

Story of the Uterotrophic Assay
in EPA's Endocrine Disruptor Screening Program
by Gary Timm OSCP, OPPTS, EPA

Background

The uterotrophic assay is an *in vivo* assay to test for estrogenicity. It is based on the principle that the growth phase of the uterus in the natural estrous cycle is under the control of estrogens. Uterine growth during the natural estrous cycle is rapid and easily measurable within two days. When the natural source of estrogen is not available, either because the animal is immature or because it has been ovariectomized (OVX), then the growth of the uterus becomes sensitive to external sources of estrogen. When exposed to such a source, the immature or OVX animal's uterus will increase in weight due to the imbibition of fluid and cell proliferation initiated by the estrogen. Therefore, the primary endpoint in this assay is uterine weight, measured using one of two techniques: dry and wet weight. Chemicals that act as oestrogen agonists would be expected to cause a statistically significant increase in uterine weight, while oestrogen antagonists, when co-administered with a potent reference oestrogen, would be expected to decrease uterine growth.

The results from the uterotrophic assay could be used in the Endocrine Disruptor Screening Program (EDSP) screening battery, along with data from other assays, by implementing a weight-of-evidence approach to determine whether a substance should be tested further or that sufficient information is available for hazard assessment purposes.

Protocol Development

The uterotrophic assay was proposed for use as a screen for identifying potential estrogenic or anti-estrogenic substances in the final report of EPA's Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC). The assay has been in use since the 1930s for pharmaceutical discovery and evaluation of estrogens, but at the time the EDSTAC report was published, it had not yet been validated as a potential screen for weak estrogens. The uterotrophic assay has been validated wholly through the Organization for Economic Cooperation and Development (OECD). The OECD's Endocrine Disruptor Testing and Assessment working group reached a consensus to make the uterotrophic assay a high priority for validation after consideration of the recommendations of EDSTAC and OECD's Detailed Review Paper on the appraisal of test methods for sex-hormone disrupting chemicals.

Although EPA participated in the design of the validation approach, the Agency has not directly managed the development of the protocol in any way. Before the validation program commenced, OECD undertook an extensive literature review and generated a supporting Background Review Document. This document not only examined the protocol options but also the chemicals recommended for use in the validation study. The Background Review Document was circulated among the Validation Management Group for Screening and Testing of

Endocrine Disruptors for Mammalian Effects (VMG-mam) for review in July 2001, and the final, revised version was made available in scientific literature in January 2003.

The Background Review Document noted that several varying designs had been used and proposed that four candidate protocols be considered for validation:

- Protocol A: Intact, immature female rats treated by oral gavage with a dosing regimen of three consecutive days;
- Protocol B: Intact, immature female rats treated subcutaneously with a dosing regimen of three consecutive days;
- Protocol C: Mature, female OVX rats treated subcutaneously with a dosing regimen of three consecutive days; and
- Protocol D: Mature, female OVX rats treated subcutaneously with a dosing regimen of seven consecutive days.

Since no consensus could be reached on the choice of one protocol over another, it was decided that all four protocols should proceed into validation. Member countries with strong animal welfare concerns opposed using OVX females and preferred the protocol using the intact, immature females. Some countries felt that the OVX animals would be more sensitive and preferred the use of the OVX protocol. Initially it was decided to focus efforts on the subcutaneous route of exposure only, but it was subsequently agreed that the oral route should be included for comparison with subsequent work on the assay using test chemicals metabolized by different routes. The rat was chosen as the test species because it was the preferred species in reproductive toxicity testing. Finally there was a question about the optimum period of exposure. This was addressed by including protocol D in the validation studies.

Phase-1

Because of the abundant historical information available on the uterotrophic assay, the assay proceeded to two phases of inter-laboratory validation studies: Phase-1 and Phase-2. Phase-1 of validation involved 19 laboratories in Europe, Japan, Korea, and the U.S. and was completed in January 2000. The purpose of Phase-1 was to determine whether multiple laboratories could execute the protocol using a potent reference oestrogen, 17 α -ethinyl oestradiol (EE). All four protocols were tested, and between four and six laboratories undertook each protocol. The results of Phase-1 indicated acceptable agreement among laboratories with respect to the magnitudes of the responses at different dose levels and the doses at which significant responses were obtained with the two versions of the assay (i.e., immature and OVX rats).

