Pure Appl. Chem., Vol. 75, Nos. 11–12, pp. 2125–2141, 2003.

Topic 3.11

Significance of experimental studies for assessing adverse effects of endocrine-disrupting chemicals*

L. Earl Gray, Jr.[1,‡] and Paul M. D. Foster[2]

[1]2525 Highway 54, MD 72, NHEERL, Endocrinology Branch, Reproductive Toxicology Division, ORD, USEPA, Research Triangle Park, NC 27711, USA; [2]Environmental Toxicology Program, NIEHS, MD E1-06, 111 TW Alexander Drive, Research Triangle Park, NC 27709, USA

Abstract: The U.S. Environmental Protection Agency (USEPA) is developing an endocrine disruptor screening and testing program to detect chemicals that alter hypothalamic-pituitary-gonadal (HPG) function, estrogen, androgen, and thyroid (EAT) hormone synthesis or metabolism and induce androgen (AR) and estrogen (ER) receptor-mediated effects in mammals and other animals. The utility of this approach is based upon the knowledge that mechanisms of endocrine-disrupting chemical (EDC) action are highly conserved at the cellular and molecular levels among vertebrates. Some EDC mechanisms also are shared with invertebrates. High-priority chemicals would be evaluated in a Tier 1 screening (T1S) battery, and chemicals that are positive in T1S would then be tested in Tier 2 (T2). T1S includes in vitro ER and AR receptor binding and/or gene expression, an assessment of steroidogenesis and mammalian (rat) and nonmammalian (fish) in vivo assays. In vivo, the uterotropic assay detects estrogens and antiestrogens, while steroidogenesis, antithyroid activity, antiestrogenicity, and HPG function are assessed in a pubertal female assay. Antiandrogens are detected in the Hershberger assay (weight of androgen-dependent tissues in castrate-immature-male rats). Fish and amphibian assays are also being developed to identify EDCs. Several alternative mammalian in vivo assays have been proposed. Of these, a short-term pubertal male rat assay appears most promising. T1S is designed to be sensitive to EAT activities, but many of the effects detected at the screening level would not be considered adverse, the dosage levels may be high, and the route of administration used may not be the most relevant. However, issues of adversity, dose response, and route(s) of exposure would be resolved in the testing phase. In addition to using an enhanced multigenerational test for Tier 2, an in utero-lactational screening protocol is also being evaluated by the USEPA for use in T2 or T1S. For T2, the numbers of endocrine-sensitive endpoints and offspring (F1) examined in multigenerational tests need to be expanded for EDCs in a thoughtful manner, based in part upon the results of T1S. In addition, for some chemicals histological examination of 10 adult F1 per sex in only the control and high-dose groups provides inadequate

*Report from a SCOPE/IUPAC project: Implication of Endocrine Active Substances for Human and Wildlife (J. Miyamoto and J. Burger, editors). Other reports are published in this issue, *Pure Appl. Chem.* **75**, 1617–2615 (2003).

‡Corresponding author: Tel.: 919-541-7750; E-mail: gray.earl@epa.gov

DISCLAIMER. The research described in this article has been reviewed by the National Health Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

statistical power to detect low-dose lesions induced during development. In these cases, we propose that all the offspring be examined after puberty for gross and histological reproductive abnormalities. Since EDCs, like the phthalates and AR-antagonists, produce characteristic profiles, or syndromes, of adverse effects, data need to be reported in a manner that clearly identifies the proportion of animals displaying one or more of the abnormalities in a syndrome. Consideration should be given to tailoring T2, based on the results of T1S, to assure that all of the effects in such chemically induced developmental syndromes are included in the study.

INTRODUCTION

The potential effects of endocrine-disrupting chemicals (EDCs) on human health and the proven effects of EDCs on wildlife are a major concern among the public and the scientific communities. Due to gaps in the current testing of chemicals, the U.S. Environmental Protection Agency (USEPA) was given a mandate in 1996 under the Food Quality Protection and Safe Drinking Water Acts to develop a screening and testing program for endocrine effects. Some of the impetus for these actions arose from a Work Session in 1991 on "Chemically Induced Alterations in Sexual Development: The Wildlife/Human Connection" [1], which stated that "many compounds introduced into the environment by human activity are capable of disrupting the endocrine system of animals, including fish, wildlife, and humans. Endocrine disruption can be profound because of the crucial role hormones play in controlling development." [2]. Scientists at this meeting "estimated with confidence" that impairments in humans have resulted from exposure to endocrine disruptors. Laboratory animal studies corroborate the abnormalities of reproductive development observed in the field and, in some cases, define toxic mechanisms causing adverse effects. Among chemicals of concern, are pesticides, "inert" ingredients, industrial chemicals, pharmaceuticals, phytochemicals, food supplements, personal care products, and "natural" products or nutraceuticals.

In response to the 1996 legislative mandate for an endocrine screening and testing program, the USEPA formed the Endocrine Disruptors Screening and Testing Advisory Committee (EDSTAC), which proposed a tiered screening and testing strategy for EDCs in its final report (1998) (<http://www.epa.gov/scipoly/oscpendo/history/finalrpt.htm>). The EDSTAC proposal includes (i) a process to prioritize chemicals for evaluation and (ii) tiered screening (Tier 1) and (iii) testing (Tier 2) batteries. The chemical "universe" to be considered includes over 80 000 chemicals, of which only a subset of high priority chemicals would initially enter the screening program. Prioritization would include an estimation of the chemicals' ability to interact with steroid hormone receptors using either quantitative structure–activity relationships (QSARs) for chemicals that bind steroid receptors or high-throughput prescreening (HTPS) using hormone-dependent gene expression assays, among several other factors. However, to date validated QSAR models or HTPS have not yet been developed for either estrogens or androgens by the USEPA.

The screening battery recommended by EDSTAC was designed to detect chemicals that alter hypothalamic-pituitary-gonadal (HPG) function, estrogen, androgen, and thyroid hormone synthesis or metabolism, or induce androgen (AR) and estrogen (ER) receptor-mediated effects in mammals and other taxa. Chemicals positive in Tier 1, based upon a "weight-of-evidence" analysis, would be considered as potential EDCs and subjected to testing (Tier 2). Tier 1 should include assays sensitive enough to detect EDCs while issues of "dose-response, relevance of the route of exposure, sensitive life stages and adversity" would be resolved in the testing phase. Although the USEPA has not yet endorsed this approach, we believe that equivocal effects in Tier 1 assays should be replicated or evaluated further in additional short-term assays before more extensive Tier 2 testing (T2T) is initiated (Fig. 1). This could prevent unnecessary testing in some cases, saving time and money and reducing animal use, and would provide useful information for tailoring Tier 2 tests.

Based in large part upon the EDSTAC recommendations, the Office of Prevention, Pesticides and Toxic Substances (OPPTS) of the USEPA designed the EPA Endocrine Disruptors Screening Program (EDSP), and the agency has begun implementing elements of this program (report to Congress, available at <http://www.epa.gov/scipoly/oscpendo/index.htm#currentstatus>), and their efforts are being coordinated internationally with the Office of Economic Cooperation and Development (OECD) (described at the following Web site: <http://www.oecd.org//ehs/ENDOCRIN.HTM>).

The first portion of this review will briefly summarize the status of some of the screening assays, while the second part will discuss approaches and limitations of different methods for testing EDCs identified as potential endocrine disruptors in T1S.

