

Original Article

Key Learnings from the Endocrine Disruptor Screening Program (EDSP) Tier 1 Rodent Uterotrophic and Hershberger Assays

M. Sue Marty[1]* and John C. O'Connor[2]

[1]Toxicology & Environmental Research and Consulting, The Dow Chemical Company, Midland, Michigan
[2]DuPont Haskell Global Centers for Health and Environmental Sciences, Newark, Delware

In 2009, companies began screening compounds using the US Environmental Protection Agency's Endocrine Disruptor Screening Program (EDSP). EDSP has two tiers: Tier 1 includes 11 assays to identify compounds with potential endocrine activity. This article describes two laboratories' experiences conducting Tier 1 uterotrophic and Hershberger assays. The uterotrophic assay detects estrogen receptor agonists through increases in uterine weight. The advantages of the uterotrophic rat models (immature vs. adult ovariectomized) and exposure routes are discussed. Across 29 studies, relative differences in uterine weights in the vehicle control group and 17α-ethynylestradiol–positive control group were reasonably reproducible. The Hershberger assay detects androgen receptor (AR) agonists, antagonists, and 5α-reductase inhibitors through changes in accessory sex tissue (AST) weights. Across 23 studies, AST weights were relatively reproducible for the vehicle groups (baseline), testosterone propionate (TP) groups (androgenic response), and flutamide + TP groups (antiandrogenic response). In one laboratory, one and four compounds were positive in the androgenic and antiandrogenic portions of the assay, respectively. Each compound was also positive for AR binding. In the other laboratory, three compounds showed potential antiandrogenic activity, but each compound was negative for AR binding and did not fit the profile for 5α-reductase inhibition. These compounds induced hepatic enzymes that enhanced testosterone metabolism/clearance, resulting in lower testosterone and decreased capacity to maintain AST weights. The Hershberger androgenic and antiandrogenic performance criteria were generally attainable. Overall, the uterotrophic and Hershberger assays were easily adopted and function as described for EDSP screening, although the mode of action for positive results may not be easily determined. *Birth Defects Res (Part B)* 101:63–79, 2014. © 2014 The Authors. *Birth Defects Research (Part B)* published by Wiley Periodicals, Inc.

Key words: ***uterotrophic; Hershberger; endocrine disruptor; endocrine screening; EDSP; endocrine disruptor screening program; estrogen; androgen; antiandrogen; 5α-reductase inhibitor***

INTRODUCTION

Over the past two decades, there has been an increasing concern over the potential of environmental chemicals to cause effects on the endocrine system. In 1996, passage of the Food Quality Protection Act and an amendment to the Safe Drinking Water Act mandated the US Environmental Protection Agency (EPA) to establish a screening program to identify compounds that have the potential to interact with the endocrine system. The US EPA implemented the Endocrine Disruptor Screening Program (EDSP), a two-tiered system for evaluating endocrine activity. Tier 1 includes five in vitro and six in vivo assays and was designed to identify compounds with the potential to interact with the estrogen-, androgen-, or thyroid-signaling pathways. Tier 1 assays were selected to minimize "false-negative" results; therefore, a corresponding increase in "false-positive" findings was deemed acceptable. Tier 2 was designed to evaluate adverse effects from potential endocrine-active compounds identified in Tier 1, as well as generate dose–response data for use in risk assessment.

In 2009, the US EPA issued test orders requiring an initial list of 67 compounds to be evaluated in Tier 1 of the EDSP. The Tier 1 battery was designed to have some redundancy across assays to enhance its sensitivity and specificity, and to aid in the identification of endocrine

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Dow; Grant sponsor: DuPont.

*Correspondence to: M. Sue Marty, Toxicology & Environmental Research and Consulting, The Dow Chemical Company, Building 1803, Midland, MI 48674. E-mail: mmarty@dow.com

Received 27 November 2013; Accepted 8 January 2014

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/bdrb) DOI: 10.1002/bdrb.21098

modes of action (MoAs). Furthermore, this approach allows regulators to apply weight of evidence (WoE) to determine whether a compound has potential endocrine activity and thus identify compounds that will subsequently require Tier 2 testing.

This article describes the experiences of two laboratories conducting two of the in vivo mammalian assays that are included as part of the EDSP Tier 1, the rodent uterotrophic and Hershberger assays. The test systems used in these assays do not have functioning hypothalamic–pituitary–gonadal (HPG) axes, and therefore, are unable to compensate for changes in estrogen or androgen signaling. This makes the uterotrophic and Hershberger assays very sensitive for detecting compounds that bind to the estrogen receptor (ER) or androgen receptor (AR) interactions, respectively, or in the case of the antiandrogenic portion of the Hershberger assay, compounds that can interfere with testosterone binding to ARs. Correspondingly, these assays do not detect compounds that act directly on the hypothalamus or pituitary. However, the Hershberger assay is able to detect some compounds that act via non-receptor-mediated MoAs such as 5α-reductase inhibitors and will yield positive results with compounds that increase androgen metabolism (discussed in this article). This article briefly describes the rationale for each assay, implementation of these assays in accordance with test guideline requirements, technical aspects encountered in the two participating laboratories during the conduct of the assays (e.g., selecting dose levels, animal models, and so on), and assay interpretation. When appropriate, we have put our experiences in the context of findings reported during validation of these assays and in other scientific publications. Additional reviews on the uterotrophic and Hershberger assays are also available (e.g., Owens and Ashby, 2002; OECD, 2003, 2008).

ASSAY CONCEPTS AND CONDUCT

Rat Uterotrophic Assay

The rodent uterotrophic assay is a short-term, in vivo screening assay designed to detect compounds with potential estrogenic activity (i.e., ER agonists) by measuring a compound's ability to produce an increase in uterine weight. The premise of the assay is based on the transient changes in uterine weight that occur during the estrous cycle; that is, the increase (or decrease) in uterine weights in response to increases (or decreases) in endogenous estrogen levels (Reel et al., 1996; Owens and Ashby, 2002). The uterotrophic assay underwent an extensive validation program coordinated by the OECD (Kanno et al., 2001, 2003a,2003b; Owens et al., 2003; Owens and Koëter, 2003; Yamasaki et al., 2003a; Kim et al., 2005). The assay has been shown to reliably detect estrogenic activity across numerous laboratories using different routes of exposure in either immature rats or adult ovariectomized rats (e.g., Kanno et al., 2001).

Test guidelines (OECD, 2007; US EPA, 2009a) are available that describe the conduct, interpretation, and performance specifications for the uterotrophic assay. The guideline assay uses immature female rats or ovariectomized adult rats (or mice), which have low levels of endogenous estrogens (Reel et al., 1996) and therefore, low baseline uterine weights. If using ovariectomized females, complete ovariectomy must be confirmed before initiation of dosing by evaluating vaginal cytology for 5 days to confirm the absence of cycling. Rats should show cytologic evidence of diestrus during this time, indicating successful ovariectomy, and subsequently indicating basal levels of endogenous estrogens are too low to induce cycling. Animals (6/dose group) are administered the test compound, by either oral gavage or subcutaneous (sc) injection, daily for 3 days (Supplemental Fig. 1). A minimum of two test compound-treatment groups are required. If immature animals are used, dosing must be completed before postnatal day (PND) 25 to complete the assay before the initiation of puberty onset and the production of endogenous estrogen, which will decrease assay sensitivity (OECD, 2003). On test day (TD) 4, approximately 24 hr after the last dose, animals are examined for vaginal patency (if immature rats are used), weighed, and euthanized. The uteri are excised, trimmed, and wet (with fluid) and blotted uterine weights are recorded. If a loss of fluid is noted, the wet weight for that sample is excluded. For blotted weights, each uterine horn is nicked and blotted to remove luminal fluid (methods for collecting blotted weights have been described; OECD, 2003). If the ovariectomized adult model is used, the absence of ovarian remnants should be confirmed at necropsy, either by visualization at the time of necropsy, or alternatively by saving the uterine stumps for subsequent microscopic evaluation. Incomplete ovariectomy can lead to marked increases in uterine weights (Zacharewski et al., 1998). Alternatively, one could continue to collect vaginal cytology data for the duration of the assay (up through TD 4), which would not only help to confirm the absence of ovarian remnants, but can also help to identify estrogenic substances by the changes in vaginal cytology (i.e., progression of vaginal smears from diestrus to either proestrus or estrus) in the ovariectomized females.

Increases in uterine weights are typically due to interaction of a compound with ERα, which can result in uterine hypertrophy, hyperplasia, and fluid imbibition. Uterine weights from compound-treated animals are compared with uterine weights in the vehicle-treated control group; a compound that causes a statistically significant increase in wet and/or blotted uterine weights is considered "positive" for potential estrogenic activity.

Hershberger Assay

The rodent Hershberger assay is a short-term, in vivo screening assay designed to detect compounds with potential to act as AR agonists, antagonists, and 5α-reductase inhibitors. Both the OECD (2009a) and US EPA (2009b) have developed test guidelines for the Hershberger assay. To conduct the Hershberger assay (Supplemental Fig. 2), male rats are castrated at approximately 42 days of age and allowed a minimum of 7 days to recover from surgery (dosing may be initiated from PND 49 to 60). During this time, the accessory sex tissues (ASTs) regress as a result of the loss of gonadal androgen synthesis. The castrated male rats (6/dose group) receive test material for 10 days by gavage or sc injection in the presence or absence of testosterone propionate (TP); the study designs employed by the two laboratories are outlined in Tables 1

Table 1
Hershberger Assay Study Design Used in Laboratory A

Group no.	Test compound dose level (mg/kg/day)[a]	TP dose level (mg/kg/day)[b]	No. of rats[c]
Androgenic study design			
1	0 (vehicle control)	0 (no sc injection)	7
2	Low-dose test compound	0 (no sc injection)	7
3	Mid-dose test compound	0 (no sc injection)	7
4	High-dose test compound[d]	0 (no sc injection)	7
Antiandrogenic study design			
5	0 (vehicle control)	0.4	7
6	3.0 flutamide[e]	0.4	7
7	Low-dose test compound	0.4	7
8	Mid-dose test compound	0.4	7
9	High-dose test compound	0.4	7

[a]Test compound administered by oral gavage (4 ml/kg body weight) on TDs 1 to 10.
[b]TP administered by sc injection (0.5 ml/kg body weight) once daily on TDs 1 to 10 in corn oil vehicle.
[c]Only six animals per group are required according to the test guidelines (OPPTS 870.1400; OECD 441).
[d]Only two dose groups of the test substance are required per the test guidelines (OPPTS 870.1400; OECD 441). For EDSP studies, three dose levels were evaluated.
[e]Flutamide administered by oral gavage (4 ml/kg body weight) in corn oil.

