Regulatory Toxicology and Pharmacology 59 (2011) 397–411

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Evaluation of EPA's Tier 1 Endocrine Screening Battery and recommendations for improving the interpretation of screening results

Christopher J. Borgert [a,b,*], Ellen M. Mihaich [c], Terry F. Quill [d], M.S. Marty [e], Steven L. Levine [f], Richard A. Becker [g]

[a] Applied Pharmacology and Toxicology, Inc., 2250 NW 24th Avenue, Gainesville, FL 32605, USA
[b] Center for Environmental and Human Toxicology, Dept. of Physiological Sciences, University of Florida College of Veterinary Medicine, Gainesville, FL 32610, USA
[c] Environmental and Regulatory Resources, LLC, Durham, NC, USA
[d] Quill Law Group, LLC, Washington, DC, USA
[e] The Dow Chemical Company, Midland, MI, USA
[f] Monsanto Company, St. Louis, MO, USA
[g] American Chemistry Council, Washington, DC, USA

ARTICLE INFO

Article history:
Received 31 August 2010
Available online 18 January 2011

Keywords:
Endocrine disrupter
EDSP screening
Screening assays
Tier 1 assays
Endocrine Screening Battery
Estrogen assay
Thyroid assay
Androgen assay
Pubertal assay
Validation

ABSTRACT

EPA's Endocrine Disruptor Screening Program (EDSP) was implemented in 2009–2010 with the issuance of test orders requiring manufacturers and registrants of 58 pesticide active ingredients and nine pesticide inert/high production volume chemicals to evaluate the potential of these chemicals to interact with the estrogen, androgen and thyroid hormone systems. The required endocrine screening will be conducted over the next 2–3 years. Based on estimates of the impacted sectors, costs are at least $750,000–$1,000,000 per substance if all of the Tier 1 assays must be conducted. The screening will entail evaluation of responses in EPA's Tier 1 Endocrine Screening Battery (EDSP ESB), consisting of 11 distinct *in vitro* and *in vivo* assays. We reviewed the details of each test method and describe the critical factors integral to the design and conduct of the EDSP ESB assays as well as the limitations related to specificity and sensitivity. We discuss challenges to evaluating each assay, identify significant shortcomings, and make recommendations to enhance interpretation of results. Factors that affect the length of time necessary to complete the EDSP ESB for any particular substance are presented, and based on the overall analysis, we recommend a sequence for running the EDSP ESB assays. It is imperative that a structured, systematic weight of evidence framework is promptly developed, subjected to peer review and adopted. This will help to ensure an objective analysis of the results of the required EDSP screening, consistent integration of results across the EDSP ESB assays, and consistent decision making as to whether subsequent testing for adverse effects is needed. Based upon the limitations of the current EPA EDSP ESB, we concur with the Agency's Scientific Advisory Panel's recommendation that after the initial set of substances has been screened, the EDSP ESB should pause so that the results can be fully analyzed to determine the value of the existing assays. After this analysis, assays that are unnecessarily redundant or that lack endocrine specificity should be eliminated and if necessary, replaced by new or revised screens that are more mechanistically specific, rapid, reliable, and cost effective.

1. Introduction

In response to the public concern that certain environmental chemicals may interfere with endocrine processes in humans, the US Congress enacted Section 408(p) of the 1996 Food Quality Protection Act (FQPA or "the Act") (US EPA, 1996) directing the US Environmental Protection Agency (EPA) to develop and implement a screening program using "validated test systems" to investigate the potential of chemicals to induce adverse health effects through endocrine pathways. Relevant assays available in 1996 varied significantly in their degree of development and validation and a screening battery had not yet been identified or adopted by EPA (EDSTAC, 1998). Several screens had an extensive history, e.g., the uterotrophic and the Hershberger screens, but others were only partially developed or were only hypothetically useful as screens, e.g., the amphibian developmental screen and the fish gonadal recrudescence screen.

Developing and validating endocrine assays and developing a screening program for endocrine-related mechanistic responses have proven to be difficult and time-consuming endeavors for

* Corresponding author at: Applied Pharmacology and Toxicology, Inc., 2250 NW 24th Avenue, Gainesville, FL 32605, USA.
E-mail address: cjborgert@apt-pharmatox.com (C.J. Borgert).

0273-2300/$ - see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.yrtph.2011.01.003

EPA. The validation of these methods coincided with the official adoption by both the US and OECD of specific and detailed guidance for test method validation and regulatory acceptance (ICCVAM, 1997; Schechtman, 2007). Both the USEPA and the OECD agreed to follow established guidance for the validation of alternative assays. The US guidance was developed by the Interagency Coordination Committee on the Validation of Alternative Methods (ICCVAM). The fundamental validation principles are to clearly state the purpose and biological basis for the assay and to verify the performance of the assay against validation criteria using a common set of test chemicals across multiple laboratories. The assay's performance is defined as its ability to correctly identify positive chemicals, where sensitivity is the rate of correct positive findings, and negative chemicals, where specificity is the rate of correct negative findings. Additional principles include using representative target chemicals of the assay, blinding the test substances to avoid laboratory bias, the use of a common protocol among laboratories, availability of data, and peer review of the findings. The simultaneous conduct of endocrine methods validation studies with the formalization of method validation criteria created a dynamic tension and certain challenges as a number of the validation studies of the endocrine screening methods, some of which were well underway when the OECD guidance was adopted, were the first to be considered in the lens of these new validation criteria.

