LICSS 2.2 Help

Contents

Overview

LICSS is a lightweight chemical spreadsheet for MS Excel for Windows easily capable of handling sheets with hundreds of thousands of molecules

You can enable Structure Display for a single Worksheet and a single associated Chart (embedded or a separate Chart sheet). You need to be connected to the network only if you don't already have the necessary JChemPaint visualisation and CDK files. Further Worksheet/Chart pairs can be enabled by running the Setup program again. If you want to enable a second chart associated with the same sheet, make sure that the 'Chart Only ' option is checked.

If you leave the Worksheet box blank then the Setup program will merely copy updated code modules to the Workbook.

On your worksheet you will also need to have a column containing Smiles strings for each chemical entry and a name "Smiles" referring to the full range (or column(s)) containing the Smiles strings. This Name needs to be local to the sheet. You can create or reset this name most easily by selecting the appropriate range and selecting <Ctrl Shift M>.

When enabled for structure display, worksheets will display structure when a cell containing a Smiles string is selected or, when a chart is active, when hovering over chart points. Selecting <Ctrl R> (or running the macro ShowRowStructure) will also display the structure for the Smiles string on the selected row (or the structure for selected Smiles String if the Smiles name is not set, even on a *non-enabled* worksheet). Selecting <Ctrl T> will toggle between showing all structures and none. If you want to use a column to provide a window title for the structure window, select <Ctrl Shift Z>. Selecting <Ctrl Shift S> will allow you to perform Substructure searches on the sheet; similarity searches can be performed with <Ctrl Shift I> When using the structure editor to draw substructures, the atom Xe will match any atom and the atom He will match any atom except H and C.

To Return a Smiles String from the Structure Edit window, select toolbar button: "Save Contents and Return to Application"

Selecting the LICSS Programs worksheet tab will display a menu of LICSS programs which can be run for enabled sheets; Selecting <Ctrl Shift R> will also display this menu. Alternatively you can use the shortcut keys below:

<Ctrl Shift E> will pick a diverse subset of compounds
<Ctrl Shift F> will calculate Fingerprints for each structure on the spreadsheet to speed up substructure searching
<Ctrl Shift T> will generate an RGroup Table for any enabled sheet
<Ctrl Shift C> will carry out Jarvis-Patrick Clustering for any enabled sheet
<Ctrl Shift A> will carry out Sammon Mapping for any enabled sheet
<Ctrl Shift N> will carry out Names to Smiles conversion
<Ctrl Shift P> will carry out Molecular Descriptor Calculation
<Ctrl Shift H> will output a window of all the shortcut keys described above

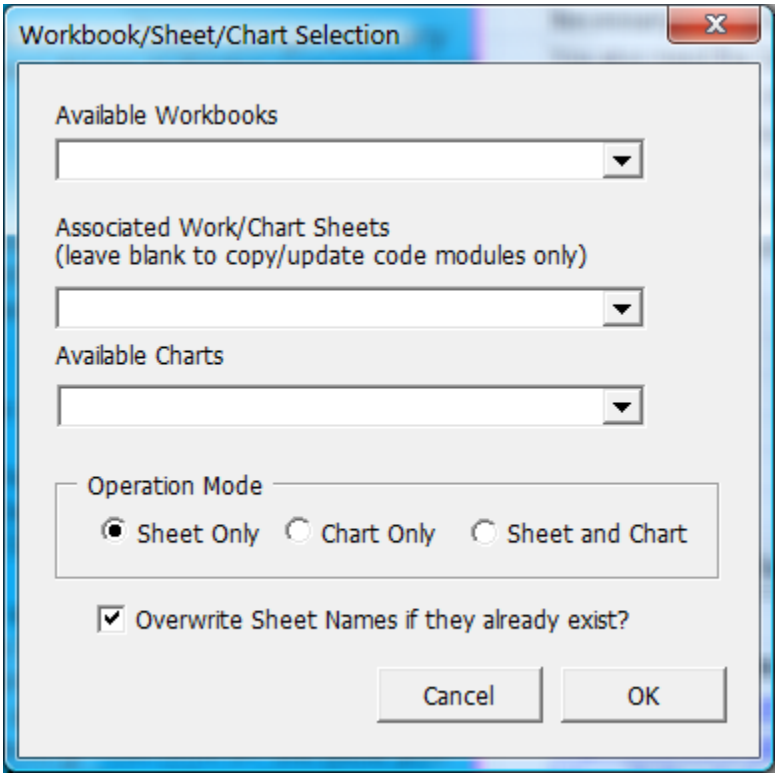
New Excel formulas: GetCDKDescriptor will calculate any CDK Molecular Descriptor or Mol Formula or Mol Weight, IsSubStructure and IsSimStructure will determine whether one Smiles string is a substructure of/is similar to another