Table 2
Hershberger Assay Study Design Used in Laboratory B

Group no.	Test compound dose level (mg/kg/day)[a]	TP dose level (mg/kg/day)[b]	No. of rats
Androgenic study design			
1	0 (vehicle control)	0 (no sc injection)	6
2	Low-dose test compound	0 (no sc injection)	6
3	Mid-dose test compound	0 (no sc injection)	6
4	High-dose test compound[c]	0 (no sc injection)	6
5	0 (no oral dosing)	0.4	6
Antiandrogenic study design			
6	0 (vehicle control)	0.4	6
7	3.0 flutamide[d]	0.4	6
8	Low-dose test compound	0.4	6
9	Mid-dose test compound	0.4	6
10	High-dose test compound	0.4	6

[a]Test compound administered by oral gavage (10 ml/kg body weight) once daily on TDs 1 to 10 in 0.1% Tween 80/0.5% methylcellulose vehicle.
[b]TP administered by sc injection (0.5 ml/kg body weight) once daily on TDs 1 to 10 in corn oil vehicle.
[c]Only two dose levels of the test compound are required per the test guidelines (OPPTS 870.1400; OECD 441). For EDSP studies, three dose levels were evaluated. For studies performed as part of internal discovery testing, either a single dose level or two dose levels per test compound were evaluated.
[d]Flutamide administered by oral gavage (4 ml/kg body weight) in corn oil.

and 2. Depending on the relevant route of exposure, the Hershberger assay is conducted using oral exposure if appropriate or sc exposure for dermal or inhalation exposures, with additional consideration given to toxicity by each route and the desire to avoid first-pass metabolism. The Hershberger assay requires a minimum of two test compound treatment groups for the androgenic portion of the assay and three test compound treatment groups for the antiandrogenic portion of the assay, in addition to the vehicle and positive control groups. Animals are euthanized approximately 24 hr after the final dose and organ weights are collected for the AST (i.e., ventral prostate, seminal vesicles with coagulating glands and fluid, levator ani-bulbocavernosus muscle [LABC], glans penis [if preputial separation {PPS} has occurred], Cowper's [bulbourethral] glands), as well as several optional organs (i.e., liver, kidneys, and adrenals). The maintenance of AST weights depends upon androgenic signals (i.e., typically, testosterone and dihydrotestosterone); therefore, the Hershberger assay detects chemicals that act as AR agonists, antagonists, or 5α-reductase inhibitors (Ashby and Lefevre, 2000; Owens et al., 2006, 2007; Freyberger et al., 2007). Compounds that significantly increase two or more AST weights in the absence of TP are considered positive

for androgenic activity, whereas compounds that significantly decrease two or more AST weights in the presence of TP are considered positive for antiandrogenicity.

The Hershberger assay underwent an extensive validation program coordinated by the OECD (Yamasaki et al., 2003a, 2003b, 2006; Owens et al., 2006, 2007; Shin et al., 2007; Moon et al., 2009), and it was demonstrated that the Hershberger assay can reliably detect androgenic and antiandrogenic activity across numerous laboratories using different routes of exposure in castrated rats. Furthermore, the assay shows good reproducibility both within and between laboratories (e.g., see Tables 5 and 6 in Owens et al. (2007)). The assay shows relatively good specificity for the androgen pathway, although thyroid hormone, growth hormone, prolactin, epithelial growth factor, and/or estrogens also can influence AST weights (OECD, 2008).

ASSAY CONDUCT–DOSE SELECTION

Uterotrophic Assay

The EPA and OECD test guidelines for the uterotrophic assay specify a limit dose of 1000 mg/kg body weight/day (mg/kg/day) for test compounds. The assay typically requires two dose levels and a vehicle control group, although more dose levels may be included. Dose levels are selected to avoid mortality, significant toxicity, or distress with some consideration of toxicokinetic factors. When using the immature model in one laboratory, the high-dose level often was selected based on a statistically significant decrease in body weight gains in the high-dose group. With only a 3-day dosing period, a significant decrease in body weight gains was more practical than a significant decrease in body weight, which would require marked effects on feed consumption, metabolism, and/or rate of growth, and could result in nonspecific effects on endocrine endpoints; thus, a significant decrease in body weight gain in immature animals was adopted as a maximum tolerated dose (MTD) criterion in one laboratory. For compounds administered by sc injection, irritation also may pose issues. Range-finding studies may be needed for dose-level selection, particularly if previously conducted studies have been performed using the dietary route of test substance administration (not oral gavage) and/or have been conducted in animals that are significantly older than those used in the uterotrophic assay. An untreated control group may also be included to ensure that the vehicle has no impact on uterine weights as this would alter assay sensitivity. An untreated control group was included in six separate studies in one laboratory; in each study, the vehicle was confirmed not to alter uterine weights. The other laboratory did not routinely include an untreated control group.

Hershberger Assay

As with the uterotrophic assay, dose levels are selected to avoid mortality, significant toxicity, or distress with some consideration of toxicokinetic factors. In addition, the highest dose level should not cause a greater than 10% reduction in terminal body weight relative to the control group. The limit dose for the Hershberger assay is 1000 mg/kg/day, but a dose level inducing an androgenic/antiandrogenic response in the assay also is considered sufficient. A previously reported study evaluating the effect of feed restriction on AST weights in the Hershberger assay showed that the assay is relatively insensitive to body weight–mediated changes in AST weights (Marty et al., 2003). As with the uterotrophic assay, irritation may pose problems for compounds administered by sc injection. Similarly, range-finding studies may be helpful for dose-level selection, particularly if previous toxicity studies have not used oral gavage.

TEST SYSTEM AND ROUTE OF EXPOSURE

Uterotrophic Assay

The OECD and US EPA differ in their preference for model and route when conducting the uterotrophic assay; the US EPA favors the use of ovariectomized adult rats with sc dosing, whereas the OECD recommends the immature (weanling) rat with either oral gavage or sc dosing. The test guidelines recommend that the investigator consider the relevant route of exposure (i.e., compounds with potential oral exposure can be given by oral gavage, whereas compounds with potential exposures by the inhalation or dermal route would require sc injection) and the potential for extensive "first-pass" metabolism, which should be avoided, when selecting the route of exposure. Other factors such as animal welfare, available toxicity information, and the physical/chemical properties of the test material also should be considered. Both animal models were reported to have comparable reliability, sensitivity, and reproducibility (e.g., Ashby et al., 1997; Kanno et al., 2001).

As discussed in the test guideline (OPPTS 890.1600; US EPA, 2009a), the US EPA recommends the use of adult ovariectomized rats with exposure via sc injection to allow direct entry of a compound into the general circulation while avoiding gut metabolism and slowing the rate of liver metabolism (US EPA, 2009a). If the ovariectomized model is used, animals must undergo surgery, then be allowed time for the uterus to regress (i.e., approximately 2 weeks) as discussed previously.

In contrast, the OECD recommends the use of immature rats due to animal welfare concerns regarding survival surgery. If the immature rat model is selected, female rats must be used between PNDs 18 and 25, with necropsy no later than PND 25 to avoid the increases in endogenous estrogen production that occur with the initiation of puberty (i.e., PND 0 is defined as the day of birth) (OECD, 2003). The OECD test guideline recommends that animal welfare and toxicologic aspects such as the relevance to the human route of exposure should be considered when selecting the route of exposure for the uterotrophic assay (i.e., oral gavage to model ingestion, sc injection to model inhalation or dermal adsorption).

While the EPA and OECD differ on which model they recommend for the uterotrophic assay (i.e., immature vs. mature ovariectomized rats), data that may already be available for a test substance from previously conducted studies can aid in the selection of the route of exposure and/or the uterotrophic model. First, the relevant route of exposure should be identified and toxicokinetic data on the test compound should be reviewed. Second, results of the EDSP Tier 1 in vitro ER binding and

transactivation assays should be considered; if the ER binding and/or transactivation assays are positive, this suggests that the parent compound may bind to the ER. In this case, sc injection of the test compound in the uterotrophic assay may avoid potential metabolic inactivation of a compound that might otherwise be estrogenic. While the use of sc injection may not represent expected exposure scenarios and/or one could challenge the relevance of identifying a substance as estrogenic in a scenario where first-pass metabolism has been circumvented, this approach is consistent with the goals of the EDSP program to avoid "false-negative" results. Similarly, if the ER binding and transactivation assays are negative up to the limit concentration (10^{-3} M), oral administration of the test compound allows an evaluation of metabolites for potential estrogenic activity, which is particularly important if oral exposure is the relevant route. Under these circumstances, this approach is also consistent with the goals of the EDSP program to avoid "false-negative" results and may provide important data for the subsequent WoE evaluation. Additional information on dose–response and adversity using a relevant route of exposure would be developed in the Tier 2 tests.

With respect to model, the adult ovariectomized model was reported to have increased specificity relative to the immature model (US EPA, 2009a), because the immature model responded to agents affecting the HPG axis rather than agents that act only at the ER (Lerner et al., 1958; Reel et al., 1996; Gray et al., 1997). In one laboratory, the adult ovariectomized model was used with the oral route of exposure. This met the recommendations of the EPA, while taking advantage of oral exposure for the reasons outlined above. In the other laboratory, the immature model was selected when using oral exposures because (1) via the oral route of exposure, the immature model was more sensitive to increases in uterine weight with estrogenic compounds than the ovariectomized model (Laws et al., 2000; Juberg et al., 2013); (2) some alternate activities (e.g., aromatizable androgens) that can result in a positive uterotrophic response in the immature model were a concern for potential endocrine activity; thus, using the immature model made the uterotrophic assay more inclusive for other endocrine MoAs that may be of concern; and (3) the immature model was more consistent with the animal use policies of the laboratory. Another noteworthy point, uterine weight is more closely related to body weight in the immature model; therefore, the variance in body weights at the start of the study must be less than ±20% of the mean body weight. Notably, under the parameters of the uterotrophic assay, if animals undergo a decrease in body weight in response to test compound treatment, an increase in uterine weight is readily discernable.

Hershberger Assay

The model system for the Hershberger assay is consistent between the OECD and EPA test guidelines, rats that have been orchi-epididyectomized at PND 42 or thereafter. An immature Hershberger model was proposed, but this model was less sensitive at detecting weak antiandrogens, and therefore was not included as part of the test guidelines (Freyberger and Schladt, 2009; OECD, 2009b).

For the Hershberger assay, the test compound can be administered by oral gavage or sc injection with consideration given to animal welfare, the physical/chemical properties of the test substance, toxicologic aspects like the relevant route for human exposure (e.g., oral gavage to model ingestion, sc injection to model inhalation or dermal adsorption), and existing data on metabolism and kinetics (e.g., need to avoid first-pass metabolism). List 1 compounds for EDSP screening were primarily pesticide-active ingredients; thus, oral ingestion of pesticide residues was often considered to be a relevant route of exposure to humans. In the two laboratories contributing to this article, all Hershberger assays used the oral route of exposure (gavage). However, if sc dosing is needed, laboratories should carefully consider the animal welfare issues related to administration of 10 daily injections of test compound and 10 daily injections of TP. Minimally, laboratories may wish to rotate injections sites to minimize pain/distress to the animals.

ASSAY IMPLEMENTATION–ASSAY LOGISTICS

Uterotrophic Assay

While not technically difficult to perform, there are preparations to be made before conducting a uterotrophic assay. Some laboratories order surgically modified (ovariectomized) rats from animal suppliers (rats ovariectomized at 6–8 weeks of age), which limits the number of laboratories required to perform these surgeries; however, assay scheduling must consider availability of the surgically modified rat model from the supplier as well as an adequate recovery period. If using immature rats, litters of rats born on specific dates are required to ensure that there are sufficient females to conduct that assay before PND 25.

Laboratories are required to demonstrate proficiency before conducting the uterotrophic assay. Both laboratories performed baseline positive control studies in which 17α-ethynylestradiol (EE) was administered with a minimum of four dose groups using the standard uterotrophic protocol adopted by each laboratory (i.e., relevant route [oral or sc] and model [immature or adult ovariectomized] as used in subsequent uterotrophic assays). Both laboratories demonstrated the responsiveness of their respective test systems (e.g., Marty, 2013) and generated dose–response curves similar to those seen during the uterotrophic validation (Kanno et al., 2001). Both laboratories include concurrent EE-treated controls in each uterotrophic assay to confirm responsiveness of the test system.

