
# Fantastic structures and how to find them

## Searching crystal structure data in the CSD with ConQuest and WebCSD

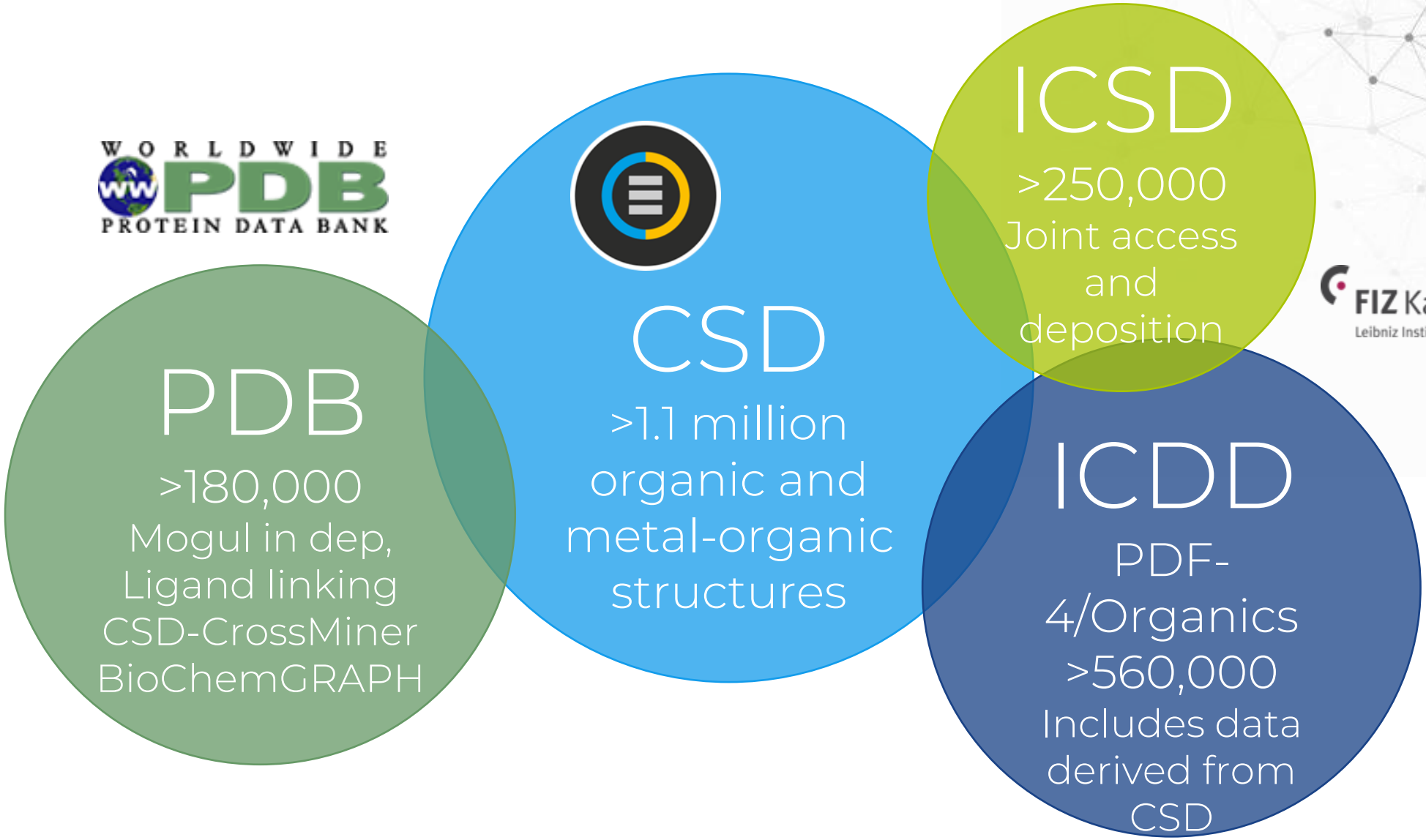
May 2022

# Learning outcomes for today

- How to search over a million published crystal structures using our desktop software ConQuest.
- How to refine your results in ConQuest to structures in targeted subsets allowing you to explore drugs, pesticides, MOFs and more in more detail.
- Tips and tricks to make your ConQuest searches more effective.
- Basics of searching using WebCSD, our web browser search component.
- Structure similarity searches in WebCSD.

CCDC

# Exploring structural databases

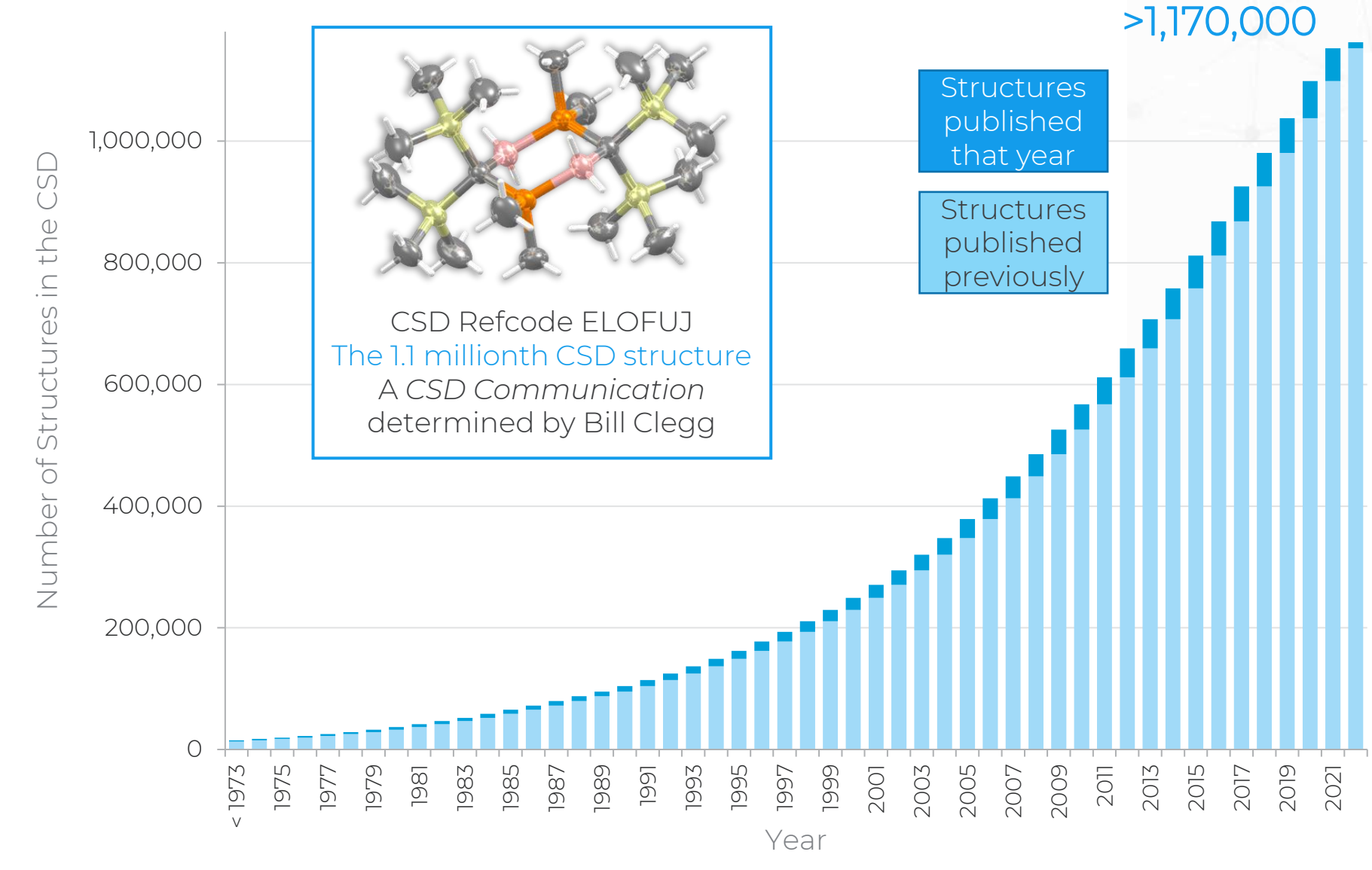



FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

ICDD

CCDC

# The Cambridge Structural Database

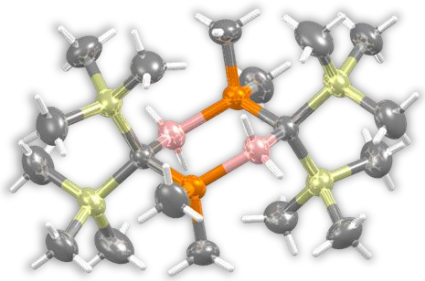



- Every published structure
  - ASAP & early view
  - CSD Communications
  - Patents
  - University repositories
  - Theses
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for over 56 years
- A CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

CCDC

# CSD Refcodes



CSD Refcode - ELOFUJ

What is ELOFUJ?

- A CSD Refcode
- A database reference code
- Containing 6-8 characters
- Used to identify entries in the CSD

## Refcode families

- The same substances are assigned the *same* 6 letter code plus an additional 2 numbers
  - Polymorphs
  - New determinations or re-refinements of the same substance
  - Determinations at different temperatures/pressures
- Stereoisomers or different solvates, co-crystals, etc are assigned *different* refcode families

CCDC

# Inside the Cambridge Structural Database

The CSD is a database of all the published organic and metal-organic experimental crystal structures

Organic
43%

Metal-Organic
57%

At least one transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

Not Polymeric
89%

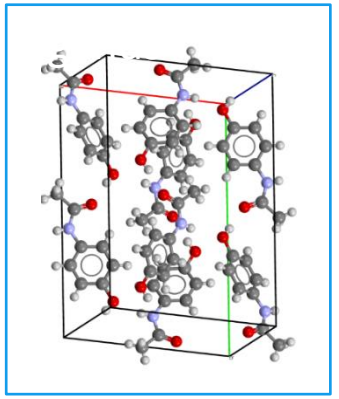
Polymeric: 11%

Single Component
56%

Multi Component
44%

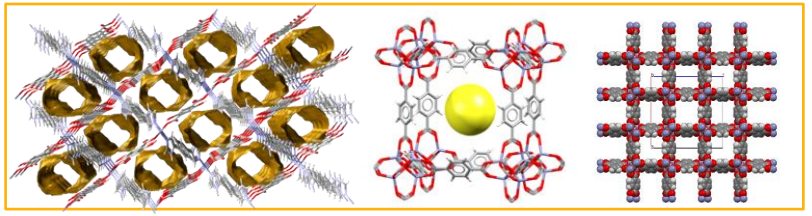
## Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



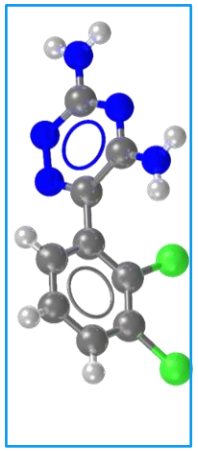
## Metal-Organic

- Metal Organic Frameworks
- Models for new catalysts
- Porous frameworks for gas storage
- Fundamental chemical bonding



## Additional data

- 11,525 polymorph families
- 171,683 melting points
- 909,992 crystal colours
- 778,663 crystal shapes
- 24,916 bioactivity details
- 11,387 natural source data
- > 250,000 oxidation states

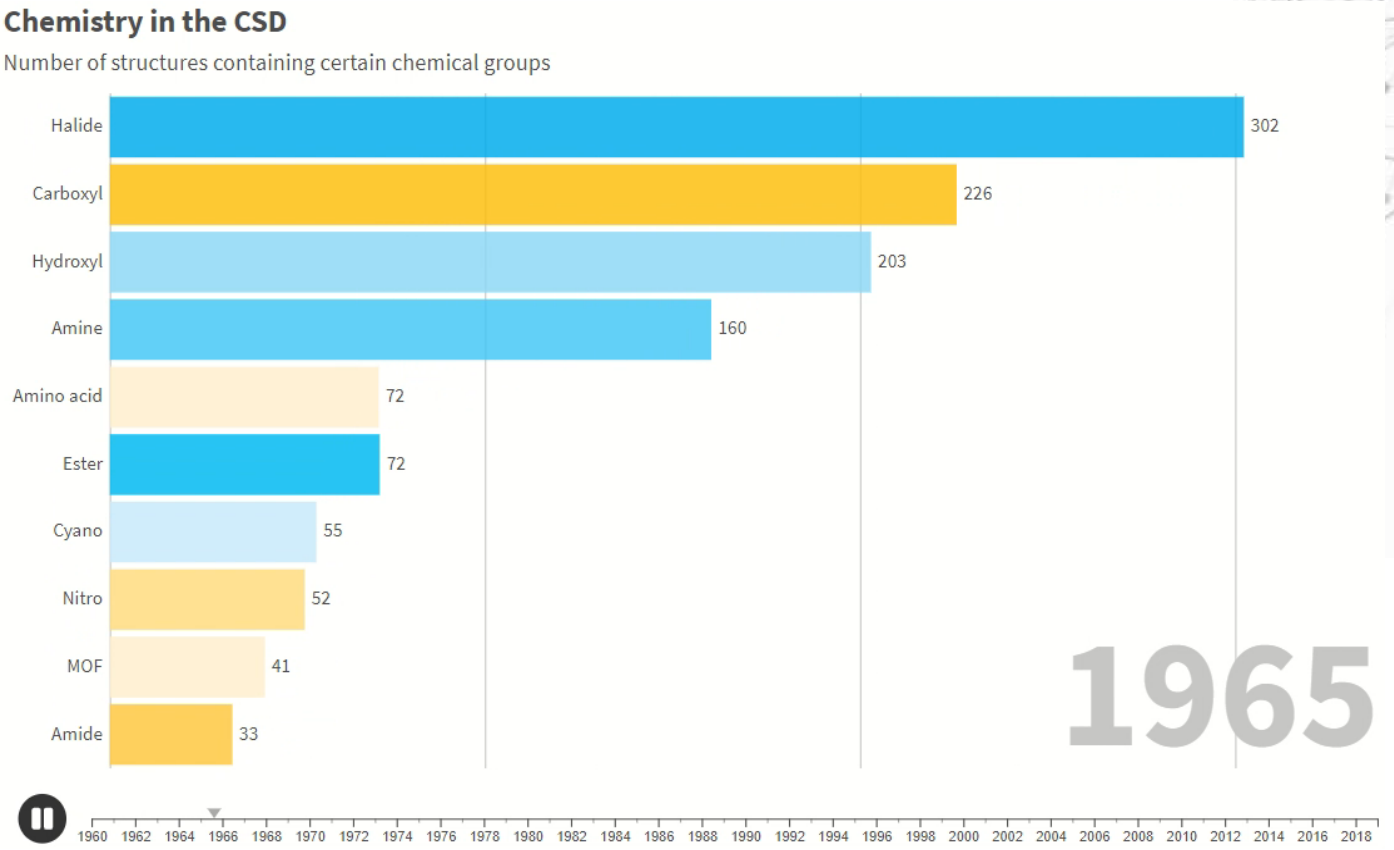


## Links and subsets

- DrugBank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticide PDB

# Exploring the CSD

- > 1 million structures
  - >94M 3D coordinates
- > 28 million bond lengths
  - >2M unique distributions
- > 40 million valence angles
  - >3M unique distributions
- > 14 million torsion angles
  - >800K unique distributions
- > 2 million rings
  - >400K unique distributions




Images and graphics created using the CSD Python API and Flourish

CCDC

# A wealth of structures, stories and possibilities

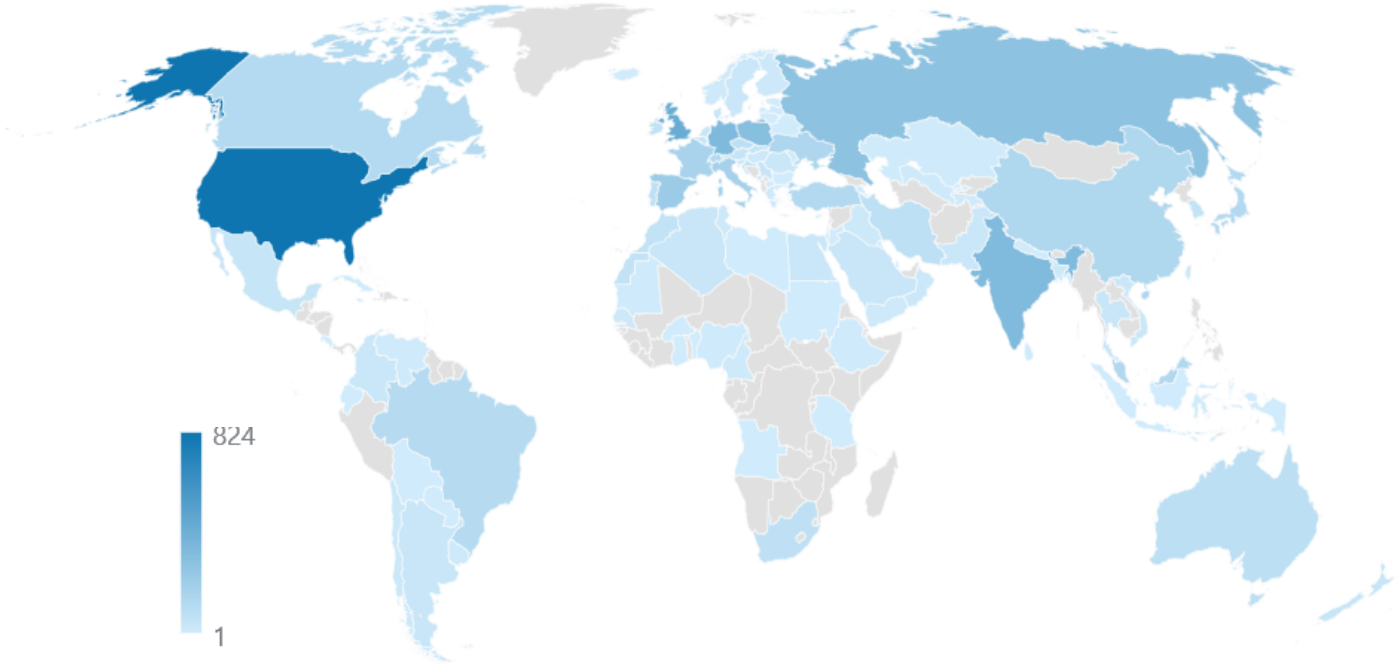
CSD one million special issue

- 33 articles from industry and academia


**1 million structures.**
**stories.**
**possibilities.**

Cambridge Structural Database, CSD

CCDC
advancing structural science



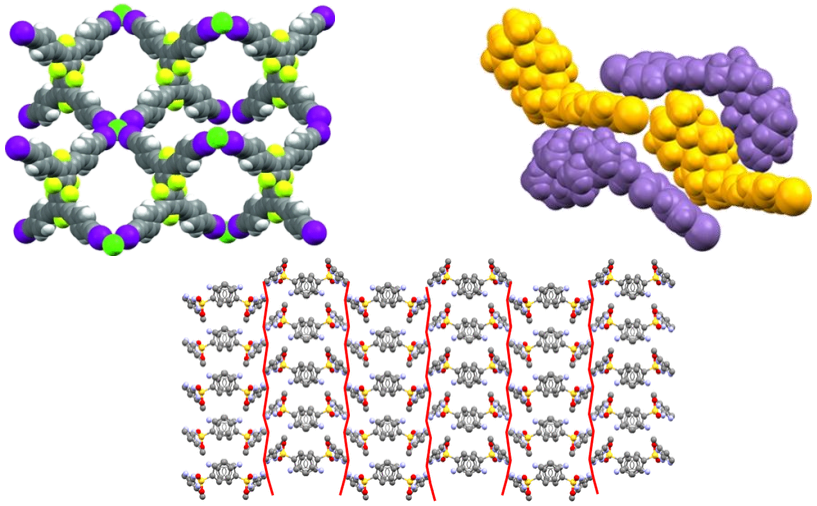

The CSD - A wealth of knowledge gained from a million structures. *CrystEngComm*, 2020,22, 7131-77502
The CSD. C.R.Groom, I.J.Bruno, M.P.Lightfoot and S.C.Ward, *Acta Cryst.* (2016). B72, 171-179 DOI: 10.1107/S2052520616003954

CCDC

# The whole is greater than the sum of its parts

## CSD Insights

- Molecular geometries
- Molecular interactions
- Molecular assemblies



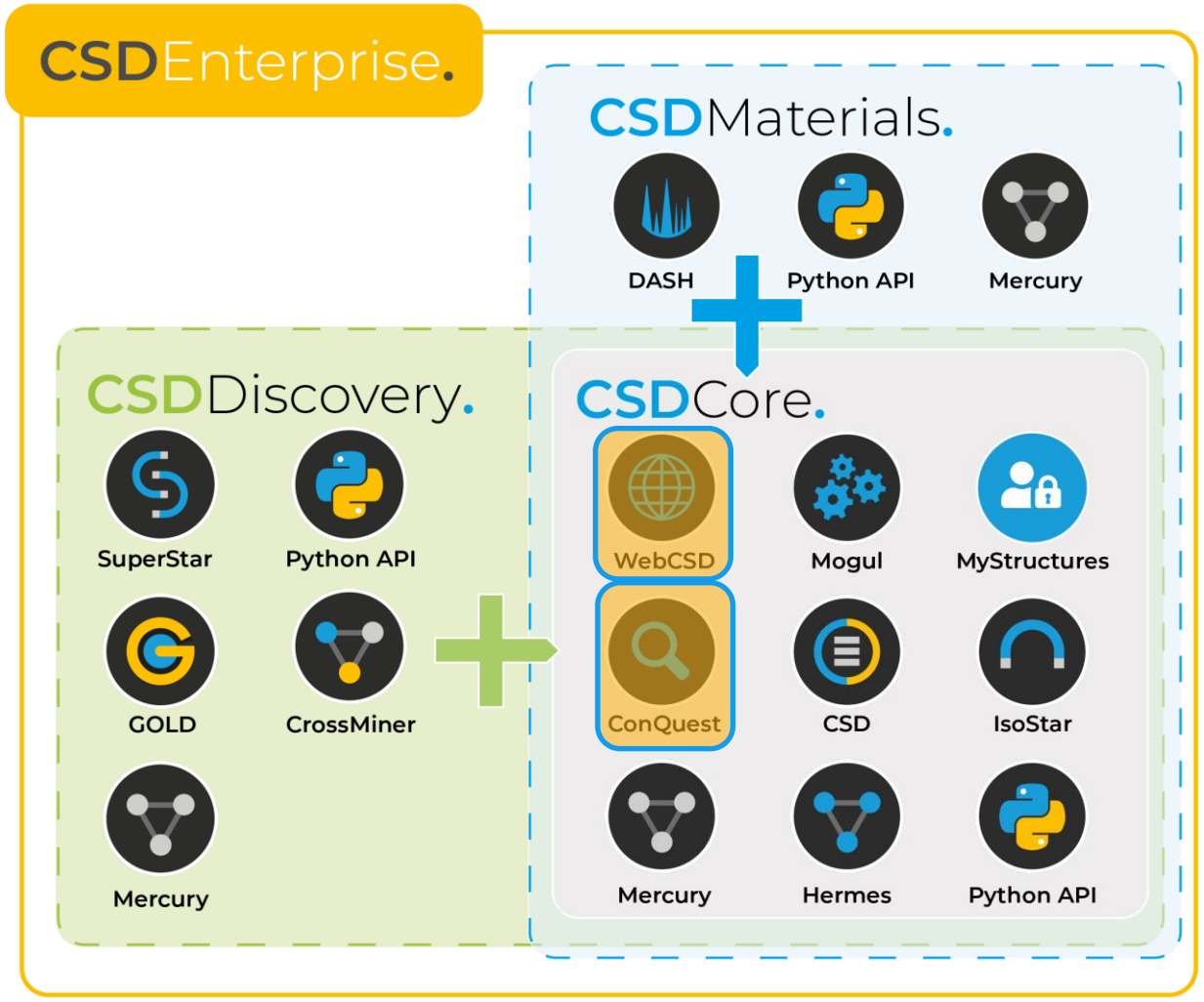
## CSD Applications

- Teaching
- Drug Design
- Materials Design
- Structure Solution
- Energetic materials
- Paints, Pigments and Dyes
- Organic Semiconductors
- Nonlinear Optical Materials
- Catalysts
- Gas Storage and Separation
- Crystal Engineering

https://doi.org/10.1021/acs.chemrev.9b00155

CCDC

# The CSD Portfolio






CCDC

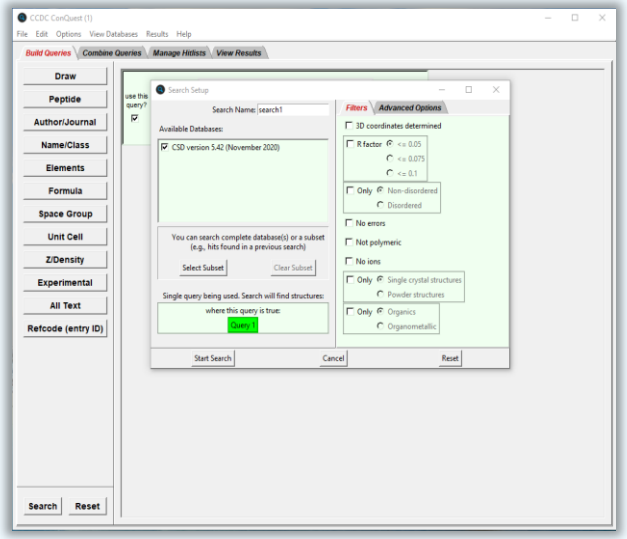
# CSD-Core – What does it enable?

