

# QSARs for Endocrine Disruption Priority Setting Database 2: The Integrated 4-Phase Model

**John D. Walker<sup>a\*</sup>, Hong Fang<sup>b</sup>, Roger Perkins<sup>b</sup> and Weida Tong<sup>c</sup>**

<sup>a</sup> TSCA Interagency Testing Committee (ITC), U.S. Environmental Protection Agency (M7401), Washington, DC 20460

<sup>b</sup> Northrop Grumman Information Technology, Jefferson, Arkansas 72079

<sup>c</sup> National Center for Toxicological Research (NCTR), Jefferson, Arkansas 72079

## Abstract

Version 2 of the Endocrine Disruption Priority Setting Database (EDPSD2) is a decision support tool developed by the U.S. Environmental Protection Agency. EDPSD2 is organized into 4 categories (exposure-related, effects-

related, combined exposure-and effects-related and specially-targeted priorities) that are supported by several compartments and information sources. The effects-related category includes a Quantitative Structure Activity Relationship compartment that is supported by two models for predicting estrogen receptor binding affinities of chemicals in EDPSD2. This paper describes the categories and compartments of EDPSD2 and the use of the Integrated 4-Phase model to predict the estrogen receptor binding affinities of chemicals in EDPSD2.

\* To receive all correspondence: Phone: (202) 564-7526, Fax: (202) 564-7528, E-mail: walker.johnd@epa.gov

**Key words:** QSARs, receptor binding, estrogen, EDPSD, Integrated 4-Phase

## Abbreviations: EDPSD

– Endocrine Disruption Priority Setting Database

QSARs	– Quantitative Structure Activity Relationships
RBAs	– receptor binding affinities
FQPA	– Food Quality Protection Act
FIFRA	– Federal Insecticide, Fungicide, and Rodenticide Act
U.S. EPA	– U.S. Environmental Protection Agency
EDSTAC	– Endocrine Disruptors Screening and Testing Advisory Committee
MDL	– Molecular Design Limited
ISIS	– Information Systems\9 Integrated Scientific Information System
SMILES	– Simplified Molecular Input Line Entry System
HQSAR	– Hologram QSAR
ER	– estrogen receptor
2D	– 2 dimensional
3D	– 3 dimensional
CAS	– Chemical Abstract Service
CoMFA	– Comparative Molecular Field Analysis
HSDB	– Hazardous Substances Data Bank
RTECS <sup>®</sup>	– Registry of Toxic Effects of Chemical Substances
ATSDR	– Agency for Toxic Substances Disease Registry
EAFUS	– Everything Added to Food in the United States
AQUIRE	– AQUatic toxicity Information RETrieval (AQUIRE) database
TERRETOX	– Terrestrial Toxicity database
OECD	– Organization for Economic Cooperation and Development
SIDS	– Screening Information Data Set
ICCA	– International Council of Chemical Associations
PLS	– partial least squares

## 1 Introduction

The 1996 Food Quality Protection Act (FQPA, 21 U.S.C. §§201–415) requires the U.S. Environmental Protection Agency (U.S. EPA) to implement programs for screening and testing of pesticide chemical active ingredients and “inert” formulation ingredients for their ability to disrupt estrogen functions. In response, the U.S. EPA established the Endocrine Disruptors Screening and Testing Advisory Committee (EDSTAC). The EDSTAC recommended that the U.S. EPA create an Endocrine Disruption Priority Setting Database (EDPSD) as a tool that could be used to assist rapid sorting and priority setting of chemicals for endocrine disruption screening and testing.

EDPSD 1 was developed for and demonstrated to EDSTAC at a December 1997 public meeting [1]. Quantitative Structure Activity Relationships (QSARs) were not used for EDPSD1 [2].

Following the development of EDPSD1, the U.S. EPA developed EDPSD2 (<http://www.epa.gov/scipoly/oscpendo/index.htm>). EDPSD2 is a multi-user client/server application developed using Visual Basic 6.0 for the “front end” screens and user interface, Microsoft Access97 for the “back end” database, and Seagate Crystal Reports 7 for the reports (<http://www.erg.com/endocrine>).

The EDPSD2 QSAR compartment includes two independently-developed QSAR-based models for predicting

estrogen receptor binding affinities (RBAs). The purpose of this paper is to discuss:

- 1) categories and compartments of EDPSD2, and
- 2) results from using the Integrated 4-Phase model to predict the estrogen RBAs of the 623 chemicals from EDPSD2.

## 2 Methods

The Integrated 4-Phase model was developed at the U.S. Food and Drug Administration's National Center for Toxicological Research (NCTR) by Drs. Hong Fang, Weida Tong, Roger Perkins and others [3, 4]. The Integrated 4-Phase model is composed of four sequential phases (Figure 1).

**Phase I: Filtering** – Two rejection filters, molecular weight <94 or >1000 and no-ring structure, were used to significantly and with high confidence eliminate those chemicals extremely unlikely to bind to an estrogen receptor (ER) [3]. These two filters were validated on ~2000 chemicals whose ER activities were available from the literature.

**Phase II: Active/Inactive Assignment** – The chemicals passing through Phase I were assigned as YES/NO for ER binding using three different methods, i.e., structural alerts, pharmacophore searching, and classification models. While structural alerts identify key 2D structural features associated with ER binding, pharmacophore searching identified 3D (3-dimensional) sub-structures important for ER binding. Classification models used pattern recognition to qualitatively categorize chemicals into active and inactive subsets on the basis of their similarity in physicochemical properties. In its current form, this Phase employs 11 models in parallel, 3 structural alerts [3], 7 pharmacophores [4] and 1 classification model [3], which are able to identify both strong and weak estrogens, some of these structures are not

similar to estradiol, such as kepones [3]. These 11 models are complementary and were designed to distinguish active from inactive chemicals. Only chemicals identified as inactive by all 11 models were eliminated from further evaluation in Phase III.

As would be expected from the application of 11 models, a chemical could be predicted to be active by none, a few, or many of these models. Since each model identifies specific structural attributes associated with activity, the number of models predicting that a chemical is active should tend to increase in direct proportion to its activity. To provide semi-quantitative activity data for these 11 models, a consensus-ranking index was developed to provide a score for a chemical by summing the model's concordance that predicts the chemical as active. The activities estimated by the consensus-ranking index correlated very well with the NCTR dataset of measured estrogen RBAs. Thus, this index was used to estimate semi-quantitatively the predicted log RBAs values in Phase II.

**Phase III: Quantitative Predictions** – In this Phase, 3D chemical structure assessments, or identification of bioactive conformers, and molecular alignments were used to develop a Comparative Molecular Field Analysis (CoMFA) model making a more accurate quantitative activity predictions for chemicals that passed Phase II [5]. Both putative bioactive conformers and proper molecular alignment in the model were determined based on the recently published 4 ligands-ER crystal structure. After alignment, the molecules in the training set were placed in a 3D cubic lattice and the steric and electrostatic descriptors at each mesh point were calculated based on the L-J (6–12) and Coulombic potentials, respectively. About 600 descriptors were analyzed using the method of partial least-squares (PLS) regression to derive the CoMFA model. The model had  $r^2 = 0.91$  in fitting and  $q^2 = 0.66$  in a leave-one-out cross validation, indicating self-consistency and high predictivity. Chemicals with higher predicted binding affinity are given higher priority for further evaluation in Phase IV. The CoMFA model demonstrated good statistical reliability using both internal and external validation.

**Phase IV: Rule-Based Decision-Making System** – A knowledge-based system can be used to foster definitive decision making and facilitate priority assessments of chemicals with high predicted estrogen RBAs. This involves a multidisciplinary effort that includes computational chemists, toxicologists, and environmental and regulatory scientists from different agencies. The system is useful only after incorporating accumulated human knowledge and expertise (i.e., rules). The following information should be considered in the final decision making:

1. Information gained at each phase of the integrated computational approach.
2. Information on human exposure, environmental fate and other effects, and chemical production volumes.
3. Chemical structure novelty, that is, when a structure is encountered that is dissimilar to all those that have been

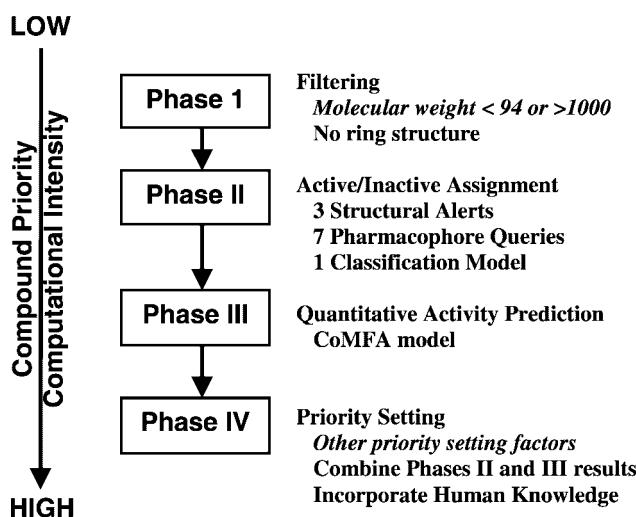


Figure 1. The Integrated 4-Phase Model.

used to train and test the models. For example, care should be taken to interpret prediction for chemicals containing metals, etc.

Phase IV was not used for this paper.

### 3 Results

The results from using the Integrated 4-Phase model to predict the estrogen RBAs of 623 EDPSD2 chemicals are summarized in Tables 1 and 2. One hundred eighty could not be screened for estrogen RBAs. The remaining 443 could be screened; 156 passed Phase I, but 287 did not (Table 1). The 156 that passed Phase I included 78 that passed Phase II and 78 that did not pass Phase II. Of the 78 that passed phase II, 35 passed Phase III and 43 did not (Table 1).

Of the 78 chemicals that passed phase II, 59 had Phase II predicted log RBA ranges  $<-2$  (Table 2). Of these 59, 39 were predicted to be not active in Phase III (Table 2). Four (4) of the 12 Phase II chemicals that had predicted log RBA ranges of  $-1 \sim 0$ , were predicted to be not active in Phase III (Table 2). None of the 7 Phase II chemicals that had predicted log RBA ranges  $>0$ , were predicted to be not active in Phase III (Table 2).

### 4 Discussion

#### 4.1 EDPSD2

The EDPSD2 characterizes the category-compartment-based approach for priority setting that was recommended by the EDSTAC. Compartment refers to a group of chemicals that share a common source, i.e., common exposure, effects, exposure and effects or a special priority.

EDPSD2 is organized into 4 categories: exposure-related, effects-related, combined exposure-and effects-related and specially-targeted priorities (Figure 2). Each category includes several compartments and information sources used to provide data for those compartments. The exposure category is the largest category; it contains 13 compartments and numerous combined and individual information sources (Figure 3). Exposure data are important because chemical

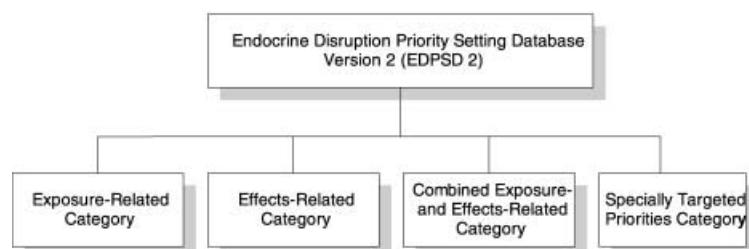
**Table 1.** Number of EDPSD2 chemicals that passed or did not pass Pre-Phase I and Phases I, II and III of the Integrated 4-Phase model

	Number of EDPSD2 chemicals		
	Total	Pass	Not pass
Pre-Phase I	623	443	180
Phase I	443	156	287
Phase II	156	78	78
Phase III	78	35	43

