U.S. Environmental Protection Agency
Endocrine Disruptor Screening Program
Universe of Chemicals and General Validation Principles

Jointly developed by the Office of Chemical Safety & Pollution Prevention, the Office of Water and the Office of Research and Development

November 2012

This document was developed by the EPA to provide guidance to staff and managers regarding the universe of chemicals and general validation principles for consideration of computational toxicology tools for chemical prioritization.

Table of Contents

Overview Statement

This document is intended to serve as a supplement to the recently published *EDSP Comprehensive Management Plan*.[1] In June 2012, the EPA published the *EDSP Comprehensive Management Plan* to provide strategic guidance to the EPA staff and managers participating in the internal activities associated with EDSP for the period from FY 2012 through FY 2017. In addition to the *Comprehensive Management Plan*, the agency is providing this document to address the universe of chemicals that may be considered by the EDSP for screening and testing within the five year time horizon of the *EDSP Comprehensive Management Plan*. Given the large number of chemicals that are regulated by the EPA and potentially subject to EDSP screening, it is important to strategically prioritize which chemicals should undergo screening so the chemicals with the greatest potential for interacting with the endocrine systems are evaluated in a timely manner to ensure public and wildlife protection. To this end, this document describes general validation principles for the use of computational toxicology tools for efficient chemical prioritization. It is important to emphasize that these are general, flexible principles and are not intended to be rigid or prescriptive.

1. Universe of Chemicals for Endocrine Disruptor Screening and Testing

Summary

One of the U.S. Environmental Protection Agency's (EPA) highest priorities is to assure that people and the environment are not exposed to dangerous levels of chemicals. The EPA developed the Endocrine Disruptor[2] Screening Program (EDSP) in response to the statutory mandate in the Federal Food, Drug, and Cosmetic Act (FFDCA) "to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or such other endocrine effects as the Administrator may designate." As part of the EDSP, the statute requires that all pesticide chemicals (active and inert ingredients) be screened, and also establishes the EPA's authority to "provide for the testing of any other substance that may have an effect that is cumulative to an effect of a pesticide chemical if the Administrator determines that a substantial population may be exposed to such a substance." In addition to FFDCA, the Safe Drinking Water Act (SDWA) provides the EPA with authority to provide for testing "of any other substance that may be found in sources of drinking water if the Administrator determines that a substantial population may be exposed to such substance." Beyond testing and determining endocrine effects, FFDCA also directs the EPA to take action: "In the case of any substance that is found…to have an endocrine effect…the Administrator shall, as appropriate, take action under such statutory authority as is available to the Administrator…to ensure the protection of public health."

The scope of the authority established by the statutory provisions in the FFDCA and SDWA (*i.e.*, pesticide chemicals and drinking water contaminants with exposure determinations for substantial populations) provide a representative chemical universe for the EDSP. While this does not necessarily include all chemicals that may ultimately be considered for prioritization and potential screening in the future of the EDSP under other authorities such as the Toxic Substances Control Act (TSCA), the agency believes that the application of TSCA would not substantially alter the current range or breadth

[1] U.S. EPA, *EDSP Comprehensive Management Plan*, June 2012.
http://www.epa.gov/endo/pubs/EDSP-comprehensive-management-plan.pdf

[2] An endocrine disruptor (ED) is defined according to the World Health Organization as "an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism" (WHO, 2002).

of unique chemical categories to be addressed during the 5-year time horizon of the *EDSP Comprehensive Management Plan*, especially after implementing the chemical prioritization scheme using exposure information, extant toxicological data and advanced computational tools. (For additional detail, see section 3 on the validation principles for use of chemical prioritization tools).

As defined by the FFDCA and SDWA statutes, the universe of chemicals potentially targeted for testing under the current five-year time horizon of the EDSP *Comprehensive Management Plan* (U.S. EPA, 2012) consists of approximately 10,000 unique chemicals.[3] This includes approximately 6,000 drinking water contaminants,[4] approximately 1,000 pesticide active ingredients and approximately 5,000 inert ingredients, with some overlap between these lists of chemicals. This scope of the universe enables the agency to estimate potential resource needs and timelines in the context of the five-year management plan. This scope also allows the EPA to focus its initial screening efforts on a manageable universe of potential chemicals for EDSP screening that reflects the priorities established by Congress.

It is important to emphasize that the identification of this universe of chemicals for potential EDSP screening should neither be interpreted as a list of chemicals that will automatically be screened, nor as a list of chemicals with potential to interfere with endocrine systems of humans or other species. As such, the universe of chemicals identified for the EDSP should not be considered a list of "known" or "likely" endocrine disruptors.

Background

In August 1998, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) issued its final report, which included recommendations regarding "the universe of chemicals to at least be considered for endocrine disruptor screening and testing." This universe was estimated to include approximately 87,000 individual chemical substances derived from the following categories:

Table 1: EDSTAC Estimate of the Universe of Chemicals, Circa 1998

Category	Previously Estimated Universe[5]
Chemicals Listed on TSCA Inventory	~75,500
Pesticide Active Ingredients	~900
Pesticide Inert Ingredients	~2,500
Chemicals Regulated by FDA	~8,000
Total	**~87,000**

In this characterization of the universe of chemicals, EDSTAC included substances beyond those regulated by the EPA, acknowledging that the Food Quality Protection Act (FQPA) [amendments to FFDCA] and SDWA "do not confer on any other agency the regulatory authority to require screening and testing for endocrine disruption potential." EDSTAC's recommendations were not based on interpretations of statutory authority but were driven by the scientific possibility of human exposure to chemicals for which there are essentially no data on the potential for endocrine disruption.