- It makes it easy to answer very specific structure questions on your desktop using highly flexible 3D searching of the CSD
- Quickly identify relevant structures across the CSD based on a wide range of properties including chemical constraints such as cyclicity
- Use interaction and structure-property knowledge gained to drive design decisions

CCDC

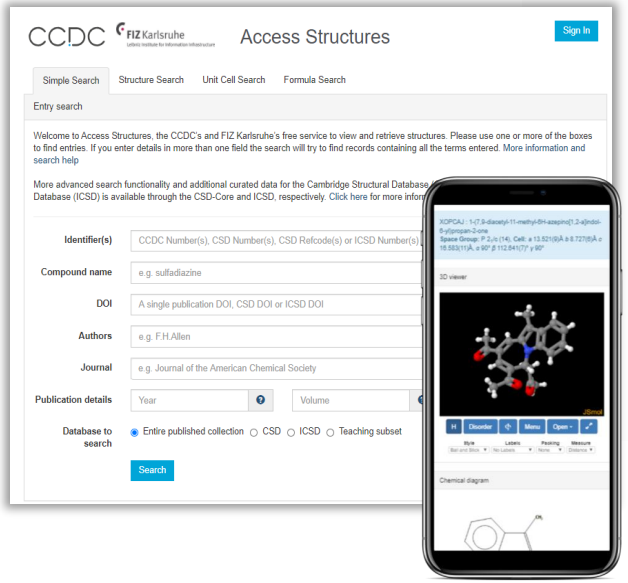
# Finding fantastic structures

## Desktop



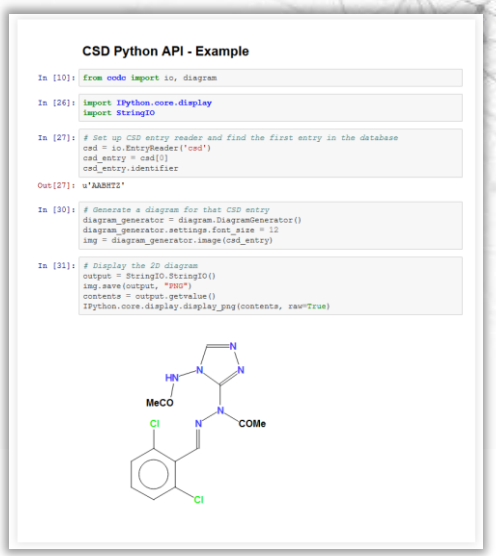

ConQuest

## Web Browser




WebCSD / Access Structures

## Programmatic

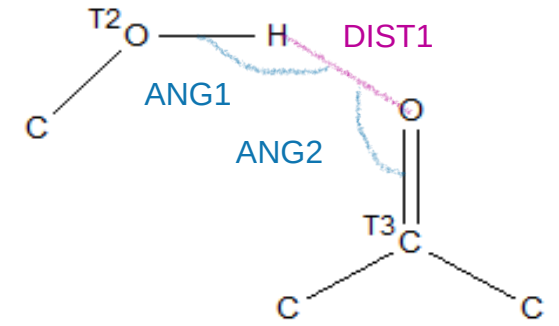


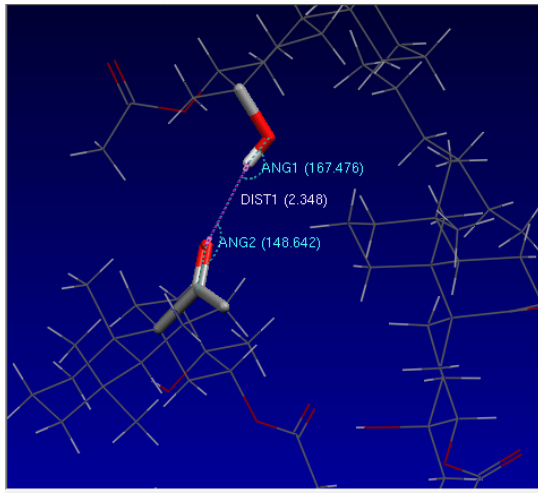

CSD Python API

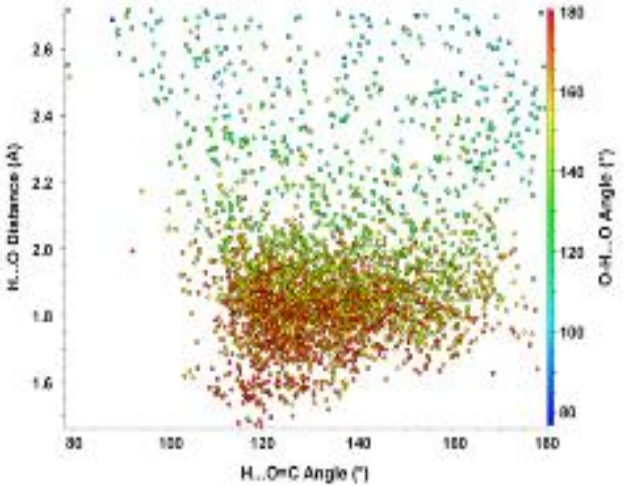
CCDC

# What is ConQuest?

- Enables search and retrieval of information from the CSD
- Provides full range of text / numeric database search options
- More complex search functionality includes:
  - Chemical structure searching
  - 3D Geometrical searching
  - Intermolecular non-bonded contact searching



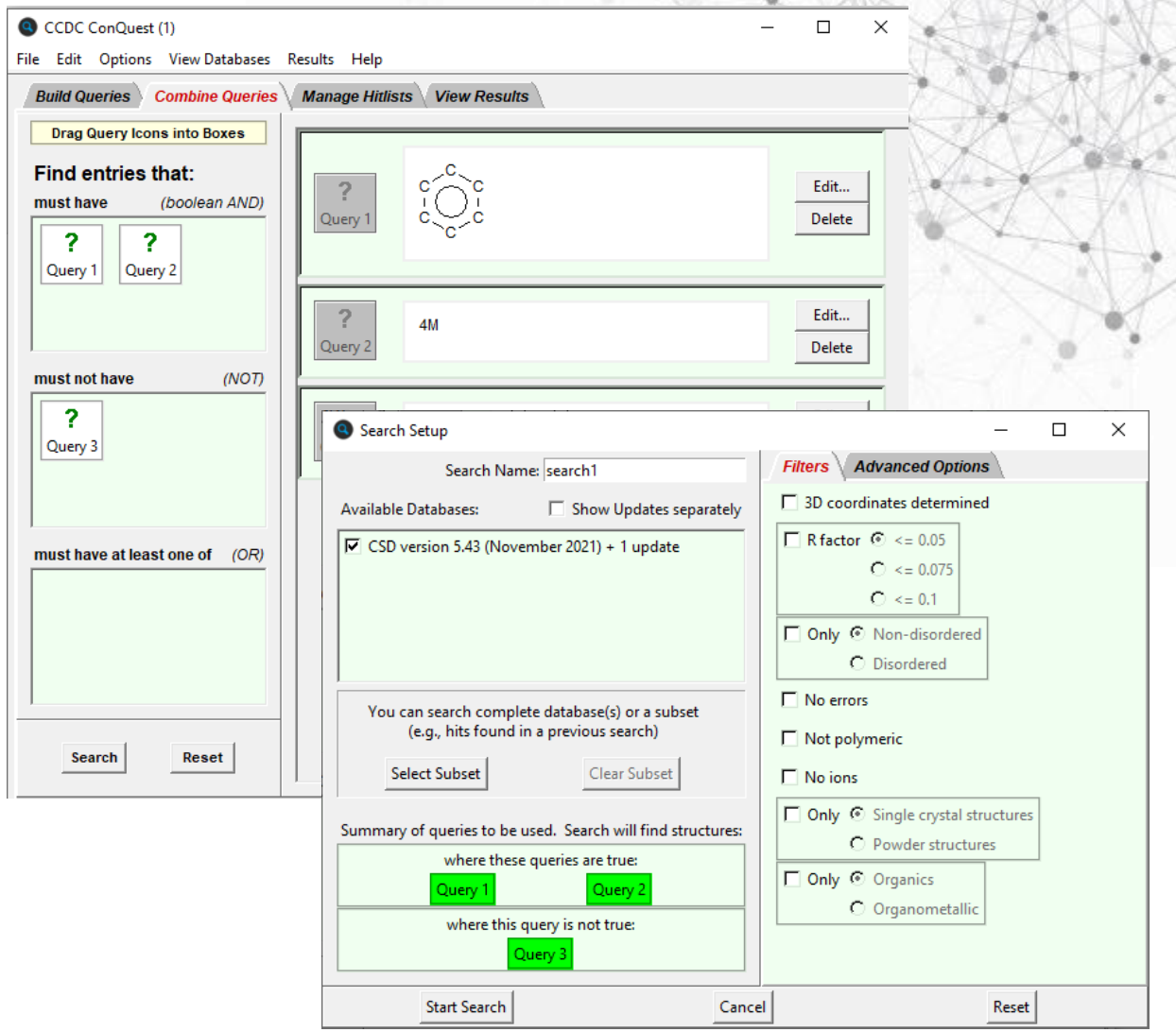



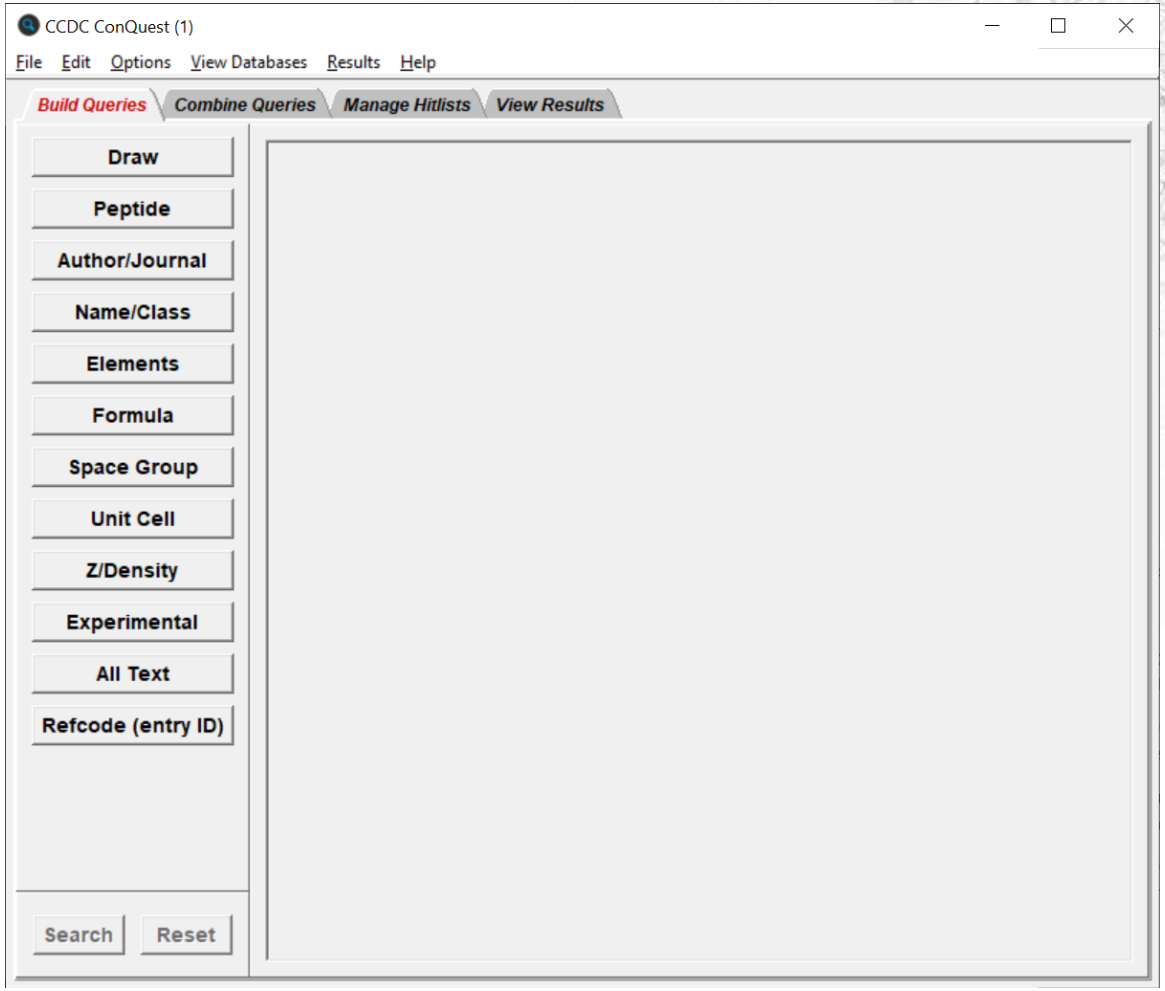

CCDC

# ConQuest searches

- **Filtering options** based on experimental or chemical considerations (*e.g.*, *R*-factor, temperature, only organics)
- **Combine search terms** with Boolean options and manage hitlists post-search
- **Export results** to Mercury for visualisation and advanced numerical analysis & plotting



# *Show One*: ConQuest – Opening and search options



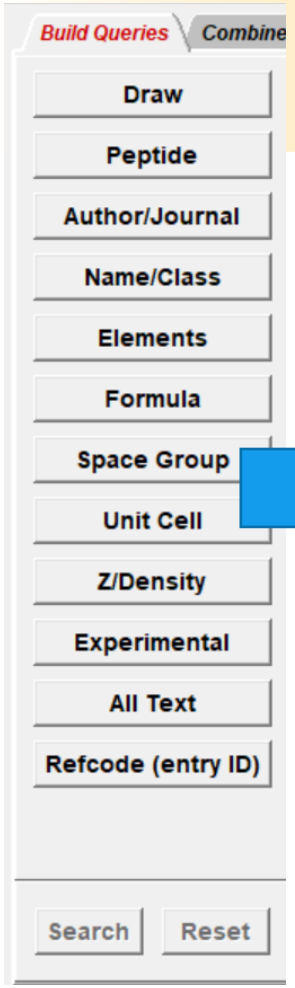

CCDC

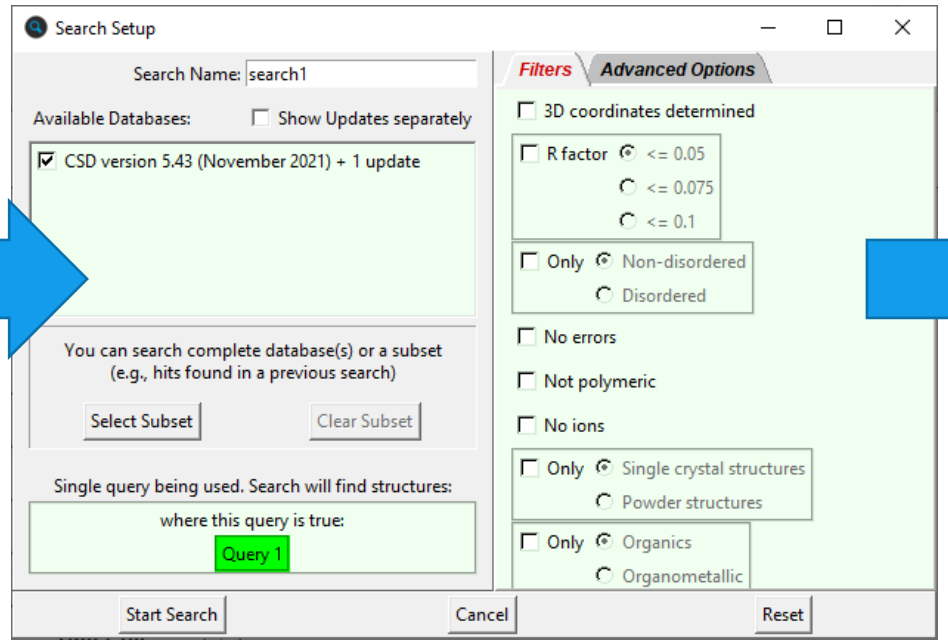
# How to search in ConQuest

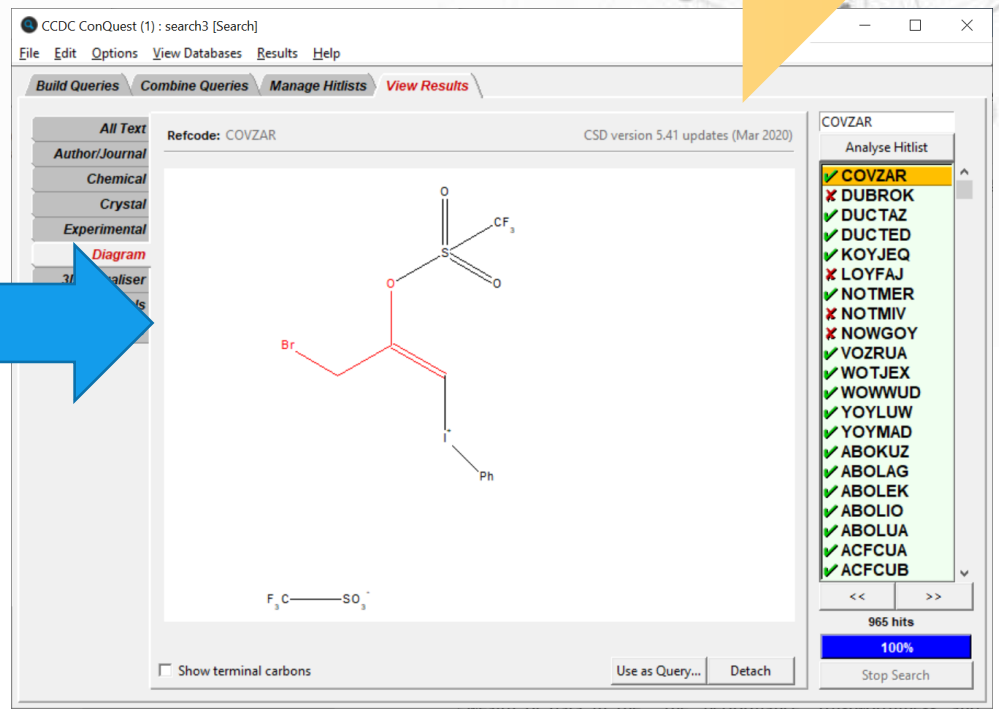
**1.** Build a Query: what do you want to find?