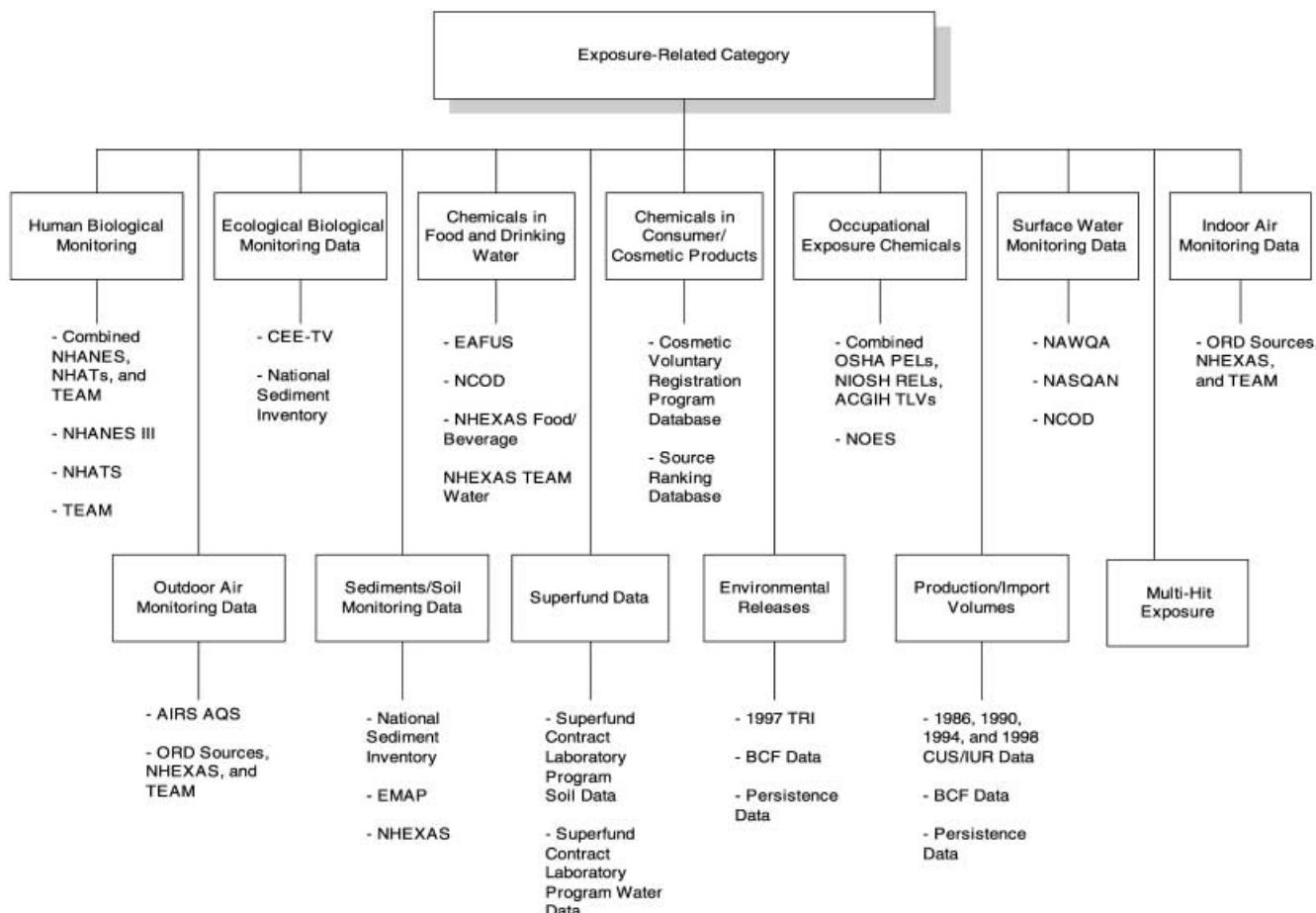
exposures have caused adverse effects in the endocrine systems of humans, domestic animals, birds, fish, and wildlife. The exposure information sources represent a variety of media encountered by humans and other species including food, drinking water, surface water, sediment, soil, etc. Exposure-related information, such as persistence in the environment, bioaccumulation potential, release to the environment, and production volume provide an indirect indication of the likelihood of exposure. The effects-related category contains 7 compartments (Figure 4). For these 7 compartments there are 6 information sources (Hazardous Substances Data Bank (HSDB), Registry of Toxic Effects of Chemical Substances (RTECS®), Agency for Toxic Substances Disease Registry (ATSDR) Toxicological Profiles, Everything Added to Food in the United States (EAFUS), AQUatic toxicity Information REtrieval (AQUIRE) database, Terrestrial Toxicity (TERRETOX) database) and 2 QSARs, the COREPA and the Integrated 4-Phase models (Figure 4). Detailed descriptions and reference sources for HSDB, RTECS®, ATSDR, EAFUS, AQUIRE and TERRETOX were provided for EDPSD1 for which they were also used as data sources [1]. The combined exposure-and effects-related category is weighted toward identifying those chemicals for which there is both exposure effects related information (Figure 5). The specially-targeted priorities category includes mixtures and naturally-occurring non steroidial estrogens (Figure 6). Mixtures include those substances found in products, the environment, and human tissues and fluids, such as contaminants in human breast milk; phytoestrogens in soy-based infant formula; mixtures of chemicals commonly found at hazardous waste sites; pesticide and fertilizer mixtures; disinfection byproducts; and gasoline. Naturally occurring non-steroidal estrogens

**Table 2.** Number of Phase II EDPSD2 chemicals and Phase III EDPSD2 chemicals associated with Phase II and Phase III predicted log RBA ranges

Phase II EDPSD2 chemicals	Phase II predicted log RBA range	Phase III EDPSD2 chemicals	Phase III predicted log RBA range
4	$>1.0$	3	$-2.0 \sim -1.0$
		1	$-3.0 \sim -2.0$
3	$0.0 \sim 1.0$	2	$-3.0 \sim -2.0$
		1	$<-3.0$
12	$-1.0 \sim 0.0$	1	$-2.0 \sim -1.0$
		4	$-3.0 \sim -2.0$
		3	$<-3.0$
		4	Not Active
30	$-3.0 \sim -2.0$	3	$-2.0 \sim -1.0$
		5	$-3.0 \sim -2.0$
		2	$<-3.0$
		20	Not Active
29	$<-3.0$	1	$-2.0 \sim -1.0$
		1	$-3.0 \sim -2.0$
		8	$<-3.0$
		19	Not Active



**Figure 2.** Categories of the Endocrine Disruption Priority Setting Database version 2 (EDPSD 2).



**Figure 3.** Compartments and information sources for the exposure category of the Endocrine Disruption Priority Setting Database version 2 (EDPSD 2).

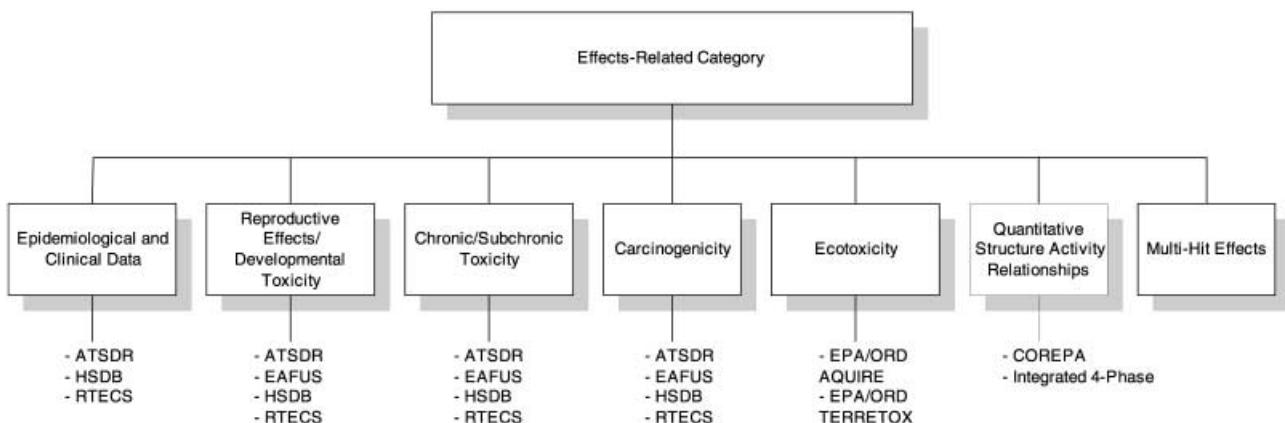
include natural products derived from plants (phytoestrogens) and fungi (mycotoxins).

#### 4.2 Integrated 4-Phase Model Predictions

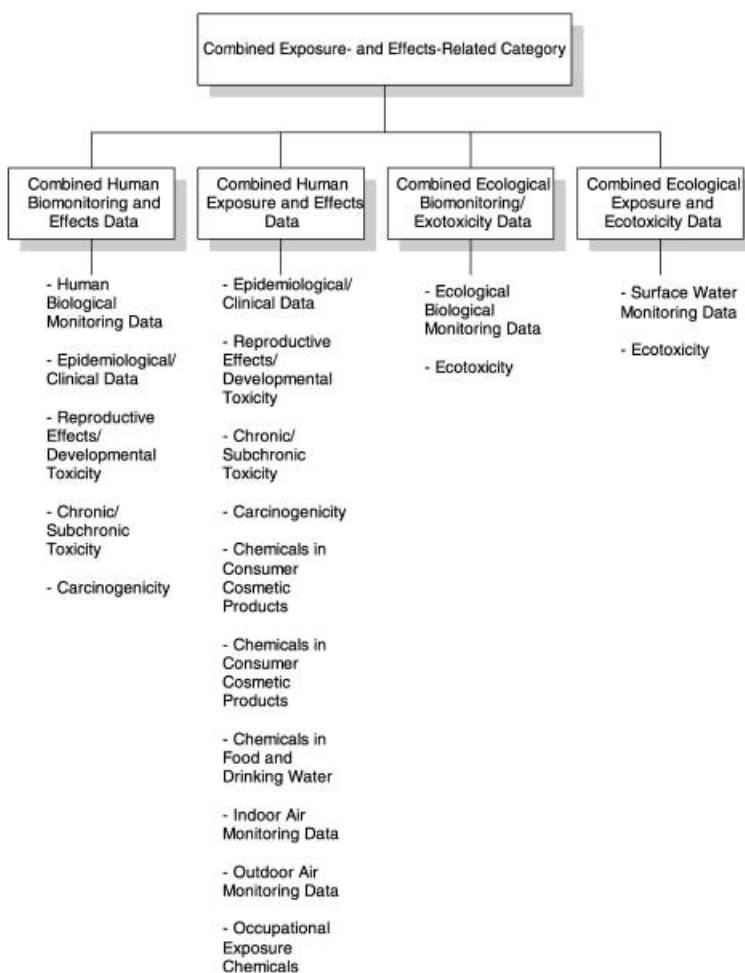
Binding to the estrogen receptor may be one process in a series of events proceeding through translation and transcription that could result in reproductive or other effects. However, because of the promiscuous nature of estrogen receptors [6] there may be some uncertainty when predicting estrogen RBAs.

#### EDPSD2 chemicals that could not be screened for estrogen RBAs

To predict estrogen RBAs, the Integrated 4-Phase QSAR-based model requires that chemicals be assigned SMILES notations [7–9]. There were 180 EDPSD2 chemicals that could not be screened for estrogen RBAs (Table 1). SMILES could not be written for 166 of these EDPSD2 chemicals; the other 14 contained structures or heavy atoms that could not be modeled. Examples of chemicals for which SMILES could not be written included tallow, lard, petroleum resins, turpentine, linseed, corn, soybean, cottonseed, coconut, peanut and castor oils.



**Figure 4.** Compartments and information sources for the effects category of the Endocrine Disruption Priority Setting Database version 2 (EDPSD 2).



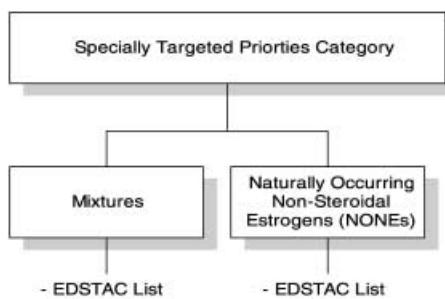
**Figure 5.** Compartments and information sources for the exposure-effects category of the Endocrine Disruption Priority Setting Database version 2 (EDPSD 2).

#### *Phase I EDPSD2 chemicals*

Of the 443 EDPSD2 chemicals that could be screened by the Integrated 4-Phase model, 156 passed Phase I (Table 1). However, 287 did not pass Phase I because they did not contain a ring structure or had a molecular weight < 94 daltons.

#### *Phase II EDPSD2 chemicals*

Of the 156 EDPSD2 chemicals that passed Phase I, 78 passed Phase II (Table 1). Seventy-eight (78) did not pass Phase II because they did not meet any of the requirements for the 3 structural alerts, 7 pharmacophore queries or the



**Figure 6.** Compartments and information sources for the special priorities category of the Endocrine Disruption Priority Setting Database version 2 (EDPSD 2).

classification and regression tree model in Phase II. For the 78 that passed Phase II qualitative estrogen RBAs were predicted (Table 2).

#### Phase III EDPSD2 chemicals

Thirty-five of the 78 EDPSD2 chemicals that passed Phase II also passed Phase III modeling so quantitative estrogen RBAs predictions could be made (Table 2). Forty-three of the 78 EDPSD2 chemicals that passed Phase II were not active in Phase III (Table 2).

#### Comparison of Phase II and Phase III predicted log RBA ranges

For the 29 Phase II EDPSD2 chemicals with predicted log RBAs  $< -3$ , 27 had predicted log RBAs  $< -3$  or lower in Phase III (Table 2). For the 30 Phase II EDPSD2 chemicals with predicted log RBAs  $-3.0 \sim -2.0$ , 27 additional EDPSD2 chemicals had predicted log RBAs  $-3.0 \sim -2.0$  or lower in Phase III. For the remaining 19 Phase II predictions, all of the Phase III predictions were lower than the Phase II predictions. These results suggested that the Phase III predictions were more conservative than the Phase II predictions. These comparisons became more conservative as the Phase II predicted log RBAs increased from  $-1.0 \sim 0.0$  to  $> 1.0$  (Table 2).