Enabled worksheets (known as LICSS sheets) are fully independent of the creator Workbook or any AddIns

Return to Contents

Setting up Worksheets and Charts

To structure-enable workbooks, open the desired workbook and then open EnableChemicalSpreadsheetV2.2.xls. Select the main button: "Setup Workbook for Structure Display and Substructure Searching". If you have already used the LICSS 2.2 system you will be prompted immediately with the dialog box:



Alternatively, if you have a company LICSS setup, the necessary files will be downloaded automatically from your network. Finally, if yours is a 'stand-alone' installation, you will be prompted to download the necessary files from the internet. A 'yes' to this prompt will result in the program automatically downloading the necessary support files from the LICSS Google Project site (can be slow/fail depending on network speed/security settings). In any case, you will ultimately be presented with the dialog box above.

Select the workbook you want to enable in the 'Available Workbooks' dropdown and then choose the worksheet you want to structure-enable. You may also choose a chart (associated with this worksheet) from the third dropdown. On pressing OK, code is created for/copied to your workbook and a success message is displayed. At this stage you can close the EnableChemicalWorkbookV2.2.xls file. You should save your enabled Workbook and close/reopen it to activate the LICSS system.

The first time you select a new cell in your newly-enabled worksheet, you will get a warning message to say that you need to set the worksheet Smiles name. Select the entire column (or just the range) containing your Smiles strings and select <Ctrl Shift M>. This will set the Smiles name for you. You may also want to set the column from which the structure window titles are created: select <Ctrl Shift Z> and follow the prompt. You can reset the Smiles range as often as you like with <Ctrl Shift M>. This can be useful if you only want to operate on a subset of structures.

If you subsequently create new sheets/charts in an enabled workbook, these can be enabled by re-use of the EnableChemicalSpreadsheetV2.2.xls file: just choose the worksheet/chart from the appropriate dropdown. If you are enabling a chart only, choose the worksheet on which it is based (in the second dropdown) and the chart (in the third dropdown) but select **Chart Only** as operation mode.

You can update workbooks created with previous versions of LICSS by using EnableChemicalSpreadsheetV2.2.xls but filling in **only** the Workbook dropdown before pressing OK. A success message will reflect the updating.

Note that every time you run EnableChemicalSpreadsheetV2.2 on a workbook you need to save/close/reopen it to activate the LICSS system.

The Smiles strings needed for LICSS can be obtained from many chemically-aware programs. LICSS has the ability to convert chemical names to Smiles and also has a formula GetSmilesFromStructure() which allows you to draw a structure and return the Smiles for it.

Return to Contents

Importing and Exporting SDF Files

SDF Import

SDF Files may be imported from an enabled worksheet, either into the worksheet itself or a new worksheet.

On selecting Import SDF File from the main LICSS Programs Menu, you will be prompted for an SDF File to import. After file selection, the SDF File is converted to Smiles format by the program then imported either into a new worksheet (default) or into the worksheet from which the program was run. In the latter case, you will also be prompted for a cell for the top-left of the new datablock.

If you import into a new worksheet, you will probably wish to run the main enabling program (from EnableChemicalWorksheetV2.2.xls) again to enable the new sheet – alternatively, manually import the new data into an existing enabled worksheet.

When SDF Files are imported, the identifier becomes the first column, the Smiles data becomes the second column and all other data fields form the subsequent columns.