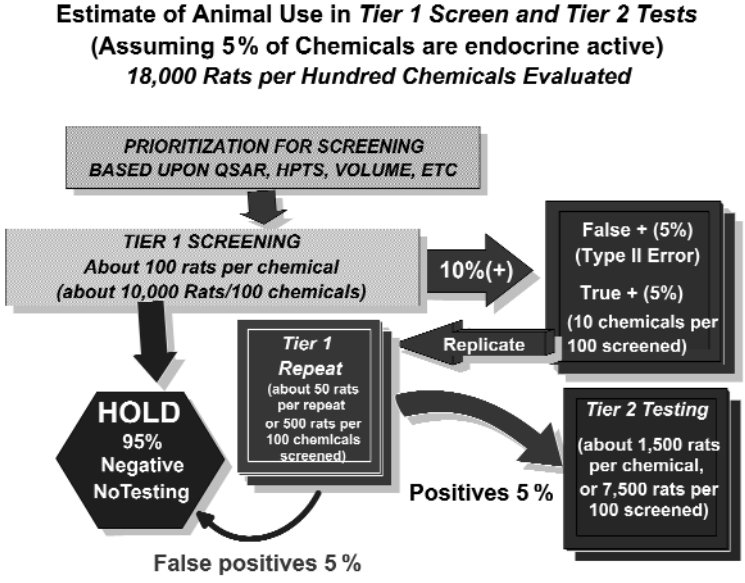

Fig. 1 In this figure, we provide an estimate the animal use in the Screening and Testing Program based upon the numbers of animals used per hundred chemicals screened. This estimate is based upon several assumptions. T1S would use about 100 animals per chemical screened. T1S includes three in vivo assays [uterotropic (n = 6/group), Hershberger (n = 6/group) and the pubertal female rat assay (n = 15/group) or a total of 27 rats per group]. We assumed that each chemical would be evaluated using three groups (control, medium and high dosage levels = 81 rats; the high dosage level should not exceed an MTD, less than a 10 % body weight reduction vs. control) with a few animals being used for limited dose-range finding for T1S. In this figure, we assumed that about 5 % of the chemicals screened would be positive because they act in an antiestrogenic, antiandrogenic, antithyroid manner or inhibit steroidogenesis or HPG maturation. We also assumed that another 5 % would be statistical false positives. We propose that false positives could be virtually eliminated from entering T2T and sent to the hold box by repeating or expanding positive T1S results before proceeding to the testing phase of the battery. T2T is the most intensive phase of the battery in terms of animal use, using about 1500 animals in a multigenerational test per chemical. Note that for each 5 % increase in the rate of false positives there is an increase of 250 rats used per 100 chemicals in the "repeat" phase we have proposed as a transition from T1S to either T2T or the hold box. Hence, if the false positive rate was 50 % rather than 5 %, animal use would increase from 18 000 only by 2250 rats to 20 250 per 100 chemicals. We suspect that initially only a few hundred high-priority chemicals will be evaluated in the EDSP Tier 1 screening battery.

TIER 1 SCREENING: IN VITRO

In vitro, cell-free, or whole-cell mammalian AR or ER binding assays were proposed to detect compounds that act as hormone agonists or antagonists [3]. The use of whole cells or recombinant human receptors for binding assays [as described by refs. 4,5,6] eliminates the need to isolate receptors from

animal tissues and may be amenable to high-throughput screening. Gene expression assays also were proposed because they can distinguish receptor agonists from antagonists. The use of cell lines containing AR or ER stably or transiently transfected with a reporter gene, or cell lines cotransfected with the reporter gene and the steroid receptor are now available to rapidly discriminate receptor agonists from antagonists [7–10]. Stable cell lines are likely to be less variable than transient transfections. and they do not require repeatedly transfecting cells, while new transfection methods have improved the utility of the transient transfection assays [11]. Recently, CV-1 and MDA-453 cells were transduced with hAR and reporter genes using adenovirus to detect antiandrogens [12]. These transduction assays provided low background and 80–100-fold induction with 1–10 nM dihydrotestosterone. Adenovirus transduced cell lines and stably transfected cells are amenable to high-throughput screening. In addition, transduction assays using adenovirus could also be used to deliver other genes to different cell types to screen for additional EDC mechanisms beyond EAT. Optimized and standardized protocols along with performance criteria (intra- and inter-assay and interlaboratory CVs) for binding and transcriptional activation assays need to be developed. While the ER and AR binding and transcriptional activation assays appear to be quite sensitive, except for some chemicals that require metabolic activation, the specificity of responses at the high concentrations that are used for weak ligands appears poor. The fact that a chemical induces a response in vitro that reaches an IC_{50} in the high micromolar to millimolar range may be due to disruption of the assay conditions for receptor stability or cell viability, rather than competitive binding to the receptor. Improving the specificity of these assays will require further in vitro experimentation to separate receptor ligands from false positives. Cytotoxicity (for whole-cell assays), and Ki [13] or Kb [14] determinations are required to enhance the specificity of these assays to acceptable levels.

Some EDCs inhibit reproduction and development by altering steroid hormone synthesis. While the presence of ER and AR in invertebrates has yet to be established, chemicals that alter steroid hormone synthesis in vertebrates can also inhibit ecdysteroid synthesis in invertebrates. Chemicals that inhibit steroidogenesis can be detected in vitro or ex vivo using a variety of methods [briefly reviewed by refs. 15,16]. Potential protocols for these assays are being evaluated under within the USEPA and in contract laboratories. A T1S in vitro assessment of steroidogenesis would complement the in vivo effects of altered steroidogenesis in the pubertal female rat assay.

Scientists are attempting to use as few animals as possible in the most precise and sensitive assays, by incorporating sensitive in vitro assays in T1S and by using QSAR models or HTPS in the prioritization of chemicals for screening. Unwarranted animal use also can be avoided as chemicals negative in T1S will not be subject to T2T (Fig. 1). In addition, testing T1S statistical "false positives" in T2 can be almost entirely eliminated by assuring that T1S assay results are replicated or further investigated in supplemental studies, before moving to T2T. Attempts to enhance T2T by adding more sensitive endpoints and a more thorough evaluation of the animals already on study should lead to improved hazard characterization and result in reductions in animal use.

TIER 1 SCREENING: IN VIVO

Three short-term in vivo mammalian assays were included in the EDSTAC Tier 1 screening battery [EDSTAC, 1998; 15,17,18]. In vivo assays are required in T1S because in vitro assays cannot account for absorption, distribution, metabolic activation, and excretion of xenobiotics, potentially resulting in "false negative" responses (positive in vivo, but negative in vitro). For example, purified methoxychlor is inactive or weakly active in vitro whereas in vivo it is almost as potent as is 17β estradiol when both are given orally. In addition, the integrated nature of the endocrine system in the developing organism and the relationship of the endocrine toxicity to other systemic effects cannot be simulated in vitro.

Fish and amphibian assays also are being developed for T1S [19–21]. The fathead minnow assay, developed by USEPA scientists, can identify EDCs displaying several mechanisms of concern, includ-

ing androgen and estrogen receptor agonists and antagonists and inhibitors of steroid hormone synthesis. An amphibian metamorphosis assay is being developed to detect thyroid-active substances.

UTEROTROPIC ASSAY

In T1S, estrogen agonists and antagonists are detected in a 3-day uterotropic assay using subcutaneous (sc) administration of the test compound. Several variations of the uterotropic assay protocol are being examined by OECD, and several phases of the interlaboratory study have been completed and published [22–24]. The described uterotropic assays for estrogens and antiestrogens uses either intact juvenile or ovariectomized adult/juvenile female rats and was found by all laboratories to be responsive to estrogenic toxicants over a wide range of potencies.

