Test Method for Assessment of Pubertal Development and Thyroid Function in Juvenile Male Rats as a Potential Screen in the Endocrine Disruptor Screening Program Tier-1 Battery

I. Purpose and Applicability

The purpose of this protocol is to quantify the effects of chemicals on pubertal development and thyroid function in the intact juvenile/peripubertal male rat. This assay detects chemicals that display anti-thyroid, or androgenic/anti-androgenic activity (e.g., alterations in receptor binding or steroidogenesis), or alter hypothalamic function or gonadotropin or prolactin secretion.

II. Endpoints

- Growth (daily body weight)
- Age and body weight at preputial separation
- Organ weights:
  - Seminal vesicle plus coagulating glands (with and without fluid)
  - Ventral prostate
  - Dorsolateral prostate
  - Levator ani/bulbocavernosus muscle complex
  - Epididymides (left and right separately)
  - Testes (left and right separately)
  - Thyroid
  - Liver
  - Kidneys (paired)
  - Pituitary
  - Adrenals (paired)
- Histology:
  - Epididymis
  - Testis
  - Thyroid (colloid area and follicular cell height)
  - Kidney
- Hormones:
  - Serum testosterone, total
  - Serum thyroxine (T_4), total
  - Serum thyroid stimulating hormone (TSH)

III. General Conditions

Rats are housed in clear plastic cages (approximately 20 x 25 x 47 cm) with heat-treated[1] laboratory-grade pine shavings as bedding. Corn cob bedding should not be used due to its potential to disrupt endocrine activity[2]. Wire-mesh-bottomed caging should not be used due to the potential for pup loss.

Prior to the onset of the study, pregnant female rats are housed individually. At weaning, pups are housed in groups of either two or three males of the same treatment group per cage.

Animals are maintained on a balanced laboratory diet[3] and water[4] *ad libitum*, in a room with a 14:10 hour light:dark photoperiod (on at 0500 hours, off at 1900 hours local time), temperature of 20 to 24° C, and a relative humidity of 40 to 50%. Small variations of this portion of the protocol except for the light cycle are permissible but must be documented[5].

IV. Animals: Juvenile Male Rats

The Sprague-Dawley or Wistar strains of rats are the preferred strains for this assay until a more-appropriate strain (or set of strains) is identified and associated performance criteria developed. Results similar to those from Sprague-Dawley rats have been produced using Wistar and Long-Evans rats in this assay or relevant modifications of this assay, suggesting that strain is not the major determinant of sensitivity in this assay.

Juvenile male rats are derived from individually housed pregnant females that were bred in-house or purchased from a supplier as "timed pregnant" dams. All dams must be pregnant for the first time and timed to deliver on the same day. If purchased from a supplier, all dams should be on the same gestation day (GD) but that GD may be GD 7, 8, 9, or 10 at the time of arrival at the performing laboratory (where GD 0 = day of sperm positive). Dams are allowed to deliver their pups naturally. Any litters with fewer

---

[1] to eliminate resins that induce liver enzymes

[2] Markaverich BM, Alejandro MA, Markaverich D, et al. 2002. Biochem Biophys Res Commun 291(3):692-700.

[3] N.B.: Totally synthetic diets are not appropriate, genistein-equivalent content of genistein plus daidzein (aglycone forms) of each batch must be less than or equal to 300 ug/g, and the same batch of feed must be used for treated and control groups at all times. ("Genistein-equivalent content" of daidzein is approximately 0.8. Owens WB, Ashby J, Odum J, Onyon L. 2003. Environ Health Perspect 111(12):1559-1567.)

[4] Deionized water is preferred. Tap water is not acceptable.

[5] EPA recognizes that the temperature and humidity ranges specified are significantly narrower than specified in guidelines published by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). However, EPA regards the AAALAC guidelines as exceptionally wide and the extremes of that range only marginally tolerable for the pubertal assay. Efforts should be made to keep the conditions within the range specified above.

than 8 total pups/litter (i.e., including both males and females) and any litters not delivered by GD 23 are excluded from the study. To maximize uniformity in growth rates, the litters are standardized to 8 to 10 pups per litter between post-natal days (PND) 3 and 5. (PND 0 is defined as the day on which the pup is first seen, assuming that the cages are checked for new births daily, in the morning.) Reducing litter size to 8-to-10 is required when dams have more than enough pups, but cross-fostering to raise litter size to the required number is not acceptable. Body weights are monitored weekly and any unthrifty litters or runted pups excluded from the study. Enough litters should be available to assure that a sufficient number of juvenile males are available for 15 male pups per treatment group. (If parallel male and female pubertal studies are being conducted, males and females from the same litter can be used in their respective studies.)