There was, however, disagreement over which experimental conditions should be standardized during prevalidation, such as rat strain, bedding, etc. The OECD observed that each country conducting the assay would be more prone to use customary strains and animal husbandry techniques and that the protocols should be robust enough to detect oestrogen activity despite these differences. Therefore, each validation laboratory was given some latitude with respect to these parameters.

Phase-1 results were approved by the VMG-mam and were subsequently approved by the Task Force on Endocrine Disruptors Testing and Assessment (EDTA).¹ The Endocrine Disruptor Methods Validation Subcommittee (EDMVS) did not play a role in prevalidation of the uterotrophic assay, but they were kept informed of its progress. Dr. Willie Owens (Proctor & Gamble, U.S.), a member of EDMVS and OECD manager of the uterotrophic validation effort, briefed the subcommittee on the execution of the Phase-1 and Phase-2 studies.

Several parameters were optimized in Phase-1, and these minor modifications were incorporated in the subsequent Phase-2 studies. The first modification involved an increase in the age range of the immature animals to a minimum of 18 days and a maximum of 20 days at the beginning of treatment. The second modification increased the period of acclimatization of the mature rats following ovariectomy from a minimum of seven days in Phase-1 to a minimum of 14 days in Phase-2.

Phase-2

Phase-2 of validation took place between June 2000 and June 2001 and involved the testing of seven test substances by 20 laboratories in Denmark, France, Germany, Italy, Japan, Korea, U.K., and the U.S. Laboratories were both public and private. The lead laboratory was the National Institute of Health Sciences in Tokyo, Japan, and statistical support was provided by the National Institute of Environmental Health Sciences (NIEHS) in Research Triangle Park, NC.

The purpose of Phase-2 was to challenge the uterotrophic assay with a number of different compounds in order to determine whether it could reliably detect strong and weak estrogens, demonstrate the transferability of the standardized protocols between laboratories, and quantify the inter- and intra-laboratory reproducibility of the assay. A secondary aim of Phase-2 was to determine whether wet weight (i.e., weight of uterus containing luminal fluid) or dry weight (i.e., weight of uterus with luminal fluid removed) provided the most reliable data, or whether either method could be used. The following experiments were conducted in order to examine the reliability and sensitivity of the four protocols:

¹ The EDTA Task Force was established to provide a focal point within OECD to identify and recommend priorities for the development and validation of new and improved methods to identify and assess substances acting through endocrine mechanisms.

- 1) Five weak oestrogen agonists having an oestrogen receptor binding affinity three or more orders of magnitude lower than EE were tested in a dose-response study. These chemicals consisted of bisphenol A, *o,p'*-DDT, genistein, methoxychlor, and nonylphenol.
- 2) A non-estrogenic chemical (dibutyl phthalate) was used along with the five weak estrogen agonists in a coded single-dose study. In order to facilitate study comparisons, the dose selected for each chemical in the coded single-dose study was also used in the dose-response study.
- 3) The reference positive control chemical EE from Phase-1 was tested at two different dose levels in both the coded single-dose² and dose-response studies. This approach provided comparison data for the Phase-1 study and also evaluated inter-laboratory reproducibility of the protocols.

A total of 17 laboratories participated in the dose-response study, and several reported unexpectedly high rates of animal mortality. These results suggested that range finding studies should be considered in the future to avoid overt toxicity, particularly with the immature animals, which were more sensitive to high doses. All four protocols were able to detect each of the five weak estrogen agonists, provided that the doses were above the minimal effective dose (MED). The MED for each of these five chemicals was substantially higher than the MED for the potent reference chemical EE, indicating that the assay was capable of detecting estrogen agonists over a substantial concentration range. No animal model (i.e., immature or OVX), protocol, or route of administration was demonstrated to be consistently superior; the specific characteristics of the test chemical seemed equally or more important. With respect to the three versus seven day OVX protocols (Protocols C and D, respectively), the longer dosing time did not indicate any significant advantage over the shorter dosing time.

