ICE Chemical Quest Tool

Introduction

The ICE Chemical Quest tool allows you to query ICE or a user-specified list for chemicals that are structurally similar to the input chemicals. You can build your query using a list of chemical identifiers. Chemical Quest will also accept drawings of chemical structures as query inputs.

The Chemical Quest tool queries ICE for available structurally similar chemicals. You can query all chemicals in ICE or limit your query to a list of chemicals of interest. Structurally similar chemicals are identified through the calculation of a Tanimoto score based on Saagar molecular descriptors (Sedykh et al., 2021). For more information on this process, refer to Appendix 1.

Located throughout the 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 below the “Help” button to generate an email to ICE Support.

Building a Chemical Quest Query

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

There are two options to run a query:

- You can query all chemicals in ICE by specifying one or more chemicals in the input fields.
- Alternatively, you can limit your query to a user-defined list of chemicals by selecting the “Search Custom Chemical List” option. You then specify one or more input chemicals and then select target chemicals from a quick list or provide your defined list of chemicals. It is also possible to provide a quick chemical list and user-defined list at the same time. More detailed information about these options, “Limiting Queries” and the tool in general is provided below.

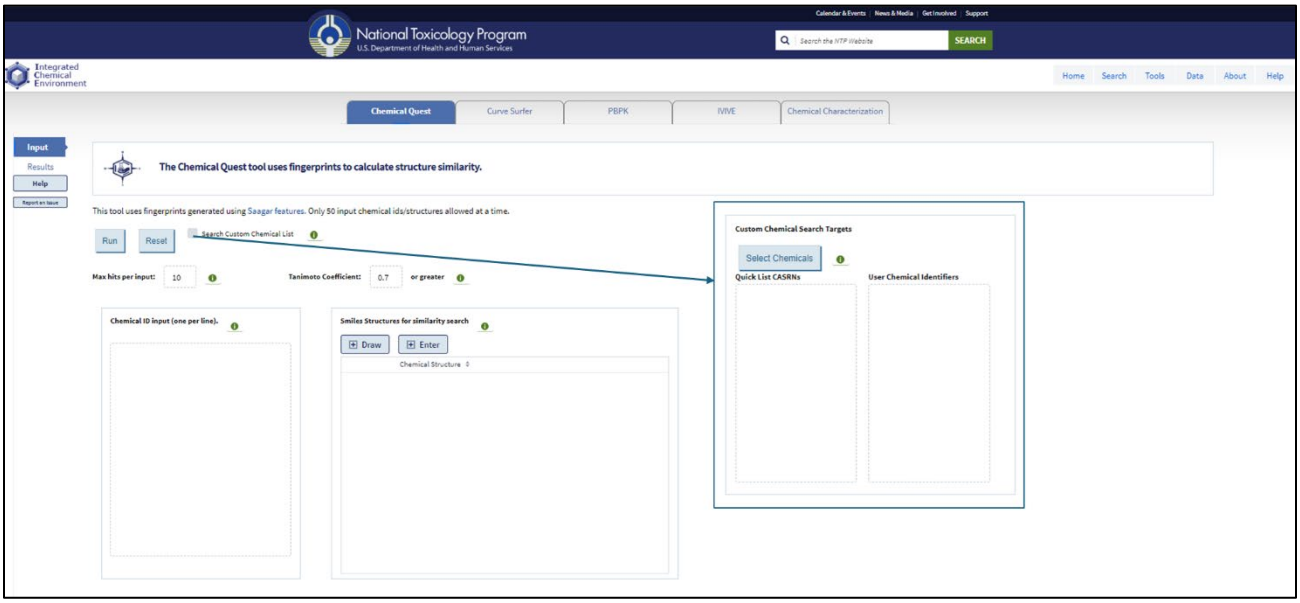

Figure 1. ICE Chemical Quest tool Input view. Note that the "Search Custom Chemical List" option is selected.

The tool has two main chemical input list fields, "Chemical ID input" and "Smiles Structures for similarity search," with a third field appearing when "Search Custom Chemical List" is selected. You must populate at least one of the two main chemical input list fields to run a query (**Figure 2**).

