



Office of Pollution Prevention and Toxics

[ChemRTK Home](#)[HPV Challenge Program](#)

The Use of Structure-Activity Relationships (SAR) in the High Production Volume Chemicals Challenge Program

I. Introduction

Under EPA's High Production Volume (HPV) Chemical Challenge Program ("Challenge Program") the chemical industry is being challenged to voluntarily compile a Screening Information Data Set (SIDS) for chemicals on the US HPV list. The SIDS, which has been internationally agreed to by member countries of the Organization for Economic Cooperation and Development (OECD), provides basic screening data needed for an initial assessment of the physicochemical properties, environmental fate, and human and environmental effects of chemicals ([Appendix A](#)). The information used to complete the SIDS can come from either existing data or from new tests conducted as part of the Challenge Program.

The Challenge Program chemical list, available online at <http://www.epa.gov/chemrtk/volchall.htm>, consists of about 2,800 HPV chemicals reported under the Toxic Substance's Control Act's 1990 Inventory Update Rule (IUR). The large number of chemicals on the list makes it important to reduce the number of tests to be conducted, where this is scientifically justifiable. Structure-activity relationships, or SAR, may be used to reduce testing in at least three different ways. First, by identifying a number of structurally similar chemicals as a group, or category, and allowing selected members of the group to be tested with the results applying to all other category members¹. Second, by applying SAR principles to a single chemical that is closely related to one or more better characterized chemicals ("analogs"). The analog data are used to characterize the specific endpoint value for the HPV candidate chemical. Third, a combination of the analog and category approaches may be used for individual chemicals. For example, one could search for a "nearest chemical class" as opposed to a nearest single chemical analog to estimate a SIDS endpoint. Such an approach is used in ECOSAR, an SAR-based computer program that generates ecotoxicity values.

EPA has developed this guidance document to assist sponsors and others in constructing and supporting SAR arguments for potential application in the Challenge Program. The guidance will draw on experience from the OECD SIDS program, the EPA Premanufacture Notification (PMN) program, and other sources available in the literature.

¹The development and use of categories in the Challenge Program is the subject of a separate [guidance document](#).

II. DEFINITIONS

A **Structure-Activity Relationship (SAR)** is the relationship of the molecular structure of a chemical with a

physicochemical property, environmental fate attribute, and/or specific effect on human health or an environmental species. These correlations may be qualitative (simple SAR) or quantitative (quantitative SAR, or QSAR).

Qualitative predictions are based on a comparison of valid measured data from one or more analogs (i.e., structurally similar compounds) with the chemical of interest. For example, terms such as “similarly toxic”, “less toxic”, or “more toxic” would be used in a qualitative SAR assessment for toxicity to humans or environmental species.

Quantitative predictions, on the other hand, are usually in the form of a regression equation and would thus predict dose-response data as part of a QSAR assessment.

Using SAR for categories offers a different situation than its use with single chemicals. Although the same SAR principles apply, having multiple chemicals in a category means that experimental data are available for two or more category members allowing for a trend analysis that, in favorable cases, can be used to interpolate or extrapolate to other category members with a certain level of confidence. On the other hand, in the case of a single chemical approach, use of data on a chemical analog requires more rigorous justification to achieve an adequate characterization of endpoints for which data gaps are present.

III. PROCEDURES FOR THE USE OF SAR IN THE CHALLENGE PROGRAM

A. General: Use of SAR for both the Category² and Individual Chemical Approaches

The OECD SIDS Program (OECD 1997), the European Union (Joint Research Centre, or JRC 1998), and the EPA Office of Pollution Prevention and Toxics, or OPPT (in the Premanufacture Notification Program, or PMN Program) have all used QSAR analyses to estimate physicochemical properties, environmental fate endpoints, and environmental (aquatic) effects. In addition, the OECD SIDS Program and OPPT have used qualitative SAR to assess human health hazard potential.

It is important to note the differences in both function and use of SARs among the OECD SIDS Program, the European Union New Chemicals/Existing Chemicals Program, and the OPPT PMN Program. The purpose of the OECD SIDS program parallels that of the Challenge Program - to collect, via a voluntary mechanism, a minimal set of hazard information on HPV chemicals. Because the OECD SIDS program has been active for more than a decade, the application of SAR in such a program is directly applicable to the Challenge Program. However, because the written OECD guidance on the use of SAR tends to be general³, EPA believes it is more useful to rely on how the OECD considered SAR in the context of chemical case histories at SIDS Initial Assessment Meetings (SIAMs). Some examples are described in [Appendix B](#).

²This subsection applies to both the category and analog approaches; however, readers are referred to the HPV Challenge [Guidance Document on Categories](#) for specific information on categories.

³The SIDS Manual (www.oecd.org/ehs/sidsman.htm) guidance on the use of SAR in the OECD SIDS program consists mainly of citations to OECD and other documents. The Manual does state that QSAR is acceptable for physicochemical properties or aquatic toxicity; although for the latter it states "...there is a preference for using measured data...however, if appropriate QSARs are available, they could be used..." There is no specific guidance for the use of SAR in assessing mammalian toxicity. The Manual also lists some examples of the potential use of SAR: groups of isomers with similar SAR profiles; close homologs; and availability of information on precursors, breakdown products, and metabolites/degradation products of specific chemicals.

The European Union has a variety of directives that regulate new and existing chemicals that are conceptually similar to

those in place in the U.S. and elsewhere. Historically, the role of SARs in these directives has been minimal. The JRC report (1998) documents how SARs will be used in European risk assessments using the new EUSES (European Union System for the Evaluation of Substances) program.