The second in vivo assay in the EDSTAC T1S is designed to detect antiandrogenic activity simply by weighing androgen-dependent tissues following 10 days of oral treatment in the immature castrated male rat [15,17,25] (Fig. 2). Both the uterotropic and Hershberger assays were standardized and validated for screening chemicals by an expert committee of endocrinologists in 1962 [26]. In this assay, weights of the ventral prostate, Cowper's glands, seminal vesicle (with coagulating glands and fluids), glans penis and levator ani/bulbocavernosus muscles are measured in castrated, sc testosterone-treated (or untreated) male rats after 10 days of oral treatment with the test compound. This is a very sensitive assay for detection of androgens and antiandrogens. Other useful endpoints that help reveal the mechanism of action include dorsolateral prostate, adrenal, liver, and kidney weights and serum (collected by cardiac puncture) levels of testosterone and LH. The Hershberger assay has been shown to be much more sensitive and specific to AR-mediated alterations than assessment of endocrine activity in the intact adult male rat, which fails to consistently detect the antiandrogenic activity of several weakly antiandrogenic pesticides like *p*,*p*-DDE and linuron that are easily detected in the Hershberger assay [6,27]. Chemicals that induce testosterone metabolism also can reduce androgen-dependent tissue weights while those that inhibit 5α reductase activity, can dramatically reduce male accessory sex gland weight with lessor effect on testosterone-mediated muscle. It appears that steroidogenesis in the

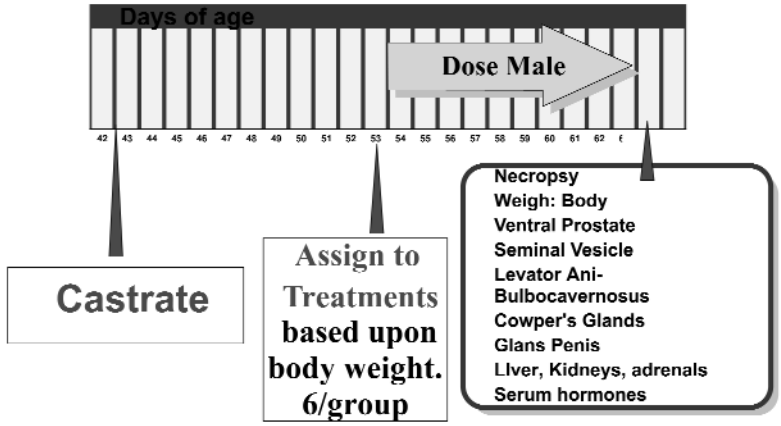

Fig. 2 OECD Hershberger assay protocol for antiandrogens, testosterone propionate (TP sc at 0.4 mg/kg/d for detection of an antiandrogen) is coadministered to male rats with the oral treatment for 10 days. To detect androgens, omit concurrent TP treatment. Castrate-immature male rats are dosed sc with TP and orally with the toxicant for 10 days, after which the animals are necropsied and androgen-dependent tissues (ventral prostate, seminal vesicles, Cowper's glands, glans penis and levator ani-bulbocavernosus muscles) and other organs (liver, kidney, adrenals) are weighed. Serum (collected by cardiac puncture) can also be taken for analysis of testosterone levels. Detects androgen receptor (AR) agonists and antagonists, and chemicals that alter AR levels, inhibitor of 5α reductase or stimulate testosterone metabolism.

adrenals together with thyroid function could also be evaluated in the Hershberger assay, but these endpoints are not routinely included by most investigators using this assay. Although changes in androgen-dependent organ weights in this assay are not necessarily considered as being adverse, we have found that chemicals that are positive in this assay often produce adverse effects during puberty and after in utero exposure. Members of the Endocrinology Branch of NHEERL, ORD, USEPA are working as the lead laboratory supporting the OECD-led effort to standardize and validate this assay in an interlaboratory (17 laboratories) study of androgens and antiandrogens of varying potencies. To date, phase 1 [an interlaboratory study using testosterone propionate sc (17 laboratories, 6 dosage levels), and flutamide (6 laboratories, 6 dosage levels)] and phase 2 (17 laboratories, 7 chemicals at several dosage levels, with 3–8 laboratories per chemical) have been completed, reported to the OECD, and the results are being prepared for publication.

PUBERTAL FEMALE RAT ASSAY

The "pubertal female rat assay" in T1S (Fig. 3) [28] was originally developed in the Endocrinology Branch in our laboratory along with the "pubertal male assay" nearly two decades ago [29,30]. In the pubertal female assay, weanling rats are dosed daily by gavage for about 20 days and the age at vaginal opening (puberty) and estrous cycles are monitored and the females are necropsied at about 42 days of age. The age at vaginal opening appears to be exquisitely sensitive to xenoestrogens like methoxychlor [30]. Additional measurements include serum thyroid hormones, and uterine and ovarian weight and histology. This assay detects alterations in thyroid hormone status, HPG function, inhibition of steroidogenesis (including aromatase), estrogens, and antiestrogens [31,32].

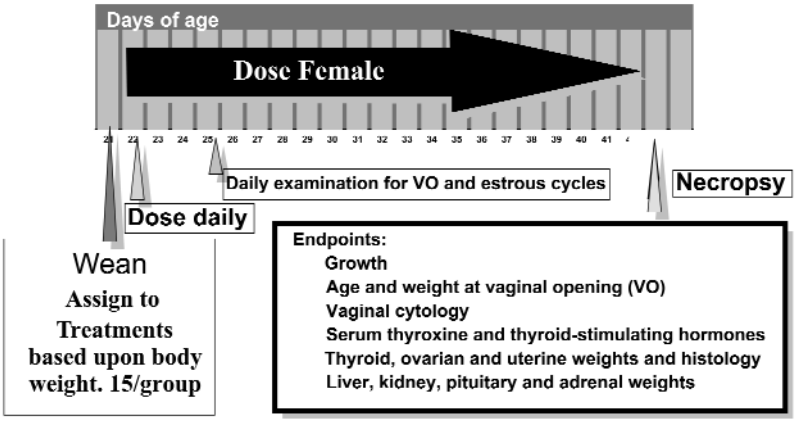

Fig. 3 Pubertal female rat protocol. Originally developed by the USEPA, ORD, NHEERL, Endocrinology Branch [28–31,33,35,37,40]. Intact weanling females are dosed by gavage from 22–43 days of age). The age and weight at vaginal opening (puberty) is determined, estrous cycles are monitored after puberty and the females are necropsied at 43 days of age. Necropsy includes an assessment of reproductive (uterus and ovaries) and nonreproductive (liver, kidney, adrenals, thyroid) organ weights and histology (uterus, ovaries, and thyroid). Serum is taken for T4 and TSH analyses. This protocol is designed to detect inhibition of steroidogenesis including aromatase, antithyroid, and antiestrogenic activities and altered HPG maturation.