Hershberger Assay

Similar preparations are needed when conducting the Hershberger assay. Some laboratories order castrated rats from animal suppliers; however, because rats are generally castrated within a small age range (i.e., 42 days of age or thereafter; one laboratory uses PND 45 while the other laboratory uses PND 43–44), there are sometimes difficulties obtaining surgically modified rats when needed. Scheduling must consider availability of the model as well as an adequate recovery period (minimum of 7 days;

both laboratories initiated testing 10 days after surgery). Sample sizes for the assay are six rats per dose, although laboratories should be cognizant of the variability in some assay endpoints to determine whether this sample size is optimal for their needs. One laboratory used seven rats per dose group to improve the likelihood of meeting the assay performance criteria, while the other laboratory used the recommended six rats per dose group.

Necropsies are scheduled approximately 24 hr after the final dose in Hershberger assays. Dissections requires some practice to conduct consistently, particularly for AST tissues from animals not exposed to TP as these tissues are very small (e.g., paired Cowper's gland weights are often 5–6 mg); without sufficient practice, variability in these endpoints may confound assay results (Ashby et al., 2004) and/or result in deviations from the acceptable coefficients of variation (CVs) as specified in the test guidelines, which has the potential to result in a repeat of the assay if multiple performance criteria are missed. To reduce variability in tissue weights, one laboratory required that the same technician dissect the same tissues across dose groups within a study to avoid variance due to different dissectors. The second laboratory found this practice to be unnecessary.

ASSAY IMPLEMENTATION–ASSAY PROFICIENCY

Uterotrophic Assay

Regardless of which animal model and route of exposure are selected, the OECD and EPA uterotrophic test guidelines require laboratories to show proficiency and verify the responsiveness of the animal model by performing an initial baseline positive control study with EE at four or more dose levels that results in the expected dose–response curve, and subsequently requiring periodic validation by either (1) re-performing the baseline positive control study with EE every 6 months or when significant changes to the assay occur, and/or (2) inclusion of an EE-treated positive control group in each assay (recommended EE dose level that achieves a 70–80% increase in uterine weight relative to the maximum uterine weight increases in the EE dose–response study (i.e., ED_{70} or ED_{80}). EE dose–response curves for uterine wet weight increases in immature and ovariectomized adult rats are shown in Kanno et al. (2001). Both laboratories include an EE-treated positive control group with each uterotrophic assay they perform.

Hershberger Assay

Concurrent control groups for androgenicity (0.2 or 0.4 mg/kg/day TP given subcutaneously) and antiandrogenicity (TP given subcutaneously + 3 mg/kg/day flutamide given orally) are included in the Hershberger assay. The TP control and TP + flutamide control should yield significant increases and decreases in AST weights, respectively. Results from all control groups (vehicle, TP and TP + flutamide) should be compared with the laboratory historical control data and/or Hershberger validation data to verify assay performance.

ANIMAL HUSBANDRY

Uterotrophic Assay

In both laboratories, animals are maintained under the conditions recommended in the *Guide for the Care and Use of Laboratory Animals* (Institute of Laboratory Animal Research, Commission on Life Sciences, National Research Council, 1996). To limit potential exposures to alternate sources of estrogens, test animals are given a low-phytoestrogen rodent diet (daidzein + genistein aglycone equivalents ranged from nondetectable to 20 μg/g diet) in accordance with the requirements of the test guidelines, where genistein equivalents must be ≤350 μg/g diet; higher phytoestrogen content may increase baseline uterine weights (OECD, 2003). In addition, corncob bedding cannot be used in the uterotrophic assay due to reports of potential antiestrogenicity (Markaverich et al., 2005); therefore, a low phytoestrogen content bedding material is needed. One laboratory used 7089 Teklad Diamond Soft paper-pulp bedding (low phytoestrogen content; Harlan Laboratories, Indianapolis, IN), while the second laboratory used Shepherd's ALPHA-dri bedding (a bedding made of pure alpha cellulose; Animal Specialties and Provisions LLC, Quakertown, PA).

Hershberger Assay

Rats are maintained under conditions as recommended by the guidelines in the *Guide for the Care and Use of Laboratory Animals* (Institute of Laboratory Animal Research, Commission on Life Sciences, National Research Council, 1996). The Hershberger assay is relatively insensitive to animal husbandry conditions, including rat strain used, diet, bedding, caging, light cycles, or animal room conditions (temperature, humidity) (Ashby and Lefevre, 2000; Owens et al., 2006).

ASSAY CONDUCT–ENDPOINTS

Uterotrophic Assay

The uterotrophic assay is straightforward to conduct, requiring the collection of the incidence of dead/moribund animals or animals showing clinical signs of toxicity, body weights/body weight gains, and wet and blotted uterine weights. Vaginal patency is examined if the immature model is used, whereas 5 days of estrous cyclicity (preexposure) and an examination for ovarian reminants are required for the ovariectomized adult model. Optional endpoints include food consumption and vaginal and uterine histopathology. Uterine histopathology can distinguish between some apparently estrogenic responses (e.g., testosterone can increase uterine weight, but the histopathology is different from estrogen; OECD (2003)). Additional endpoints (e.g., target organ) may also be included if there is a desire to better characterize toxicity and/or stress. For example, one laboratory routinely collects vaginal cytology data for the duration of the assay (up through TD 4), which helps to confirm the absence of ovarian remnants but can also help to identify estrogenic substances by changes in vaginal cytology (i.e., progression of vaginal smears from diestrus to either proestrus or estrus) in the ovariectomized females.

Hershberger Assay

AST weights are the cornerstone of the Hershberger assay; however, there may be difficulty obtaining glans penis weights in all animals. For the Hershberger assay, male rats are castrated at approximately 42 days of age. Control data from three laboratories showed the mean age at PPS was between 42 and 46 days of age in CD rats (Stump et al., 2014). Thus, rats generally are castrated shortly before completion of PPS. At the end of dosing, PPS is examined in the Hershberger assay because glans penis weight cannot be collected in animals that have not completed PPS. If some animals have not achieved PPS, statistical analysis of PPS incidence is required according to the Hershberger test guidelines. Given the long interval between castration (~PND 42) and dosing (PND 49–60), most animals achieve PPS before initiation of treatment as mesenchymal-cell cornification of the balanopreputial epithelium was initiated before castration. However, even intact control animals can occasionally fail to achieve complete PPS (e.g., preputial threads may remain; Marty et al. (2003)). If an animal in a Hershberger assay fails to achieve PPS, it is unclear whether this is related to test material administration or some other delay in development that was present at the time of castration. Without additional data, it may not be possible to conclude whether such a finding is treatment related. To limit the potential impact of incomplete PPS, it may be useful to castrate animals at a slightly older age (i.e., PND 43–45 was used by the two laboratories participating in this article) and/or it may be useful to evaluate PPS before dosing to determine whether differences exist between groups before treatment. In both laboratories contributing to this article, there were no instances of incomplete PPS in the Hershberger assays conducted.

Many optional endpoints may also be measured in the Hershberger assay, including liver, kidney, and adrenal weights, and hormone levels (testosterone, luteinizing hormone [LH], follicle-stimulating hormone [FSH], triiodothyronine [T3], and thyroxine [T4]). Liver, kidney, and adrenal weights can provide additional information on systemic toxicity and/or metabolic enzyme induction. Other target organs also may be evaluated if there is a desire to better characterize systemic toxicity. Hormone measurements may provide additional information about MoA; however, it is important to recognize the small sample sizes may or may not be sufficient for accurate assessment of hormone levels due to interanimal variability (Owens and Ashby, 2002; Yamada et al., 2004), and furthermore, the animals are castrated, and therefore do not possess a normal-functioning HPG axis.

STATISTICAL ANALYSIS

Uterotrophic Assay

The uterotrophic assay requires a one-tailed statistical analysis of wet and blotted uterine weight. This analysis is appropriate because the uterotrophic assay, as described in the test guidelines, is designed to examine estrogenic responses only (uterine weight increases). Statistical significance was considered $p < 0.05$.

Hershberger Assay

The Hershberger test guideline advocates that body weights and organ weights should be evaluated for homogeneity of variance, transformed if appropriate, then analyzed by analysis of variance (ANOVA) followed by Dunnett's test. Statistical significance is considered $p < 0.05$. Once again, Dunnett's test is one-tailed to examine either androgenic responses (upper tail increases) or antiandrogenic responses (lower tail decreases). The test guidelines advocate the use of a combined analysis of all AST weights in a multivariate analysis to improve the detection of androgenic or antiandrogenic activity. However, when using the multivariate analysis of variance (MANOVA) on negative control data from the interlaboratory assay validation, an additional positive result was identified for nonylphenol with TP (personal communication), indicating that MANOVA increased the false-positive rate. Notably, these validation datasets had only one treatment group and a control group; multiple treatment groups would increase statistical power and presumably increase the false discovery rate to a greater extent.

ASSAY INTERPRETATION: HISTORICAL CONTROL DATA AND VARIANCE

Uterotrophic Assay

Generally, body weights and body weight gains were consistent among the multiple uterotrophic assays that were performed, regardless of whether they used intact immature rats or ovariectomized adult rats (Table 3). For immature rats, PND 22 terminal body weights generally ranged from 56 to 62 g except for one study with a lower value (52 g); despite this outlier, the mean CV for terminal body weight was 7.7% for all 13 studies. For the adult ovariectomized rats (dosing initiated at ~8 weeks of age), mean terminal body weights ranged from 247 to 325 g in rats dosed by the sc route of administration and from 245 to 347 g in rats dosed by oral gavage. While the absolute range in the adult rats was slightly greater for those studies using oral gavage compound administration, the small number of studies performed by the sc route likely contributed to this apparent difference. In support of this conclusion, the mean CV for the uterotrophic assays was approximately 6% for the uterotrophic assay using adult ovariectomized rats, regardless of the route of compound administration.

For one testing laboratory, initial experiments were conducted to confirm that there was no impact of the vehicle (corn oil) by comparing uterine weights in untreated controls with vehicle-treated controls. Data showed that neither sc injection nor oral gavage of corn oil vehicle altered baseline uterine weights. For both the ovariectomized adult model and the immature models, mean absolute wet and blotted uterine weights were comparable between the vehicle-treated control groups and the untreated control groups, confirming that there was no estrogenic activity in the vehicle. In the other laboratory, all studies were performed with test compounds prepared in a vehicle of 0.1% Tween 80/0.5% methylcellulose vehicle, while the EE-positive control was prepared in corn

Table 3
Uterotrophic Assay Historical Control Data for Uterine Weights

Endpoint	Untreated control[a]		Vehicle control							Mean increase with EE[b]	Range of increases with EE[b]
	Mean	SD	No. of studies	No. of animals	Mean	SD	Min value	Max value	Mean CV		
Ovariectomized adult model with sc dosing (corn oil vehicle)[c]											
Terminal body weight (g)	279.4	25.3			281.3	24.4	246.7	325.2	5.9%	NA	NA
Absolute wet uterine wt (g)	0.0869	0.0224	2	12	0.0761	0.0132	0.058	0.1080	15.1%	221%	86–357%
Absolute blotted uterine wt (g)	0.0736	0.0235			0.0726	0.0136	0.0601	0.1017	14.2%	180%	86–273%
Intact immature rat model with oral gavage dosing (corn oil vehicle)[d]											
Terminal body weight (g)	58.2	2.39			57.7	2.51	52.3	62.3	7.7%	NA	NA
Absolute wet uterine wt (g)	0.0254	0.0032	13	86	0.0254	0.0028	0.0210	0.0318	14.5%	795%	475–1086%
Absolute blotted uterine wt (g)	0.0235	0.0022			0.0236	0.0031	0.0195	0.0308	13.9%	441%	349–553%
Ovariectomized adult model with oral gavage dosing (0.1% Tween 80/0.5% methylcellulose vehicle)[c]											
Terminal body weight (g)	NA	NA			287.6	16.9	245.0	346.7	5.8%	NA	NA
Absolute wet uterine wt (g)	NA	NA	14	84	0.1000	0.0135	0.0763	0.1225	13.8%	196%	97–315%
Absolute blotted uterine wt (g)	NA	NA			0.0982	0.0131	0.0543	0.1326	13.7%	148%	87–238%

[a]Two untreated control groups for ovariectomized adult model; four untreated control groups for immature model.
[b]Animals dosed with 0.27–0.3 μg/kg/day EE for the ovariectomized adult model (sc), 100 μg/kg/day EE for the ovariectomized adult model (oral), or 10 μg/kg/day EE for the immature model (oral).
[c]Animals were ovariectomized at approximately 6 weeks of age and dosing was initiated at approximately 8 weeks of age, allowing a 2-week recovery period. Animals were dosed for 3 days by either subcutaneous injection or oral gavage and necropsied approximately 24 hr after the last dose.
[d]Animals were weaned on PND 18, dosed for 3 days from PND 19 to 21 and necropsied approximately 24 hr after the last dose on PND 22.
NA, not applicable.

oil (Table 3). A comparison of uterine weights in vehicle-treated and untreated control animals was not conducted in this laboratory, but the consistency of control uterine weights across studies suggests that it is unlikely that the vehicle altered baseline uterine weights. Among all three models (Table 3), mean uterine weights (wet and blotted) were consistent across different assays, with mean CV values that range from 13 to 15%. These results were consistent with findings in the OECD validation studies (Kanno et al., 2001; Owens and Koëter, 2003).