The purpose of the OPPT PMN Program is to screen new chemicals for potential hazard *and* potential risk before they are manufactured or used⁴. The PMN Program has had almost two decades of experience using SAR. PMN submitters are not required to generate new data in support of a new chemical submission. Consequently, the use of SARs for new chemicals plays a larger role than its use for existing - and especially HPV - chemicals. The PMN program use of SAR is based on a “nearest analog” assessment to estimate health effects where test data are lacking, and the use of a chemical class/statistical-based QSAR method to assess ecotoxicity (ECOSAR). The PMN Program has developed a database of over 50 chemical classes that represent potential health and/or ecological concern (available online at www.epa.gov/opptintr/newchms/chemcat.htm).

It is important to introduce the concept of how and why SARs may be used in the HPV Challenge Program ([Table 1](#)). Methods available to estimate physicochemical properties generally assign values to atoms, bonds, and their placement in a molecule. These QSARs yield regression equations that estimate a given endpoint.

⁴OPPT was involved in a collaborative study with the European Commission to compare the PMN SAR techniques and predictions with actual data developed in Europe. The study included over a hundred chemicals and covered some SIDS endpoints (USEPA 1994a and OECD 1994).

Approach	SIDS Endpoint	Comment
Category	All ¹	Assemble information on all endpoints for all category members to determine whether trends exist that would allow adequate characterization
Nearest Analog	Health	Depends upon existing data for analog chemical to estimate the effect of the HPV candidate chemical.
“Nearest Chemical Class”	Ecotoxicity, Degradation	Depends upon the placement of the HPV candidate chemical in an existing chemical class that is part of a QSAR.
Other QSAR	Physicochemical	Estimations based on chemical bonds and where located in the candidate chemical.

¹ In some cases, there may be an opportunity to use nearest analog, nearest chemical class, or other QSAR approaches within a category (see example 4 in Appendix B).

The environmental fate and aquatic toxicity SARs rely heavily on physicochemical properties as inputs, and are similarly structured in terms of models, chemical classes, and regression equations. However, “accepted QSARs” (cases in which ample data are available for a given chemical class) are not available for certain chemical classes for either ecotoxicity endpoints estimated using ECOSAR or biodegradation endpoints estimated using BIOWIN (see [Section IV](#) for details).

SARs for health effects are different from the other SIDS endpoints. This is due to the variety of scenarios (acute vs.

chronic exposure conditions, in vitro vs. in vivo tests) and endpoints (e.g., general toxicity, organ-specific effects, mutagenicity, developmental effects, effects on fertility). Therefore, generic QSAR models are either not readily available or not widely accepted (see [Hulzebos](#) et al. 1999 for review), and an analog approach is a reasonable way to proceed.

B. Scope and Applications in the Use of SAR/QSAR in the U.S. HPV Challenge Program

The use of SAR/QSAR in the HPV Challenge Program is expected to decrease the number of new tests required to develop a SIDS for each HPV chemical. Their use, by either the category or individual chemical approach, will necessarily be limited by the nature of the SIDS endpoint, the amount and adequacy of the existing data, and the type of SAR/QSAR analysis performed. ***Measured data developed using acceptable methods are preferred over estimated values.***

The development and use of SAR/QSAR in the Challenge Program will be different for each of the major categories of SIDS (i.e., physicochemical properties, environmental fate, ecotoxicity, and health effects). ***In the final analysis, because the goal of the Program is to adequately characterize the hazard of HPVs, a careful, reasonable, and transparent argument using measured data and estimation techniques will need to be presented.***

Physicochemical properties. It is anticipated that melting point, boiling point, vapor pressure, octanol/water partition coefficient, and water solubility data will be available for most HPVs. In some cases, this will be in the form of values taken from standard reference books (e.g., the Merck Index, CRC Handbook of Chemical and Physical Properties). In the event that neither measured data nor reference book values are available, estimations using an appropriate model (see [Section IV](#)) will be accepted for all physicochemical endpoints.

Environmental fate. Acceptable estimation techniques are available for photodegradation and hydrolysis, whereas biodegradation models are less available and less well-accepted. The fourth SIDS endpoint in this category is a model (fugacity models to estimate transport/distribution), and so there is no measured data requirement to fulfill. Thus, estimations will be acceptable in lieu of photodegradation and hydrolysis tests, but not for biodegradation.

Ecotoxicity. ECOSAR is an established QSAR program which estimates toxicity to fish, invertebrates, and algae. Even though this approach represents a screening-level characterization, it is of a higher order than either physicochemical or environmental fate tests. This is not to diminish the importance of physicochemical/environmental fate tests, but there are layers of complexity not present in these endpoints when toxicity is the entity being measured/estimated. Therefore, *some* measured data must be available to strengthen the use of ECOSAR to characterize aquatic toxicity for an HPV chemical in the Challenge Program. For example, if an ECOSAR (or other aquatic toxicity SAR estimation procedure) is to be presented for any one endpoint, it must be accompanied by experimental data on that endpoint with a close analog.

Health Effects. As stated above for ecotoxicity, the use of SARs to estimate toxicity is more complicated than its use in estimating physicochemical/environmental fate. The estimation of toxicity to mammals is even more complicated than the estimation of aquatic toxicity because there is a variety of endpoints

(mutagenicity vs. general toxicity vs. reproductive/developmental toxicity) and exposure (in vitro vs. in vivo and acute vs. chronic) conditions. Also, unlike ecotoxicity, the available SAR programs are very different from each other, unique to certain endpoints, and most are not validated (see [Hulzebos et al., 1998](#) for review). Therefore, *in all cases*, SAR estimations for a health endpoint must be accompanied by experimental data with a close analog.

C. Individual Chemical Approach

For individual chemicals, SAR is applied in two ways: (1) by the use of (usually quantitative) predictive models based on well-validated data sets (QSAR); (2) by comparing the chemical to one or more closely related chemicals, or **analogs**, and using the analog data in place of testing the chemical. In the case of models, the comparison has essentially been incorporated into the model.