The pups are weaned on PND 21. Also on day 21, male pups are marked by litter, and then all of the male pups from all of the litters are weighed individually to the nearest 0.1 g and ranked by body weight. A population of male pups that is as homogeneous as possible is selected for the study by eliminating an equal number of pups from the heavy end and the light end of the distribution, leaving the number of animals needed for the study in the middle. In this way, one nuisance variable (viz., body weight at weaning) is experimentally controlled. The male pups are assigned to treatment groups such that the mean body weights and variances for all groups are similar. Avoid placing littermates in the same group.

As described in more detail later in this document, the preferred procedure is to kill all the animals on a single day to close the in-life portion of the study, but kills may be conducted over two days rather than one, with half of each group killed on each day. If kills over two days are planned, the kill days are assigned at the time of initial distribution to treatment groups and the body weights are distributed across kill groups such that the mean body weights and variances for all groups are similar, and, when possible, litter mates are not in the same group.

After assignment to treatment groups, male offspring belonging to the same treatment group are housed in groups of either 2 or 3 per cage, such that each cage has the same number of animals. (In the case of housing two per cage and a planned N of 15, it will be necessary to add an additional rat to the last cage.)

V. Experimental Design

This protocol uses a randomized complete block design (time-separated necropsy is the blocking factor) with fifteen male rats in each treatment group. The treatment conditions are (1) vehicle-treated; and (2) xenobiotic-treated (two dose levels). The highest dose level should be at or just below the Maximum Tolerated Dose (MTD) level but need not exceed the limit dose of 1 g/kg/day. A dose level will be considered to be at or just below the Maximum Tolerated Dose level if it causes a statistically significant reduction in terminal body weight in treated animals vs. controls, the reduction is no

greater than approximately 10% of the mean for the controls, and no clinical signs of toxicity associated with the dose level are observed throughout the study. In addition, abnormal blood chemistry values at termination (particularly creatinine and blood urea nitrogen (BUN)) may indicate that MTD was exceeded, even in the absence of a reduction in terminal body weight compared to controls. Histopathology of the kidney (or any other organ where gross observations indicate damage) may be used as evidence that MTD was exceeded. Blood chemistry and histopathology of the kidney are not required, however.

The second dose level should be one half of the highest dose level being tested unless justification is provided for testing at a different level.

If necessary, the study can be conducted in time-separated blocks rather than at one time. In this case, each block should contain all treatment groups and be balanced with respect to numbers of animals and body weight at weaning.

VI. Test substance

Chemical purity and stability in vehicle must be known prior to testing so that dose levels are correctly prepared.

Corn oil is the preferred vehicle for all treatment groups. If corn oil is used, it must be clear and free of sediment. It should have a bland odor, free from rancid, musty, metallic, putrid or any other undesirable odor. Other solvents such as water or carboxymethylcellulose may be used where appropriate. If the test substance is not soluble in any of the conventional solvents, it is administered as a suspension in corn oil. It is important that the dosing solution or suspension be well-mixed to keep the chemical well-distributed prior to and throughout dosing, and care must be taken to ensure that the particle size of insoluble substances does not interfere with delivery of the full dose through the gavage tube or needle tip.

VII. Treatment

Each animal is weighed daily, prior to treatment, and the body weight recorded. Clinical observations are also recorded daily. Animals which are found dead or which must be euthanized *in extremis* are removed from the cage. Endpoint measures (organ weights, hormone levels, histology, etc.) are not taken from these animals.

