all wells. The wells containing the greatest two concentrations are visually assessed for formation of precipitates or cloudiness as an indication of incomplete solubility of the test chemical by using a stereo microscope. If such conditions (cloudiness, formation precipitates) are observed, wells containing the next lesser concentrations are examined as well (and so forth) and concentrations that did not completely go into solution are to be excluded from further evaluation and analysis. The plate is returned to the incubator at 37°C under a 5% CO_2 in air atmosphere for 48 hours. The test chemical plate layout is shown in Table 6. Stocks 1 -7 show placement of increasing doses of test chemical.

Table 6: Dosing schematic for the exposure of H295R cells to test chemicals in a 24 well plate

	1	2	3	4	5	6
A	DMSO	DMSO	DMSO	Stock 4	Stock 4	Stock 4
B	Stock 1	Stock1	Stock 1	Stock 5	Stock 5	Stock 5
C	Stock 2	Stock 2	Stock 2	Stock 6	Stock 6	Stock 6
D	Stock 3	Stock 3	Stock 3	Stock 7	Stock 7	Stock 7

40. After 48 hours the exposure plates are removed from the incubator and every well is checked under the microscope for cell condition (attachment, morphology, degree of confluence) and signs of cytotoxicity. The medium from each well is split into two equal amounts (approximately 490 µl each) and transferred to two separate vials appropriately labelled (i.e. one aliquot to provide a spare sample for each well). To prevent cells from drying out, medium is removed a row or column at a time and replaced with the medium for the cell viability/cytotoxicity assay. If cell viability/cytotoxicity is not to be measured immediately, 200 µl PBS with Ca^{2+} and Mg^{2+} is added to each well. The media are frozen at -80ºC until further processing to analyse hormone concentrations (see paragraphs 44-46). While T and E2 in medium kept at -80°C are generally stable for at least 3 months, hormone stability during storage should be documented within each laboratory.

41. Immediately after removing the medium, cell viability/cytotoxicity is determined for each exposure plate.

Cell Viability Determination

42. A cell viability/cytotoxicity assay of choice can be used to determine the potential impact of the test chemical on cell viability. The assay should be able to provide a true measure of the percentage of viable cells present in a well, or it should be demonstrated that it is directly comparable to (a linear function of) the Live/Dead® Assay (see Appendix III of the validation report (4)). An alternative assay that has been shown to work equally well is the MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide] test (18). The assessment of cell viability using the above methods is a relative measurement that does not necessarily exhibit linear relationships with the absolute number of cells in a well. Therefore, a subjective parallel visual assessment of each well by the analyst should be conducted, and digital pictures of the SCs and the two

greatest non-cytotoxic concentrations are to be taken and archived to enable later assessment of true cell density if this should be required. If by visual inspection or as demonstrated by the viability/cytotoxicity assay there appears to be an increase in cell number, the apparent increase needs to be verified. If an increase in cell numbers is verified, this should be stated in the test report. Cell viability will be expressed relative to the average response in the SCs, which is considered 100% viable cells, and is calculated as appropriate for the cell viability/cytotoxicity assay that is used. For the MTT assay, the following formula may be used:

% viable cells = (response in well – average response in MeOH treated [=100% dead] wells) ÷ (average response in SC wells - average response in MeOH treated [=100% dead] wells)

43. Wells with viability lower than 80%, relative to the average viability in the SCs (=100% viability), should not be included in the final data analysis. Inhibition of steroidogenesis occurring in the presence of almost 20% cytotoxicity should be carefully evaluated to ensure that cytotoxicity is not the cause for the inhibition.

Hormone Analysis

44. Each laboratory can use a hormone measurement system of its choice for the analysis of T and E2. Spare aliquots of medium from each treatment group may be used to prepare dilutions to bring the concentration within the linear part of the standard curve. As noted in paragraph 29, each laboratory should demonstrate the conformance of their hormone measurement system (*e.g.* ELISA, RIA, LC-MS, LC-MS/MS) with the QC criteria by analysing supplemented medium spiked with an internal hormone control prior to conducting QC runs or testing of chemicals. In order to ensure that the components of the test system do not interfere with measurement of hormones, the hormones may need to be extracted from the media prior to their measurement (see paragraph 30 for the conditions under which an extraction is or is not required). It is recommended to conduct extraction following the procedures in Appendix III of the validation report (4).

45. If a commercial test kit is being used to measure the hormone production, the hormone analysis should be conducted as specified in the manuals provided by the test kit manufacturer. Most manufacturers have a unique procedure by which the hormone analyses are conducted. Dilutions of samples need to be adjusted such that expected hormone concentrations for the solvent controls fall within the centre of the linear range of the standard curve of the individual assay (Appendix III of the validation report (4)). Values outside of the linear portion of the standard curve should be rejected.

46. Final hormone concentrations are calculated as follows:

Example:

Extracted:	450 µl medium
Reconstituted in:	250 µl assay buffer
Dilution in Assay:	1:10 (to bring the sample within the linear range of the standard curve)

Hormone Concentration in Assay:	150 pg/ml (already adjusted to concentration per ml sample assayed)
Recovery:	89 %
Final hormone concentration =	(Hormone concentration (per ml) ÷ recovery) (dilution factor)
Final hormone concentration =	(150 pg/ml) ÷ (0.89) × (250 µl/450 µl) × 10 = 936.3 pg/ml

Selection of test concentrations

47. A minimum of two independent runs of the assay should be conducted. Unless prior information such as information on solubility limits or cytotoxicity provides a basis for selecting test concentrations, it is recommended that the test concentrations for the initial run be spaced at $\log_{10}$ intervals with 10^{-3} M being the maximum concentration. If the chemical is soluble, and not cytotoxic at any of the tested concentrations, and the first run was negative for all concentrations, then it is to be confirmed in one more run using the same conditions as the first run was conducted (Table 7). If the results of the first run are *equivocal* (i.e. the fold-change is statistically significant from the SC at only one concentration) or positive (i.e. the fold change at two or more adjacent concentrations is statistically significant), the test should be repeated as indicated in Table 7 by refining the selected test concentrations. Test concentrations in runs two and three (if applicable) should be adjusted on the basis of the results of the initial run bracketing concentrations that elicited an effect using ½-log concentration spacing (*e.g.* if the original run of 0.001, 0.01, 0.1, 1, 10, 100, 1000 µM resulted in inductions at 1 and 10 µM, the concentrations tested in the second run should be 0.1, 0.3, 1, 3, 10, 30, 100 µM), unless lower concentrations need to be employed to achieve a LOEC. In the latter case, at least five concentrations below the lowest concentration tested in the first run should be used in the second run using a ½-log scale. If the second run does not confirm the first run (*i.e.* statistical significance does not occur at the previously positively tested concentration ± 1 concentration-increment), a third experiment is to be conducted using the original testing conditions. Equivocal results in the first run are considered negative if the observed effect could not be confirmed in any of the two subsequent runs. Equivocal results are considered as positive responses (effect) when the response can be confirmed in at least one more run within a ± 1 concentration increment (see section 55 for the Data Interpretation Procedure).

Table 7: Decision matrix for possible outcome scenarios

Run 1	Run 2		Run 3		Decision	
Scenario	Decision	Scenario	Decision	Scenario	Positive	Negative
Negative	Confirm[a]	Negative	Stop			X
Negative	Confirm[a]	Positive	Refine[b]	Negative		X
Equivocal[c]	Refine[b]	Negative	Confirm[a]	Negative		X
Equivocal[c]	Refine[b]	Negative	Confirm[a]	Positive	X	
Equivocal[c]	Refine[b]	Positive			X	

Positive	Refine[b]	Negative	Confirm[a]	Positive	X
Negative	Confirm[a]	Positive	Refine[b]	Positive	X
Positive	Refine[b]	Positive	Stop		X

[a] Confirm previous run using the same experimental design.

[b] Re-run assay at ½-log concentration spacing (bracketing the concentration that tested significantly different in the preceding experiment).