[3] See http://www.epa.gov/endo for the current complete list of chemicals.
[4] Contaminant Candidate List 3 (CCL 3) universe of chemicals and contaminants with National Primary Drinking Water Regulations (NPDWRs). For more details on the CCL 3 universe, see http://water.epa.gov/scitech/drinkingwater/dws/ccl/upload/CCL3_Chemicals_Universe_08-31-09_508_v3.pdf. For more details on contaminants with NPDWRs see: http://water.epa.gov/drink/contaminants/index.cfm.
[5] Based on estimates in 1997 and 1998

Later that year, the EPA published the *EDSP Proposed Statement of Policy*[6] in which the agency stated its concern about the endocrine disrupting potential of more than 87,000 chemical substances, including pesticide chemicals (active and inert ingredients), commercial chemicals, cosmetics ingredients, food additives, nutritional supplements and certain mixtures. The agency noted the impracticality of testing 87,000 chemicals. Furthermore, in order to ensure the EDSP screening and testing of these substances, the authority would need to extend beyond FFDCA and SDWA and the effort would need to include other federal agencies and departments. In recognition of the large volume of chemicals for potential screening and the availability of new, advanced computational toxicological methods, the agency announced the establishment of the National Center for Computational Toxicology (NCCT) to develop high-throughput (HTP) assays for chemical prioritization and screening. The agency subsequently proposed conducting HTP assays on approximately 15,000 chemicals captured in the EDSP universe of chemicals domain (commercial chemicals produced in amounts greater than or equal to 10,000 pounds per year and all pesticides).

In 1999, the EPA convened a joint meeting of the agency's Science Advisory Board (SAB) and Federal Insecticide, Fungicide, and Rodenticide (FIFRA) Scientific Advisory Panel (SAP) to review the proposed EDSP. By 1999, the agency concluded that the HTP and other computational toxicology tools were not yet ready for regulatory implementation. The SAB/SAP concurred with this assessment and expressed concerns regarding the ambitious scope of the 87,000 chemical universe for the EDSP. Furthermore, the SAB/SAP advised that developing massive amounts of screening data, even on 15,000 chemicals, would not necessarily expedite the development of the appropriate underpinnings that the agency needed before it proceeded with the screening of the large universe of chemicals anticipated to be included in the EDSP. The panels recommended that the EPA not expand the set of agents beyond those captured in FFDCA and SDWA until the agency developed or adopted validated systems and provided clear decision criteria.

In September 2005, the agency published its approach for selecting chemicals for the initial round of screening in the EDSP, effectively adopting the SAB/SAP's joint recommendations.[7] In the approach, chemicals were selected based on their relatively high potential for human exposure. The scope of the first group included pesticide active ingredients and High Production Volume (HPV) chemicals used as inert ingredients in pesticide formulations. This scope allowed the EPA "to focus its initial screening efforts on a smaller and more manageable universe of chemicals that emphasizes the early attention to the pesticide chemicals that Congress specifically mandated the EPA to test for possible endocrine effects."

In the FY 2010 House Appropriations Committee Report,[8] the committee directed the EPA to "publish within one year of enactment a second list of no less than 100 chemicals for screening that includes drinking water contaminants, such as halogenated organic chemicals, dioxins, flame retardants (PBDEs, PCBs, PFCs), plastics (BPA), pharmaceuticals and personal care products." In response, on November 17, 2010, the agency published a proposed second list of chemicals, including pesticides and drinking water contaminants, for EDSP Tier 1 screening. The House committee also directed the EPA to engage in a timely re-evaluation of the Tier 1 battery of assays, replacing outdated ones with updated, more efficient screens that have been validated.

[6] 63 *Federal Register* (FR) 71542-71568 (December 28, 1998), EDSP: Statement of Policy. http://www.epa.gov/endo/pubs/122898frnotice.pdf
[7] 70 *FR* 56449 56449 -56465 (September 27, 2005), EDSP; Chemical Selection Approach for Initial Round of Screening https://federalregister.gov/a/05-19260
[8] U.S. Congress, House Appropriations Committee Report 111-180, pp. 105-106. http://www.gpo.gov/fdsys/pkg/CRPT-111hrpt180/pdf/CRPT-111hrpt180.pdf

Also in FY 2010, the EPA announced plans to evolve the EDSP to make greater use of computational toxicology tools. The FY 2012 President's budget notes that, "In FY 2012 EPA will begin a multi-year transition from the Endocrine Disruptor Screening Program (EDSP) to validate and more efficiently use computational toxicology methods and high-throughput screens that will allow the agency to more quickly and cost-effectively assess potential chemical toxicity."[9] To help implement this transition, the EPA developed and published an EDSP21 work plan summary[10] that details three phases of Tox21 tool application: 1) chemical prioritization, 2) screening and 3) data replacement. These three phases are intended to build knowledge and increase confidence in the use of computational toxicology tools in the regulatory decision making process.