#### Comparing predicted estrogen RBAs with experimental data

To compare the predicted estrogen RBAs with experimental data, 3 data sources were reviewed:

- 1) rat estrogen RBA data from Blair et al. [10],
- 2) E-Screen data from Soto et al. [11, 12], and
- 3) yeast two-hybrid assay data from Nishihara et al. [13].

**Table 3.** Comparison of Phase III EDPSD2 chemicals predicted log RBA ranges with experimental data

Phase III EDPSD2 chemicals	Phase III predicted log RBA range	Blair et al. [10] data	Soto et al. [11, 12] data	Nishihara et al. [13] data
8	$-2.0 \sim -1.0$	3/ $-1.5, -1.8, -2.3$	2/ $-1.5, -3.5$	2/ $-0.8, \text{NB}$
13	$-3.0 \sim -2.0$	2/ $-2.1, -3.6$	1/ $-3.5$	3/ $-2.0, -3.0, \text{NB}$
14	$< -3.0$	2/ $-3.4, \text{NB}$	1/ $\text{NB}$	1/ $-4.1$
43	Not Active	6/ $-4.5, 5\text{NB}$	3/ $-3.5, -4.2, \text{NB}$	10/ $-4.0, -4.2, 8\text{NB}$

The experimental rat estrogen RBA data developed by Blair et al. [10] were developed using the rat uterine cytosol estrogen receptor competitive binding assay. The E-Screen assay was developed to assess the estrogenicity of environmental chemicals using the proliferative effects of estrogens on MCF-7 cells as an endpoint [11, 12]. The Nishihara et al. yeast two-hybrid assay measures ER binding-dependent transcriptional and translational activities [13].

As expected, the experimental rat estrogen RBA data developed by Blair et al. [10] compared most favorably with the Phase III predicted log RBA ranges because these data were used to develop the Integrated 4-phase model (Table 3). Comparisons of experimental rat estrogen RBA data correlated with the E-Screen MCF-7 cell proliferation assay data [11, 12] and the yeast two-hybrid assay data [13], especially for the lower Phase III predicted log RBA ranges (Table 3). These findings suggest that estrogen receptor binding may be related to different levels of biological complexity, e.g., cell proliferation (E-Screen assay) and ER binding-dependent transcriptional and translational activities (yeast two-hybrid assay). Additional comparisons of these assays were described previously [14].

## 5 Conclusions

Including the QSAR-based Integrated 4-Phase model in EDPSD2 provided opportunities to predict estrogen RBAs. Estrogen RBA predictions from the Integrated 4-Phase model indicated that 50% of the EDPSD2 chemicals that passed Phase I, passed Phase II, but <50% of those passed Phase III. Of the 35 EDPSD2 chemicals that passed Phase III, 17 had predicted log RBAs  $< -3$ , suggesting their estrogen receptor binding affinity potential was very weak at best. The remaining 18 EDPSD2 chemicals that passed Phase III had predicted log RBAs ranging from  $-3$  to  $-1$ , suggesting that estrogen receptor binding affinity potential was only slightly stronger.

Since SMILES, mole files or some other method for describing chemical structures are needed to develop QSARs, some other non-QSAR modeling approach will have to be used to screen the estrogen RBAs of the 166 EDPSD2 chemicals for which SMILES could not be written and the 14 EDPSD2 chemicals that could not be modeled. To provide other investigators with opportunities to measure or predict the RBAs of the 623 EDPSD2 chemicals the

**Table 4.** Chemical Abstract Service (CAS) number, chemical name, applicable Simplified Molecular Input Line Entry System (SMILES) and molecular weight (MW) of EDPSD2 chemicals for which estrogen receptor binding affinities were predicted using the Integrated 4-Phase model

CAS No.	Chemical Name	SMILES	MW
000050-00-0	Formaldehyde	O=C	30.03
000050-21-5	Lactic acid	O=C(O)C(O)C	90.08
000050-70-4	Glucitol, D-	OCC(O)C(O)C(O)C(O)CO	182.17
000050-81-7	L-Ascorbic acid	O1C(=O)C(O)=C(O)C1C(O)CO	176.13
000050-99-7	D-Glucose	OCC1C(O)C(O)C(O)C(O)O1	180.16
000054-21-7	Sodium salicylate	Oc1ccccc1C(=O)O[Na]	160.11
000056-81-5	Glycerol	OCC(O)CO	92.1
000057-10-3	Palmitic acid	O=C(O)CCCCCCCCCCCOCCC	256.43
000057-11-4	Stearic acid	O=C(O)CCCCCCCCCCCOCCC	284.49
000057-13-6	Urea	O=C(N)N	60.06
000057-50-1	Sucrose	O(C(C(O)C(O)C1O)CO)C1OC(OC(C2O)CO)(C2O)CO	342.3
000057-55-6	1,2-Propanediol	OCC(O)C	76.1
000059-67-6	Nicotinic acid	O=C(O)c(ccn1)c1	123.11
000060-00-4	Acetic acid, (ethylenedinitrilo)tetra-	O=C(O)CN(CCN(CC(=O)O)CC(=O)O)CC(=O)	292.25
000060-12-8	Phenethyl alcohol	OCCc(cccc1)c1	122.17
000060-33-3	Linoleic acid	O=C(O)CCCCCCCC=CCC=CCCCCC	280.45
000062-54-4	Calcium acetate	[Ca](OC(=O)C)OC(=O)C	158.17
000064-02-8	Acetic acid, (ethylenedinitrilo)tetra-, tetrasodium salt	[Na]OC(=O)CN(CCN(CC(O[Na])=O)CC(O[Na])=O)CC(O[Na])=O	380.17
000064-17-5	Ethyl alcohol	OCC	46.07
000064-18-6	Formic acid	O=CO	46.03
000064-19-7	Acetic acid	O=C(O)C	60.05
000065-85-0	Benzoic acid	O=C(O)c(cccc1)c1	122.12
000067-56-1	Methanol	OC	32.04
000067-63-0	Isopropyl alcohol	OC(C)C	60.1
000067-64-1	Methyl ketone	O=C(C)C	58.08
000067-68-5	Methyl sulfoxide	O=S(C)C	78.13
000068-04-2	Citric acid, trisodium salt	[Na]OC(=O)CC(O)(C(=O)O[Na])CC(=O)O[Na]	258.07
000069-65-8	Mannitol, D-	OCC(O)C(O)C(O)C(O)CO	182.17
000069-72-7	Salicylic acid	O=C(O)c(c(O)cccc1)c1	138.12
000071-23-8	Propyl alcohol	OCCC	60.1
000071-36-3	Butyl alcohol	OCCCC	74.12
000071-41-0	1-Pentanol	OCCCCCC	88.15
000071-55-6	Ethane, 1,1,1-trichloro	C(Cl)(Cl)(Cl)C	133.41
000074-84-0	Ethane	CC	30.07
000074-86-2	Ethyne	C#C	26.04
000074-98-6	Propane	C(C)C	44.1
000075-00-3	Ethane, chloro-	C(=C)C	64.52
000075-05-8	Acetonitrile	N#CC	41.05
000075-28-5	Propane, 2-methyl-	C(C)(C)C	58.12
000075-31-0	2-Propanamine	NC(C)C	59.11
000075-37-6	Ethane, 1,1-difluoro-	FC(F)C	66.05
000075-45-6	Methane, chlorodifluoro-	FC(F)Cl	86.47
000075-52-5	Methane, nitro-	O=N(=O)C	61.04
000075-65-0	tert-Butyl alcohol	OC(C)(C)C	74.12
000075-68-3	Ethane, 1-chloro-1,1-difluoro-	FC(F)(Cl)C	100.5
000075-69-4	Methane, trichlorofluoro-	FC(Cl)(Cl)Cl	137.37
000075-71-8	Methane, dichlorodifluoro-	FC(F)(Cl)Cl	120.91
000076-13-1	Ethane, 1,1,2-trichloro-1,2,2-trifluoro-	FC(F)(C(F)(Cl)Cl)Cl	187.38
000077-71-4	Hydantoin, 5,5-dimethyl-	O=C(NC(Cl)=O)(C)C)N	128.13
000077-73-6	4,7-Methanoindene, 3a,4,7a-tetrahydro-	C(C=C(C12)C1)(C2=C3)C3	132.21
000077-86-1	1,3-Propanediol, 2-amino-2-(hydroxymethyl)-	OCC(N)(CO)CO	121.14
000077-90-7	Citric acid, tributyl ester, acetate	O=C(OC(C(=O)O)CCCC)(CC(=O)O)CCCC)CC(=O)OCCCC)C	402.49
000077-92-9	Citric acid	O=C(O)C(O)(CC(=O)O)CC(=O)O	192.13
000077-99-6	1,3-Propanediol, 2-ethyl-2-(hydroxymethyl)-	OCC(CC)(CO)CO	134.18
000078-40-0	Phosphoric acid, triethyl ester	O=P(OCC)(OCC)OCC	182.16
000078-51-3	Ethanol, 2-butoxy-, phosphate (3:1)	O=P(OCCOCCCC)(OCCOCCCC)OCCOCCCC	398.48
000078-59-1	2-Cyclohexen-1-one, 3,5,5-trimethyl-	O=C(C=C(CC1)C)C1	138.21
000078-70-6	1,6-Octadien-3-ol, 3,7-dimethyl-	OC(C=C)CCC=C(C)C	154.25
000078-78-4	Butane, 2-methyl-	C(CC)(C)C	72.15
000078-83-1	1-Propanol, 2-methyl-	OCC(C)C	74.12
000078-92-2	sec-Butyl alcohol	OC(CC)C	74.12
000078-93-3	2-Butanone	O=C(CC)C	72.11

