



Answering the Endocrine Test Questions

The concern that some chemicals in the environment could disturb the endocrine systems of humans and wildlife has emerged as one of the decade's highest-profile environmental health issues. Evidence suggesting that certain chemicals could potentially bind to endogenous hormone receptors and disturb normal endocrine functioning in a way that increases the risk of reproductive problems and cancer has alarmed the public and led to international efforts focused on screening chemicals for endocrine activity and potential health effects. Screening chemicals for endocrine activity is an undertaking with a tremendous scope: the U.S. Environmental Protection Agency (EPA) has recommended that its Endocrine Disruptor Screening Program evaluate 87,000 commercial chemicals found in industrial products, pesticides, dietary supplements, and feed additives. The majority of these chemicals lack even the most basic

toxicity information, let alone sophisticated data corresponding to mechanisms of endocrine disruption.

The Legislative Background

In the United States, both the 1996 Food Quality Protection Act and the 1996 amendments to the Safe Drinking Water Act addressed public concerns over endocrine disruption by directing the EPA to develop a comprehensive screening program to evaluate chemicals for hormonal effects.

The screening program that is currently in place was designed by a multi-stakeholder group comprising industry, environmental groups, government, and academia. The program, known as the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), was formed in 1996 to advise the EPA on how best to achieve the goals of the Food Quality Protection Act and the Safe Drinking Water Act. In its final report, published in the 28 December 1998 *Federal Register*, EDSTAC recommended that the

screening program evaluate chemicals in a manner that sequentially eliminates them from further testing based on their performance in a tiered system of assays.

The basic components of the program involve starting with an initial sorting of chemicals based on currently available information. A "Tier 1" screening battery identifies chemicals with the potential to interact with the endocrine system and a subsequent "Tier 2" testing battery provides dose-response data and information on whether the endocrine activity causes adverse effects in humans, fish, and wildlife. Finally, a hazard assessment measures the magnitude of a chemical's potential threat to human and ecological health. EDSTAC also proposed a bypass mechanism that would allow chemicals to circumvent the screening phase and move directly to testing or hazard assessment according to manufacturer desires. This could save the manufacturer's time and money, particularly if they already suspect that a chemical may be an endocrine disruptor.

Test Validation

A key EDSTAC requirement is that candidate test systems used in the endocrine disruptor program be extensively validated to ensure they provide relevant, reliable, reproducible data. The question boils down to whether a test does what it is supposed to do and whether it is reproducible across labs.

Validation is proving to be a monumental task. The process involves confirming that candidate tests can detect chemical effects of numerous hormones, including estrogens, antiestrogens, androgens, antiandrogens, and thyroid and antithyroid hormones. Validation requires reproduction of results in several laboratories, each using dozens of different chemical standards of varying hormonal potency to test for both weak and strong hormonal activity. Tests also need to be validated using chemicals known to lack endocrine activity to the extent that they generate false positive results. The ultimate goal of validation is to provide researchers with reliable standardized test systems they can use to screen and evaluate chemicals for endocrine activity under clearly defined laboratory conditions.

The amount of effort required for validation is so great that some stakeholders are concerned that it will overwhelm resources and consume the screening and testing program altogether. Peter de Fur, an affiliate associate professor of environmental studies at Virginia Commonwealth University in Richmond, who represents several public-sector environmental groups to EDSTAC, says, "We don't want validation to be the enemy of progress. We're concerned that validation could

drag on for many years [and] stand in the way of setting regulatory standards that are protective of public health."

Anthony Maciorowski, a senior technical advisor in the EPA's Office of Prevention, Pesticides, and Toxic Substances and chair of the agency's ad hoc Endocrine Disruption Standardization and Validation Task Force, which currently oversees validation efforts in the United States, says that although these con-

We don't want validation to be the enemy of progress. We're concerned that validation could drag on for many years [and] stand in the way of setting regulatory standards that are protective of public health.

Peter de Fur

cerns are understandable, the EPA must use validated tests if it wants to gather data that will ultimately be useful for regulatory purposes and risk assessment. "I think that some of the assays can be standardized and validated in a reasonable time frame," he says. "And as they come on line, we can begin [screening chemicals]." The task force Maciorowski heads is now coordinating the various groups involved in validation and will process validation data on each of the candidate tests for submission to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), which will conduct the final peer review of the assays and the tier system. ICCVAM was established in 1994 by the NIEHS with the goal of achieving domestic and international harmonization of criteria for the validation and acceptance of alternative test methods. Maciorowski says that the current goal of the task force is to complete Tier 1 screening validation by 2001 and Tier 2 testing validation within two to five years. The total expected cost of validation is \$50 million, a figure that some stakeholders fear may be less than adequate.