Between 2010 and 2011, the EPA's Office of the Inspector General (OIG) evaluated the EDSP and determined that, without a better-defined universe of chemicals, the agency will not be able to estimate longer-term resource needs for completion of milestones for the program. Therefore, the OIG recommended that the agency define and identify the universe of chemicals for potential EDSP screening and testing. In response, the agency has incorporated a discussion of the universe of chemicals into both the *EDSP21 Work Plan* and the *EDSP Comprehensive Management Plan*. Both documents are intended to provide primary guidance regarding the strategic direction and management of the EDSP for an annually reviewed period of at least five years.

Agency Statutory Obligations

As defined by the statutory provisions in the FFDCA and SDWA (Table 2), the universe of approximately 10,000 chemicals is appropriate for EDSP prioritization and potential screening. This chemical universe allows the agency to continue to focus its prioritization and screening efforts on a chemical universe that reflects the priorities established by Congress: the pesticide chemicals that Congress specifically required to be screened and the more recent congressional directive to begin screening drinking water contaminants with exposure determinations for substantial populations.

Estimates of the universe of chemicals, defined in Table 2, are expected to change over time. For example, each year, the Office of Pesticide Programs (OPP) registers new pesticide active ingredients and approves new inert ingredients for incorporation into pesticide product formulations. OPP may also cancel registrations for certain active ingredients or discontinue approvals for inert ingredients.

In considering this universe of chemicals, the agency plans to develop a prioritization scheme built on the broader, general concept articulated in the National Academy of Sciences *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy* (Report in Brief), which speaks to the need to integrate all existing knowledge and multiple tools to generate a more practical, scientifically-based prioritized list of chemicals for EDSP screening. These prioritization tools include the consideration of physicochemical properties (*e.g.*, exclusions of polymers, strong acids and bases, reactive and unstable compounds, undefined chemicals like coconut, oils and kaolin, etc.), structure activity relationship and high-throughput computational methods. This chemical prioritization scheme will be submitted for external scientific peer review in January 2013.[11]

[9] U.S. EPA, FY 2012, *Justification of Appropriation Estimates for the Committee on Appropriations*, EPA-190-R-11-003, p. 53. http://www.epa.gov/planandbudget/archive.html
[10] U.S. EPA, *EDSP for the 21st Century (EDSP21): Summary Overview*, September 2011. http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf
[11] 77 *Federal Register* 68773-68775 (November 16, 2012) http://www.gpo.gov/fdsys/pkg/FR-2012-11-16/pdf/2012-27816.pdf

In the integrated prioritization scheme, the agency will apply both exposure and effect-based methods and consider inherent chemical properties to formulate a process for prioritizing chemicals based on their likelihood to potentially interact with the endocrine system. Some potential outcomes from this prioritization scheme may be to place certain chemicals on a low priority list for screening because their physicochemical properties indicate a low likelihood for potential to interact with the endocrine system or cause systemic effects; or chemicals may be excluded from EDSP screening based on other factors (*e.g.*, no significant human exposure, or too reactive or unstable). Consideration of these factors is expected to further prioritize candidate chemicals and reduce the number of chemicals for Tier 1 screening under the Endocrine Disruptor Screening Program to those with a higher potential probability of interacting with the endocrine system.

Note: The universe of chemicals is not a static list of chemicals; it represents a dynamic universe of chemicals that will change over time. The prioritization status of this universe of chemicals will be updated with the annual update of the *EDSP Comprehensive Management Plan*, and as new data becomes available. See http://www.epa.gov/endo for the current list of chemicals that comprise the universe of chemicals for potential endocrine disruptor screening and testing.

Table 2: The Numerical Estimates of Chemicals Associated with Each Authority[12]

Citation	Statutory Required	Defined Universe
FFDCA §408(p)(3)(A) [21 USC 346a(p)(3)(A)]	(3) SUBSTANCES - In carrying out the screening program…the Administrator — (A) shall provide for the testing of all pesticide chemicals;	Pesticide Active Ingredients = ~**1,000** Chemicals Pesticide Inert Ingredients = ~**4,000** Chemicals
	Discretionary Authority	
FFDCA §408(p)(3)(B) [21 USC 346a(p)(3)(B)]	(3) SUBSTANCES - In carrying out the screening program…the Administrator — (B) may provide for the testing of any other substance that may have an effect that is cumulative to an effect of a pesticide chemical if the Administrator determines that a substantial population may be exposed to such substance.	Will depend on case-by-case determinations regarding cumulative effects and exposure.[13]
SDWA §1457 [42 USC 300j–17]	In addition to the substances referred to in section 408(p)(3)(B) of the Federal Food, Drug, and Cosmetic Act (21 USC 346a(p)(3)(B)) the Administrator may provide for testing…of any other substance that may be found in sources of drinking water if the Administrator determines that a substantial population may be exposed to such substance.	Regulated Contaminants =~**90** Chemicals CCL 3 Universe =~**6,000** Chemicals
Universe of Chemicals for Prioritization and Screening		**~10,000 Chemicals***

*Due to overlap among the different chemical regulatory inventories the total number, *i.e.* 10,000, in the universe is not equal to the sum of the individual inventories and should not be construed as a list of chemicals that will automatically be submitted for EDSP screening.