Treatments are administered daily by oral gavage from PND 23 through PND 53. This duration of treatment is unnecessary to detect androgenic chemicals, but is required for the detection of pubertal delay and antithyroid effects. Test chemicals are administered in 2.5 to 5.0[6] ml corn oil/kg body weight at 0700-0900 daily using an 18 gauge gavage

[6] Dosage volume per kg body weight must be the same for all treated animals in the experiment but the

needle (1 to 1½ inch length, 2.25 mm ball) and a 1 cc (disposable) tuberculin syringe for each treatment. Needle gauge may be optimized to animal size but must be constructed of metal to avoid the potential for absorption by or leaching of substances from rubber or plastic tubing. The treatments are administered on a mg/kg body weight basis using the current day's weight, and volume of the dose administered is recorded each day.

In the absence of other clinical signs that would normally lead to removal of an animal from the study, failure to gain weight at the same rate as controls is generally not a reason to remove a treated animal during the course of the study. However, it is recognized that severe failure to grow may be a reason to disqualify an animal even in the absence of other signs of toxicity. As general guidance, EPA suggests that a reduction in body weight when compared to controls of more than 20% in the absence of other signs of toxicity *may* justify removal.

VIII. Preputial Separation

Beginning on PND 30, males are examined daily for preputial separation (PPS). The appearance of partial and complete preputial separation, or a persistent thread of tissue between the glans and prepuce are all recorded on the days they are observed. The day of complete preputial separation is the endpoint used in the analysis for the age at preputial separation. However, if any animal within any treatment group shows incomplete separation (including persistent threads) for greater than three days, a separate analysis is conducted using the ages at which partial separation was first observed. Documentation of a thread even if PPS otherwise appears complete is important. It is also critical that "initiation" of PPS be recorded. It is preferred but not critical that PPS observations be taken after the daily dosing. Whether collected before or after dosing, the PPS observations must be collected at approximately the same time each day.

IX. Necropsy

Males are killed on PND 53. If necessary, one half of the males may be killed on PND 53 and the remaining males on PND 54 as long as the animals in each treatment group are equally dispersed between the two necropsy days. Animals killed on PND 54 are dosed and treated on the day of kill just like the animals killed on PND 53 with regard to time of dosing, collection of PPS information, etc. All animals are dosed between 0700 and 0900 hours local time, and killed beginning 2 hours following dosing. It is critical that kills are completed by 1300 hours due to normal diurnal fluctuation in thyroid hormone levels.

---

value chosen for the study may be anywhere in the specified range.

The preferred method of kill is by decapitation without any form of anesthesia, to minimize the potential for release of testosterone during CO_2 asphyxiation or interference by other anesthetics. Decapitation is also considered more humane than CO_2 asphyxiation. If CO_2 is used, it must be given for no more than 60 seconds prior to decapitation, even if the animal has not fully succumbed in that time. Decapitation has generally not be found to interfere with the integrity of the thyroid, which must be maintained in order to obtain thyroid weight and histology sections.

The order of necropsy is randomized or otherwise evenly distributed across *all* groups being necropsied that day. That is, do *not* necropsy all animals in one group before moving to the next group. When two or more test chemicals use the same control group, it is particularly important to intersperse the control animal necropsies across the entire time span in which all of the necropsies for all the test chemicals and dose levels are conducted.

Blood from the trunk of the animal is collected immediately (e.g., by inversion over a funnel)[7]. After collection, the blood is centrifuged. The serum is pipetted into siliconized microcentrifuge tubes[8] and stored at -20° C or colder for subsequent thyroxine (T_4), TSH, and testosterone measurements.

At necropsy, the testes, epididymides, ventral prostate, dorsolateral prostate, seminal vesicle with coagulating glands and fluid, levator ani plus bulbocavernosus muscles, thyroid (with attached portion of trachea), liver, kidneys, pituitary, and adrenals are removed and the weights of each except the thyroid/trachea recorded in milligrams to one decimal place with the exception of kidney and liver, which are recorded in grams, to two decimal places. (Kidneys and adrenals are weighed as pairs. The left and right testes, and the left and right epididymides, are weighed individually.) Care must be taken to remove mesenteric fat with small surgical scissors from these tissues such that the fluid in the sex accessory glands is retained. After weighing the seminal vesicle with coagulating glands and fluid, the fluid is removed and the seminal vesicle weight with coagulating glands but without fluid is also recorded. Small tissues such as the adrenals and pituitary, as well as tissues that contain fluid, should be weighed immediately to prevent tissues from drying out prior to weighing. Measures to prevent drying out may be necessary if such organs cannot be weighed immediately.