PUBERTAL MALE ASSAY: AN ALTERNATIVE SCREENING ASSAY

Alternative T1 and T2 in vivo assays were discussed by EDSTAC, and some of these are being evaluated by the USEPA in the EDSP. If these assays prove to be of sufficient sensitivity and equivalent utility, they could replace or augment current T1S assays. One promising alternative assay is the "pubertal male rat assay" (Fig. 4) [15,29,30,34–37], which detects alterations in thyroid function, HPG matura-

tion, steroidogenesis (but not aromatase), and steroid hormone function (androgen). Intact weanling males are exposed orally to the test substance for about 30 days, the age at puberty is determined, and reproductive tissues are evaluated and serum taken for hormonal analyses. Although this is a screening assay, the USEPA has previously used altered puberty [a two-day delay in preputial separation (vinclozolin) or a similar delay in vaginal opening in the female] as adverse effects in their risk assessments. In addition, for vinclozolin the extra 10 X factor mandated by 1996 Food Quality Protection Act for chemicals of special concern for the child was retained in the risk assessment. In this case, a delay in preputial separation was accompanied by other endocrine and reproductive alterations, in the absence of an effect on body weight or growth. Furthermore, all of the in vivo reproductive effects of vinclozolin are consistent with the AR antagonism seen in vitro and altered androgen-dependent gene expression in vivo [13,38] and other in vivo effects [39,40].

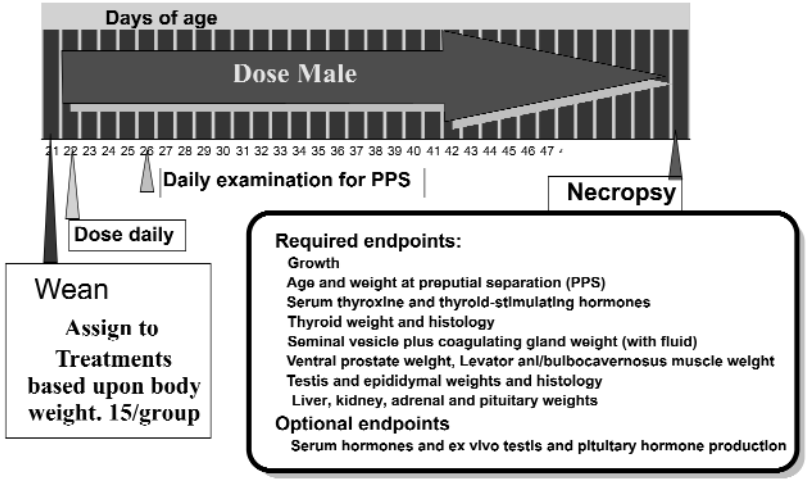

Fig. 4 Pubertal male rat protocol. Originally developed by the USEPA, ORD, NHEERL, Endocrinology Branch [28–31,33,35,37,40]. Weanling males are dosed by gavage from 22–53 days of age. The age and weight at preputial separation (puberty) is determined and the males are necropsied at 52–55 days of age. Necropsy includes an assessment of reproductive (testes, epididymides, ventral prostate, seminal vesicles, Cowper's glands, levator ani plus bulbocavernosus muscles) and nonreproductive (liver, kidney, adrenals), pi organ weights and histology. Serum is taken for T4 and TSH analyses. Thyroid weight and histology also are evaluated. This protocol is designed to detect inhibition of steroidogenesis, antithyroid, and antiandrogenic activities and altered HPG maturation.

TIER 2 TESTING

In a tiered screening and testing approach, only chemicals displaying positive, reproducible responses in T1S would be evaluated further in full life cycle, or multigenerational tests (Fig. 1). It is in T2 testing, not T1S, that issues of dose response, relevance of the route of exposure, sensitive life stages, and adversity are resolved [15,41,42]. For some endocrine activities, the numbers of sensitive endpoints and (F1) offspring examined in these assays should be expanded on a case-by-case basis. For example, endpoints like anogenital distance at birth and nipple/areola retention in infant female and male rats should be included in testing of androgens and antiandrogens, respectively, because they are sensitive, potentially permanent effects that are highly correlated with malformations and reproductive organ weight changes later in life. These early alterations constitute part of antiandrogen-induced developmental syndromes. As described in detail earlier [40–47], a careful evaluation of the male rat offspring allows one to distinguish the "phthalate syndrome" where effects on reproductive development involve a decrease in fetal testicular testosterone and insulin like-3 peptide hormone biosynthesis [46,48,49] from the "AR

antagonist syndrome" typified by vinclozolin or flutamide [50]. The main distinction being that the phthalate syndrome includes testicular, epididymal, and gubernacular cord agenesis [40,47,51]. These lesions are rarely seen in the AR antagonist syndrome even when all the males display hypospadias. In the AR-antagonist syndrome, DHT-dependent tissues are more severely than T-dependent tissues and the dose–response curves display an orderly cascade of effects in the male offspring (listed in order of low to high dosage levels of 3–200 mg vinclozolin kg/d) including reduced anogenital distance, retained areolas, reduced androgen-dependent organ weights and retained nipples, hypospadias, vaginal pouch, agenesis of the sex accessory glands, and undescended testes) [15,39,44,50,52]. In the risk assessment on vinclozolin, the acute dietary risk was based upon a permanent reduction in ventral prostate weight (retaining an extra 10 X factor under FQPA). Of the all the chemicals that interfere with androgen signaling pathway in the fetal male rat, only the phthalates affect Leydig cell insl-3 peptide hormone synthesis and cause undescended testes with gubernacular agenesis [53].

When conducting a transgenerational study, one must evaluate all the components of a syndrome so that affected animals are not misidentified and classified as "normal" and the data should be summarized in a manner that clearly delineates the proportion of animals that are affected, indicating that they display any lesion (histological or gross pathology) consistent with the syndrome. Teratology/developmental toxicity studies typically present and analyze data in this manner, indicating the number malformed/number observed on an individual and litter basis, while multigenerational studies frequently do not, even when clear teratogenic and other developmental responses are noted postnatally. In the multigeneration study, one cannot easily determine how many animals are adversely affected because the histopathological data (usually on limited numbers of F1 animals from some, but not all of the treatments) are not analyzed with the gross pathology data. It is important to note that the endpoints altered by exposure to classes of antiandrogens are not necessarily as sensitive to disruption by other important endocrine mechanisms. For EDCs that display androgenic or antithyroid activity, alter steroidogenesis, estrogenic or display antiestrogenic activity, the syndromes (diagnostic profiles of developmental effects) will differ markedly from those described for the AR-antagonist and phthalate syndromes.

As discussed above, if a chemical produced positive results in T1S and T1-Repeat (as described in Fig. 1) it would proceed to T2T in "enhanced" multigenerational tests. Multigenerational protocols are used in T2 because these are the only protocols that expose the animals during all critical stages of development and examine reproductive function of offspring after they mature (Fig. 5). In contrast, the developmental toxicity/teratology test is too insensitive to be used to detect the reproductive effects of EDCs (Fig. 6) because it only examines fetal animals a period when most reproductive tract malformations cannot be detected by a gross examination. For example, trenbolone-induced vaginal agenesis and linuron-induced epididymal agenesis were not detected in developmental toxicity tests on these chemicals. This can be problematic because this assay is often used by the USEPA to establish NOAELs, RfDs, and aPADs for acute dietary exposures.