In developing an SAR, proposers need to consider the following steps for each HPV chemical they are interested in sponsoring (presented schematically in [Figure 1](#) and discussed more fully below):

Step 1: Conduct literature search

Step 2: Determine data adequacy by SIDS endpoint

Step 3: Identify data gaps by SIDS endpoint

Step 4: Use SAR or perform test, by SIDS endpoint

STEP 1: Conduct Literature Search

Gather published and unpublished literature on physicochemical properties, environmental fate and effects, and health effects for the HPV chemical of interest. This should include all existing relevant data and not be limited to the SIDS endpoints (e.g., metabolism and cancer studies are relevant but not formally part of SIDS). (LINK to [literature search strategy guidance document](#)).

STEP 2: Determine Data Adequacy by SIDS Endpoint

Evaluate available data for adequacy. Please see EPA guidance document on [Data Adequacy](#).

STEP 3: Identify Data Gaps by SIDS Endpoint

Determine if adequate, available data have been identified for a given SIDS endpoint. If not, then there is a data gap for that endpoint. Because SIDS represents a base hazard data set, any data gap must be filled to meet the Challenge Program commitment.

STEP 4: Use SAR or Test to Fill the Data Gaps For Each SIDS Endpoint⁵

If the chemical can be rationally placed in a category for a category-type SAR analysis, or if there is a desire to use

either a QSAR model or available information on an analog, EPA suggests the following procedure:

- A. If the chemical can be placed in a category, see EPA guidance [document on categories](#)
- B. If a QSAR model is available (e.g., models available to estimate certain environmental fate properties, or ECOSAR for aquatic ecotoxicity), it may be used with the appropriate rationale for its applicability to the HPV candidate chemical. It is important to consider whether the model has been validated for the structural class to which the compound in question belongs. (See [Section III.B.](#) and [Section IV.](#))
- C. If the analog approach is used, the following guidance is offered:

1. **Identify analog(s) for each SIDS endpoint.**

Identification of the appropriate analog for the HPV candidate chemical is complicated by the likelihood that the SAR may differ for different SIDS endpoints. Thus, it is necessary to look for an analog for each SIDS endpoint for which there is a data gap.

The most likely analogs are chemicals that resemble the candidate chemical in terms of: (1) molecule structure/size; (2) some substructure that may play a critical functional role (including whether the chemical belongs to a series of well-studied structural analogs known to produce a particular kind of effect); (3) some molecular property (i.e., lipophilicity, electronic and steric parameters); and/or (4) some precursor, metabolite, or breakdown product. **Sponsors must include the rationale for their choice of analog(s).**

An obvious but important point is that analogs need not themselves be HPV chemicals, as the focus is on the analog data available, their adequacy, and whether they can support an SAR.

2. **Conduct literature search on the analog and evaluate for data adequacy.** Data used for SAR purposes must be scientifically sound and unambiguous. Just as the available data on the HPV chemical must be adequate to obviate testing for an endpoint, analog data must meet data adequacy criteria in order to support an SAR claim (i.e., the data must be adequate to support a no test decision for the analog endpoint just as if it was an HPV chemical). See EPA guidance on [data adequacy](#).

3. **Evaluate the relationship of the analog to the HPV chemical for each SIDS endpoint.** The fundamental basis for an SAR lies in the structural, metabolic and other relationships between the chemical and its analog(s). These relationships must be substantial and unambiguous in order to be acceptable in the HPV Challenge Program.

For example, where the postulated SAR relies on a metabolic transformation, consider such factors as whether pharmacokinetic studies exist that demonstrate the conversion and that the rate of conversion supports the use of metabolite data to represent the parent compound. Conversely, consider whether there are structural features that may interfere

with conversion to the analog and thus nullify the SAR argument.

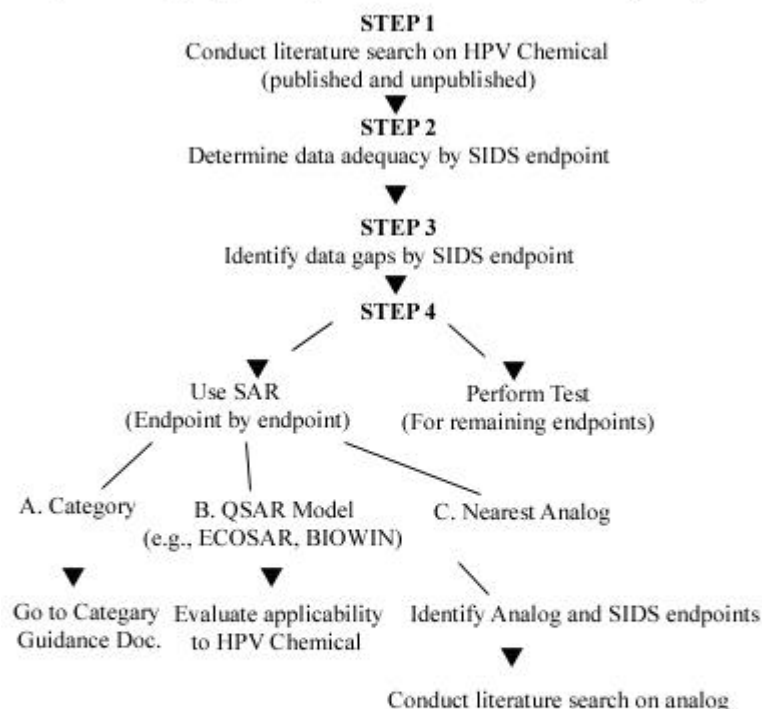
4. **Develop SAR Proposal in Test Plan.** It is essential to construct a logical, tightly reasoned, convincing written proposal. This is not to discourage creativity but to emphasize the importance of generating reliable information as the principal purpose of the program.

Sponsors will need to make an SAR proposal and rationale available to EPA and others for review, indicating proposed tests and SAR predictions in the finalized test plan (see Appendices in Data Adequacy and Category guidance document for a discussion of test plans). While sponsors are ultimately responsible for the success of their proposals, EPA's position on individual proposals will reflect its need to anticipate the acceptability of the results in EPA's own chemical assessment programs and in OECD SIDS as appropriate.