The Proposed Methods

At this point in the validation process, 10 *in vivo* and *in vitro* assays for both mammalian and ecological effects have been slated for validation to be completed over the next two

I think that some of the assays can be standardized and validated in a reasonable time frame.

Anthony Maciorowski

years. These include the following: high throughput prescreening (HTPS) estrogen and androgen assays, bench method assays for estrogen and androgen, a rodent 3-day uterotrophic assay, a rodent 5- to 7-day Hershberger assay, a rodent 20-day thyroid/pubertal male assay (as well as a similar

assay for females), a frog metamorphosis assay, a fish gonadal recrudescence assay, an *in utero* developmental assay, and a two-generational mammalian reproductive toxicity study. Of these, the first nine are to be used for Tier 1 screening; the only Tier 2 test currently being validated is the two-generational toxicity study.

Most attention to date has focused on the uterotrophic and Hershberger assays (both *in vivo* mammalian tests) and the HTPS, which

is an automated robotic system initially developed by the pharmaceutical industry to screen drugs for hormonal activity. Validation of the *in vitro* HTPS is being led by the EPA, whereas validation efforts for the uterotrophic and Hershberger assays are being coordinated by the Paris-based Organisation for Economic Co-operation and Development (OECD), with which the United States is working closely to coordinate mutual goals on endocrine disruptor screening. Each of the candidate tests poses a number of difficult challenges. To begin with, even though several of the assays have a long history of laboratory use, none were designed with the explicit intent of evaluating the endocrine activity of a large number of chemicals. Most environmental chemicals are likely to have only weak endocrine activity, which raises some questions among stakeholders regarding the extent to which validation should focus on low-dose testing. "Can we get away with testing in the parts-per-million range or will we have to go to parts per trillion and beyond?" asks de Fur. "We won't know until we get some data."

Another problem is that none of these tests have ever been performed using standardized protocols, meaning that laboratories using them in the past have traditionally applied their own unique approaches. According to Maciorowski, accommodating the new endocrine end points and designing standardized protocols will be among the most challenging aspects of the validation process.

Describing some preliminary data obtained from the HTPS, Maciorowski says, "Some recent reviews indicated that it isn't 'ready for prime time.' We've had problems with low tissue inducibility and high signal-to-noise ratios, [which make it] hard to detect chemicals with weak estrogenic activity." The HTPS determines a chemical's estrogen and androgen receptor binding affinity by measuring a response called transcriptional activation, which is proportional to the degree of receptor

binding. Not surprisingly, using the HTPS in a nonpharmaceutical setting is presenting some difficulties. For example, a much wider range of structural, physical, and chemical properties is likely to be encountered with environmental chemicals, which cut across all classes of chemicals, than in specific classes of commercial drugs. Also, the EPA is more concerned about detecting environmental compounds with low potency, and the HTPS was designed to identify highly active compounds. Maciorowski says he is nonetheless optimistic that the HTPS will find use as a prescreening tool that can be used to prioritize large numbers of chemicals in a short period of time—particularly the 15,000 chemicals produced at volumes of 10,000 pounds or greater that the EPA would like to see screened first. The task force is currently looking to implement a so-called “Challenge Program” whereby 10–12 contract laboratories will evaluate several HTPS methods and gradually weed out the ones that don’t work.

In vitro systems such as the HTPS (and analogous bench method assays) are important, but scientists caution that they won’t be able to provide any information about how metabolism might influence receptor binding; *in vivo* tests are needed to obtain this important information. Because the uterotrophic and Hershberger assays have over 30 years of use in the pharmaceutical industry, they are among the first in line for *in vivo* test validation. Both tests are designed to detect hormonal activity by evaluating changes in specific organ systems. The uterotrophic assay measures changes in uterine weight following exposure to estrogens by female rodents that have undergone ovariectomy. The Hershberger assay measures increased weight of sex glands upon exposure to androgens in castrated male rats. Herman Koëter, the principal administrator of the

We want to be sure the tests capture chemicals with weak [hormonal] activity, because even weak activity can be harmful.