- Populate the left-hand "Chemical ID input" field by entering your own list of chemical identifiers, one per row. Users are required to limit the input size to 50 chemicals maximum to optimize query speed. You can use any combination of the following identifiers:
    - Chemical Abstracts Service Registry Numbers (CASRNs).
    - Chemical names.
    - Distributed Structure-Searchable Toxicity Substance Identifiers (DTXSIDs).
    - Simplified molecular-input line-entry system (SMILES) strings.
    - Hashed International Chemicals Identifiers (InChIKeys).
- Populate the next input field to the right, "Smiles Structures for similarity search," by drawing a structure or entering a SMILES string.
    - To draw a structure, click the "Draw" button to open a drawing window. A dialog box will open; click the tools in the margins of the window to build a chemical structure. Any invalid component of an input structure will be underlined in red. When you are finished drawing, click "Close." Your drawn chemical structure will be converted to a SMILES string that will appear in the input box below the "Draw" button.
    - You can also manually enter a SMILES string in this field by clicking the "Enter" button. A dialog box will open and let you manually enter a SMILES string. When you are finished with your string, click "Close" and your SMILES string will appear in the input box.

- To edit a structure you have entered, click on the pencil icon to the right of the SMILES string.
- To remove a structure you have entered, click on the “X” icon to the right of the SMILES string.

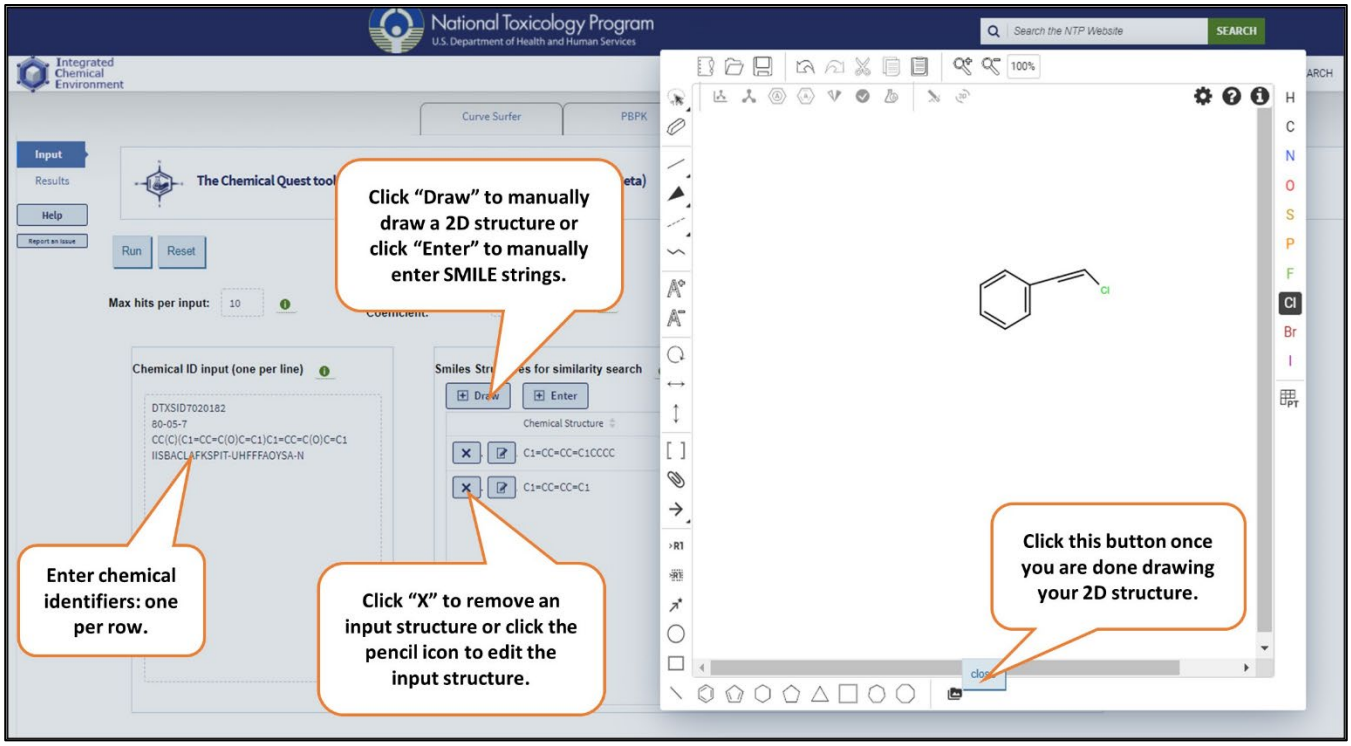

Figure 2. Input chemicals by typing chemical identifiers into the left-hand text box or manually draw and enter SMILES strings in the right-hand input box.

Limiting Queries

If you have selected “Search Custom Chemical List”, an input field for Custom Chemical Search Targets will appear on the far right of the screen. You have two options to populate this field with a list to query.