In the current studies, the required criteria for assay sensitivity were met in all cases. In all 13 studies using the immature model, blotted uterine weights in the vehicle control group were <40 mg, meeting the OECD 440 guideline criterion for acceptable control uterine weights. Furthermore, the mean blotted vehicle control uterine weight was typically 0.03 to 0.04% of terminal body weight for the immature animals. For the adult ovariectomized animals, the mean blotted vehicle control uterine weight was typically 0.025 to 0.026% of terminal body weight for one testing laboratory and 0.030 to 0.037% for the other testing laboratory. Thus, these baseline uterine weights always met the required criteria to yield sufficient assay sensitivity (<0.09% of terminal body weight for immature rats and less than <0.04% of terminal body weight for ovariectomized adult rats per OECD 440 and OPPTS 890.1600). High control uterine weights can compromise sensitivity of the uterotrophic assay such that weakly estrogenic compounds may not be detected. Historical data for vehicle control uterine weights and EE uterine weights (positive control) can also confirm the sensitivity of future uterotrophic assays; deviations from the historical control data should be examined carefully.

Overall, control values in the uterotrophic assay show good reproducibility between assays, which is consistent with the findings reported in the OECD validation (e.g., see Table 1 in Owens and Koëter (2003)). In the immature animals, the CVs for the vehicle control groups ranged from 5.8 to 30.8% (mean = 14.5%) for uterine wet weights and from 8.6 to 30.2% (mean = 13.9%) for blotted uterine weights. Similar CV values were observed with the adult ovariectomized model from both laboratories, where mean CV values were 15.1% for wet uterine weights and 14.2% for blotted uterine weights from one testing laboratory (ranges between 9 and 20% for both wet and blotted weights), and 13.8% for wet uterine weights and 13.7% for blotted uterine weights in the other testing laboratory (ranges between 5 and 28% for both wet and blotted weights).

In the OECD validation study, there was reasonable power (81% probability) of detecting a 35% increase in uterine weight at the high dose with six animals per group if the CV remained low (i.e., 15.0%). Across experiments from both testing laboratories, mean CV values for absolute wet and blotted uterine weight were ≤15% for vehicle-treated control animals (Table 3). Within individual assays, 50% of the wet and blotted uterine weight had CV values ≤15% with the adult ovariectomized model in both laboratories, which was a lower proportion of assays than reported in the uterotrophic validation program

Table 4
Hershberger Assay Historical Control Data for AST Weights

	Cowper's glands (g)	Percentage change[a] Cowper's glands	Glans penis (g)	Percentage change[a] glans penis	LABC (g)	Percentage change[a] LABC	Ventral prostate (g)	Percentage change[a] ventral prostate	Seminal vesicles with coagulating glands (g)	Percentage change[a] seminal vesicles
Group #1: Oral corn oil/sc corn oil vehicle control ($n = 3$ assays)[d]										
Mean ± SD	0.0066 ± 0.0020	NA	0.0532 ± 0.0075	NA	0.1436 ± 0.0243	NA	0.0242 ± 0.0100	NA	0.0439 ± 0.0139	NA
Mean range	0.0056–0.0083	NA	0.0479–0.0594	NA	0.1393–0.1498	NA	0.0235–0.0246	NA	0.0317–0.0598	NA
Mean CV (range)	26% (23–29%)	NA	11% (10–12%)	NA	16% (8–26%)	NA	47%[b](29–70%)[b]	NA	19% (17–20%)	NA
Group #2: Oral corn oil vehicle control (no sc dosing, $n = 7$ assays)[d]										
Mean ± SD	0.0060 ± 0.0025	NA	0.0521 ± 0.0075	NA	0.1574 ± 0.0377	NA	0.0190 ± 0.0068	NA	0.0418 ± 0.0086	NA
Mean range	0.0049–0.0094	NA	0.0490–0.0546	NA	0.1191–0.1911	NA	0.0147–0.0249	NA	0.0351–0.0493	NA
Mean CV (range)	36%[c](12–88%)[c]	NA	15% (12–17%)	NA	19% (10–34%)	NA	33% (21–49%)	NA	22% (17–25%)	NA
Group #3: Oral 0.1% Tween 80/0.5% methylcellulose vehicle (no sc dosing, $n = 12$ assays)[e]										
Mean ± SD	0.0067 ± 0.0022	NA	0.0516 ± 0.0085	NA	0.1817 ± 0.0270	NA	0.0131 ± 0.0045	NA	0.0816 ± 0.0211	NA
Mean range	0.0044–0.0100	NA	0.0396–0.0590	NA	0.1440–0.2176	NA	0.0044–0.0206	NA	0.0442–0.1071	NA
Mean CV (range)	33% (15–60%)	NA	16% (7–31%)	NA	15% (6–22%)	NA	32% (9–63%)	NA	27% (14–45%)	NA
Group #4: No oral dosing/sc 0.4 mg/kg/day TP control ($n = 12$ assays)[e]										
Mean ± SD	0.0407 ± 0.0078	+ 548	0.0829 ± 0.0106	+ 61	0.5535 ± 0.0547	+ 205	0.1743 ± 0.0349	+ 1358	0.8005 ± 0.1194	+ 943
Mean range	0.0350–0.0457	280–789	0.0741–0.0935	37–105	0.4466–0.6478	165–258	0.1085–1920	554–3196	0.6963–9823	680–1881
Mean CV (range)	19% (8–33%)	NA	13% (8–19%)	NA	10% (5–16%)	NA	21% (10–47%)	NA	15% (10–22%)	NA
Group #5: Oral corn oil/sc 0.4 mg/kg/day TP control ($n = 8$ assays)[d]										
Mean ± SD	0.0407 ± 0.0095	+ 584	0.0892 ± 0.0089	+ 70	0.4238 ± 0.1334	+ 159	0.2012 ± 0.0555	+ 858	0.5614 ± 0.1290	+ 1222
Mean Range	0.0316–0.0503	399–806	0.0847–0.0912	52–83	0.2773–0.6223	96–226	0.1393–0.2683	493–1328	0.4155–0.7104	855–1760
Mean CV (Range)	23% (12–40%)	NA	10% (6–16%)	NA	13% (8–17%)	NA	19% (13–26%)	NA	15% (8–26%)	NA
Group #6: Oral 0.1% Tween 80/0.5% methylcellulose vehicle/sc 0.4 mg/kg/day TP control ($n = 12$ assays)[e]										
Mean ± SD	0.0419 ± 0.0089	+ 555	0.0910 ± 0.0144	+ 78	0.5757 ± 0.0729	+ 220	0.1751 ± 0.0388	+ 1242	0.7889 ± 0.1473	+ 919
Mean range	0.0334–0.0475	316–715	0.0830–0.0956	53–110	0.4771—0.6932	159–292	0.1422–0.2093	750–3670	0.6977–0.9500	671–1611
Mean CV (range)	21% (14–30%)	NA	16% (5–27%)	NA	13% (7–22%)	NA	22% (2–39%)	NA	19% (11–28%)	NA
Group #7: Oral 3 mg/kg/day flutamide/sc 0.4 mg/kg/day TP ($n = 7$ assays)[d]										
Mean ± SD	0.0172 ± 0.0053	−58	0.0698 ± 0.0097	−22	0.2259 ± 0.0738	−44	0.0543 ± 0.0221	−73	0.1155 ± 0.0525	−79
Mean range	0.0134–0.0216	−51 to −64	0.0642–0.0773	−14 to −29	0.1715–0.3369	−36 to −57	0.0371–0.0866	−66 to −77	0.0801–0.1991	−70 to −83
Mean CV (range)	27% (15–38%)	NA	13% (10–17%)	NA	19% (12–30%)	NA	27% (16–43%)	NA	34%[f](17–60%)[f]	NA
Group #8: Oral 3 mg/kg/day flutamide/sc 0.4 mg/kg/day TP control ($n = 12$ assays)[e]										
Mean ± SD	0.0174 ± 0.0046	−58	0.0687 ± 0.0120	−24	0.3265 ± 0.0482	−43	0.0568 ± 0.0158	−67	0.2564 ± 0.0598	−67
Mean range	0.0142–0.0206	−50 to −70	0.0605–0.0764	−15 to −35	0.2719–0.3817	−36 to −54	0.0428—0.0691	−54 to −76	0.1867–0.3427	−56 to −77
Mean CV (range)	26% (14–43%)	NA	18% (7–30%)	NA	15% (6–31%)	NA	28% (7–37%)	NA	24% (2–42%)	NA

[a]Percentage change for 0.4 mg/kg/day TP control is relative to vehicle control values; percentage change for 0.4 mg/kg/day TP + 3 mg/kg/day flutamide is relative to 0.4 mg/kg/day TP values.

[b]Highest CV was due to one unusually high ventral prostate weight in one study; mean and maximum CVs were 37 and 43%, respectively, when this value was removed.

[c]Highest CV was due to enlarged Cowper's glands in one animal in one study (visually confirmed, but cause unknown); mean and maximum CVs were 28 and 45%, respectively, when this value was removed.

[d]Studies conducted by Laboratory A.

[e]Studies conducted by Laboratory B.

[f]Highest CV was due to one unusually high seminal vesicle weight in one study; mean and maximum CVs were 28 and 46%, respectively, when this value was removed.

NA, not applicable.

(78 and 89% of assays had wet weight and blotted weight CVs ≤15%, respectively; Kanno et al. (2001)). In contrast, 60% (wet weights) and 67% (blotted weights) of the 13 immature assays had CV values ≤15% for uterine weights, which was a higher proportion of studies than reported in the uterotrophic validation program (50 and 38% of assays had wet weight and blotted weight CVs ≤ 15%, respectively; Kanno et al. (2001)). Generally, blotted uterine weights have similar or slightly less variability than wet uterine weights.

