ICE Chemical Characterization Tool

Introduction

The ICE Chemical Characterization tool allows you to view the predicted physicochemical properties, absorption, distribution, metabolism, and excretion (ADME) properties, functional use categories, and curated product use categories of a chemical set. The tool allows you to either explore the properties and use categories for one set of chemicals or compare two sets of chemicals. You can examine the properties and use categories of ICE Chemical Quick Lists or provide your own list of chemical identifiers.

The Chemical Characterization tool queries ICE for available properties. These are either experimental values or predictions derived from the Open Structure-activity/property Relationship App (OPERA; Mansouri et al.2018). Results are returned as tables, interactive boxplots for visual comparisons, and principal component analysis (PCA) plots.

Chemical functional use and curated product use categories in ICE are derived from the U.S. Environmental Protection Agency's (EPA's) Chemicals and Products Database (CPDat v3; Williams et al. 2017). You can visualize functional use data through heat maps in the "Functional Use Explorer" in the "Results" view. There are separate heat maps for reported functional use harmonized to functional use categories established by the Organisation for Economic Co-operation and Development (OECD) and functional use predicted by quantitative structure-use relationship (QSUR) models (Phillips et al. 2018). You can also visualize curated product use data through interactive circle graphs and downloadable tables in the "Curated Product Use Explorer" within the "Results" view.

Located throughout the Chemical Characterization tool are green information buttons to help you better understand key features and results. When you hover over a button, brief explanatory text will appear. When you click a button, more details will appear in a text box that can be resized and relocated on the screen.

The "Help" button on the left side of the display opens a text box with a brief description of the tool and links to the webpage and downloadable user guides. If you encounter a problem using the tool, click the "Report an Issue" button on the left side of the display to generate an email to ICE Support. Click the "Cite ICE" button to view example citation formats for referencing ICE in your research publication that utilizes ICE data or outputs from ICE tools.

Building a Chemical Characterization Query

Figure 1 shows the Chemical Characterization tool "Input" view. You can toggle between the "Input" view and "Results" view by clicking on the tabs on the left side of the screen. The tool window defaults to "Input" view when it is first opened.

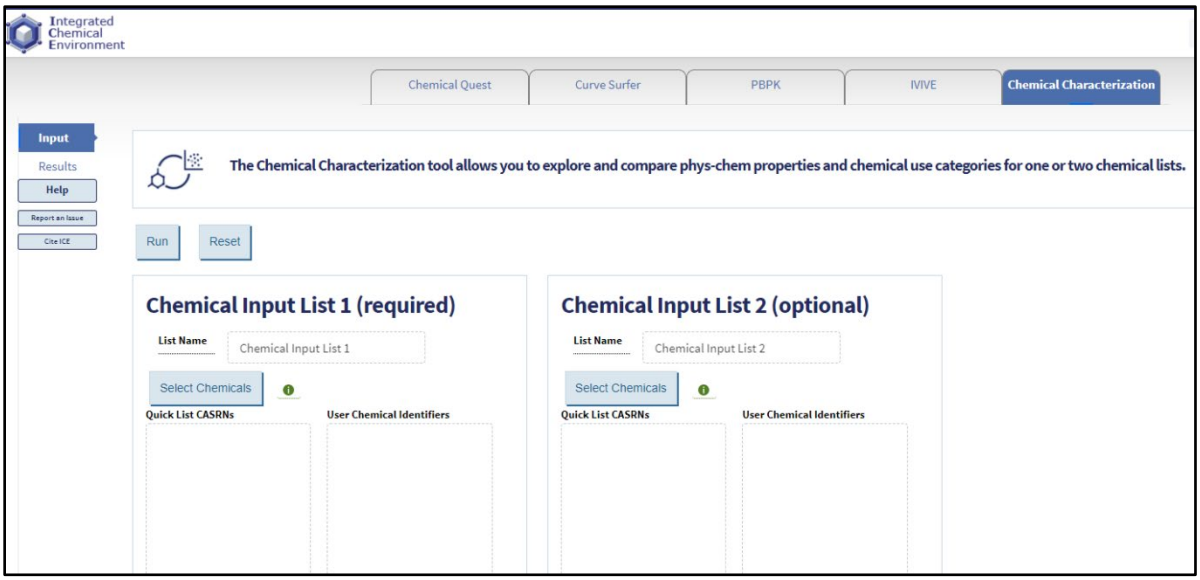

Figure 1. ICE Chemical Characterization tool Input view.

The tool has two "Chemical Input List" fields. The first, on the left, is required, and the second, on the right, is optional.You can populate the text boxes within the fields using one or both of two input methods (**Figure 2**):

1. Populate the left-hand text box in each field ("Quick List CASRNs") by selecting one or more Chemical Quick Lists. Select one or more Chemical Quick Lists by clicking "Select Chemicals", which opens a dialog box with ICE Chemical Quick Lists. Select the checkboxes in the dialog box to choose one or more chemical lists; click "Finished" when you are done. Chemical Abstracts Service Registry Numbers (CASRNs) from chemicals in the selected list(s) will appear in the left-hand text box under "Quick List CASRNs".

2. Populate the right-hand text box in each field ("User Chemical Identifiers") by entering your own list of chemical identifiers, one per row. You can use any combination of the following identifiers:

    - CASRNs.
    - Distributed Structure-Searchable Toxicity Substance Identifiers (DTXSIDs).
    - Simplified molecular-input line-entry system (SMILES) strings.
    - Hashed International Chemicals Identifiers (InChIKeys).
    - Chemical names.