- Populate the left-hand "Quick List CASRNs" text box by selecting one or more ICE Chemical Quick Lists.
    - Click "Select Chemicals," which brings up a list of ICE Chemical Quick Lists.
    - Select the checkboxes in the dialog box to choose one or more chemical lists; click "Finished" to close the dialog box when you are done.
    - CASRNs from the selected ICE Chemical Quick Lists will populate the “Quick List CASRNs” text box.
- Populate the right-hand text box (User Chemical Identifiers) by entering your own list of chemical CASRNs or DTXSIDs (one per line).

All input boxes may contain some of the same chemicals.

Input fields above the chemical input fields (**Figure 3**) allow you to set criteria to limit your search results.

- "Max Hits per Input" allows you to limit the number of returned results per input chemical. Enter a whole number; default is 10. This feature is only available when querying the entire ICE database. Please note that the tool only returns the selected number of chemicals, even if more chemicals meet the Tanimoto criteria. To overcome this limitation, increase the Tanimoto Coefficient criteria.
- "Tanimoto Coefficient" allows you to specify a minimum Tanimoto score for your returned results. Enter a number between 0 and 1; default is 0.7.

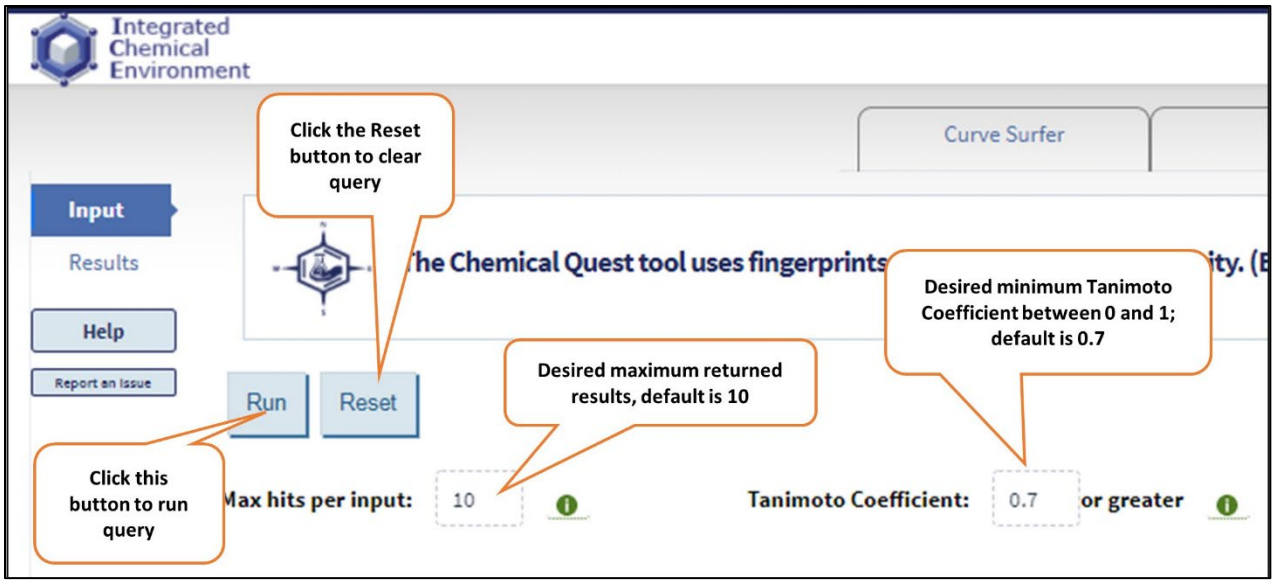

Figure 3. Options to limit search results.

Click the "Run" button at the top of the page to run your query. To reset your query parameters, click "Reset".

Viewing Chemical Quest Results

The window will switch to Results view. The Results view displays each query chemical in a clickable summary tile (**Figure 4**). Each tile displays the name, CASRN, DTXSID, and 2D chemical structure of a query chemical.

- Click on a chemical's CASRN to be directed to the available test article page about that chemical in the NTP Chemical Effects in Biological Systems (CEBS) database.
- Click on a chemical's DTXSID to be directed to the U.S. Environmental Protection Agency's CompTox Chemicals Dashboard entry for that chemical, which provides the chemical structure and other physicochemical and experimental properties.

In addition, each tile lists:

- Tanimoto: limiting criteria included in your query.

- Hit Count: number of returned results. If no similar chemicals are identified, then the count will be zero.
- Passed Filter(s): number of filtered results that can be sent to other ICE tools (see "Using Results to Query Other ICE Tools" below). This will always initially be the same as the total number of hits. You will filter items in the Results detailed view described below.
- Selected Item(s): number of individually selected results that will be sent to other ICE tools or downloaded (see "Using Results to Query Other ICE Tools" below). This number will always start at "0." You will select items in the Results detailed view described below.
- Figure 4. Screenshot shows summary tiles for two input chemical results with features described in the main text.