[12] FFDCA §408(p)(3)(A) and (B) are both subject to the exemptions described at §408(p)(4): "EXEMPTION.—Notwithstanding paragraph (3), the Administrator may, by order, exempt from the requirements of this section a biologic substance or other substance if the Administrator determines that the substance is anticipated not to produce any effect in humans similar to an effect produced by a naturally occurring estrogen."

[13] The first step in understanding what chemicals may have endocrine effects that are cumulative to those of pesticides is to screen pesticides for their potential to interact with endocrine systems.

References for the Universe Defined in Table 2

Pesticide Chemicals

Antimicrobials: http://www.epa.gov/oppsrrd1/registration_review/2011-14-antimicrobial.pdf
Biochemicals: http://www.epa.gov/oppsrrd1/registration_review/2011-14-biochemical.pdf
Conventionals: http://www.epa.gov/oppsrrd1/registration_review/2011-14-conventional.pdf
Microbials: http://www.epa.gov/oppsrrd1/registration_review/2011-14-microbial.pdf
Inerts: http://www.epa.gov/opprd001/inerts/

SDWA Chemicals

Regulated Contaminants: http://water.epa.gov/drink/contaminants/#List

CCL 3 Universe:
http://water.epa.gov/scitech/drinkingwater/dws/ccl/upload/CCL3_Chemicals_Universe_08-31-09_508_v3.pdf

2. Public Participation in Prioritizing Chemicals for Endocrine Screening

The agency is committed to an open and transparent public process for selecting chemicals to undergo EDSP screening under FFDCA and SDWA. The EPA routinely incorporates public comment periods in the 1) CCL nomination and pesticide registration review processes[14] and 2) EDSP proposed chemical list *Federal Register* notices.

There are several ways the public can comment on the chemicals within the universe covered by the EDSP. For instance, the EPA provides an opportunity for public review and comment on chemicals considered under the SDWA CCL as part of the development process. The EPA also provides the public with an opportunity to comment during the registration review of pesticides. In addition, the process for regulating pesticide active and inert ingredients under the FFDCA, all of which will be evaluated with Tox21 computational tools for potential EDSP screening, provides numerous opportunities for public input, as part of the statutorily mandated process of establishing and revoking pesticide tolerances.

The public can also directly comment on the EDSP lists of chemicals. The EPA has routinely provided an opportunity for public comment on all proposed lists of the chemicals included for screening through EDSP. During these open public comment periods, the public may provide additional information to support or refute the proposal of chemicals for EDSP screening, and may suggest additional chemicals that warrant screening. The EPA believes the existing public participation processes adequately ensure public input is considered in the prioritization of chemicals for EDSP screening.

3. Validation Principles for Computation Tools used in Prioritizing Chemicals for Endocrine Screening

The FFDCA[15] and the SDWA amendments[16] contain provisions for screening chemicals for their potential to affect the endocrine system. Thus, given the large number of chemicals that are regulated by the EPA and potentially subject to potential endocrine screening, it is important to strategically prioritize which chemicals should undergo screening, so the chemicals with greatest potential for interacting with

[14] EPA provides opportunity for public participation and comment on docket openings, proposed decisions and significant assessments in the registration review process. http://www.epa.gov/oppsrrd1/registration_review/reg_review_process.htm
[15] 21 U.S.C. § 346a(p)
[16] 42 U.S.C. 1457-a

endocrine systems are evaluated ahead of chemicals that are not anticipated to interact with the endocrine system due to inherent chemical properties. The EDSTAC recommended that, to the extent possible, prioritization should involve a scheme that combines exposure and effects information.[17] With regard to effects information to support prioritization, the EDSTAC encouraged the agency to evaluate the use of HTP *in vitro* assays and computational (*in silico*) models, including (quantitative) structure-activity relationships [(Q)SARs]. Furthermore, the National Research Council (NRC) provided a strategy for the use of new *in vitro* and *in silico* technologies to reduce the use of animals and accelerate the pace of testing and assessment.[18] Since the NRC report, the EPA has invested in the development of HTP *in vitro* assays and *in silico* models for the rapid screening of potential chemical targets, including endocrine-related endpoints. Hereafter within this document, HTP *in vitro* effects-based assays and *in silico* models are referred to jointly as "computational toxicology tools."

Consistent with the NRC 21st century vision, the agency developed an EDSP21 Work Plan for modernizing the EDSP, which outlines an incremental, phased approach for integrating computational toxicology tools in the EDSP as a component of testing and assessment.[19] This workplan contemplates validation of computational toxicology tools in each of the proposed phases prior to implementation:

- Prioritization: determining the order of chemicals to enter Tier 1 screening;
- Screening: using effects-based validated *in vitro* HTP assays and chemical category-based predictive models to replace *in vitro* Tier 1 assays and inform relevant *in vivo* Tier 1 assays;
- Replacement: replacing validated *in vivo* Tier 1 assays with effects-based validated *in vitro* HTP assays and chemical category-based predictive models.