The thyroid (with attached portion of the trachea) and a single testis and epididymis from each animal are prepared for histological examination. Either the left or the right testis/epididymis may be chosen for a study but the choice must be applied consistently to all animals in the study and the choice must be reported. The testis and epididymis should both be from the same side but if that is not possible, use the other of the pair

[7] Heart puncture after no more than 60 seconds of CO_2 administration is an alternative method of obtaining blood, but is less preferred due to the inhumane stress of asphyxiation, the potential for testosterone release, and the smaller volume of blood collected.

[8] If there is a greater volume of blood than will fit in one microcentrifuge tube, prepare as many separate aliquots as appropriate before freezing. Do not freeze in large aliquots, to avoid excessive freeze/thaw cycles.

and note the reason. (The testis and epididymis not chosen for histology are discarded after weighing.) Testis and epididymis are placed in Bouin's fixative overnight (no longer than 24 hours)[9], after which they are rinsed and stored in 70% ethanol until embedded in paraffin. They are then stained with hematoxylin and eosin (H&E) for subsequent histological evaluations. The thyroid, with attached trachea, is fixed in 10% buffered formalin for at least 24 hours. Then the thyroid (with parathyroids) is dissected from the trachea, blotted and weighed to the nearest 0.01 mg, placed in 70% ethanol until embedded in paraffin, stained with H&E, and histologically evaluated. Kidney, like thyroid, is fixed in 10% buffered formalin for at least 24 hours, then placed in 70% ethanol until embedded in paraffin, stained with H&E, and histologically evaluated.

X. Hormonal Assays

Hormonal measurements can be conducted using radioimmunoassay (RIA), immunoradiometric assay (IRMA), enzyme-linked immunosorbent assay (ELISA), or time-resolved immunofluorescent procedures. Regardless of which is used, always include multiple quality control (QC) samples run in duplicates that are dispersed among the test samples. Any measurement kit that is used must be shown to yield appropriate values *for control rats at the laboratory performing the pubertal assay*. The lab's criteria for evaluating the kit's performance must be included in the study report. If the laboratory has never had experience with the kit for making measurements specifically in the rat, it might be appropriate to test the kit in one or more untreated rats outside of the pubertal assay before relying on it for the full study.

XI. Histology

Testis, epididymis, thyroid, and kidney are evaluated for pathologic abnormalities and potential treatment-related effects.

Thyroid sections are subjectively evaluated for follicular cell height and colloid area using a five point grading scale (1 = shortest/smallest; 5 = tallest/largest)[10] and any abnormalities/lesions noted. A minimum of two sections of each of the two lobes of the thyroid are evaluated. Example photomicrographs are attached. The examples illustrate the magnitude of differences that are typically evaluated as separate scores, but the reader will need to establish the range appropriate for the particular study being evaluated.

[9] Fixing the testis and epididymis in 10% buffered formalin for at least 24 hours may be used in place of fixing in Bouin's solution for less than 24 hours, but Bouin's is preferred due to the relative ease of handling that results.

[10] see Capen CC, Martin SL. 1989. The effects of xenobiotics on the structure and function of thyroid follicular and C-cells. Toxicol Pathol 17(2):266-93.

Guidance for histological evaluation of the testis and epididymis is given in EPA's Health Effects Test Guideline OPPTS 870.3800: Reproduction and Fertility Effects[11]: "Besides gross lesions such as atrophy or tumors, testicular histopathological examination should be conducted in order to identify treatment-related effects such as retained spermatids, missing germ cell layers or types, multinucleated giant cells, or sloughing of spermatogenic cells into the lumen .... Examination of the intact epididymis should include the caput, corpus, and cauda, which can be accomplished by evaluation of a longitudinal section, and should be conducted in order to identify such lesions as sperm granulomas, leukocytic infiltration (inflammation), aberrant cell types within the lumen, or the absence of clear cells in the cauda epididymal epithelium."

XII. Statistical Analysis

Consideration should be given to whether there are any data points that should be excluded from the data set, and whether any data points that are identified as outliers by an appropriate statistical test should actually not be excluded, based on toxicological judgment. Values due to obvious technical errors are excluded. Justification for exclusion of each data point must be given. Outliers must be specified in the raw data. Do not test incidence data, e.g., from histopathology evaluation, for outliers.