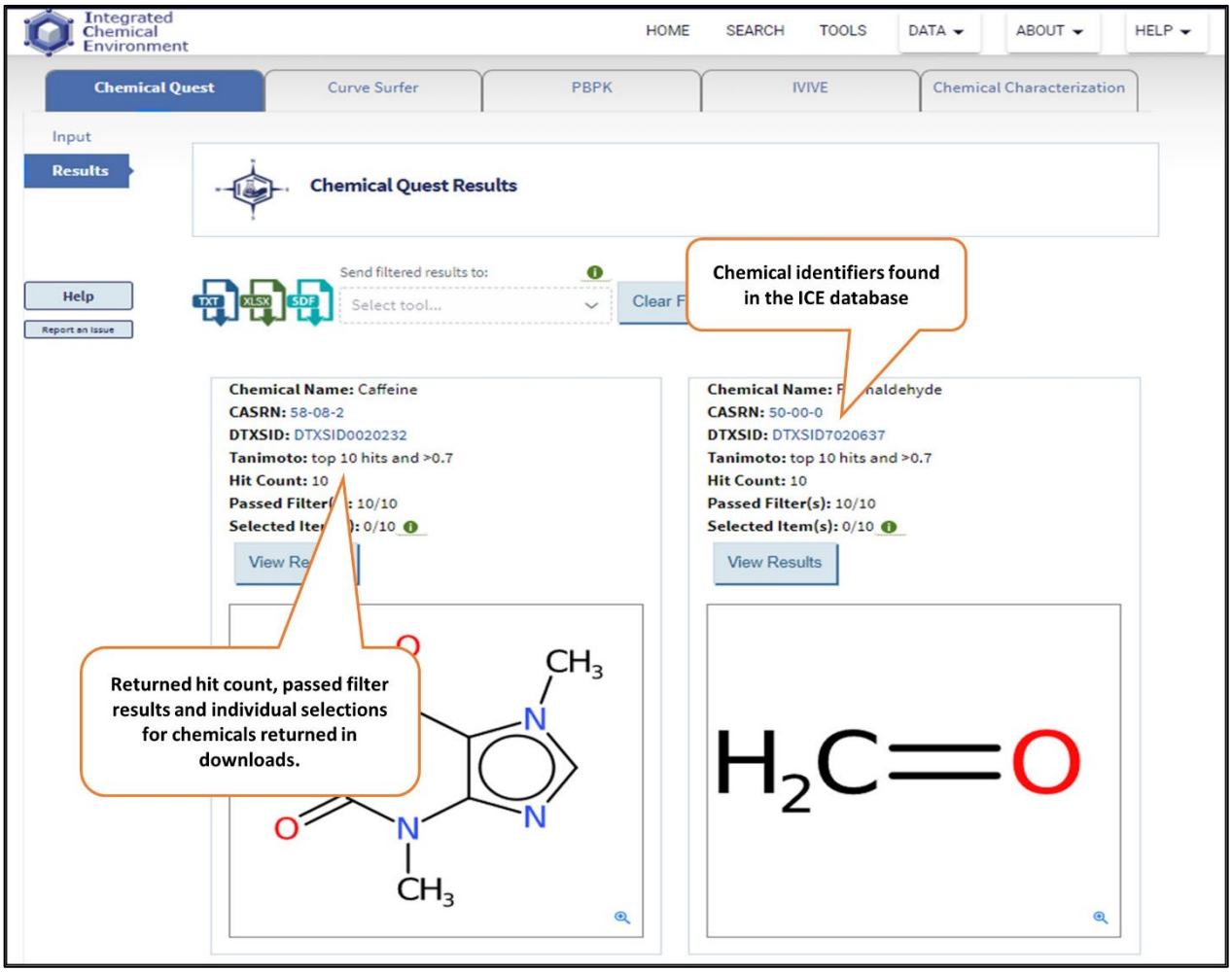

Figure 4. Example of input chemical summary tile displayed on the Results page.

Detailed Query Results

To view detailed query results for a chemical, click the "View Results" button in that chemical's tile. A dialog box will open (**Figure 5**) showing query results with the following information for each returned chemical:

- CASRN.

- DTXSID.
- Chemical name.
- Tanimoto value: "1.0" indicates that the chemical shares the same structural substructures, included in the fingerprint of the query chemical.
- Has Bioactivity: this indicates if there are bioactivity data available for this chemical in the Curated High-throughput Screening Data (cHTS).
- 2D chemical structure

Results may be on multiple pages depending on the number of hits. You can view the page count and select or browse pages using the arrows and dropdown list under "Select Page." Click the "Sort Results By" dropdown list to select from four sorting criteria, and "Direction" to display returned chemicals in ascending or descending order.