The responsiveness of the positive control animals was confirmed as shown by the historical "EE-positive control" data (Table 3). For uterotrophic assays conducted in both testing laboratories, an EE-positive control group (10 μg/kg/day for oral dosing in immature rats; 100 μg/kg/day for oral dosing in adult ovariectomized rats; 0.3 μg/kg/day for sc dosing in adult ovariectomized rats) was run concurrently with the treatment groups in each assay. Within assays, the response of the uterotrophic assay to EE was relatively consistent, although there were some differences in EE responsiveness across assays. On average, sc administration of EE to ovariectomized adults induced a 221% increase in uterine wet weights and a 180% increase in blotted uterine weights across two assays. These EE responses were somewhat greater than the mean EE-induced increases in wet and blotted uterine weight in the OECD validation study at 0.3 μg/kg/day (100 and 94% increases, respectively; Kanno et al. (2001)). This difference is likely related to interanimal variability and small sample sizes. Oral administration of EE induced a 196% increase in uterine wet weights and 148% increase in blotted uterine weights across 14 assays with the adult ovariectomized model. Oral EE to immature rats induced a 795% increase in uterine wet weights and a 441% increase in blotted uterine weights across 13 assays. The EE-induced increases in uterine weight were

Table 5
In Vitro Formation of Hydroxylated Testosterone Metabolites from Control and Treated Livers in the Hershberger Assay

	Hydroxylated testosterone metabolites (pmol/min/mg microsomal protein)											
	2α-OHT		6α-OHT[a]		6β-OHT		7α-OHT		16α-OHT		16β-OHT[a]	
Dose (mg/kg/day)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	275.1	129.0	21.0	7.9	488.5	286.3	61.8	16.4	141.0	143.1	7.1	2.7
Mid dose	214.1	47.1	43.1*	4.3	922.0*	127.5	70.1	16.5	307.3*	112.6	198.0*	27.8
High dose	261.4	49.1	99.8*	21.2	936.0*	171.3	149.2*	37.5	253.3	50.1	249.0*	41.4

[a]Analysis performed on square-root transformed data.
$n = 7$/dose. *Significant at $\alpha = 0.05$.
OHT, hydroxytestosterone.

Table 6
Hershberger Assay Androgenic Study Design Maximum Permissible CV Results[a]

		No. (%) of studies where CV exceeded in the control group		No. (%) of studies where CV exceeded in the high-dose group	
Endpoint	Recommended maximum CV (%)	Laboratory A[b]	Laboratory B[c]	Laboratory A[b]	Laboratory B[d]
Cowper's gland	55	0 (0)	**1 (8)**	0 (0)	0 (0)
Glans penis	22	0 (0)	**2 (17)**	0 (0)	**4 (25)**
LABC	30	0 (0)	0 (0)	0 (0)	**1 (6)**
Ventral prostate	45	**2 (29)**	**3 (25)**	**1 (14)**	**2 (13)**
Seminal vesicles with coagulating glands	40	0 (0)	**2 (17)**	0 (0)	**1 (6)**

[a]When performance criteria were not met, values were generally just outside the acceptable range.
[b]Number of studies was 7.
[c]Number of studies was 12.
[d]Number of studies was 16.
Values in bold are the number of studies (percentage of studies) that exceeded the maximum permissible CV value specified in the test guidelines.

Table 7
Hershberger Assay Antiandrogenic Study Design Maximum Permissible CV Results[a]

		No. (%) of studies where CV exceeded in the control group		No. (%) of studies where CV exceeded in the high-dose group	
Endpoint	Recommended maximum CV (%)	Laboratory A[b]	Laboratory B[c]	Laboratory A[b]	Laboratory B[d]
Cowper's gland	35	**1 (13)**	0 (0)	**1 (13)**	0 (0)
Glans penis	17	0 (0)	**4 (33)**	0 (0)	**5 (31)**
LABC	20	0 (0)	**1 (8)**	**3 (38)**	0 (0)
Ventral prostate	40	0 (0)	**1 (8)**	**1 (13)**	**1 (6)**
Seminal vesicles with coagulating glands	40	0 (0)	0 (0)	0 (0)	0 (0)

[a]When performance criteria were not met, values were generally just outside the acceptable range.
[b]Number of studies was 8.
[c]Number of studies was 12.
[d]Number of studies was 16.
Values in bold are the number of studies (percentage of studies) that exceeded the maximum permissible CV value specified in the test guidelines.

similar to the uterotrophic validation study, where imbibed and blotted uterine weights were increased up to 754 and 400%, respectively, with oral administration of 10 μg/kg bw/day EE to immature rats (Kanno et al., 2001). EE-induced increases in uterine weights were less variable for blotted weights than wet weights, suggesting greater variability in the amount of fluid imbibition across animals.

Hershberger Assay

Across assays, terminal body weight and body weight gain were typically increased in the TP-treated animals compared to the control group (e.g., overall mean = 335 g in the TP-treated group compared with 315 g in the "oral corn oil" group for one testing laboratory and 339 g in the TP-treated group compared with 313 g in the methylcellulose control group for the other testing

laboratory). While body weights/body weight gains were slightly lower in the flutamide + TP group compared with the TP-treated animals, these differences were minor and seldom achieved statistically significance (e.g., overall mean = 328 g in the flutamide + TP-treated group vs. 335 for the TP-treated group for one testing laboratory, and 331 g in the flutamide + TP-treated group vs. 339 for the TP-treated group for the other testing laboratory). These TP effects on body weight/body weight gains also were seen in the Hershberger validation study (Owens et al., 2006).

For one of the testing laboratories participating in this article, initial experiments were conducted to confirm that there was no impact of sc injection of the TP vehicle (corn oil) by comparing AST weights in animals given corn oil orally versus animals given corn oil both orally and by sc injection (i.e., vehicle for test material and vehicle for TP, respectively). As shown in Table 4, the "oral corn oil only" group (Group 2) demonstrated that sc administration of corn oil did not alter body weights or baseline AST weights when compared with "oral and sc corn oil" group (Group 1). Given the lack of significant effects on body weights or organ weights, the minimal impact of body weight on Hershberger assay endpoints (Marty et al., 2003), and animal welfare concerns, these data support the conclusion that it is unnecessary to inject all animals with corn oil vehicle during the androgenic portion of the study. In subsequent Hershberger assays, the vehicle controls (oral vehicle with no sc injection) were compared with the treated groups (oral test compound with no sc injection) to examine whether the test compound induced AST weight increases (androgenic portion); thus, the absence of sc injection in these groups was unlikely to affect the outcome or interpretation of the assays and was judged to be more humane to the test animals. These results are consistent with the Hershberger validation study (Owens et al., 2006), where it was shown that sc administration of corn oil at 0.5 ml/kg/day did not alter AST weights, other organ weights, or body weights.

In the second testing laboratory, all studies were performed with test compounds and flutamide prepared in a vehicle of 0.1% Tween 80/0.5% methylcellulose (Table 4; Groups 3, 6, and 8), while the TP-positive control was prepared in corn oil (Group 4). This laboratory examined whether oral gavage dosing of vehicle affected AST weights in TP-treated animals. As shown for Groups 4 and 6, responses to TP treatment were similar across these groups, indicating that the vehicle did not affect TP responsiveness of the AST.

Table 4 shows laboratory historical control data for means, variances, ranges, and CVs of the vehicle control groups for both laboratories (Groups 1, 2, and 3), 0.4 mg/kg/day TP control group (Groups 4, 5, and 6), and 0.4 mg/kg/day TP plus 3 mg/kg/day flutamide group (Groups 7 and 8). Table 4 also shows the relative change in AST weights when animals were treated with 0.4 mg/kg/day TP (change relative to vehicle control) or 0.4 mg/kg/day TP with 3 mg/kg/day flutamide (change relative to TP control). The magnitude of the TP-induced increases in AST weights varied by tissue. While the increases were fairly consistent between both laboratories, one laboratory reported the greatest relative increase in seminal vesicle weights, whereas the other laboratory reported the greatest relative increase in ventral prostate weights. Across both laboratories, the TP increases in tissue weights ranged from a 61% increase for glans penis weights to a 1222% increase for seminal vesicles with coagulating glands and fluid and a 1358% increase for ventral prostate. This finding was reported in phases 2 and 3 of the Hershberger validation (Yamasaki et al., 2003a, 2003b, 2006; Owens et al., 2007). Flutamide treatment decreased the TP-induced increase in AST, ranging from a 22% decrease in glans penis weights to a 79% decrease in seminal vesicle weights. These results are consistent with the lower and higher dynamic range for glans penis and seminal vesicle weights, respectively, and similar results were observed in both laboratories. Historical data for vehicle control, TP-treated and flutamide + TP-treated AST weights can confirm the sensitivity of future Hershberger assays; deviations from the historical control data should be examined carefully.

As reported in the validation studies, control values in the Hershberger assay showed reasonably good reproducibility across studies. Generally, the lowest CV values were observed with glans penis and LABC weights (Table 4), which was consistent with the Hershberger validation results (Yamasaki et al., 2003a, 2003b, 2006; Owens et al., 2007). Furthermore, fluid-filled tissues (ventral prostate, seminal vesicles, and Cowper's glands) generally showed greater variability and higher CVs (Yamasaki et al., 2003a, 2003b, 2006; Owens et al., 2007). Across tissues, the TP-treated group had the lowest CV values, which may improve the sensitivity of the assay to detect antiandrogenic responses (i.e., this is the group against which treated values are compared). The test guidelines state that organ weight values that differ by >3 SDs from the treatment group mean should be carefully examined and possibly discarded, which can improve CV values.

ASSAY INTERPRETATION: PRACTICAL EXPERIENCE WITH ENDPOINT SENSITIVITY AND SPECIFICITY

Uterotrophic Assay

Despite relative consistency in the conduct and performance of the uterotrophic assay, false-negative results are possible. Issues with reproducibility may arise when increases in uterine weights are in the lower portion of the dose–response curve and, therefore, may not achieve statistical significance and/or baseline uterine weights are outside the normal range, altering assay sensitivity (OECD, 2003). Thus, modest increases in uterine weight should be interpreted carefully using a WoE approach. Suspect results warrant assay replication, particularly in cases where a maximum-tolerated dose was not achieved, and a higher dose level is warranted.

The uterotrophic assay also may yield false-positive results under some circumstances. For example, in one study in one laboratory, there was a significant increase in wet and blotted uterine weights at the mid-dose level, but not at other dose levels. Due to the minimal increase in uterine weights (23–24% with uterine weights within the laboratory historical control data range), the atypically low control uterine weights (i.e., outside the historical control data range), and the absence of a dose–response relationship, the assay was repeated twice to verify this

result. In the second and third replicates, which included larger sample sizes (10/group), there were no significant increases in wet or blotted uterine weights at any dose level of the test compound; thus, the result from the first study was not reproducible. The second laboratory has not experienced this over the 14 uterotrophic assays that are included in the historical control database.

In previous reports (Tinwell et al., 2000; Owens and Ashby, 2002; Kanno et al., 2003) where the uterotrophic assay was not reproducible, there was typically a minimal 15 to 40% increase in uterine weight that was statistically significant in an initial study. When using data from control (untreated) groups, group-to-group variation in uterine weights indicates that occasional false-positive results are possible. As stated in the OECD guidance document (OECD, 2003), "The difference between the highest and the lowest value data sets for the vehicle controls would yield an apparent difference of 15–40% over the low (control) values." This statement of intergroup variability is lower than the intergroup variability in the historical control data for our laboratories, where blotted uterine weights in controls ranged from 0.0195 to 0.0308 g for immature animals and ranged from 0.0601 to 0.1017 g and from 0.0543 to 0.1326 g for adult ovariecomized rats in the two laboratories (Table 3).

Supplemental Figure 3A and B shows laboratory historical control values for wet and blotted uterine weights across studies in immature female rats from one laboratory and for adult ovariectomized female rats from the other laboratory. Across assays, there is more inherent variability in control uterine weights from immature animals than adult ovariectomized animals, although within assay variance does not differ between the test models (Table 3). Control uterine weight data can be useful when interpreting equivocal results; thus, identification of outlier values may be more straightforward when examining adult ovariectomized data. Each laboratory should establish its own historical control data for baseline uterine weights to determine whether uterine weights in the vehicle control group are atypically high and a repeat of the assay should be considered.