A common sequence is to chemically-enable a new worksheet then activate it followed by SDF File import into the worksheet. This is equivalent to creating a new LICSS spreadsheet from an SDF File. Before import of the SDF File you should name the 2nd column of the blank worksheet “Smiles” by selecting the entire column then <Ctrl-Shift-M>. This ensures that you can use the LICSS programs tab and that the new worksheet is ready for structure display after SDF File import (select cell A1 when prompted for a cell to import data into).

SDF Export

On selecting Export SDF File from the main LICSS Programs Menu, you will be prompted for an SDF File to export. After file selection, the program exports the data as an SDF file to the directory you have chosen. Data is chosen for export based on your range selection in the active sheet. If a single cell is chosen, the entire region around the active cell is used otherwise, the selection is used. In either case, the first row must contain column headings.

Return to Contents

Structure Display

Structures may be displayed for enabled worksheets by selecting the individual Smiles strings. A popup Java window shows the structure; this window may be resized and/or repositioned and will maintain the new settings in the current session.

For enabled charts, hovering over chart points will display the structure window as described above. Note that if the worksheet is filtered, for example by substructure, the chart will only display (by default) the filtered points and their corresponding structures if hovering over them.

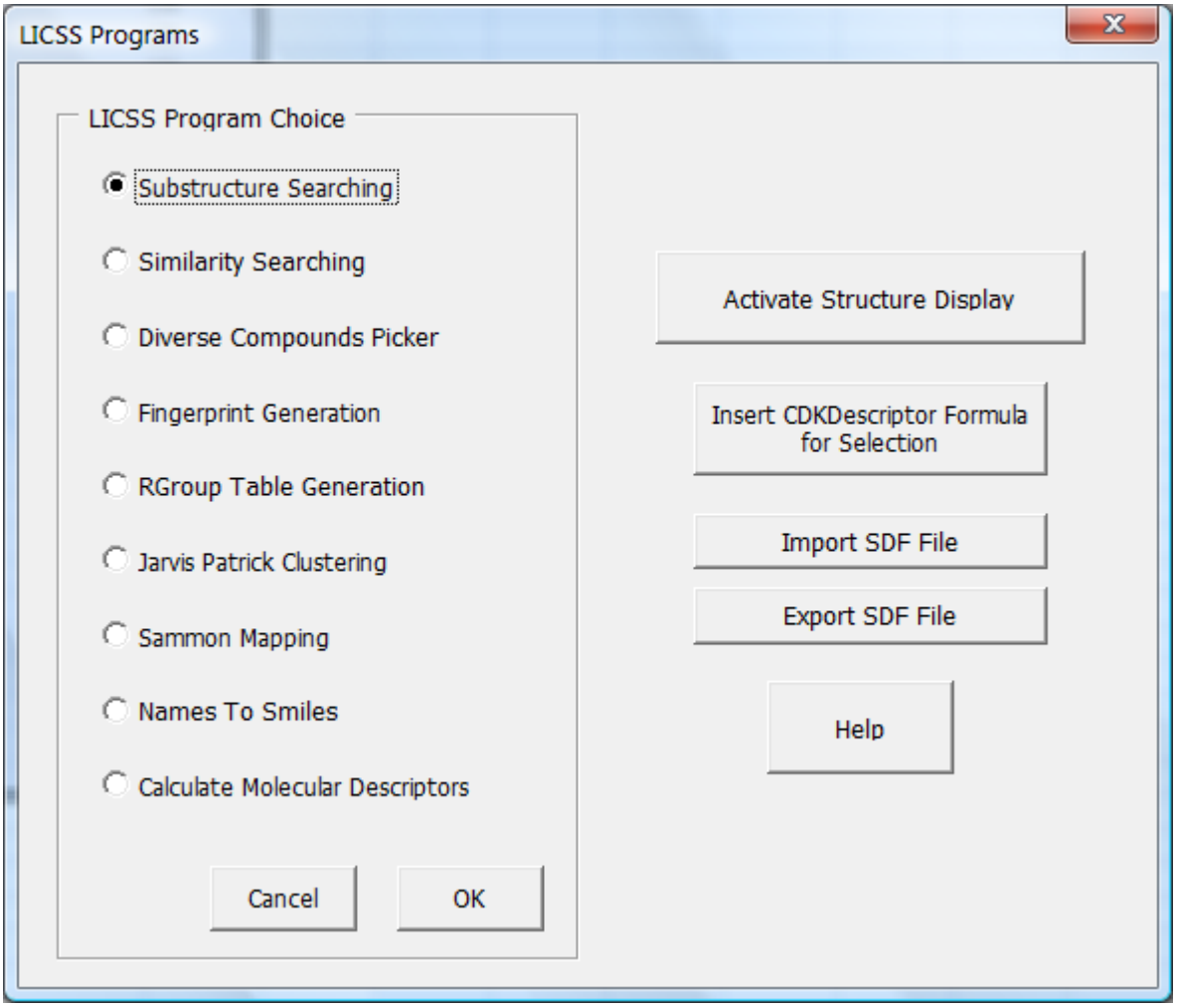
For non-enabled worksheets (but in enabled work**books**), selecting <Ctrl R> will display the structure window for any **selected** Smiles string. For enabled sheets, <Ctrl R> will display the structure for the Smiles string on that **row**.