**2.** Click search and select filters: do you need to restrict your search?

**3.** Visualise and analyse results: what can you learn from this data?

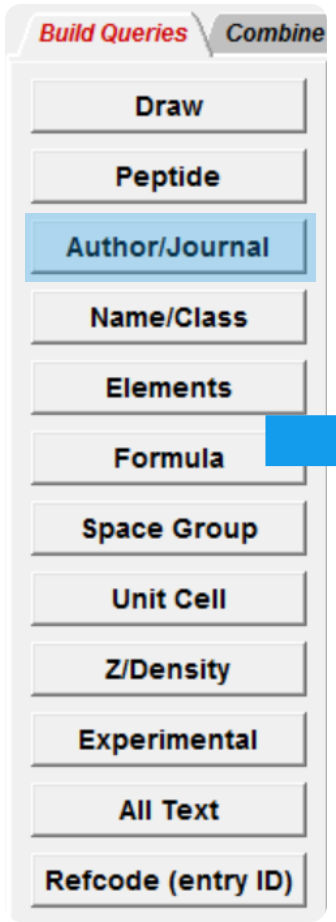


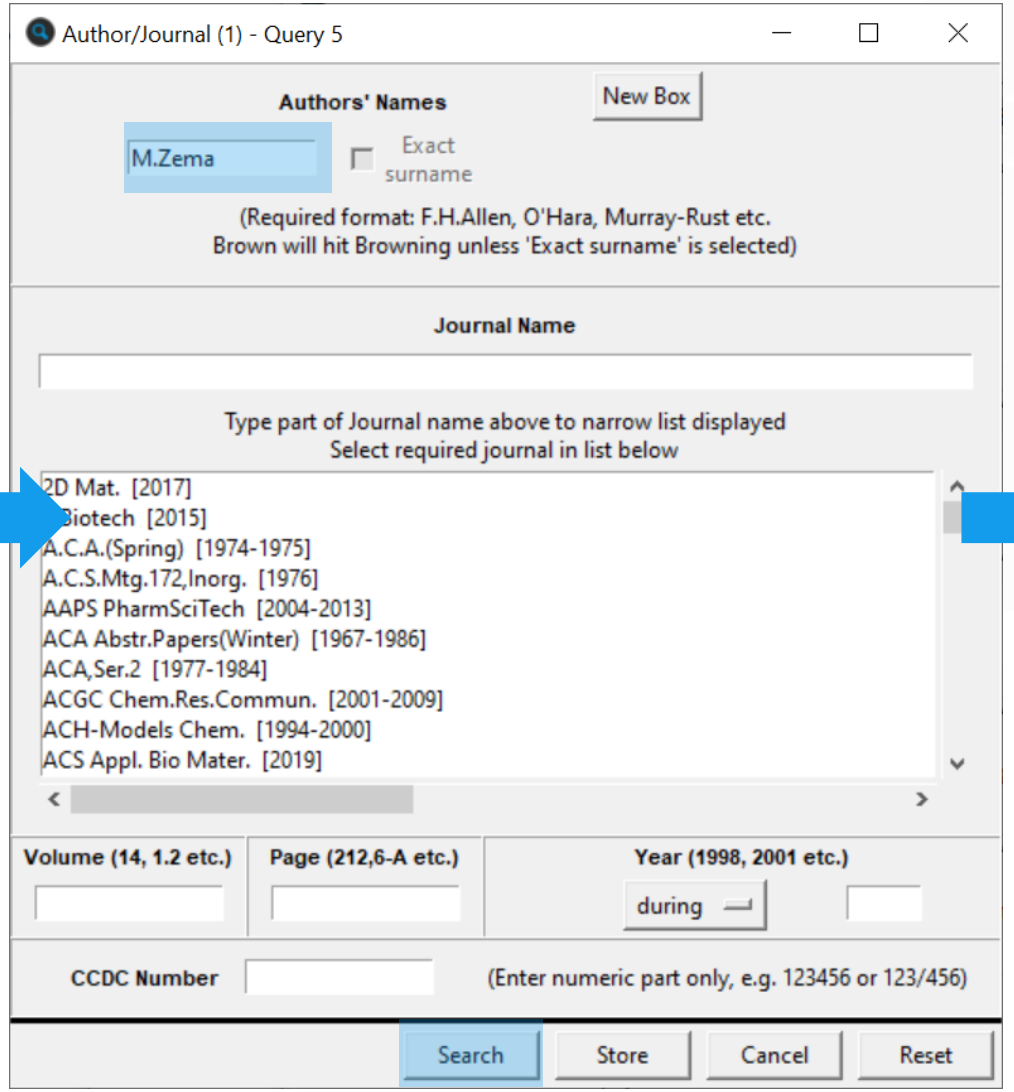



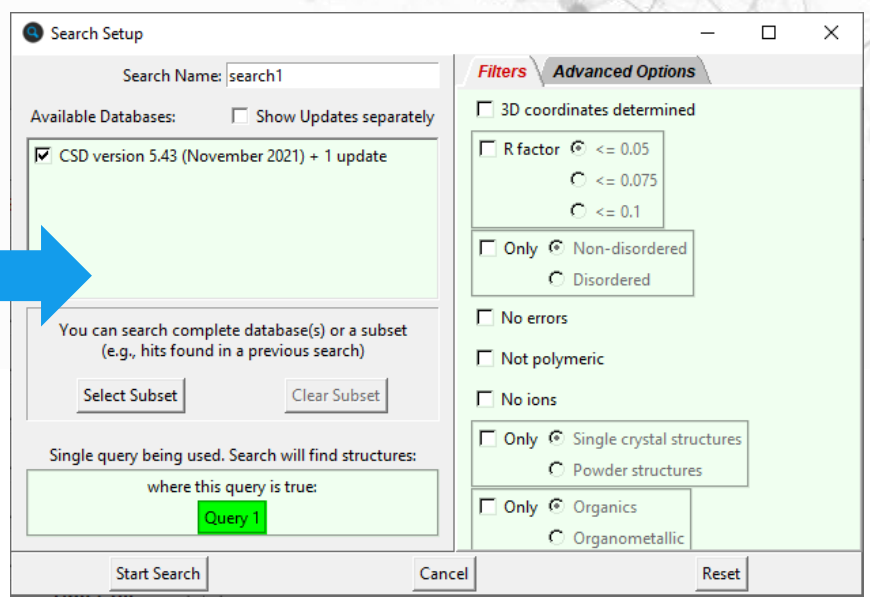




CCDC

# Author and journal search



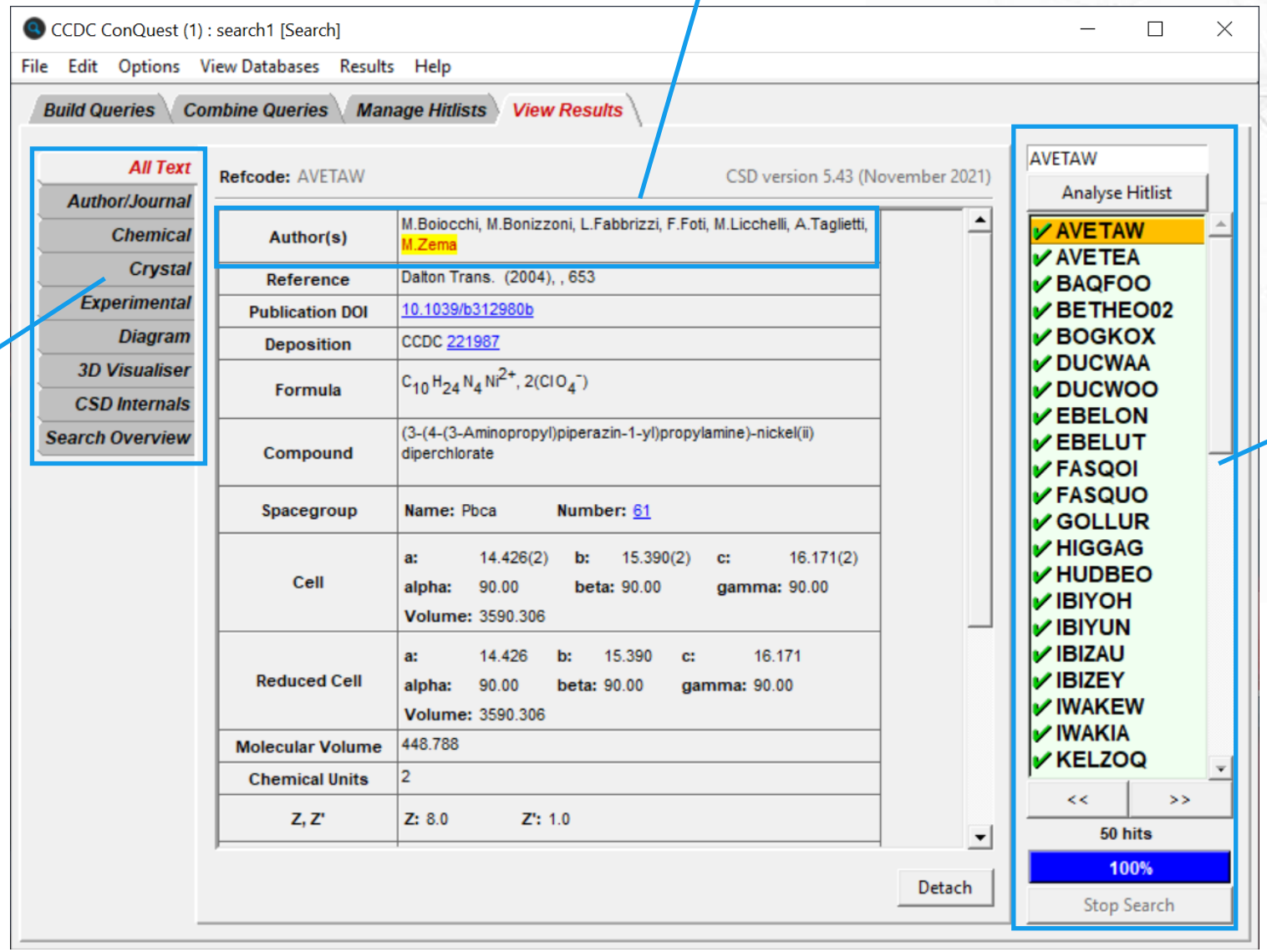




CCDC

# Results

Query highlighted

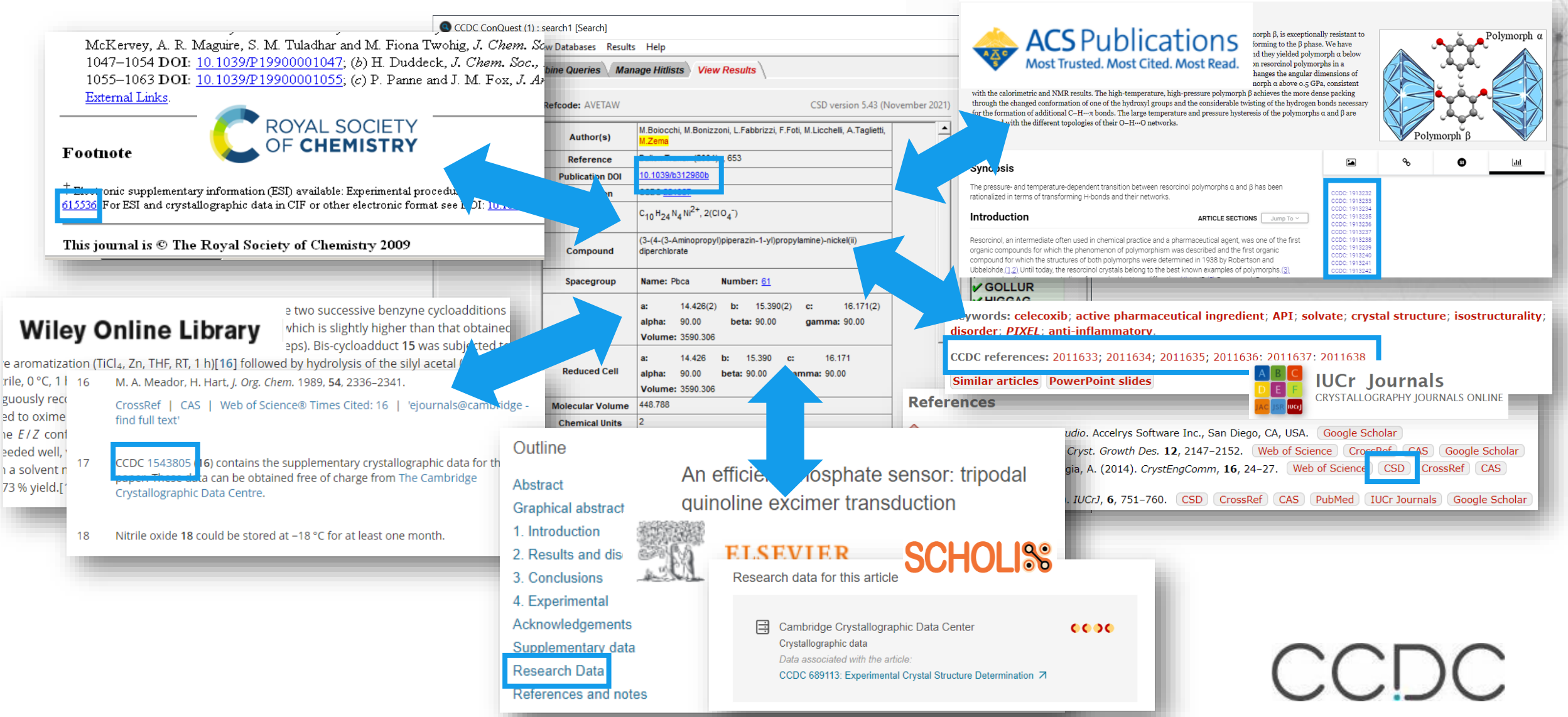



Left clicking each tab will display different information

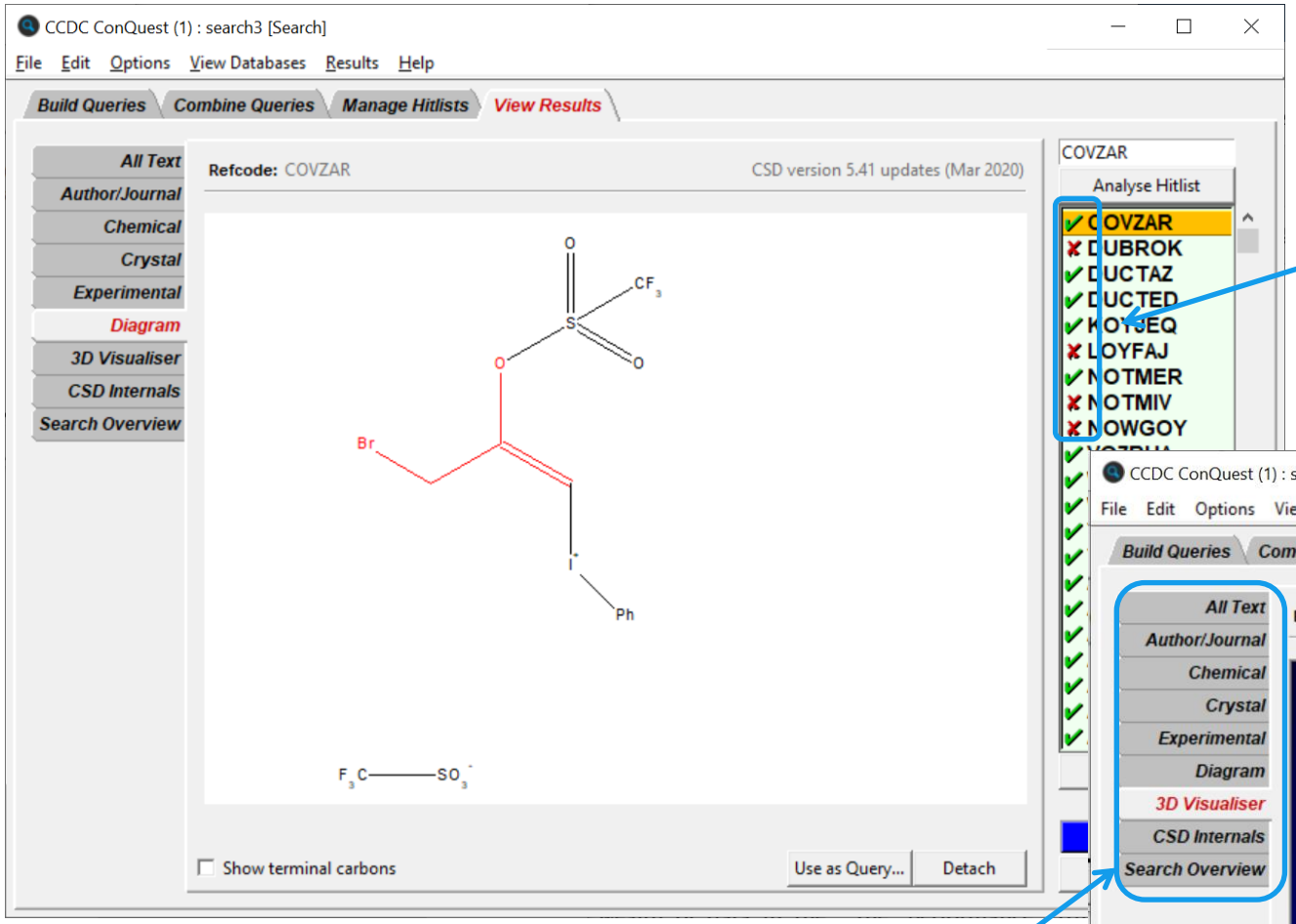
A hit list will be returned within the View Results tab of ConQuest. Left clicking on each REFCODE will toggle on and off that hit for future manipulation.

CCDC

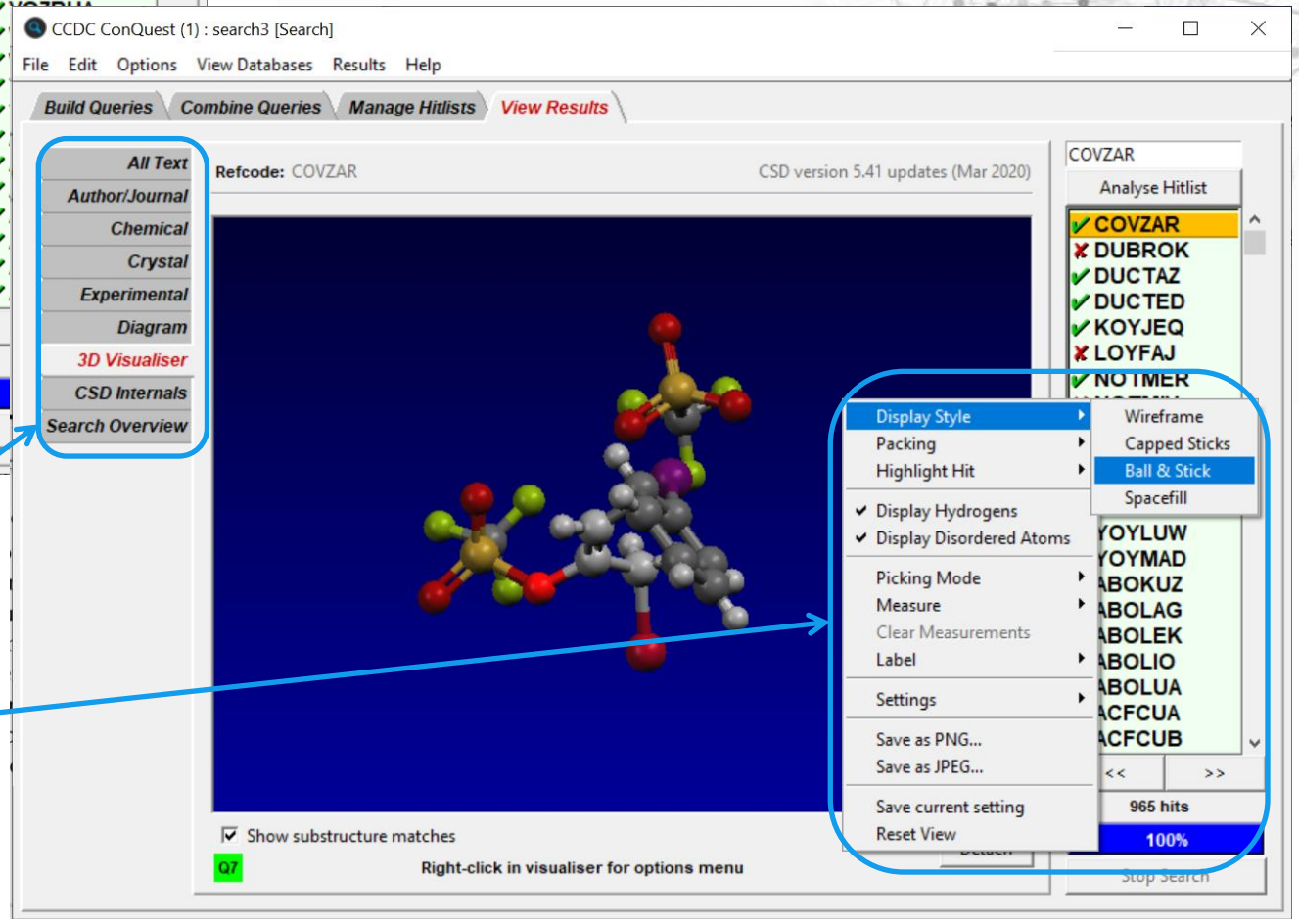
# From data to publication




CCDC

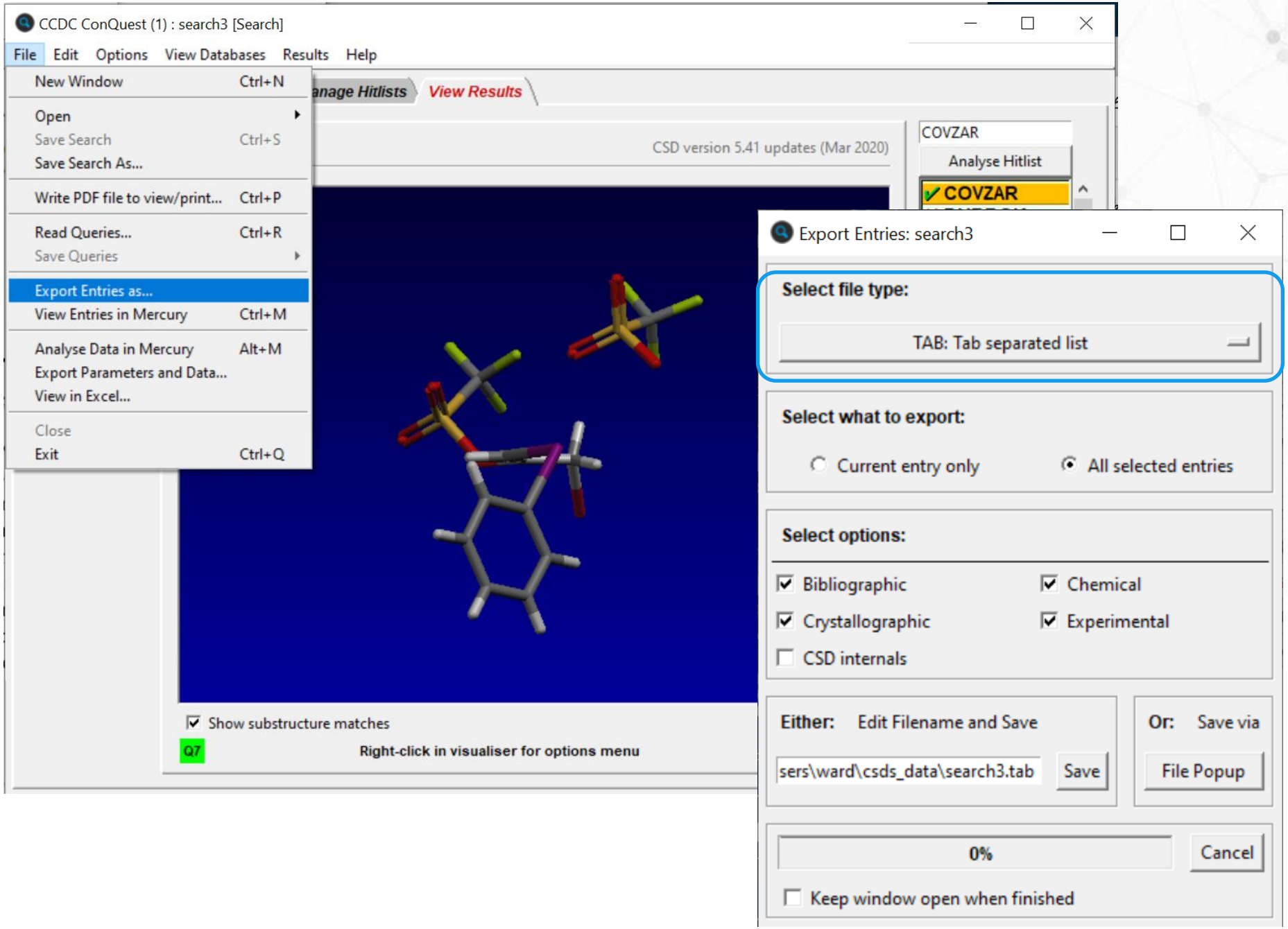



A hit list will be returned within the "results" tab of ConQuest.
Left clicking on each REFCODE will toggle on and off that hit for future manipulation.




Left clicking each tab will display different information

Right clicking on 3D visualiser enables you to change the display style



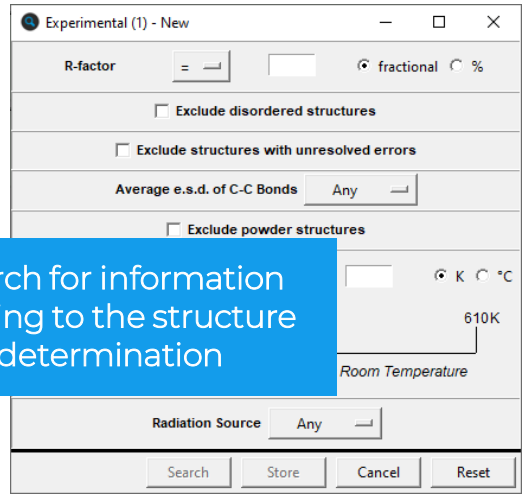




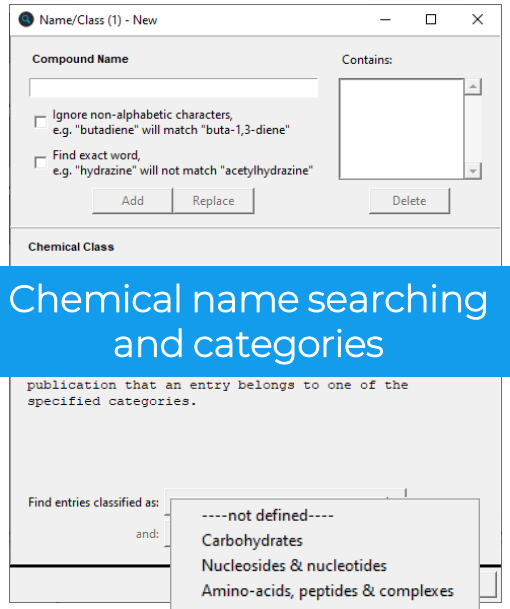
You can export results in a range of different formats to save data for future manipulation.