Participants should bear in mind that new information generated by testing might in turn be used to confirm or support SAR arguments that are currently uncertain.

⁵Examples of this step are provided in Appendix B

Figure 1: Process for Developing SAR Proposals for Single HPV Chemicals (The "Analog Approach") in the HPV Chemical Challenge Program



NAME

IV UNDERLYING PRINCIPLES OF SAR BY SIDS ENDPOINTS

In this section, brief reviews of the SAR/QSAR methods used by OPPT for each of the major SIDS categories are presented. This review is not intended to be comprehensive, but is provided for illustrative/guidance purposes only. [Table 2](#) at the end of this section lists the SAR models discussed.

A. Physicochemical Estimation Techniques

Methods exist for estimating most of the physicochemical properties required to develop a basic understanding of the behavior of a chemical released to the environment and its potential environmental exposure pathways. Some of the methods require input as simple as chemical structure, while others require much less readily available information such as water solubility values, octanol/water partition coefficient, etc. Estimation methods for key physicochemical properties have been reviewed by Howard and Meylan (1997) and are discussed briefly below.

Boiling Point, Melting Point and Vapor Pressure. Most comprehensive estimation methods for boiling point, melting point, and vapor pressure are “group contribution” methods, where values assigned to atoms, bonds, and their placement in a molecule are used to estimate their contribution to the inherent physicochemical properties of that molecule. The Stein and Brown (1994) method for estimating boiling points was developed and validated on a large database (>10,000 chemicals) and has been integrated into a computer program (MPBPVP) used by OPPT. In contrast, melting points are not very well estimated by this method so the group contribution method is combined with an algorithm that relates melting point with boiling points to estimate melting point. This method is used in MPBPVP. Recently, attempts have been made to use molecular symmetry (Simamora and Yalkowski 1994; and Krzyzaniak et al. 1995), but the methods have not been well documented or validated.

A limited number of methods are available for estimating vapor pressure. Most rely on estimating the vapor pressure from the boiling point and use melting points when the chemical is a solid at room temperature, which is the method used by OPPT in MPBPVP.

Octanol/water partition coefficient. The octanol/water partition coefficient describes the lipophilic properties of a chemical. Since measured values range from $<10^{-4}$ to $>10^{+8}$, the logarithm (log P) is commonly used to express its value.

The literature contains many methods for estimating log P. The most common are classified as "fragment constant" methods in which a structure is divided into fragments (atom or larger functional groups) and values of each group are summed together (sometimes with structural correction factors) to yield the log P estimate (Meylan and Howard 1995; Hansch and Leo 1979, 1995; Hansch et al. 1995). OPPT's KOWIN model is based on the fragment constant method. General estimation methods based upon molecular connectivity indices (Niemi et al. 1992), UNIFAC-derived activity coefficients (Banerjee and Howard 1988), and properties of the entire solute molecule (charge densities, molecular surface area, volume, weight, shape, and electrostatic potential) (Bodor et al. 1989; Bodor and Huang 1992; Sasaki et al. 1991) have also been developed.

Water Solubility. Water solubility is a determining factor in the fate and transport of a chemical in the environment as well as the potential toxicity of a chemical. Yalkowsky and Banerjee (1992) have reviewed most of the recent literature on aqueous solubility estimation and concluded that, at present, the most practical means of estimating water solubility involves regression-derived correlations using log P. OPPT uses the log-P based WSKOW model to estimate water solubility. Recently, direct fragment constant approaches to estimating water solubility have been developed (Myral et al. 1995; Meylan and Howard 1996; Kuhne et al. 1995).

B. Environmental Fate Estimation Techniques

Biodegradation. Biodegradation (i.e., complete mineralization, or conversion to carbon dioxide and water) is an important environmental degradation process for organic chemicals. Prediction of biodegradability is severely limited because of the lack of reproducibility of biodegradation data (Howard et al. 1987) as well as the numerous protocols that have been used for biodegradation tests (Howard and Banerjee 1984). As a result, quantitative prediction of biodegradation rates has only been attempted on very limited numbers of structurally related chemicals (Howard et al. 1992). A number of comprehensive approaches using fragment constants have been attempted to qualitatively predict biodegradability.

Many of the models have used a weight-of-evidence biodegradation database (BIODEG) that was specifically developed for structure/biodegradability correlations (Howard et al. 1986). Boethling et al. (1994) used the experimental BIODEG database as well as results of an expert survey to develop four models (these models are in the OPPT program called BIOWIN) that all used the same structural fragments; these structural fragments were selected from previously known "rules of thumb" (e.g., increasing the number of chlorines on aromatic ring results in increased persistence). The structural fragments in the other models were mostly selected by statistical significance, rather than previous indication of correlation to biodegradability.

Hydrolysis Rates. Hydrolysis is the reaction of a substance with water in which the water molecule or the hydroxide ion displaces an atom or group of atoms in the substance. Chemical hydrolysis at a pH normally found in the environment (i.e., pH 5 to 9) can be important for a variety of chemicals that have functional groups that are potentially hydrolyzable, such as alkyl halides, amides, carbamates, carboxylic acid esters and lactones, epoxides, phosphate esters, and sulfonic acid esters (Neely 1985). Only a method to predict hydrolysis rate constants for esters, carbamates, epoxides, and halogenated alkanes has been developed using LFER (Taft and Hammett constant) methodology. A computer program (HYDROWIN) that uses this methodology is available and is used by OPPT. Also, Ellenrieder and Reinhard (1988) have developed a spreadsheet program that allows hydrolysis rates to be calculated at different pHs and temperatures if adequate data are available in the companion database.