[c] Fold-change at one concentration is statistically significantly different from the SC

Quality Control of the Test Plate

48. In addition to meeting the criteria for the QC plate, other quality criteria that pertain to acceptable variation between replicate wells, replicate experiments, linearity and sensitivity of hormone measurement systems, variability between replicate hormone measures of the same sample, and percentage recovery of hormone spikes after extraction of medium (if applicable; see Paragraph 30 regarding extraction requirements) should be met and are provided in Table 8. Data should fall within the acceptable ranges defined for each parameter to be considered for further evaluation. If these criteria are not met, the spreadsheet should note that QC criteria were not met for the sample in question, and the sample should be re-analysed or dropped from the data set.

Table 8: Acceptable ranges and/or variation (%) for H295R assay test plate parameters. LOQ: Limit of Quantification of the hormone measurement system. CV: Coefficient of variation; SC: Solvent Control; DPM: Disintegrations per minute.

	Comparison Between	T	E2
Basal hormone production in SCs	Fold-greater than LOQ	$\geq$ 5-fold	$\geq$ 2.5-fold
Exposure Experiments - Within Plate CV for SCs (Replicate Wells)	Absolute Concentrations	$\leq$ 30%	$\leq$ 30%
Exposure Experiments - Between Plate CV for SCs (Replicate Experiments)	Fold-Change	$\leq$ 30%	$\leq$ 30%
Hormone Measurement System – Sensitivity	Detectable fold-decrease relative to SC	$\geq$ 5-fold	$\geq$ 2.5-fold
Hormone Measurement System – Replicate Measure CV for SCs[a]	Absolute Concentrations	$\leq$ 25%	$\leq$ 25%
Medium Extraction – Recovery of Internal 3H Standard (If Applicable)	DPM	$\geq$ 65% Nominal	

[a] Refers to replicate measures of the same sample

## DATA ANALYSIS AND REPORTING

Data Analysis

49. To evaluate the relative increase/decrease in chemically altered hormone production, the

results should be normalised to the mean SC value of each test plate, and results expressed as changes relative to the SC in each test plate. All data are to be expressed as mean ± 1 standard deviation (SD).

50. Only hormone data from wells where cytotoxicity was less than 20% should be included in the data analysis. Relative changes should be calculated as follows:

    Relative Change = (Hormone concentration in each well) ÷ (Mean hormone concentration in all solvent control well).

51. If by visual inspection of the well or as demonstrated by the viability/cytotoxicity assay described in paragraph 42 there appears to be an increase in cell number, the apparent increase needs to be verified. If an increase in cell numbers is verified, this should be stated in the test report.

52. Prior to conducting statistical analyses, the assumptions of normality and variance homogeneity should be evaluated. Normality should be evaluated using standard probability plots or other appropriate statistical method (*e.g.* Shapiro-Wilk's test). If the data (fold changes) are not normally distributed, transformation of the data should be attempted to approximate a normal distribution. If the data are normally distributed or approximate a normal distribution, differences between chemical concentration groups and SCs should be analysed using a parametric test (*e.g.* Dunnett's Test) with *concentration* being the independent, and *response* (fold-change) being the dependent variable. If data are not normally distributed, an appropriate non-parametric test should be used (*e.g.* Kruskal Wallis, Steel's Many-one rank test). Differences are considered significant at *p ≤ 0.05*. Statistical evaluations are done based on average values for each well that represent independent replicate data points. It is anticipated that due to the large spacing of doses in the first run ($\log_{10}$ scale) in many cases it will not be possible to describe clear concentration-response relationships where the two greatest doses will be on the linear portion of the sigmoid curve. Therefore, for the first run or any other data sets where this condition occurs (*e.g.* where no maximum efficacy can be estimated) type I fixed variable statistics as described above will be applied.

53. If more than two data points lie on the linear portion of the curve and where maximum efficacies can be calculated - as is anticipated for some of the 2nd runs that are conducted using a semi-log spacing of exposure concentrations - a probit, logit or other appropriate regression model should be utilized to calculate effective concentrations (*e.g.* EC50 and EC20).

54. Results should be provided both in graphical (bar graphs representing mean +/- 1 SD) and tabular (LOEC/NOEC, direction of effect, and strength of maximum response that is part of the dose-response portion of the data) formats (see Figure 3 for an example). Data assessment is only considered valid if it has been based on at least two independently conducted runs. An experiment or run is considered independent if it has been conducted at a different date using a new set of solutions and controls. The concentration range used in runs 2 and 3 (if necessary) may be tailored on the basis of the results of run 1 to better define the dose response range containing the LOEC (see paragraph 47).
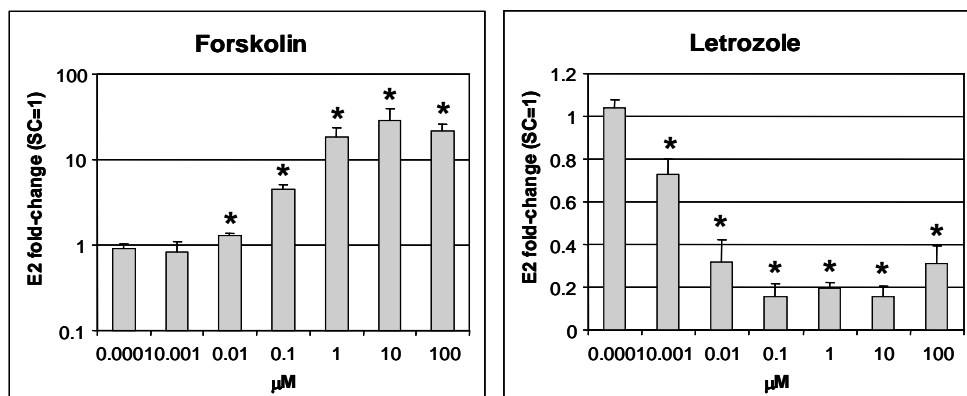
Chemical	LOEC	Max Change
Forskolin	0.01	0.15 fold
Letrozole	0.001	29 fold

Figure 3: Example of the presentation and evaluation of data obtained during the conduct of the H295R Assay in graphical and tabular format. Asterisks indicate statistically significant differences from the solvent control (*p<0.05*). LOEC: Lowest observed effective concentration; Max Change: Maximum strength of the response observed at any concentration relative to the average SC response (=1).

Data Interpretation Procedure

55. A test chemical is judged to be positive if the fold induction is statistically different ($p \leq 0.05$) from the solvent control at two adjacent concentrations in at least two independent runs (Table 7). A test chemical is judged to be negative following two independent negative runs, or in three runs, comprising two negative runs and one equivocal or positive run. If the data generated in three independent experiments does not meet the decision criteria listed in Table 7, the experimental results are not interpretable. Results at concentrations exceeding the limits of solubility or at cytotoxic concentrations should not be included in the interpretation of results.

Test Report

56. The test report should include the following information:

Testing facility

- Name of facility and location;
- Study director and other personnel and their study responsibilities;
- Dates the study began and ended;

Test chemical, reagents and controls

- Identity (name/CAS no. as appropriate), source, lot/batch number, purity, supplier, and characterisation of test chemical, reagents, and controls;
- Physical nature and relevant physicochemical properties of test chemical;

132

- Storage conditions and the method and frequency of preparation of test chemicals, reagents and controls;

- Stability of test chemical;

Cells

- Source and type of cells;

- Number of cell passages (cell passage identifier) of cells used in test;

- Description of procedures for maintenance of cell cultures;

Pre-test requirements (if applicable)

- Description and results of chemical hormone-assay interference test;

- Description and results of hormone extraction efficiency measurements;

- Standard and calibration curves for all analytical assays to be conducted;

- Detection limits for the selected analytical assays;

Test conditions

- Composition of media;

- Concentration of test chemical;

- Cell density (estimated or measured cell concentrations at 24 hours and 48 hours)

- Solubility of test chemical (limit of solubility, if determined);

- Incubation time and conditions;

Test results

- Raw data for each well for controls and test chemicals--each replicate measure in form of the original data provided by the instrument utilised to measure hormone production (*e.g.* OD, fluorescence units, DPM, etc.);

- Validation of normality or explanation of data transformation;

- Mean responses +/- 1 SD for each well measured;

- Cytotoxicity data (test concentrations that caused cytotoxicity);

- Confirmation that QC requirements were met;

- Relative change compared with solvent control corrected for cytotoxicity;

- A bar graph showing relative (fold change) at each concentration, SD and statistical significance as stated in paragraph 49-54;

Data interpretation

- Apply the data interpretation procedure to the results and discuss findings;

Discussion

- Are there any indications from the study regarding the possibility that the T/E2 data could be influenced by indirect effects on the gluco-, and mineral-corticoid

pathways?