<Ctrl T> will toggle all structure display on/off. In this case, the structures for all visible cells are displayed in a new column next to the Smiles column. Although it is possible to edit Excel whilst structures are displayed, there is a limit to what operations are consistent with successful structure display and, in general, you should toggle structure display off where possible. Thus, for example, you can filter a worksheet and then toggle structure display on (and structures will display correctly for filtered data). However, if you try to filter **whilst** structure display is on you will get some unpredictable results. In particular you should never leave a sheet whilst structure display is toggled on

Return to Contents

LICSS Programs Menu

Is displayed when the LICSS Programs Tab is selected from a LICSS-enabled worksheet (or by selecting <Ctrl Shift R>)

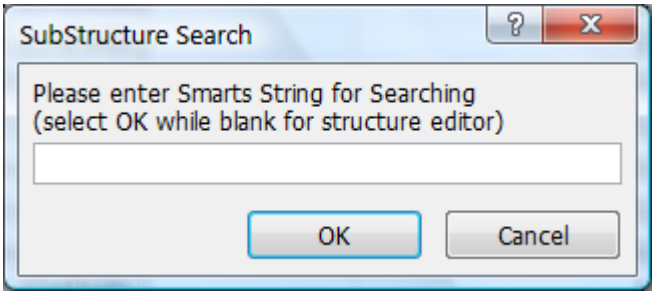


Select one of the options on the Left + OK to carry out an operation on your Smiles range (with the exception of Names to Smiles which works on the current selection). Alternatively, Enter a CDK Descriptor formula into the current selection or Activate Structure Display or display Help with one of the buttons on the Right