Atmospheric Oxidation Rates (An assessment of Photodegradation). For most chemicals in the vapor phase in the atmosphere, reaction with photochemically generated hydroxyl radicals is the most important degradation process (Atkinson 1989). Methods for estimating reactivity with hydroxyl radicals have generally relied on fragment constant approaches or molecular orbital calculations. The method validated on the largest number of chemicals (641) is the Atkinson fragment and functional approach method (the method used in AOPWIN, the model used by OPPT), although molecular orbital methodology gives promising results on a much more limited number of chemicals.

C. Ecological Endpoint Estimation Techniques

(Q)SARs for aquatic toxicity to fish, aquatic invertebrates, and algae have been developed and used by OPPT since 1979 (USEPA 1994b,c). These (Q)SARs have been incorporated into a software program (ECOSAR) available free from the EPA website at www.epa.gov/opptintr/newchms (click on the ECOSAR button).

ECOSAR uses molecular weight and structure and $\log K_{ow}$ to predict aquatic toxicity. The predictions are based on actual data of at least one member of a chemical class. The data (measured toxicity values) are correlated with molecular weight and $\log K_{ow}$ to derive a regression equation that may be used to predict aquatic toxicity of another chemical that belongs to the same chemical class. ECOSAR contains equations for many chemical classes (>50 - the full list can be found at www.epa.gov/opptintr/newchms/chemcat.htm) which can be categorized into four main areas:

- A. Neutral organics that are nonreactive and nonionizable;
- B. Organics that are reactive and ionizable and that exhibit "excess toxicity" (toxicity beyond narcosis associated with neutral organic toxicity);
- C. Surface-active organic compounds such as surfactants and polycationic polymers; and
- D. Inorganic compounds including organometallics.

Therefore, to use ECOSAR for a particular chemical it is necessary to select an appropriate SAR based on the following: chemical structure, chemical class, predicted $\log K_{ow}$, molecular weight, physical state, water solubility, number of carbons, ethoxylates or both, and percent amine nitrogen or number of cationic charges or both, per 1000 molecular weight. Because the regression equations are chemical-specific, and because they may vary by species (fish vs. daphnid vs. algae), the most important factor is the identification of the chemical class (USEPA 1994b).

The following presents some guidance on the approach for evaluating the aquatic toxicity (to fish, plants, and invertebrates) of a candidate HPV chemical using ECOSAR:

1. Identify the chemical structure and convert it to SMILES⁶ notation;
2. Identify appropriate physicochemical properties: physical state, melting point, water solubility, vapor pressure, and K_{ow} are required to predict effect concentrations (i.e., EC_{50}). If a chemical is highly water-reactive (for example, a hydrolysis half-life less than one hour) consider estimating toxicity for the hydrolysis products (s);

3. Decide what ECOSAR chemical class best fits your chemical⁷ ; and
4. Run the ECOSAR program to develop an aquatic toxicity profile for the candidate chemical.

⁶SMILES (Simplified Molecular Input Line Entry System) converts chemical structures into a string of characters that are easily entered into a computer program. For more information see Weininger (1998) or either of the following websites: www.daylight.com or <http://esc.syrres.com>.

⁷There is a range of data points that support each ECOSAR chemical class. Users are encouraged to review these background data to determine the applicability of the ECOSAR results for their particular chemical and chosen chemical class.

D. Health Endpoint Estimation Techniques

Hulzebos et al. (1999) reviewed the literature on QSARs for human toxicological endpoints and divided the available estimation techniques into three groups: rule-based systems (e.g., HazardExpert, DEREK); statistically-based systems (TOPKAT, MULTICASE); and systems that are a combination of the two (RASH). Rule-based SARs rely on placing chemicals into categories by presumed mechanism of action, and statistical-based SARs use statistically-derived descriptors to predict the activity of a chemical and thus may be applicable to a more heterogenous group of chemicals.

Hulzebos et al. noted that more validation is needed to correlate SAR with individual health endpoints. For the purposes of the U.S. HPV Challenge Program - to adequately characterize the hazard of an HPV- the above mentioned models could not replace an actual test.

However, there is an opportunity to use SAR for health endpoints in the Challenge Program. Given the complexity of health endpoints, and the amount of uncertainty in many models, OPPT has historically used an expert judgment/nearest analog approach to SAR for predicting such effects in assessing new chemicals. OPPT suggests that a similar approach be applied in the Challenge Program.

The goal is to find toxicity data for an analog that can be used to address the testing needs of an HPV chemical. This is best done on an endpoint-by-endpoint and case-by-case basis.

Valid analogs should have close structural similarity and the same functional groups. In addition, the following parameters should be compared between the chemical and its analog(s): physicochemical properties - physical state, molecular weight, log Kow, water solubility; absorption potential; mechanism of action of biological activity; and metabolic pathways/kinetics of metabolism. A high correlation between the HPV chemical and the putative analog for most of these parameters improves the chance that an SAR approach will be reasonable and acceptable.

A more convincing argument can be made for the use of surrogate data if there are toxicity studies in common (i.e., ones that are not necessarily SIDS endpoints, but have been done with both the analog and the HPV candidate chemical) that demonstrate the toxicological similarity of the chemicals.

The following presents possible examples of the use of surrogate data to characterize individual chemicals:

- 1) Chemicals that are essentially the same *in vivo*. For example different salts of the same anion or

cation. The salts must fully dissociate *in vivo* and the counter ion must not contribute any more (or less) toxicity.

2) A chemical that metabolizes to one (or more) compounds that have been tested. The metabolism must be rapid and complete.

3) Chemicals that have only minor structural differences that are not expected to have an impact on toxicity. All functional groups must be the same.

E. Summary

Table 2 provides a summary of the SAR models discussed above.