CCDC

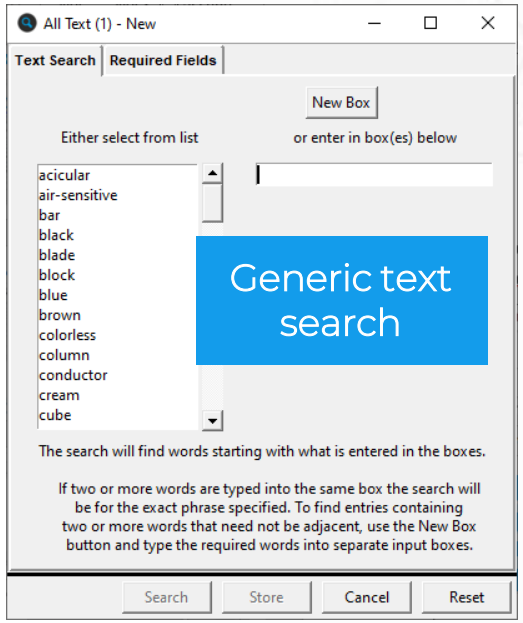
# ConQuest search features



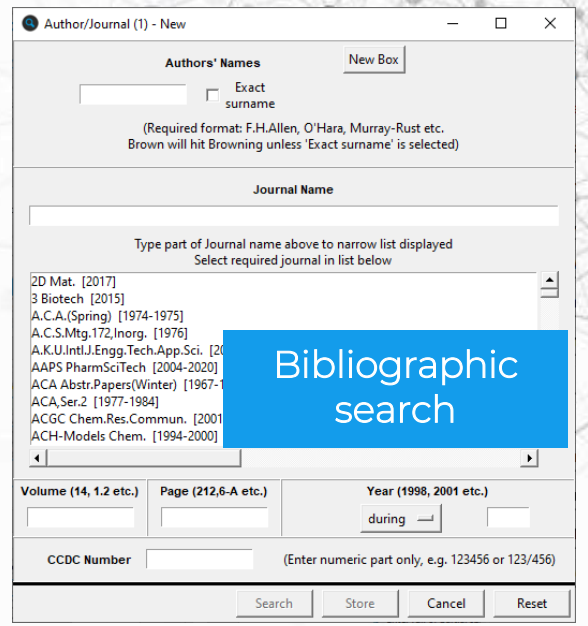

Search for information relating to the structure determination



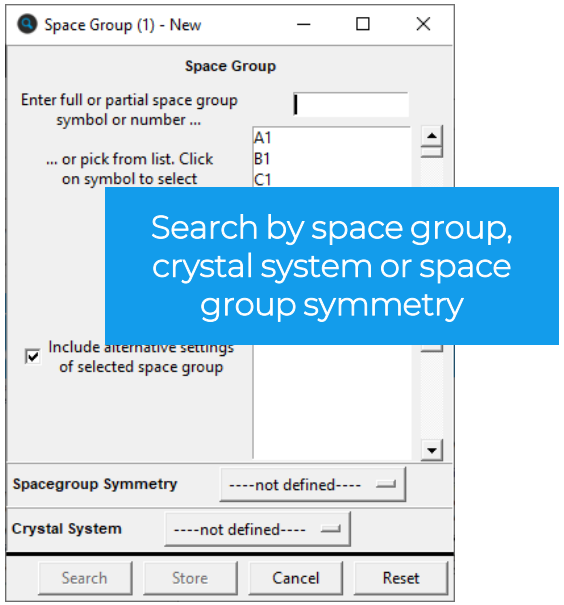

Chemical name searching and categories



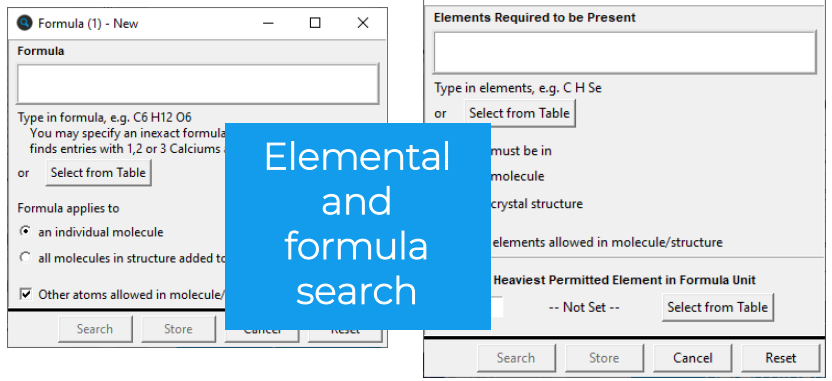

Generic text search



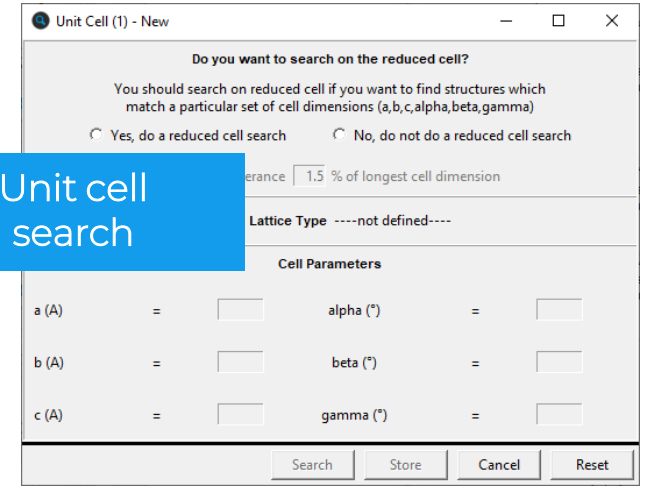

Bibliographic search




Search by space group, crystal system or space group symmetry



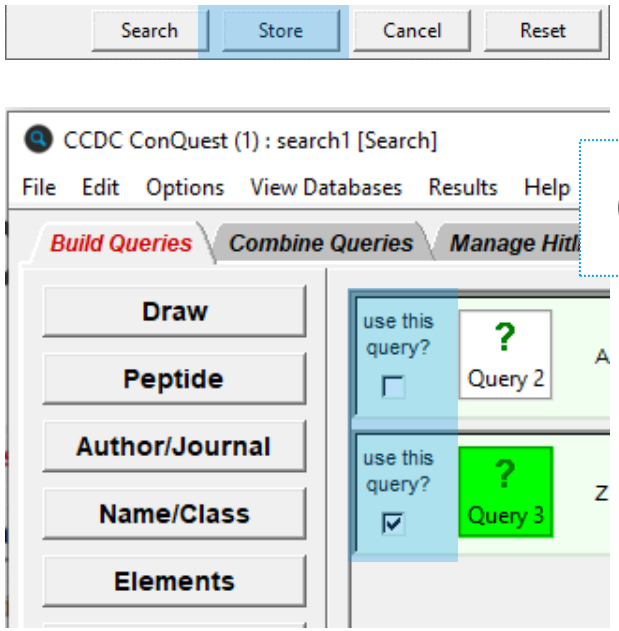

Elemental and formula search




Unit cell search

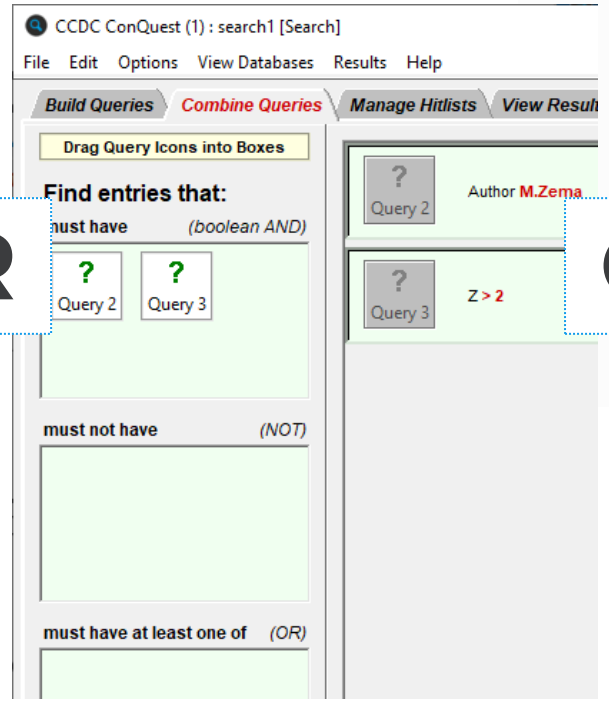
# Combining searches and results

## Store and Search



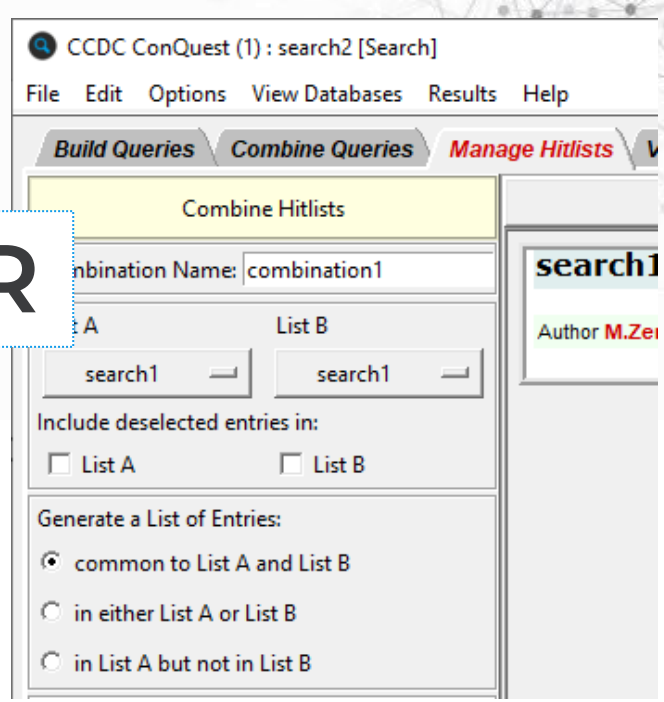

OR

## Combine Queries

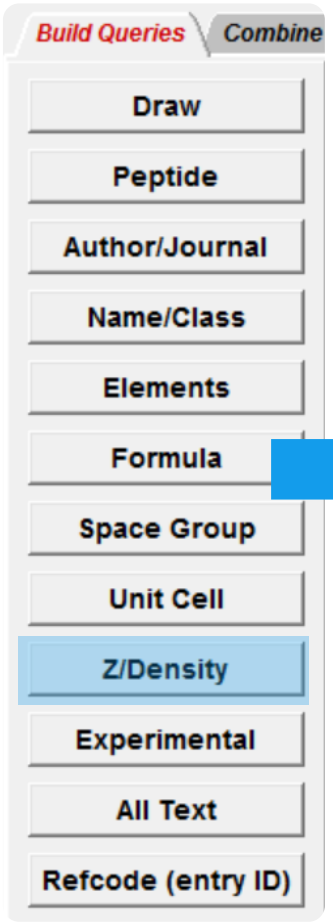


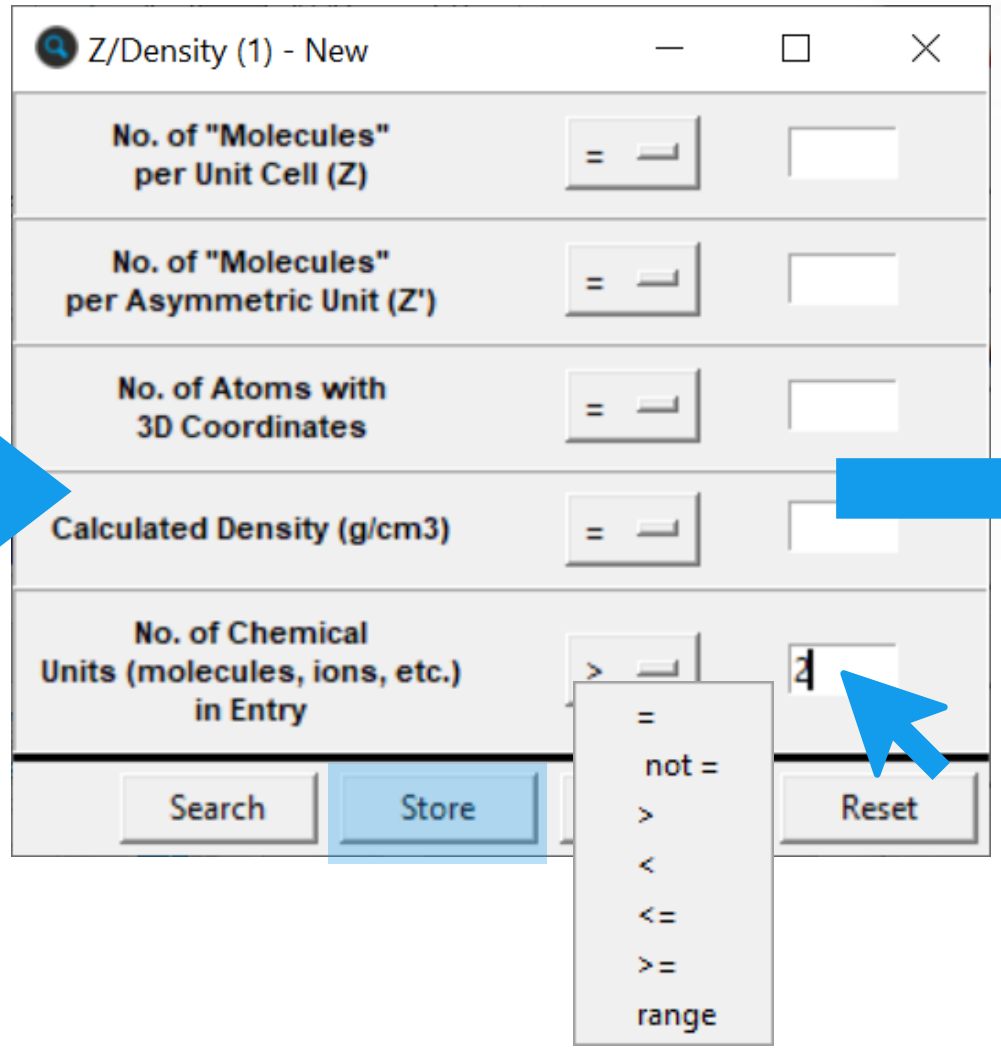

OR

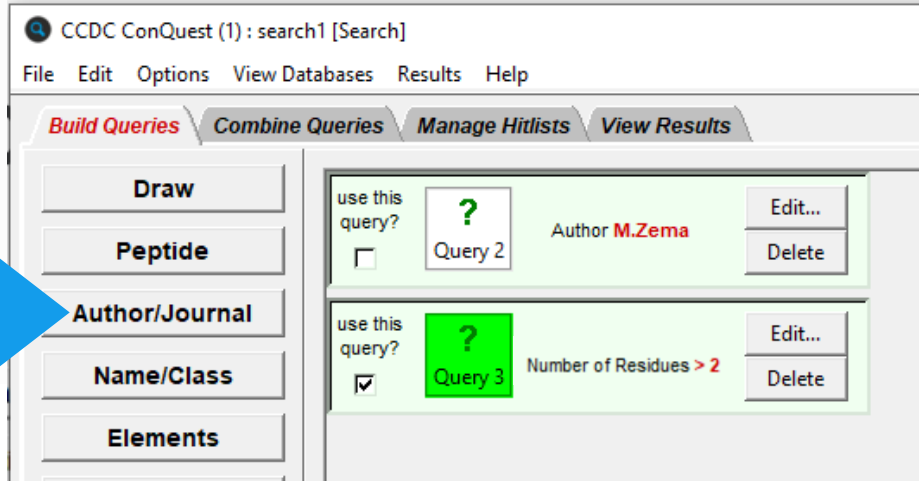
## Manage Hitlists



# Store a query shown on Z/Density search



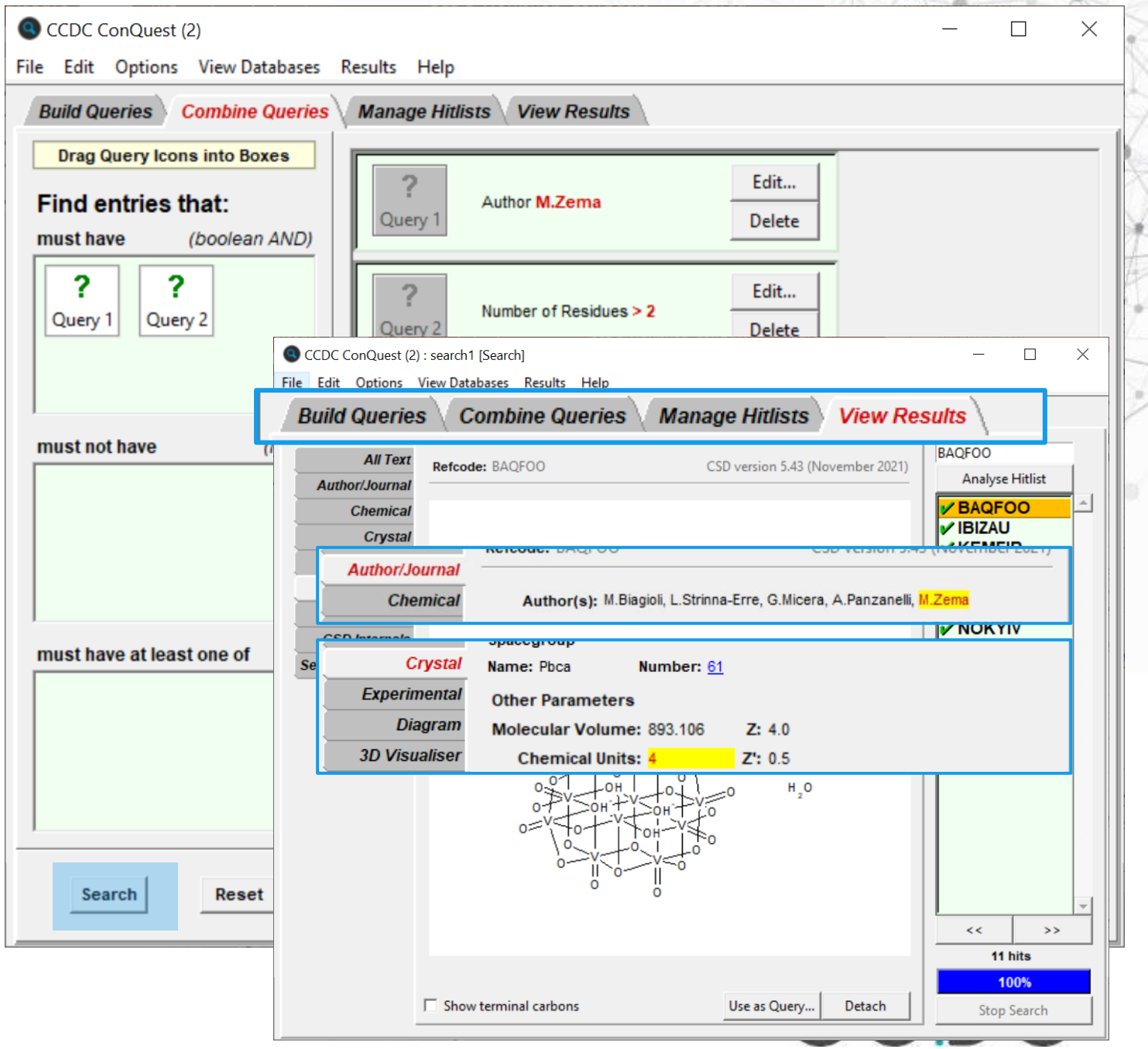

CCDC

# Combine Queries


Allows you to combine different queries using boolean AND, NOT, OR

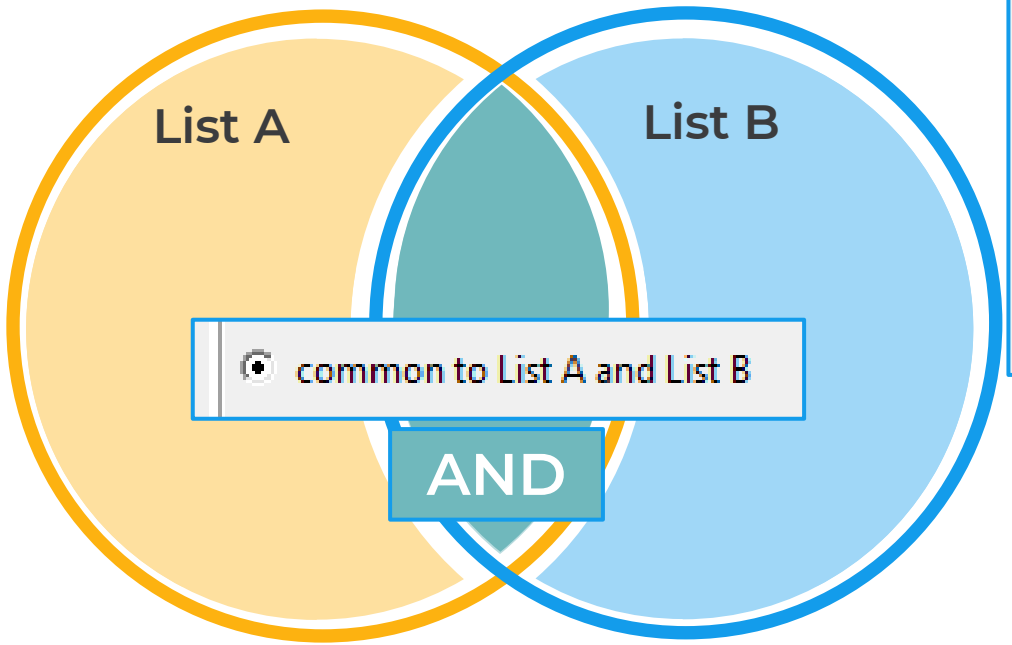
Query 1: Authors include M.Zema
Query 2: Number of Residues > 2