Return to Contents

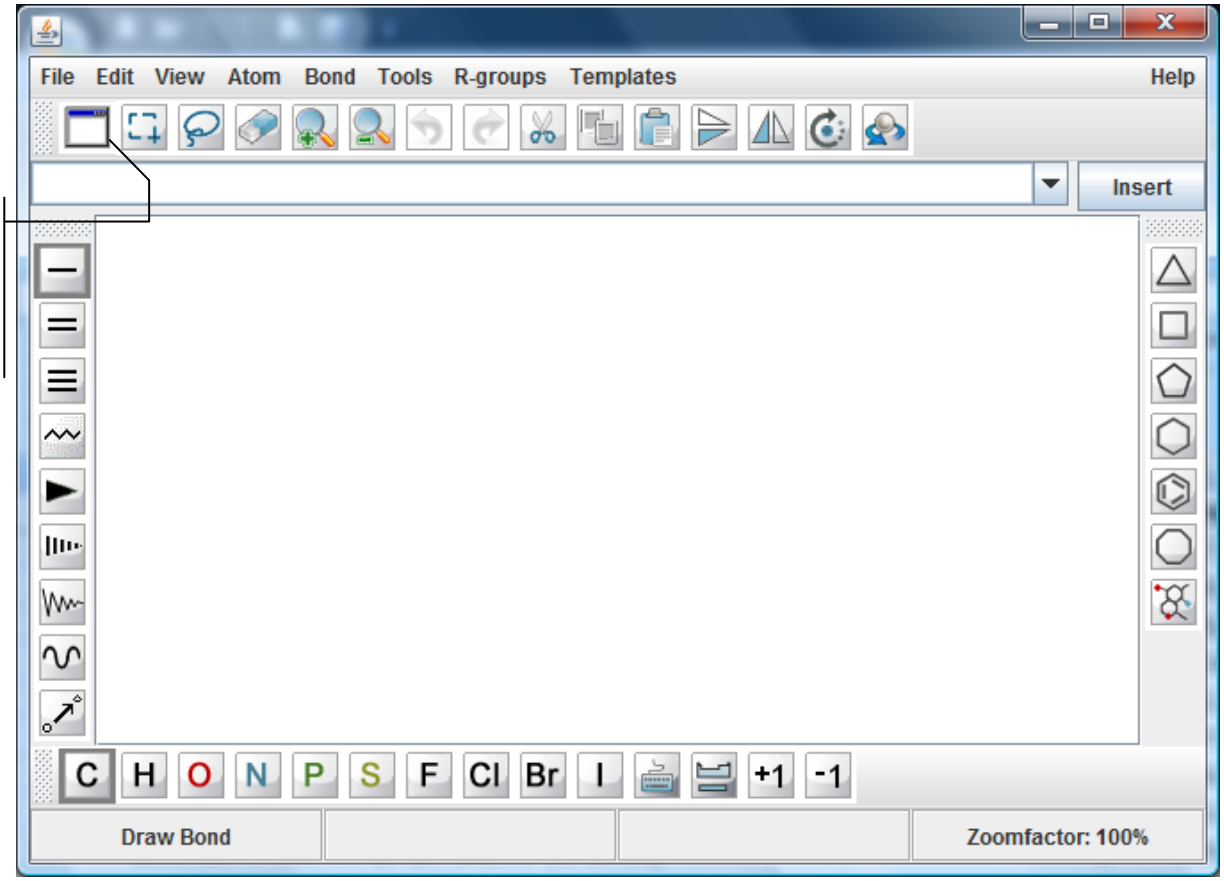
Substructure Searching

On selecting Substructure searching from the main LICSS Programs Menu, if you have previously calculated fingerprints for the compounds on the sheet, you will first be asked whether you want to carry out a fingerprint search. Select 'Yes' for big sheets with > 5000 compounds. You will then be presented with the dialog box:



You can enter Smiles or Smarts at this prompt (except when you are carrying out a fingerprint search first). Normally, however, just select OK to be shown the main J Chem Paint edit window:

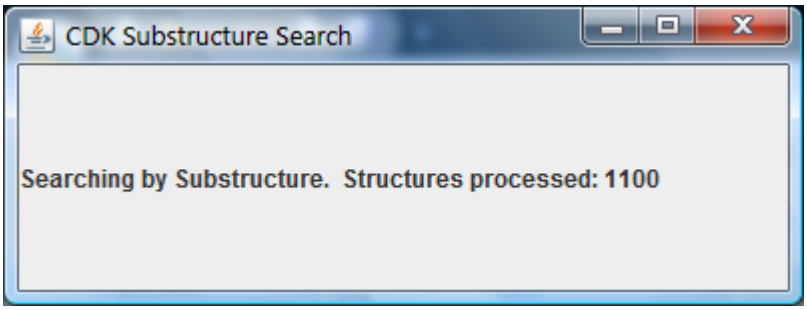
Select when finished drawing



You can then use all the drawing tools as usual. The Xe atom will match any atom and the He atom will match any except Hydrogen and Carbon. If you want to carry out a partial or exact search, put in explicit bonds to Hydrogen as required. You can hover over atoms to be changed from Carbon and just type the atom required (eg N/O).