All data except histology evaluation (i.e., initial body weight [PND 23], body weight gain[12], age and body weight at preputial separation, body and organ weights at necropsy, and serum hormones) are analyzed by Analysis of Variance (ANOVA). If the study was conducted in blocks, then the analysis is a two-way ANOVA with Block and Treatment as main effects. Age and body weight at PPS, and all organ weights should also be analyzed by Analysis of Covariance (ANCOVA) using the body weight at PND 21 as the covariate[13]. When statistically significant effects are observed ($p < 0.05$), treatment means are examined further using appropriate pairwise comparison tests to compare the control with each dose group[14]. Where there is heterogeneity of variance, data should be transformed appropriately prior to ANOVA/ANCOVA, or analyzed using an appropriate nonparametric test. Non-parametric analysis should be the method of last resort since it does not allow analysis of covariation. In addition to ANOVA and ANCOVA, examine the unadjusted and adjusted values for linear trend with dose level.

---

[11] USEPA, Office of Prevention, Pesticides, and Toxic Substances. August 1998. Health Effects Test Guidelines 870.3800: Reproduction and fertility effects. EPA 712-C-98-208.

[12] Use the body weights on the last day all the animals were weighed. Specifically, if kills were performed over two days, do not use the day when only the last half of the animals were available.

[13] The covariate is body weight on the day of weaning rather than on the day of kill because ANCOVA assumes that the covariate that is being adjusted for is not affected by the treatment, whereas in this assay endocrine-active substances may affect the overall body weight gain and thus body weight at kill. Using body weight at kill as covariate could mask a potentially endocrine effect on an organ. The Agency understands that using body weight at kill as covariate might identify which organs are more sensitive (or less sensitive) than body weight gain to potentially endocrine effects, but has chosen to maximize the potential of identifying organ-specific effects rather than relying on bodyweight as an indicator of potential endocrine activity.

[14] Comparison of the means of dose groups to each other is not required.

In cases where preputial separation (PPS) has not occurred prior to necropsy, use the last day of observation +1 as the age at PPS when determining the mean for each group. For example, if the animal was killed on PND 53 without PPS, use PND 54 as the value for that animal when determining the mean for the treatment group.

When possible, appropriate statistical analysis should be applied to the histology results.

XIII. Data Summary

Provide the following figure and tables for each test chemical along with the respective control: **Be sure to use the units shown in the example tables.** Provide values to one decimal place for the organs reported in milligrams, and to two decimal places for those organs that are required to be reported in grams.

Figure 1- Graph the mean body weight +/- Standard Deviation (SD) for each day during dosing for each treatment group, including vehicle control. (If animals were necropsied over 2 days, do not include the body weight from the last day of necropsy since only half of the animals are available.) Place an arrow at the mean age of the controls at PPS. All three groups (control, dose level 1, and dose level 2) are plotted on the same graph but are distinguished from each other by point and line styles.

Table 1 - Preputial separation; general growth
Report the mean, standard deviation, coefficient of variation, number of animals (N), and p-value, for the following endpoints, for each test-chemical dose group and control, both unadjusted (U) and adjusted (A) for body weight on PND 21:

a) age at PPS
b) body weight at the age of PPS
c) initial body weight (PND 23)
d) body weight on the last day all the animals were weighed (i.e., if kills were performed over two days, do not use the day when only the last half of the animals were available)
e) final body weight as percent of control (leave control column blank)
f) body weight gain from first dose to the (first) day of necropsy.

Mark endpoints that show an effect (by ANOVA/ANCOVA or a non-parametric test, as appropriate) with an asterisk in the "Effect" column. List the transformation (if any) used to eliminate heterogeneity of variance, or the non-parametric test used, in the "Transform or nonparam" column. Name the pairwise test used to compare the means of dosed groups to the mean of controls in the "Pairwise test" column.

Mark means that are significantly different from control means ($p < 0.05$) by shading the cell (rather than by the traditional asterisk).

Show the proportion of animals in which PPS had not occurred by the time of necropsy (e.g., X/15), and explain that age-at-necropsy-plus-one (e.g., 54) was used for those animals when calculating the mean.