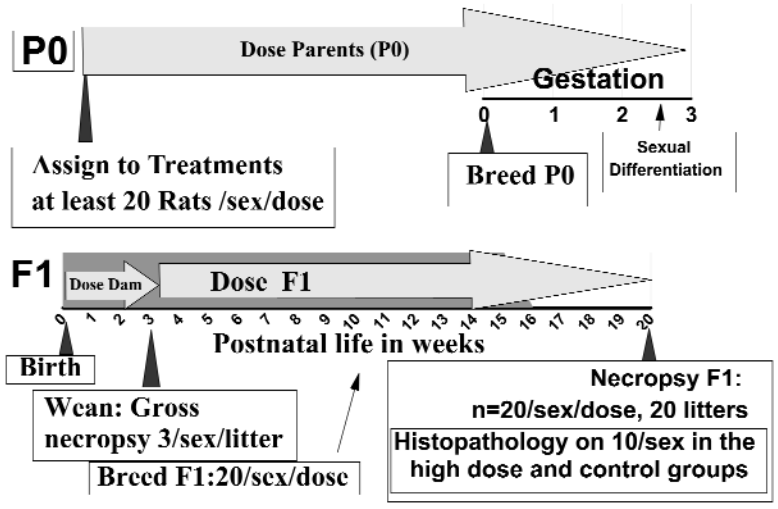

Fig. 5 A brief description of the standard USEPA 1996 multigenerational reproductive test guideline. This protocol uses more litters in the P0 generation than does the transgenerational study (Fig. 5), using at least 20 per sex per dose group. Litters can be standardized as neonates. At weaning three males and females per litter are examined for gross reproductive lesions, one animal per sex per litter is retained for further assessment (with dosing continued, usually oral, either gavage or dietary) and the rest of the pups are discarded. After generation of the F1a, and in some cases and F1b or F1c, the P0 animals are necropsied. Puberty, and fertility are assessed in the F1 and they too are necropsied (20 per sex per dose). Histology of the reproductive organs is required only on 10 animals per sex from the control and high-dose groups. If puberty was altered in the F1 generation, anogenital distance is measured in the F2. F2 pups are usually discarded at weaning.

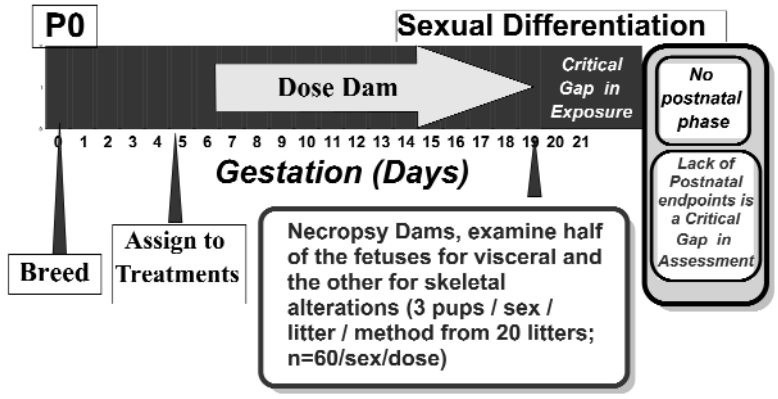

Fig. 6 General USEPA teratology/developmental toxicology test guideline study. Pregnant rats are dosed after implantation to a day or so before parturition. Previous to the current updated test guidelines, as found in most publications and pesticide studies submitted to the agency, dosing was terminated at day 15 of gestation. After dosing is terminated, the dams are euthanized, fetuses removed, and all the pups are examined for malformations, with one-half being used for visceral examination and the other half undergoing skeletal exam. These studies typically do not include any postnatal assessment of the exposed F1, while some studies examine the F1 animals during neonatal life.

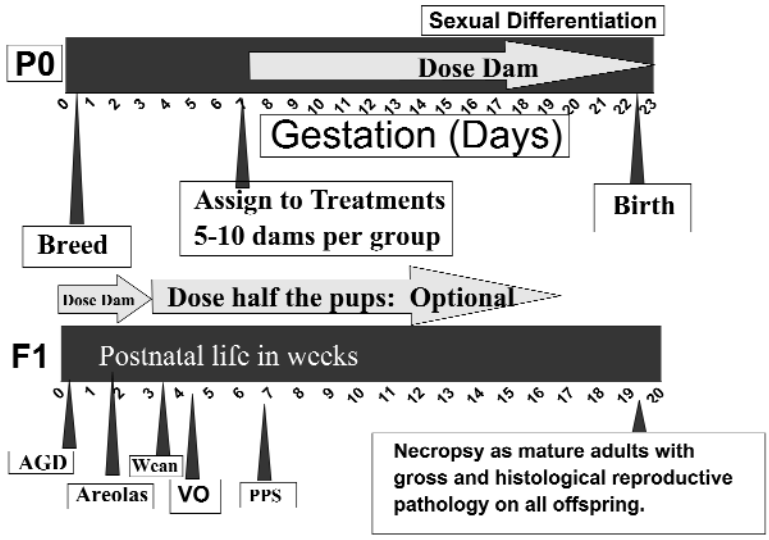

Fig. 7 General "transgenerational protocol". These protocols vary slightly from laboratory to laboratory and sometimes from study to study within a laboratory with regard to (i) the specific day of gestation that dosing is initiated, (ii) the day of gestation or lactation on which dosing is terminated, (iii) whether or not dosing is continued after weaning, with the F1 being directly exposed (rarely so), and (iv) the numbers of litters used in each dose group (5–10 is typical). However, in general the exposure is oral gavage and always includes the period of fetal sexual differentiation, the litters are not culled or standardized and all of the pups of the affected sex(es) are examined through puberty, young adulthood (5–9 months of age) and all individuals are examined thoroughly (gross morphology, histology, and weights of reproductive tissues) at necropsy. These tests are tailored to emphasize the endpoints known to be affected or suspected to be target tissues, based on other in vitro or in vivo screening studies.

TRANSGENERATIONAL TESTING FOR ANTIANDROGENS AND ANDROGENS

While the new USEPA multigenerational test provides for a comprehensive evaluation of the P0 or parental generation, too few F1 animals (offspring with developmental exposure) are examined after maturity to detect anything but the most profound reproductive teratogens [14,54]. While P0 animals within a dose group typically respond in a similar fashion to the chemical exposure, developmental events can be "all or none" with just a few animals displaying reproductive malformations in the lower dosage groups. Thus, a standard multigenerational protocol which examines one F1 animal per sex, per litter after maturity from 20 litters per dosage group can only detect statistically significant alterations if they are displayed by 25 % or more of the offspring. Histopathological alterations would have to be displayed by 50 % or more of the offspring since a histological examination of the reproductive tract is only required in 10 F1 animals per sex, per dose, far too few to detect anything but the most profound effects on reproductive development. As shown in Fig. 8, we have found that the antiandrogenic effects of linuron (epididymal and testicular hypoplasia) and vinclozolin (retained nipples) are displayed in the low-dose groups at rates well below these levels and, hence, would either go undetected at necropsy or would be described as nonsignificant. In a study of the phthalate ester di-*iso*-nonyl phthalate (DINP) [47], we found that a high dose administered to the pregnant dam induced a significant incidence of severe malformations in only 7 % of the male offspring (based upon effects consistent with the phthalate syndrome). In this regard, there is considerable uncertainty associated with LOAELs and NOAELS identified in multigenerational studies of phthalate esters that do not include either a complete assessment of the endpoints that constitute the phthalate syndrome or those included in the USEPA's new multigenerational test guidelines [55,56].