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
000078-96-6	2-Propanol, 1-amino-	OC(CN)C	75.11
000079-09-4	Propionic acid	O=C(O)CC	74.08
000079-11-8	Acetic acid, chloro-	O=C(O)CCl	94.5
000079-14-1	Glycolic acid	O=C(O)CO	76.05
000079-24-3	Ethane, nitro-	O=N(=O)CC	75.07
000079-34-5	Ethane, 1,1,2,2-tetrachloro-	C(C(Cl)Cl)(Cl)Cl	167.85
000079-41-4	Methacrylic acid	O=C(O)C(=C)C	86.09
000080-05-7	Phenol, 4,4'-isopropylidenedi-	Oc(ccc1)C(c(ccc(O)c2)c2)(C)C)c1	228.29
000080-26-2	p-Menth-1-en-8-ol, acetate	O=C(OC(C(CCC(=C1)C)C1)(C)C)C	196.29
000080-56-8	2-Pinene	C(CC1C2)C1(C)C)(=C2)C	136.24
000080-62-6	Methacrylic acid methyl ester	O=C(OC)C(=C)C	100.12
000081-07-2	1,2-Benzisothiazolin-3-one, 1,1-dioxide	O=C(NS(=O)(=O)c1cccc2)c12	183.18
000084-66-2	Phthalic acid, diethyl ester	O=C(OCC)C(c(ccc1)C(=O)OCC)c1	222.24
000084-74-2	Phthalic acid, dibutyl ester	O=C(OC(OC)C)c(eccc1)C(=O)OCCCC)c1	278.35
000085-40-5	4-Cyclohexene-1,2-dicarboximide	O=C(NC(=O)C1CC=CC2)C12	151.17
000085-41-6	Phthalimide	O=C(NC(=O)c1cccc2)c12	147.13
000085-44-9	Phthalic anhydride	O=C(OC(=O)c1cccc2)c12	148.12
000085-68-7	Phthalic acid, benzyl butyl ester	O=C(OCc(ccc1)C)e(ccc2)C(=O)OCCCC)c2	312.37
000088-12-0	2-Pyrrolidinone, 1-ethyl-	O=C(N(C=C)CC1)C1	111.14
000089-65-6	D-erythro-Hex-2-enonic acid, .gamma.-lactone	OCC(O)C1C(O)=C(O)C(=O)O1	176.13
000090-80-2	Gluconic acid, .delta.-lactone, D-	O=C(OC(C(O)C1O)CO)C1O	178.14
000091-20-3	Naphthalene	c(c(cc1)ccc2)(c1)c2	128.18
000091-53-2	Quinoline, 6-ethoxy-1,2-dihydro-2,4-trimethyl-	Oc(ccc(NC(=C1C)(C)C)c12)c2)CC	217.31
000093-83-4	Oleamide, N,N-bis(2-hydroxyethyl)-	O=C(N(CCO)CCO)CCCCCCCC=CCCCCC	369.59
000094-36-0	Benzoyl peroxide	O=C(OOC(=O)c(ccc1)C1)c(ccc2)c2	242.23
000095-14-7	1H-Benzotriazole	c1cc2nnnc2c1	119.13
000095-48-7	o-Cresol	Oc(ccc1)C)c1	108.14
000095-49-8	Toluene, o-chloro-	c(eccc1)Cl)(c1)C	126.59
000096-29-7	2-Butanone, oxime	N(O)=C(CC)C	87.12
000096-48-0	2(3H)-Furanone, dihydro-	O=C(OCC1)C1	86.09
000097-63-2	Methacrylic acid, ethyl ester	O=C(OCC)C(=C)C	114.15
000097-64-3	Lactic acid, ethyl ester	O=C(OCC)C(O)C	118.13
000097-85-8	Isobutyric acid, isobutyl ester	O=C(OCC(C(C)C)C(C)C)	144.22
000097-88-1	Methacrylic acid, butyl ester	O=C(OCCCC)C(=C)C	142.2
000097-99-4	Furfuryl alcohol, tetrahydro-	O(C(CC1)CO)C1	102.13
000098-54-4	Phenol, p-tert-butyl-	Oc(ccc1)C(C)(C)C)c1	150.22
000098-55-5	p-Menth-1-en-8-ol	OC(C(CCC(=C1)C)C1)(C)C	154.25
000098-82-8	Cumene	c(ccc1)(c1)C(C)C	120.2
000098-86-2	Acetophenone	O=C(ecccc1)C	120.15
000098-94-2	Cyclohexylamine, N,N-dimethyl-	N(C(CCCC1)C1)(C)C	127.23
000099-76-3	Benzoic acid, p-hydroxy-, methyl ester	O=C(OC)C(ccc(O)c1)c1	152.15
000100-02-7	Phenol, p-nitro-	O=N(=O)c(ccc(O)c1)c1	139.11
000100-37-8	Ethanol, 2-(diethylamino)-	OCCN(CC)CC	117.19
000100-41-4	Benzene, ethyl-	c(ccc1)(c1)CC	106.17
000100-51-6	Benzyl alcohol	OCc(ccc1)C1	108.14
000100-52-7	Benzaldehyde	O=Cc(ccc1)C1	106.13
000100-97-0	1,3,5,7-Tetraazatricyclo[3.3.1.13,7]decane	N(CN(CN1CN23)C3)(C1)C2	140.19
000101-02-1	Phosphorous acid, triphenyl ester	O(c(ccc1)C1P(Oc(ccc2)c2)Oc(ccc3)c3	310.29
000101-68-8	Isocyanic acid, methylenedi-p-phenylene ester	O=C=N(ccc1)Cc(ccc(N=C=O)c2)c2)c1	250.26
000101-86-0	Cinnamaldehyde, .alpha.-hexyl-	O=CC(=Cc(ccc1)C1)CCCC	216.33
000102-60-3	2-Propanol, 1,1',1'''-(ethylenedinitrilo)tetra-	OC(C)CN(CCN(CC(O)C)CC(O)C)CC(O)C	292.42
000102-71-6	Ethanol, 2,2',2''-nitrilotri-	OCCN(CCO)CCO	149.19
000102-76-1	Acetin, tri-	O=C(OCC(OC(=O)C)COC(=O)C)C	218.21
000103-11-7	Acrylic acid, 2-ethylhexyl ester	O=C(OCC(CCCC)CC)C=C	184.28
000103-23-1	Adipic acid, bis(2-ethylhexyl) ester	O=C(OCC(CCCC)CC)CCCC(=O)OCC(CCCC)CC	370.58
000103-24-2	Azelaic acid, bis(2-ethylhexyl) ester	O=C(OCC(CCCC)CC)CCCCCCCC(=O)OCC(CCCC)CC	412.66
000104-15-4	p-Toluenesulfonic acid	O=S(=O)(O)c(ccc(c1)C)c1	172.2
000104-46-1	Anisole, p-propenyl-	O(c(ccc(c1)C=CC)c1)C	148.21
000104-55-2	Cinnamaldehyde	O=CC=Cc(ccc1)C1	132.16
000104-76-7	1-Hexanol, 2-ethyl-	OCC(CCCC)CC	130.23
000106-22-9	6-Octen-1-ol, 3,7-dimethyl-	OCCC(CCC=C(C)C)C	156.27
000106-24-1	2,6-Octadien-1-ol, 3,7-dimethyl-, (E)-	OCC=C(CCC=C(C)C)C	154.25
000106-43-4	Toluene, p-chloro-	c(cc(c1)Cl)(c1)C	126.59

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
000106-44-5	p-Cresol	Oc(cc(c1C)c1	108.14
000106-88-7	Butane, 1,2-epoxy-	O(C1CC)C1	72.11
000106-97-8	Butane	C(CC)C	58.12
000107-15-3	Ethylenediamine	NCCN	60.1
000107-18-6	2-Propen-1-ol	OCC=C	58.08
000107-21-1	Glycol	OCCO	62.07
000107-22-2	Glyoxal	O=CC=O	58.04
000107-41-5	2,4-Pentanediol, 2-methyl-	OC(CC(O)(C)C)C	118.18
000107-64-2	Dimethyldioctadecylammonium chloride	N(Cl)(C)(C)CCCCCCCCCCCCCCCCCCCCCCCCCCCC	586.52
000107-88-0	1,3-Butanediol	OCCC(O)C	90.12
000107-92-6	Butyric acid	O=C(O)CCC	88.11
000107-98-2	2-Propanol, 1-methoxy-	O(CC(O)C)C	90.12
000108-10-1	2-Pentanone, 4-methyl-	O=C(CC(C)C)C	100.16
000108-21-4	Acetic acid, isopropyl ester	O=C(OC(C)C)C	102.13
000108-24-7	Acetic anhydride	O=C(OC(=O)C)C	102.09
000108-31-6	Maleic anhydride	O=C1OC(=O)C=C1	98.06
000108-32-7	Carbonic acid, cyclic propylene ester	O=C(OCC1C)O1	102.09
000108-39-4	Phenol, 3-methyl-	Oc(ccc1C)c1	108.14
000108-46-3	1,3-Benzenediol	Oc(ccc1O)c1	110.11
000108-63-4	Adipic acid, bis(1-methylheptyl) ester	O=(OC(CCCCCC)C)CCCC(=O)OC(CCCCCC)C	370.58
000108-65-6	2-Propanol, 1-methoxy-, acetate	COCC(OC(=O)C)C	132.16
000108-80-5	1,3,5-Triazine-2,4,6(1H,3H,5H)-trione	n(c(ne(n1O)O)O)c1O	129.08
000108-83-8	4-Heptanone, 2,6-dimethyl-	O=C(CC(C)C)CC(C)C	142.24
000108-88-3	Toluene	c(ccc1c1C)	92.14
000108-93-0	Cyclohexanol	OC(CCCCCC)C1	100.16
000108-94-1	Cyclohexanone	O=C(CCCCCC)C1	98.15
000108-95-2	Phenol	Oc(ccc1c1)	94.11
000109-52-4	Pentanoic acid	O=C(O)CCCC	102.13
000109-89-7	Diethylamine	N(CC)CC	73.14
000109-99-9	Furan, tetrahydro-	O(CCC1C)C1	72.11
000110-05-4	tert-Butyl peroxide	O(OC(C)(C)C)C(C)(C)C	146.23
000110-12-3	2-Hexanone, 5-methyl-	O=C(CCC(C)C)C	114.19
000110-16-7	Maleic acid	O=C(O)C=CC(=O)O	116.07
000110-17-8	Fumaric acid	O=C(O)C=CC(=O)O	116.07
000110-19-0	Acetic acid, isobutyl ester	O=C(OCC(C)C)C	116.16
000110-27-0	Myristic acid, isopropyl ester	O=C(OC(C)C)CCCCCCCCCCCC	270.46
000110-30-5	Octadecanamide, N,N'-ethylenebis-	O=C(NCCNC(=O)CCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCC	593.04
000110-43-0	2-Heptanone	O=C(CCCCCC)C	114.19
000110-44-1	Sorbit acid	O=C(O)=CC=CC	112.13
000110-54-3	Hexane	C(CCCC)C	86.18
000110-63-4	1,4-Butanediol	OCCCCO	90.12
000110-64-5	2-Butene-1,4-diol	OCC=CCO	88.11
000110-80-5	Ethanol, 2-ethoxy-	O(CCO)CC	90.12
000110-82-7	Cyclohexane	C(CCCCCC)C1	84.16
000110-91-8	Morpholine	O(CCNC1C)1	87.12
000110-97-4	2-Propanol, 1,1'-iminodi-	OC(C)CNCC(O)C	133.19
000111-27-3	Hexyl alcohol	OCCCCCC	102.18
000111-30-8	Pentanodial	O=CCCCCC=O	100.12
000111-40-0	Diethylenetriamine	N(CC)CCN	103.17
000111-41-1	Ethanol, 2-[(2-aminoethyl)amino]-	OCCNN	104.15
000111-42-2	Ethanol, 2,2'-iminodi-	OCCNCO	105.14
000111-46-6	Ethanol, 2,2'-oxybis-	O(CC)CCO	106.12
000111-60-4	Stearic acid, 2-hydroxyethyl ester	O=(OCOO)CCCCCCCCCCCCCCCC	328.54
000111-70-6	Heptyl alcohol	OCCCCCC	116.21
000111-76-2	Ethanol, 2-butoxy-	O(CCCC)CCO	118.18
000111-77-3	Ethanol, 2-(2-methoxyethoxy)-	O(CCOC)CCO	120.15
000111-82-0	Lauric acid, methyl ester	O=C(OC)CCCCCCCCCCCC	214.35
000111-87-5	Octyl alcohol	OCCCCCC	130.23
000111-90-0	Ethanol, 2-(2-ethoxyethoxy)-	O(CCOCC)CCO	134.18
000112-02-7	Ammonium, hexadecyltrimethyl-, chloride	CCCCCCCCCCCCCCCCN(Cl)(C)(C)C	320.01
000112-05-0	Nonanoic acid	O=C(O)CCCCCCCC	158.24
000112-07-2	Ethanol, 2-butoxy-, acetate	O=C(OCCOCCCC)C	160.21
000112-24-3	Triethylenetetramine	N(CCNCN)CCN	146.24