Conclusions

LITERATURE

(1) OECD (2002), OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals, in Appendix 2 to Chapter B.54 of this Annex

(2) Hecker, M., Newsted, J.L., Murphy, M.B., Higley, E.B., Jones, P.D., Wu, R. and Giesy, J.P. (2006), Human adrenocarcinoma (H295R) cells for rapid in vitro determination of effects on steroidogenesis: Hormone production, *Toxicol. Appl. Pharmacol.,* 217, 114-124.

(3) Hecker, M., Hollert, H., Cooper, R., Vinggaard, A.-M., Akahori, Y., Murphy, M., Nellemann, C., Higley, E., Newsted, J., Wu, R., Lam, P., Laskey, J., Buckalew, A., Grund, S., Nakai, M., Timm, G., and Giesy, J. P. (2007), The OECD validation program of the H295R steroidgenesis assay for the identification of in vitro inhibitors or inducers of testosterone and estradiol production, Phase 2: inter laboratory pre-validation studies. *Env. Sci. Pollut. Res*., 14, 23−30.

(4) OECD (2010), *Multi-Laboratory Validation of the H295R Steroidogenesis Assay to Identify Modulators of Testosterone and Estradiol Production*, OECD Series of Testing and Assessment No. 132, ENV/JM/MONO(2010)31, Paris. Available at [http://www.oecd.org/document/30/0,3746,en_2649_34377_1916638_1_1_1_1,00.html]

(5) OECD (2010), *Peer Review Report of the H295R Cell-Based Assay for Steroidogenesis,* OECD Series of Testing and Assessment No. 133, ENV/JM/MONO(2010)32, Paris. Available at: [http://www.oecd.org/document/30/0,3746,en_2649_34377_1916638_1_1_1_1,00.html]

(6) Battelle (2005), Detailed Review Paper on Steroidogenesis, Available at: [http://www.epa.gov/endo/pubs/edmvs/steroidogenesis_drp_final_3_29_05.pdf]

(7) Hilscherova, K., Jones, P. D., Gracia, T., Newsted, J. L., Zhang, X., Sanderson, J. T., Yu, R. M. K., Wu, R. S. S. and Giesy, J. P. (2004), Assessment of the Effects of Chemicals on the Expression of Ten Steroidogenic Genes in the H295R Cell Line Using Real-Time PCR, *Toxicol. Sci.,* 81, 78-89.

(8) Sanderson, J. T., Boerma, J., Lansbergen, G. and Van den Berg, M. (2002), Induction and inhibition of aromatase (CYP19) activity by various classes of pesticides in H295R human adrenocortical carcinoma cells, *Toxicol. Appl. Pharmacol.,* 182, 44-54.

(9) Breen, M.S., Breen, M., Terasaki, N., Yamazaki, M. and Conolly, R.B. (2010), Computational model of steroidogenesis in human H295R cells to predict biochemical response to endocrine-active chemicals: Model development for metyrapone, *Environ. Health Perspect.*, 118: 265-272.

(10) Higley, E.B., Newsted, J.L., Zhang, X., Giesy, J.P. and Hecker, M. (2010), Assessment of chemical effects on aromatase activity using the H295R cell line, *Environ. Sci. Poll. Res.*, 17:1137-1148.

(11) Gazdar, A. F., Oie, H. K., Shackleton, C. H., Chen, T. R., Triche, T. J., Myers, C. E., Chrousos, G. P., Brennan, M. F., Stein, C. A. and La Rocca, R. V. (1990), Establishment and characterization of a human adrenocortical carcinoma cell line that expresses Multiple pathways of steroid biosynthesis, *Cancer Res.*, 50, 5488-5496.

(12) He, Y.H., Wiseman, S.B., Zhang, X.W., Hecker, M., Jones, P.D., El-Din, M.G., Martin, J.W. and Giesy, J.P. (2010), Ozonation attenuates the steroidogenic disruptive effects of sediment free oil sands process water in the H295R cell line, *Chemosphere*, 80:578-584.

(13) Zhang, X.W., Yu, R.M.K., Jones, P.D., Lam, G.K.W., Newsted, J.L., Gracia, T., Hecker, M., Hilscherova, K., Sanderson, J.T., Wu, R.S.S. and Giesy, J.P. (2005), Quantitative RT-PCR methods for evaluating toxicant-induced effects on steroidogenesis using the H295R cell line, *Environ. Sci. Technol.*, 39:2777-2785.

(14) Higley, E.B., Newsted, J.L., Zhang, X., Giesy, J.P. and Hecker, M. (2010), Differential assessment of chemical effects on aromatase activity, and E2 and T production using the H295R cell line, *Environ. Sci. Pol. Res.*, 17:1137-1148.

(15) Rainey, W. E., Bird, I. M., Sawetawan, C., Hanley, N. A., Mccarthy, J. L., Mcgee, E. A., Wester, R. and Mason, J. I. (1993), Regulation of human adrenal carcinoma cell (NCI-H295) production of C19 steroids, *J. Clin. Endocrinol. Metab.*, 77, 731-737.

(16) Chapter B.55 of this Annex: Hershberger Bioassay in Rats: A short-term Screening Assay for (Anti)Androgenic Properties.

(17) Shapiro, R., and Page, L.B. (1976), Interference by 2,3-dimercapto-1-propanol (BAL) in angiotensin I radioimmunoassay, *J. Lab. Clin. Med.*, 2, 222-231.

(18) Mosmann, T. (1983), Rapid colorimetric assay for growth and survival: application to proliferation and cytotoxicity assays, *J. Immunol. Methods.*, 65, 55-63.

(19) Brock, B.J., Waterman, M.R. (1999). Biochemical differences between rat and human cytochrome P450c17 support the different steroidogenic needs of these two species, Biochemistry. 38:1598-1606.

(20) Oskarsson, A., Ulleras, E., Plant, K., Hinson, J. Goldfarb, P.S., (2006), Steroidogenic gene expression in H295R cells and the human adrenal gland: adrenotoxic effects of lindane in vitro, J. Appl. Toxicol., 26:484-492.

DEFINITIONS

Confluency refers to the coverage or proliferation that the cells are allowed over or throughout the culture medium.

Chemical means a substance or a mixture.

CV refers to the coefficient of variation, and is defined as the ratio of the standard deviation of a distribution to its arithmetic mean.

CYP stands for cytochrome P450 mono-oxygenases, a family of genes and the enzymes produced from them that are involved in catalyzing a wide variety of biochemical reactions including the synthesis and metabolism of steroid hormones.

DPM are disintegration per minute. It is the number of atoms in a given quantity of radioactive material that is detected to have decayed in one minute.

E2 is 17β-oestradiol, the most important oestrogen in mammalian systems.

H295R cells are human adreno-carcinoma cells which have the physiological characteristics of zonally undifferentiated human foetal adrenal cells and which express all of the enzymes of the steroidogenesis pathway. They are available from the ATCC.

Freeze medium is used to freeze and to store frozen cells. It consists of stock medium plus BD NuSerum and dimethyl sulfoxide.

Linear Range is the range within the standard curve for a hormone measurement system where the results are proportional to the concentration of the analyte present in the sample.

LOQ stands for "*Limit of Quantification*", and is the lowest quantity of a chemical that can be distinguished from the absence of that chemical (a blank value) within a stated confidence limit. For the purpose of this method, the LOQ is typically defined by the manufacturer of the test systems if not specified differently.

LOEC is the Lowest Observed Effect Concentration, the lowest concentration level at which the assay response is statistically different from that of the solvent control.

NOEC is the No Observed Effect Concentration, which is the highest concentration tested if the assay does not provide a positive response.