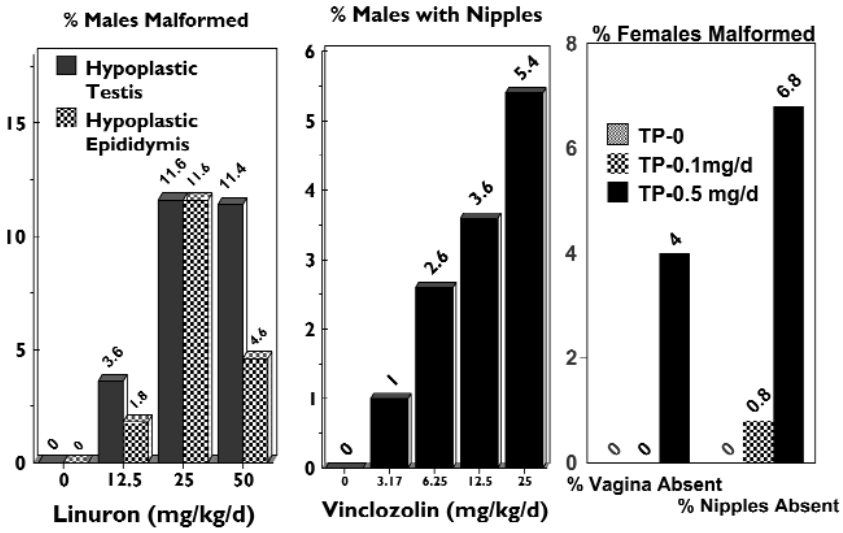

Fig. 8 In the standard multigenerational reproduction test, malformations induced by low-dosage levels of EDCs may go undetected (not observed and/or not being statistically significant) because too few F1 animals are thoroughly examined after reproductive maturity. Presented are examples of malformations that we have found in our studies that occur at a low incidence in lower-dosage groups. We are concerned that such effects will go undetected with a gross examination of only 20 F1 adult animals per sex per group and a histological examination of only 10 F1 adult animals per sex from the control and high-dose groups. Treatment with an androgen or antiandrogens induces low levels of malformations in the lower-dosage groups. [Data from refs. 14,40,61.]

When the female rat fetus is exposed to testosterone in utero, agenesis of the vagina and nipples also are seen at a low rate in the lower-dosage groups (Fig. 8). In fact, most of the low-dose effects of androgens in the female offspring (retained prostatic and vesicular tissues, nipple agenesis) are effects that are likely to be missed in a standard necropsy. There are additional factors besides detection of adverse effects at necropsy or during data analysis and interpretation that limit interpretation of data from the standard multigenerational reproduction test. The life-long exposure of both males and females in the F1 generation, which allows one to detect effects induced in utero, during lactation or from direct exposure after puberty, can confound the identification of when the effect was induced, i.e., during adulthood versus development or even the affected sex. For example, the current RED on linuron exemplifies how the registrant [57] and the USEPA (RED, 1995, TRED, 2002) have misinterpreted the stage of life and the mode of action for the effects of this pesticide on the testis and epididymis. This confusion arose from the fact that the F1 animals were dosed continuously and the reproductive effects, though actually induced in utero [6,14,40,58,59], were incorrectly attributed to a direct effect of linuron in the adult via increased serum LH and Leydig cell hyperplasia. In fact, linuron does not increase serum LH or induce Leydig cell hyperplasia in adult male rats [40] and the testicular effects result from pressure atrophy after puberty as the Sertoli cells form tight junctions and the fluid secreted into the lumen of the tubules is unable to flow normally out of the testis in animals presenting epididymal agenesis [58,59]. In our studies, the dosing period normally is terminated near birth, or at weaning which precludes misinterpretation of the developmental origin of reproductive effects. For DBP, the effects on fertility in P0 female rats is obscured by the more obvious effects on testis histology and sperm numbers in their breeding partners [40]. On the other hand, it is clear that our transgenerational protocols would not be appropriate for EDCs that induced low-dose alterations in the pubertal or adult animal. It is ironic that the teratology test examines all the animals carefully in a litter, but at an inappropriate stage of life, while the multigenerational test, which includes exposure during sexual differentiation and then examines these offspring as adults, discards most of the F1 animals before maturity.

In our "transgenerational" protocols [14,39,43–47,54,59], we typically use fewer litters (7–10 per dose group) but examine all the animals in each litter. A detailed methods paper describing how we execute this protocol is now available [60]. These protocols actually use fewer animals, but provide more statistical power to detect reproductive effects in the F1 generation. In addition, the measurement of anogenital distance (AGD) at 1–2 days of age and areolas/nipples at 12–13 days of age in the F1 generation is included. These endpoints are not only extremely useful "biomarkers" of effect, but the changes in AGD and nipples can be permanent. Moreover, these endpoints are highly correlated with malformations and other alterations in androgen-dependent tissues [59; Hotchkiss et al., in preparation]. Unfortunately, neither of these markers are measured in the F1 generation under the 1996 USEPA multigenerational test guidelines. In this regard, we propose that testing protocols should be tailored based upon the results of T1S such that more F1 male offspring are examined thoroughly (Table 1) as adults than is now required if the chemical appears to be antiandrogenic, while the examination of F1 females can be similarly enhanced (Table 2) for androgens.

It is important to reiterate that the above endpoints that are sensitive to antiandrogens or androgens in utero have not been shown to be sensitive to xenoestrogens or some other EDCs. We are not suggesting that all EDCs should be tested like the androgens or antiandrogens. Testing should be tailored based upon the pharmacological activity demonstrated in T1S. In addition, the developing fetus is not always the most sensitive life stage. Some EDCs, like fenarimol, ketoconazole, and other inhibitors of steroidogenesis, disrupt pregnancy by altering maternal maternal endocrine function in P0 dams at dosage levels that appear to be without direct effect on the offspring. In such cases, the standard USEPA multigenerational protocol, with minor enhancements, or a transgenerational protocol with exposure continued after weaning should be employed. For example, with xenoestrogens, the most consistent low-dose effect seen in rat multigenerational studies has been pseudoprecocious puberty in the female offspring. We believe that the transgenerational or in utero/lactational protocols fill a gap in the testing program for EDCs that should be used only on a case-by-case basis, as indicated by the results of T1S and any T1-Repeat. In cases where one is uncertain about how T2 should be tailored, a transgenerational study, including dosing of half the F1 after weaning, could be employed.

Table 1 Detecting antiandrogen syndromes.

Multitude of effects of antiandrogens in male rat offspring that should be evaluated in Tier 2 testing studies displaying this activity in Tier 1 screening or other assays.

Neonatal-infantile data

1. Anogenital distance at birth (1–3 days of age)
2. Areola/nipple retention in infant male rats at 13–14 days of age

External necropsy endpoints on all male rat offspring at maturity

1. Body weight any unusual malformations or anomalies, euthanize
2. Shave ventral surface from inguinal region to neck and count nipples and areolas (observer blind to treatment), record position of areolas and nipples
3. Check animals for hypospadias, epispadias, cleft phallus, and measure AGD
4. Note if testes obviously undescended
5. Note if inguinal region soiled with urine
6. Note if prepuce partially or entirely detached from glans penis, especially if a persistent thread of tissue is present along frenulum

Internal endpoints on all male rat offspring at maturity

1. Location of each testis (scrotal, abdominal, gubernaclum attached to abdominal wall)
2. Gubernacular cords, present or absent, and length in mm if abnormal
3. Note if present, cranial suspensory ligaments
4. Note if testes are small, absent, fluid filled, enlarged, appear infected or other
5. Note if epididymides are small, absent, or infected (record region of effects)
6. Note if ventral prostate is small, absent, or infected
7. Note if dorsolateral prostate is small, absent, or infected
8. Note if seminal vesicles are small, absent, infected, or one side larger than the other
9. Note if coagulating glands are small absent, infected, one side larger than the other or detached from seminal vesicles
10. Note if kidneys display hydronephrosis, calcium deposits
11. Note presence of hydroureter
12. Note presence of bladder stones or bloody in bladder

Weigh the following organs on all male rat offspring at maturity

1. Each testis individually (examine histology of each testis)
2. Each corpus plus caput epididymis (examine histology of each segment)
3. Each cauda epididymis (examine histology of each segment)
4. Entire seminal vesicle, plus coagulating glands with fluid as a unit, if possible
5. Entire ventral prostate, if possible
6. Each kidney
7. Paired adrenals
8. Liver
9. Levator ani plus bulbocavernosus
10. Cowper's glands as a pair, if possible
11. Glans penis
12. Pituitary
13. Brain

Histology on all male F1 offspring at maturity

1. Both testes
2. Both epididymides
3. Prostate glands
4. Any grossly abnormal reproductive tissues

Table 2 Detecting the androgenized female syndrome [61].