When drawing is finished select the top-left tool button indicated (tooltip: Save Contents and Return to Application). If you have opted for a fingerprint search it will now be carried out (with

appropriate status bar info). A file of Smiles strings which need atom-by-atom matching is written out and a Java window will pop-up indicating the search progress:



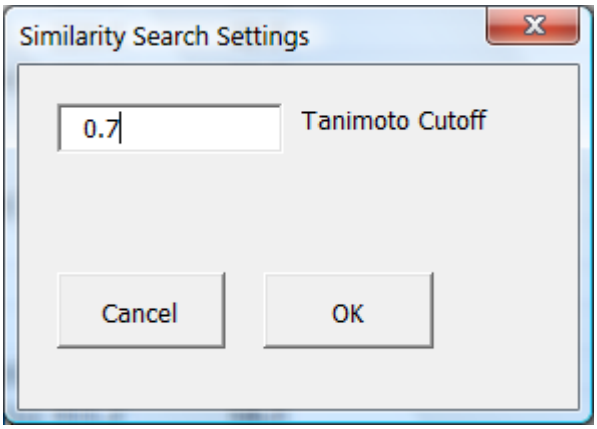
If you close this window the search is aborted. When complete, the hit file is read back into excel and a new column created for hits (1) or misses (0). A title cell is added: SSS: + the Smiles string used for searching. You can then filter on the search results using Excel's normal filtering facilities.

For sheets containing fingerprints, Substructure searching on several hundred thousand structures is eminently feasible in a short time.

Return to Contents

Similarity Searching

On selecting Similarity searching from the main LICSS Programs Menu you will be presented with the Similarity Search Settings dialog. Select the Tanimoto coefficient (between 0 and 1) you would like to use as a cut-off. CDK Fingerprints are such that you may wish to use a relatively low value between 0.5 and 0.7 to obtain many hits. On, Okaying this box, you will be presented with the J Chem Paint editor after which the Java program will run as with Substructure Searching. A new column is inserted to the right of your Smiles names with 0 (for miss), 1 (for hit) and a column title with the Smiles string searched for @ level of similarity chosen



Return to Contents

Name to Smiles Conversion

This program converts chemical names or identifiers to Smiles strings. On selecting this option you will be prompted for which names you want a conversion (default is current selection). A second prompt will ask you to choose between OPSIN or CIR for conversion. OPSIN is designed primarily for IUPAC names (which should exactly follow IUPAC rules including all spelling/capitalisation). CIR is a web service which will convert a variety of formats (including CAS numbers) to Smiles. This option will only work for compounds with entries in the CIR database. Note also that internet access is required for this to work.

Output Smiles strings are put in the Smiles column (if it exists) or (if not) the column to the right of the source names.

If you are using the CIR service and you need a proxy for internet access this should be entered as a Name in your workbook: proxy with =”proxy_address:proxy_port” as the “Refers to” value. This can also be entered as a default in EnableChemicalSpreadsheetV2.2.xls in which case it will be copied to all enabled workbooks.

Return to Contents

Fingerprint Generation

After a warning giving a time estimate for Fingerprint collection, the Java program to calculate fingerprints is launched with a progress window (as in Substructure Searching). Fingerprints are stored in a column to the left of the Smiles column. This program may also be conveniently used to calculate fingerprints for new compounds on an existing, enabled sheet – only the new compounds are calculated.

Fingerprints are stored as Hex strings corresponding to 512 bit fingerprints. They are used to speed up Substructure Searching (perhaps most useful on sheets > 2000 compounds)

Return to Contents

R Group Table Creation

This program analyses a compound set as a core structure with a substituent table. On selection, you will first be prompted whether or not you want to include compounds in which the core is not found. You will then be prompted to draw the core (as in Substructure Searching). Every position bearing implicit Hydrogen atoms is used as a potential site for substitution. If you want to block a position from substitution, draw an explicit bond to H.