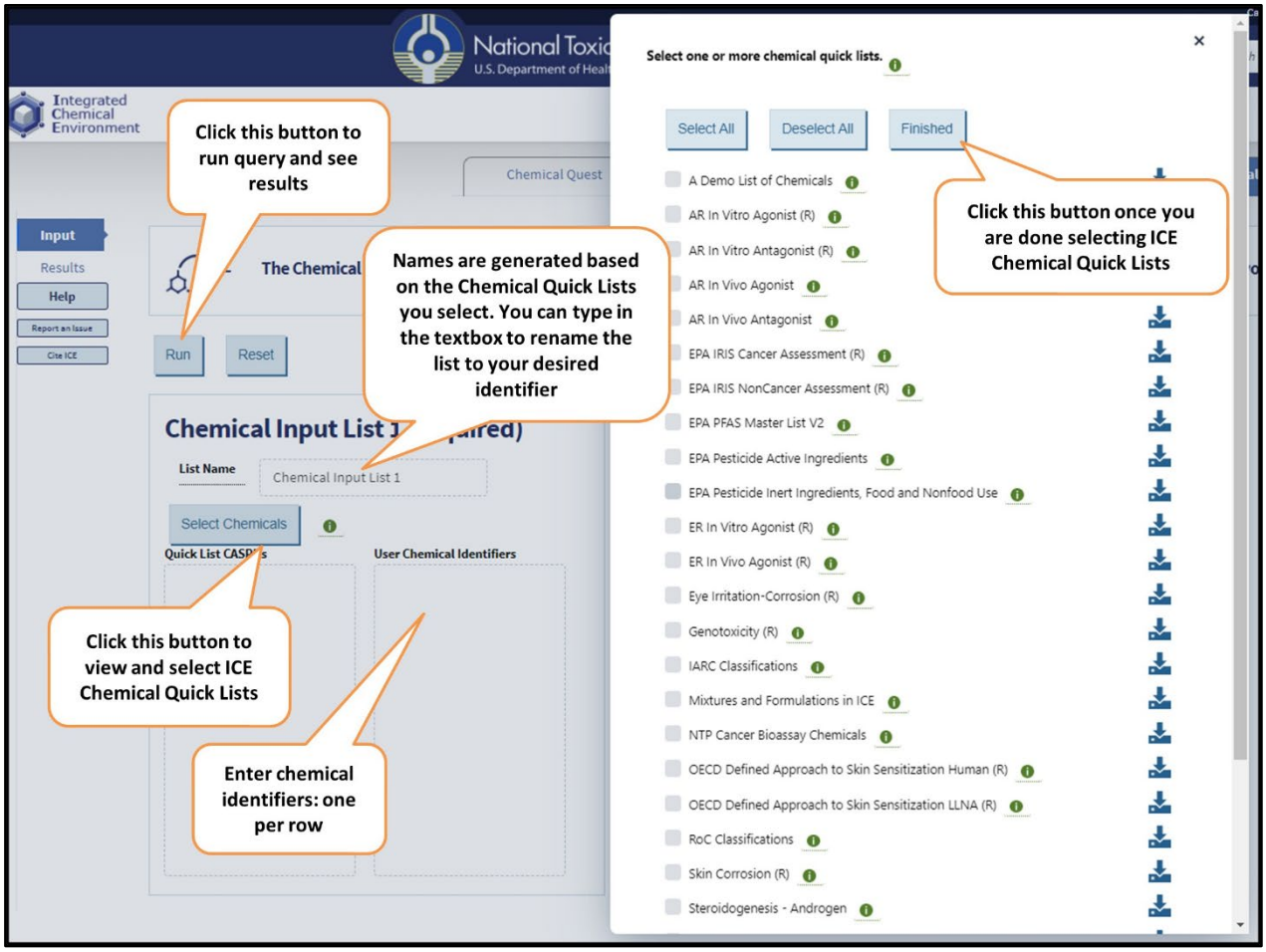

Figure 2. Input chemicals by selecting ICE Chemical Quick Lists or typing chemical identifiers into the right-hand text box.

For either chemical input method a list name will be generated.

3. If you populate your query using Chemical Quick Lists, the name of each list will appear in the list name box separated by a comma.
    - If you enter your own chemical identifiers, “Custom Chemical List” will appear as the list name.
    - You may edit the automatically generated list name to a name of your choice in the text box next to “Enter List Name”

Run Query

Once you have selected chemicals, click the “Run” button at the top of the page (**Figure 2**) to run your query. To reset the chemical input lists, click on “Reset”.

Viewing Chemical Characterization Results

The window will switch to “Results” view upon clicking the run button. The “Results” view displays the “Curated Product Use Explorer”; four clickable headings above this provide views of chemical properties or functional use categories.

The ICE Chemical Characterization tool "Results" view provides a variety of tables and plots through which you can explore your data. Hover over any plot to display a menu in the top right corner of the plot, providing options for adjusting the view and exporting the plot. To zoom in and highlight an area of interest in the plot, click and drag to draw a rectangle around the desired area. After zooming in, double-click on the plot to zoom out. For more information about interactive graph visualization options, consult the Interactive Graphs User Guide.

Chemical Properties Summary

Results under "Chemical Properties Summary" are displayed in two tables. If your query included two chemical lists, the default display shows results from both chemical lists. Use the dropdown list above the tables to change the view to display results from either of the chemical lists. Click the "TXT" icon above either table to export results as a tab-delimited text file; click the "XLSX" icon to export as an Excel workbook. The current table view with any filters applied will be downloaded.

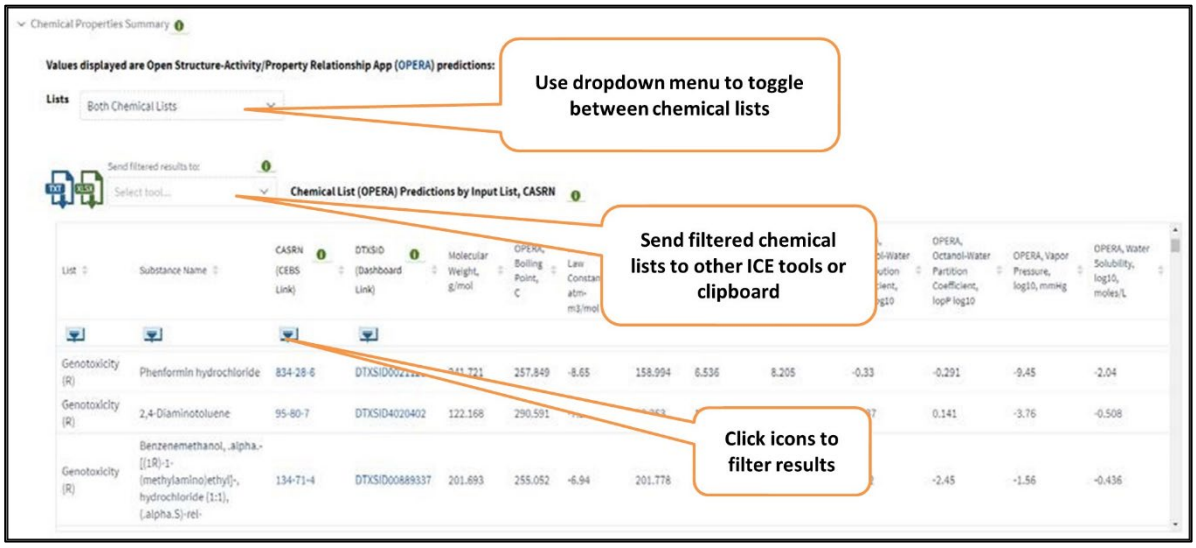

Figure 3. ICE Chemical Characterization tool Results view showing table with detailed view.

The first table (**Figure 3**) provides a detailed view of all the physicochemical property data returned for each chemical. The first column specifies a chemical's inclusion in an input list. If a chemical is present in both input lists, it will be listed twice.