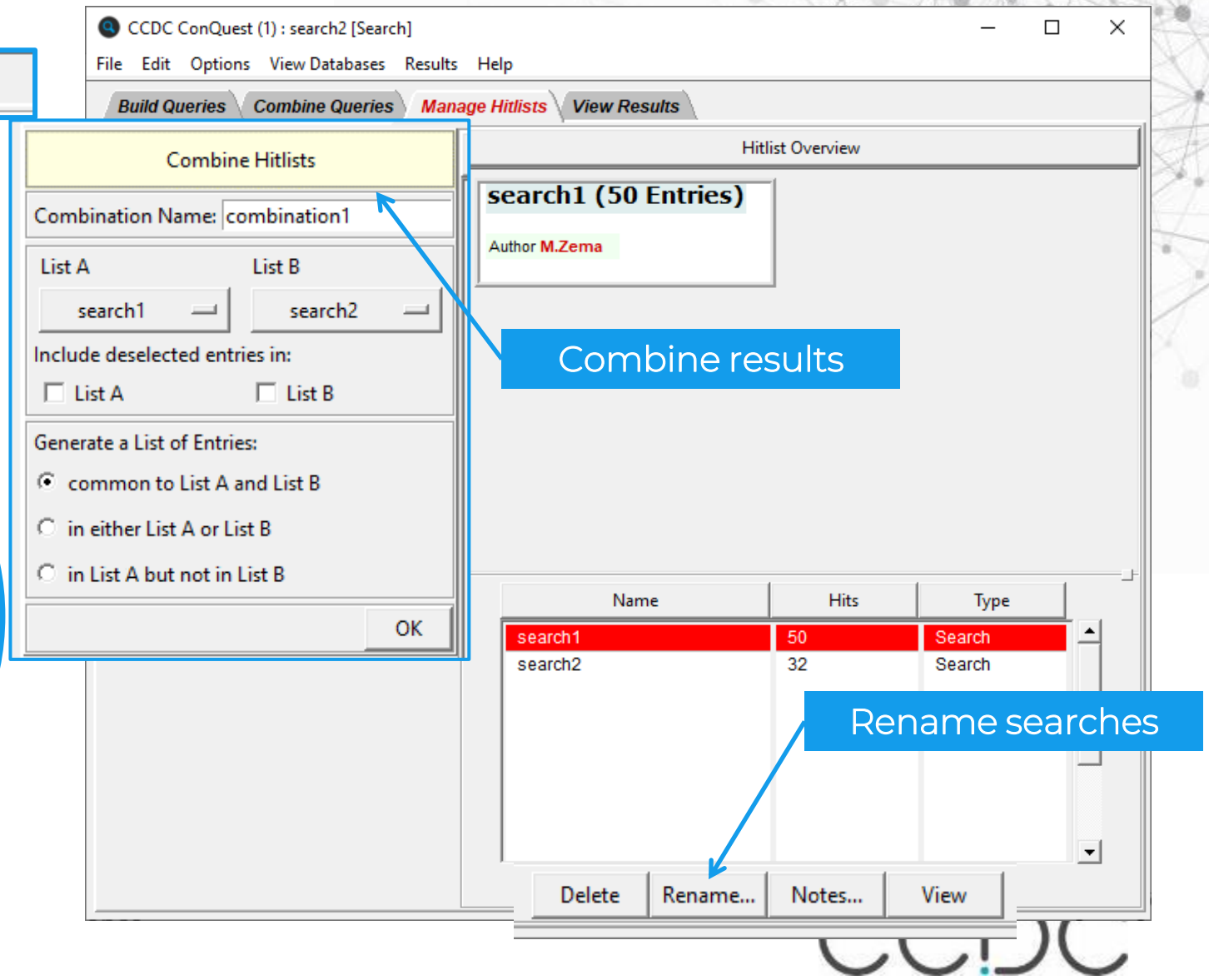
**Must have:** Query 1 **AND** Query 2



# Manage Hitlists

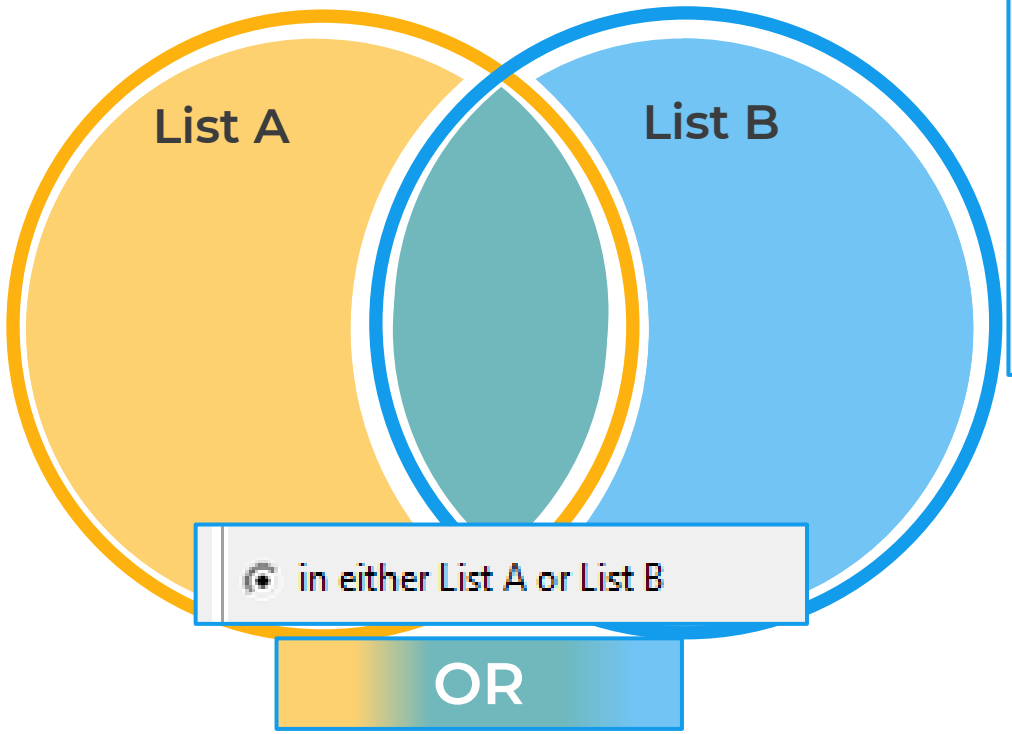

- Hitlists = results from searches that you have already run

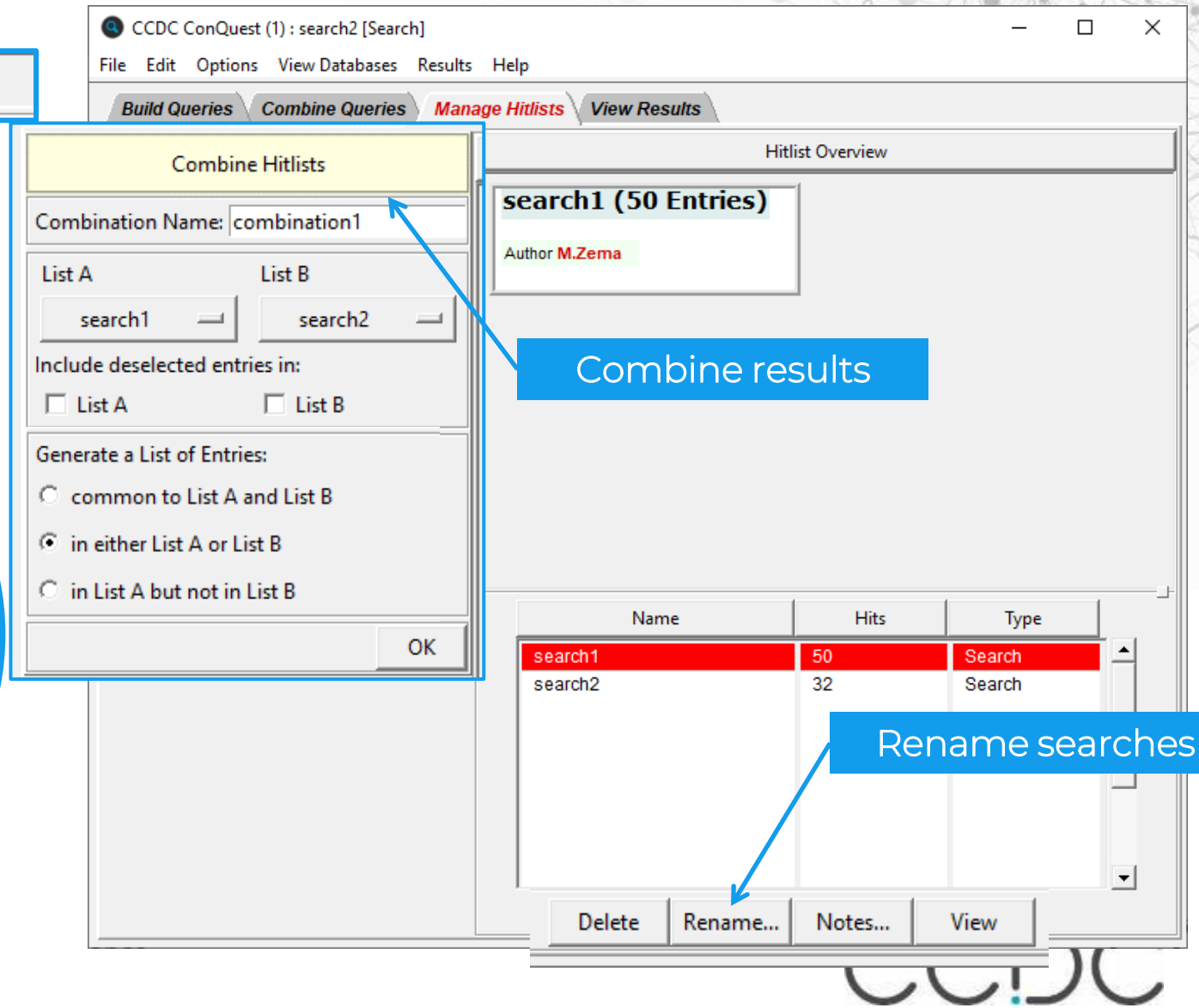


# Manage Hitlists

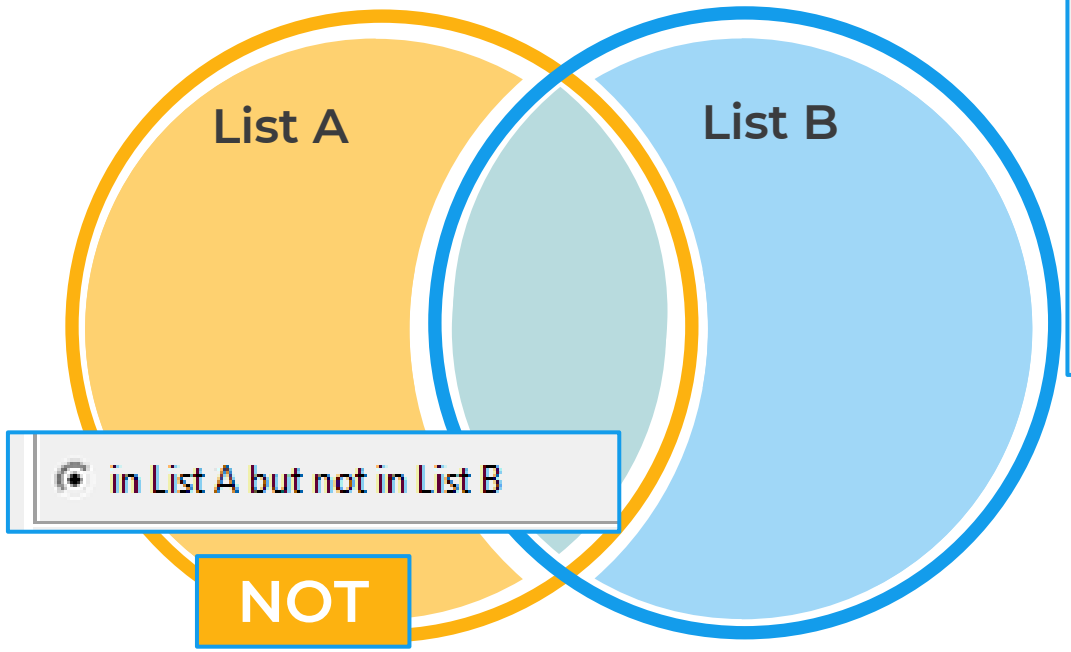

- Hitlists = results from searches that you have already run

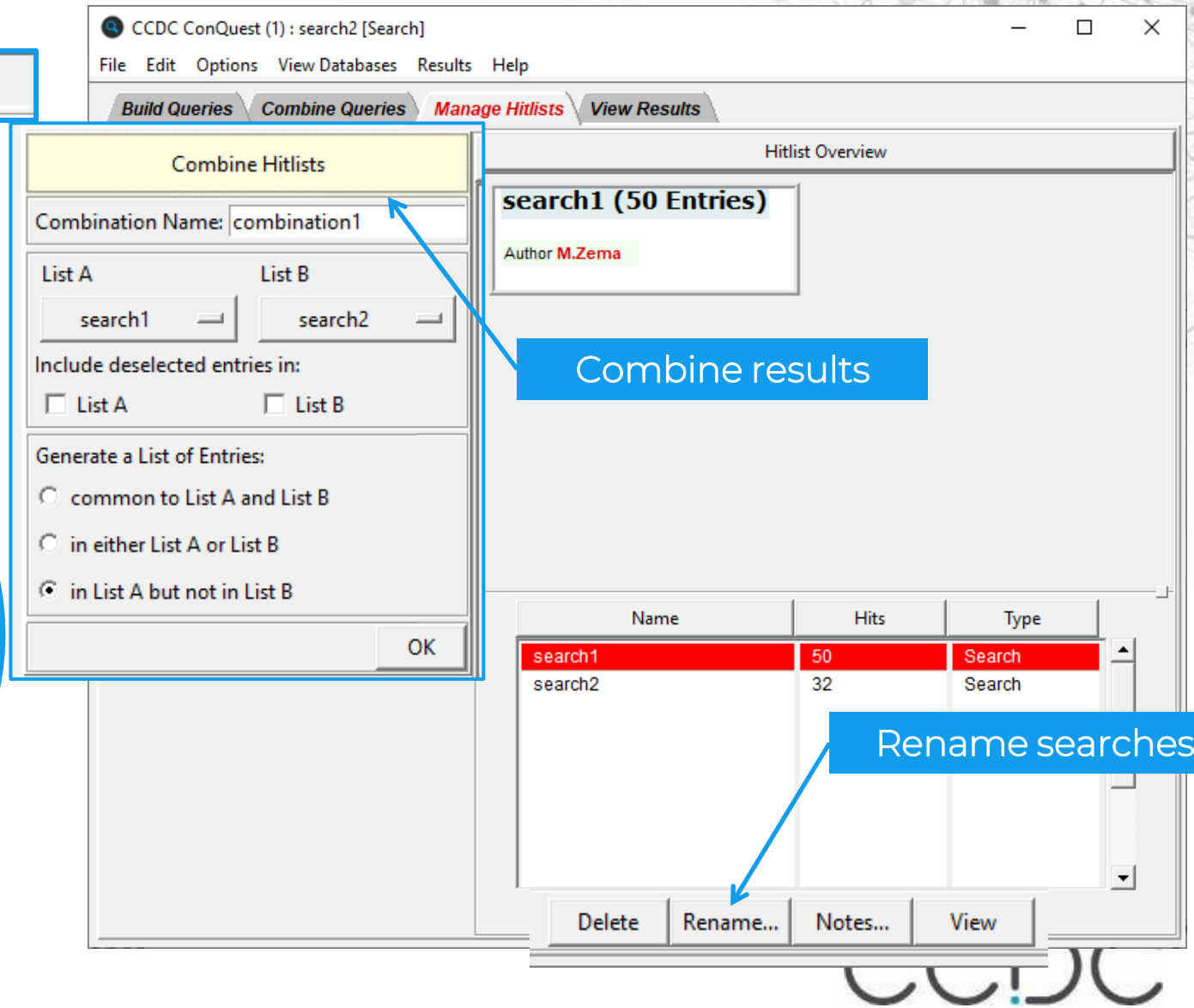


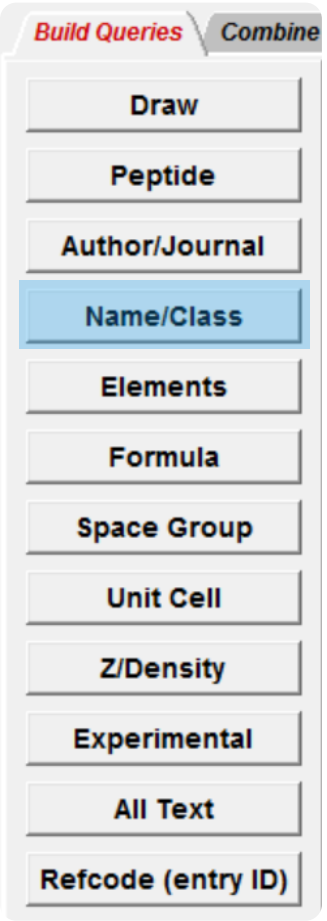
# Manage Hitlists

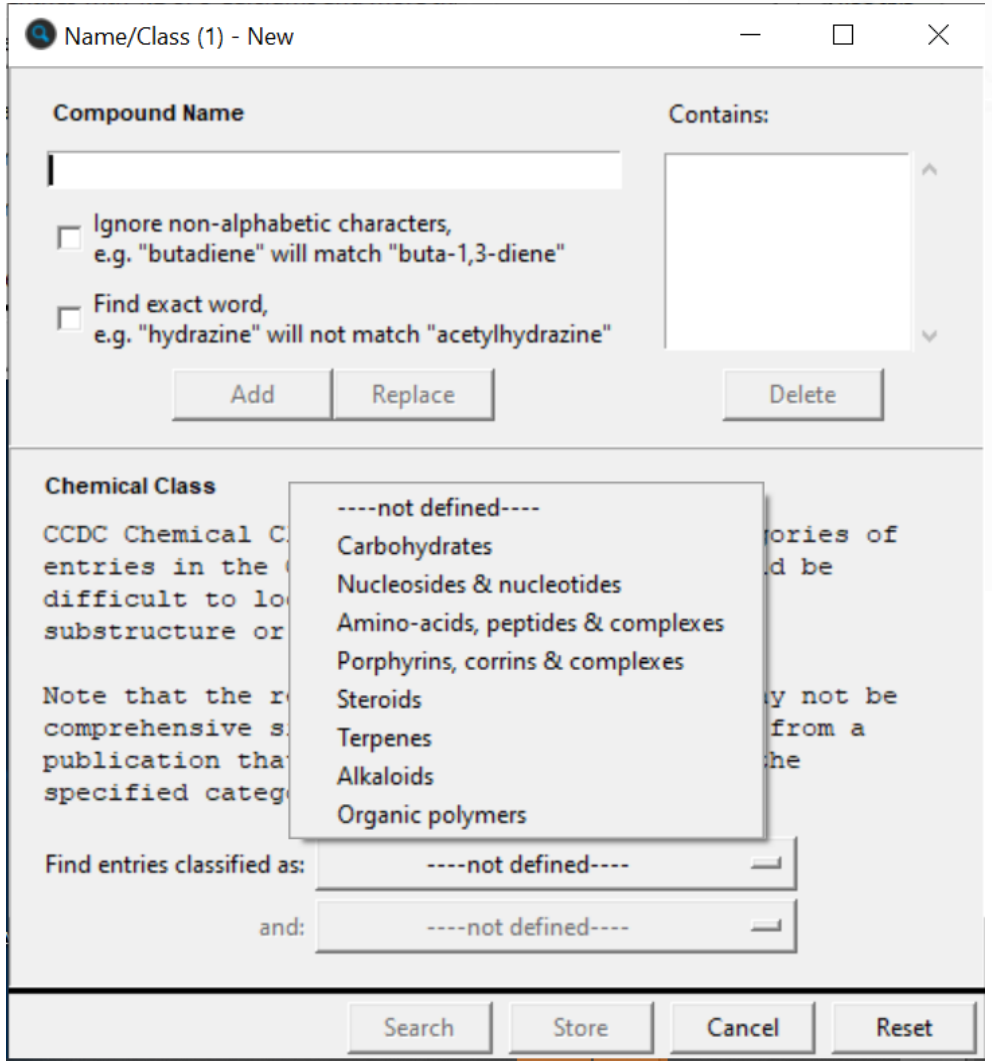

- Hitlists = results from searches that you have already run



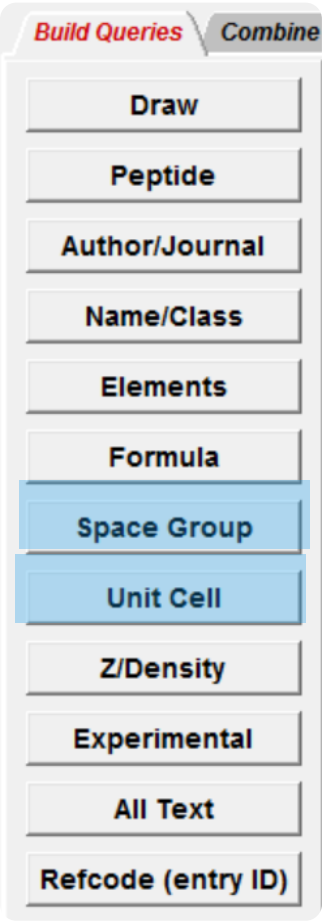



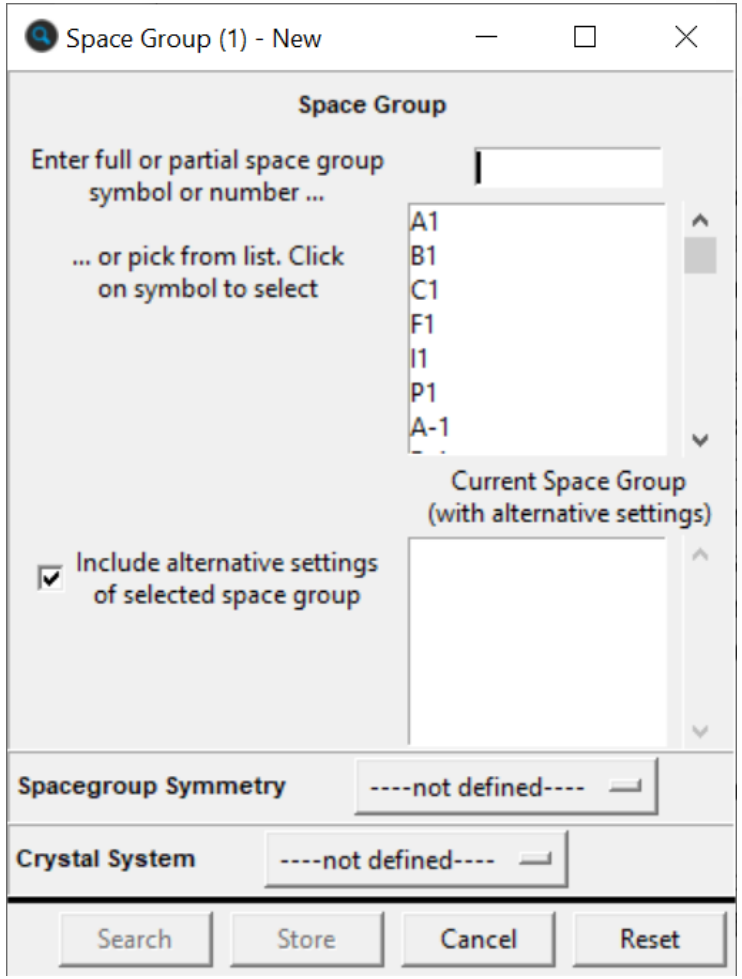
# Compound name search

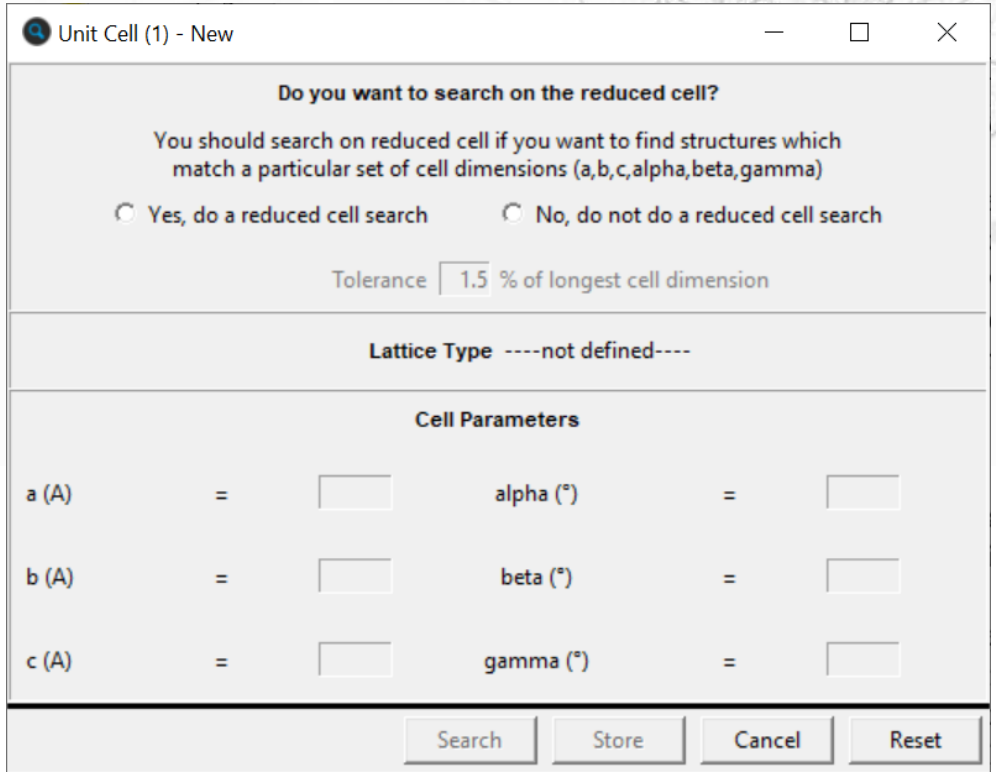


# Space Group and cell search

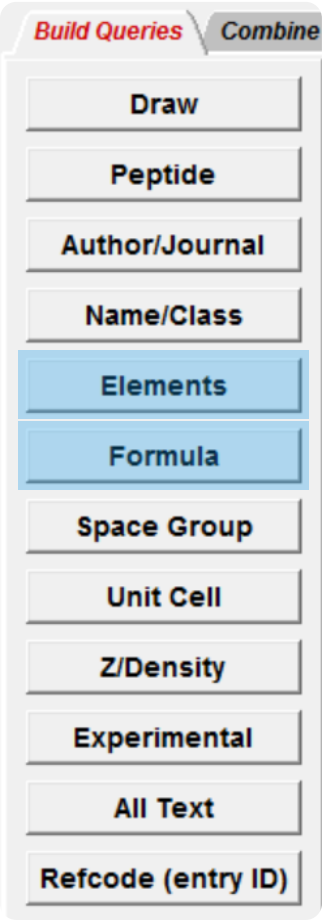


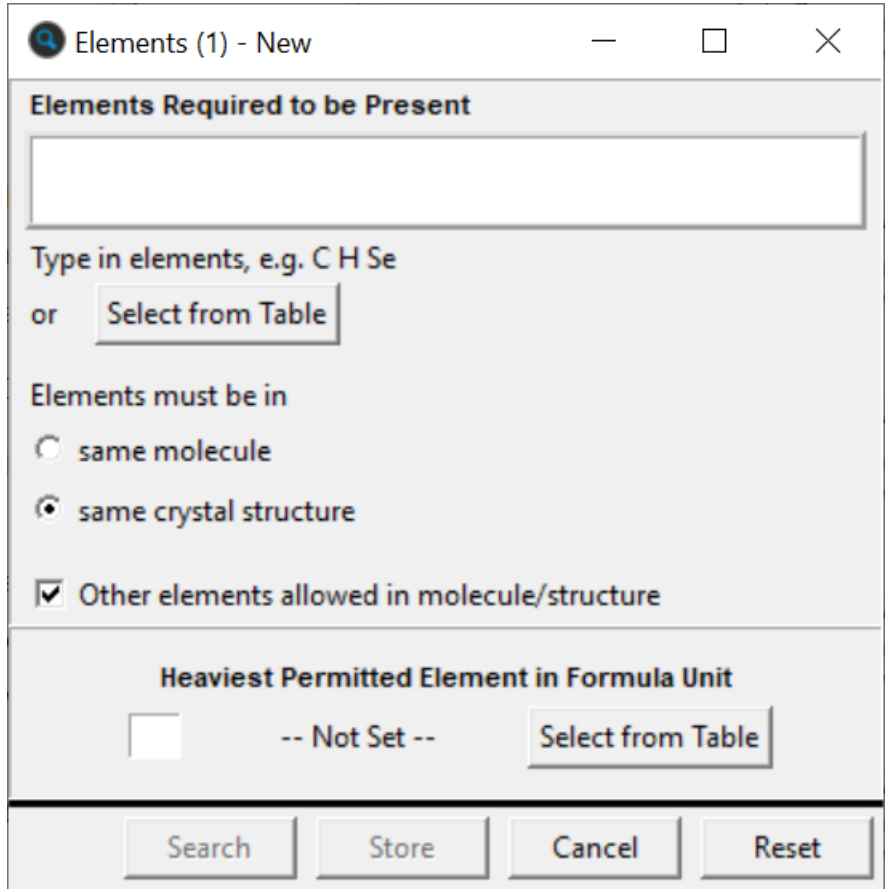



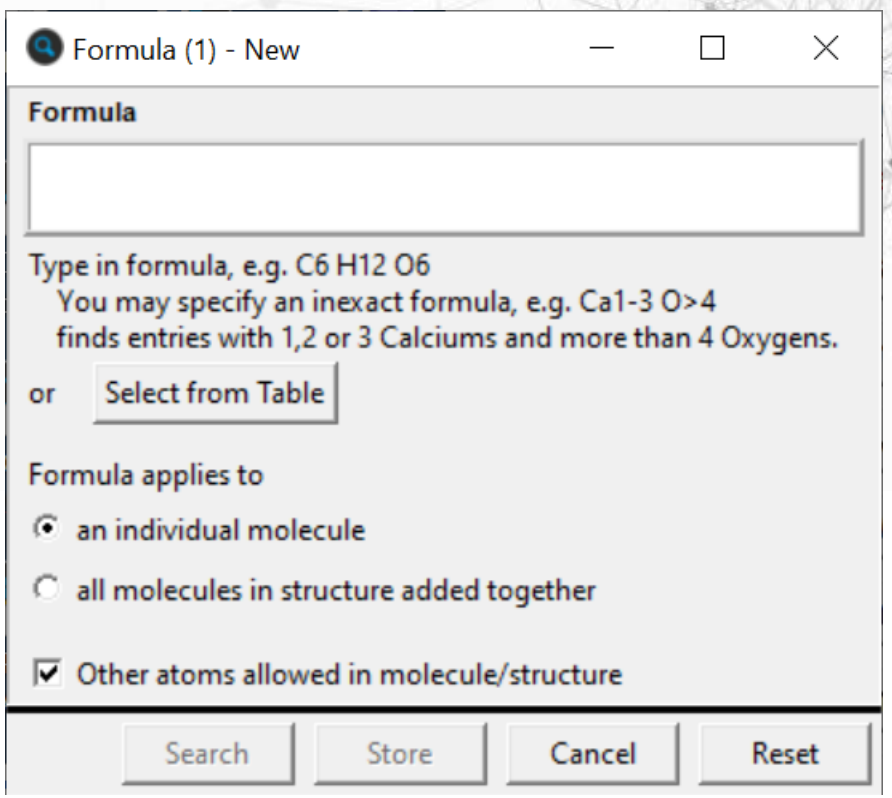




CCDC

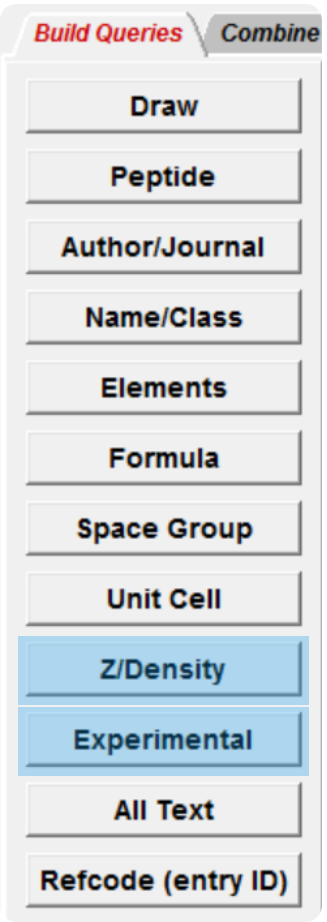
# Elements and formula search

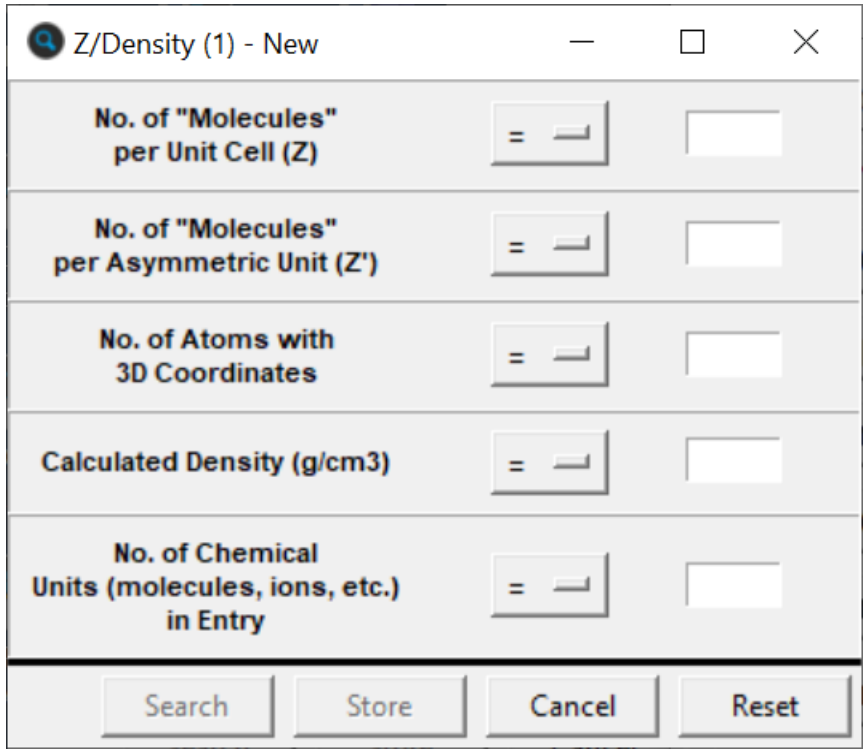


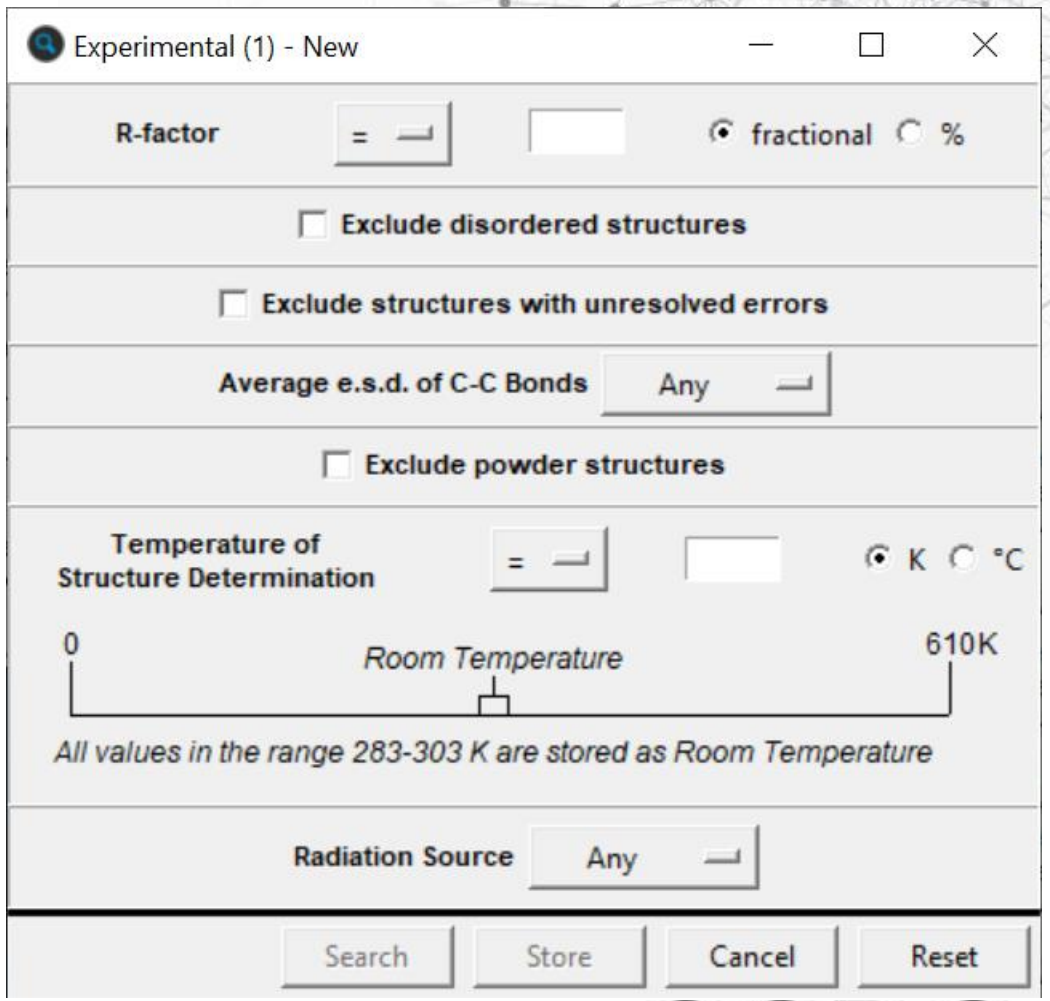



# Z/Density and experimental info search

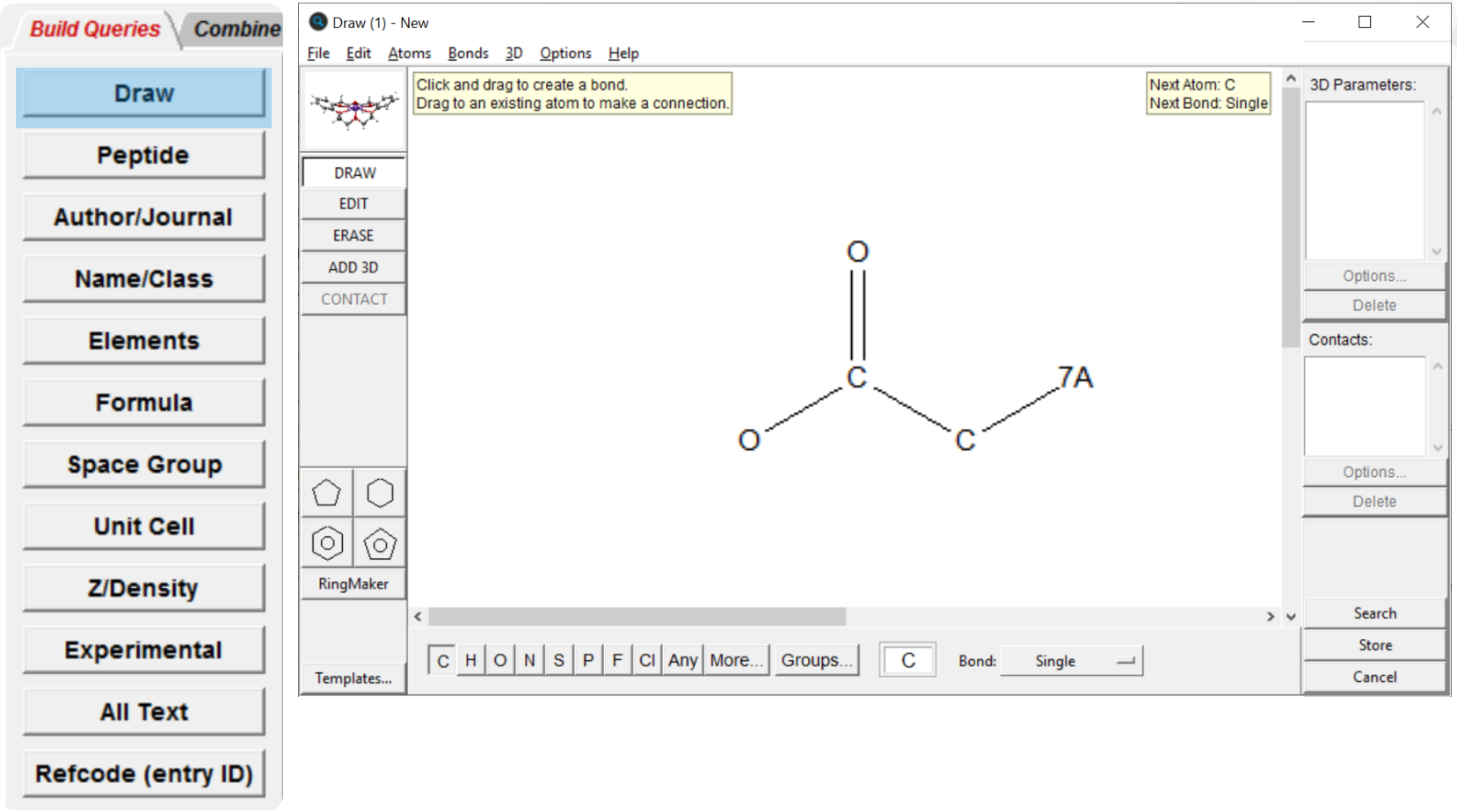


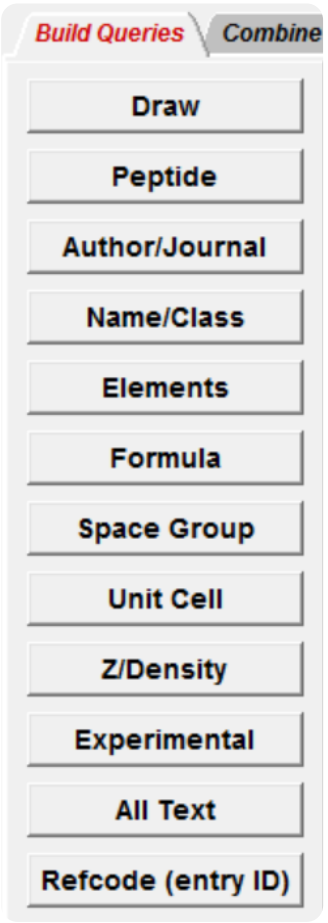

CCDC

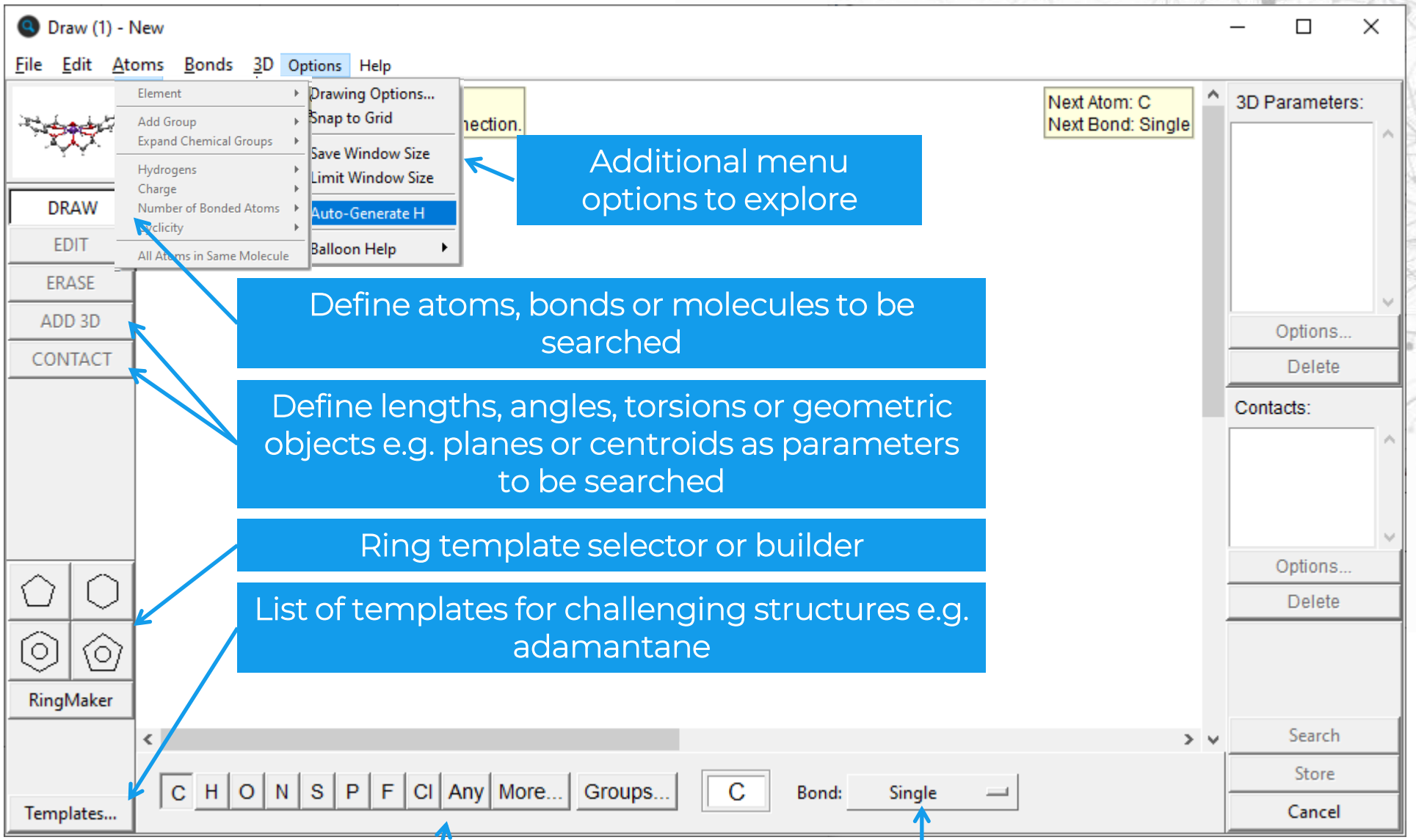
# Draw/Structure search




Example of a search, where 7A represents any halogen

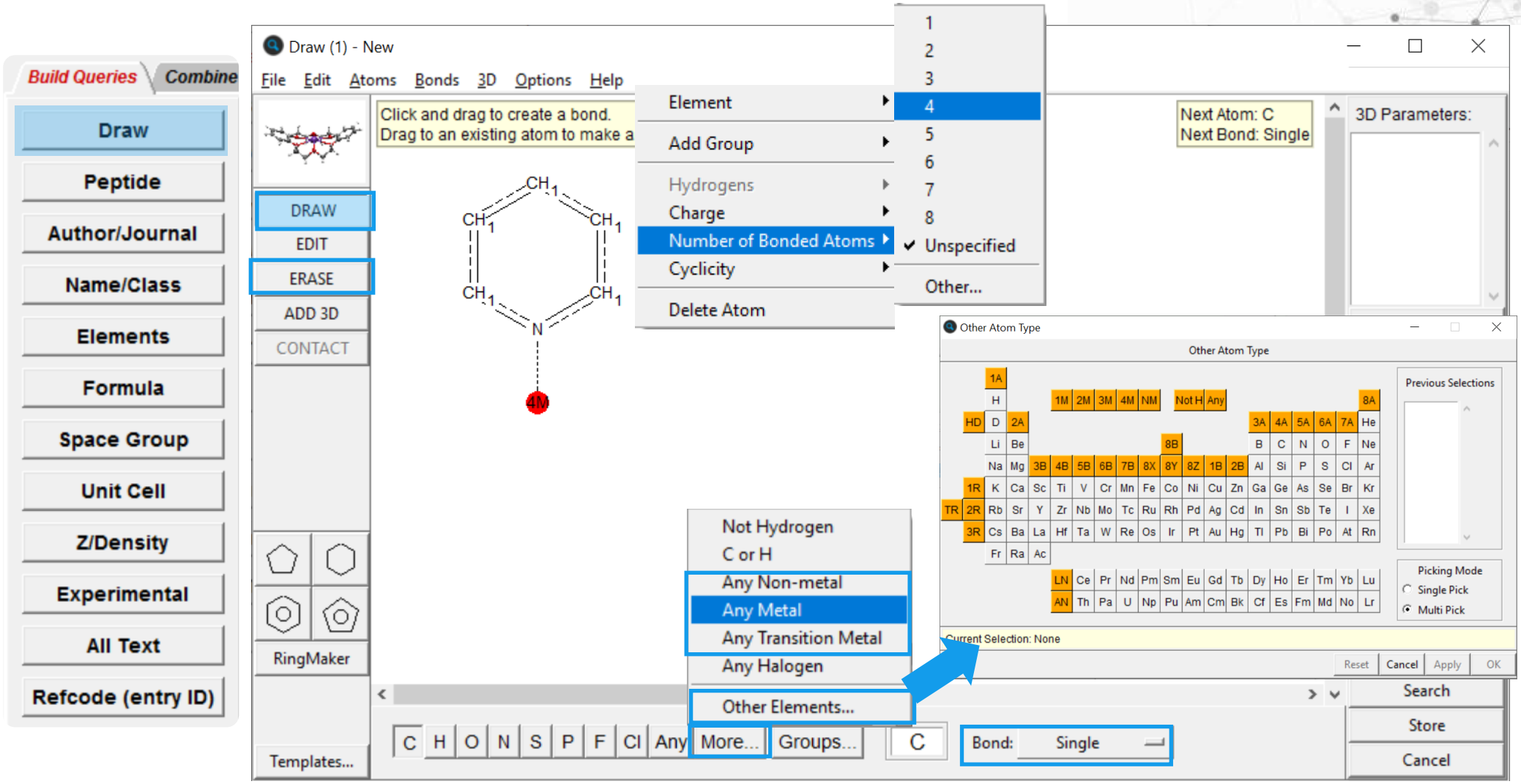
# Draw


Build Queries
Combine
Draw
Peptide
Author/Journal
Name/Class
Elements
Formula
Space Group
Unit Cell
Z/Density
Experimental
All Text
Refcode (entry ID)


Draw (1) - New
File
Edit
Atoms
Bonds
3D
Options
Help
Element
Add Group
Expand Chemical Groups
Hydrogens
Charge
Number of Bonded Atoms
Cyclicity
All Atoms in Same Molecule
Drawing Options...
Snap to Grid
Save Window Size
Limit Window Size
Auto-Generate H
Balloon Help
DRAW
EDIT
ERASE
ADD 3D
CONTACT
RingMaker
Templates...
Next Atom: C
Next Bond: Single
3D Parameters:
Options...
Delete
Contacts:
Options...
Delete
Search
Store
Cancel
C H O N S P F Cl Any More... Groups...
C
Bond: Single
Additional menu options to explore
Define atoms, bonds or molecules to be searched