Table 2: SAR Models Used by EPA for Each SIDS Endpoint				
SIDS Category	SIDS Endpoint	SAR Model	Required Input	Model Availability
Chemical and Physical Properties ¹	Melting point	MPBPVP	CAS # and/or SMILES Notation	Available from Syracuse Research Corp. (SRC) at: http://esc.syrres.com/~esc1/
	Boiling point			
	Vapor pressure			
	Partition coefficient (log K _{ow})	KOWWIN		
Water solubility	WSKOW			
Environmental Fate and Pathways ^{1,2}	Photodegradation	AOPWIN		
	Stability in Water	HYDROWIN		
	Biodegradation	BIOWIN		
Ecotoxicity Tests	Acute toxicity to fish, aquatic invertebrates, and algae	ECOSAR		May be downloaded from: www.epa.gov/opptintr/newchms
Human Health Effects	Acute Toxicity			Nearest analog analysis using expert judgment (see text).
	General Toxicity (repeated dose)			
	Genetic Toxicity (effects on the gene and chromosome)			
	Reproductive/Developmental Toxicity			
¹ The Estimations Programs Interface program for Windows (EPIWIN) is used by OPPT to run selected estimations programs for a variety of endpoints. The chemical structure or CAS number is entered only once, and EPIWIN executes all of the programs and captures their output. (Appendix C has a sample output).				
² Transport/distribution is another SIDS endpoint in this category, but no experimental studies are required, only use of the EQC model (see Mackay et al., 1996, <i>Env. Tox. Chem.</i> [15][9]: 1627-1637).				

References:

Atkinson, R. 1989. Kinetics and mechanisms of the gasphase reactions of the hydroxyl radical with organic compounds. *J. Phys. Chem. Ref. Data Monograph No. 1*. American Institute of Physics & American Chemical Society, New York, NY, USA.

Banerjee, S. and P.H. Howard. 1988. Improved estimation of solubility and partitioning through correction of UNIFAC-derived activity coefficients. *Environ. Sci. Technol.* 22:839-841.

Boethling, R.S., P.H. Howard, W.M. Meylan, W. Stiteler, J. Beauman and N. Tirado. 1994. Group contribution method for predicting probability and rate of aerobic biodegradation. *Environ. Sci. Technol.* 28:459-465.

Bodor, N., Z. Gabanyi and C.K.Wong. 1989. A new method for the estimation of partition coefficient. *J. Amer. Chem. Soc.* 111:3783-3786.

Bodor, N. and M.J. Huang. 1992. An extended version of a novel method for the estimation of partition coefficients. *J. Pharm. Sci.* 81:272-281.

Ellenrieder, W. and M. Reinhard. 1988. Athias - an information system for abiotic transformations of halogenated hydrocarbons in aqueous solution. *Chemosphere* 17:331-44.

Hansch, C. and A.J. Leo. 1979. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York, NY, USA.

Hansch, C. and A. Leo. 1995. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society, Washington, DC, USA.

Hansch, C., A. Leo and D. Hoekman. 1995. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*. American Chemical Society, Washington, DC, USA.

Hilal, S.H., L.A. Carreira and S.W. Karickhoff. 1994. Estimation of chemical reactivity parameters and physical properties of organic molecules using SPARC. In *Quantitative Treatments of Solute/Solvent Interactions: Theoretical and Computational Chemistry* Vol. 1 Pub. City, St, USA pp 291-353. Elsevier, New York, NY, USA.

Howard, P.H. and S. Banerjee. 1984. Interpreting results from biodegradability tests of chemicals in water and soil. *Environ. Toxicol. Chem.* 3:551562.

Howard, P.H. and W.M. Meylan. 1997. *Prediction of Physical Properties, Transport, and Degradation for Environmental Fate and Exposure Assessments*. IN: Quantitative Structure-Activity Relationships in Environmental Sciences VII., edited by F. Chen and G. Schuurmann. SETAC Press, Pensacola, FL. Pages 185-205.

Howard, P.H., A.E. Hueber and R.S. Boethling. 1987. Biodegradation data evaluation for structure/biodegradability relations. *Environ. Toxicol. Chem.* 6: 110.

Howard, P.H., R.S. Boethling, W.M. Stiteler, W.M. Meylan, A.E. Hueber, J.A. Beauman and M.E. Larosche. 1992. Predictive model for aerobic biodegradability developed from a file of evaluated biodegradation data. *Environ. Toxicol. Chem.* 11:593-603.

Howard, P.H., A.E. Hueber, B.C. Mulesky, J.C. Crisman, W.M. Meylan, E. Crosbie, D.A. Gray, G.W. Sage, K. Howard, A. LaMacchia, R.S. Boethling and R. Troast. 1986. BIOLOG, BIODEG, and fate/expos: new files on microbial degradation and toxicity as well as environmental fate/exposure of chemicals. *Environ. Toxicol. Chem.* 5:977-80.

Hulzebos, E.M., P.C.J.I. Schielen, and L. Wijkhuizen-Maslankiewicz. 1999. (Q)SARs for human toxicological endpoints: a literature search. A report by the RIVM (Research for Man and Environment), The Netherlands, RIVM Report 601516.001

Joint Research Centre, 1998. *Technical Guidance Documents in Support of The Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and The Commission Regulation (EC) 1488/94 on Risk Assessment for Existing Substances*. A report by the Joint Research Centre, European Chemicals Bureau, European Commission. (No report number or other type of identifier in the report). Chapter 4, pp. 505-566.

Kuhne, R., R.-U. Ebert, F. Kleint, G. Schmidt and G. Schuurmann. 1995. Group contribution methods to estimate water solubility of organic chemicals. *Chemosphere* 30:2061-2077.

Krzyzaniak, J.F., P.B. Myrdal, P. Simamora and S.H. Yalkowsky. 1995. Boiling and melting point prediction for aliphatic, non-hydrogen-bonding compounds. *Ind. Eng. Chem. Res.* 34:2530-2535.

Lyman, W.J., W.F. Reehl and D.H. Rosenblatt. 1990. *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds*. American Chemical Society, Washington, DC, USA.