Passage is the number of times that cells are split after initiation of a culture from frozen stock. The initial passage that was started from the frozen stock is assigned the number one (1). Cells that were split 1 time are labelled passage 2, etc.

PBS is Dulbecco's phosphate buffered saline.

Quality Control, abbreviated QC, refers to the measures needed to assure valid data.

Quality control plate is a 24 well plate containing two concentrations of the positive and negative controls to monitor the performance of a new batch of cells or to provide the positive controls for the assay when testing chemicals.

Run is an independent experiment characterised by a new set of solutions and controls.

Stock medium is the base for the preparation of other reagents. It consists of a 1:1 mixture of Dulbecco's Modified Eagle's Medium and Ham's F-12 Nutrient mixture (DMEM/F12) in 15 mM HEPES buffer without phenol red or sodium bicarbonate. Sodium bicarbonate is added as the buffer, see Appendix II of the validation report (4).

Supplemented medium consist of stock medium plus BD Nu-Serum and ITS+ premium mix, see Appendix II of the validation report (4).

Steroidogenesis is the synthetic pathway leading from cholesterol to the various steroid hormones. Several intermediates in the steroid synthesis pathway such as progesterone and testosterone are important hormones in their own right but also serve as precursors to hormones farther down the synthetic pathway.

T stands for testosterone, one of the two most important androgens in mammalian systems.

Test chemical is any substance or mixture tested using this test method.

Test plate is the plate on which H295R cells are exposed to test chemicals. Test plates contain the solvent control and the test chemical at seven concentration levels in triplicate.

Trypsin 1X is a dilute solution of the enzyme trypsin, a pancreatic serine protease, used to loosen cells from a cell cultivation plate, see Appendix III of the validation report (4).

B.58. Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays

INTRODUCTION

1. This test method is equivalent to OECD Test Guideline (TG) 488 (2013). EU test methods are available for a wide range of *in vitro* mutation assays that are able to detect chromosomal and/or gene mutations. There are test methods for *in vivo* endpoints (i.e. chromosomal aberrations and unscheduled DNA synthesis); however, these do not measure gene mutations. Transgenic Rodent (TGR) mutation assays fulfil the need for practical and widely available *in vivo* tests for gene mutations.

2. The TGR mutation assays have been reviewed extensively (24) (33). They use transgenic rats and mice that contain multiple copies of chromosomally integrated plasmid or phage shuttle vectors. The transgenes contain reporter genes for the detection of various types of mutations induced *in vivo* by test chemicals.

3. Mutations arising in a rodent are scored by recovering the transgene and analysing the phenotype of the reporter gene in a bacterial host deficient for the reporter gene. TGR gene mutation assays measure mutations induced in genetically neutral genes recovered from virtually any tissue of the rodent. These assays, therefore, circumvent many of the existing limitations associated with the study of *in vivo* gene mutation in endogenous genes (e.g. limited tissues suitable for analysis, negative/positive selection against mutations).

4. The weight of evidence suggests that transgenes respond to mutagens in a similar manner to endogenous genes, especially with regard to the detection of base pair substitutions, frameshift mutations, and small deletions and insertions (24).

5. The International Workshops on Genotoxicity Testing (IWGT) have endorsed the inclusion of TGR gene mutation assays for *in vivo* detection of gene mutations, and have recommended a protocol for their implementation (15) (29). This test method is based on these recommendations. Further analysis supporting the use of this protocol can be found in (16).

6. It is anticipated that in the future it may be possible to combine a TGR gene mutation assay with a repeat dose toxicity study (Chapter B.7 of this Annex). However, data are required to ensure that the sensitivity of TGR gene mutation assays is unaffected by the shorter one day period of time between the end of the administration period and the sampling time, as used in the repeat dose toxicology study, compared to 3 days used in TGR gene mutation assays. Data are also required to indicate that the performance of the repeat dose assay is not adversely affected by using a transgenic rodent strain rather than traditional rodent strains. When these data are available, this test method will be updated.

7. Definitions of key terms are set out in the Appendix.

INITIAL CONSIDERATIONS

8. TGR gene mutation assays for which sufficient data are available to support their use in this test method are: *lacZ* bacteriophage mouse (Muta™Mouse); *lacZ* plasmid mouse; *gpt* delta (*gpt* and Spi⁻) mouse and rat; *lacI* mouse and rat (Big Blue®), as performed under standard conditions. In addition, the *cII* positive selection assay can be used for evaluating mutations in the Big Blue® and Muta™Mouse models. Mutagenesis in the TGR models is normally assessed as mutant frequency; if required, however, molecular analysis of the mutations can provide additional information (see paragraph 24).

9. These rodent *in vivo* gene mutation tests are especially relevant to assessing mutagenic hazard in that the assays' responses are dependent upon *in vivo* metabolism, pharmacokinetics, DNA repair processes, and translesion DNA synthesis, although these may vary among species, among tissues and among the types of DNA damage. An *in vivo* assay for gene mutations is useful for further investigation of a mutagenic effect detected by an *in vitro* system, and for following up results of tests using other *in vivo* endpoints (24). In addition to being causally associated with the induction of cancer, gene mutation is a relevant endpoint for the prediction of mutation-based non-cancer diseases in somatic tissues (12) (13) as well as diseases transmitted through the germline.

10. If there is evidence that the test chemical, or a relevant metabolite, will not reach any of the tissues of interest, it is not appropriate to perform a TGR gene mutation assay.

PRINCIPLE OF THE TEST

11. In the assays described in paragraph 8, the target gene is bacterial or bacteriophage in origin, and the means of recovery from the rodent genomic DNA is by incorporation of the transgene into a λ bacteriophage or plasmid shuttle vector. The procedure involves the extraction of genomic DNA from the rodent tissue of interest, *in vitro* processing of the genomic DNA (i.e. packaging of λ vectors, or ligation and electroporation of plasmids to recover the shuttle vector), and subsequent detection of mutations in bacterial hosts under suitable conditions. The assays employ neutral transgenes that are readily recoverable from most tissues.

12. The basic TGR gene mutation experiment involves treatment of the rodent with a chemical over a period of time. Chemicals may be administered by any appropriate route, including implantation (e.g. medical device testing). The total period during which an animal is dosed is referred to as the administration period. Administration is usually followed by a period of time, prior to sacrifice, during which the chemical is not administered and during which unrepaired DNA lesions are fixed into stable mutations. In the literature, this period has been variously referred to as the manifestation time, fixation time or expression time; the end of this period is the sampling time (15) (29). After the animal is sacrificed, genomic DNA is isolated from the tissue(s) of interest and purified.

13. Data for a single tissue per animal from multiple packaging/ligations are usually aggregated, and mutant frequency is generally evaluated using a total of between 10^5 and 10^7 plaque-forming or colony-forming units. When using positive selection methods, total plaque-forming units are determined with a separate set of non-selective

plates.

14. Positive selection methods have been developed to facilitate the detection of mutations in both the *gpt* gene [*gpt* delta mouse and rat, *gpt*⁻ phenotype (20) (22) (28)] and the *lacZ* gene [Muta^{TM}Mouse or *lacZ* plasmid mouse (3) (10) (11) (30)]; whereas, *lacI* gene mutations in Big Blue® animals are detected through a non-selective method that identifies mutants through the generation of coloured (blue) plaques. Positive selection methodology is also in place to detect point mutations arising in the *cII* gene of the λ bacteriophage shuttle vector [Big Blue® mouse or rat, and Muta^{TM}Mouse (17)] and deletion mutations in the λ *red* and *gam* genes [Spi⁻ selection in *gpt* delta mouse and rat (21) (22) (28)]. Mutant frequency is calculated by dividing the number of plaques/plasmids containing mutations in the transgene by the total number of plaques/plasmids recovered from the same DNA sample. In TGR gene mutation studies, the mutant frequency is the reported parameter. In addition, a mutation frequency can be determined as the fraction of cells carrying independent mutations; this calculation requires correction for clonal expansion by sequencing the recovered mutants (24).

