Module 6: Comparing and Searching Chemical Entities

Questions

1. Conceptually, data in a database are stored in the same way as we would record them in a table or excel spreadsheet. The rows in the table correspond to compounds, and the columns correspond to properties or descriptions for those compounds (e.g., melting and boiling points, chemical names, toxicity, bioactivity, target proteins, and so on). These columns are commonly called "data fields". You may want to perform a search against all data fields or only a particular field. To search the chemical name field of the records in the PubChem Compound database, a chemical name query needs to be suffixed with either of the "[synonym]" or "[completesynonym]" index. The "[synonym]" index will search for molecules whose names contain the query chemical name as a part (that is, partial matching), and the "[completesynonym]" index will search for those whose names completely match the query (that is, exact matching). If no index is given after the query, PubChem will search all data fields.

   Go to the PubChem homepage (https://pubchem.ncbi.nlm.nih.gov) and select the "Compound" tab above the search box. Provide the following queries in the search box and click the "Go" button. How many hits do you get for each search? Clicking the image of each compound will direct you to the Compound Summary page of that compound, which provides comprehensive information on the compound. On the Compound Summary page of each compound, check the "Depositor-Supplied Synonyms" section to see if any of the chemical names of the molecule contain the string "zyrtec".

   (1) zyrtec
   (2) zyrtec[synonym]
   (3) zyrtec[completesynonym]

2. To perform an identity search for Cymbalta (CID 60835), go to the Chemical Structure Search page (https://pubchem.ncbi.nlm.nih.gov/search/search.cgi) and select the "Identity/Similarity" tab. Expand the "Options" section by clicking the "plus" button and select the "Identical Structures" with "same connectivity" from the drop-down menus. Expand the Filters section and limit the number of covalent units to 1 (by setting the range to "from 1 to 1"). Provide the query CID in the search box and run the search. Repeat the search with the "same isotopical labels" option selected. Explain how the two different options affect the identity search results.

3. Perform a 2-D similarity search using CID 5090 as a query. Select the "Identity/Similarity" tab and expand the Options sections by clicking the "plus" button next to the "Options" section heading. Select the "Similar Structures" and "95%" from the drop-down menus. Expand the Filters section and limit the number of covalent units to 1. Provide the CID query in the search box and press the "search" button. Repeat the search with the following similarity search threshold: 90%, 85%, and 80%. How many records are returned for each search?

The right column of the last search result page (for threshold >= 80%) shows what kind of information is available for the returned compounds. Click the "Pharmacological Actions" link under "BioMedical Annotation" to choose the compounds with the Pharmacological Action annotations. For each compound, check the information under the "Pharmacology and Biochemistry" section. What pharmacological actions do these compouns have?

4. Select the “3D Conformer” tab to perform a 3-D similarity search using CID 5090 as a query. Expand the Options section and select the “(Sort results by) Shape-then-feature” and “(output to) NCBI Entrez” options from the drop-down menus. Expand the Filters section and limit the covalent unit count to 1. Type the query CID in the search box and press the “search” button. How many compounds are returned? How many CIDs have pharmacological action annotations. Compare the results from 3-D similarity search with those from 2-D similarity search.