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
000112-27-6	Triethylene glycol	O(CCOCCO)CCO	150.18
000112-30-1	1-Decanol	OCCCCCCCCCC	158.29
000112-34-5	Ethanol, 2-(2-butoxyethoxy)-	O(CCOC(O)CCCC)	162.23
000112-39-0	Palmitic acid, methyl ester	O=C(OC)CCCCCCCCCCCC	270.46
000112-53-8	Dodecyl alcohol	OCCCCCC	186.34
000112-55-0	Dodecyl mercaptan	SCCCCCCCC	202.4
000112-57-2	Tetraethylenepentamine	N(CCNCNCCN)CCN	189.31
000112-62-9	Oleic acid, methyl ester	O=C(OC)CCCCCCC=CCCCCCC	296.5
000112-69-6	Hexadecylamine, N,N-dimethyl-	N(CCCCCC)C(C)	269.52
000112-70-9	1-Tridecanol	OCCCCCC	200.37
000112-80-1	Oleic acid	O=C(O)CCCCCCC=CCCCCCC	282.47
000112-90-3	9-Octadecenylamine, (Z)-	NCCCCCCCC=CCCCCCC	267.5
000112-92-5	1-Octadecanol	OCCCCCC	270.5
000115-10-6	Dimethyl ether	O(C)C	46.07
000115-77-5	Pentaerythritol	OCC(CO)(CO)CO	136.15
000115-83-3	Stearic acid, neopentanetetrayl ester	O=C(OCC(COC(=O)O)CCCCCCCCCCCC)(COC(=O)CCCCCCCC)	1202
000115-95-7	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate	O=C(O(C(=C)(CCC=C(C)C)C)C	196.29
000117-84-0	Phthalic acid, diethyl ester	O=C(OC(=O)CCCC)C(c(cc1C(=O)OCCCC)C)c1	390.57
000118-58-1	Salicylic acid, benzyl ester	O=C(OCc(ccc1c1)c(c(O)ccc2)c2	228.25
000119-36-8	Salicylic acid, methyl ester	O=C(OCc(c1O)cc1)c1	152.15
000119-47-1	p-Cresol, 2,2'-methylenebis[6-tert-butyl-	Oc(ccc1c1)Cc(c(O)cc2C)C(C)(C)C)c2c1C(C)(C)C	340.51
000119-61-9	Benzophenone	O=C(c(ccc1c1)c(ccc2)c2	182.22
000120-40-1	Dodecanamide, N,N-bis(2-hydroxyethyl)-	O=C(N(CCO)CCO)CCCCCCC	287.45
000120-51-4	Benzoic acid, benzyl ester	O=C(OCc(ccc1c1)c(ccc2)c2	212.25
000121-33-5	Vanillin	O=C(ccc(O)c1OC)c1	152.15
000121-44-8	Tritylamine	N(CC)(CC)CC	101.19
000121-91-5	Isophthalic acid	O=C(O)c(ccc1C(=O)O)c1	166.13
000122-19-0	Ammonium, benzylidemethyloctadecyl-, chloride	CCCCCCCCCCCCCCCCCN(Cl)(C)(C)Cc1cccc1	424.16
000122-20-3	2-Propanol, 1,1',1"-nitrilotri-	OC(C)CN(CC(O)C)CC(O)C	191.27
000122-40-7	Cinnamaldehyde, ,alpha.-pentyl-	O=C(C=C(c(ccc1c1)c(ccc2)c2)C)C	202.3
000123-17-1	4-Nonanol, 2,6,8-trimethyl-	OC(C)C(C(C)C)CC(C)C	186.34
000123-31-9	Hydroquinone	Oc(ccc(O)c1c1)	110.11
000123-38-6	Propionaldehyde	O=CCC	58.08
000123-42-2	2-Pentanone, 4-hydroxy-4-methyl-	O=C(CC(O)(C)C)C	116.16
000123-77-3	Formamide, 1,1'-azobis-	O=C(N=Nc(=O)N)N	116.08
000123-79-5	Adipic acid, diethyl ester	O=C(OC(=O)CCCC)CCCC(=O)OCCCC	370.58
000123-86-4	Acetic acid, butyl ester	O=C(OC(=O)C)C	116.16
000123-95-5	Stearic acid, butyl ester	O=C(OC(=O)CCCC)CCCCCCCC	340.6
000124-04-9	Adipic acid	O=C(O)CCCC(=O)O	146.14
000124-07-2	Octanoic acid	O=C(O)CCCCCCC	144.22
000124-10-7	Myristic acid, methyl ester	O=C(O)CCCCCCCCCCCC	242.41
000124-38-9	Carbon dioxide	O=C=O	44.01
000124-40-3	Dimethylamine	N(C)C	45.08
000124-68-5	1-Propanol, 2-amino-2-methyl-	OCC(N)(C)C	89.14
000126-30-7	1,3-Propanediol, 2,2-dimethyl-	OCC(CO)(C)C	104.15
000126-86-3	5-Decyne-4,7-diol, 2,4,7,9-tetramethyl-	OC(C#CC(O)(CC(C)C)C)(CC(C)C)C	226.36
000126-92-1	Sulfuric acid, mono(2-ethylhexyl) ester, sodium salt	CCCC(C)COS(O[Na])=O	232.27
000127-08-2	Potassium acetate	[K]OC(=O)C	98.14
000127-09-3	Sodium acetate	CC(O[Na])=O	82.03
000127-39-9	Succinic acid, sulfo-, 1,4-diisobutyl ester, sodium salt	CC(C)COC(=O)C(S(=O)(=O)O[Na])CC(=O)OCC(C)C	332.35
000128-37-0	p-Cresol, 2,6-di-tert-butyl-	Oc(ccc1c(C)C(C)(C)c1C(C)C)C	220.36
000128-44-9	1,2-Benzisothiazolin-3-one, 1,1-dioxide, sodium salt	O=C1N([Na])S(=O)(=O)c2cccc12	205.16
000131-11-3	Phthalic acid, dimethyl ester	O=C(OC)c(c(cc1C(=O)OC)c1)	194.19
000131-17-9	Phthalic acid, diallyl ester	O=C(OCC=C)c(c(ccc1c(c(O)cc1OC)c2)c2)	246.27
000131-57-7	Benzophenone, 2-hydroxy-4-methoxy-	[Na]Oc1cccc1c2cccc2	228.25
000132-27-4	2-Biphenylol, sodium salt	Oc(ccc(c1ccc2c)c1)	192.19
000135-19-3	2-Naphthol	[Ca](OC(=O)C(CCCC)CC)OC(=O)C(CCCC)C	144.17
000136-51-6	Hexanoic acid, 2-ethyl-, calcium salt	[Zn](OC(=O)C(CC)CCCC)OC(=O)C(CC)CCCC	312.47
000136-53-8	Hexanoic acid, 2-ethyl-, zinc salt	CCCCCCCC=CCCCCCCC(=O)N(C)CCS(=O)(=O)O[Na]	351.81
000137-20-2	Taurine, N-methyl-N-oleoyl-, sodium salt	CCC(=O)O[Na]	425.61
000137-40-6	Propionic acid, sodium salt	OC(=O)CN(CC(N(CC(O)=O)CC(O[Na])=O)CC(O[Na])=O	96.06
000139-33-3	Acetic acid, (ethylenedinitrilo)tetra-, disodium salt	[Na]OC(=O)CN(CCO)CCN(CC(=O)O[Na])CC(=O)O[Na]	336.21
000139-89-9	Glycine, N-(carboxymethyl)-N'-(2-hydroxyethyl)-N,N'-ethylenedi-, trisodium salt		344.21

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
000139-96-8	Sulfuric acid, monododecyl ester, compd. with 2,2',2"-nitrilotriethanol (1:1)	CCCCCCCCCCCO=C=O.OCCN(OC)CCO	415.66
000140-01-2	Glycine, N,N-bis[2-(bis(carboxymethyl)amino)ethyl]-, pentasodium salt	[Na]OC(CN(CCNC(CN(CC(O[Na])=O)CC(O[Na])=O)CC(O[Na])=O)CC(O[Na])=O	503.26
000140-11-4	Acetic acid, benzyl ester	O=C(OCc(ccc1)c1)C	150.18
000140-66-9	Phenol, p-(1,1,3,3-tetramethylbutyl)-	Oc(ccc1)C(CC(C)(C)C)(C)C)c1	206.33
000141-32-2	Acrylic acid butyl ester	O=C(OCCC)C=C	128.17
000141-43-5	Ethanol, 2-amino-	OCCN	61.08
000141-78-6	Acetic acid ethyl ester	O=C(OCC)C	88.11
000141-79-7	3-Penten-2-one, 4-methyl-	O=C(C=C(C)C)C	98.15
000141-97-9	Acetoacetic acid, ethyl ester	O=C(OCC)CC(=O)C	130.14
000142-47-2	Glutamic acid, monosodium salt, L-	NC(CCC(O[Na])=O)C(O)=O	169.11
000142-78-9	Dodecanamide, N-(2-hydroxyethyl)-	O=C(NCCO)CCCCCCCCCCC	243.39
000142-82-5	Heptane	C(CCCCC)C	100.21
000142-91-6	Palmitic acid, isopropyl ester	O=C(OC(C))CCCCCCCCCCCC	298.51
000143-07-7	Lauric acid	O=C(O)CCCCCCCC	200.32
000143-18-0	Oleic acid, potassium salt	[K]OC(=O)CCCCCCCC=CCCCCCCC	320.56
000143-28-2	9-Octadecen-1-ol, (Z)-	OCCCCCC=CCCCCC	268.49
000144-62-7	Oxalic acid	O=C(O)C(=O)O	90.04
000147-14-8	Copper, [phthalocyaninato(2-)]-	[Cu]6N2C3=Nc8nc(c0c8cccc0)=Ne9c5cccc5c(n69)N=c7nc(c	578.1
000149-30-4	2(3H)-Benzothiazolethione	N(c(c(S1)ccc2)c2)=C1S	167.24
000149-44-0	Sodium formaldehydesulfoxylate	OCS(O[Na])=O	118.08
000149-57-5	Hexanoic acid, 2-ethyl-	O=C(O)C(CCCC)CC	144.22
000150-76-5	Phenol, p-methoxy-	O(c(ccc(O)c1)c1)C	124.14
000151-21-3	Sulfuric acid monododecyl ester sodium salt	CCCCCCCCCCCCCO(=O)(=O)O[Na]	288.38
000298-07-7	Phosphoric acid, bis(2-ethylhexyl) ester	O=P(OCC(CCCC)CC)(OCC(CCCC)CC)O	322.43
000300-92-5	Aluminum, hydroxybis(stearato)-	O=C(O[Al](OC(=O)CCCCCCCCCCCC)O)CCCCCCCCCCCCCCCC	610.95
000334-48-5	Decanoic acid	CCCCCCCCCCCC(=O)O	172.27
000461-58-5	Guanidine, cyano-	NC(=NC(#N))N	84.08
000504-60-9	1,3-Pentadiene	C(=CC=C)C	68.12
000506-93-4	Guanidine, mononitrate	O=N(=O)ON(H)=C(N)N	121.08
000526-95-4	Gluconic acid, D-	O=C(O)C(O)C(O)C(O)C(O)CO	196.16
000527-07-1	Gluconic acid, monosodium salt, D-	[Na]OC(=O)C(O)C(O)C(O)C(O)CO	218.14
000532-32-1	Sodium benzoate	[Na]OC(=O)c1ccccc1	144.11
000533-74-4	2H-1,3,5-Thiadiazine-2-thione, tetrahydro-3,5-dimethyl-	N(CSC(N1C)=S)(Cl)C	162.27
000541-02-6	Cyclopentasiloxane, decamethyl-	C[Si]1(C)O[Si](C)(C)O[Si](C)(C)O[Si](C)(C)O[Si](C)(C)	370.78
000544-63-8	Myristic acid	O=C(O)CCCCCCCCCCCC	228.38
000552-30-7	1,2,4-Benzenetricarboxylic acid, cyclic 1,2-anhydride	O=C(OC(=O)c1ccc(C(=O)O)e2)c12	192.13
000555-31-7	2-Propanol, aluminum salt	[Al](OC(C)C)(OC(C)C)OC(C)C	204.25
000556-67-2	Cyclotetrasiloxane, octamethyl-	C[Si]1(C)O[Si](C)(C)O[Si](C)(C)O[Si](C)(C)O1	296.62
000557-04-0	Magnesium stearate	[Mg](OC(=O)CCCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCC	591.26
000557-05-1	Stearic acid, zinc salt	[Zn](OC(=O)CCCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCC	632.35
000577-11-7	Succinic acid, sulfo-, 1,4-bis(2-ethylhexyl) ester, sodium salt	CCCC(CC)OC(=O)CC(C(=O)OCC(CC)CCCC)S(=O)(=O)O	422.58
000586-62-9	o-Menth-1,4(8)-diene	C(=C(C)C)(CCC=C1)C1	136.24
000593-81-7	Trimethylamine, hydrochloride	CN(H)(Cl)(C)C	95.57
000614-45-9	Peroxybenzoic acid, tert-butyl ester	O=C(OOC(C)(C)c(ccc1)c1	194.23
000627-83-8	Stearic acid, ethylene ester	O=C(OCCOC(=O)CCCCCCCCCCCCCCCC)CCCCCCCCCCCCCCCC	595.01
000628-63-7	Acetic acid, pentyl ester	O=C(OCCCC)C	130.19
000631-61-8	Acetic acid, ammonium salt	N(H)(H)(H)(H)OC(=O)C	77.08
000637-12-7	Aluminum stearate	[Al](OC(=O)CCCCCCCCCCCCCCCC)(OC(=O)CCCCCCCCCCCCCCCC	877.42
000763-69-9	Propionic acid, 3-ethoxy-, ethyl ester	O=C(OCC)CCOCC	146.19
000822-16-2	Stearic acid, sodium salt	[Na]OC(=O)CCCCCCCCCCCCCCCC	306.47
000872-50-4	2-Pyrrolidinone, 1-methyl-	O=C1CCCN1C	99.13
001002-89-7	Stearic acid, ammonium salt	N(H)(H)(H)(H)OC(=O)CCCCCCCCCCCCCCCC	301.52
001047-16-1	Quinol[2,3-b]acridine-7,14-dione, 5,12-dihydro-	Oc(e(c(n1cc(e(nc(c2ccc3)c3)c4)c2O)ccc5)c5)c14	312.33
001111-78-0	Ammonium carbamate	NC(=O)ON(H)(H)H	78.07
001241-94-7	Phosphoric acid, 2-ethylhexyl diphenyl ester	O=P(Oc1cccc1)(Oc2cccc2)OCC(CC)CCCC	362.41
001300-72-7	Xylenesulfonic acid, sodium salt	[Na]OS(=O)(=O)c1ccc(cc1)C	208.21
001319-77-3	Cresol	Oc1cccc1C	108.14
001321-94-4	Naphthalene, methyl-	Cc1cccc2cccc12	142.2
001328-53-6	C.I. Pigment Green 7	[Cu]763[N+]2=C8c1e(c(c(c1Cl)Cl)Cl)C2=Ne4n3(c5c4c(c(c(c5Cl)Cl)Cl)Cl)N=C%11[N+]6=C(N=C9N7C(=N8)c%10c9(c(c(c%10Cl)Cl)Cl)Cl)c%12c%11c(c(c(c%12Cl)Cl)Cl)Cl	106.17
001330-20-7	Xylene	Cc1cccc1C	274.87
001333-86-4	Carbon black	C(=O)(O[Cu])CCCCCCC1CC(CC1)C	
001338-02-9	Naphthenic acids, copper salts		