15. The mutations scored in the *lacI, lacZ, cII* and *gpt* point mutation assays consist primarily of base pair substitution mutations, frameshift mutations and small insertions/deletions. The relative proportion of these mutation types among spontaneous mutations is similar to that seen in the endogenous *Hprt* gene. Large deletions are detected only with the Spi⁻ selection and the *lacZ* plasmid assays (24). Mutations of interest are *in vivo* mutations that arise in the mouse or rat. *In vitro* and *ex vivo* mutations, which may arise during phage/plasmid recovery, replication or repair, are relatively rare, and in some systems can be specifically identified, or excluded by the bacterial host/positive selection system.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

16. A variety of transgenic mouse gene mutation detection models are currently available, and these systems have been more widely used than transgenic rat models. If the rat is clearly a more appropriate model than the mouse (e.g. when investigating the mechanism of carcinogenesis for a tumour seen only in rats, to correlate with a rat toxicity study, or if rat metabolism is known to be more representative of human metabolism) the use of transgenic rat models should be considered.

Housing and feeding conditions

17. The temperature in the experimental animal room ideally should be 22°C (±3°C). Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning, the goal should be to maintain a relative humidity of 50-60%. Lighting should be artificial, with a daily sequence of 12 hours light, followed by 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this route. Animals should be housed in small groups (no more than five) of the same sex if no aggressive

behaviour is expected. Animals may be housed individually if scientifically justified.

Preparation of the animals

18. Healthy young sexually mature adult animals (8-12 weeks old at start of treatment) are randomly assigned to the control and treatment groups. The animals are identified uniquely. The animals are acclimated to the laboratory conditions for at least five days. Cages should be arranged in such a way that possible effects due to cage placement are minimised. At the commencement of the study, the weight variation of animals should be minimal and not exceed ± 20% of the mean weight of each sex.

Preparation of doses

19. Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles or admixed in diet or drinking water prior to dosing of the animals. Liquid test chemicals may be dosed directly or diluted prior to dosing. For inhalation exposures, test chemicals can be administered as gas, vapour or a solid/liquid aerosol, depending on their physicochemical properties. Fresh preparations of the test chemical should be employed unless stability data demonstrate the acceptability of storage.

Test Conditions

Solvent/vehicle

20. The solvent/vehicle should not produce toxic effects at the dose volumes used, and should not be suspected of chemical reaction with the test chemical. If other than well-known solvents/vehicles are used, their inclusion should be supported with reference data indicating their compatibility. It is recommended that wherever possible, the use of an aqueous solvent/vehicle should be considered first.

Positive Controls

21. Concurrent positive control animals should normally be used. However, for laboratories that have demonstrated competency (see paragraph 23) and routinely use these assays, DNA from previous positive control treated animals may be included with each study to confirm the success of the method. Such DNA from previous experiments should be obtained from the same species and tissues of interest, and properly stored (see paragraph 36). When concurrent positive controls are used, it is not necessary to administer them by the same route as the test chemical; however, the positive controls should be known to induce mutations in one or more tissues of interest for the test chemical. The doses of the positive control chemicals should be selected so as to produce weak or moderate effects that critically assess the performance and sensitivity of the assay. Examples of positive control chemicals and some of their target tissues are included in Table 1.

Table 1: Examples of positive control chemicals and some of their target tissues

Positive control chemical and CAS No.	EINECS name and	Characteristics	Mutation Target Tissue	
			Rat	Mouse

	EINECS No.			
N-Ethyl-N-nitrosourea [CAS no. 759-73-9]	N-Ethyl-N-nitrosourea [212-072-2]	Direct acting mutagen	Liver, lung	Bone marrow, colon, colonic epithelium, intestine, liver, lung, spleen, kidney, ovarian granulosa cells, male germ cells
Ethyl carbamate (urethane) [CAS no. 51-79-6]	Urethane [200-123-1]	Mutagen, requires metabolism but produces only weak effects		Bone marrow, forestomach, small intestine, liver, lung, spleen
2,4-Diaminotoluene [CAS no. 95-80-7]	4-Methyl-m-phenylenediamine [202-453-1]	Mutagen, requires metabolism, also positive in the Spi⁻ assay	Liver	Liver
Benzo[a]pyrene [CAS no. 50-32-8]	Benzo[def]chrysene [200-028-5]	Mutagen, requires metabolism	Liver, omenta,	Bone marrow, breast, colon, forestomach, glandular stomach, heart, liver, lung, male germ cells

Negative controls

22. Negative controls, treated with solvent or vehicle alone, and otherwise treated in the same way as the treatment groups, should be included for every sampling time. In the absence of historical or published control data showing that no deleterious or mutagenic effects are induced by the chosen solvent/vehicle, untreated controls should also be included for every sampling time in order to establish acceptability of the vehicle control.

Verification of laboratory proficiency

23. Competency in these assays should be established by demonstrating the ability to reproduce expected results from published data (24) for: 1) mutant frequencies with positive control chemicals (including weak responses) such as those listed in Table 1, non-mutagens, and vehicle controls; and 2) transgene recovery from genomic DNA (e.g. packaging efficiency).

Sequencing of mutants

24. For regulatory applications, DNA sequencing of mutants is not required, particularly where a clear positive or negative result is obtained. However, sequencing data may be useful when high inter-individual variation is observed. In these cases, sequencing can be used to rule out the possibility of jackpots or clonal events by identifying the proportion of unique mutants from a particular tissue. Sequencing approximately 10 mutants per tissue per animal should be sufficient for simply determining if clonal mutants contribute to the mutant frequency; sequencing as many as 25 mutants may be necessary to correct mutant frequency mathematically for clonality. Sequencing of mutants also may be considered when small increases in mutant frequency (i.e. just exceeding the untreated control values) are found. Differences in the mutant spectrum between the mutant colonies from treated and untreated animals may lend support to a mutagenic effect (29). Also, mutation spectra may be useful for developing mechanistic hypotheses. When sequencing is to be included as part of the study protocol, special care should be taken in the design of such studies, in particular with respect to the number of mutants sequenced per sample, to achieve adequate power according to the statistical model used (see paragraph 43).

PROCEDURE

Number and Sex of Animals

25. The number of animals per group should be predetermined to be sufficient to provide statistical power necessary to detect at least a doubling in mutant frequency. Group sizes will consist of a minimum of five animals; however, if the statistical power is insufficient, the number of animals should be increased as required. Male animals should normally be used. There may be cases where testing females alone would be justified; for example, when testing human female-specific drugs, or when investigating female-specific metabolism. If there are significant differences between the sexes in terms of toxicity or metabolism, then both males and females will be required.

Administration Period

26. Based on observations that mutations accumulate with each treatment, a repeated-dose regimen is necessary, with daily treatments for a period of 28 days. This is generally considered acceptable both for producing a sufficient accumulation of mutations by weak mutagens, and for providing an exposure time adequate for detecting mutations in slowly proliferating organs. Alternative treatment regimens may be appropriate for some evaluations, and these alternative dosing schedules should be scientifically justified in the protocol. Treatments should not be shorter than the time required for the complete induction of all the relevant metabolising enzymes, and shorter treatments may necessitate the use of multiple sampling times that are suitable for organs with different proliferation rates. In any case, all available information (e.g. on general toxicity or metabolism and pharmacokinetics) should be used when justifying a protocol, especially when deviating from the above standard recommendations. While it may increase sensitivity, treatment times longer than 8 weeks should be explained clearly and justified, since long treatment times may produce an apparent increase in mutant frequency through clonal expansion (29).

Dose Levels

27. Dose levels should be based on the results of a dose range-finding study measuring general toxicity that was conducted by the same route of exposure, or on the results of pre-existing sub-acute toxicity studies. Non-transgenic animals of the same rodent strain may be used for determining dose ranges. In the main test, in order to obtain dose response information, a complete study should include a negative control group (see paragraph 22) and a minimum of three, appropriately-spaced dose levels, except where the limit dose has been used (see paragraph 28). The top dose should be the Maximum Tolerated Dose (MTD). The MTD is defined as the dose producing signs of toxicity such that higher dose levels, based on the same dosing regimen, would be expected to produce lethality. Chemicals with specific biological activities at low non-toxic doses (such as hormones and mitogens), and chemicals which exhibit saturation of toxicokinetic properties may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis. The dose levels used should cover a range from the maximum to little or no toxicity.