Multitude of effects of androgens in female rat offspring that should be evaluated in Tier 2 testing studies

Neonatal-infantile data in female rat offspring
1. Anogenital distance at birth (1–3 days of age)
2. Areola/nipple agenesis (complete or faint) in infant female rats at 13–14 days of age

External necropsy endpoints on all female rat offspring at maturity
1. Body weight any unusual malformations or anomalies, euthanize
2. Shave ventral surface from inguinal region to neck and count nipples and areolas (observer blind to treatment), record position of areolas and nipples
3. Check animals for cleft phallus and measure AGD and position of vaginal opening
4. Note if inguinal region soiled with urine

Internal endpoints on all female rat offspring at maturity
1. Location of ovaries in relationship to kidneys
2. Note if absent, cranial suspensory ligaments
3. Note if ovaries are small, cystic-fluid filled, enlarged, appear infected or other
4. Note if oviducts, uterus, or upper or lower vagina are small, absent, or infected (record region of effects) or fluid-filled
5. Note if ventral prostate tissue is present
6. Note if seminal vesicle tissue is present
7. Note if levator ani/bulbocavernosus muscle tissues are present
8. Note if other male tissues are present
9. Note if kidneys display hydronephrosis, calcium deposits
10. Note presence of hydroureter
11. Note presence of bladder stones or blood in bladder

Weigh the following organs in all female offspring at maturity
1. Paired ovaries (histology)
2. Uterus with fluid (histology)
3. Vagina
4. Each kidney
5. Paired adrenals
6. Liver
7. Pituitary
8. Brain

Histology on all female rat offspring at maturity
1. Both ovaries
2. Uterus
3. Vagina
4. Any suspected male reproductive tissues
5. Any grossly abnormal reproductive tissues

REFERENCES

1. T. Colborn and C. Clement. *Advances in Modern Environmental Toxicology. Vol XXI. Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Princeton Scientific Publishing, Princeton, NJ (1992).
2. T. Colborn, H. Bern, P. Blair et al. *Toxicol. Ind. Health* **14**, 1–8 (1998).
3. H. R. Andersen, A. M. Andersson, S. F. Arnold, H. Autrup, M. Barfoed, N. A. Beresford, P. Bjerregaard, L. B. Christiansen, B. Gissel, R. Hummel, E. B. Jorgensen, B. Korsgaard, R. Le Guevel, H. Leffers, J. McLachlan, A. Moller, J. B. Nielsen, N. Olea, A. Oles-Karasko, F. Pakdel, K. L. Pedersen, P. Perez, N. E. Skakkebaek, C. Sonnenschein, A. M. Soto et al. *Environ. Health Perspect*. **107** (Suppl. 1), 89–108 1999.
4. E. R. Bauer, A. Daxenberger, T. Petri, H. Sauerwein, H. H. Meyer. *APMIS.* **108** (12), 838–846 (2000).
5. C. Wong, W. R. Kelce, M. Sar, E. M. Wilson. *J. Biol. Chem*. **25; 270** (34), 19998–20003 (1995).
6. C. Lambright, J. Ostby, K. Bobseine, V. Wilson, A. K. Hotchkiss, P. C. Mann, L. E. Gray, Jr. *Toxicol. Sci.* **56** (2), 389–399 (2000).
7. V. S. Wilson, K. Bobseine, C. R. Lambright, L. E. Gray, Jr. *Toxicol. Sci.* **66** (1), 69–81 (2002).
8. C. Sultan, F. Paris, B. Terouanne, P. Balaguer, V. George, N. Poujol, S. Jeandel, J. Nicolas. *Horm. Reprod. Update* **7** (3), 314–322 (2001).
9. J. Legler, C. E. van den Brink, A. Brouwer, A. J. Murk, P. T. van der Saag, A. D. Vethaak, B. van der Burg. *Toxicol. Sci*. **48** (1), 55–66 (1999).
10. J. M. Rogers and M. S. Denison. *In Vitro Mol. Toxicol*. **13** (1), 67–82 (2000).
11. A. M. Vinggaard, V. Breinholt, J. C. Larsen. *Food Addit. Contam*. **16** (12), 533–542 (1999).
12. P. C. Hartig, K. L. Bobseine, B. H. Britt, M. C. Cardon, C. R. Lambright, V. S. Wilson, L. E. Gray, Jr. *Toxicol. Sci*. **66** (1), 82–90 (2002).
13. W. R. Kelce, E. Monosson, M. P. Gamcsik, S. C. Laws, L. E. Gray, Jr. *Toxicol. Appl. Pharmacol.* **126** (2), 276–285 (1994).
14. B. S. McIntyre, N. J. Barlow, D. G. Wallace, S. C. Maness, K. W. Gaido, P. M. D. Foster. *Toxicol. Appl. Pharmacol.* **167** (2), 87–99 (2000).
15. L. E. Gray, Jr., W. R. Kelce, T. Wiese, R. Tyl, K. Gaido, J. Cook, G. Klinefelter, D. Desaulniers, E. Wilson, T. Zacharewski, C. Waller, P. Foster, J. Laskey, J. Reel, J. Giesy, S. Laws, J. McLachlan, W. Breslin, R. Cooper, R. Di Giulio, R. Johnson, R. Purdy, E. Mihaich, S. Safe, T. Colborn et al. *Reprod. Toxicol*. **11** (5), 719–50 (1997).
16. P. Mak, F. D. Cruz, S. Chen. *Environ, Health Perspect*. **107** (11), 855–860 (1999).
17. L. E. Gray, Jr. *Toxicol. Lett*. **102–103**, 677–680 (1998).
18. L. E. Gray, Jr., J. Ostby, V. Wilson, C. Lambright, K. Bobseine, P. Hartig, A. Hotchkiss, C. Wolf, J. Furr, M. Price, L. Parks, R. L. Cooper, T. E. Stoker, S. C. Laws, S. J. Degitz, K M. Jensen, M. D. Kahl, J. J. Korte, E. A. Makynen, J. E. Tietge, G. T. Ankley. *Xenoendocrine Disrupters Tiered Screening and Testing: Filling Key Data Gaps Toxicology*, Dec 27, **181–182**, 371–382. Review (2002).
19. G. Ankley, E. Mahiach, R. Stahl, D. Tillitt, T. Colborn, S. McMaster, R. Miller, J. Bantle, P. Campell, R. Dickerson, M. Fry, J. Giesy, L. E. Gray, Jr. et al. *Environ. Toxicol. Chem*. **17**, 68–87 (1998).
20. G. T. Ankley, K. M. Jensen, M. D. Kahl, J. J. Korte, E. A. Makynen. *Environ. Toxicol. Chem.* **20**, 1276–1290 (2001).
21. G. T. Ankley, M. D. Kahl, K. M. Jensen, M. W. Hornung, J. J. Korte, E. A. Makynen, R. L. Leino. *Toxicol. Sci.* **67** (1), 121–130 (2002).
22. J. Kanno, L. Onyon, J. Haseman, F. Fenner-Crisp, J. Ashby, W. Owens, Organization for Economic Cooperation and Development. *Environ. Health Perspect*. **109** (8), 785–794 (2001).