Validation of computational toxicology tools will be based on its "fit for purpose." This paper focuses only on the "validation principles" that the agency will use in the development and implementation of a chemical prioritization approach based on computational toxicology tools. These tools will be used to help prioritize candidate chemicals for EDSP Tier 1 screening. The validation approach presented within this paper is based on internationally adopted and harmonized scientifically peer-reviewed principles by the Organization for Economic Co-operation and Development (OECD) (OECD, 2004; OECD, 2007). These principles can be found in the OECD document entitled *Principles for the Validation, for Regulatory Purposes of (Quantitative) Structure-Activity Relationship Models* (2007).[20]

In its report, *Integrated Approaches to Testing and Assessment Strategies: Use of New Computational and Molecular Tools*, the FIFRA SAP recommended the use of the OECD principles in evaluating acceptance and utility of computational tools or models based in part on their flexibility.[21] The SAP also highlighted the importance of understanding how each model should be developed with a "fit for purpose" paradigm. Furthermore, the FIFRA SAP used the same OECD validation principles to evaluate, and support the use of, an Estrogen Receptor (ER) Expert System for the prioritization of

[17] U.S. EPA, *Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), Final Report*, August 1998.
[18] National Academy of Sciences, *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy (Report in Brief)*, 2007, pp. 196. http://dels.nas.edu/resources/static-assets/materials-based-on-reports/reports-in-brief/Toxicity_Testing_final.pdf
[19] U.S. EPA, *EDSP for the 21st Century (EDSP21): Summary Overview*, September 2011. http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf
[20] http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en
[21] U.S. EPA, SAP: *Integrated Approaches to Testing and Assessment Strategies: Use of New Computational and Molecular Tools*, May 2011. http://www.epa.gov/scipoly/sap/meetings/2011/may/052411minutes.pdf

chemicals for screening in EDSP Tier 1.[22] (See also the appendix.) Thus, the validation principles presented below are generally applicable to a diverse range of computational approaches and can be extended, for purposes of chemical prioritization, to validating HTP and other *in vitro* assays that provide data inputs for computational tools.

Validation

Validation encompasses the process necessary to develop methods for use in a regulatory program. In general, validation is a process of building confidence that the results obtained from a method can be reproduced, that the test method accurately predicts or measures the defined endpoint and that the test is appropriately grounded with reference compounds reflective of the universe of chemicals to which the methods will be applied. As discussed in the OECD "*Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models*" (OECD, 2007), validation is defined as:

> *The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.*
>
> *Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability.*
>
> *Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method.*

Additionally, evaluation of a new method should include a review of the assumptions, relevance, reliability, sensitivity, and specificity of the method for the intended use prior to regulatory acceptance. There is a need to understand the uncertainties and limitations associated with new technologies. While these technologies and methods may be new, there are existing frameworks for validation of methods which ensure the transparency and usefulness of the method to answer a regulatory question.

Principles of Validation for Chemical Prioritization in EDSP

The agency will continue to follow internationally accepted science principles when developing chemical prioritization methods using computational toxicology tools, including but not necessarily limited to methods and tools that may be used to support the EDSP. Consistent with OECD (2007), EDSP validation will establish that a tool has the following:

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness and predictivity;
5. a mechanistic interpretation.

[22] U.S. EPA, SAP: *The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing,* August 2009. http://www.epa.gov/scipoly/sap/meetings/2009/august/082509minutes.pdf

These principles are consistent with and complement the recommendations of the EDSTAC. The specifics of these principles are not prescriptive and are flexible enough to be broadly applicable to predictive models or computational tools which may be based on HTP or other *in vitro* data.

Beyond the importance of validating model equations and algorithms for regulatory needs, two important characteristics of model validations are addressed by OECD to enhance regulatory acceptance for priority-setting. The first characteristic is transparency of a model estimate, not only with respect to the methods used, but also in terms of how the estimate can be explained mechanistically, the data it is built on, and whether the resulting model estimate is reasonable when evaluated against data for comparable chemicals. The second major characteristic for model acceptance is usefulness of a particular model for estimating endpoints of regulatory relevance for all compounds within specified chemical inventories. Since the OECD principles of validation seek to describe the domain of the model in terms of the chemical structures used to create the model, usefulness can be evaluated by comparing the chemical domain of the test set (*e.g.*, the chemical tested *in vitro* HTP or any other *in vitro* data used to build a model) and the chemical domain of the regulatory inventory assessed in a specific regulatory context.

Principle 1: Defined Endpoint

The intent of Principle 1, a well-defined endpoint, is to ensure clarity in the endpoint data used as the basis of comparison. This applies equally to any data collection representing the regulatory inventory of chemicals whether that data is *in vitro* lower through-put, *in vitro* HTP data, or *in vivo* data. It also applies whether the data is used as is, or further used to build predictive models, chemical categories, or read-across.[23] Any predictions made from the data will inherently contain all uncertainties and limitations in the data measurements, and may also include additional uncertainties due to built-in model assumptions. Thus, all aspects of measurements and the reported endpoint of an assay should be well-defined.

The endpoint could be a physicochemical property, biological effect or environmental parameter related to chemical structure that can be measured in an assay and then modeled. The defined endpoint is a characterization of how the assay was conducted and the endpoint measured so that strengths, limitations and uncertainties in the data are defined for the types of chemicals and chemical properties found in the regulatory inventory of interest. The endpoint definition should include what is or is not understood (well characterized) about the endpoint as it is measured in the assay. Performance of the assay with appropriate reference chemicals, reproducibility of the assay, signal-to-noise ratios, background subtraction, interferences, etc. should all be characterized when describing the strengths and limitations of the assay.