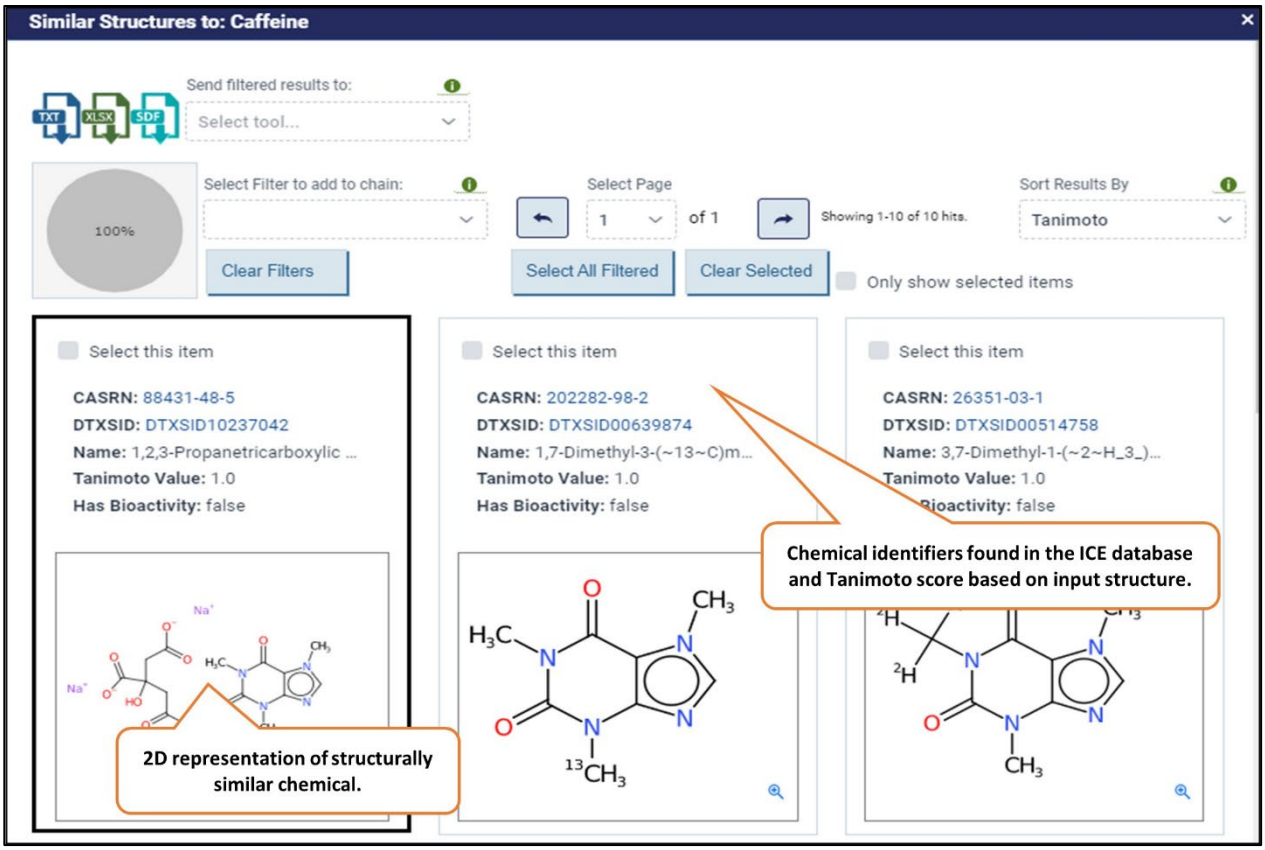

Figure 5. Returned results for a structurally similar chemical with specific chemical identifiers and Tanimoto score.

Selecting and Filtering in Detailed Results View

You can select chemicals of interest within the main detailed view by clicking on the "Select this item" box in the upper left corner of a chemical's tile. This will allow them to be downloaded (see below). To clear all selected results, click "Clear Selected." To view your selected results, select the "Only show selected items" box, next to "Clear Selected."

You can also use a variety of filter options to identify chemicals of interest. To apply a filter, select the

dropdown list under “Select Filter to add to chain.” Filters available are:

- CASRN.
- Chemical Name.
- DTXSID.
- SMARTS (SMILES arbitrary target specification; see below for more details).
- Has Bioactivity: chemical does or does not have bioactivity data in the cHTS.
- Tanimoto.

For each filter you select, a separate dialog box will appear allowing you to specify your filter criteria. Once you have finished click “Close.”