Upon completion of the Java program (with status window as in Substructure Searching), a new sheet is created in the sheet with the Core molecule and substituent table. Substituent connection points are marked with a Xe atom. The whole sheet may, of course, be enabled by a further use of the EnableChemicalStructureV2.2 workbook. However, each Smiles string may be individually visualised without this activation by merely selecting it and choosing <Ctrl R>

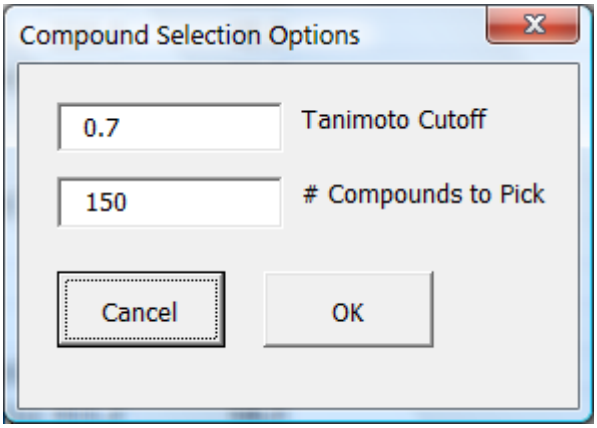
If you wish to calculate an RGroupTable for very large sheets it is worth breaking up the task in to batches of, say 50,000 structures.

Return to Contents

Diverse Compound Picker

This option is used to pick a diverse set from a bigger compound set.

On selecting this option you are first presented with the Compound Selection Dialog:



The # Compounds to Pick box defaults to approx 10% of the set (to the nearest 50). You can set this box to the number required. Note that although the set picked for a desired number of hits will be diverse, it will not necessarily be representative as the compounds are picked at random and may not fully sample the chemical space represented by the full set. If you want to pick the full set of compounds at a particular similarity threshold, enter 0 in this box.

The Tanimoto cut-off ensures that no two compounds picked will be more similar than represented by this value. Note that if a very low value is used (< 0.5), then it may not be possible to pick sufficient compounds.

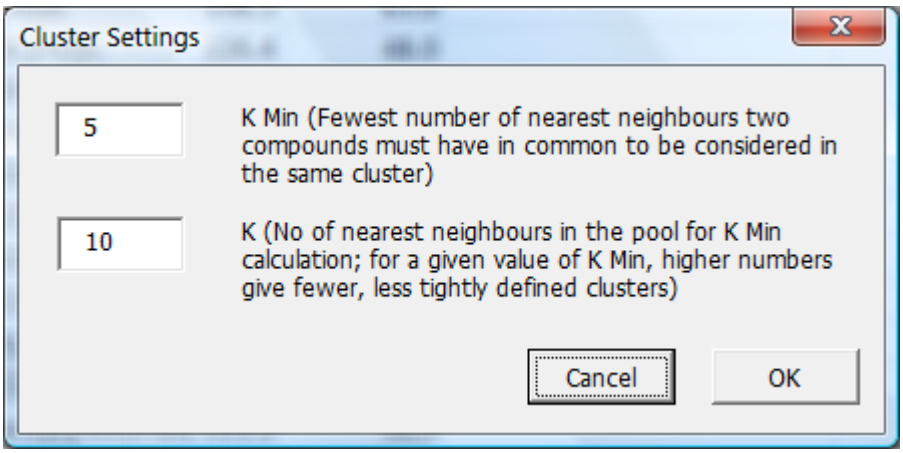
Once the OK button is pressed, the Smiles strings are written out to file and the Java program launched with a status window as in Substructure Searching giving the number of compounds which have been tried.

A Column title is produced showing the number of compounds chosen for picking and the similarity threshold. If it was not possible to pick sufficient compounds (either because 0 was used a number to pick or because the similarity threshold is too low for the number chosen), then the column title will reflect this by prefacing the number actually picked with the text: “Max Pick:”

Return to Contents

Jarvis Patrick Clustering

On selecting this option you will be presented with the Cluster Settings dialog box:



The default settings don't normally need changing. If you want fewer clusters, increase K.