Table 1

Chemical Name		Effect	Transform or nonparam	Pairwise test	Vehicle Control				(Dose Level 1)					(Dose Level 2)				
					Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Age at PPS (PND)	U																	
	A																	
Body weight at PPS (g)	U																	
	A																	
Initial body weight (PND 23, g)	U																	
	A																	
Final body weight (g)	U																	
	A																	
Final body weight (% of control)	U																	
	A																	
Body weight gain (final minus initial body weight) (g)	U																	
	A																	
Proportion unseparated (#/N)																		

*Means different from controls at p<0.05 are marked by a shaded cell.

Table 2- Organ weights at necropsy:
Report the mean, standard deviation, coefficient of variation, number of animals (N), and p-value for liver, kidneys, pituitary, adrenals, seminal vesicle plus coagulating glands (with and without fluid), ventral prostate, dorsolateral prostate, levator ani/bulbocavernosus muscle complex, epididymides (left and right separately), testes (left and right separately), and thyroid weights, for each treatment group, both unadjusted (U) and adjusted (A) for body weight on PND 21. Do not use relative organ to body weight ratios, and do not adjust for body weight at necropsy.

Table 3 - T_4, TSH, and Testosterone levels:
Report the mean, standard deviation, coefficient of variation, number of animals, and p-value for the T_4, TSH, and testosterone levels, for each treatment group, including vehicle control.

Prepare an executive summary describing the number and strain of rats used in the study, the dose levels and chemicals tested, and the effects with levels of statistical significance for all endpoints except histology. Include a summary of the histological findings.

Electronic and hard copies of spreadsheets containing the raw data from all animals for each endpoint are to be submitted to the EPA. The full report of the histological findings with photomicrographs of significant observations must also be provided.

Table 2

Chemical name					Vehicle Control				(Dose Level 1)					(Dose Level 2)				
Organ weights		Effect	Transform or nonparam	Pairwise test	Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Liver (g)	U																	
	A																	
Kidneys (g)	U																	
	A																	
Pituitary (mg)	U																	
	A																	
Adrenals (mg)	U																	
	A																	
Seminal vesicle + coagulating gland, with fluid (mg)	U																	
	A																	
Seminal vesicle + coagulating gland, without fluid (mg)	U																	
	A																	
Ventral prostate (mg)	U																	
	A																	
Dorsolateral prostate (mg)	U																	
	A																	
LABC (mg)	U																	
	A																	
Epididymis (left, mg)	U																	
	A																	

Epididymis (right, mg)	U A																	
Testis (left, mg)	U A																	
Testis (right, mg)	U A																	
Thyroid (mg)	U																	
	A																	

*Means different from controls at p<0.05 are marked by a shaded cell.

Table 3

Chemical name				Vehicle Control				(Dose Level 1)					(Dose Level 2)				
	Effect	Transform or nonparam	Pairwise test	Mean	SD	CV	N	Mean*	SD	CV	N	p-value	Mean*	SD	CV	N	p-value
Serum T_4, total (µg/dl)																	
Serum TSH, (ng/ml)																	
Serum testosterone, total (ng/ml)																	

*Means different from controls at p<0.05 are marked by a shaded cell.

XIV. Performance criteria

The following performance criteria have been established for the vehicle-control animals. See the Data Interpretation Procedure for use of the performance criteria. Units for the endpoints are shown in the table. Coefficients of variation (CVs) are in percent. The "mean", "2 SDs", "CV", and "1.5 CV" columns describe the mean, two standard deviations, coefficient of variation, and 1.5 times the coefficient of variation for that endpoint in historical controls. Mean values and CVs for the vehicle control group must fall in the acceptable range of each to be considered fully acceptable.