Herman Koëter

OECD’s test guideline program, says that the assays “seem to be appropriate for endocrine screening,” but he adds that standardizing the tests poses continuing challenges. “We want to be sure the tests capture chemicals with weak [hormonal] activity,” he says, because even weak activity can be harmful.

Concern over Animal Welfare

Some stakeholder sectors are concerned that validation and testing for endocrine activity will increase the use of animals in environmen-

tal research. This concern is especially pervasive in Europe, where the European Centre for the Validation of Alternative Methods (ECVAM) is pushing the OECD to reduce the use of animals in its endocrine disruptor program. ECVAM was created in the early 1990s by the European Union with a mandate to coordinate the validation of alternative methods among the union’s 15 member states. Leslie Onyon, a coordinator for endocrine disruptor test valida-

Validation has to be sensible. Stakeholders have to be reasonable in what they demand because you could validate indefinitely. But . . . I think the tests are going to perform well. They may not be right 100% of the time, but I think they will be nearly so.

Lynn Goldman

tion in the OECD’s Environmental Health and Safety Division, says that because ECVAM’s basic purpose is to develop alternative test methods, it is naturally critical of the OECD’s use of *in vivo* methods. But she adds that the actual identification of alternative methods is in the early stage of development. “There’s no international agreement on how much weight should be placed on alternative methods, so right now countries prefer short-term *in vivo* tests,” she says. Onyon says that the OECD is continuing to negotiate with ECVAM, which she says participates fully in all discussions regarding validation. (ECVAM is represented on the OECD’s 15-member Validation Management Group, which is similar in function to the EPA’s validation task force.)

Alternative testing is also a concern in the United States, especially given that ICCVAM (which also has a mandate to develop alternative methods) is a key player in the EPA’s test validation efforts. A current question under debate is whether to use surgically altered animals in the uterotrophic and Hershberger assays, as opposed to using immature animals with low hormonal activity. William Stokes, cochair of ICCVAM and a laboratory animal veterinary specialist, says that the use of immature animals is preferable from an animal welfare perspective because the animals do not run the risk of the potential pain and distress that could result from undergoing castration or ovariectomy. However, the model that is ultimately used will be the one that best performs its intended function. Stokes emphasizes that whenever the surgical procedures are performed, appropriate anesthetics and analgesics are used to minimize potential pain or distress. “One of the criteria for acceptance [of a proposed method] is that there’s adequate consideration of reduction, replacement, and refinement of animal use,” he says. “This is required by federal laws in the United States and in other countries.” But

Stokes adds that the overriding goal is to implement tests that are more predictive than current methods. “This is more important for public health,” he says.

Risk Assessment Issues

Public health objectives will ultimately depend on the identification of candidate tests that can be modified to fit screening goals. Furthermore, Tier 2 tests that can provide data

adequate for use in risk assessment will also be

needed so that public health officials can set regulatory standards that are protective of these end points. In anticipation of the screening exercise that looms ahead, industry stakeholders are adamant that Tier 1 screens that are likely to implicate certain chemicals as potential endocrine disruptors be validated concurrently with Tier 2 tests that can determine whether the effects are actually adverse.

According to Ronald Miller, a senior toxicology consultant with the Environment, Health, and Safety Division at Dow Chemical, industry representatives remain concerned that some chemicals could be prematurely designated as harmful and then be left hanging without more complete information on health effects, information that could reduce public alarm and suspicion. Lynn Goldman, former EDSTAC chair and now an adjunct professor at the Johns Hopkins School of Public Health in Baltimore, Maryland, says such industry fears are not unreasonable. But she cautions that there are “forces whose interest . . . is to make the process move as slowly as possible”—industry, for example, will be footing the screening bill and will therefore benefit from any delay in the validation process. Says Goldman, “Validation has to be sensible. Stakeholders have to be reasonable in what they demand because you could validate indefinitely. But I’m optimistic about validation. I think the tests are going to perform well. They may not be right 100% of the time, but I think they will be nearly so,” she says. The challenge of the task force is to design a testing system that minimizes false positive results without holding the screening process hostage to a validation effort that goes on indefinitely. In this respect, Maciorowski suggests that even once the formal validation process is completed, time and experience with the tests will gauge their true effectiveness once the screening process is under way.

Charles W. Schmidt