Selecting OK launches the Java program with a status window similar to that for Substructure Searching. Fingerprints are first calculated then the clustering algorithm runs. A full table of inter-compound distances must be calculated and the algorithm scales with N^2. There is therefore a practical limit of *ca* 5000 compounds for clustering using this algorithm (depending on machine memory available)

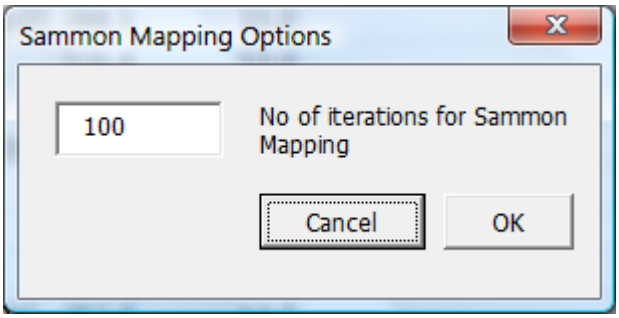
Once clustering is complete, a new sheet is created with all the clusters described. A new column is also inserted into the original sheet describing which cluster each compound belongs to.

Return to Contents

Sammon Mapping

Sammon mapping is a technique for projecting fingerprint multi-dimensional space into two dimensions. The results can be displayed using Excel's scatter plot facility whence similar compounds cluster together.

On choosing this menu option, you will first be presented with the Sammon Options dialog:



The default value is normally fine. As with Jarvis Patrick clustering, Fingerprints are first calculated and then the Sammon Mapping algorithm runs with a practical limit of about 5000 compounds. As in Substructure Searching a Java status window pops up displaying calculation progress.

Results are presented in Excel as Sammon X and Sammon Y coordinate columns.

Return to Contents

Keyboard Shortcuts

<Ctrl R> - display Row structure (for enabled sheets)/Structure for selected cells (non-enabled sheet in enabled workbook)

<Ctrl Shift Z> - select a column for title of structure display window

<Ctrl Shift S> - carry out substructure searching

<Ctrl Shift I> - carry out similarity searching

<Ctrl Shift R> - display the Run Menu

<Ctrl T> - toggle display of all structures on/off

<Ctrl Shift E> - pick a diverse subset of compounds

<Ctrl Shift F> - calculate Fingerprints for each structure on the spreadsheet to speed up substructure searching

<Ctrl Shift T> - generate an RGroup Table for any enabled sheet

<Ctrl Shift C> - carry out Jarvis-Patrick Clustering for any enabled sheet

<Ctrl Shift A> - carry out Sammon Mapping for any enabled sheet

<Ctrl Shift N> - carry out Names to Smiles conversion

<Ctrl Shift P> - carry out Molecular Descriptors calculation

<Ctrl Shift H> - output a window of all the shortcut keys described above

Return to Contents

Calculate Molecular Descriptors

This option leads to a prompt for which CDK Descriptor to calculate and this is followed by writing of Smiles strings to file and activation of the Java program to calculate the descriptors. As with Substructure Searching, a Java popup window informs you of progress.

The results of all descriptors are returned in a new column as text, preceded with the name of each descriptor type (some types return several values). The text can be removed by excel using search and replace if the raw data is required (or a Molecular Descriptor Formula can be used)

Return to Contents

Insert Molecular Descriptor Formula (and other new Excel formulas)

This menu option enters a formula: GetCDKDescriptor in to the current selected cell. You can also enter the formula manually, but the menu looks after the parameter syntax for you. Note that some descriptors (such as ALogP) return arrays rather than single values; the last formula parameter is the array index you want to return – if entered *via* the menu, this is assumed to be 1.

There are also formulas for IsSubStructure(smiles) and IsSimStructure(smiles, cutoff_value), each returning 1or 0 for hits/misses respectively.

A formula ReformatSmiles(Smiles, kekule, expHs) is available which takes a Smiles string as first parameter and reformats it according to the two remaining parameters. Kekule (true/false) determines whether you want kekulised structures or aromatic Smiles (C1=CC=CC=C1 vs c1ccccc1). ExpHs determines whether you want Smiles returned with explicit hydrogens recorded.

If you want to sketch a structure and find out the equivalent Smiles string enter the formula: GetSmilesFromStructure(). A structure edit window will pop up and, on selecting the transfer button (top left) the Smiles string equivalent to the structure drawn will be shown

Return to Contents