- Click on a chemical's CASRN to be directed to the available test article page about that chemical in the National Toxicology Program's Chemical Effects in Biological Systems (CEBS) database.
- Click on a chemical's DSSTox Substance Identifier (DTXSID) to be directed to the EPA CompTox Chemicals Dashboard entry for that chemical, which provides the chemical structure and other physicochemical and experimental properties.

You can sort the table by clicking on any column heading. Click the filter icon below any of the first four column headings to filter on specific chemicals or chemical lists (**Figure 3**). In the dialog box that opens, select chemicals or lists of interest; click "Close" when done. The filter icon will turn red to indicate a filter is applied. To remove a filter, open the dialog box again and select individual checkboxes or the box at the top of the column to unselect all boxes. Any applied filters will be applied to both summary data

tables and their downloads.

The second table (**Figure 4**), provides mean, median, quartiles, and range for the physicochemical data displayed in the first table.

Endpoint	Min	25th	Median	Mean	75th	Max
OPERA, Water Solubility	-10.4	-4.41	-2.37	-2.82	-0.955	1.355
Molecular Weight	30.026	122.992	212.25	217.53	275.386	1202.611
OPERA, Octanol-Air Partition Coefficient	0.969	3.991	7.947	7.047	9.464	11.757
OPERA, Henry's Law Constant	-11.1	-7.72	-6.48	-6.24	-4.32	-1.27
OPERA, Negative Log of Acid Dissociation Constant	-1.66	0.0	1.0	1.923	2.026	11.326
OPERA, Vapor Pressure	-10.7	-7.39	-4.54	-3.9	-0.523	4.385
OPERA, Melting Point	-160.0	6.729	108.43	91.542	168.208	374.84
OPERA, Octanol-Water Distribution Coefficient	-11.6	0.0895	1.671	1.683	3.327	7.978
OPERA, Boiling Point	-75.8	196.006	294.854	267.744	356.028	535.958

Figure 4. ICE Chemical Characterization tool Results view showing table with summary statistics.

Visualization of Chemical Properties

Below the tables, click on the “Visualization of Chemical Properties” heading to display boxplots that provide visual summaries for each parameter (**Figure 5**).

If there were chemicals in your query that did not have property data accessible by the ICE database, you can find them by clicking the subheading “All Chemical Identifiers Not Returned by Query”. Under the subheading there is a link to a text file to download the missing CASRNs. Clicking on additional subheadings will display CASRNs not returned for each individual list.

You can use the boxplots to view the distribution of ten chemical parameters for chemicals in your list(s). However, only two boxplots can be viewed at a time. Click the dropdown lists above the plots to choose the parameters to be displayed.

- Within each figure, boxplots are split by chemical list with the first list in red and optional second list in blue.

An additional boxplot titled "ICE Properties" displays the property data for all chemicals in the ICE database, allowing users to compare their selected chemical list to the ICE chemical space. Hover over a

boxplot to view summary statistics for all chemicals in the chemical list.

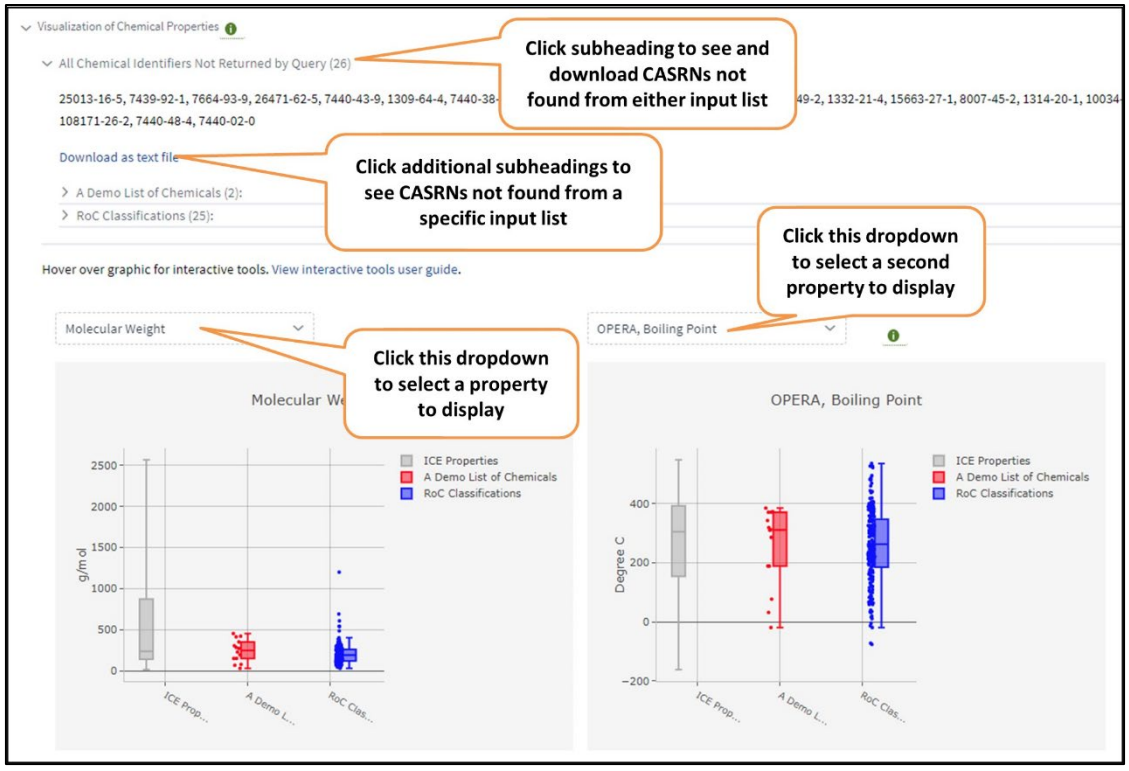

Figure 5. ICE Chemical Characterization tool "Results" view showing boxplots for two chemical input lists.

Point values for the user-provided chemicals are plotted adjacent to the boxplots. Hover over an individual data point to display the CASRN for that chemical as well as the value for that individual data point. Hover over the plot to display a menu in the top right corner of the plot, providing options to adjust the view and export the plot. Zoom in on an area of interest in the plot by clicking and dragging to draw a rectangle around the area (**Figure 6**)**.** Double-click on the plot to zoom back out.