A total of 86 chemical/laboratory/protocol dose-response combinations were performed, and the results were statistically analyzed. Results showed good agreement among laboratories across protocols, and the magnitude and shape of the dose-response curve for each of the five weak oestrogen agonists was similar within a particular protocol. Furthermore, the response to the EE reference doses was similar within a protocol.

Sixteen laboratories participated in the coded single-dose studies. The single-doses were selected based on presumed positions on the midpoint or upper half of the dose-response curve, but in several cases, the selected doses were erroneously at or near the MED. Therefore, some laboratories did not achieve statistical significance for certain chemicals. An assessment of intra- and inter-laboratory reproducibility indicated that relative increase in uterine weights

² EE was tested in the single-dose study, but it was not under code because the laboratories were required to prepare it at two different concentrations. All of the five weak oestrogen agonists were tested under code.

within and across laboratories were reproducible both with all five weak oestrogen agonists and with the reference chemical EE, taking into account that the magnitude of the relative increase was dose-dependent. The negative control was marginally positive in three out of 36 studies, indicating a false positive rate of approximately eight percent. As with the dose-response studies, the single-dose studies did not show that one protocol or route of administration was clearly superior.

Several sources of assay variability were identified during Phase-2. The most important variable related to the expertise and care within a particular laboratory. For instance, if the laboratory technician did not remove the entire ovary from the OVX rats, then a higher level of background estrogen was produced, making the dynamic range of response smaller and an estrogenic effect harder to detect. With the immature females, it was critical to begin dosing the rats when they were young enough (i.e., before puberty); otherwise, the uterus did not begin the protocol at baseline, and the dynamic range of response was smaller. Cautionary statements related to these factors will be placed in the final Test Guideline.

Phytoestrogen contents in feed were also analyzed as a source of potential variability. At sufficiently large levels, phytoestrogen intake led to a gradual loss in dynamic range of response. Data suggested that the immature rat protocols may be impacted when genestein intakes exceed 50 mg/kg/d. Therefore, the Phase-2 report recommended that experiments with immature rats should limit dietary content of phytoestrogens to approximately 350 µg phytoestrogens/g diet.

Other sources of error identified in the Phase-2 report include the following:

- Dosing errors and losses. The smaller physical size of the immature rat presents a greater challenge for oral gavage and subcutaneous injection.
- The dissection and tissue preparation of the uterus. Because the immature uterus is approximately one-third the weight of the young adult size after OVX, differences in the skill levels of laboratory technicians increases both inter- and intra-laboratory variability.
- Tissue dessication. The small, moist uterine tissue may experience weight loss through dessication before the wet weight can be accurately measured; therefore, precautionary procedures may be required.

The Phase-2 report stated that the observed statistically significant increases in uterine weight compared to the control were consistent with the oestrogen mode of activity of the test chemicals. The overall results of the Phase-2 studies demonstrated that the protocols were robust and reliable for identifying oestrogen agonists and antagonist and that they were transferable across laboratories. The protocols also allowed for variations in several experimental conditions, including rat strain, diet, housing and husbandry practices (e.g., cage

bedding), and the route of exposure. Because the assay is not directly based on absolute uterine weights but rather on the uterine weight increase relative to vehicle controls, the data were presented and primarily analyzed as the ratio of treated to vehicle control uterine weights, adjusting for the final body weight.

The OECD EDTA Task Force agreed that no further work beyond Phase-2 was necessary to demonstrate that the uterotrophic assay was reliable and relevant. It approved the Phase-2 report without changes and agreed to its submission to the Working Group of the National Coordinators of the Test Guidelines Programme (WNT). The Phase-2 report recommended that a draft Test Guideline be prepared for the uterotrophic assay after overall endorsement by the Joint Meeting of the OECD Chemicals Group and Committee and Working Party on Chemicals, Pesticides, and Biotechnology (the Joint Meeting). EPA will permit the use of the four uterotrophic protocols included in the OECD Test Guideline, thus allowing for harmonization and mutual acceptance of data.