The SMARTS filter allows you to enter multiple SMARTS strings. SMARTS strings are case sensitive; see SMARTS Examples for assistance. Click “Add SMARTS Query” after entering each string. As you enter individual strings, the chemical structure diagrams displayed below the text box will change to show elements that match characters in the SMARTS string highlighted in yellow. The query string will also appear in a list under “Select SMART Query,” and the display to the right will change to show the proportion of the original chemical list represented by the filter. To clear a SMARTS string selection, click the “X” to the right of the listed string or uncheck the box. When done click “Apply Filter.” If you add several SMARTS queries to a filter, the resulting set of filtered chemicals will be the union of the chemicals filtered by each individual SMARTS query. If you desire to filter result chemicals using an intersection of several SMARTS, you will need to define a filter for each SMARTS.

Returning to the main detailed view (**Figure 6**), one or more colored boxes will now appear that display the filter(s) applied, the number of selected criteria, and the number of chemicals that passed or did not pass the filter. The pie chart to the right shows the number of chemicals meeting the criteria in each layer of filters.

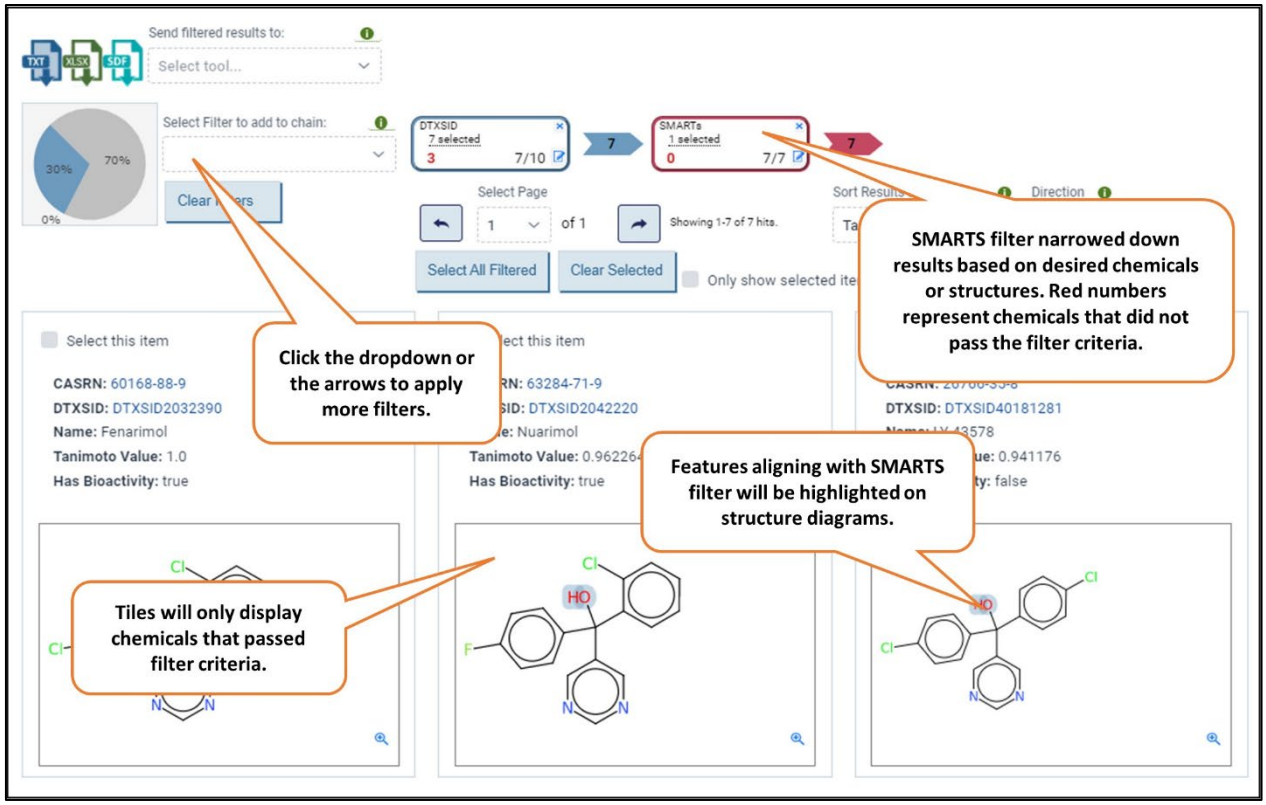

Figure 6. Detailed view display showing filters applied. In this example, a SMARTS filter was used to identify chemicals with an oxygen and sulfur so O and S features in structures are highlighted.

Note that applying filters will restrict the tiles displayed in the main detailed view to those chemicals passing any filters that have been applied. Therefore, you may wish to use the “select” option described above before applying any filters.