In accordance with the OECD's "Detailed Background Review of the Uterotrophic Bioassay" (OECD, 2003), a positive uterotrophic compound should show a dose–response curve for at least two dose levels while reaching an increase in uterine weights that is greater than 40%. For these marginal increases (15–40%) in uterine weight: (1) the uterotrophic assay results could be compared with results from other EDSP Tier 1 assays (i.e., supported by the WoE: ER binding, ER transactivation, female pubertal, and fish short-term reproduction assays) to determine whether a pattern of altered estrogen signaling exists; (2) the uterotrophic assay could be repeated, possibly with larger sample sizes; or (3) a specific ER antagonist (e.g., ICI-182,780) could be used to block the compound-induced increase in uterine weight and demonstrate response specificity. Invariably, a robust WoE approach should be employed, using all available data on the test compound, to characterize the response from the uterotrophic assay. For example, in studies performed in one laboratory, two proprietary compounds have been evaluated in the uterotrophic assay using adult ovariectomized female rats and were found to result in marginal (<29%) increases in uterine wet and blotted weights. In these cases, receptor binding data showed no interaction with the ER, and the incidence of uterine fluid and estrous conversion (conversion from diestrus to proestrous or estrous) were 0%. Therefore, while there were marginal increases in uterine weight, the WoE supported a negative finding for estrogenicity. In contrast, four proprietary compounds evaluated by this same laboratory were found to be positive in the uterotrophic assay with increases in uterine weight >55% (and ranged up to a 179% increase compared to the vehicle control group), and invariably, the incidence of uterine fluid imbibition and estrous conversion were also increased to 17 and 50% for the substance inducing the 55% increase in uterine weight, respectively. Therefore, this illustrates the caution that must be taken when increases in uterine weight are small and the importance of the WoE evaluation in determining potential estrogenic activity, and this also illustrates the utility that the addition of estrous conversion evaluation can have as a supporting parameter for the uterorophic assay.

Hershberger Assay

In our laboratories, 23 test compounds have been examined in the Hershberger assays. Of those 23 test compounds, the 7 evaluated by one laboratory showed no signs of androgenicity as there were no significant increases in any AST weights by test compounds, and this finding was consistent with the negative AR binding results for these compounds. For the second laboratory, 16 test compounds have been evaluated in the Hershberger assay. Of those 16 test compounds (3 compounds as part of the EDSP and 13 proprietary compounds), 4 of the 16 test compounds showed statistically significant increases in at least one of the AST weights. Of those four compounds, one showed increased ventral prostate and seminal vesicle with coagulating gland weights, and it was subsequently confirmed to show binding to the AR. The other three compounds showed increased weights for one of the ASTs (one each for Cowper's gland, LABC, and seminal vesicles with coagulating glands) and were either negative in the AR binding assay or showed equivocal binding results. This suggests that the guideline recommendation of positive responses in two AST weights is a prudent approach for evaluating the results from the Hershberger assay.

For the antiandrogenic portion of the assay, one laboratory had six of seven compounds that showed statistically significant decreases in at least one absolute AST weight in the antiandrogenic portion of the assay; three of the seven compounds caused significant decreases in at least two absolute AST weights, the requirement for a "positive" Hershberger assay. All three compounds were negative for AR binding and none showed a pattern consistent with 5α-reductase inhibition. The endpoints that were most commonly decreased in response to test compound treatment were LABC (decreased in 4/7 assays), ventral prostate (decreased in 3/7 assays), and seminal vesicles with coagulating glands (decreased in 3/7 assays). Glans penis weights were significantly decreased in one assay (in which it was the only AST weight significantly altered) and Cowper's gland weights were not

significantly affected in any of the Hershberger assays conducted. For the second laboratory, 8 of 16 compounds evaluated showed statistically significant decreases in at least one AST in the antiandrogenic portion of the study. Of those eight compounds, four test compounds caused statistically significant decreases in at least two ASTs, and all four of these test compounds were subsequently confirmed to show AR binding in the AR binding assay. Of the remaining four test compounds, all showed statistically significant decreases in one AST (one each for glans penis and Cowper's gland, and two for LABC).

5α-reductase inhibitors, which block the conversion of testosterone to dihydrotestosterone, can be detected in the Hershberger assay by differential effects on AST weights. During the OECD validation of the Hershberger assay, the 5α-reductase inhibitor, finasteride, induced significant decreases in all five ASTs in a dose-dependent manner (Owens et al., 2007). The ventral prostate, seminal vesicles, and Cowper's gland showed similar sensitivity to finasteride (benchmark doses for these tissues were 0.87, 1.4, and 0.77 mg/kg bw/day, respectively), whereas LABC and glans penis weights were not as sensitive across laboratories (benchmark doses were 8.8 and 20.6 mg/kg bw/day, respectively) (Owens et al., 2007). In other Hershberger studies, seminal vesicle weights were among the most sensitive to finasteride treatment (Ashby et al., 2004; Kennel et al., 2004). In the eight antiandrogenic Hershberger assays conducted in this laboratory, none of the decreases in AST weights corresponded to the expected pattern for a 5α-reductase inhibitor and did not show binding in AR binding assays.

Generally, organ weights measured in the Hershberger assay are relatively insensitive to terminal body weight decreases (Marty et al., 2003). A 9 to 11% decrease in terminal body weight decreased absolute and relative weights of the LABC and relative ventral prostate weights in untreated, castrated rats; these AST decrements would not be misinterpreted as an androgenic response. In TP-treated animals, relative weights were altered for some organs (LABC, seminal vesicles with coagulating glands, glans penis, Cowper's glands) and the absolute weight of the Cowper's gland was affected at one time point (66 days of age). Overall, absolute AST weights in castrated rats supplemented with exogenous TP were not highly sensitive to body weight changes.

Hershberger Assay and Hepatic Enzyme Induction

The Hershberger assay is designed to detect compounds that interact with the AR or inhibit 5α-reductase; however, at the time of assay validation, it was recognized that positive results could be attained in the antiandrogenic portion of the Hershberger assay with compounds that induce liver enzymes (Freyberger et al., 2007; Freyberger and Schladt, 2009). Induction of liver enzymes can enhance metabolism and clearance of the exogenously administered TP; therefore, less TP is available to interact with ARs and subsequently maintain the weights for these ASTs. In the castrated male rat used in the Hershberger assay, there is no endogenous synthesis of androgens and therefore, no means to compensate for the increase in testosterone metabolism by a test compound. This can produce a seemingly antiandrogenic response due to a decrease in AST weights in the compound-treated groups by an indirect MoA, unrelated to the AR antagonism or 5α-reductase inhibition. The 10 days of bolus gavage dosing that are required in the Hershberger assay design likely promote enzyme induction by compounds that have this potential.

Optional endpoints such as liver weights and serum testosterone concentrations may provide some evidence to support an enzyme-induction MoA; however, it has been our experience that not all compounds that enhance TP clearance show significant decreases in serum testosterone levels. This may be related to the timing and variability in serum testosterone measurements. In the Hershberger assay, animals are sent to necropsy approximately 24 hr after the last dose of TP and test compound; if blood samples are collected at necropsy, this may not be the optimum time to detect changes in TP metabolism/clearance. Plasma testosterone can be rapidly metabolized in male rats (Coulson et al., 2003) with some interanimal variance in the rate of testosterone clearance (Fig. 1). Thus, at 24 hr post dosing, TP values may be at low levels in both treated and control animals which, when coupled with interanimal variability, may make it difficult to detect significant changes in serum testosterone levels. However, when significant decreases in serum testosterone are present, additional evidence is useful to support the hypothesis of enhanced TP metabolism/clearance, such as: (1) treatment-related increases in liver weights; (2) established enzyme induction in previous toxicity studies with the compound in question or its chemical class; and (3) the absence of AR binding. In some cases, additional data may be needed to support this MoA. This scenario was encountered in both of the participating laboratories.

To further support that an antiandrogenic result in a Hershberger assay was related to hepatic enzyme induction and enhanced testosterone clearance, a testosterone clearance study was conducted with a test compound evaluated in one laboratory. For this study, castrated male rats with indwelling jugular vein cannulas were used in a Hershberger assay design. Using a method similar to Coulson et al. (2003), rats were administered an intravenous dose emulsion of ^{14}C-testosterone in ~10% Intralipid on TD 1 (targeted dose levels: ~1 μCi/g and ~5.5 μg ^{14}C-testosterone/g; targeted radioactivity was ~0.85 μCi/rat). Following administration of ^{14}C-testosterone, blood was collected at 0.08, 0.17, 0.25, 0.5, 1, 2, 3, and 6 hr post dosing to ensure that there were no differences between groups in testosterone clearance rates before test material administration. On TD 1 to 10, daily oral gavage doses of control vehicle or test material (high dose only) were administered (animals were dosed on TD 1 after the initial ^{14}C-testosterone time course data were collected). On TD 2 to 9, TP was also administered by sc injection in accordance with the Hershberger design. On TD 10, animals with patent cannulas (≥3/dose group) were administered ^{14}C-testosterone via their jugular vein cannulas and a complete blood time course was collected as described for TD 1. Plasma was obtained from the blood via centrifugation and ^{14}C-testosterone derived radioactivity determined by liquid scintillation spectrometry (Packard Tri-Carb 2900TR) from weighed aliquots (~50 μl).

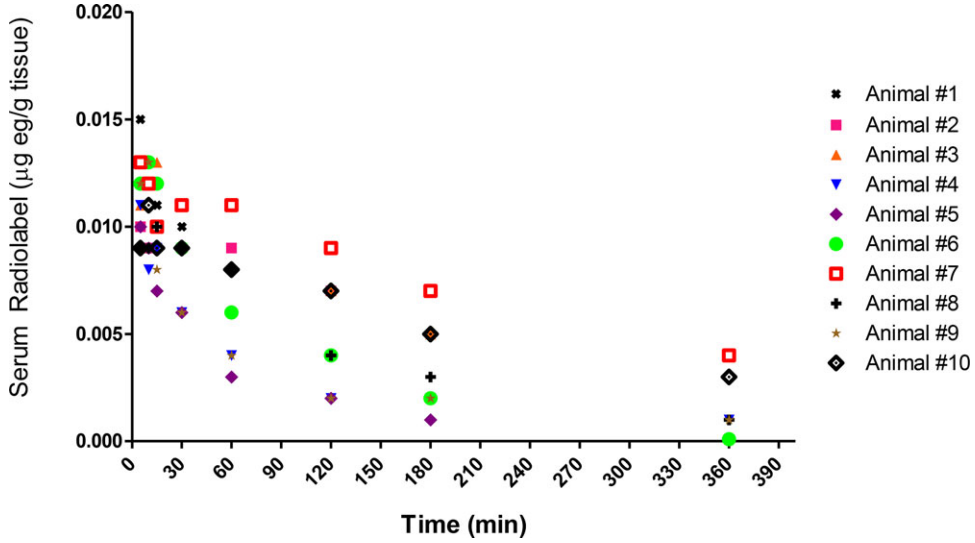

Fig. 1. Individual rates of clearance of radioactivity from serum in castrated control male rats given ^{14}C-testosterone by intravenous (iv) injection. Within a control population, some interanimal variability in testosterone clearance rates is apparent.