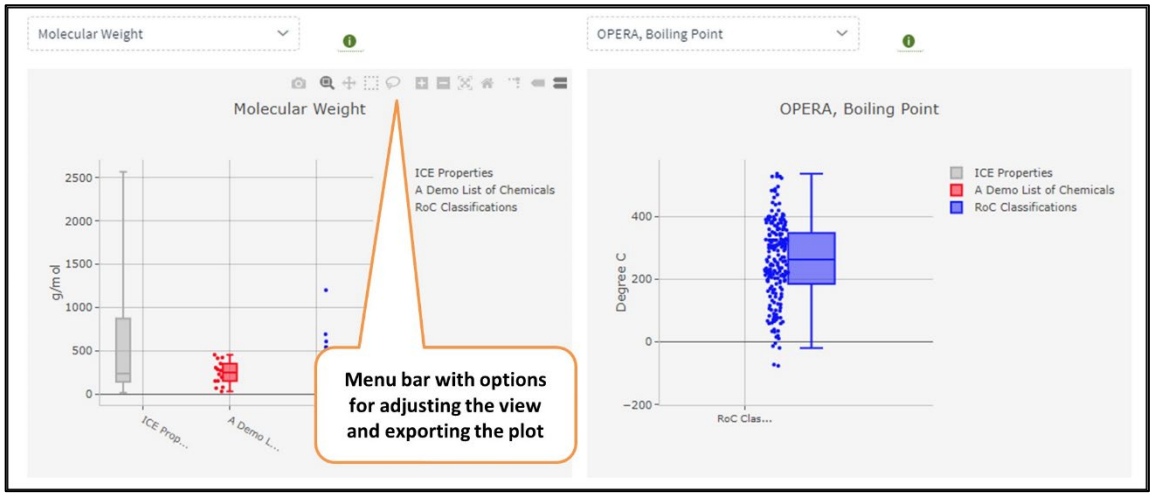

Figure 6. The OPERA Boiling Point plot (right side) zoomed in to look at a specific chemical list in more detail.

Interactive PCA Plots

Below the boxplots, click on the "Interactive PCA" heading to view two PCA plots from your selected lists. These plots are generated by performing PCA on the physicochemical properties or molecular descriptors for your chemical lists. The PCA plots visualize all chemicals in your query that have the selected properties or descriptors, with the position of each point representing a numeric summary of the selected parameters for a single chemical. The distance between points corresponds to the similarity between chemical parameters. For more details about PCA and interpretation of the PCA plots, see **Appendix 2**.

These plots have a variety of display options and allow you to explore parameters of interest. For both PCA plots, you can hover over an individual data point to display the CASRN and name for a chemical.

Static PCA Plots

The static plot shows your query chemicals relative to a background of a larger chemical set. The chemicals in your first list are red and those in the optional second list are blue. If chemicals appear in both lists, the points are purple.

Use the "Selected Background" dropdown list (left) above the static plot to select a backgroundchemical set for your display, which will be displayed in gray (**Figure 7**):

- DSSTox: displays your chemical lists against the background of over 800,000 chemicals in DSSTox with physicochemical properties derived from OPERA (this may take a few seconds to load due to the large chemical set).
- Tox21: displays your chemical lists against the background of over 9,000 chemicals tested in the EPA's Toxicology in the 21st Century (Tox21).

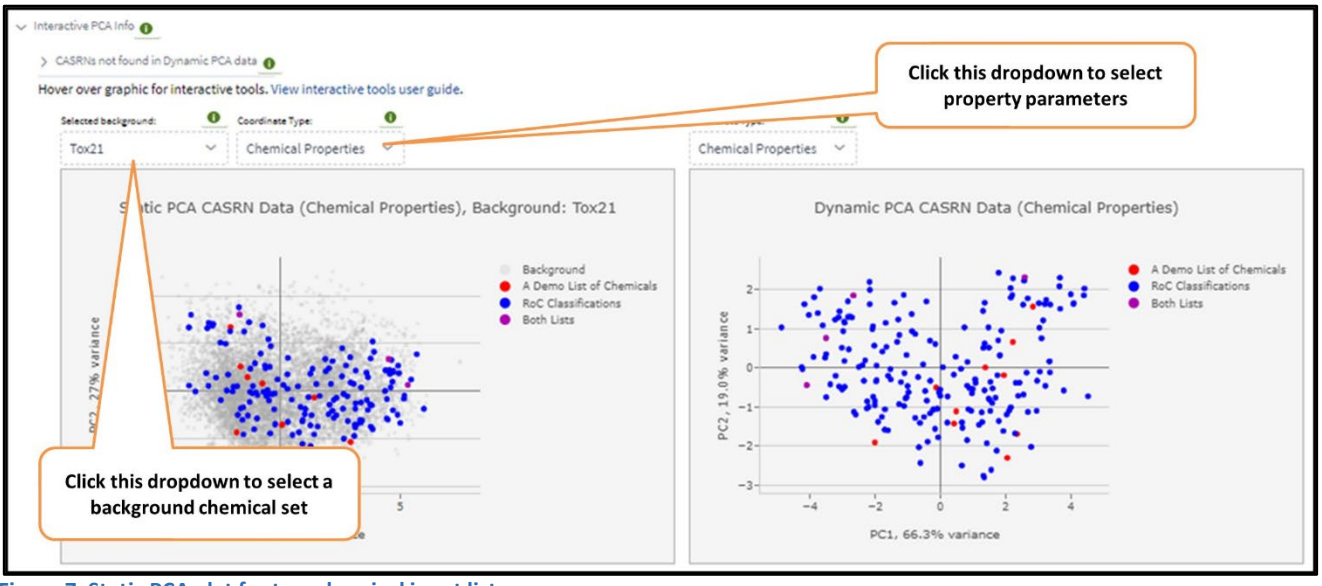

Figure 7. Static PCA plot for two chemical input lists.

Use the "Coordinate Type" dropdown list (right) above the static plot to select property parameters for your display (**Figure 7**):

- Chemical Properties: 8 properties that describe characteristics of a chemical substance during a reaction or a chemical change, such as boiling point, partition coefficients, and molecular weight.
- Molecular Descriptors: 76 descriptors calculated using PaDEL software that produce 1D, 2D, and 3D mathematical representations of chemical structures.

For a complete list of chemical properties and molecular descriptors used in the static and dynamic PCA plots see **Appendix 2**.

Dynamic PCA Plots

The dynamic PCA plot (**Figure 8**) visualizes property similarities among the members of each chemical list. Dynamic plots are ideal when you want to compare the characteristics of your two chemical lists to each other. Use the "Coordinate Type" dropdown list above the dynamic plot to select property parameters for your display (**Figure 8**):

- Chemical Properties: 8 properties that describe characteristics of a chemical substance during a reaction or a chemical change, such as boiling point, partition coefficients, and molecular weight.
- Molecular Descriptors: 76 descriptors calculated using PaDEL software that produce 1D, 2D, and 3D mathematical representations of chemical structures.

For a complete list of chemical properties and molecular descriptors used in the static and dynamic PCA plots see **Appendix 2**.