Click the “Select All Filtered” option if you want to send filtered chemicals to query other tools. After applying all desired filters and selections, click "Close" in the bottom right corner of the window to return to the main Results page.

Returning to the main Results view (**Figure 7**), the number of chemicals meeting the filter criteria will be noted as "Passed Filter(s)" and the number of chemicals that were selected will appear as "Selected Item(s)." To clear filters on an individual chemical, click the "View Results" button to return to the Results detailed view and click “Clear Filters.” Similarly, to clear selections click “Clear Selected” in the Results detailed view. To clear filters on all chemicals, you can click “Clear Filter” at the top of the main Results page. This will clear filters, but any chemicals selected using the “Select All Filtered” option above will remain selected.

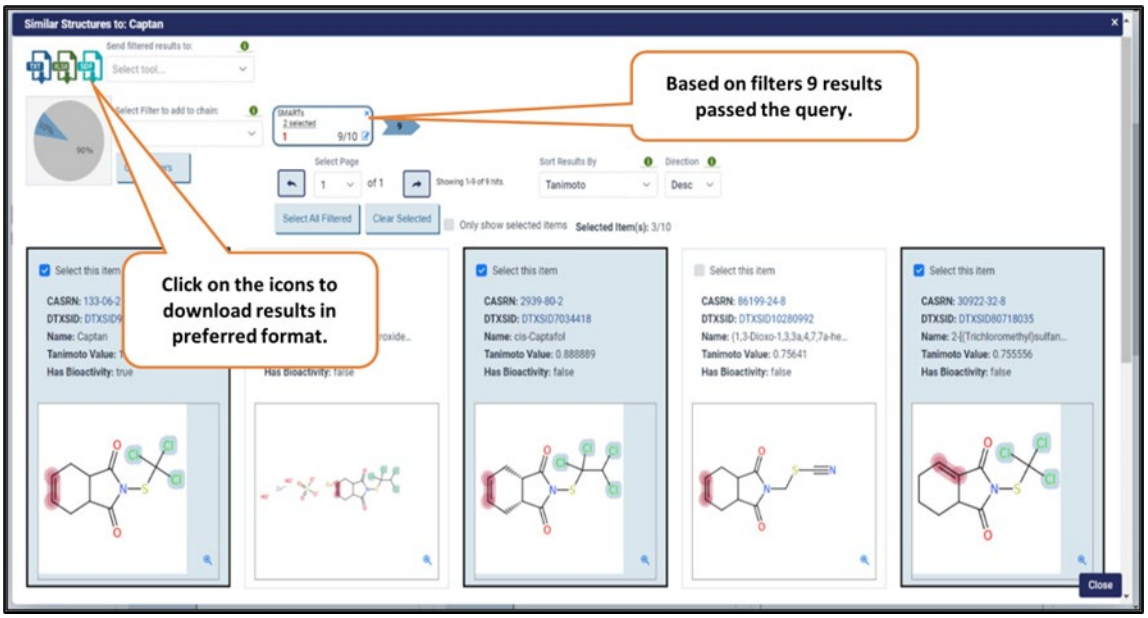

Figure 7. Tile on main Results page after applying filters; download tools shown at top of image.

Downloading Results

Query results can be downloaded using the download icons at the top of the main Results page (**Figure 7**).

- If any chemicals were selected among the returned results, only the selected chemicals will be sent to the download.
- If no chemicals were selected, all returned chemicals will be sent to the download.
- If filters were applied to the results for a particular chemical, only those chemicals that meet the filter criteria will be sent to the download for that chemical, but all chemicals returned for other chemicals will be sent.

There are three format options for downloading.

- Clicking the "TXT" icon will export your results to a tab-delimited text file. Data in this file are organized as one row for each “hit” returned by the query. They include descriptive information for the input chemical (name, SMILES, quantitative structure–activity relationship [QSAR]-ready SMILES, CASRN, and DTXSID) and the same information plus the Tanimoto score for the hit chemical.
- Clicking the “XLSX” icon will export your results to an Excel file. This file includes the same data in the same formatting as the tab-delimited text file described above.
- Clicking the “SDF” icon will export your results to a file in spatial data file (SDF) format. This file includes the structural information and chemical identifiers (DTXSID, QSAR-ready SMILES, CASRN, QSAR-ready InChiKey, name) of returned results. The file can be viewed through a text editor or SDF reader.