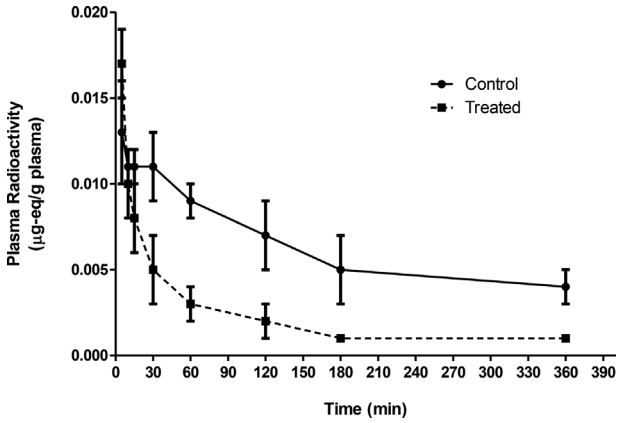

Fig. 2. The plasma time course for the clearance of radioactivity in control versus treated castrated CD rats. Radioactivity is derived from iv administration of ^{14}C-testosterone. Treated rats cleared radiolabel 3.7× faster than control rats, resulting in a 65% decrease in AUC.

The testosterone clearance study was used for one test compound, which was shown to induce hepatic enzymes in previous toxicity studies. This test material significantly decreased LABC and seminal vesicle weights and induced a 65% increase in liver weight at the high-dose level in the Hershberger assay. Terminal blood samples showed a 29% decrease in serum testosterone level in the high-dose group, a decrease that was not statistically significant. While there were no differences in initial testosterone clearance rates before test compound treatment, results indicated that the treated animals cleared ^{14}C-testosterone-derived radioactivity significantly faster (3.7-fold) than the TD 10 control animals (931 vs. 251 ml kg^{-1} hr^{-1}) due to significantly faster elimination ($t_{1/2\alpha}$) in the treated animals (Fig. 2). This was also reflected in the plasma AUC values, where the treated animals had a mean $AUC_{0\text{-}t}$ significantly lower (65%) than the controls (0.014 vs. 0.039 µg hr g^{-1}). Animals with nonpatent cannulas on TD 10 were administered ^{14}C-testosterone via the tail vein on TD 10 and limited time course sampling was collected from these animals. The limited time course data were consistent with the animals having a complete dataset. This compound was negative for AR binding at ≤1 mM in the Tier 1 AR binding assay using rat ventral prostate cytosols as well as ToxCast AR binding and AR transactivation assays (US EPA, 2013). The profile of AST weight changes was not consistent with 5α-reductase inhibition (see previous description). In addition to these results, a previous toxicity study with this compound indicated a 2.5-fold increase in hepatic uridine diphosphate glucuronosyltransferases (UGTs), a key enzyme for testosterone glucuronide conjugation and subsequent urinary excretion. Together, these data suggest that this compound produced a positive antiandrogenic response in the Hershberger assay via an MoA that does not directly involve the endocrine system. Of the seven Hershberger assays conducted in our laboratory as part of EDSP Tier 1 screening, three assays that were positive for antiandrogenicity (i.e., significant decreases in two or more AST weights) showed enhanced TP clearance secondary to hepatic enzyme induction. The remaining four Hershberger assays were negative.

An alternative method also was used in the same laboratory to evaluate hepatic enzyme induction and enhanced testosterone metabolism due to test compound administration; this method required the collection of liver samples at necropsy for subsequent in vitro testosterone metabolism studies. Liver weights were collected and livers were quick frozen in liquid nitrogen. These Hershberger liver samples were used to prepare liver microsomes to examine testosterone metabolism in vitro in control versus treated animals using a method similar to Freyberger et al. (2007). Briefly, frozen stored liver samples were thawed on ice, and homogenized using a Potter-Elvehjem apparatus. Microsomes were isolated using the method outlined by Guengerich (1982). Microsomal

incubation with testosterone was conducted according to published procedures (Williams and Borghoff, 2000; Li and Letcher, 2002), whereby one concentration of testosterone (0.25 mM) was incubated for 30 min with rat liver microsomes (1 mg/ml) in the presence of cofactor NADPH (1 mM) under physiologic conditions (pH 7.4 with 0.1 M phosphate buffer at 37°C). The reaction was stopped by transferring the reaction vials to an ice bath and adding 10% kill solution (2% formic acid in CH_3CN (v/v)). The supernatant was analyzed for hydroxylated testosterone metabolites (2α-, 6α-, 6β-, 7α-, 16α-, 16β-hydroxytestosterone) and androstenedione by LC/MS-MS methods according to published procedures (Li and Letcher, 2002; Wang and Zhang, 2007). Control incubations (i.e., with and without NADPH or microsomes or using heat-inactivated microsomes) were included.

Using this approach, one compound was examined in vitro for enhanced testosterone metabolism. This test material decreased AST weights and induced 49 and 102% increases in absolute liver weights at the mid- and high-dose levels, respectively, in the Hershberger assay. Using blood samples collected at necropsy, serum testosterone levels were decreased by 50% at the high-dose level, a value that was not statistically different from the control group. Serum testosterone was not decreased in the mid-dose animals relative to the control group. Previous MoA toxicity data indicated that this compound likely induces liver enzymes that would enhance testosterone metabolism. Consequently, in vitro testosterone metabolism was examined as described above. The metabolic capacity of liver microsomes isolated from treated animals in the antiandrogenic portion of the Hershberger assay (TP + test compound groups) was compared to liver microsomes from the control animals (TP-treatment only). Over the 30-min incubation period, microsomes isolated from treated animals rapidly metabolized testosterone, showing a dose-dependent increase in total hydroxytestosterone metabolite formation (Table 5). The hydroxylation of testosterone in the 16β position was most sensitive with 28- and 35-fold increases at the mid- and high-dose levels, respectively. In addition, there were significant, dose-related increases (two- to fivefold) in 6α- and 6β-hydroxytestosterone and 7α-hydroxytestosterone formation at the mid- and/or high-dose levels. 16α-hydroxytestosterone was significantly increased (twofold) at the mid-dose level, but not at the high-dose level; however, this biphasic response has been reported previously (Freyberger and Schladt, 2009) and may represent an increase in CYP 2B1 with a concurrent decrease in CYP 2C11. Testosterone hydroxylation to 2α-hydroxytestosterone or dehydrogenation to androstenedione was not changed with treatment. The profile of testosterone metabolism was consistent with other results examining enzyme expression in the liver of rodents exposed to this test compound. Furthermore, ToxCast data support enzyme induction of isozymes involved in steroid hormone metabolism and clearance (e.g., rat CYP 2A1, human CYP 2B6 [human isozyme corresponding with rat CYP 2B1; Imaoka et al. (1996)], CYP 3A4, UGT1A1, etc.). While not all potential testosterone metabolites were included in the analysis, the data clearly show an increase in the formation of hydroxylated testosterone metabolites, resulting in testosterone inactivation. The enhanced metabolism implies that testosterone was less available to interact with ARs and increase AST weights. Lastly, this compound did not alter AR binding.

PERFORMANCE CRITERIA

Uterotrophic Assay

There are no specific performance criteria provided in the test guidelines for the uterotrophic assay; however, baseline uterine weights must be below specifications given in the test guidelines to ensure assay sensitivity. Criteria for acceptable baseline uterine weights were as follows: (1) baseline uterine weights generally range from 20 to 35 mg for immature rats and from 80 to 110 mg in the ovariectomized young adults (OECD, 2003, 2007); (2) mean blotted uterine weight for the vehicle control group should be <0.04% of terminal body weight (ovariectomized model) or <0.09% of terminal body weight (immature model) to yield sufficient assay sensitivity (US EPA, 2009a); and (3) immature blotted uterine weights of 40 to 45 mg in the control group may warrant concern for assay sensitivity and weights greater than 45 mg may require the assay to be rerun (US EPA, 2009a). These criteria were met for 26 of 29 compounds evaluated in uterotrophic assays conducted in both testing laboratories. In one instance where these criteria were not met, the mean blotted uterine weight for the vehicle control group was 0.041% of terminal body weight, just outside the recommended range (<0.040%) for adult ovariectomized rats. In the other two instances where these criteria were not met, the mean uterine weights (wet and blotted weights) were slightly above the 80 to 110 mg range that is generally recommended.

Hershberger Assay

The US EPA and OECD provide performance criteria (i.e., CVs) for AST weights in the control and high-dose groups in the Hershberger assay. Per the test guidelines, it is important for negative assays to meet the performance criteria to demonstrate sufficient sensitivity to detect an AST weight change if one is present. Performance criteria results for our laboratories are shown in Tables 6 and 7. In the androgenic portion of the Hershberger assay, most recommended CV values were achieved across studies. Ventral prostate weight was the AST that most often exceeded the recommended CVs in both testing laboratories. In one laboratory, ventral prostate weight CV values exceed the recommended range in 3 of 14 measurements (21%) when looking at the control and high-dose groups combined. Similarly, in the second laboratory, ventral prostate weight CV values exceed the recommended range in 3 of 15 measurements (20%). The second laboratory also had studies where CV values for the Cowper's gland (7% for 15 combined groups for control and high-dose animals) and glans penis (13% for 15 combined groups for control and high-dose animals), and seminal vesicles with coagulating glands (13% for 15 combined groups for control and high-dose animals) exceeded the recommended range, although these incidences were lower than that observed for the ventral prostate. For the antiandrogenic portion of the Hershberger assay, instances where recommended CV values

were exceeded were more evenly split among all AST weights, with the only exception being the seminal vesicles, where the recommended CV range was not exceeded in either laboratory. For one laboratory, recommended CV values were missed twice for Cowper's gland weights (13% of 16 measurements), three times for LABC weights (19% of measurements), and once for ventral prostate (6% of measurements), while recommended CV ranges were not exceeded for glans penis weights in any studies. For the second laboratory, recommended CV values were missed four times for glans penis weights (27% of 15 measurements), one time for LABC weights (7% of measurements), and twice for ventral prostate weights (13% of measurements), while recommended CV ranges were not exceeded for Cowper's gland weight in any studies. Generally, when performance criteria were not met, values were just outside the acceptable range.

RECOMMENDATIONS

Uterotrophic Assay

The uterotrophic assay has good sensitivity and reasonable specificity; however, increases in uterine weight are not always due to estrogenic responses, particularly when using the immature model. Caution is warranted when interpreting minor increases in uterine weights ($\leq$ 40%) as increases of this magnitude are within the variance of the control population and may be spurious. When these minor increases in uterine weight are encountered, complete a full WoE evaluation of the test compound. A comparison of uterine weights with the laboratory historical control values also may aid in assay interpretation. If necessary, consider repeating the uterotrophic assay to determine whether potential estrogenic activity is present. Lastly, uterotrophic assay results, as with all of the Tier 1 EDSP assays, should be evaluated in a WoE approach that considers all available toxicity data and the results of other Tier 1 assays, particularly results for ER binding and transactivation, the female pubertal assay, and the fish short-term reproduction assay, the other four assays included in the EDSP that are designed to detect potential estrogenic activity.

Hershberger Assay

Based on the validation data, the Hershberger assay appears to be sensitive and specific for androgenic responses. The Hershberger assay is sensitive and reasonably specific for antiandrogenic responses, although it can detect hepatic enzyme-inducing compounds as well as AR antagonists and 5α-reductase inhibitors. Some optional endpoints (e.g., liver weights, serum testosterone levels) can be included in the Hershberger assay to shed light on the MoA. In some cases, alternate or complimentary approaches may be needed to verify which MoA is occurring (e.g., AR binding data, ex vivo testosterone metabolism using isolated liver microsomes from Hershberger rats, testosterone clearance studies, etc.). To assist in assay interpretation, laboratories also should maintain historical control data for vehicle controls, testosterone-treated controls, and testosterone- plus flutamide-treated controls to verify that assay results are consistent with previous findings. Lastly, Hershberger assay results should be evaluated in a WoE approach that considers all available toxicity data and the results of other Tier 1 assays, particularly results for AR binding, the male pubertal assay, and the fish short-term reproduction assay, the other three assays included in the EDSP that are designed to detect potential androgenic/antiandrogenic activity.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the following individuals for assistance during these studies: V. Marshall, C. Zablotny, J. Passage, K. Brooks, M. Lawson, J. Thomas, R. Sura, R. Hukkanen, K. Stebbins, Suzanne Snajdr, and the TERC pathology and animal care groups. The authors also wish to recognize the contributions of Dr. Craig Rowlands in reviewing this manuscript. These studies were conducted to meet data call-in requirements as part of the US EPA EDSP, and therefore, the studies were funded by Dow and DuPont in response to test orders for this regulatory program.