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
001338-24-5	Naphthenic acids	O=C(OCC(C1OCC(C1O)O)O)CCCCCCCCCCC	346.47
001338-39-2	Sorbitan, monolaurate	O(CC1O)C(C1O)C(O)COC(=O)CCCCCCCCCC	430.63
001338-41-6	Sorbitan, monostearate	O(CC1O)C(C1O)C(O)COC(=O)CCCCCCC=CCCCCCC	428.61
001338-43-8	Sorbitan, monooleate	O=C(Cl(C=C(O)C(O)=C(O)C=I))OC4(C(O)C(OC(=O)C2(=CC(O)CCC(=O)[O]C(=O)CC	636.48
001401-55-4	Tannins	O(CC(O)C)CCC	205.08
001560-69-6	Propionic acid, cobalt(2+) salt	CCCCCCCCCCCCCCCC(=O)[O][Ca]OC(=O)CCCCCCCCCCCC	118.18
001569-01-3	2-Propanol, 1-propoxy-	O=N(CCCCCCCCCCCC)C(C)C	607.04
001592-23-0	Stearic acid, calcium salt	O(C1COc(cce2)C(cccc(OCC(O3)C3)c4)(C)C)c2)C1	229.41
001643-20-5	Dodecylamine, N,N-dimethyl-, N-oxide	Oc(cce1)CCCCCCCC	340.42
001675-54-3	Propane, 2,2-bis[p-(2,3-epoxypropoxy)phenyl]-	O=C(ecccc1e)(e(O)cc(OCCCCCCCC)c2)c2	206.33
001806-26-4	Phenol, p-octyl-	[Na]Oc2c(N=Nc1cccc1)(O[Na])=O)c(nn2c3ccc(cc3)	326.44
001843-05-6	Benzophenone, 2-hydroxy-4-(octyloxy)-	Oc(ccc1)C(C)C	556.34
001934-21-0	C.I. Yellow 23	Oc(ccc1)C(C)C	166.22
001948-33-0	Hydroquinone, tert-butyl-	[Na]OC(=O)CCCCCCC	166.2
001984-06-1	Octanoic acid, sodium salt	O=C(OCCCCCCC)e(c(O)ccc1)c1	208.26
002050-08-0	Salicylic acid, pentyl ester	O=C(OCCCCCCCCCCCCC)CCc(cc(c(O)c1C(C)(C)C)C(C)C)C(S(=O)(=O)(O)O)CCCCCCCCCCCC	530.88
002082-79-3	Hydrocinnamic acid, 3,5-di-tert-butyl-4-hydroxy-, octadecyl ester	S(=O)(=O)(C)OCCCCCCCCCCCC	266.4
002235-54-3	Sulfuric acid, monododecyl ester, ammonium salt	Oc1ccc(C)c1n2nc3cccc3n2	225.25
002440-22-4	p-Cresol, 2-(2H-benzotriazol-2-yl)-	c12N=C(S[Na])Sc1cccc2	189.23
002492-26-2	(2H)-Benzothiazolethione, sodium salt	CO[Si](CCCC)(OC)OC	198.72
002530-87-2	Silane, (3-chloropropyl)trimethoxy-	N(H)(CCO)(CCO)(CCO)OC(=O)CCCCCCC=CCCCCCCC	431.66
002717-15-9	Oleic acid, compd. with 2,2',2"-nitrilotriethanol (1:1)	Oc2ccc1cc(cce2N=Nc3ccc(cc3)S(O[Na])=O)=OS(O[Na])	452.37
002783-94-0	Sunset Yellow FCF	O=c(Nc(c(OCC)cce1)c1)c(O)c(N=Nc(ccc(C(=O)N)c2)c2)c3	454.49
002786-76-7	2-Naphtho-o-phenetidine, 4-[(p-carbamoylphenyl)azo]-3-hydroxy-	O(CCO)CCC	104.15
002807-30-9	Ethanol, 2-propoxy-	O=P(O)(O)C(O)(P(=O)(O)O)C	206.03
002809-21-4	Phosphonic acid, (1-hydroxyethylidene)di-	O=C(OCCCCCCC)C=CC(=O)OCCCCCCC	340.51
002915-53-9	Maleic acid, dioctyl ester	Oc(c(n(nc1cccc2)nc1)cc(c3)C(CC(C)(C)C)C(C)c3	323.44
003147-75-9	Phenol, 2-(2H-benzotriazol-2-yl)-4-(1,1,3,3-tetramethylbutyl)-	[K]OC(=O)C(CCCC)CC	182.31
003164-85-0	Hexanoic acid, 2-ethyl-, potassium salt	O=C(OCC(CCCC)CC)e(cc(c1C(=O)O)OCC(CCCC)CC)C(=O)OCC(CC	546.79
003319-31-1	1,2,4-Benzenetricarboxylic acid, tris(2-ethylhexyl) ester	O=N(=O)e(c(N=N(e(c(ccc1cc2)c1)c2)c3)N(=O)=O)c3	338.28
003468-63-1	C.I. Pigment Orange 5	O=C(OCCCCCCCC)CCCCCCC=CCCCCCCC	422.74
003687-46-5	Oleic acid, decyl ester	ON(CC)CC	89.14
003710-84-7	Ethanamine, N-ethyl-N-hydroxy-	[Na]OS(=O)(=O)c2cccc2C(c3ccc(cc3)N(Cc4cccc(c4)S(=O)	794.87
003844-45-9	C.I. Acid Blue 9, disodium salt	[Na]Oc(=O)CCl	116.48
003926-62-3	Acetic acid, chloro-, sodium salt	[Ca]([OC(=O)CC)OC(=O)CC	186.22
004075-81-4	Propionic acid, calcium salt	CIC=CCN(C1)CN(C1C2)(CN1C3)CN32	251.16
004080-31-3	3,5,7-Triaza-1-azoniaadamantane, 1-(3-chloroallyl)-, chloride	[Na]Oc(=O)CNC	111.08
004316-73-8	Sarcosine, monosodium salt	[Na]OS(=O)(=O)c2c(cc(c2)Nc3nc(nc(n3)N(CC)CC)CC)Nc4cc	960.95
004404-43-7	C.I. Fluorescent Brightening Agent 28	[Na]Oc(=O)CN(CC(O[Na])=O)CC(O[Na])=O	257.09
005064-31-3	Acetic acid, nitrilotri-, trisodium salt	O(CC(O)C)CCCC	132.2
005131-66-8	2-Propanol, 1-butoxy-	c12c(N=Nc3c(S(=O)(=O)O)cc(Cc3)c(O)c(C(=O)O)cc1cccc	386.38
005281-04-9	C.I. Pigment Red 57, calcium salt (1:1)	CCCCCCCCS(O[Na])=O	216.27
005324-84-5	1-Octanesulfonic acid, sodium salt	O=C=C(CCC=C(C)C)C	152.24
005392-40-5	2,6-Octadienal, 3,7-dimethyl-	O=C(OCC(CCCC)CC)C=C(cccc(OC)c1)c1	290.41
005466-77-3	2-Propenoic acid, 3-(4-methoxyphenyl)-, 2-ethylhexyl ester	C(=CCC(C(=C)C)C1)C	136.24
005989-27-5	p-Mentha-1,8-diene, (R)-(+)-	O=C(NC(NC(=O)N)C(C)C)N	174.2
006104-30-9	Urea, 1,1'-isobutylidenedi-	O=C(Ne(c(O)cc1c1)C(N=Ne(c(OC)cc(N(=O)=O)c2)c2)C(=O)C1O)C(=O)C1O	386.37
006358-31-2	C.I. Pigment Yellow 74	OCC(O)C1OC(=O)C(=C1O)O[Na]	198.11
006381-77-7	D-erythro-Hex-2-enonic acid, gamma-lactone, monosodium salt	O=P(O)(O)CN(CP(=O)(O)O)CP(=O)(O)O	299.05
006419-19-8	Phosphonic acid, [nitrilotris(methylene)]tri-	O=C(OCC(C)C(OC(=O)C(C)C)C(C)C)C(C)C	286.42
006846-50-0	Isobutyric acid, 1-isopropyl-2,2-dimethyltrimethylene ester	O=C(O)CC(O)C(=O)O	134.09
006915-15-7	Malic acid	O=C(OC(C)C)CCCC(=O)OC(C)	230.31
006938-94-9	Adipic acid, diisopropyl ester	c12c(N=Nc3cc(Cl)c(C)cc3S(=O)(=O)O[Ca]5)c(O)c(C(=O)O)S(N(=O)Cc6ccccc6)CCCN	458.89
007023-61-2	C.I. Pigment Red 48, calcium salt (1:1)	O=C(C(CC)CCCC)O[Ni]OC(=O)C(CC)CCCC	324.6
007173-62-8	1,3-Propanediamine, N-9-octadecenyl-, (Z)-	008001-20-5	12.01
007440-44-0	Carbon	Tung oil	345.11
007580-31-6	Hexanoic acid, 2-ethyl-, nickel salt	Soybean oil	
008001-22-7		Linseed oil	
008001-26-1		Cottonseed oil	
008001-29-4		Corn oil	
008001-30-7		Coconut oil	
008001-31-8			

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
008001-78-3	Castor oil, hydrogenated		
008001-79-4	Castor oil		
008002-03-7	Peanut oil		
008002-09-3	Oils, pine		
008002-13-9	Rape oil		
008002-26-4	Tall oil		
008002-43-5	Lecithins		
008002-50-4	Fats and Glyceridic oils, menhaden		
008002-74-2	Paraffin waxes and Hydrocarbon waxes		
008006-54-0	Lanolin		
008006-64-2	Turpentine, oil		
008007-43-0	Sorbitan, sesquioleate		
008007-45-2	Tar, coal		
008008-20-6	Kerosine, (petroleum)		
008008-57-9	Oils, orange, sweet		
008009-03-8	Petrolatum		
008013-07-8	Soybean oil, epoxidized		
008016-11-3	Linseed oil, epoxidized		
008016-70-4	Soybean oil, hydrogenated		
008029-43-4	Syrups, hydrolyzed starch		
008030-30-6	Naphtha		
008030-76-0	Lecithins, soya		
008030-78-2	Quaternary ammonium compounds, trimethyltallow alkyl, chlorides		
008032-32-4	Ligroine		
008042-47-5	White mineral oil, (petroleum)		
008050-09-7	Rosin		
008050-15-5	Resin acids and Rosin acids, hydrogenated, Me esters		
008050-31-5	Resin acids and Rosin acids, esters with glycerol		
008052-10-6	Tall-oil rosin		
008052-41-3	Stoddard solvent		
008052-42-4	Asphalt		
008052-48-0	Fatty acids, tallow, sodium salts		
008061-52-7	Lignosulfonic acid, calcium salt	[Ca]OS(=O)(=O)CC(Oc3cc(cc3)CCCS(=O)(=O)O)OC)Cc1cc(OCCOCCOCCCC)C	530.62
009004-77-7	Poly(oxy-1,2-ethanediyl), .alpha.-butyl-.omega.-hydroxy-	[Na]OS(=O)(=O)OC=COC=COCCCCCCCCCC	162.23
009004-82-4	Glycols, polyethylene, mono(hydrogen sulfate), dodecyl ether, sodium salt	C(=CCCCCCCCCC)CCCCCCCCCO	372.46
009004-96-0	Glycols, polyethylene, monooleate	C(=CCCCCCCCCC)CCCCCCCCCC	268.49
009004-98-2	Glycols, polyethylene, mono-9-octadecenyl ether, (Z)-	C(=CCCCCCCCCC)CCCCCCCCCC	314.56
009004-99-3	Glycols, polyethylene, monostearate	c1(ccc(cc1)CCCCCCCCCC)OCCO	264.41
009005-00-9	Poly(oxy-1,2-ethanediyl), .alpha.-octadecyl-.omega.-hydroxy-	[Na]OS(=O)(=O)c1ccc(cc1)C	194.18
009005-07-6	Glycols, polyethylene, dioleate	[Na+]OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CC(O)=O	334.25
009005-67-8	Sorbitan, monostearate, polyoxyethylene derivs.	[Mn](OC(=O)C(CCCC)CC)OC(=O)C(CCCC)CC	341.35
009005-90-7	Turpentine	OC(CCCC(C(=C)C)C)C	156.27
009016-45-9	Poly(oxy-1,2-ethanediyl), .alpha.-(nonylphenyl)-.omega.-hydroxy-	O(CC(OCC(O)C)C)C	148.2
009046-01-9	Poly(oxy-1,2-ethanediyl), .alpha.-tridecyl-.omega.-hydroxy-, phosphate	O(CC(CC(OCC(O)C)C)C)C	206.28
012001-85-3	Naphthenic acids, zinc salts	OC(=O)CN(CC(=O)O)CCN(CC(=O)O)CC(=O)O	292.25
012068-03-0	ar-Toluenesulfonic acid, sodium salt	O=C(OCC(CCCC)CC)c1ccc(N(C)C)c1Cl	277.41
015708-41-5	Ferrate(1-), [(ethylenedinitrilo)tetraacetato]-, sodium	[Zr](OC(=O)C(CCCC)CC)OC(=O)C(CCCC)CC	377.64
015956-58-8	Hexanoic acid, 2-ethyl-, manganese salt	[K]OC(=O)C=CC=CC	150.22
018479-58-8	7-Octen-2-ol, 2,6-dimethyl-	CC(O)COC(C)COC(C)CO	192.26
020324-32-7	2-Propanol, 1-(2-methoxy-1-methylethoxy)-	c1ccc(cc1)C(=C)C	118.18
020324-33-8	2-Propanol, 1-[2-(2-methoxy-1-methylethoxy)-1-methylethoxy]-	COc1ccc(O)c1C(C)C	180.25
020824-56-0	Acetic acid, (ethylenedinitrilo)tetra-, diammmonium salt	Oc1ccc(cc1)CCCCCCCC	220.36
021245-02-3	Benzoinoic acid, p-(dimethylamino)-, 2-ethylhexyl ester	[Na]OS(=O)(=O)c1ccc(cc1)CCCCCCCCCCCC	348.48
022464-99-9	Hexanoic acid, 2-ethyl-, zirconium salt	CC(O)COCC(O)C	134.18
024634-61-5	Sorbic acid, potassium salt	O=C(OCC(C(C)C)C)C(C)C	216.32
024800-44-0	Tripropylene glycol	O=S(=O)(c1ccc(cc1)C)O	186.23
025013-15-4	Styrene, ar-methyl-		
025013-16-5	Phenol, tert-butyl-4-methoxy-		
025154-52-3	Phenol, nonyl-		
025155-30-0	Benzenesulfonic acid, dodecyl-, sodium salt		
025265-71-8	Propanol, oxybis-		
025265-77-4	Propanoic acid, 2-methyl-, monoester with 2,2,4-trimethyl-1,3-pentanediol		
025321-41-9	Benzenesulfonic acid, dimethyl-		