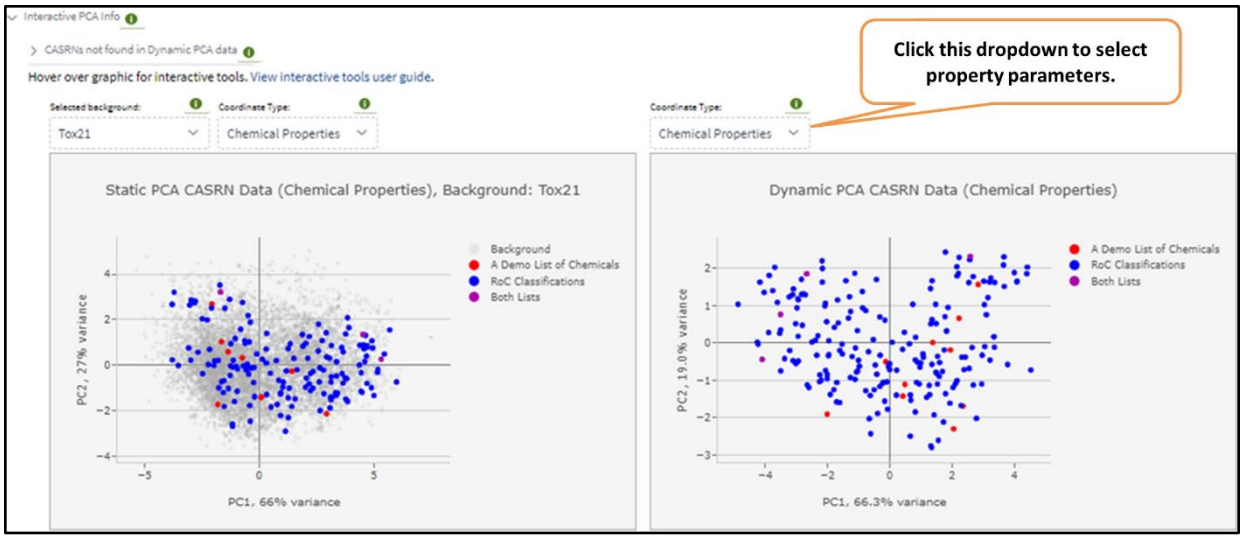

Figure 8. Dynamic (right) PCA plot for two chemical input lists.

Functional Use Explorer

Below the PCA plots, click on the "Functional Use Explorer" heading to explore OECD Functional Use Categories and Predicted Functional Use data from EPA's Chemicals and Products Database (CPDat v3). The tool consists of two heatmaps (**Figure 9**) that represent how the chemicals in your list(s) are used in products. "OECD Functional Use" is reported functional use harmonized to OECD categories. "Predicted

Functional Use" is predicted from one-vs.-all random forest models (Phillips et al. 2018). Each heatmap has a title identifying the source of the data and how many of your queried chemicals were present in that data set. The data set's functional use categories are listed on the x-axis and the shared y-axis lists substances in your query.

Within each heatmap, a cell is colored if the chemical on the y-axis has the given functional use on the x-axis. A cell is colored red if the chemical is in the first chemical set, blue if the chemical is in the optional second chemical set, and purple if the chemical is in both chemical sets. A grayed-out, empty cell means that the chemical does not have a given functional use.

As chemical names on the y-axis are truncated to provide a larger view of the graph, the best way to view the full chemical name and functional use is to hover your mouse over a cell. When you hover over a cell the resulting text box will include the functional use, chemical name, and the name of the chemical input list.

The heatmap will be paginated if more than 50 chemicals in your query have a functional use. To view the results for additional chemicals, click the arrows at the bottom of the heatmap.

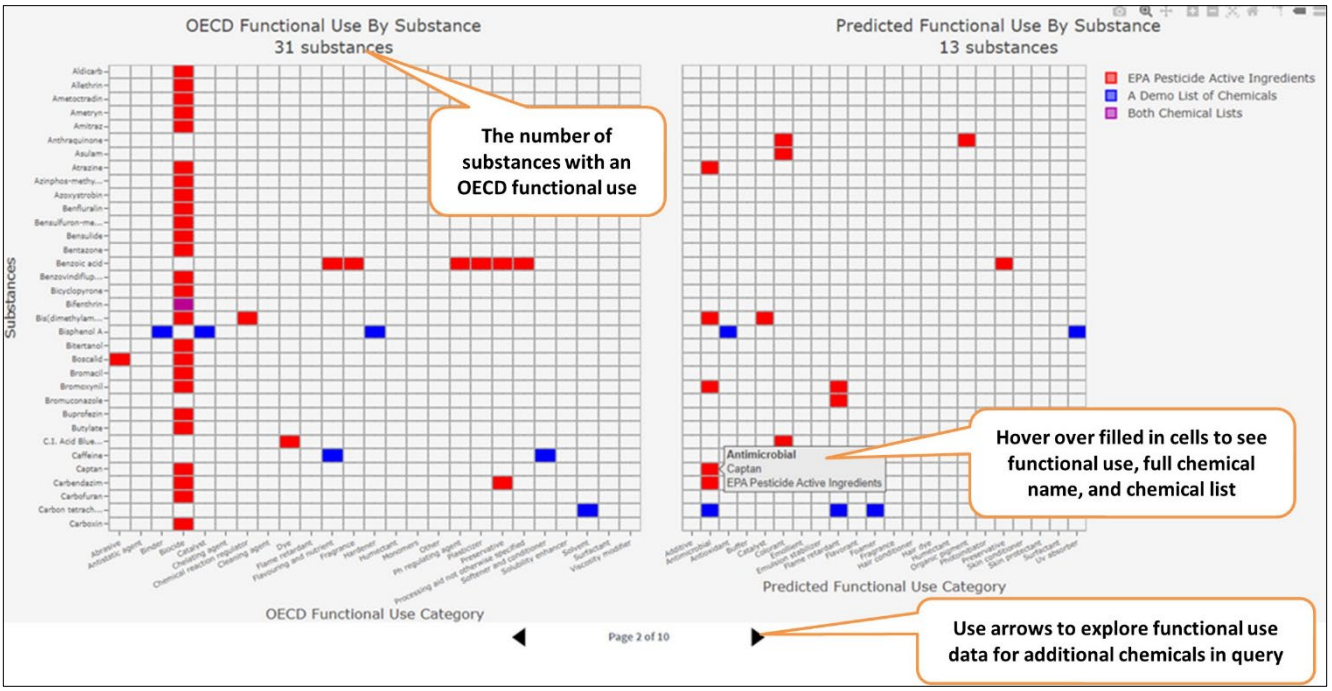

Figure 9: Functional Use Explorer heatmaps for two chemical input lists.