Using Results to Query Other ICE Tools

Use the “Send filtered results to” dropdown list next to the download icons to send your query results to other ICE tools (**Figure 8**). Any selections or filters applied to your results will be treated as for downloads (see above). 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 “Curve Surfer” to display activity curves for chemicals that have cHTS assay data in ICE.
- Click “PBPK” to send data to the ICE Physiologically Based Pharmacokinetics (PBPK) tool. This tool generates predictions of tissue-specific chemical concentration profiles following a dosing event.
- Click "IVIVE" to send data to the ICE In Vitro to In Vivo Extrapolation (IVIVE) tool. This tool allows you to estimate the daily equivalent administered dose that would result in the plasma concentration of a chemical equal to the activity concentration in a given in vitro assay.
- Click "Chem Characterization" to view physicochemical and absorbance, distribution, metabolism, and excretion (ADME) properties and chemical use categories for these chemicals.

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

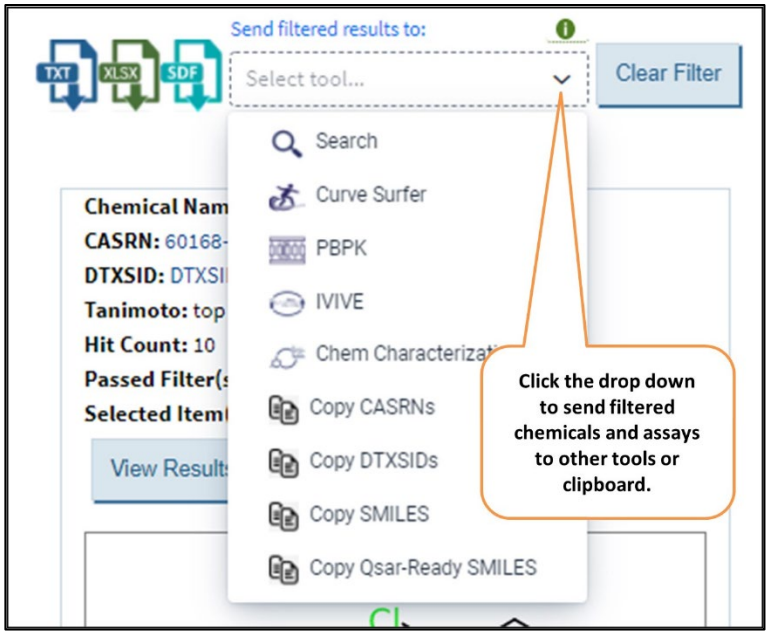

Figure 8. Options to send filtered results to run queries in other ICE tools.

Appendix 1: How Chemical Quest Finds Similar Chemicals

The Chemical Quest tool determines the structural similarity of chemicals through comparing molecular descriptors of query chemicals to target chemicals. For easy representation and statistical analysis, molecular descriptors can be described as groupings of binary codes (that is, consisting of 0s and 1s)

called fingerprints. A fingerprint represents the presence of a particular structural feature within a molecule such as double bond count or hydroxyl group. Molecular descriptors play a key role in characterizing a chemical structure through experimental measurements, such as octanol–water partition coefficient and other physical properties, as well as theoretical atom placement. Descriptors based on structural properties are calculated using different algorithms which result in descriptors that contain representations of feature counts such as rings and heavy atoms in addition to polar surface area and structural binding.

The Chemical Quest tool utilizes Saagar fingerprints (Sedykh et al. 2021), which identify 834 features. The features of Saagar fingerprints include chemical elements, structural classes, rings of various sizes and substituents and their positions on rings, separation of certain hetero atoms, chemical structures often used as scaffolds for endogenous biochemicals, and scaffolds present in common medicinal and industrial chemicals. More information about Saagar fingerprints can be found on the Sciome website.

Chemical Quest identifies structurally similar chemicals by assigning a Tanimoto score to a target chemical. Tanimoto scores are derived by comparing fingerprints between two chemicals and identifying shared molecular descriptors. Tanimoto scores range from 0 to 1, with higher scores representing more structurally similar chemicals. However, while a Tanimoto score of 1 can indicate an exact match, it will also be returned for chemicals that contain your input structure and additional substructures. For example, running a query on benzene will return benzene itself but also salts of benzene, structures containing different carbon and hydrogen isotopes, and structures such as 4-phenylazulene that contain a benzene ring and other structures.

Appendix 2: Abbreviations

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

CASRN: Chemical Abstracts Service Registry Number

DTXSID: DSSTox Substance identifier used by the U.S. Environmental Protection Agency CompTox Chemicals Dashboard

InChiKey: hashed International Chemical Identifier

IVIVE: in vitro to in vivo extrapolation

PBPK: physiologically based pharmacokinetics

QSAR: quantitative structure–activity relationship

SDF: spatial data file

SMARTS: SMILES arbitrary target specification; SMARTS differ from SMILES by specifying substructures instead of entire structures

SMILES: simplified molecular-input line-entry system

TXT: text file

XLSX: Excel file