Define lengths, angles, torsions or geometric objects e.g. planes or centroids as parameters to be searched
Ring template selector or builder
List of templates for challenging structures e.g. adamantane
Select specific or general atom types/functional groups
Select bond type

CCDC

# ConQuest – Draw/Structure search

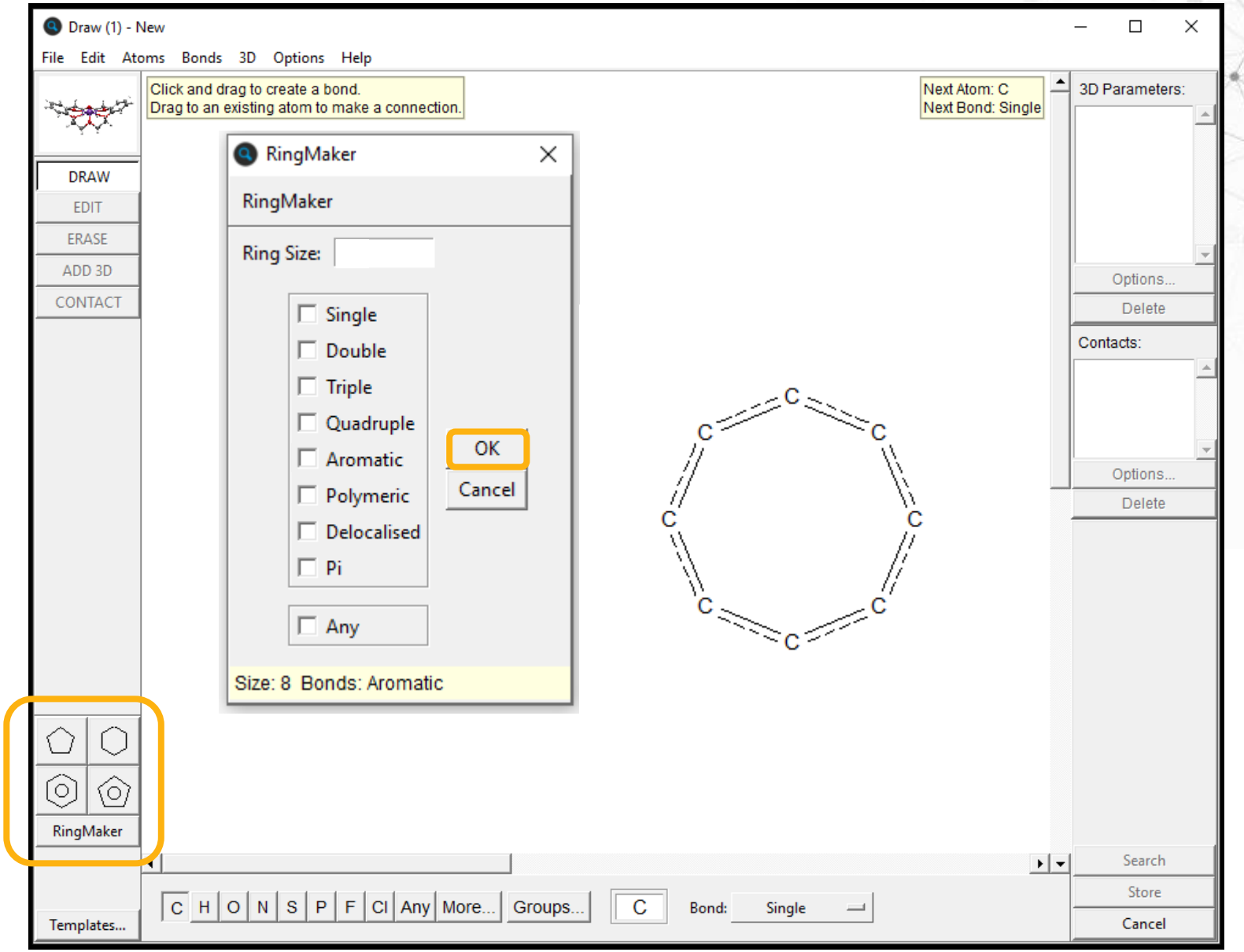



Left click in space to add an atom

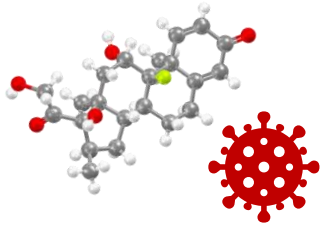
Left click and drag from an atom to attach an atom

Right click on an atom or bond to edit or add extra properties to it

# Making Drawing Easier: RingMaker

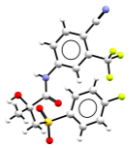


COVID-19

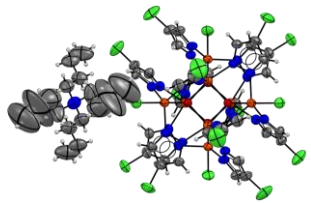


Pesticides

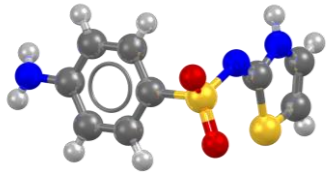
Drugs



ADPs



Best representative

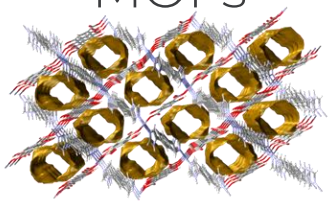


Teaching

MOF Dimensionality

- 1D
- 2D
- 3D

MOFs



# CSD Subsets

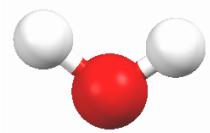
*Groups of structures that may be more difficult to find in CSD from searching alone*

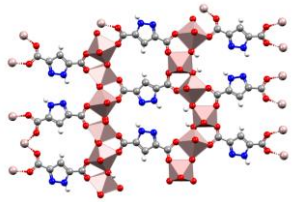
**Best representative lists** provide a single example of every structure in the CSD (including polymorphs) under a particular condition;

- low temperature,
- room temperature,
- with 3D determined hydrogen atoms,
- or the lowest R factor measurement

**New in CSD v5.43**

Hydrates

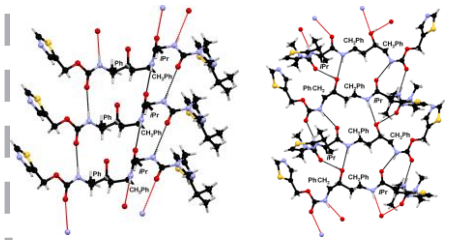




Electron Diffraction

High Pressure

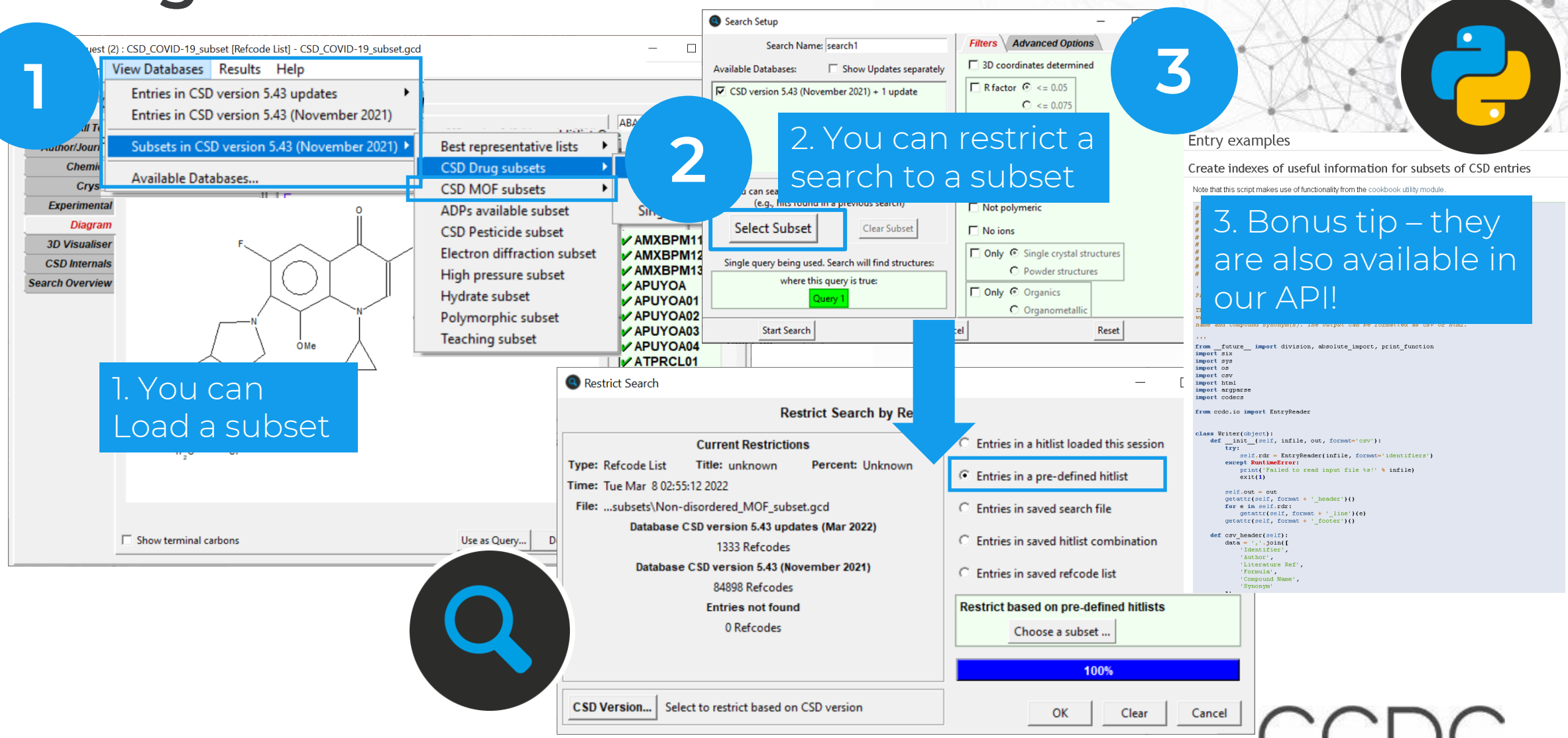
Polymorphs



Included structures must have an R<10%, no disorder and 3-D coordinates present (amongst other criteria).
*van de Streek, J. (2006). Acta Cryst. B62, 567-579. DOI: 10.1107/S0108768106019677*

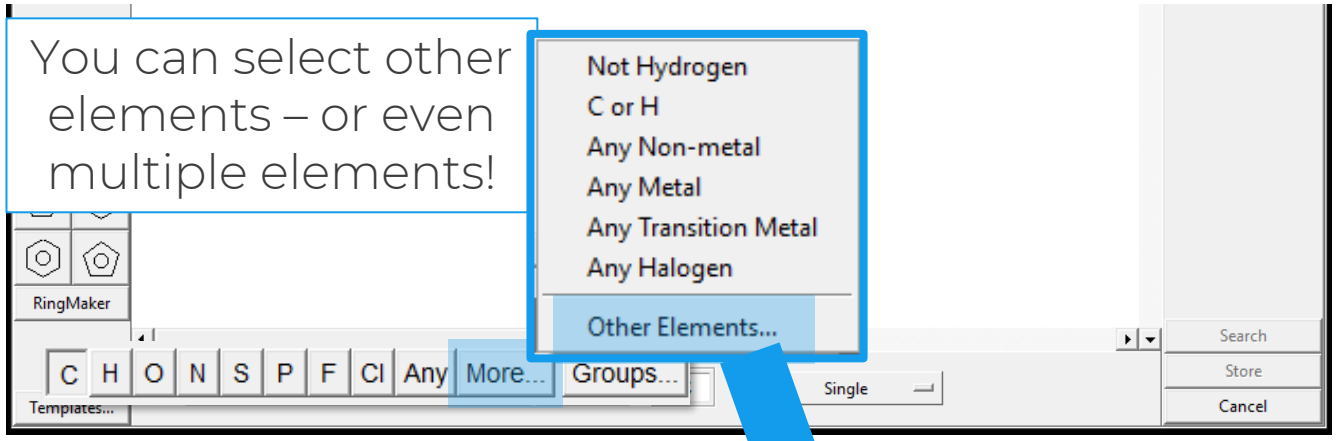
CCDC

# Using the subsets



# Making Drawing Easier: Other Elements

You can select other elements – or even multiple elements!

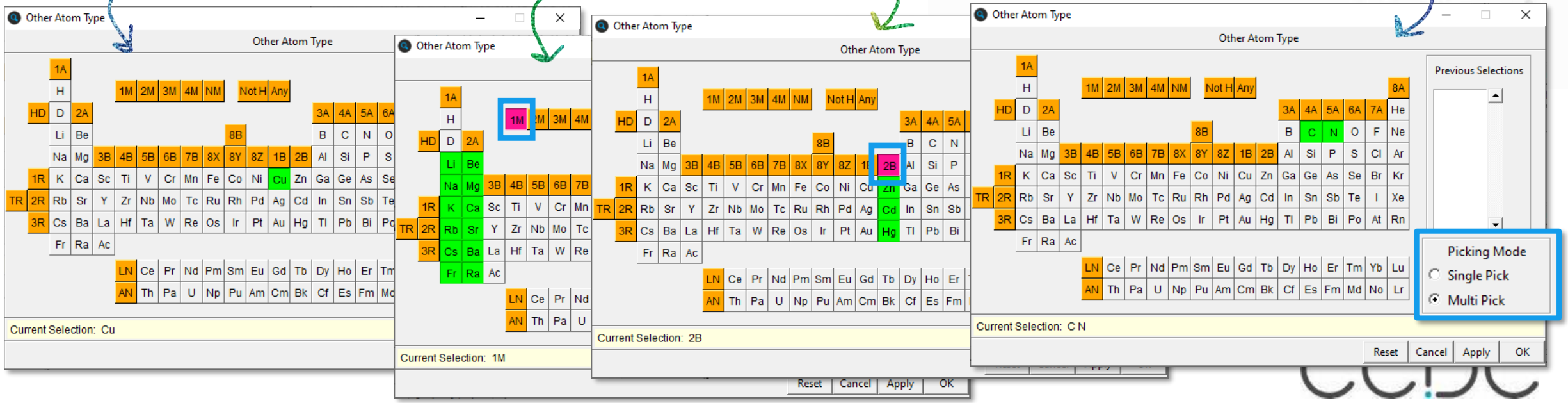



On the periodic table
Single Element: click on the element you need.
Pre-defined selection (e.g. groups, periods, metals, ...): click on the corresponding orange tile.
Personalised selection: tick Multi Pick, then select the elements you need.