Defining the endpoint includes determining the degree to which chemical form and bioavailability within the assay system has been measured and is generally known, or whether these types of measurements have seldom been made and are largely unknown. Similarly, the degree of understanding surrounding the biological endpoint collected in HTP or otherwise and upon which a model is based is also defined. For example, to what extent all the biological components of the assay have been described, especially with respect to where the chemical-biological interaction is occurring and the

[23] Read-across is a method of filling in data gaps for a substance by using surrogate data from another substance. Read-across can be between two substances or through a group or category of chemicals. The groups are selected on the assumption that the properties of a series of chemicals with common structural features will show similar trends in their physico-chemical properties and in their toxicological effects or environmental fate properties.

proximity of this site of initial chemical interaction in relation to where the assay's measurement endpoint is taken. Information must also be provided about important experimental conditions that affect the measurements, and therefore the predictions/data and the applicability to the regulatory question the data is used for. The goal of this principle is not to exclude a particular test system from use, but rather to describe the degree of understanding surrounding a test system (both chemical and biological aspects), to assess what regulatory questions the information can or cannot answer, and to define the degree of uncertainty in the answer. The potential user can then decide on the appropriateness of the use of the particular endpoint data and the model derived from it for their question at hand.

As discussed in OECD (2007), the "endpoint being modeled can be described in the following terms:"

1. The endpoint should be defined by providing detailed information on the test components and the protocols used to generate the assay data used in a HTP data collection or upon which a model is built (*i.e.*, training set data), especially with respect to factors that impact variability, knowledge of uncertainties, and possible deviations from standardized test guidelines if available.

2. Alternative means of measuring the described endpoint (*e.g.*, alternative signal detection methods) should not lead to markedly different values of the endpoint when the same chemical is tested at the same concentration relative to positive and negative controls with each alternative detection method.

3. Differences within a protocol (*e.g.*, media, reagents) should not lead to differences that cannot be rationalized (*e.g.*, impact of increasing total protein decreasing free and biologically-effective chemical concentration in an *in vitro* assay).

4. A well-defined bioassay domain will report the responses in the bioassay to the types of chemicals (*i.e.*, range of chemical properties) that are anticipated to encompass those properties of the chemicals for which model predictions are needed when addressing the regulatory question. The attributes of reference chemicals in the context of the biological response/endpoint being measured is important, as well as the degree to which the training set (*i.e.*, reference chemicals) cover the chemicals (or properties) to which the assay or model built from assay training sets will be applied; this information is used to define assay strengths and limitations, a necessary component in determining applicability to the regulatory question. [Also relevant to Principle 3: Defined Domain of Applicability]

5. A well-defined endpoint describes the degree to which it is known whether the test chemical may be altered within the test system by biotic (*e.g.*, metabolism) or abiotic (*e.g.*, hydrolysis) processes within the assay system prior to the chemical-biological interaction (molecular initiating event) leading to the measured endpoint.

6. A well-defined endpoint should provide confidence that measured endpoint differences are attributable to differences between chemical structures tested and are not confounded by how the biological or physical aspects of the assay system may affect chemicals differently (adsorption, volatility affecting localized chemical availability/concentration); a well-defined system will identify the potential for some chemicals to interfere with the signal output (directly or indirectly) apart from the biological activity being measured (*e.g.*, a fluorescent chemical interfering with a fluorescent reporter system output).

Principle 2: Unambiguous Algorithm

The intent of Principle 2, an unambiguous algorithm, is to ensure transparency in each part of the system that generates predictions of an endpoint so that others can reproduce the model and understand how the model predictions are derived. In this instance there will be transparency in any algorithms (or decision frameworks) used to prioritize a chemical. Along with transparency of any *in silico* models used in prioritization there will be transparency in any algorithm used for correcting or filtering HTP assay data. For *in silico* models, all equations will be explicitly defined, including definitions of all descriptors used. For *in vitro* HTP effects-based assays the processes of data normalization, outlier identification, curve fitting, background subtractions, determinations of assay interferences and all other data manipulation techniques will be adequately and thoroughly described in a comprehensive and transparent manner.

Principle 3: Defined Domain of Applicability

The intent of Principle 3, defined domain of applicability, expresses the need to establish the scope and limitations of a HTP data collection, or model based on the assay data (*i.e.*, training sets), with respect to the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions. The importance of the principle lies in the fact that data collections and resultant models can only be expected to give reliable predictions for chemicals that are similar to those measured in the dataset and used in model building. In regards to HTP assays, the domain of applicability for prioritization must incorporate limitations of "testability" which include issues related to solubility, volatility, detection interference, etc.

In application of this principle to an expert system using decision logic trees, a chemical is within the effects-based expert system domain if the chemical class is represented and, if the target chemical falls within the values of the pertinent chemical property or properties of the training set chemicals. The pertinent chemical properties are those associated with the biological activity being predicted as described in OECD Expert Consultation (OECD 2009a; and U.S. EPA 2009).