Limit Test

28. If dose range-finding experiments, or existing data from related rodent strains, indicate that a treatment regime of at least the limit dose (see below) produces no observable toxic effects, and if genotoxicity would not be expected based upon data from structurally related chemicals, then a full study using three dose levels may not be considered necessary. For an administration period of 28 days (i.e. 28 daily treatments), the limit dose is 1000 mg/kg body weight/day. For administration periods of 14 days or less, the limit dose is 2000 mg/kg/body weight/day (dosing schedules differing from 28 daily treatments should be scientifically justified in the protocol; see paragraph 26).

Administration of Doses

29. The test chemical is usually administered by gavage using a stomach tube or a suitable intubation cannula. In general, the anticipated route of human exposure should be considered when designing an assay. Therefore, other routes of exposure (such as drinking water, subcutaneous, intravenous, topical, inhalation, intratracheal, dietary, or implantation) may be acceptable where they can be justified. Intraperitoneal injection is not recommended since it is not a physiologically relevant route of human exposure. The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not exceed 2 ml/100g body weight. The use of volumes greater than this should be justified. Except for irritating or corrosive chemicals, which will normally reveal exacerbated effects at higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure a constant volume at all dose levels.

Sampling Time

Somatic Cells

30. The sampling time is a critical variable because it is determined by the period needed for mutations to be fixed. This period is tissue-specific and appears to be related to the turnover time of the cell population, with bone marrow and intestine being rapid responders and the liver being much slower. A suitable compromise for the

measurement of mutant frequencies in both rapidly and slowly proliferating tissues is 28 consecutive daily treatments (as indicated in paragraph 26) and sampling three days after the final treatment; although the maximum mutant frequency may not manifest itself in slowly proliferating tissues under these conditions. If slowly proliferating tissues are of particular importance, then a later sampling time of 28 days following the 28 day administration period may be more appropriate (16) (29). In such cases, the later sampling time would replace the 3 day sampling time, and would require scientific justification.

Germ Cells

31. TGR assays are well-suited for the study of gene mutation induction in male germ cells (7) (8) (27), in which the timing and kinetics of spermatogenesis have been well-defined (27). The low numbers of ova available for analysis, even after super-ovulation, and the fact that there is no DNA synthesis in the oocyte, preclude the determination of mutation in female germ cells using transgenic assays (31).

32. The sampling times for male germ cells should be selected so that the range of exposed cell types throughout germ cell development is sampled, and so that the stage targeted in the sampling has received sufficient exposure. The time for the progression of developing germ cells from spermatogonial stem cells to mature sperm reaching the vas deferens/cauda epididymis is ~49 days for the mouse (36) and ~70 days for the rat (34) (35). Following a 28-day exposure with a subsequent three day sampling period, accumulated sperm collected from the vas deferens/cauda epididymis (7)(8) will represent a population of cells exposed during approximately the latter half of spermatogenesis, which includes the meiotic and postmeiotic period, but not the spermatogonial or stem cell period. In order to adequately sample cells in the vas deferens/cauda epididymis that were spermatogonial stem cells during the exposure period, an additional sampling time at a minimum of 7 weeks (mice) or 10 weeks (rat), after the end of treatment is required.

33. Cells extruded from seminiferous tubules after a 28 + 3 day regimen comprise a mixed population enriched for all stages of developing germ cells (7) (8). Sampling these cells for gene mutation detection does not provide as precise an assessment of the stages at which germ cell mutations are induced as can be obtained from sampling spermatozoa from the vas deferens/cauda epididymis (since there is a range of germ cell types sampled from the tubules, and there will be some somatic cells contaminating this cell population). However, sampling cells from seminiferous tubules in addition to spermatozoa from the vas deferens/cauda epididymis following only a 28 + 3 day sampling regimen would provide some coverage of cells exposed across the majority of phases of germ cell development, and may be useful for detecting some germ cell mutagens.

Observations

34. General clinical observations should be made at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. The health condition of the animals should be recorded. At least twice daily, all animals should be observed for morbidity and mortality. All animals should be weighed at least once a week, and at sacrifice. Measurements of food consumption should be made at least weekly. If the test chemical is administered via the drinking water, water

consumption should be measured at each change of water and at least weekly. Animals exhibiting non-lethal indicators of excess toxicity should be euthanatised prior to completion of the test period (23).

Tissue Collection

35. The rationale for tissue collection should be defined clearly. Since it is possible to study mutation induction in virtually any tissue, the selection of tissues to be collected should be based upon the reason for conducting the study and any existing mutagenicity, carcinogenicity or toxicity data for the chemical under investigation. Important factors for consideration should include the route of administration (based on likely human exposure route(s)), the predicted tissue distribution, and the possible mechanism of action. In the absence of any background information, several somatic tissues as may be of interest should be collected. These should represent rapidly proliferating, slowly proliferating and site of contact tissues. In addition, spermatozoa from the vas deferens/cauda epididymis and developing germ cells from the seminiferous tubules (as described in paragraphs 32 and 33) should be collected and stored in case future analysis of germ cell mutagenicity is required. Organ weights should be obtained, and for larger organs, the same area should be collected from all animals.

Storage of Tissues and DNA

36. Tissues (or tissue homogenates) should be stored at or below −70°C and be used for DNA isolation within 5 years. Isolated DNA, stored refrigerated at 4 ºC in appropriate buffer, should be used optimally for mutation analysis within 1 year.

Selection of Tissues for Mutant Analysis

37. The choice of tissues should be based on considerations such as: 1) the route of administration or site of first contact (e.g. glandular stomach if administration is oral, lung if administration is through inhalation, or skin if topical application has been used); and 2) pharmacokinetic parameters observed in general toxicity studies, which indicate tissue disposition, retention or accumulation, or target organs for toxicity. If studies are conducted to follow up carcinogenicity studies, target tissues for carcinogenicity should be considered. The choice of tissues for analysis should maximise the detection of chemicals that are direct-acting *in vitro* mutagens, rapidly metabolised, highly reactive or poorly absorbed, or those for which the target tissue is determined by route of administration (6).

38. In the absence of background information and taking into consideration the site of contact due to route of administration, the liver and at least one rapidly dividing tissue (e.g. glandular stomach, bone marrow) should be evaluated for mutagenicity. In most cases, the above requirements can be achieved from analyses of two carefully selected tissues, but in some cases, three or more would be needed. If there are reasons to be specifically concerned about germ cell effects, including positive responses in somatic cells, germ cell tissues should be evaluated for mutations.

Methods of Measurement

39. Standard laboratory or published methods for the detection of mutants are available for the recommended transgenic models: *lacZ* lambda bacteriophage and plasmid (30); *lacI*

mouse (2) (18); *gpt* delta mouse (22); *gpt* delta rat (28); *cII* (17). Modifications should be justified and properly documented. Data from multiple packagings can be aggregated and used to reach an adequate number of plaques or colonies. However, the need for a large number of packaging reactions to reach the appropriate number of plaques may be an indication of poor DNA quality. In such cases, data should be considered cautiously because they may be unreliable. The optimal total number of plaques or colonies per DNA sample is governed by the statistical probability of detecting sufficient numbers of mutants at a given spontaneous mutant frequency. In general, a minimum of 125 000 to 300 000 plaques is required if the spontaneous mutant frequency is in the order of 3 x 10^{-5} (15). For the Big Blue® *lacI* assay, it is important to demonstrate that the whole range of mutant colour phenotypes can be detected by inclusion of appropriate colour controls concurrent with each plating. Tissues and the resulting samples (items) should be processed and analysed using a block design, where items from the vehicle/solvent control group, the positive control group (if used) or positive control DNA (where appropriate), and each treatment group are processed together.

DATA AND REPORTING

Treatment of Results

40. Individual animal data should be presented in tabular form. The experimental unit is the animal. The report should include the total number of plaque-forming units (pfu) or colony-forming units (cfu), the number of mutants, and the mutant frequency for each tissue from each animal. If there are multiple packaging/rescue reactions, the number of reactions per DNA sample should be reported. While data for each individual reaction should be retained, only the total pfu or cfu need be reported. Data on toxicity and clinical signs as per paragraph 34 should be reported. Any sequencing results should be presented for each mutant analysed, and resulting mutation frequency calculations for each animal and tissue should be shown.