Endpoint	Rat strain	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
Ventral prostate (grams)							
	Wistar	0.223	0.072	0.151 to 0.295	16.67	5.65	22.32
	SD	0.246	0.086	0.160 to 0.332			
LABC (grams)							
	Wistar				15.77	11.33	27.100
	SD	0.651	0.204	0.447 to 0.855			
Epididymis (grams)							
	Wistar	0.474	0.124	0.350 to 0.598	10.94	5.45	16.39
	SD	0.446	0.082	0.364 to 0.528			
Seminal vesicle (grams)							
	Wistar	0.576	0.234	0.342 to 0.810	20.61	0.45	21.06
	SD	0.507	0.212	0.295 to 0.719			
Testis (grams)							
	Wistar	1.341	0.250	1.091 to 1.591	9.27	8.35	17.62
	SD						
T4 (ug/dl)							
	Wistar	5.478	2.164	3.314 to 7.642	18.27	9.20	27.46
	SD	5.716	1.660	4.056 to 7.376			
Thyroid weight (milligrams)							
	Wistar				15.39	8.24	23.63
	SD	20	6	14 to 26			
TSH (ng/ml)[15]							
	Wistar				34.04	24.26	58.29
	SD	14.162	9.950	4.212 to 24.112			
Age at PPS (postnatal day, where day of birth = PND 0)							
	Wistar	43.124	2.948	40.176 to 46.072	3.64	2.03	5.67
	SD	43.147	3.366	39.781 to 46.513			
Weight at PPS (grams)							
	Wistar	209.142	31.850	177.292 to 240.992	7.54	0.03	7.57

[15] TSH values were derived using the kit from Amersham.

Endpoint	Rat strain	Mean	2 SDs	Acceptable range	CV	1.5 CV	Top of acceptable range[a]
	SD	222.223	33.946	188.277 to 256.169			
Testosterone (ng/ml)							
	Wistar	2.118	2.540	0 to 4.658	58.82	30.88	89.70
	SD	2.110	1.850	0.260 to 3.960			
Final body weight (grams)							
	Wistar	291.818	41.578	250.24 to 333.396	6.62	0.85	7.47
	SD	295.647	36.412	259.235 to 332.059			
Adrenals (milligrams)							
	Wistar	54.597	13.768	40.829 to 68.365	15.42	7.34	22.77
	SD	46.478	14.636	31.842 to 61.114			
Kidneys (grams)							
	Wistar	2.516	0.550	1.966 to 3.066	9.56	5.20	14.76
	SD	2.646	0.404	2.242 to 3.050			
Liver (grams)							
	Wistar	14.070	2.874	11.196 to 16.944	10.24	4.69	14.93
	SD	12.670	2.680	9.990 to 15.350			
Pituitary (milligrams)							
	Wistar	8.051	1.934	6.117 to 9.985	12.14	3.83	15.98
	SD	10.354	2.544	7.810 to 12.898			
Weaning weight (grams)							
	Wistar	58.238	11.058	47.180 to 69.296	8.04	2.21	10.25
	SD	52.642	7.170	45.472 to 59.812			

[a] Bottom of the acceptable range for coefficient of variation is zero.

XV. Data Interpretation Procedure

The male pubertal assay is intended to be one of a suite of *in vitro* and *in vivo* assays for determining the potential of a substance to interact with the endocrine system. Therefore, it is important to emphasize that the data interpretation of a specific chemical will be a combination of the results from a number of these Tier-1 screening assays taken as a whole and not merely the sum of results of assays interpreted in isolation. That said, there are certain guidelines that can be given for interpreting data from a male pubertal assay.

First, the dose levels tested should be examined to see if a Maximum Tolerated Dose was used. Indications that MTD was approached but not exceeded include clinical observations and/or body weight loss compared to controls at termination that does not exceed approximately 10%. Histopathology of the kidney and/or other organs, and/or significant deviations from standard blood chemistry values may be indications that MTD was exceeded.

Negative results for interaction with the endocrine system in the pubertal assay will generally require demonstration that the highest dose level tested was at or near the MTD. Positive results in the assay generally require no such proof, but will generally require demonstration that interference due to body weight loss *per se* was not a factor in generating the results. Thus, studies which suggest thyroid activity only at a dose level causing more than approximately 6% body weight loss at termination compared to controls may need to be repeated at a lower dose level. Similarly, studies which suggest interaction with non-thyroid endocrine systems only at a dose level that causes more than approximately 10% body weight loss at termination compared to controls may need to be repeated.

The endpoint values for the control group should be compared to the performance criteria. Comparison should be made on the basis of the measured values, not adjusted values. Any endpoints which do not meet the performance criteria in controls will generally be disregarded for the test chemicals if they are negative but may provide useful information if they are positive.

Information that is missing due to inability to meet a performance criterion is not the same as a negative result. The more endpoints that are missing, the less likely the study will be regarded as adequate. No firm rules can be given for the minimum number of endpoints that must be available for evaluation since some of the endpoints are somewhat redundant (e.g., androgen-dependent tissue weights) while others are not (thyroid-related endpoints). In general, however, missing one or two performance criteria will not be regarded as fatal to the study.