Meylan, W.M. and P.H. Howard. 1995. Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* 84:83-92.

Meylan, W.M. and P.H. Howard. 1996. Water Solubility Estimation by Base Compound Modification: Current Status. Syracuse Research Corp., Environ. Sci. Center. Prepared for U.S. Environ. Protection Agency: Contract No. 68D20141. Washington, DC.

Myrdal, P.B., A.M. Manka and S.H. Yalkowsky. 1995. AQUAFAC 3: Aqueous functional group activity coefficients; Application to the estimation of aqueous solubility. *Chemosphere* 30:1619-1637.

Neely WB. 1985. Hydrolysis. In: W.B. Neely and G.E. Blau, eds. *Environmental Exposure from Chemicals Vol I*. CRC Press, Boca Raton, FL, USA. pp. 157-73.

Niemi, G.J., S.C. Basak, G.D. Veith and G. Grunwald. 1992. Prediction of octanol-water partition coefficient (K_{ow}) with algorithmically derived variables. *Environ. Toxicol. Chem.* 11:893-900.

OECD Organisation for Economic Co-Operation and Development. 1994. U.S. EPA/EC Joint

Project on the Evaluation of (Quantitative) Structure Activity Relationships, OECD Report No. OECD/GD/(94) 28, Paris, France.

Perrin, D.D., B. Dempsey and E.P. Serjeant. 1981. *pK_a Prediction for Organic Acids and Bases*. Chapman and Hall, New York, NY, USA.

Sasaki, Y., H. Kubodera, T. Matsuzaki and H. Umeyama. 1991. Prediction of octanol/water partition coefficients using parameters derived from molecular structures. *J. Pharmacobio.-Dyn.* 14:207-214.

Simamora, P. and S.H. Yalkowsky. 1994. Group contribution methods for predicting the melting points and boiling points of aromatic compounds. *Ind. Eng. Chem. Res.* 33:1405-1409.

Stein, S.E. and R.L. Brown. 1994. Estimation of normal boiling points from group contribution. *J. Chem. Inf. Comput. Sci.* 34:581-587.

USEPA. 1994a. U.S. EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships, Washington, DC: Office of Pollution Prevention and Toxics, US EPA, EPA Report No. EPA 743-R-94-001.

USEPA. 1994b. Estimating toxicity of industrial chemicals to aquatic organisms using SAR, 2nd edition. OPPT. EPA 748-R-93-001. Available from National Center for Environmental Publication and Information, 1-800-490-9198.

USEPA. 1994c. ECOSAR: A Computer Program for Estimating the Ecotoxicity of Industrial Chemicals (EPA-748-R-93-002). Available from National Center for Environmental Publication and Information, 1-800-490-9198.

Weininger, D. 1988. A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comp. Sci.* (28):31-36.

Yalkowsky, S.H. and S. Banerjee. 1992. *Aqueous Solubility Methods of Estimation for Organic Compounds*. Marcel Dekker, Inc. New York, NY, USA.

APPENDIX A

SCREENING INFORMATION DATA SET (SIDS)

SIDS Endpoints		
SIDS Category	Test/Estimation Endpoint	OECD Guideline (or equivalent) ¹
Chemical and Physical Properties	Melting point	OECD 102
	Boiling point	OECD 103
	Vapor pressure	OECD 104
	Partition coefficient (log K _{ow})	OECD 107, 117
	Water solubility	OECD 105, 112
Environmental Fate and Pathways	Photodegradation	-
	Stability in Water	OECD 111
	Biodegradation	OECD 301, 302
	Transport/Distribution	EQC Model ²
Ecotoxicity Tests	Acute toxicity to fish	OECD 203
	Acute toxicity to aquatic invertebrates	OECD 202 ³
	Toxicity to aquatic plants	OECD 201 ³
	Chronic aquatic invertebrate test (When appropriate)	OECD 211 ³
Human Health Effects	Acute Toxicity	OECD 401-403, 420, 423, 425
	General Toxicity (repeated dose)	OECD 407-413, 422
	Genetic Toxicity (effects on the gene and chromosome)	OECD 471-486
	Reproductive Toxicity	OECD 415, 416, 421, 422
	Developmental Toxicity	OECD 414, 421, 422
<p>¹ EPA recognizes that alternate, equivalent test guidelines exist for some of the listed endpoints. For example, guidelines listed by EPA, ASTM, etc. The OECD Guidelines are presented here for both illustration purposes and because the Challenge Program is based on the OECD SIDS Program.</p> <p>² This model is available online from the University of Trent, Ontario, Canada at http://www.trentu.ca/envmodel.</p> <p>³ The OECD is in the process of updating these Guidelines.</p>		

APPENDIX B

The examples presented in this section are from OECD SIDS cases and represent steps that could be taken in Step 4 of the process discussed in the text of this document and shown schematically in Figure 1.

1. Acid-salt pairs.**Chloroacetic acid/sodium salt.**

In this case, both the acid and salt were identified as HPV chemicals. Separate existing data packages (dossiers) were prepared for each chemical in the data collection step. However, available data supported the position that the acid and salt were equivalent for most endpoints; for example, the pKa of 2.8 for the acid suggested that dissociation of the substances in aqueous systems at environmentally relevant pH values was virtually complete. Observed and potential differences were commented upon when appropriate, such as the skin corrosivity reported for the acid. (NOTE: Skin irritation is not a formal SIDS endpoint, but this illustrates the value of considering non-SIDS information in evaluating hazard.)