REFERENCES

Ashby J, Lefevre PA. 2000. Preliminary evaluation of the major protocol variables for the Hershberger castrated male rat assay for the detection of androgens, antiandrogens, and metabolic modulators. Regul Toxicol Pharmacol 31:92–105.

Ashby J, Odum J, Foster JR. 1997. Activity of raloxifene in immature and ovariectomized rat uterotrophic bioassays. Regul Toxicol Pharmacol 25:226–231.

Ashby J, Tinwell H, Odum J, Lefevre P. 2004. Natural variability and the influence of concurrent control values on the detection and interpretation of low-dose or weak endocrine toxicities. Environ Health Perspect 112:847–853.

Coulson M, Gibson GG, Plant N, Hammond T, Graham M. 2003. Lansoprazole increases testosterone metabolism and clearance in male Sprague-Dawley rats: implications for Leydig cell carcinogenesis. Toxicol Appl Pharmacol 192:154–163.

Freyberger A, Schladt L. 2009. Evaluation of the rodent Hershberger bioassay on intact juvenile males—testing of coded chemicals and supplementary biochemical investigations. Toxicology 262:114–120.

Freyberger A, Ellinger-Ziegelbauer H, Krötlinger F. 2007. Evaluation of the rodent Hershberger bioassay: testing of coded chemicals and supplementary molecular-biological and biochemical investigations. Toxicology 239:77–88.

Gray LE Jr, Kelce WR, Wiese T, Tyl R, Gaido K, Cook J, Klinefelter G, Desaulniers D, Wilson E, Zacharewski T, Waller C, Foster P, Laskey J, Reel J, Giesy J, Laws S, McLachlan J, Breslin W, Cooper R, di Giulio R, Johnson R, Purdy R, Mihaich E, Safe S, Sonnenschein C, Welshons W, Miller R, McMaster S, Colborn T. 1997. Endocrine screening methods workshop report: detection of estrogenic and androgenic hormonal and antihormonal activity for chemicals that act via receptor or steroidogenic enzyme mechanisms. Reprod Toxicol 11:719–750.

Guengerich FP. 1982. Microsomal enzymes involved in toxicology – analysis and separation. In: Hayes AW, editor. Principles and methods of toxiocology. New York: Raven Press. p. 609–634.

Imaoka S, Yamada T, Hiroi T, Hayashi K, Sakaki T, Yabusaki Y, Funae Y. 1996. Multiple forms of human P450 expressed in *Saccharomyces cerevisiae*. Systematic characterization and comparison with those of the rat. Biochem Pharmacol 51:1041–1050.

Juberg DR, Borghoff S, Becker RA, Casey W, Hartung T, Holsapple M, Marty S, Mihaich E, van der Kraak G, Wade MG, Willett K, Andersen M, Borgert C, Coady K, Dourson M, Fowle JR III, Gray E, Lamb J, Ortego L, Schug TT, Toole C, Zorrilla L, Kroner O, Patterson J, Rinckel L, Jones B. 2013. t^4 workshop report: lessons learned, challenges, and opportunities: the U.S. Endocrine Disruptor Screening Program. ALTEX Available at: http://www.altex.ch/resources/WR_Juberg_epub.pdf. Accessed October 27, 2013.

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 109:785–794.

Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003a. The OECD program to validate the rat uterotrophic bioassay. Phase 2: coded single-dose studies. Environ Health Perspect 111:1550–1558.

Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003b. The OECD program to validate the rat uterotrophic bioassay. Phase 2: dose-response studies. Environ Health Perspect 111:1530–1549.

Kennel PF, Pallen CT, Bars RG. 2004.Evaluation of the rodent Hershberger assay using three reference endocrine disrupters (androgen and antiandrogens). Reprod Toxicol 18:63–73.

Kim HS, Kang TS, Kang IH, Kim TS, Moon HJ, Kim IY, Ki H, Park KL, Lee BM, Yoo SD, Han SY. 2005. Validation study of OECD rodent uterotrophic assay for the assessment of estrogenic activity in Sprague-Dawley immature female rats. J Toxicol Environ Health A 68:2249–2262.

Laws SC, Carey SA, Ferrell JM, Bodman GJ, Cooper RL. 2000. Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. Toxicol Sci 54:154–167.

Lerner LJ, Holthaus FJ, Thompson, CR. 1958. A non-steroidal oestrogen antagonist 1-(*p*-2-diethylaminoexthoxyphenyl)-1-phenyl-2-*p*-methoxyphenylethanol. Endocrinology 63:295–318.

Li H, Letcher RJ. 2002. A high-performance-liquid-chromatography-based method for the determination of hydroxylated testosterone metabolites formed in vitro in liver microsomes from gray seal (Halichoerus grypus). J Chromatogr Sci 40:397–402.

Markaverich BM, Crowley JR, Alejandro MA, Shoulars K, Casajuna N, Mani S, Reyna A, Sharp J. 2005. Leukotoxin diols from ground corncob bedding disrupt estrous cyclicity in rats and stimulate MCF-7 breast cancer cell proliferation. Environ Health Perspect 113:1698–1704.

Marty MS. 2013. Mammalian methods for detecting and assessing endocrine-active compounds. In: Matthiessen P, editor. Methods of measuring the hazards and predicting the risks of endocrine disrupters for wildlife. Oxford, England: Wiley. p. 304–340.

Marty MS, Johnson KA, Carney EW. 2003. Effect of feed restriction on Hershberger and pubertal male assay endpoints. Birth Defects Res B 68:363–374.

Moon HJ, Kang TS, Kim TS, Kang IH, Ki HY, Kim SH, Han SY. 2009. OECD validation of phase 3 Hershberger assay in Korea using surgically castrated male rats with coded chemicals. J Appl Toxicol 29:350–355.

Institute of Laboratory Animal Research, Commission on Life Sciences, National Research Council. 1996. Guide for the care and use of laboratory animals. Washington, DC: National Academy Press.

OECD. 2003. Detailed background review of the uterotrophic bioassay. Series on testing and assessment. Number 38. ISSN: 2007–7876. Available at: http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2003)1&doclanguage=en. Accessed November 14, 2013.

OECD. 2007. OECD test guideline 440 uterotrophic bioassay in rodents: a short-term screening test for oestrogenic properties. Available at: http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents˙9789264067417-en;jsessionid=12ic71yg7bopl.delta. Accessed November 14, 2013.

OECD. 2008. Background review document on the rodent Hershberger bioassay. Series on testing and assessment. Number 90. ENV/JM/MONO(2008)17. July 24, 2008. Available at: http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2008)17&doclanguage=en. Accessed November 13, 2013.

OECD. 2009a. OECD test guideline 441 Hershberger bioassay in rats: a short-term screening assay for (anti) androgenic properties. Available at: http://www.oecd-ilibrary.org/environment/test-no-441-hershberger-bioassay-in-rats˙9789264076334-en. Accessed November 14, 2013.

OECD. 2009b. Guidance document on the weanling Hershberger bioassay in rats: a short-term screening assay for (anti)androgenic properties. Series on testing and assessment. Number 115. ENV/JM/MONO(2009)41. November 26, 2009. Available at: http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2009)41&doclanguage=en. Accessed November 13, 2013).

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. Organisation for Economic Co-operation and Development. Crit Rev Toxicol 32:445–520.

Owens W, Koëter HB. 2003. The OECD program to validate the rat uterotrophic bioassay: an overview. Environ Health Perspect 111:1527–1529.

Owens W, Ashby J, Odum J, Onyon L. 2003. The OECD program to validate the rat uterotrophic bioassay. Phase 2: dietary phytoestrogen analyses. Environ Health Perspect 111:1559–1567.

Owens W, Zeiger E, Walker M, Ashby J, Onyon L, Gray LE. 2006. The OECD program 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. Environ Health Perspect 114:1259–1265.

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 115:671–678.

Reel JR, Lamb JC IV, Neal BH. 1996. Survey and assessment of mammalian estrogen biological assays for hazard characterization. Fundam Appl Toxicol 34:288–305.

Shin JH, Moon HJ, Kang IH, Kim TS, Lee SJ, Ahn JY, Bae H, Jeung EB, Han SY. 2007. OECD validation of the rodent Hershberger assay using three reference chemicals; 17alpha-methyltestosterone, procymidone, and p,p'-DDE. Arch Toxicol 81:309–318.

Stump DG, O'Connor JC, Lewis JM, Marty MS. 2014. Endocrine Disruptor Screening Program (EDSP) Tier 1 male and female pubertal assays. Birth Defects Res Part B

Tinwell H, Joiner R, Pate I, Soames A, Foster J, Ashby J. 2000. Uterotrophic activity of bisphenol A in the immature mouse. Regul Toxicol Pharmacol 32:118–126.

US EPA. 2009a. Series 890—endocrine disruptor screening test OPPTS 890.1600: uterotrophic assay. Available at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0012. Accessed January 29, 2012.

US EPA. 2009b. Series 890—endocrine disruptor screening test OPPTS 890.1400: Hershberger bioassay. Available at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0008. Accessed January 29, 2012.

US EPA. 2013. ToxCast™. Computational toxcology research. Washington, DC: United States Environmental Protection Agency. Available at: http://www.epa.gov/comptox/toxcast/. Accessed November 13, 2013.

Wang D, Zhang M. 2007. Rapid quantitation of testosterone hydroxyl metabolites by ultra-performance liquid chromatography and mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 855:290–294.

Williams TM, Borghoff SJ. 2000. Induction of testosterone biotransformation enzymes following oral administration of methyl tert-butyl ether to male Sprague-Dawley rats. Toxicol Sci 57:147–155.

Yamada T, Kunimatsu T, Miyata K, Yabushita S, Sukata T, Kawamura S, Seki T, Okuno Y, Mikami N. 2004. Enhanced rat Hershberger assay appears reliable for detection of not only (anti-)androgenic chemicals but also thyroid hormone modulators. Toxicol Sci 79:64–74.

Yamasaki K, Takeyoshi M, Sawaki M, Imatanaka N, Shinoda K, Takatsuki M. 2003a. Immature rat uterotrophic assay of 18 chemicals and Hershberger assay of 30 chemicals. Toxicology 183: 93–115.

Yamasaki K, Sawaki M, Ohta R, Okuda H, Katayama S, Yamada T, Ohta T, Kosaka T, Owens W. 2003b. OECD validation of the Hershberger assay in Japan: Phase 2 dose response of methyltestosterone, vinclozolin, and p,p'-DDE. Environ Health Perspect 111: 1912–1919.

Yamasaki K, Ohta R, Okuda H. 2006. OECD validation of the Hershberger assay in Japan: phase 3. Blind study using coded chemicals. Toxicol Lett 163:121–129.

Zacharewski TR, Meek MD, Clemons JH, Wu ZF, Fielden MR, Matthews JB. 1998. Examination of the *in vitro* and *in vivo* estrogenic activities of eight commercial phthalate esters. Toxicol Sci 46: 282–293.