Thirteen years after passage of the FQPA, EPA began issuing the first set of testing orders for the Endocrine Disruptor Screening Program (EDSP). Even as EPA proceeds with requiring EDSP screening, the utility, validation, and interpretability of specific assays and of the EDSP Tier 1 Endocrine Screening Battery (EDSP ESB) as a whole continue to be debated (US EPA, 2008a). Despite great efforts by EPA, the scientific community and regulated parties, numerous questions and uncertainties remain as to the usefulness and limitations of the specific assays selected by EPA and of the EDSP ESB. Understanding these strengths and limitations is critical for interpretation of the screening results and for decision making based on those results.

The purpose of this paper is to critically review the EDSP battery. The review will identify which assays are or are not sufficiently robust to clearly identify chemicals interacting with endocrine mechanisms. One focus will be the extensive problems the anticipated high rate of false positives presents to regulatory interpretation of the assays. This is because the screening results are collectively intended to identify chemicals for which subsequent Tier 2 testing is necessary. Tier 2 tests could include a host of expensive assays in terms of cost, time and animal use, such as rodent reproduction and development assays and fish full life cycle assays. Limitations related to assay specificity and sensitivity, and to interpreting assay results are discussed because these issues are important to consider when both planning and interpreting EDSP screening. Although EPA has provided formal test guidelines for each individual EDSP screening assay (US EPA, 2009a), it has not yet provided specific guidance on interpreting results of the individual assays or a much-needed, detailed and specific guidance on using weight of evidence to evaluate the overall EDSP ESB results. Understanding the strengths and weaknesses of the overall EDSP ESB may help determine whether enhancing certain study protocols beyond basic regulatory requirements could provide greater insight when evaluating results. Enhancements might include increasing the number of dose groups, adjusting the number of animals per dose group and including positive, negative and/or pair-fed controls, with due consideration given to animal welfare. In addition, based on the strengths and limitations of specific assays, we suggest a sensible sequence for conducting the assays in the EDSP ESB that might help with interpreting results obtained when testing chemicals in specific assays and when evaluating a chemical's profile across the EDSP ESB as a whole.

The effort to develop, standardize and validate endocrine screens and tests has been challenging and time-consuming. The time that has elapsed from passage of the FQPA in 1996 to EPA's issuance of EDSP test orders has led some to voice concern that potential endocrine-related adverse effects have gone unregulated. Although EPA has just implemented the EDSP, significant scientific data on potential endocrine activity, or lack thereof, are already available in the open scientific literature, government authoritative reviews or in pesticide registration assessments available on the Internet for a number of high profile chemicals; including phthalate esters (David, 2006), bisphenol A (Birth Defects Research (Part B) 83: 157–395 (2008) NTP-CERHR Expert Panel Report on the Reproductive and Developmental Toxicity of Bisphenol A), conazole pesticides (http://www.epa.gov/oppsrrd1/REDs/propiconazole_red.pdf; http://www.epa.gov/hhrp/files/2009_posters/ltg_1-01_nesnow.pdf) triazine pesticides (http://www.epa.gov/oppsrrd1/reregistration/status_triazines.htm) and a number of other agents employed in standardizing and validating endocrine screens and tests (Freyberger et al., 2007; Freyberger and Schladt, 2009; Kanno et al., 2001, 2003a,b; Laws et al., 2006; O'Connor et al., 1996, 1999a,b, 2000, 2002a,b; Owens and Ashby, 2002; Owens and Koëter, 2003; Owens et al., 2003, 2006, 2007, plus many of the EPA and OECD documents included in the Reference section).

1.1. EDSP development

When promulgating the FQPA in 1996, Congress apparently believed that EPA could quickly and easily develop an inexpensive, fast, accurate and validated "estrogenic substance screening program" because it imposed a two-year timeframe for development using validated test systems and other scientifically relevant information, with implementation to occur not later than 3 years after enactment of the law. In the Act, Congress focused narrowly on screening pesticide chemicals for estrogenic effects in humans, although it allowed EPA to require additional testing of other substances and for other endocrine effects. Developing an endocrine screening program, however, proved to be neither quick nor easy, and the program will be very expensive and time consuming as it is currently being conducted. Based on an analysis of the program costs conducted by industry stakeholders, the EDSP ESB alone could range from $750,000 to $1,000,000 per substance (analysis available electronically on EPA Docket; see US EPA, 2008b). Costs could be higher, if, as discussed below, it is necessary to add dose groups (e.g., positive and/or negative controls or an antagonist arm) to some EDSP ESB assays, conduct range-finding studies, and repeat some assays to clarify equivocal results. At the same time, certain EDSP ESB assays may lack accuracy and value for differentiating potential endocrine-mediated responses from responses via other modes of action (e.g., cytotoxicity in *in vitro* assays and hepatotoxicity in *in vivo* assays) or systemic toxicity. To some extent, these problems arise from difficulties in developing useful assays for the intended breadth of EPA's EDSP.

After the FQPA was enacted, EPA convened a federal advisory committee, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), to assess the current state of the science and assist the Agency in developing an endocrine screening program (The Keystone Center, 1996). EDSTAC consisted of scientists and others representing various interests, including advocates of the endocrine disruption theory and the regulated community. EDSTAC concluded that the assays necessary to determine the potential endocrine activity of chemical substances varied significantly in their degree of development and validation. At the same time, EDSTAC recommended that EPA develop an extensive program that would subject all chemicals to screening and testing