Principle 4: Appropriate measures of Goodness-of-fit, Robustness and Predictivity

Principle 4 expresses the need to report what is known about the performance of any type of model developed using assay data. Measurements used will be appropriate to the type of model built and for the intended use. What is appropriate for one type of model and use case is not necessarily appropriate for all, *e.g.*, regressions models are different from decision logic systems. Goodness-of-fit and robustness are used as measures of internal performance of a model, while predictivity is determined by external performance. Robustness refers to the stability of the model in response to perturbations or varying inputs. OECD recognizes that implementation of this principle is context-specific (*e.g.*, what is recommended for quantitative approaches are likely not appropriate for qualitative approaches).

Principle 5: Mechanistic Interpretation

The intent of Principle 5, mechanistic interpretation, is to ensure a mechanistic association between the attributes of a chemical and its interaction with the biological system resulting in the measured endpoint, to the degree possible. An aspect of mechanistic interpretation relates to showing the evidence of why a chemical structure/substructure is being focused on and how the substituents in the substructure are related to a biological activity. The intent of Principle 5, however, is not to reject models that have no apparent mechanistic basis, but to ensure that consideration is given to the mechanistic association between the descriptors used in a model and the endpoint being predicted, and to ensure that the extent of this association is documented.

4. Peer Review and Public Participation

The agency is committed to basing its decisions on sound science and to developing its methods and policies through an open, public participatory, and transparent process. New scientific methods used to prioritize chemicals are expected to focus further screening efforts on a significantly smaller subset of chemicals with a higher probability of interacting with the endocrine system.

This integrated process will undergo an expert external peer review process in January 2013 to ensure scientific rigor and provide opportunities for public participation. Existing practice at the EPA is to ensure that the scientific peer review of any new method or guidance that is of critical importance to the agency's work is consistent with the principles in the U.S. Environmental Protection Agency Peer Review Handbook (U.S. EPA, 2000).

An overriding point of the validation principles, as defined in OECD (2007) and as interpreted by the agency, is to document how much is known about the data, how they were generated, the mechanism of interest, and the chemical domain of applicability within which these data or model(s) may be applied to make predictions. Transparency is achieved by providing to the public the documentation around the computational toxicology tools, including HTP assays and computational models (*e.g.*, (Q)SARs and Expert Systems) proposed for use in chemical prioritization. Adherence to the general validation principles promotes confidence in the reliability and relevance of the tools and their outputs, which, in turn, ensures their "fit-for-purpose" application in a given regulatory context.

The agency will continue to observe these validation principles, in the development and implementation of chemical prioritization methods based on computational toxicology tools. As these tools are incrementally phased into the EDSP21 initiative, users will have a clear understanding of the decision framework used for prioritization and the level of confidence associated with various predictions or data inputs that inform the decision making process.

Appendix - Application of the Validation Principles Using An Effects-Based Expert System to Predict Estrogen Receptor Binding Affinity

In August 2009, the EPA convened an SAP to review the use of a structure activity relationship expert system to support prioritization of chemicals for screening and testing in EDSP (U.S. EPA 2009). This system is termed the Estrogen Receptor (ER) Expert System, which was developed using two specific *in vitro* assays: i) measured chemical binding to the rainbow trout ER to detect the potential for a chemical to initiate the ER-mediated adverse outcome pathway (AOP); and ii) ER-mediated vitellogenin induction in rainbow trout liver slices to confirm that ER binding translates to an effect at a point further along the ER-mediated AOP (Schmieder *et al.* 2004). The ER Expert system was reviewed as part of a larger Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox (OECD 2009b) and was the subject of an OECD Expert Consultation (OECD, 2009a) held February 16, 2009, demonstrating how the (Q)SAR validation principles are interpreted with regard to expert system decision logic. The appropriateness of the ER expert system for prioritizing food-use inert and antimicrobial pesticide chemicals from those two groups for Tier 1 screening was evaluated by the SAP using the OECD Principles for the Validation of (Q)SARs (OECD, 2007).

This chemical prioritization model used a (Q)SAR-based approach that maintains mechanistic transparency and allows the domain of the knowledge base to be aligned with the domains of specific regulatory inventories of interest. This ER binding affinity expert system will be used to illustrate the application of the validation principles for chemical prioritization.

Table A - Application of the OECD (Q)SAR Validation Principles Using An Effects-Based Expert System to Predict Estrogen Receptor Binding Affinity

Defined Endpoint	The U.S. EPA is faced with large numbers of chemicals that need to be assessed for their potential to cause endocrine disruption. Time and resources dictate that all chemicals cannot be evaluated at once. The challenge is to determine which chemicals should be tested first. Ideally a hypothesis-based approach would be used, which focuses on and prioritizes those chemicals for screening that most likely have the potential to interact with endocrine systems. In this case, the approach used aims to identify chemical structures that can bind and activate the estrogen receptor (ER). This model is based on the concept of an adverse outcome pathway that is initiated through direct chemical binding to the ER and that could plausibly lead to reproductive impairment. The expert system was developed based on measuring chemical-ER binding and ER-mediated gene activation using *in vitro* assays. The two assays used are: (i) a rainbow trout competitive binding assay to directly measure chemical-biological interaction and (ii) a trout liver slice assay in which the consequences of ER activation or inhibition are measurable as a result of chemical-tissue uptake and partitioning in the presence of xenobiotic metabolism. Experimental conditions were established for these *in vitro* assays that would detect low affinity ER binders of the type found on the chemical inventories the model was built to predict.
An Unambiguous Algorithm	This principle is not applicable to the ER expert system because it is not a regression model but is instead a model based on decision tree logic approach that did not employ algorithms.