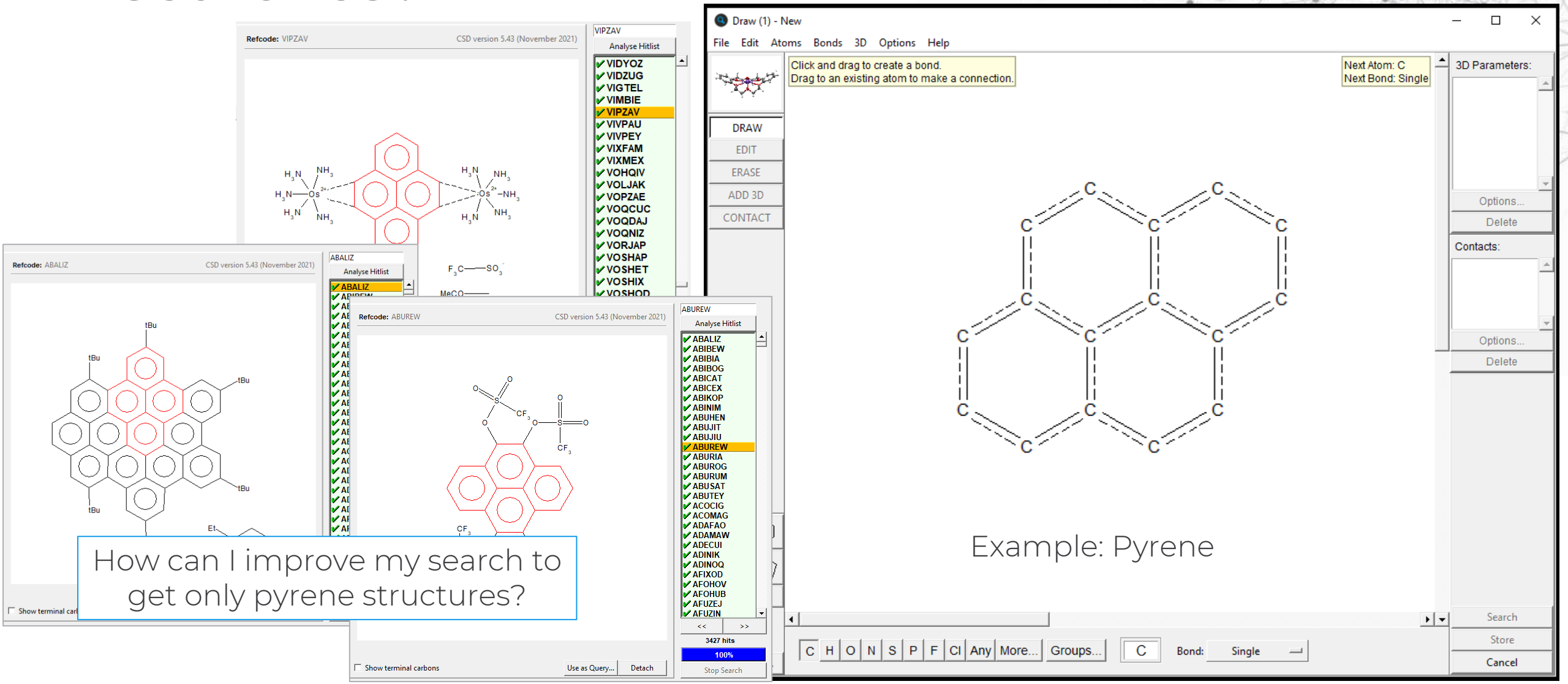
single element

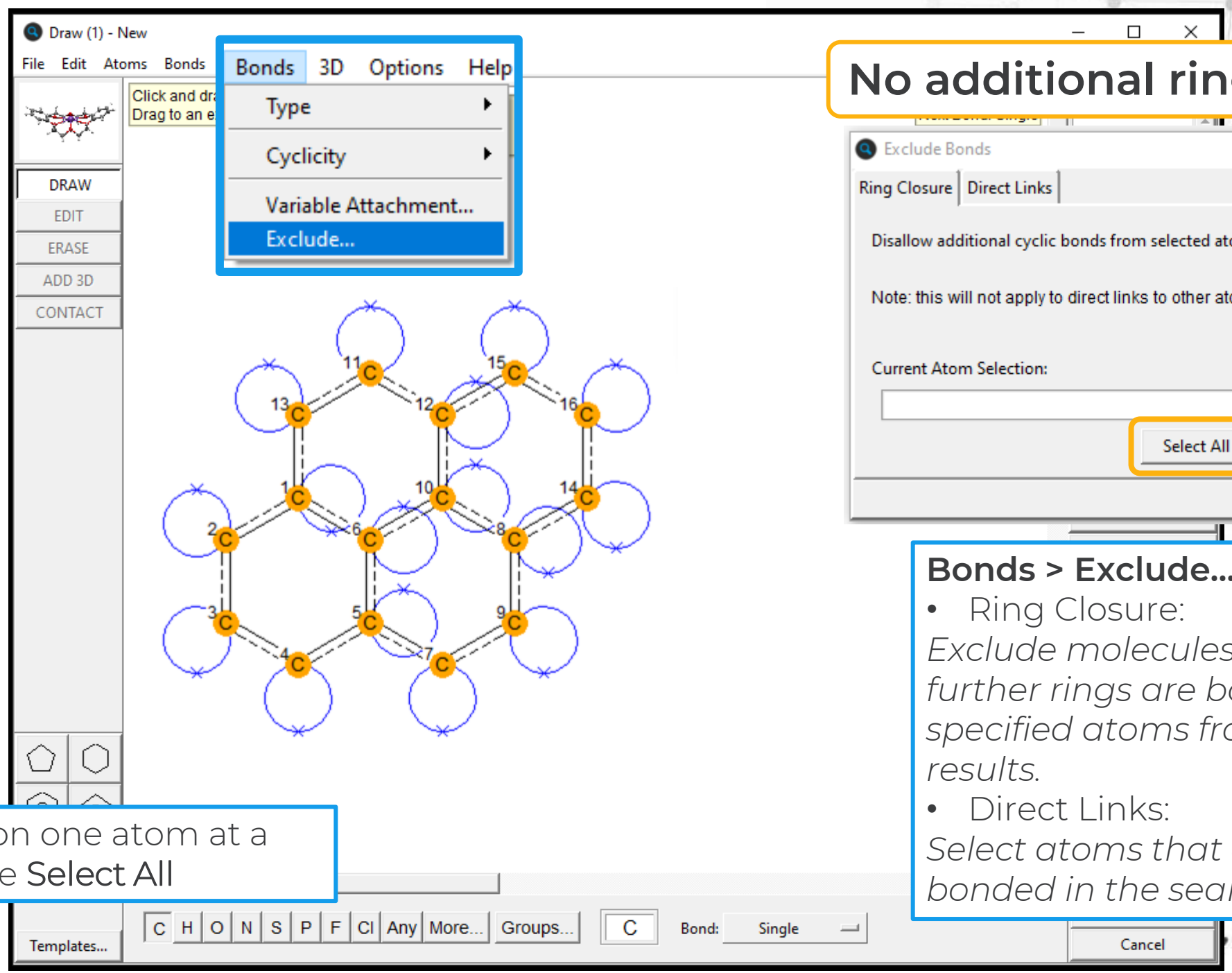
pre-defined selections

personalised selection



# How can I be more specific in my draw searches?






# No additional rings




**Bonds > Exclude...**

- Ring Closure:

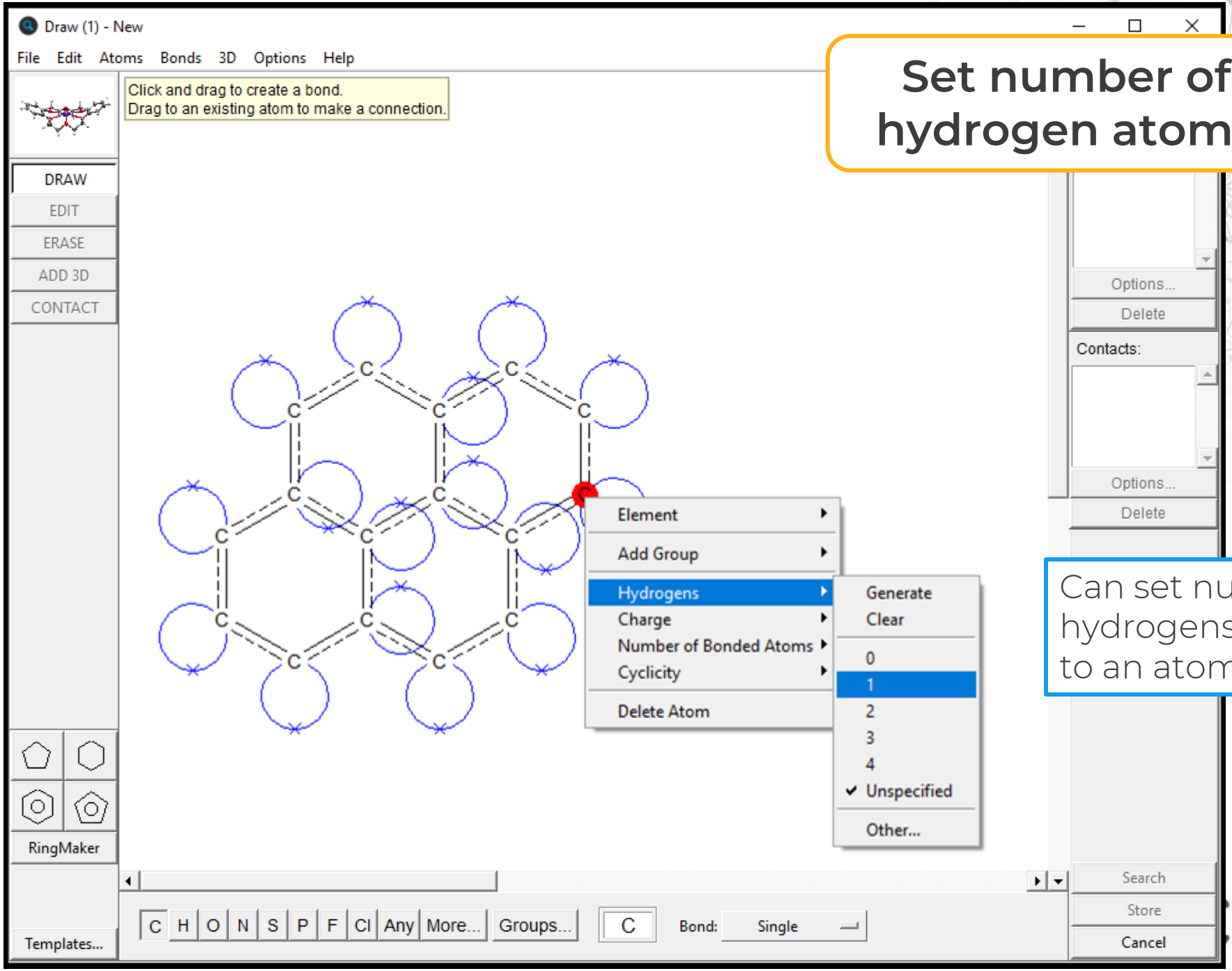
*Exclude molecules where further rings are bonded to specified atoms from search results.*

- Direct Links:

*Select atoms that cannot be bonded in the search results.*

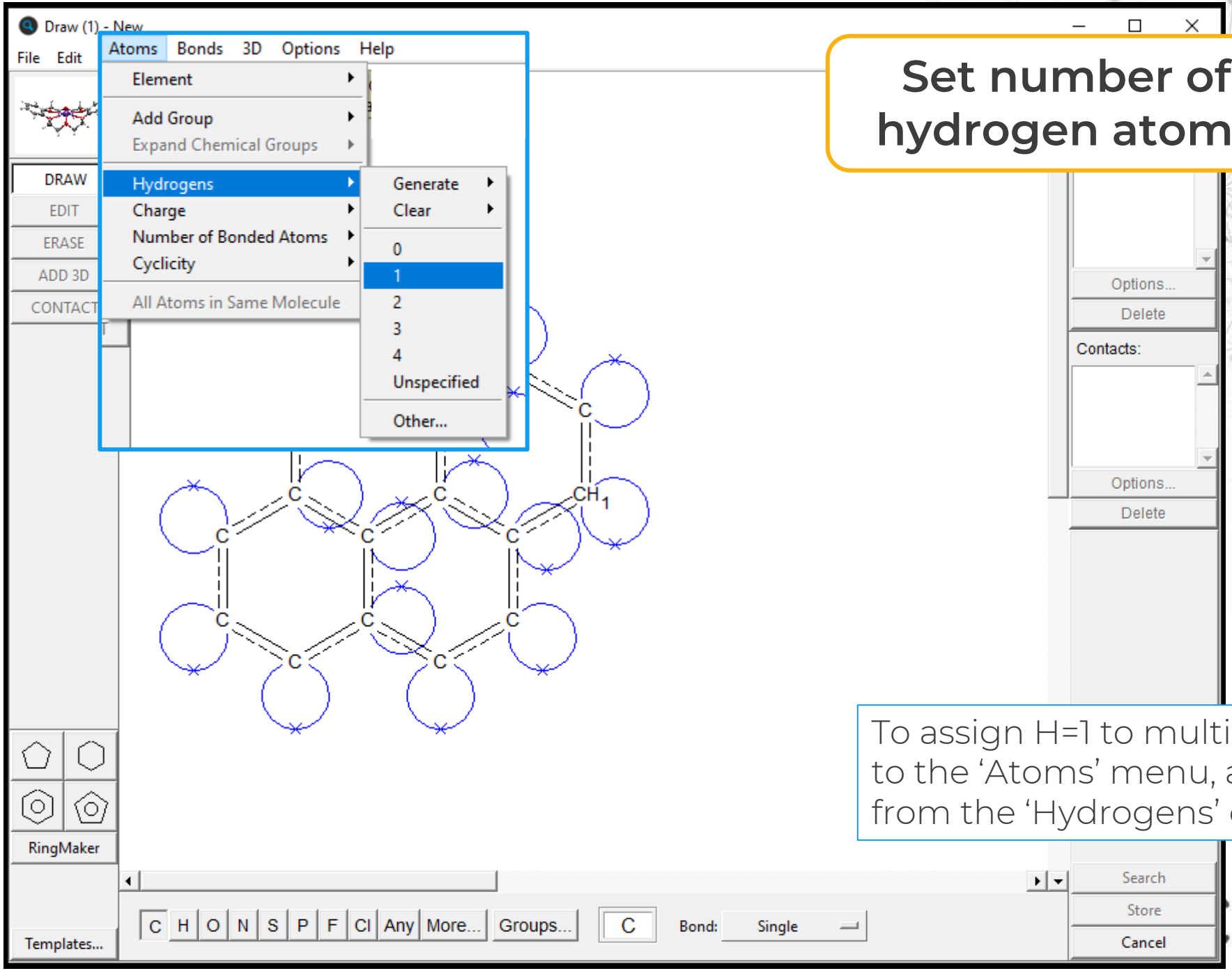
Can click on one atom at a time or use **Select All**

# Set number of hydrogen atoms



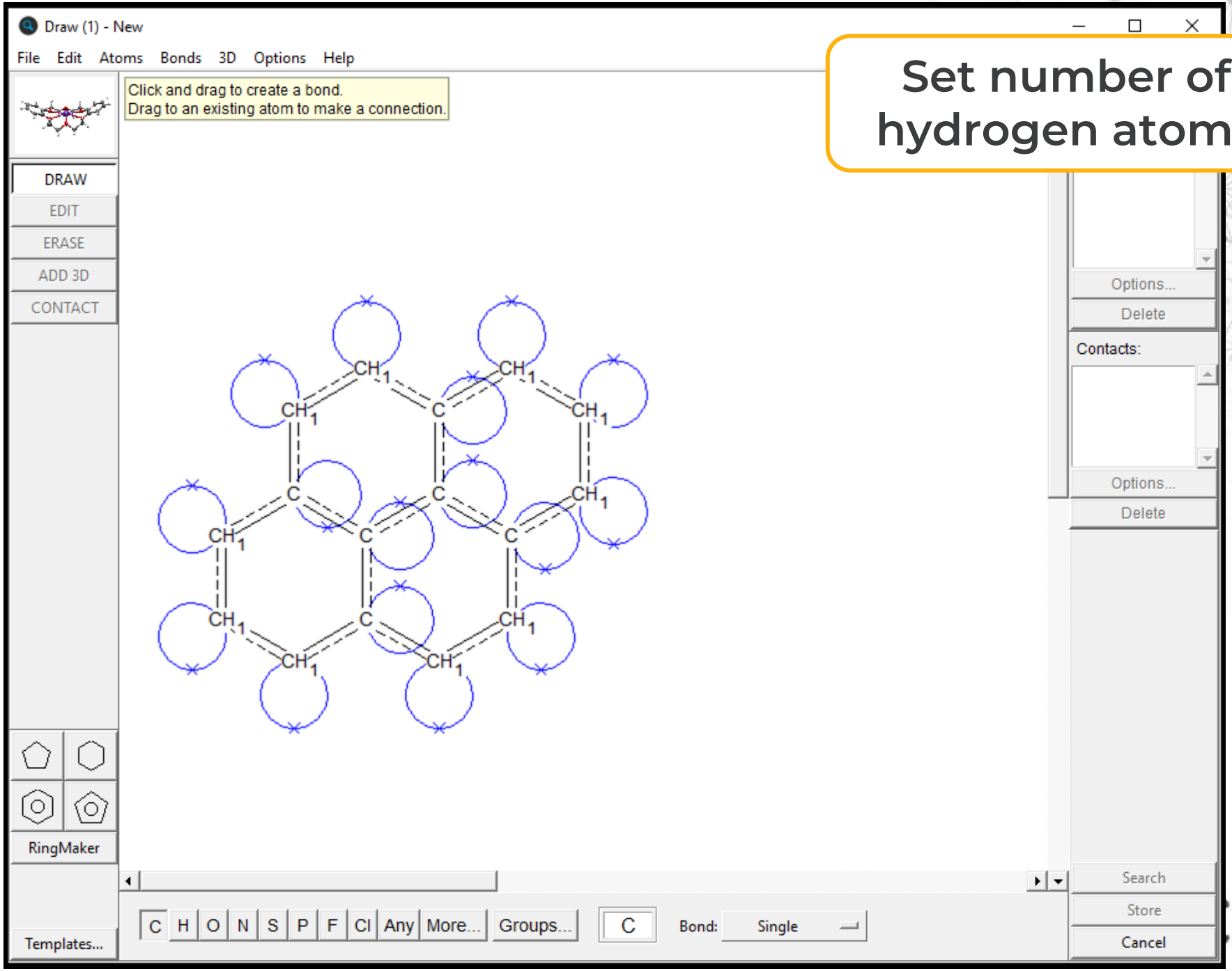

Can set number of hydrogens bonded to an atom.

# Set number of hydrogen atoms

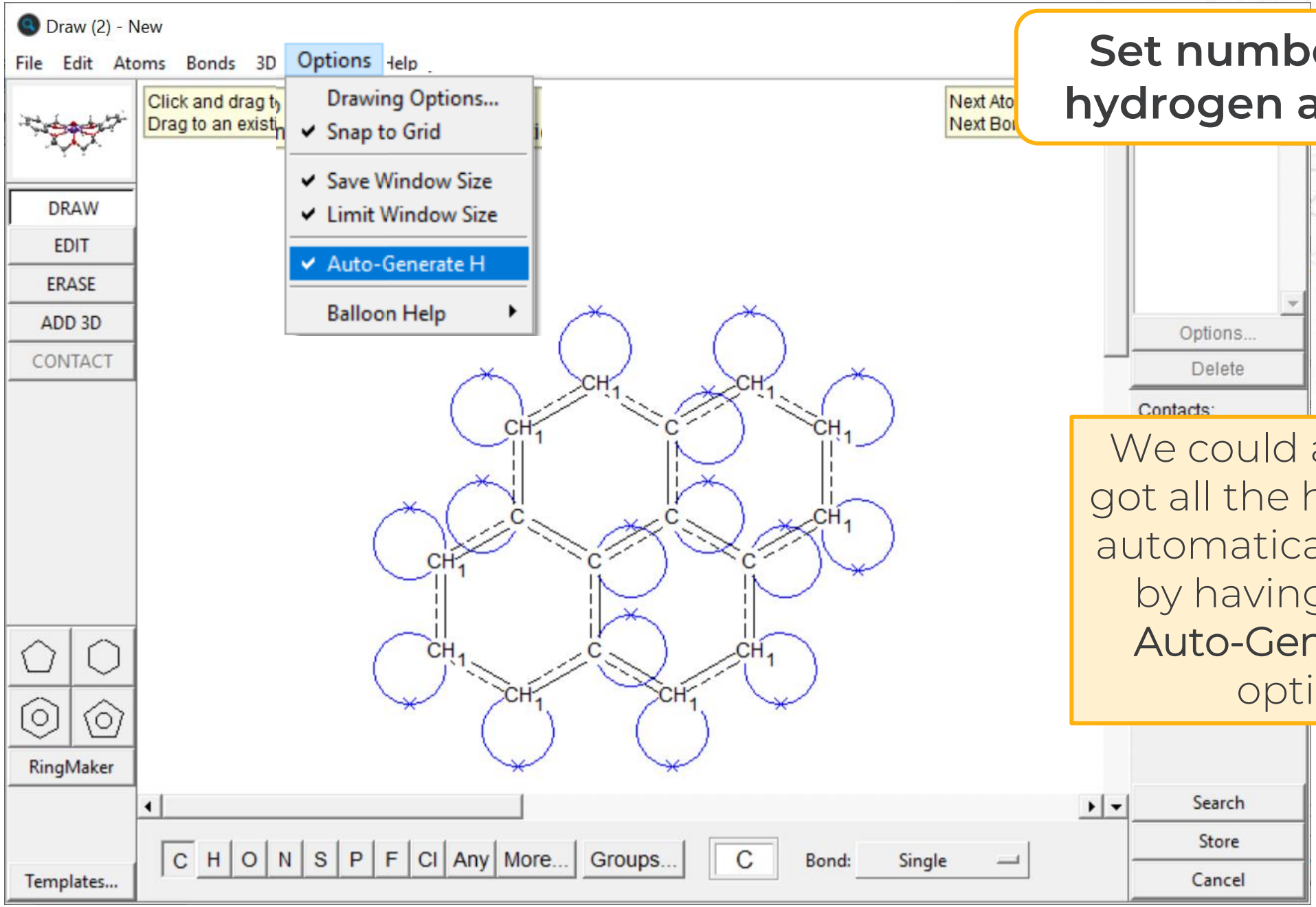



To assign H=1 to multiple atoms, go to the 'Atoms' menu, and select 1 from the 'Hydrogens' option.

# Set number of hydrogen atoms



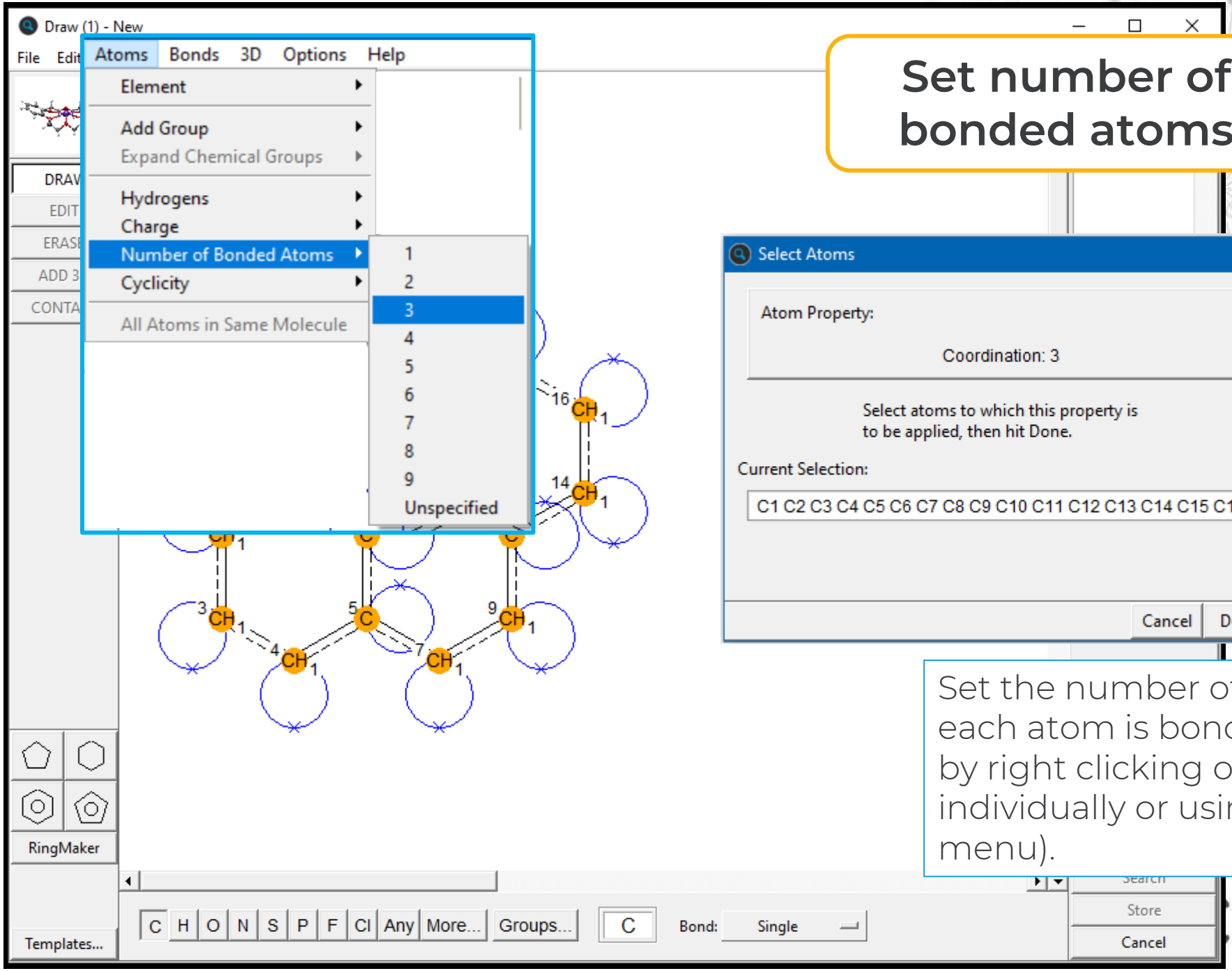
# Set number of hydrogen atoms



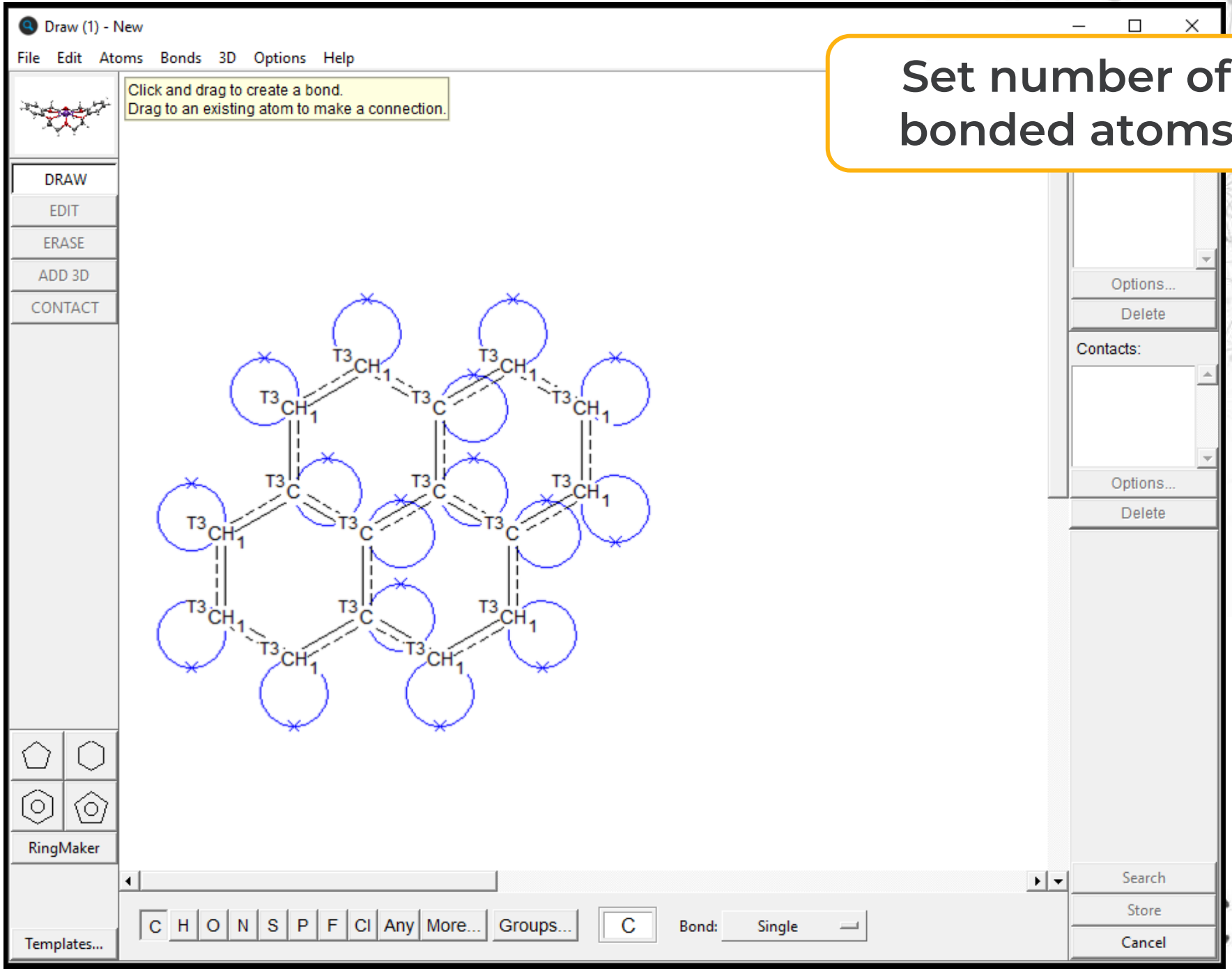

We could also have got all the hydrogens automatically added by having on the **Auto-Generate H** option