Peer Review

The 35th Joint Meeting discussed options for how a peer review may be undertaken prior to the development of the Test Guidelines. The 13th WNT endorsed an independent peer review test-case, including the establishment of a small independent peer review panel with a defined task and minimal face-to-face meetings.

The peer review stage for the uterotrophic assay has been contentious. OECD member countries and non-member stakeholders were permitted to nominate members for the Peer Review Panel (PRP). Member countries participated as panel members, while non-member stakeholders observed as non-voting participants. At times, these roles were difficult to distinguish, especially when differences in opinion arose. The PRP was charged with developing a consensus on the following overall question: *“Has the Uterotrophic Bioassay been sufficiently evaluated and has its performance been satisfactorily characterized by the OECD validation program to support its proposed use for screening the potential of substances to act as oestrogen agonists and antagonists in vivo?”*

The PRP was divided on whether the validation program for the uterotrophic protocols was adequate. Some members felt that the validation program was unsatisfactory and that it resembled more of a prevalidation program. They contended that a single protocol, rather than all four, should have been chosen to proceed to validation. They felt that incorporating all four protocols introduced too many degrees of freedom and sources of variability. In addition, some members critiqued the validation process because it did not include enough negative control compounds. Overall, a consensus on conclusions related to the repeatability/reproducibility of the assay was not reached among PRP members and observers.

An analysis and response to the draft peer review report has been developed by the OECD EDTA Task Force containing a point-by-point rebuttal of the criticisms raised by members of the PRP. EDTA determined that none of the criticisms raised by the PRP would support the conclusion that the validation was unsuccessful in demonstrating that the uterotrophic assay would detect the potential of substances to act as oestrogen agonists and antagonists *in vivo*, and therefore, they have recommended that the OECD develop a draft Test Guideline that permits the flexibility to use includes elements from all four protocols. This recommendation has been presented to the WNT, which will consider it in April 2005. The European Centre for the Validation of Alternative Methods (ECVAM) is critical of this action and has expressed interest in reviewing OECD's peer review and potentially conducting an additional peer review of the assay.

Many of the considerations involved in the design and execution of the validation of the uterotrophic assay are similar to other *in vivo* assay validations. While most of the other *in vivo* assays are not considering several protocols as they proceed to validation, the question of which experimental conditions to standardize (e.g., strain, feed, bedding, etc.) and the number and nature of the reference chemicals to use has been raised with several assays. Thus, the ultimate conclusions as to the adequacy of the validation of the uterotrophic assay may set a precedent for the remaining EDSP assays as they continue through validation and peer review.

The EPA has requested that the Endocrine Disruptor Methods Validation Advisory Committee (EDMVAC) also examine the issues raised by the peer review and provide their opinion on the adequacy of the validation process. This topic is scheduled for the April 2005 EDMVAC meeting. The discussion of these issues will follow a presentation on the validation program and issues raised by the PRP by Dr. Willie Owens.

References

Draft Peer-Review Report for the Uterotrophic Bioassay. Task Force on Endocrine Disrupters Testing and Assessment (EDTA) of the Test Guidelines Programme. Organization of Economic Cooperation and Development, ENV/JM/TG/EDTA(2004)1. 05 January 2005.

OECD Draft Report of the Validation of the Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories. Task Force on Endocrine Disrupters Testing and Assessment (EDTA) of the Test Guidelines Programme. Organization of Economic Cooperation and Development, ENV/JM/TG/EDTA(2003)1. 05 March 2003.

Timm, Gary. United States Environmental Protection Agency, Office of Science Coordination and Policy. Personal communication (meeting) on 16 February 2005.

Additional Information

For the detailed reports on this assay, please see the *Assay table* on the EDSP website at <http://www.epa.gov/scipoly/oscpendo>.