Defined Domain of Applicability	The ER expert system was developed to cover two specific EPA chemical inventories - food use pesticide inert ingredients (FI), and antimicrobial active ingredients (AM). A systematic study of chemicals with low binding affinity was undertaken to establish (Q)SAR training sets (*in vitro* assays) that reflected current understanding of the ER binding domain, and that also was representative of the chemical groups in the specific FI and AM chemical inventories to which the expert system was to be applied. The system was then used to extrapolate from a smaller number of chemicals shown to bind ER *in vitro* to the larger chemical space of a defined regulatory chemical inventory through strategic testing within chemicals classes present in the inventories. Representative chemicals in > 30 chemical classes were tested to cover the >600 structures in the FI and AM inventories as well as gain a basic understanding of chemical structural attributes associated with rainbow trout ER binding. The information gained from the testing and examination of relationships between Log K_{ow} and binding affinity was used to develop a decision logic Expert System to prioritize the FI and AM chemicals.
Appropriate Measures of Goodness-of–Fit, Robustness, and Predictivity	This principle is not applicable to the ER expert system because this model is based on a decision tree logic approach and not a regression model, therefore these summary statistics are not appropriate. The decision tree provides localized training sets (assay data) for chemical classes and subclasses covering the entire FI and AM inventories.
Mechanistic Interpretation	This expert system is based on the concept of an adverse outcome pathway which links the molecular event that initiates the pathway (*i.e.*, chemical binding to the ER) with a series of measures that can be made at successively higher and more complex levels of biological organization that are plausibly linked to an adverse outcome. The key events in this pathway include: • Initiation of events by a chemical binding the ER as a result of sufficient chemical uptake into the organism and partitioning to a target tissue with ER-containing cells; • Cell and tissue level gene transcription and translation, *e.g.*, activation of ER responsive genes indicated by vitellogenin (Vtg; an egg-yolk pre-curser) protein production in fish liver; • Organ effects (*e.g.*, appearance of ova in male fish testicular tissue); and • Adverse reproductive and developmental outcome(s) observed in the individual (*e.g.*, change in secondary sex characteristics (feminization of males); cessation of spawning in females; complete sex reversal (genetic males with fully developed and functioning ovaries). This ER expert system is based on empirical data from the first two key events. An understanding of the energetic and steric characteristics of the ER binding domain provides the means to establish a mechanistic basis for defining a chemical structure space associated with ER ligands. Rainbow trout liver assay results provide confirmation that the ER binding translates to gene activation or antagonism at the next, higher level of biological organization along the ER-mediated AOP (*e.g.*, tissue/organ level) and increases confidence in the linkage between low affinity ER binding and gene activation.

References

National Academy of Sciences, *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy* (Report in Brief), 2007, pp. 196. http://dels.nas.edu/resources/static-assets/materials-based-on-reports/reports-in-brief/Toxicity_Testing_final.pdf

Organization for Economic Cooperation and Development (OECD), 2004, *The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs*. Series on Testing and Assessment. No. 49. December 17, 2004. ENV/JM/MONO(2004)24.

OECD, 2007, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SARs] Models*. Series on Testing and Assessment. No. 69. ENV/JM/MONO(2007)2. March 30, 2007.

OECD, 2009a, *Report of the Expert Consultation to Evaluate an Estrogen Receptor Binding Affinity Model for Hazard Identification*. Task Force on Hazard Assessment. ENV/JM/HA/RD(2009)1. March 2, 2009, pp. 107.

OECD, 2009b, *Report of the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox*, Series on Testing and Assessment, No. 101. February 2009. ENV/JM/MONO(2009)4.

Schmieder PK, Tapper MA, Denny JS, Kolanczyk RC, Sheedy BR, Henry TR, Veith GD. 2004. *Use of trout liver slices to enhance mechanistic interpretation of estrogen receptor binding for cost-effective prioritization of chemicals within large inventories*. Environmental Science and Technology. 38(23):6333-42.

U.S. EPA, *Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report*, August 1998. http://www.epa.gov/endo/pubs/edspoverview/finalrpt.htm

U.S. EPA, *SAP: The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing*, August 2009. http://www.epa.gov/scipoly/sap/meetings/2009/august/082509minutes.pdf

U.S. EPA, *Endocrine Disruptor Screening Program for the 21st Century (EDSP21): Summary Overview*, September 2011. http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf

U.S. EPA, *SAP: Integrated Approaches to Testing and Assessment Strategies: Use of New Computational and Molecular Tools*, May 2011. http://www.epa.gov/scipoly/sap/meetings/2011/may/052411minutes.pdf

U.S. EPA, *Endocrine Disruptor Screening Program Comprehensive Management Plan*, June 2012. http://www.epa.gov/endo/pubs/EDSP-comprehensive-management-plan.pdf