Statistical Evaluation and Interpretation of Results

41. There are several criteria for determining a positive result, such as a dose-related increase in the mutant frequency, or a clear increase in the mutant frequency in a single dose group compared to the solvent/vehicle control group. At least three treated dose groups should be analysed in order to provide sufficient data for dose-response analysis. While biological relevance of the results should be the primary consideration, appropriate statistical methods may be used as an aid in evaluating the test results (4) (14) (15) (25) (26). Statistical tests used should consider the animal as the experimental unit.

42. A test chemical for which the results do not meet the above criteria in any tissue is considered non-mutagenic in this assay. For biological relevance of a negative result, tissue exposure should be confirmed.

43. For DNA sequencing analyses, a number of statistical approaches are available to assist in interpreting the results (1) (5) (9) (19).

44. Consideration of whether the observed values are within or outside of the historical control range can provide guidance when evaluating the biological significance of the

response (32).

Test report

45. The test report should include the following information:

Test chemical:

- identification data and CAS n°, if known;
- source, lot number if available;
- physical nature and purity;
- physiochemical properties relevant to the conduct of the study;
- stability of the test chemical, if known;

Solvent/vehicle:

- justification for choice of vehicle;
- solubility and stability of the test chemical in the solvent/vehicle, if known;
- preparation of dietary, drinking water or inhalation formulations;
- analytical determinations on formulations (e.g., stability, homogeneity, nominal concentrations);

Test animals:

- species/strain used and justification for the choice;
- number, age and sex of animals;
- source, housing conditions, diet, etc.;
- individual weight of the animals at the start of the test, including body weight range, mean and standard deviation for each group;

Test conditions:

- positive and negative (vehicle/solvent) control data;
- data from the range-finding study;
- rationale for dose level selection;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route of administration;
- methods for measurement of animal toxicity, including, where available, histopathological or haematological analyses and the frequency with which animal observations and body weights were taken;
- methods for verifying that the test chemical reached the target tissue, or general circulation, if negative results are obtained;
- actual dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentration (ppm) and consumption, if applicable;

- details of food and water quality;

- detailed description of treatment and sampling schedules and justifications for the choices;

- method of euthanasia;

- procedures for isolating and preserving tissues;

- methods for isolation of rodent genomic DNA, rescuing the transgene from genomic DNA, and transferring transgenic DNA to a bacterial host;

- source and lot numbers of all cells, kits and reagents (where applicable);

- methods for enumeration of mutants;

- methods for molecular analysis of mutants and use in correcting for clonality and/or calculating mutation frequencies, if applicable;

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;

- body and organ weights at sacrifice;

- for each tissue/animal, the number of mutants, number of plaques or colonies evaluated, mutant frequency;

- for each tissue/animal group, number of packaging reactions per DNA sample, total number of mutants, mean mutant frequency, standard deviation;

- dose-response relationship, where possible;

- for each tissue/animal, the number of independent mutants and mean mutation frequency, where molecular analysis of mutations was performed;

- concurrent and historical negative control data with ranges, means and standard deviations;

- concurrent positive control (or non-concurrent DNA positive control) data;

- analytical determinations, if available (e.g. DNA concentrations used in packaging, DNA sequencing data);

- statistical analyses and methods applied;

Discussion of the results

Conclusion

LITERATURE

(1) *Adams,* W.T. and T.R. *Skopek* (1987), "Statistical *Test* for the Comparison of Samples from Mutational Spectra", *J. Mol. Biol.*, 194: 391-396.

(2) Bielas, J.H. (2002), "A more Efficient Big Blue® Protocol Improves Transgene Rescue and Accuracy in an Adduct and Mutation Measurement", *Mutation Res.*, 518: 107–112.

(3) Boerrigter, M.E., M.E. Dollé, H.-J. Martus, J.A. Gossen and J. Vijg (1995), "Plasmid-based Transgenic Mouse Model for Studying *in vivo* Mutations" *Nature*, 377(6550): 657–659

(4) Carr, G.J. and N.J. Gorelick (1995), "Statistical Design and Analysis of Mutation Studies in Transgenic Mice*", Environ. Mol. Mutagen*, 25(3): 246–255.

(5) Carr, G.J. and N.J. Gorelick (1996), "Mutational Spectra in Transgenic Animal Research: Data Analysis and Study Design Based upon the Mutant or Mutation Frequency", *Environ. Mol. Mutagen,* 28: 405–413.

(6) Dean, S.W., T.M. Brooks, B. Burlinson, J. Mirsalis, B. Myhr, L. Recio and V. Thybaud (1999), "Transgenic Mouse Mutation Assay Systems can Play an important Role in Regulatory Mutagenicity Testing *in vivo* for the Detection of Site-of-contact Mutagens", *Mutagenesis*, 14(1): 141–151.

(7) Douglas, G.R., J. Jiao, J.D. Gingerich, J.A. Gossen and L.M. Soper (1995), "Temporal and Molecular Characteristics of Mutations Induced by Ethylnitrosourea in Germ Cells Isolated from Seminiferous Tubules and in Spermatozoa of *lacZ* Transgenic Mice", *Proc. Natl. Acad. Sci. USA*, 92: 7485-7489.

(8) Douglas, G.R.*,* J.D. Gingerich, L.M. Soper and J. Jiao (1997), "Toward an Understanding of the Use of Transgenic Mice for the Detection of Gene Mutations in Germ Cells", *Mutation Res.*, 388(2-3): 197-212.

(9) Dunson, D.B. and K.R. Tindall (2000), "Bayesian Analysis of Mutational Spectra", *Genetics*, 156: 1411–1418.

(10) Gossen, J.A., W.J. de Leeuw, C.H. Tan, E.C. Zwarthoff, F. Berends, P.H. Lohman, D.L. Knook and J. Vijg (1989), "Efficient Rescue of Integrated Shuttle Vectors from Transgenic Mice: a Model for Studying Mutations *in vivo"*, *Proc. Natl. Acad. Sci. USA,* 86(20): 7971–7975.

(11) Gossen, J.A. and J. Vijg (1993), "A Selective System for lacZ-Phage using a Galactose-sensitive *E. coli* Host", *Biotechniques*, 14(3): 326, 330.

(12) Erikson, R.P. (2003), "Somatic Gene Mutation and Human Disease other than Cancer", *Mutation Res.*, 543: 125-136.

(13) Erikson, R.P. (2010), "Somatic Gene Mutation and Human Disease other than Cancer: an Update", *Mutation Res.*, **705: 96-106**.

(14) Fung, K.Y., G.R. Douglas and D. Krewski (1998), "Statistical Analysis of *lacZ* Mutant Frequency Data from Muta^{TM}Mouse Mutagenicity Assays", *Mutagenesis*, 13(3): 249–255.

(15) Heddle, J.A.*, S. Dean, T. Nohmi, M. Boerrigter, D. Casciano, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.C. Mirsalis, H.-J Martus, T.R. Skopek, V. Thybaud, K.R.Tindall and N. Yajima (2000), "In vivo Transgenic Mutation Assays", *Environ. Mol. Mutagen.*, 35: 253-259.

(16) Heddle, J.A., H.-J. Martus and G.R. Douglas (2003), "Treatment and Sampling Protocols for Transgenic Mutation Assays", *Environ. Mol. Mutagen.*, 41: 1-6.

(17) Jakubczak, J.L., G. Merlino, J.E. French, W.J. Muller, B. Paul, S. Adhya and S. Garges (1996), "Analysis of Genetic Instability during Mammary Tumor Progression using a novel Selection-based Assay for *in vivo* Mutations in a Bacteriophage λ Transgene Target", *Proc. Natl. Acad. Sci. USA*, 93(17): 9073–9078.

(18) Kohler, S.W., G.S. Provost, P.L. Kretz, A. Fieck, J.A. Sorge and J.M. Short (1990), "The Use of Transgenic Mice for Short-term, in vivo Mutagenicity Testing", Genet. Anal. Tech. Appl., 7(8): 212–218.