Curated Product Use Explorer

The "Curated Product Use Explorer" tool (**Figure 10**) is at the bottom of the Chemical Characterization tool Results view. This tool lets you view the major product categories for your chemical lists using information drawn from EPA's CPDat. This ICE tool was previously called the "Consumer Use Explorer" but was renamed for ICE version 4.0.1.

If there were chemicals in your query that did not have property data in the ICE database, you can find them by clicking the subheading "All DTXSIDs Not Returned by Query." There is also an option to

download the missing DTXSIDs as a text file, and additional subheadings will display DTXSIDs not returned for the individual lists.

The "All Chemicals Popup" button above the plots (**Figure 10**) opens a window that allows you to explore all chemicals in CPDat via the same graphs and tables described below. This allows you to compare categories of your chemical set to the full set of chemicals in CPDat. (The window will take a few seconds to open.)

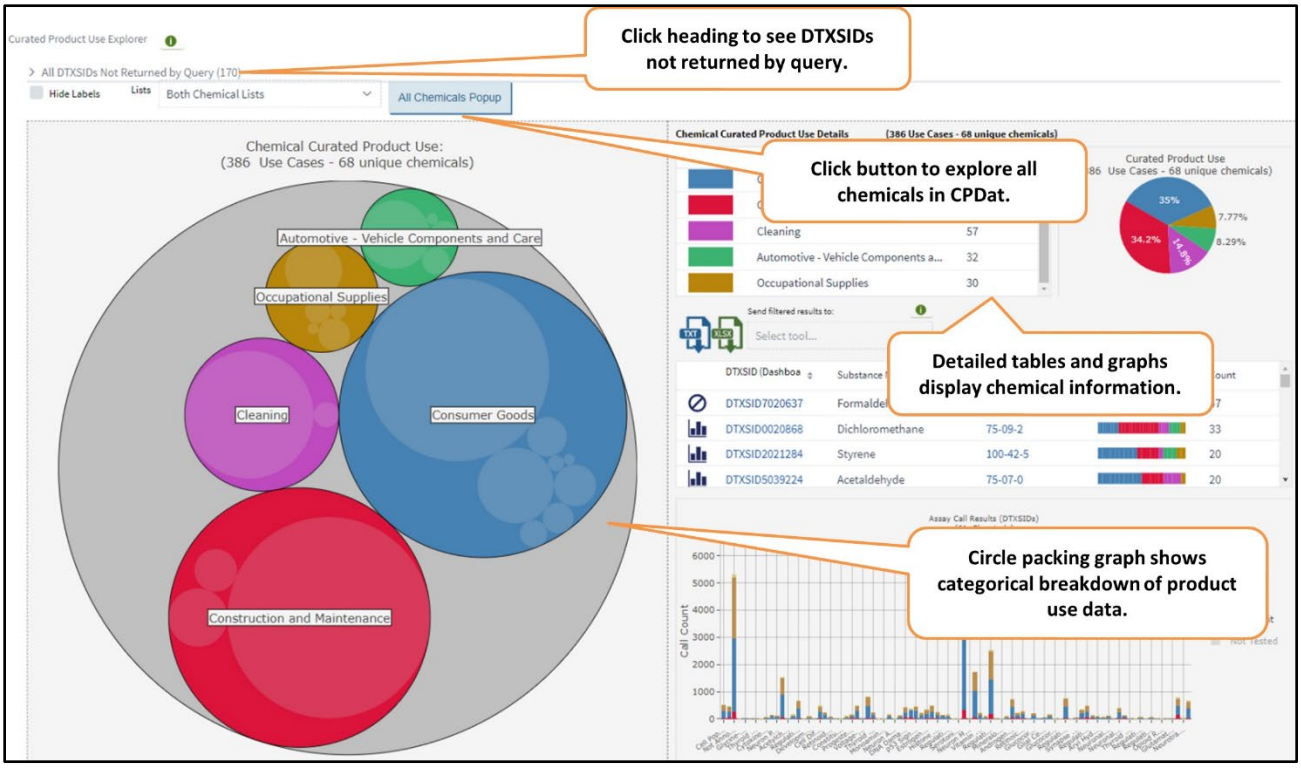

Figure 10. Curated Product Use Explorer for two chemical input lists.

Curated Product Use Circle Plot

This plot lets you view categories of chemicals within interactive circle graphs. The initial display shows the parent category as a large bubble. This contains subcategories as small circles packed within the parent category bubble. The size of each circle is proportional to the number of chemicals associated with that subcategory. To explore these subcategories, click on a small circle to zoom in and display further divisions within the subcategory. To help you keep oriented, the graph title changes as you zoom in to list the category and subcategories you are currently exploring (**Figure 11**).

Options above the table allow you to hide category labels and choose to display data from both or either of your chemical lists (**Figure 11**). You can hover your mouse over a circle to view the category name, number of uses, and number of chemicals represented in that set whether the labels are displayed or not. These values are different due to the possibility of chemicals falling within several subcategories.

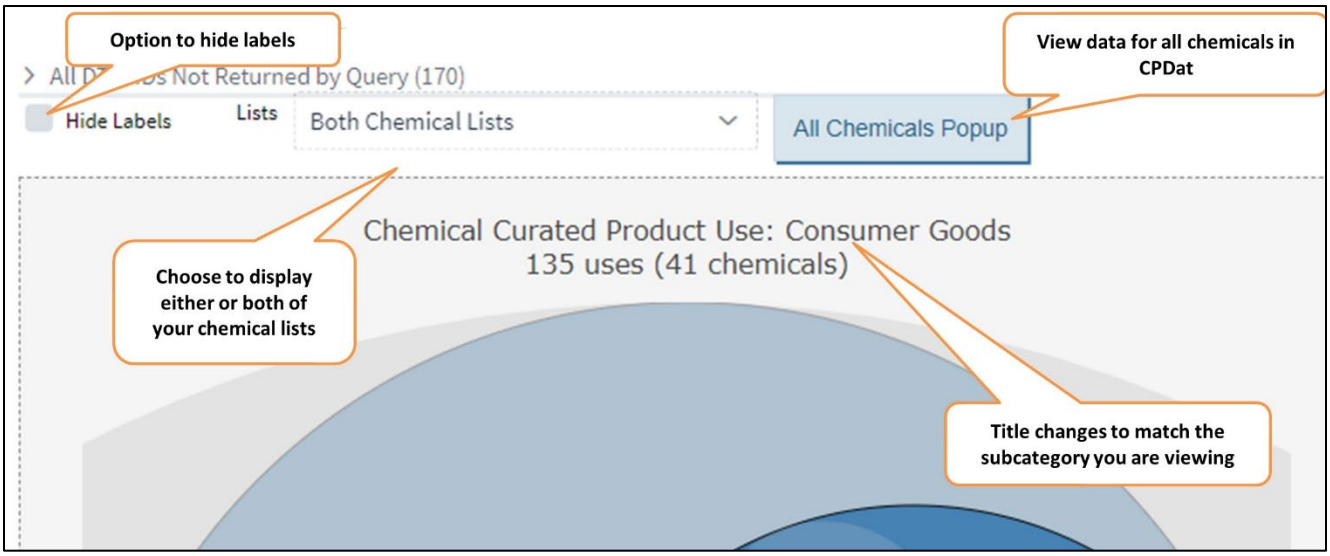

Figure 11. Using the Curated Product Use Explorer circle plots to explore subcategories.