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
025322-68-3	Glycols, polyethylene	OC(COC(CO)C)C	
025322-69-4	Glycols, polypropylene	OCCCCCC(C)C	158.29
025339-17-7	Isodecyl alcohol	O=C(OCC(O)CO)CCCCCCCC=CCCCCC	356.55
025496-72-4	Olein, mono-	OCC(OCC(COC(COC)C)C)C	206.28
025498-49-1	Tripropylene glycol monomethyl ether	O=P(c1cccc1)(OCCCCCCC(C)C)OCCCCCCC(C)C	438.64
025550-98-5	Phosphorous acid, diisodecyl phenyl ester	COc1cc(c(C)c1N=Nc2c(O)ccc3cc(c23)S(=O)[Na])=O)S	496.42
025956-17-6	2-Naphthalenesulfonic acid, 6-hydroxy-5-[2-(methoxy-5-methyl-4-sulfophenyl)azo]-, disodium salt	Oc(cc(c1)C(CC)(C)C(CC)(C)C)c1n(nc(c2ccc3)c3)n2	351.5
025973-55-1	Phenol, 2-(2H-benzotriazol-2-yl)-4,6-di-tert-pentyl-		
026183-52-8	Poly(oxy-1,2-ethanediyl), .alpha.-decy1-.omega.-hydroxy-		
026264-05-1	Benzenesulfonic acid, dodecyl-, compd. with isopropylamine (1:1)		
026264-06-2	Benzenesulfonic acid, dodecyl-, calcium salt	CCCCCCCCCCCCc1ccc(S(=O)(=O)O[Ca]OS(=O)(=O)c2ccc(CCCC	691.06
026266-57-9	Sorbitan, monopalmitate	O=C(OCC(C(C(C(O)CO)O)O)CCCCCCCCCCCC)OC(=O)CCCCCCC=C	420.59
026266-58-0	Sorbitan, trioleate	OCC(C(C(COC(=O)CCCCCCC)OC(=O)CCCCCCC=C	1504.5
026401-27-4	Phosphorous acid, isoctyl diphenyl ester	O=P(Oc1cccc1)Oc2cccc2CCCC(C)C	346.41
026447-10-9	Xylenesulfonic acid, ammonium salt	Cc1cc(S(=O)(=O)ON(H)(H)H)cc(C)c1	203.26
026523-78-4	Phenol, nonyl-, phosphite (3:1)	CCCCCCCCCc1ccc(cc1)OP(Oc2cccc2)CCCCCCCC)Oc3ccc(cc	689.02
026544-23-0	Phosphorous acid, isodecyl diphenyl ester	CC(CCCCCCOP(Oc1cccc1)Oc2cccc2)C	374.46
026635-93-8	Glycols, polyethylene, (9-octadecenylimino)diethylene ether, (Z)-		
026761-40-0	Phthalic acid, diisodecyl ester	CC(C)CCCCCCCCOC(=O)c1cccc1C(=O)OCCCCCCC(C)C	446.68
027138-31-4	Dipropylene glycol, dibenzoate	O=C(c1cccc1)OCC(OC(COC(=O)c2cccc2)C)C	342.39
027176-87-0	Benzenesulfonic acid, dodecyl-	O=S(=O)(c1ccc(cc1)CCCCCCCCCCCC)O[K]	326.5
027177-77-1	Benzenesulfonic acid, dodecyl-, potassium salt	Oc(cc1CCCCCCC)cc1	364.59
027193-28-8	Phenol, octyl-	Oc1cccc(c1)CCCCCCCCCCCC	206.33
027193-86-8	Phenol, dodecyl-	O=C(c1cc(c1C(=O)OCCCCCCC(C)C)C(=O)OCCCCCCC(C)C)OCCCC	262.44
027251-75-8	1,2,4-Benzenetricarboxylic acid, triisooctyl ester	[Co](OC(=O)CC(C)(C)CC(C)(C)C)OC(=O)CC(C)(C)CC(C)(C)	546.79
027253-31-2	Neodecanoic acid, cobalt salt		401.46
027323-41-7	Benzenesulfonic acid, dodecyl-, compd. with 2,2',2"-nitrilotris[ethanol] (1:1)		
028188-24-1	Stearic acid, triester with pentaerythritol	O=C(CCCCCCCCCCCCCCCCC)OCC(COC(=O)CCCCCCCCCCCCCCCC)(	935.56
028348-53-0	ar-Cumenesulfonic acid, sodium salt	[Na]OS(=O)(=O)c1ccc(cc1)C(C)C	222.24
028519-02-0	Benzenesulfonic acid, dodecyloxydi-, disodium salt	c1cc(S(=O)(=O)O[Na])cc(c1)Oc2cc(S(=O)(=O)O[Na])cc(CC	542.62
028553-12-0	Phthalic acid, diisononyl ester	O=C(c1cccc1C(=O)OCCCCCCC(C)C)OCCCCCCC(C)C	418.62
028987-17-9	Phenol, nonyl-, barium salt	[Ba](Oc2cc(ccc2)CCCCCCCC)Oc1cc(c1)CCCCCCCC	576.03
029385-43-1	1H-Benzotriazole, methyl-	Cc1ccc2c1N=NN2	133.15
029761-21-5	Phosphoric acid, isodecyl diphenyl ester	O=P(Oc1cccc1)Oc2cccc2OCCCCCCC(C)C	390.46
029806-73-3	Palmitic acid, 2-ethylhexyl ester	O=C(OCC(CCCC)CC)CCCCCCCCCCCC	368.65
029911-28-2	2-Propanol, 1-(2-butoxy-1-methylethoxy)-	O(C(OCC(O)C)CCCC	190.29
031566-31-1	Octadecanoic acid, monoester with 1,2,3-propanetriol	OCC(O)OC(=O)CCCCCCCCCCCCCCCC	358.57
032687-78-8	Hydrazine, 1,2-bis(3,5-di-tert-butyl-4-hydroxyhydrocinnamoyl)-	O=C(NNC(=O)CC(c(c(O)c1C(C)(C)C(C)(C)C)c1)CCc(cc(COC(C)CO)C)C	552.8
034590-94-8	Dipropylene glycol monomethyl ether		148.2
036445-71-3	Benzenesulfonic acid, decyl(sulfophenoxy)-, disodium salt	OCCCCCC	242.45
036653-82-4	1-Hexadecanol	CCCCCCCCCCCCCCCC=	356.55
037220-82-9	9-Octadecenoic acid (Z)-, ester with 1,2,3-propanetriol	O=C(N(CC=C)CC=C(Cl)Cl)C	208.09
037764-25-3	Acetamide, 2,2-dichloro-N,N-di-2-propenyl-	O=C(O)C(P(=O)(O)O)(CC(=O)O)CCC(=O)O	270.13
037971-36-1	1,2,4-Butanetricarboxylic acid, 2-phosphono-	OP(=O)(OCOc1cccc1)O	218.15
039464-70-5	Poly(oxy-1,2-ethanediyl), .alpha.-phenyl-.omega.-hydroxy-, phosphate		
051274-00-1	C.I. Pigment Yellow 42	P(=O)(Oc1ccc(cc1)CCCCCCCC)(OCC(O)CO)O	374.42
051811-79-1	Poly(oxy-1,2-ethanediyl), .alpha.-(nonylphenyl)-.omega.-hydroxy-, phosphate	O=C(CC(C(C(NC1(C)C(C)C1)CCCCCCCC(=O)OC(CC(NC2(C)C	480.74
052829-07-9	Decanedioic acid, bis(2,2,6,6-tetramethyl-4-piperidinyl) ester	O=C(N(CC(O1)C1(C)C(C)C1)CCCCCCCC(=O)O)C	226.1
052836-31-4	Oxazolidine, 3-(dichloroacetyl)-2,2,5-trimethyl-	O=C(C1CC(CCC1CCCCCCC(=O)O)CCCC)O	354.53
053980-88-4	2-Cyclohexene-1-octanoic acid, 5(or 6)-carboxy-4-hexyl-	O=C(OCC=C1NCCCCC)C	281.09
055406-53-6	Carbamic acid, butyl-, 3-iodo-2-propynyl ester	Oc1c(ccc1)C=C	120.15
061788-44-1	Phenol, styrenated		
061788-47-4	Fatty acids, coco		
061788-59-8	Fatty acids, coco, Me esters	O=C(C=CCCCCCCCCCCC=CC(=O)O)O	310.44
061788-89-4	Fatty acids, C18-unsatd., dimers		
061789-01-3	Fatty acids, tall-oil, epoxidized, 2-ethylhexyl esters		
061789-30-8	Fatty acids, coco, potassium salts		
061789-31-9	Fatty acids, coco, sodium salts		
061789-40-0	1-Propanaminium, 3-amino-N-(carboxymethyl)-N,N-dimethyl-, N-coco acyl derivs., inner salts		
061789-51-3	Naphthenic acids, cobalt salts		
061789-65-9	Resin acids and Rosin acids, aluminum salts		
061789-72-8	Quaternary ammonium compounds, benzyl(hydrogenated tallow alkyl)dimethyl, chlorides		
061789-73-9	Quaternary ammonium compounds, benzylbis(hydrogenated tallow alkyl)methyl, chlorides		

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
061789-77-3	Quaternary ammonium compounds, dicoco alkyldimethyl, chlorides		
061789-81-9	Quaternary ammonium compounds, bis(hydrogenated tallow alkyl)dimethyl, Me sulfates		
061789-86-4	Sulfonic acids, petroleum, calcium salts		
061789-97-7	Tallow		
061789-99-9	Lard		
061790-12-3	Fatty acids, tall-oil		
061790-50-9	Resin acids and Rosin acids, potassium salts		
061790-51-0	Resin acids and Rosin acids, sodium salts		
061790-63-4	Fatty acids, coco, compds. with diethanolamine		
061791-01-3	Fatty acids, tall-oil, diesters with polyethylene glycol		
061791-31-9	Ethanol, 2,2'-iminobis-, N-coco alkyl derivs.		
061791-47-7	Ethanol, 2,2'-iminobis-, N-coco alkyl derivs., N-oxides		
061791-48-8	Fatty acids, tall-oil monoesters with sorbitan		
063231-60-7	Paraffin waxes and Hydrocarbon waxes, microcryst.		
063449-39-8	Paraffin waxes and Hydrocarbon waxes, chlоро		
064147-40-6	Castor oil, dehydrated		
064365-17-9	Resin acids and Rosin acids, hydrogenated, esters with pentaerythritol		
064665-57-2	1H-Benzotriazole, 4(or 5)-methyl-, sodium salt	Cc1cc2N([Na])N=Nc2cc1	155.14
064741-41-9	Naphtha, (petroleum), heavy straight-run		
064741-50-0	Distillates, (petroleum), light paraffinic		
064741-51-1	Distillates, (petroleum), heavy paraffinic		
064741-52-2	Distillates, (petroleum), light naphthenic		
064741-53-3	Distillates, (petroleum), heavy naphthenic		
064741-59-9	Distillates, (petroleum), light catalytic cracked		
064741-65-7	Naphtha, (petroleum), heavy alkylate		
064741-67-9	Residues, (petroleum), catalytic reformer fractionator		
064742-14-9	Distillates, (petroleum), acid-treated light		
064742-16-1	Petroleum resins		
064742-42-3	Hydrocarbon waxes, (petroleum), clay-treated microcryst.		
064742-43-4	Paraffin waxes, (petroleum), clay-treated		
064742-46-7	Distillates, (petroleum), hydrotreated middle		
064742-47-8	Distillates, (petroleum), hydrotreated light		
064742-51-4	Paraffin waxes, (petroleum), hydrotreated		
064742-54-7	Distillates, (petroleum), hydrotreated heavy paraffinic		
064742-55-8	Distillates, (petroleum), hydrotreated light paraffinic		
064742-56-9	Distillates, (petroleum), solvent-dewaxed light paraffinic		
064742-61-6	Slack wax, (petroleum)		
064742-65-0	Distillates, (petroleum), solvent-dewaxed heavy paraffinic		
064742-81-0	Kerosine, (petroleum), hydrodesulfurized		
064742-88-7	Solvent naphtha, (petroleum), medium alph.		
064742-89-8	Solvent naphtha, (petroleum), light alph.		
064742-94-5	Solvent naphtha, (petroleum), heavy arom.		
064742-95-6	Solvent naphtha, (petroleum), light arom.		
064742-96-7	Solvent naphtha, (petroleum), heavy alph.		
064743-02-8	Alkenes, C > 10 .alpha.-	CCCCCCCCCC=C	140.27
064771-72-8	Paraffins, (petroleum), normal C5-20		
065997-06-0	Rosin, hydrogenated		
067701-05-7	Fatty acids, C8-18 and C18-unsatd.	O=C(C=CC=CC=CC=CC=C)O	176.22
067701-09-1	Fatty acids, C8-18 and C18-unsatd., potassium salts		
067701-10-4	Fatty acids, C8-18 and C18-unsatd., sodium salts		
067762-27-0	Alcohols, C16-18	CCCCCCCCCCCCCCCCCO	242.45
067891-79-6	Distillates, (petroleum), heavy arom.		
067891-80-9	Distillates, (petroleum), light arom.		
068081-81-2	Benzenesulfonic acid, mono-C10-16-alkyl derivs., sodium salts		
068122-86-1	Imidazolium compounds, 4,5-dihydro-1-methyl-2-nortallow alkyl-1-(2-tallow amidoethyl), Me sulfates		
068139-30-0	1-Propanaminium, N-(3-aminopropyl)-2-hydroxy-N,N-dimethyl-3-sulfo-, N-coco acyl derivs., hydroxides, inner salts		
068140-00-1	Amides, coco, N-(hydroxyethyl)		
068153-22-0	Paraffin waxes and Hydrocarbon waxes, oxidized		
068154-36-9	Fatty acids, coco, monoesters with sorbitan		
068155-09-9	Amides, coco, N-[3-(dimethylamino)propyl], N-oxides		
068171-29-9	Ethanol, 2,2',2"-nitrilotri-, tris(dihydrogen phosphate) (ester), sodium salt	O=P(OCCN(CCOP(=O)(O)O)CCOP(=O)(O)O[Na])(O)O	411.11
068187-51-9	C.I. Pigment Yellow 119		