Data were considered adequate for hazard assessment purposes if available on either chemical for a given endpoint. Thus, developmental toxicity data available only for the salt were considered adequate for assessment of the pair. No testing was considered necessary, because the combined available data for the acid/salt pair covered all the SIDS endpoints. (SIDS Initial Assessment Profile for Monochloroacetic acid and Sodium monochloroacetate, available at the United Nations Environmental Program (UNEP), International Registry of Potentially Toxic Chemicals (IRPTC) website: <http://irptc.unep.ch/irptc/sids/sidspub.html>)

2. Use of Metabolites**Ethyl acetate.**

Reproductive and developmental effects data were not available on this substance. However, there were adequate studies with ethanol for these endpoints. The sponsors supplied data showing that ethyl acetate administered intravenously to rats was rapidly hydrolyzed to ethanol (Deisinger and English, 1998⁸). Ethyl acetate had a half life of less than one minute, with the majority of it being converted to ethanol. EPA accepted the sponsor's argument that available, adequate data on ethanol were sufficient to satisfy the reproductive/developmental endpoints for ethyl acetate.

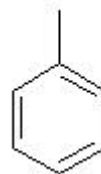
3. Homologous Series**Glycol ethers (Triethylene glycol monomethyl and -ethyl ethers)(TGME, TGEE).**

Where R = CH₃ for TGME and CH₂CH₃ for TGEE and R' = H

These compounds had considerable data available, but TGEE was missing reproductive and genetic toxicity data. Sponsors supplied data for these ethers and a third analog (the monobutyl ether) showing very slow dermal uptake (a major human exposure route) and low overall toxicity for all three chemicals. This SAR argument was based on data from three related chemicals and was accepted by EPA and OECD and no further testing was deemed necessary.

4. Class 2 (Mixture)

Linear alkylbenzenes (LABs).



Where $x + y = 7-13$ and $X = 0-7$

The LABs have been presented as an example of a category analysis in the U.S. [HPV Challenge Program](#). They are also presented here to illustrate how data on one mixture may be useful to fulfill a SIDS endpoint on a "similar mixture". There are nine LAB formulations currently available in commerce. These nine products fall under eight CAS numbers. The individual formulations vary only in the proportion of the chain lengths of the alkyl derivatives present (see [Category document](#) for a more detailed explanation). From an individual chemical SAR standpoint, there are a number of nearest analog opportunities (see Tables B-1, B-2, and B-3 in the [Category document](#)). For example, adequate and available data on Alkylate 215 could be used to evaluate either Nalkylene 500 or Nalkylene 500L because all three have similar (high percentage of smaller chain length alkyl group) makeup.

⁸Deisinger, PJ and JC English. 1998. Pharmacokinetics of ethyl acetate in rats after intravenous administration. Final Report. Laboratory Project ID 97-0300BT01. Sponsored by the Chemical Manufacturers Association and performed at the Toxicological Sciences Laboratory at Eastman Kodak Co., Rochester, NY.

APPENDIX C

SAMPLE OUTPUT FROM PHYSICAL CHEMICAL PROPERTIES / FATE ESTIMATION PROGRAM INTERFACE (EPI)

SMILES : c1(C(CC)C)cc(C(CCC)C)cc(CCCC)c1

CHEM :

MOL FOR: C19 H32

MOL WT : 260.47

-----EPI SUMMARY (v2.30) -

Physical Property Inputs:

Water Solubility (mg/L): -----

Vapor Pressure (mm Hg) : -----

Henry LC (atm-m3/mole) : -----

Log Kow (octanol-water): -----

Boiling Point (deg C) : -----

Melting Point (deg C) : -----

Log Octanol-Water Partition Coef (SRC):

Log Kow (KOWWIN v1.57 estimate) = 8.40

Boiling Pt, Melting Pt, Vapor Pressure Estimations (MPBPWIN v1.26):

Boiling Pt (deg C): 321.72 (Adapted Stein & Brown method)

Melting Pt (deg C): 63.19 (Mean or Weighted MP)

VP(mm Hg,25 deg C): 0.000269 (Modified Grain method)

Water Solubility Estimate from Log Kow (WSKOW v1.27):

Water Solubility at 25 deg C (mg/L): 0.00139

log Kow used: 8.40 (estimated)

no-melting pt equation used

Henrys Law Constant (25 deg C) [HENRYWIN v3.00]:

Bond Method : 1.23E-001 atm-m3/mole

Group Method: 2.62E-001 atm-m3/mole

Probability of Rapid Biodegradation (BIOWIN v2.62):

Linear Model : 0.8960

Non-Linear Model : 0.9471

Expert Survey Biodegradation Results:

Ultimate Survey Model: 2.6974 (weeks-months)

Primary Survey Model : 3.5354 (days-weeks)

Atmospheric Oxidation (25 deg C) [AopWin v1.85]:

Hydroxyl Radicals Reaction:

OVERALL OH Rate Constant = 40.3011 E-12
cm3/molecule-sec

Half-Life = 0.265 Days (12-hr day; 1.5E6

OH/cm3)

Half-Life = 3.185 Hrs

Ozone Reaction:

No Ozone Reaction Estimation

Soil Adsorption Coefficient (PCKOCWIN v1.62):

Koc : 2.958E+005

Log Koc: 5.471

Aqueous Base/Acid-Catalyzed Hydrolysis (25 deg
C) [HYDROWIN v1.62]:

Rate constants can NOT be estimated for this
structure!

BCF Estimate from Log Kow (BCFWIN v2.0):

Log BCF = 2.894 (BCF = 782.5)

log Kow used: 8.40 (estimated)

Volatilization from Water:

Henry LC: 0.262 atm-m3/mole (estimated by
Group SAR Method)

Half-Life from Model River: 4.721 hours

Half-Life from Model Lake : 153.3 hours (6.389
days)

Removal In Wastewater Treatment:

Total removal: 94.14 percent

Total biodegradation: 0.77 percent

Total sludge adsorption: 92.62 percent

Total to Air: 0.76 percent



[OPPT Home](#) | [EPA Home](#) | [Search](#) | [Comments](#)

Last Revision: 8/26/99

URL: <http://www.epa.gov/opptintr/chemrtk.htm>