23. J. Kanno, L. Onyon, S. Peddada, J. Ashby, E. Jacob, W. Owens. *Environ. Health Perspect*. **111** (12), 1550–1558 (2003).
24. J. Kanno, L. Onyon, S. Peddada, J. Ashby, E. Jacob, W. Owens. *Environ. Health Perspect.* **111** (12), 1530–1549 (2003).
25. L. Hershberger, E. Shipley, R. Meyer. *Proc. Soc. Exp. Biol. Med.* **83**, 175–180 (1953).
26. R. Dorfman. "Androgens and anabolic agents", in *Methods in Hormone Research, Vol. II, Bioassays*, R. I. Dorfman (Ed.), pp. 275–313, Academic Press, New York (1962).
27. J. C. O'Connor, S. R. Frame, L. G. Davis, J. C. Cook. *Toxicol. Sci*. **51** (1), 44–53 (1999).
28. J. M. Goldman, S. C. Laws, S. K. Balchak, R. L. Cooper, R. J. Kavlock. *Crit. Rev. Toxicol*. **30** (2), 135–196 (2000).
29. L. E. Gray Jr, J. Ostby, R. Sigmonn, J. Ferrell, G. Rehnberg, R. Linder, R. Cooper, J. Goldman, J. Laskey. *Reprod. Toxicol*. **2**, 281–287 (1988).
30. L. E. Gray, Jr., J. Ostby, J. Ferrell, G. Rehnberg, R. Linder, R. Cooper, J. Goldman, V. Slott, J. Laskey. *Fundam. Appl. Toxicol.* **12**, 92–108 (1989).
31. S. C. Laws, J. M. Ferrell, T. E. Stoker, J. Schmid, R. L. Cooper. *Toxicol. Sci*. **58** (2), 366–376 (2000).
32. M. S. Marty, J. W. Crissman, E. W. Carney. *Toxicol. Sci*. **52** (2), 269–277 (1999).
33. W. R. Kelce, C. R. Stone, S. C. Laws, L. E. Gray, J. A. Kemppainen, E. M. Wilson. *Nature* **375**, 581–585 (1995).
34. J. Ashby and P. A. Lefevre. *Regul. Toxicol. Pharmacol*. **26**, 330–337 (1997).
35. E. Monosson, W. R. Kelce, C. Lambright, J. Ostby, L. E. Gray, Jr. *Toxicol. Ind. Health* **15** (1–2), 65–79 (1999).
36. M. S. Marty, J. W. Crissman, E. W. Carney. *Toxicol. Sci*. **60** (1), 63–76 (2001).
37. T. E. Stoker, L. G. Parks, L. E. Gray, R. L. Cooper. *Crit. Rev. Toxicol*. **30** (2), 197–252 (2000).
38. W. Kelce, C. Lambright, L. Gray, K. Roberts. *Toxicol. Appl. Pharmacol*. **142**, 192–200 (1997).
39. L. E. Gray, Jr., J. S. Ostby, W. R. Kelce. *Toxicol. Appl. Pharmacol*. **129** (1), 46–52 (1994).
40. L. E. Gray, Jr., C. Wolf, C. Lambright, P. Mann, M. Price, R. L. Cooper, J. S. Ostby. *J. Toxicol. Ind. Health* **15** (1–2), 94–118 (1999).
41. P. M. D. Foster and B S. McIntyre. *Toxicol. Pathol*. **30** (1), 59–65 (2002).
42. L. E. Gray, J. Ostby, J. Furr, C. J. Wolf, C. Lambright, L. Parks, D. N. Veeramachaneni, V. Wilson, M. Price, A. Hotchkiss, E. Orlando, L. Guillette. *Hum. Reprod. Update*. **7** (3), 248–264 (2001).
43. E. Mylchreest, R. C. Cattley, P. M. D. Foster. *Toxicol. Sci.* **43** (1), 47–60 (1998).
44. E. Mylchreest, M. Sar, R. C. Cattley, P. M. D. Foster. *Toxicol. Appl. Pharmacol*. **15; 156** (2), 81–95 (1999).
45. E. Mylchreest, D. G. Wallace, R. C. Cattley, P. M. D Foster. *Toxicol. Sci*. **55** (1), 143–151 (2000).
46. E. Mylchreest, M. Sar, D. G. Wallace, P. M. D. Foster. *Reprod. Toxicol*. **16** (1), 19–28 (2002).
47. L. E. Gray, Jr., J. Ostby, J. Furr, M. Price, D. N. Veeramachaneni, L. Parks. *Toxicol. Sci.* **58** (2), 350–365 (2000).
48. V. D. Shultz, S. Phillips, M. Sar, P. M. D. Foster, K. W. Gaido. *Toxicol. Sci*. **64** (2), 233–242 (2001).
49. L. G. Parks, J. S. Ostby, C. R. Lambright, B. D. Abbott, G. R. Klinefelter, N. J. Barlow, L. E. Gray, Jr. *Toxicol. Sci.* **58** (2), 339–249 (2000).
50. B. S. McIntyre, N. J. Barlow, P. M. Foster. *Toxicol. Sci*. **62** (2), 236–249 (2001).
51. P. M. D. Foster, E. Mylchreest, K. W. Gaido, M. Sar. *Hum. Reprod. Update* **7** (3), 231–235 Review (2001).
52. J. Ostby, W. R. Kelce, C. Lambright, C. J. Wolf, P. Mann, L. E. Gray, Jr. *Toxicol. Ind. Health* **15** (1–2), 80–93 (1999).
53. V. S. Wilson, C. Lambright, J. Furr, J. Ostby, C. Wood, G. Held, L. E. Gray, Jr. *Toxicol. Lett.* (2003). In press.

54. P. M. D. Foster, R. C. Cattley, E. Mylchreest. *Food Chem. Toxicol*. **38** (1 Suppl.), S97–99, Review (2000).
55. R. H. McKee, R. T. Przygoda, M. A. Chirdon, G. Engelhardt, M. Stanley. *J. Appl. Toxicol*. **20**, 491–497 (2000).
56. L. J. Hushka, S. J. Waterman, L. H. Keller, G. W. Trimmer, J. J. Freeman, J. L. Ambroso, M. Nicolich, R. H. McKee. *Reprod. Toxicol*. **15** (2), 153–169 (2001).
57. J. C. Cook, L. S. Mullin, S. R. Frame, L. B. Biegel. *Toxicol. Appl. Pharmacol*. **119** (2), 195–204 (1993).
58. B. S. McIntyre, N. J. Barlow, M. Sar, D. G. Wallace, P. M. D. Foster. *Reprod. Toxicol*. **16** (2), 131–139 (2002).
59. B. S. McIntyre, N. J. Barlow, P. M. D. Foster. *Toxicol. Sci*. **65** (1), 62–70 (2002).
60. J. Ostby and L. E. Gray. *Curr. Protocol. Toxicol*. In press.
61. C. J. Wolf, A. Hotchkiss, J. S. Ostby, G. A. LeBlanc, L. E. Gray, Jr. *Toxicol. Sci.* **65** (1), 71–86 (2002).