More emphasis will be placed on meeting performance criteria for the coefficients of variation than for the endpoint control means. Laboratories may submit historical data for their own colonies to substantiate claims that tissue weights or other endpoints in the study being evaluated are in line with historical values of controls in that laboratory. Once the usable data set has been identified through application of the performance criteria, it is evaluated to see if there is evidence of interaction of the test chemical with the endocrine system.

Due to the covariance of certain organ weights with body weight, care should be taken in interpreting pituitary, liver, and kidney weight changes. Only if a change in the organ weight relative to body weight is significant for these particular organs (i.e., not all the organs) should the weights adjusted for covariance with body weight at weaning for these particular organs be interpreted as relevant.

Endpoints other than pituitary, liver, and kidney weights should not be evaluated based on their values relative to terminal body weight, nor should an analysis of covariance with terminal body weight be used for interpretation. Since endocrine-active agents themselves may have an effect on body weight, it is most appropriate to adjust for covariance with body weight at weaning, before chemical treatment began.

The male pubertal data provide general profiles of changes in the assay endpoints for various modes of action such as androgen agonism, androgen antagonism, alteration of steroidogenesis, thyroid toxicity, and interference with HPG function. These profiles can be used to establish a "weight of evidence" for general mechanisms of interaction of a test chemical with the endocrine system. For example, an antiandrogen such as vinclozolin delays puberty, impairs reproductive tract development (e.g., decreased VP, SV, LABC, epididymis weight) and increases testosterone at higher doses, so a test chemical with similar responses would likely be suspected of having an antiandrogenic interaction. A similar profile would be expected if the compound inhibits testosterone synthesis. One way to discern a compound that inhibits steroidogenesis from one that is antiandrogenic is to evaluate serum testosterone (a required endpoint) as this endpoint will obviously be decreased.

The pubertal male assay includes redundant androgen-dependent endpoints and in general, all would be expected to respond similarly. However, it is possible that a chemical may not alter all the endpoints measured, or the effect may not be dose dependent, or it may occur only at the high dose. In these kinds of cases, data from the other proposed assays in the Tier 1 battery would provide added insight into a potential effect.

Another example of how a chemical would produce a particular profile would be the way in which the thyroid homeostasis is disrupted. For example, the polybrominated diphenyl ether mixture, DE-71, disrupts thyroid hormones by inducing the clearance of thyroxin by hepatic enzyme induction. The ensuing profile of effects includes a decrease in T_4, decrease in T_3 (although not required in the male pubertal) and a subsequent increase in TSH. In many cases, thyroid weight and thyroid histology appear less sensitive than the changes in the hormone concentration because the hormonal changes can occur more rapidly than any frank change in histology.

Other chemicals may target the HPG axis and there are certain profiles that may indicate altered brain-pituitary function. For example, a chemical may delay puberty, lead reduced androgen dependent tissues, yet be negative in a androgen receptor binding assay, negative for alterations in steroidogenesis and the Hershberger.

It should be noted that consistency with a known "profile" is not a requirement for determining that a test chemical interacts with the endocrine system, nor is consistency among supposedly redundant endpoints.

In addition to the redundancy of the endpoints in the protocol to detect reproductive and thyroid effects, the requirement for a minimum of two doses provides an opportunity to examine the relationship between dose and response. If an endpoint is positive at the lower dose in just one endpoint and no effect is seen at the higher dose, then the effect and the overall conclusions would need to be questioned. Thus, the dose response informs the weight of evidence approach discussed above.

Endocrine profiles which can be identified in the male pubertal protocol

Androgen antagonist	Steroidogenesis inhibitor or HPG suppression	Hypothyroidism
↑Age of Puberty	↑ Age of Puberty	↓ T_4
↓Ventral prostate, seminal vesicles, LABC, epididymis	↓Ventral prostate, seminal vesicles, LABC, epididymis	↑TSH
↑Testosterone	↓Testosterone or no effect	↑Thyroid wt.
		↑Follicular cell height ↓Colloid area
		↑Liver wt (for compounds which induce hepatic clearance of thyroxine) or no effect