Chemical Curated Product Use Details

To the right of the circle plot, additional tables and graphs allow you to further explore the product use characteristics of your chemical lists. These tables and graphs show characteristics of whatever part of your query set is being displayed in the circle plot (**Figure 12**).

- The table in the top left of this section shows the categories available for viewing along with the chemical count and use case count, which is the number of curated product use and chemical pairs returned by your chemical query.
- The circle in the top right represents the chemical use categories shown in the circle plot as a pie chart and lists the proportion of chemicals in each category as a percentage of the entire set.
- The "Curated Product Use Categories" by DTXSID, CASRN table in the middle lists each chemical represented in the set displayed in the circle plot. Hover over the colored segments of the bars in the "Sub Categories" column to display what subcategories that chemical is included in.
    - Click on a chemical's CASRN to be directed to the test article page about that chemical in the CEBS database.
    - Click on a chemical's DTXSID to be directed to the EPA CompTox Chemicals Dashboard entry for that chemical.
- The "Assay Call Results" histogram at the bottom shows the collective activity of the chemicals within the set in different types of high-throughput assays. If you want to only view specific bioactivities, click the labels you wish to remove in the legend to the right and the graph will update accordingly. To restore the deselected bioactivity, click on the label. Hover over colored areas of the bars to show the type of assay, call, number of assay results in that set, and the proportion of all assay results. The menu in the upper right of the plot provides other options for adjusting the view and exporting the plot.

All data in the tables and graphs to the right of the bubble plot will change to reflect the information being viewed in the interactive circle graph.

To download and save your list of chemicals and the categories they belong to, click the icons below the category list in the right panel (**Figure 12**) to export either a text file or an Excel file. The data exported will reflect the chemicals being displayed in the circle graph.

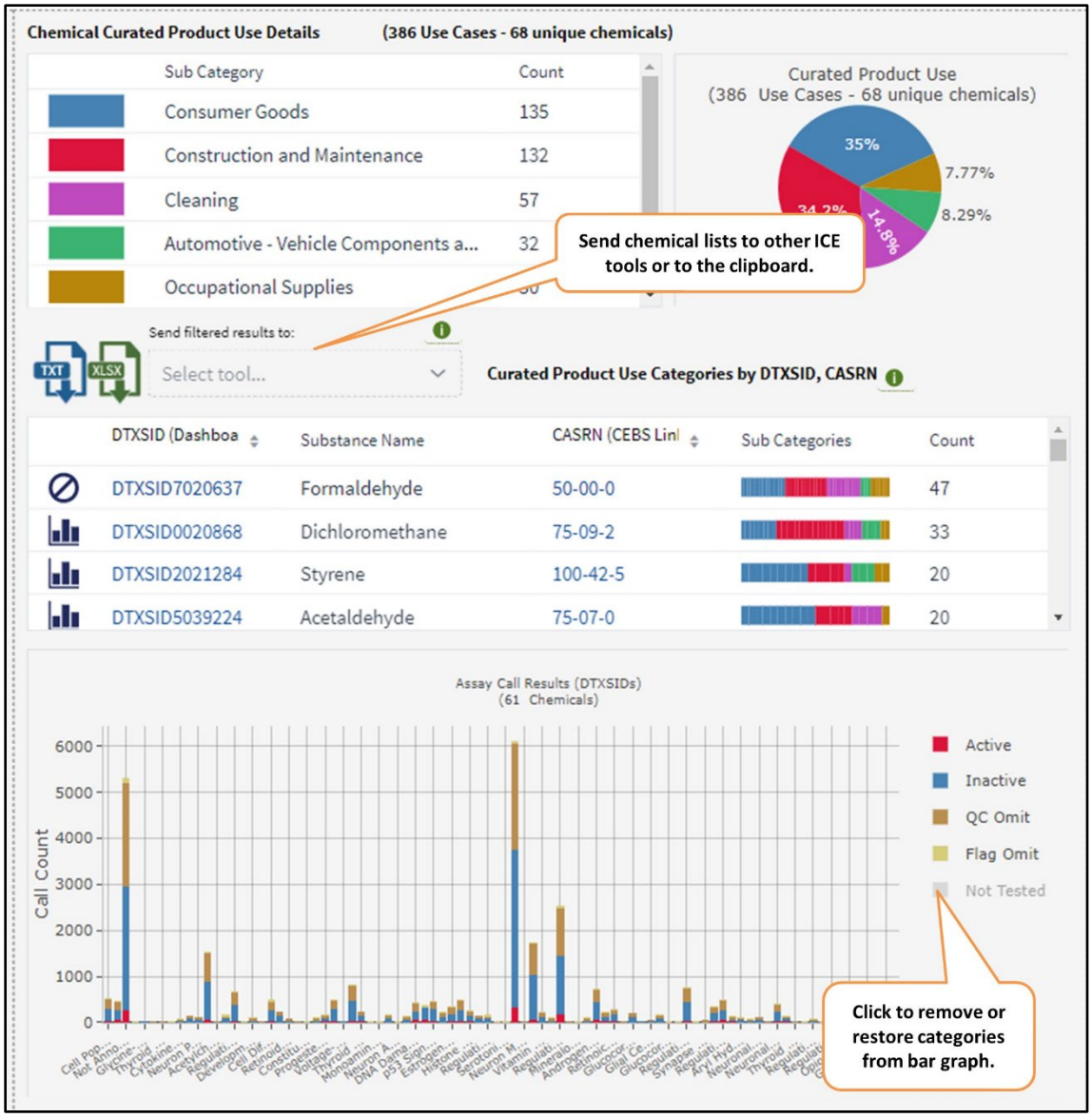

Figure 12. Detailed view of data within a Curated Product Use Explorer category.

Using Results to Query Other ICE Tools

Both the Chemical Properties Summary table at the top of the “Results” view and the "Consumer Use Categories by DTXSID, CASRN" table within the “Consumer Use Explorer” include a "Send filtered results to" dropdown list next to the download buttons. This list allows you to send the chemicals in your current view to other ICE tools. For details on the use and outputs of these tools, refer to their user guides.