(19) Lewis P.D., B. Manshian, M.N. Routledge, G.B. Scott and P.A. Burns (2008), "Comparison of Induced and Cancer-associated Mutational Spectra using Multivariate Data Analysis", Carcinogenesis, 29(4): 772-778.

(20) Nohmi, T., M. Katoh, H. Suzuki, M. Matsui, M. Yamada, M. Watanabe, M. Suzuki, N. Horiya, O. Ueda, T. Shibuya, H. Ikeda and T. Sofuni (1996), "A new Transgenic Mouse Mutagenesis Test System using Spi− and 6-thioguanine Selections", Environ. Mol. Mutagen., 28(4): 465–470.

(21) Nohmi, T., M. Suzuki, K. Masumura, M. Yamada, K. Matsui, O. Ueda, H. Suzuki, M. Katoh, H. Ikeda and T. Sofuni (1999), "Spi⁻ Selection: an Efficient Method to Detect γ-ray-induced Deletions in Transgenic Mice", *Environ. Mol. Mutagen.*, 34(1): 9–15.

(22) Nohmi, T., T. Suzuki and K.I. Masumura (2000), "Recent Advances in the Protocols of Transgenic Mouse Mutation Assays", *Mutation Res.*, 455(1–2): 191–215.

(23) OECD (2000), Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental

Animals Used in Safety Evaluation, Series on Testing and Assessment, N°19, ENV/JM/MONO(2000)7, OECD, Paris.

(24) OECD (2009), Detailed Review Paper on Transgenic Rodent Mutation Assays, Series on Testing and Assessment, N° 103, ENV/JM/MONO(2009)7, OECD, Paris.

(25) Piegorsch, W.W., B.H. Margolin, M.D. Shelby, A. Johnson, J.E. French, R.W. Tennant and K.R. Tindall (1995), "Study Design and Sample Sizes for a lacI Transgenic Mouse Mutation Assay", Environ. Mol. Mutagen., 25(3): 231–245.

(26) Piegorsch, W.W., A.C. Lockhart, G.J. Carr, B.H. Margolin, T. Brooks, ... G.R. Douglas, U.M. Liegibel, T. Suzuki, V. Thybaud, J.H. van Delft and N.J. Gorelick (1997), "Sources of Variability in Data from a Positive Selection lacZ Transgenic Mouse Mutation Assay: an Interlaboratory Study", Mutation. Res., 388(2–3): 249–289.

(27) Singer, T.M., I.B. Lambert, A. Williams, G.R. Douglas and C.L. Yauk (2006), "Detection of Induced Male Germline Mutation: Correlations and Comparisons between Traditional Germline Mutation Assays, Transgenic Rodent Assays and Expanded Simple Tandem Repeat Instability Assays", Mutation. Res., 598: 164-193.

(28) Toyoda-Hokaiwado, N., T. Inoue, K. Masumura, H. Hayashi, Y. Kawamura, Y. Kurata, M. Takamune, M. Yamada, H. Sanada, T. Umemura, A. Nishikawa and T. Nohmi (2010), "Integration of in vivo Genotoxicity and Short-term Carcinogenicity Assays using F344 gpt delta Transgenic Rats: in vivo Mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene Structural Isomers", Toxicol. Sci., 114(1): 71-78.

(29) Thybaud, V., S. Dean, T. Nohmi, J. de Boer, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.A. Heddle, R.H. Heflich, I. Lambert, H.-J. Martus, J.C. Mirsalis, T. Suzuki and N. Yajima (2003), "*In vivo* Transgenic Mutation Assays", *Mutation Res.*, 540: 141-151.

(30) Vijg, J. and G.R. Douglas (1996), "Bacteriophage λ and Plasmid *lacZ* Transgenic Mice for studying Mutations *in vivo*" in: G. Pfeifer (ed.), *Technologies for Detection of DNA Damage and Mutations, Part II*, Plenum Press, New York, NY, USA, pp. 391–410.

(31) Yauk, C.L., J.D. Gingerich, L. Soper, A. MacMahon, W.G. Foster and G.R. Douglas (2005), "A *lacZ* Transgenic Mouse Assay for the Detection of Mutations in Follicular Granulosa Cells", *Mutation Res.*, 578(1-2): 117-123.

(32) Hayashi, M., K. Dearfield, P. Kasper, D. Lovell, H.-J. Martus, V. Thybaud (2011), "Compilation and Use of Genetic Toxicity Historical Control Data", Mutation Res., doi:10.1016/j.mrgentox.2010.09.007.

(33) OECD (2011), *Retrospective Performance Assessment of OECD Test Guideline on Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays*, Series on Testing and Assessment, N° 145, ENV/JM/MONO(2011)20, OECD, Paris.

(34) Clermont, Y. (1972), "Kinetics of spermatogenesis in mammals seminiferous epithelium cycle and spermatogonial renewal". Physiol. Rev. 52: 198-236.

(35) Robaire, B., Hinton, B.T., and Oregbin-Crist, M.-C. (2006), "The Epididymis", in Neil, J.D., Pfaff, D.W., Chalis, J.R.G., de Kretser, D.M., Richards, J.S., and P. M, Wassarman (eds.), Physiology of Reproduction, Elsevier, the Netherlands, pp. 1071-1148.

(36) Russell, L.B. (2004), "Effects of male germ-cell stage on the frequency, nature, and spectrum of induced specific-locus mutations in the mouse", Genetica, 122: 25–36.

DEFINITIONS

Administration period: the total period during which an animal is dosed.

Base pair substitution: a type of mutation that causes the replacement of a single DNA nucleotide base with another DNA nucleotide base.

Capsid: the protein shell that surrounds a virus particle.

Chemical: a substance or a mixture.

Clonal expansion: the production of many cells from a single (mutant) cell.

Colony-forming unit (cfu): a measure of viable bacterial numbers.

Concatamer: a long continuous biomolecule composed of multiple identical copies linked in series.

Cos site: a 12-nucleotide segment of single-stranded DNA that exists at both ends of the bacteriophage lambda's double-stranded genome.

Deletion: a mutation in which one or more (sequential) nucleotides is lost by the genome.

Electroporation: the application of electric pulses to increase the permeability of cell membranes.

Endogenous gene: a gene native to the genome.

Extrabinomial variation: greater variability in repeat estimates of a population proportion than would be expected if the population had a binomial distribution.

Frameshift mutation: a genetic mutation caused by insertions or deletions of a number of nucleotides that is not evenly divisible by three within a DNA sequence that codes for a protein/peptide.

Insertion: the addition of one or more nucleotide base pairs into a DNA sequence.

Jackpot: a large number of mutants that arose through clonal expansion from a single mutation.

Large deletions: deletions in DNA of more than several kilobases (which are effectively detected with the Spi⁻ selection and the lacZ plasmid assays).

Ligation: the covalent linking of two ends of DNA molecules using DNA ligase.

Mitogen: a chemical that stimulates a cell to commence cell division, triggering mitosis (i.e. cell division).

Neutral gene: a gene that is not affected by positive or negative selective pressures.

Packaging: the synthesis of infective phage particles from a preparation of phage capsid and tail proteins and a concatamer of phage DNA molecules. Commonly used to package DNA cloned onto a lambda vector (separated by cos sites) into infectious lambda particles.

Packaging efficiency: the efficiency with which packaged bacteriophages are recovered in host bacteria.

Plaque forming unit (pfu): a measure of viable bacteriophage numbers.

Point mutation: a general term for a mutation affecting only a small sequence of DNA including small insertions, deletions, and base pair substitutions.

Positive selection: a method that permits only mutants to survive.

Reporter gene: a gene whose mutant gene product is easily detected.

Sampling time: the end of the period of time, prior to sacrifice, during which the chemical is not administered and during which unprocessed DNA lesions are fixed into stable mutations.

Shuttle vector: a vector constructed so that it can propagate in two different host species; accordingly, DNA inserted into a shuttle vector can be tested or manipulated in two different cell types or two different organisms.

Test chemical: Any substance or mixture tested using this test method.

Transgenic: of, relating to, or being an organism whose genome has been altered by the transfer of a gene or genes from another species."