**Table 4.** (cont.)

CAS No.	Chemical Name	SMILES	MW
068187-76-8	Castor oil, sulfated, sodium salt		
068187-84-8	Castor oil, oxidized		
068308-53-2	Fatty acids, soya		
068308-74-7	Amides, tall-oil fatty, N,N-di-Me		
068334-00-9	Cottonseed oil, hydrogenated		
068334-30-5	Fuels, diesel		
068391-01-5	Quaternary ammonium compounds, benzyl-C12-18-alkyldimethyl, chlorides		
068409-75-6	Bone meal		
068410-69-5	Poly(oxy-1,2-ethanediyl), .alpha.-[2-[bis(2-aminoethyl)methylammonio]ethyl]-.omega.-hydroxy-, N,N'-di-tallow acyl derivs., Me sulfates (salts)		
068424-85-1	Quaternary ammonium compounds, benzyl-C12-16-alkyldimethyl, chlorides		
068425-47-8	Amides, soya, N,N-bis(hydroxyethyl)		
068439-57-6	Sulfonic acids, C14-16-alkane hydroxy and C14-16-alkene, sodium salts		
068439-70-3	Amines, C12-16-alkyldimethyl	CN(CCCCCCCCCCC)C	241.46
068476-30-2	Fuel oil, no. 2		
068476-31-3	Fuel oil, no. 4		
068476-33-5	Fuel oil, residual		
068476-40-4	Hydrocarbons, C3-4	CCC	44.1
068476-78-8	Molasses		
068476-80-2	Fats and Glyceridic oils, vegetable, deodorizer distillates		
068477-31-6	Distillates, (petroleum), catalytic reformer fractionator residue, low-boiling		
068477-33-8	Gases, (petroleum), C3-4, isobutane-rich		
068514-74-9	Palm oil, hydrogenated		
068515-49-1	1,2-Benzenedicarboxylic acid, di-C9-11-branched alkyl esters, C10-rich	O=C(c1ccccc1C(=O)OCCCCCC(C)(C)C)OCCCCCC(C)(C)C	446.68
068526-82-9	Alkenes, C6-10, hydroformylation products, high-boiling		
068527-23-1	Naphtha, (petroleum), light steam-cracked arom.		
068551-07-5	Alcohols, C8-18	OCCCCCC	200.37
068551-19-9	Alkanes, C12-14-iso-	CC(CCCCCCCCC)C	184.37
068553-00-4	Fuel oil, no. 6		
068584-22-5	Benzenesulfonic acid, C10-16-alkyl derivs.	O=S(=O)(c1ccc(cc1)CCCCCCCC)O	298.44
068584-25-8	Benzenesulfonic acid, C10-16-alkyl derivs., compds. with triethanolamine		
068585-34-2	Poly(oxy-1,2-ethanediyl), .alpha.-sulfo-.omega.-hydroxy-, C10-16-alkyl ethers, sodium salts		
068602-80-2	Distillates, (petroleum), C12-30-arom.		
068603-15-6	Alcohols, C6-12	OCCCCCC	102.18
068603-18-9	Alcohols, C10-16, distn. residues		
068603-42-9	Amides, coco, N,N-bis(hydroxyethyl)		
068608-26-4	Sulfonic acids, petroleum, sodium salts		
068611-44-9	Silane, dichlorodimethyl-, reaction products with silica		
068611-55-2	Sulfuric acid, mono-C10-16-alkyl esters		
068647-72-3	Terpenes and Terpenoids, sweet orange-oil		
068650-39-5	Imidazolium compounds, 1-[2-(carboxymethoxy)ethyl]-1-(carboxymethyl)-4,5-dihydro-2-norcoco alkyl, hydroxides, inner salts, disodium salts		
068783-78-8	Quaternary ammonium compounds, dimethyltallow alkyl, chlorides		
068891-38-3	Poly(oxy-1,2-ethanediyl), .alpha.-sulfo-.omega.-hydroxy-, C12-14-alkyl ethers, sodium salts		
068920-06-9	Hydrocarbons, C7-9		
068937-83-7	Fatty acids, C6-10, Me esters	O=C(OC)CCCCCC	158.24
068938-15-8	Fatty acids, coco, hydrogenated		
068953-36-6	Fatty acids, tall-oil, reaction products with tetraethylenepentamine		
068955-64-6	Hexanedinitrile, hydrogenated, high-boiling fraction, phosphonomethylated		
068990-53-4	Glycerides, C14-22 mono-	O=C(C(C(OC(=O)CCCCCCCC)O)O)O	290.36
069009-90-1	1,1'-Biphenyl, bis(1-methylethyl)-	CC(c1ccc(cc1)c2ccc(cc2)C(C)C)C	238.38
069012-32-4	Slags, phosphorus-manufg.		
070131-50-9	Bentonite, acid-leached		
070592-80-2	Amines, C10-16-alkyldimethyl, N-oxides	CCCCCCCCCN(=O)(C)C	201.36
071011-27-3	Quaternary ammonium compounds, bis(hydrogenated tallow alkyl)dimethyl, salts with Hectorite		
078491-02-8	Urea, N-[1,3-bis(hydroxymethyl)-2,5-dioxo-4-imidazolidinyl]-N,N'-bis(hydroxymethyl)-	O=C(NCO)N(C1N(C(=O)N(C1(=O))CO)CO)CO	278.22
088230-35-7	Hexanol, acetate, branched and linear	CCCCCCOC(=O)C	144.22
090438-79-2	Acetic acid, C6-8-branched alkyl esters	O=C(OCCCCCC)C	158.24
097676-23-8	Oils, licorice	c1ccc2N(C(=O)C(Cl)Cl)C(C)COc2c1	260.12
098730-04-2	2H-1,4-Benzoxazine, 4-(dichloroacetyl)-3,4-dihydro-3-methyl-		
119345-04-9	Benzene, 1,1'-oxybis-, tetrapropylene derivs., sulfonated, sodium salts		

Chemical Abstract Service (CAS) number, name, SMILES and molecular weights (if available) for these chemicals are listed in Table 4.

## Disclaimer

Reference herein to any specific commercial product, process, or service by trademark, manufacturer, or otherwise, does not necessarily constitute its endorsement, recommendation or favoring by the Toxic Substances Control Act (TSCA) Interagency Testing Committee (ITC) or any of the 16 U.S. Government organizations represented on the ITC, including the U.S. Environmental Protection Agency (U.S. EPA). Views expressed in this paper do not necessarily reflect policies of the ITC or any of the U.S. Government organizations represented on the ITC, including the U.S. EPA.

## Acknowledgments

The authors thank Norma Williams for formatting and proofing the manuscript and Gary Timm from the U.S. EPA's Office of Science Coordination and Policy for reviewing the manuscript.

## References

- [1] Walker, J. D., Waller, C., and Kane, S., The Endocrine Disruption Priority Setting Database (EDPSD): A tool to rapidly sort and prioritize chemicals for endocrine disruption screening and testing, Chapter 4, in: Walker J. D. (Ed.), *QSARs for Predicting Biological Activities Related to Endocrine Disruption and Environmental-Human Health Interactions*, SETAC Press, Pensacola (Florida), 2003.
- [2] Schmieder, P., Ankley, G., Mekenyany, O., Walker, J. D., and Bradbury, S., QSARs for predicting estrogen receptor binding affinity of structurally diverse chemicals, *Environ. Toxicol. Chem.* 22, In Press (2003).
- [3] Shi, L., Tong, W., Fang, H., Perkins, R., Wu, J., Tu, M., Blair, R. M., Branham, W. S., Waller, C., Walker, J. D., and Sheehan, D. M., An Integrated "4-Phase" Approach for Setting Endocrine Disruption Screening Priorities – Phase I and II Predictions of Estrogen Receptor Binding Affinities, *SAR QSAR Environ. Res.* 13, 69–88 (2002).
- [4] Hong, H., Tong ,W., Fang, H., Shi, L., Xie, R. Q., Wu, J., Perkins, R., Walker, J. D., Branham, W. S., and Sheehan, D. M., Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts, *Environ. Health Perspect.* 110, 29–36. (2002).
- [5] Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R., Branham, W. and Sheehan, D., QSAR models using a large diverse set of estrogens, *J. Chem. Inf. Comput. Sci.* 41, 186–195 (2001).
- [6] Elsby, R., Ashby, J., Sumpter, J. P., Brooks, A. N., Pennie, W. D., Maggs, J. L., Lefevre, P. A., Odum, J., Beresford, N., Paton, D., Park, B. K., Obstacles to the prediction of estrogenicity from chemical structure: Assay-mediated metabolic transformation and apparent promiscuous nature of the estrogen receptor, *Biochem. Pharmacol.* 60, 1519–1530 (2000).
- [7] Anderson, E., Veith, G. D., and Weininger, D., SMILES: A line notation and computerized interpreter for chemical structure, EPA/600/M-87-021, Technical Report, U.S. Environmental Protection Agency, Duluth (Minnesota) 1987.
- [8] Weininger, D., SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules, *J. Chem. Inf. Comput. Sci.* 28, 31–36 (1988).
- [9] Weininger, D., Weininger, A., and Weininger, J. L., SMILES. 2. Algorithm for generation of unique SMILES notation, *J. Chem. Inf. Comput. Sci.* 29, 97–101 (1989).
- [10] Blair, R., Fang, H., Branham, W. S., Hass, B., Dial, S. L., Moland, C. L., Tong, W., Shi, L., Perkins, R., and Sheehan, D. M., Estrogen Receptor Relative Binding Affinities of 188 Natural and Xenochemicals: Structural Diversity of Ligands, *Toxicol. Sci.* 54, 138–153 (2000).
- [11] Soto, A. M., Justicia, H., Wray, J. W., and Sonnenschein, C., *p*-Nonylphenol: an estrogenic xenobiotic released from "modified" polystyrene, *Environ. Health Perspect.* 92, 167–173 (1991).
- [12] Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N., and Serrano, F. O., The E-SCREEN Assay as a Tool to Identify Estrogens: an update on Estrogenic Environmental Pollutants, *Environ. Health Perspect.* 103, Suppl. 7, 113–122 (1995).
- [13] Nishihara, T., Nishikawa, J., Kanayama, T., Dakeyama, F., Saito, K., Imagawa, M., Takatori, S., Kitagawa, Y., Hori, S., and Utsumi, H., Estrogenic activities of 517 chemicals by yeast two-hybrid assay, *J. Health Sci.* 46, 282–298 (2000).
- [14] Fang, H., Tong, W., Perkins, R., Soto, A., Prechtel, N., and Sheehan, D. M., Quantitative comparison of *in vitro* assays for estrogenic activity, *Environ. Health Perspect.* 108, 723–729 (2000).

Received on ■;