- Click "Search" to query and retrieve all data in ICE for the selected chemicals.
- Click "Chemical Quest" to query ICE for chemicals that are structurally similar to your chemicals.
- Click "Curve Surfer" to display activity curves for chemicals that have curated high-throughput screening (cHTS) assay data in ICE.
- Click "PBPK" to send data to the Physiologically Based Pharmacokinetics (PBPK) tool. This toolgenerates predictions of tissue-specific chemical concentration profiles following a dosing event.
- Click "IVIVE" to send data to the In Vitro to In Vivo Extrapolation (IVIVE) tool. This tool allows you to estimate the daily equivalent administered dose (EAD) that would result in the plasma concentration of a chemical equal to the activity concentration in a given in vitro assay.

You can also copy the CASRNs, DTXSIDs, Original SMILES or QSAR-ready SMILES to the clipboard.

Appendix 1: Physicochemical Property and ADME Predictions in the OPERA Model

OPERA (Mansouri et al. 2018) is a free and open-source/open-data suite of quantitative structure activity relationship (QSAR) models. OPERA can predict physicochemical properties and environmental fate and toxicity endpoints. It can also provide assessments of applicability domain and prediction accuracy. All OPERA models were built on curated data and standardized QSAR-ready chemical structures. OPERA documentation, downloadable code, and other resources are available at https://github.com/NIEHS/OPERA.

Appendix 2: PCA Plots

What is PCA?

PCA is a method for transforming a large numeric data set into a set of variables, or principal components (PCs) by projecting the raw data onto a new coordinate system (Jolliffe and Cadima 2016). The PCs are linear combinations of the original data variables and are designed to capture variability across multiple parameters. Visualizing the PCA results can reveal patterns, clusters, and outliers among individual data points that otherwise would be difficult to identify within the raw data.

PCA Plots in ICE

The PCA plots in ICE can be generated with a set of chemical properties or molecular descriptors. The variables that are used within the PCA are as follows:

- **Chemical properties**: BP_pred (predicted boiling point), LogHL_pred (predicted Henry's Law), LogWS_pred (predicted water solubility at 25°C), LogKOA_pred (predicted log octanal/air partition coefficient), MP_pred (predicted melting point), LogVP_pred (predicted log vapor pressure), MW_total (total molecular weight)

- **Molecular descriptors**: Zagreb, XLogP, WPOL, MW, VAdjMat, VABC, TopoPSA, LipinskiFailures, nRotB, topoShape, MLogP, nAtomLAC, nAtomP, nAtomLC, HybRatio, nHBDon, nHBAcc, bpol, nB, nBase, nAtom, nAromBond, naAromAtom, apol, ALogP, ALogp2, AMR, nAcid, nHeavyAtom, nH, nC, nN, nO, nS, nP, nF, nCl, nBr, nI, nX, nBonds, nBonds2, nBondsS, nBondsS2, nBondsS3, nBondsD, nBondsD2, nBondsT, nBondsM, Sv, Sse, Spe, Sare, Sp, Si, Mv, Mse, Mpe, Mare, Mp, Mi, nRing, n3Ring, n4Ring, n5Ring, n6Ring, n7Ring, n8Ring, n9Ring, n10Ring, n11Ring, RotBFrac, nRotBt, RotBtFrac, topoRadius, topoDiameter

ICE provides two types of PCA plots:

- The static PCA plot is derived from a PCA calculated with your query chemicals and a background chemical set (DSSTox or Tox21). Query chemicals are highlighted against the background chemical set to contextualize the diversity of your queried chemicals. If your chemicals are not part of the background chemical set, they will not be displayed in the PCA.
- The dynamic PCA plot is derived from a PCA calculated with only your query chemicals. The dynamic PCA plot can be used to evaluate the similarity of your query chemicals or to compare two chemical lists.

Both the static and dynamic PCA plots visualize the first two PCs from the underlying PCA. In a PCA, each PC describes a percentage of the data variability. The first PC (PC1) describes the highest percentage of variability within the data and the second PC (PC2) describes the second highest percentage of variability. The coordinates of each chemical on PC1 and PC2 are derived from their defined linear combination of the raw data, such that the distance between points corresponds to similarity of the chemical properties or molecular descriptors.

Appendix 3: References and More Information

Jolliffe IT, Cadima J. 2016. Principal component analysis: a review and recent developments. Phil TransR Soc A 374: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202.

Mansouri K, Grulke CM, Judson RS et al. 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform 10:10. https://doi.org/10.1186/s13321-018-0263-1

Phillips KA, Wambaugh JF, Grulke CM, Dionisio KL, Isaacs KK. 2017. High-throughput screening of chemicals as functional substitutes using structure-based classification models. Green Chem 19(4):1063–1074. https://doi.org/10.1039%2FC6GC02744J

Williams A. 2017. Chemical and Products Database (CPDat) MySQL Data File. U.S. Environmental Protection Agency Center for Computational Toxicology and Exposure. Dataset. https://doi.org/10.23645/epacomptox.5352997

Appendix 4: Abbreviations

This list includes both abbreviations used within this User Guide and abbreviations used in the ICE Chemical Characterization tool interface.

ADME: Absorption, Distribution, Metabolism, and Excretion

AR: androgen receptor

BP: boiling point

BP_pred: predicted boiling point

CASRN: Chemical Abstracts Service Registry Number

CEBS: Chemical Effects in Biological Systems

cHTS: curated high-throughput screening

CPDat: Chemicals and Products Database

DSSTox: Distributed Structure-Searchable Toxicity database (U.S. Environmental Protection Agency)

DTXSID: DSSTox Substance Identifier

EAD: equivalent administered dose

EPA: U.S. Environmental Protection Agency

ER: estrogen receptor

HL: Henry's Law constant (air/water partition coefficient)

IARC: International Agency for Research on Cancer

InChIKey: hashed International Chemicals Identifier

IVIVE: in vitro to in vivo extrapolation

KOA: octanol-air partition coefficient

LogD, pH 7.4: octanol-water distribution constant at pH 7.4

LogHL_pred: predicted Henry’s Law

LogKOA_pred: predicted log octanal/air partition coefficient)

LogP: octanol-water partition coefficient

LogVP_pred: predicted log vapor pressure

LogWS_pred: predicted water solubility at 25°C

MP_pred: predicted melting point

MP: melting point

MW: molecular weight

MW_total: total molecular weight

OECD: Organisation for Economic Co-operation and Development

OPERA: Open Structure-activity/property Relationship App

PBPK: physiologically based pharmacokinetic

PC: principal component

PCA: principal component analysis

pKa: acid dissociation constant

QSAR: quantitative structure-activity relationship

QSUR: quantitative structure-use relationship

RoC: Report on Carcinogens

SMILES: Simplified molecular-input line-entry system

TopoPSA: topological polar surface area

Tox21: Toxicology in the 21st Century

VP: vapor pressure

WS: water solubility