CCDC

# Set number of bonded atoms



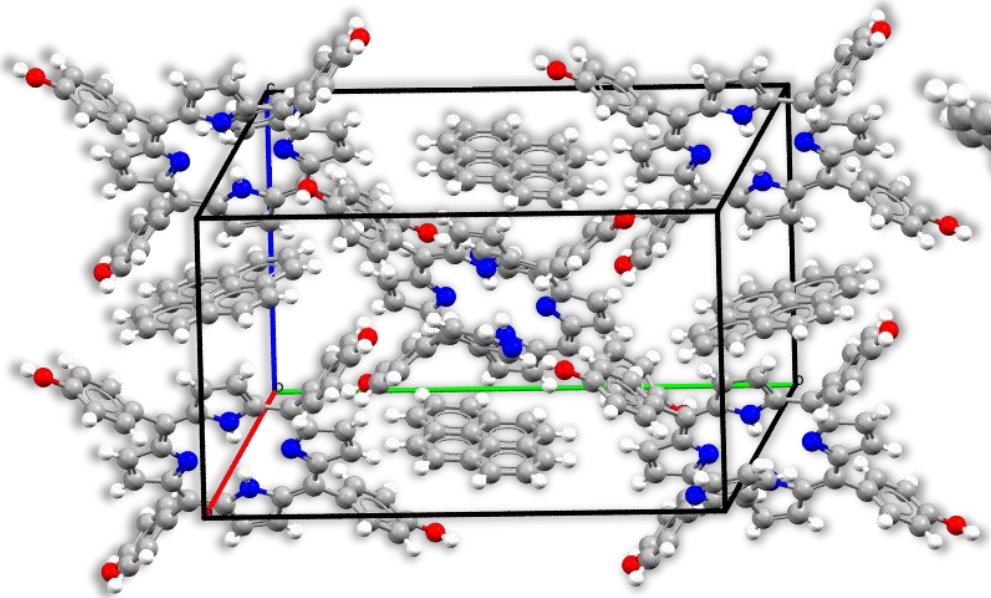

Set the number of other atoms each atom is bonded to (either by right clicking on atoms individually or using the atoms menu).



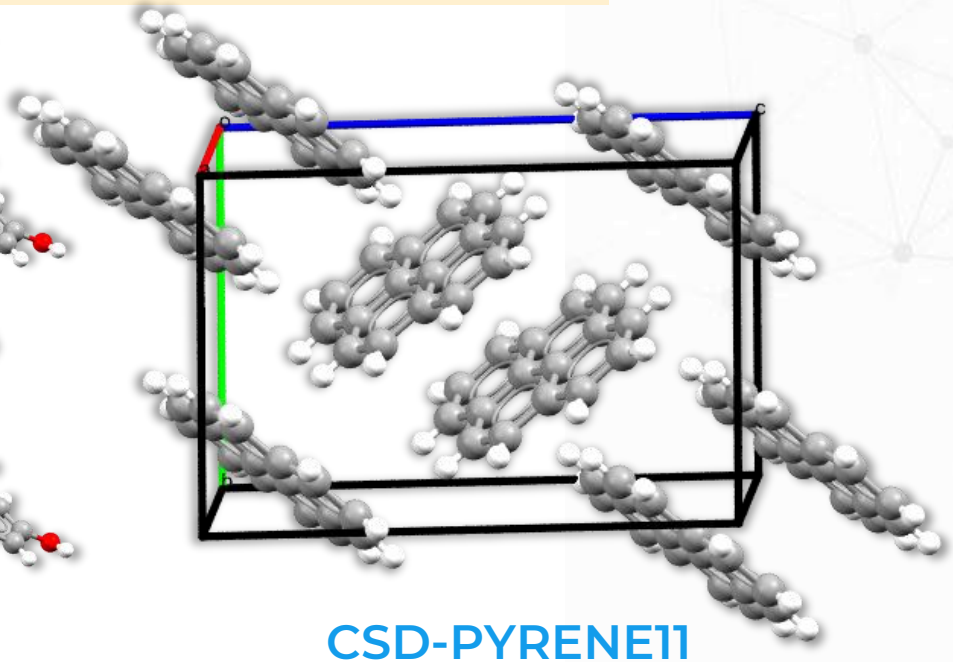

# Set number of bonded atoms

CCDC

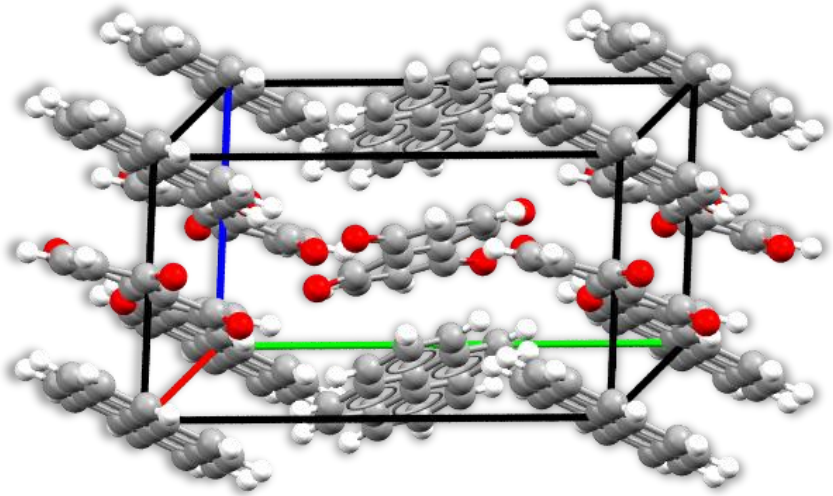
We have found pyrene and multicomponent substances containing pyrene.



CSD-ELUGUP



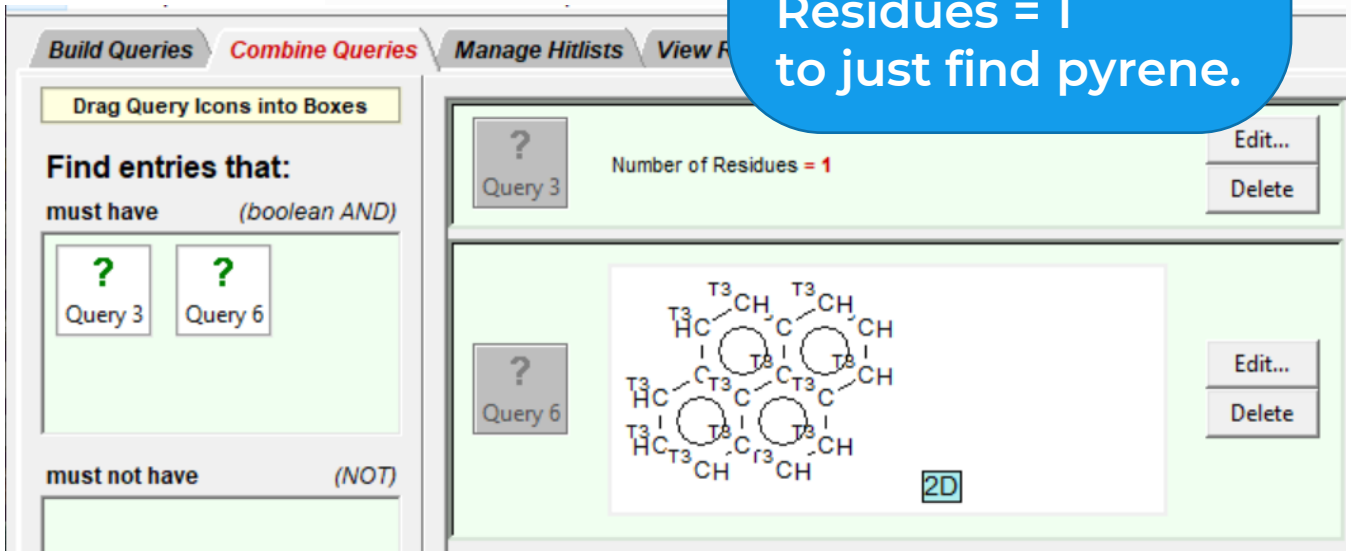
CSD-PYRENE11



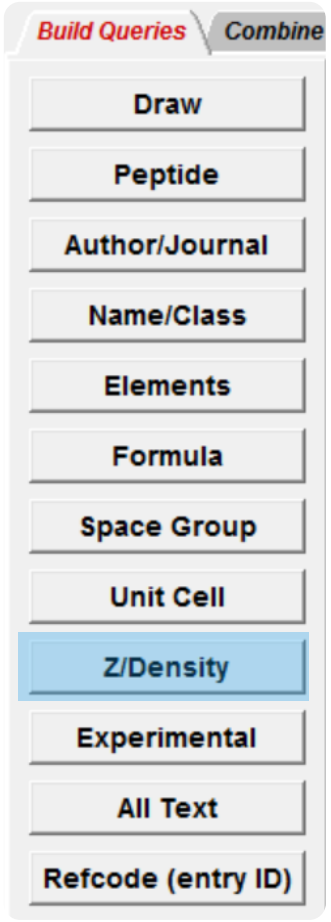
CSD-CEKBUP

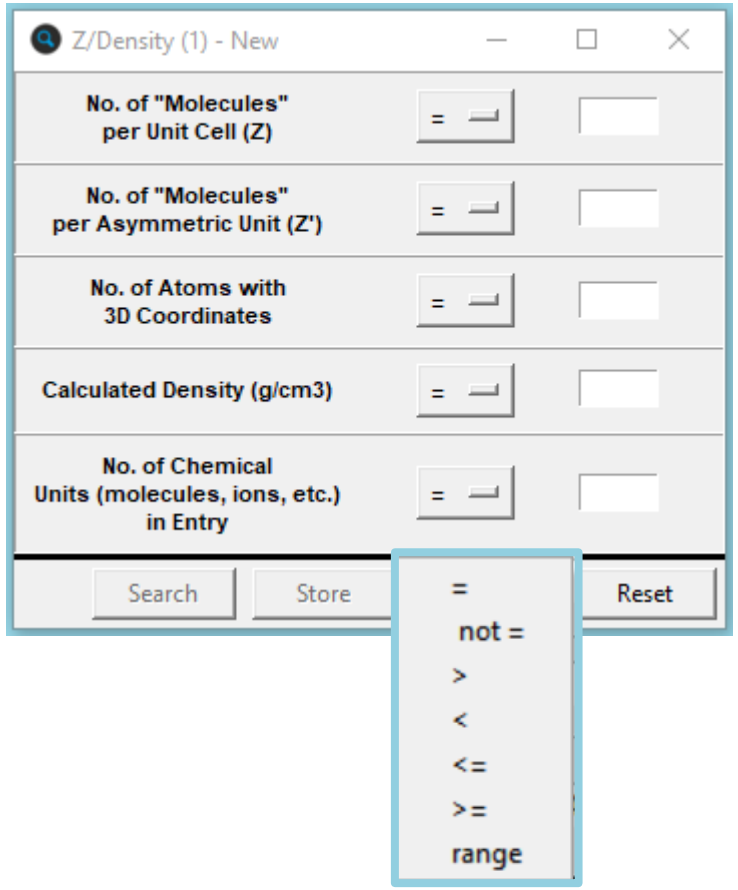
## Tip

Do you need only pyrene structures?
Use
Combine Queries
to combine this query with a query for
Number of Residues = 1
to just find pyrene.



# What if I only want a certain number of molecules in the crystal structure?




You can use the Z/Density search to set the number of molecules found in the structure.

Then use the Combine Queries tab to create a search that includes both the Pyrene molecule and the Z/Density restriction.

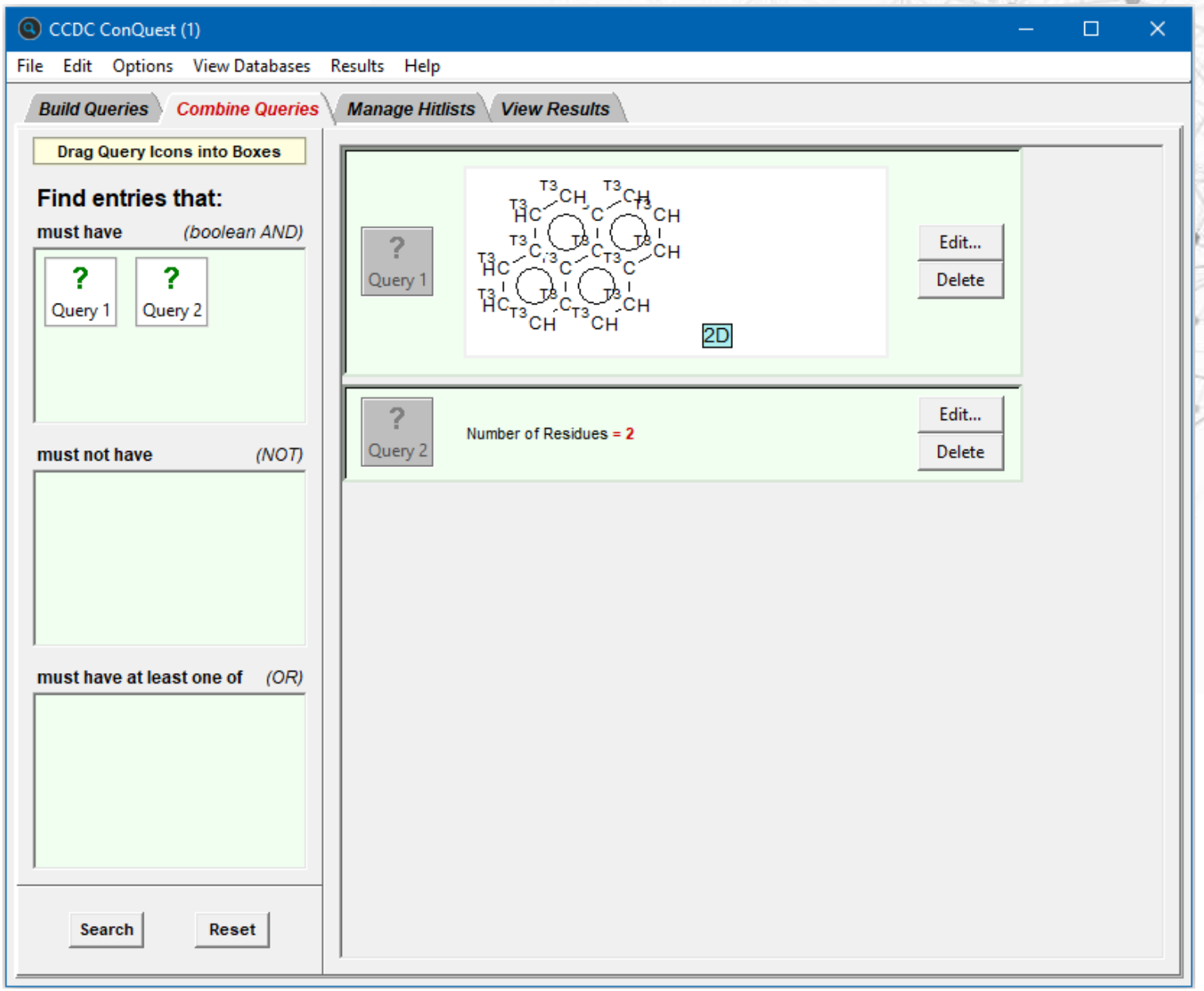
CCDC

# Combine Queries


Allows you to combine different queries using boolean AND, NOT, OR

Query 1: Pyrene query
Query 2: Number of residues = 2

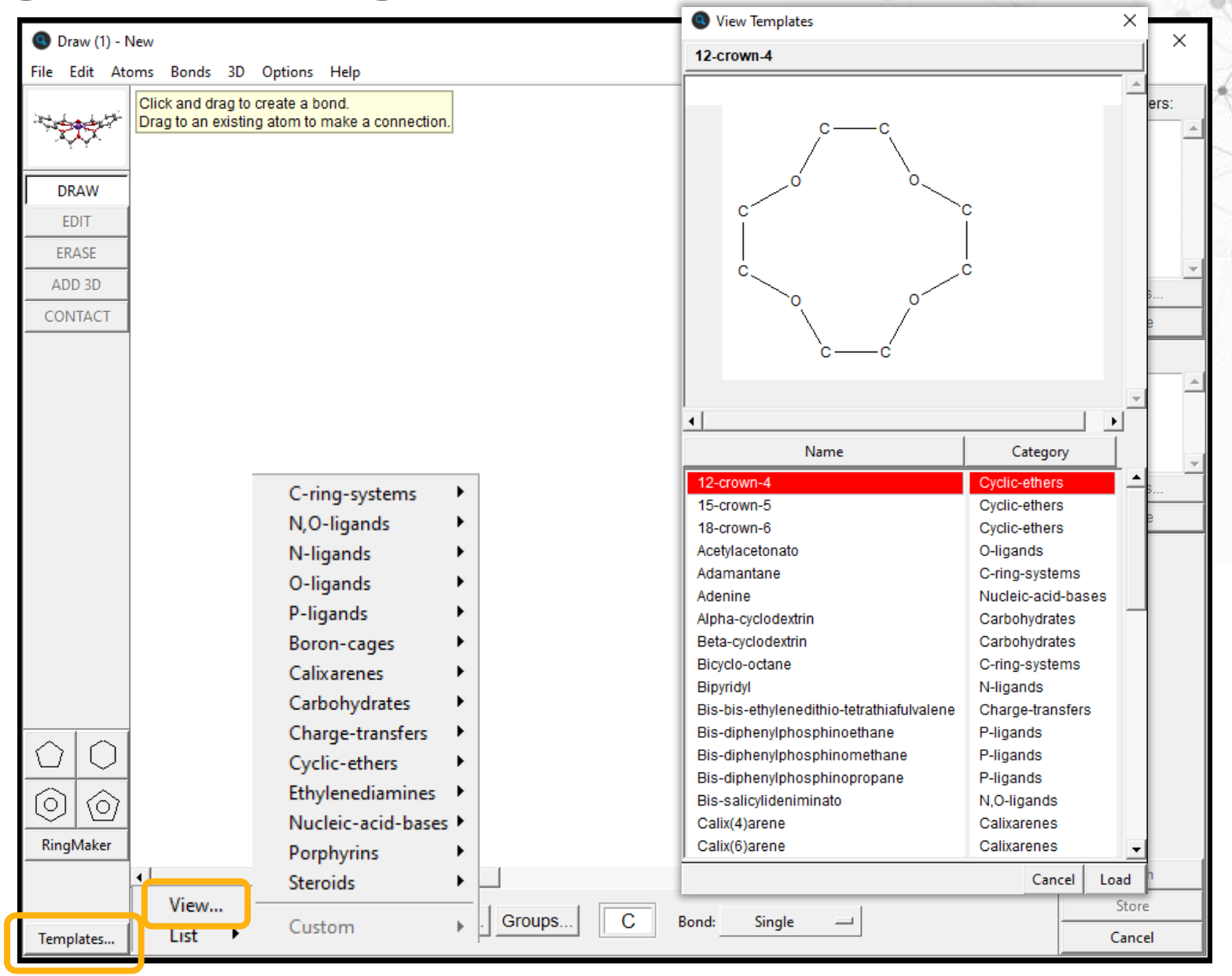
**Must have:** Query 1 **AND** Query 2



# *Explore More:* Tips & Tricks

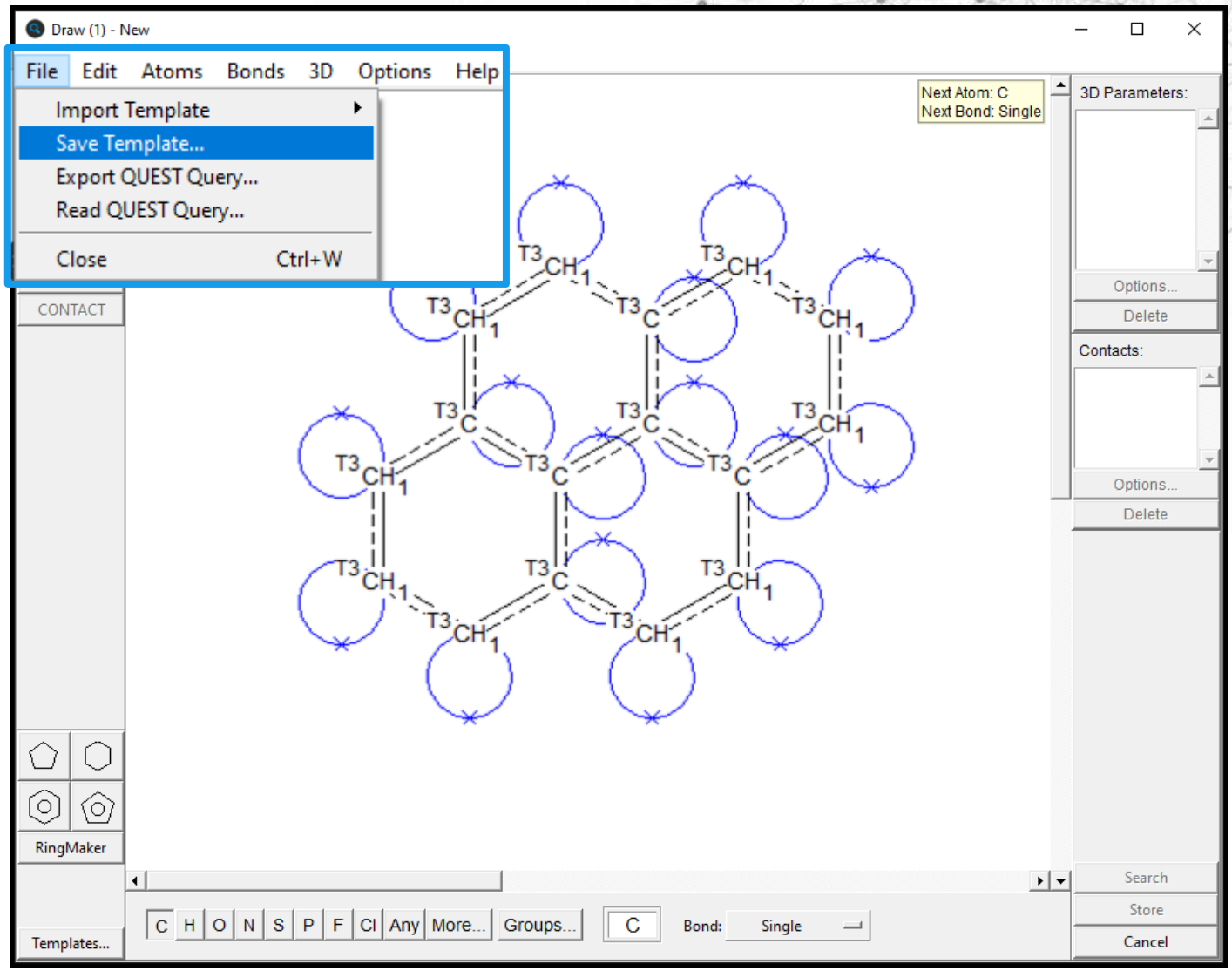
- Using and creating templates
- Finding structures with different functional group positions
- Customising your searching
- Searches for structures on the web
- Finding similar structures of interest

# Making Drawing Easier: Templates



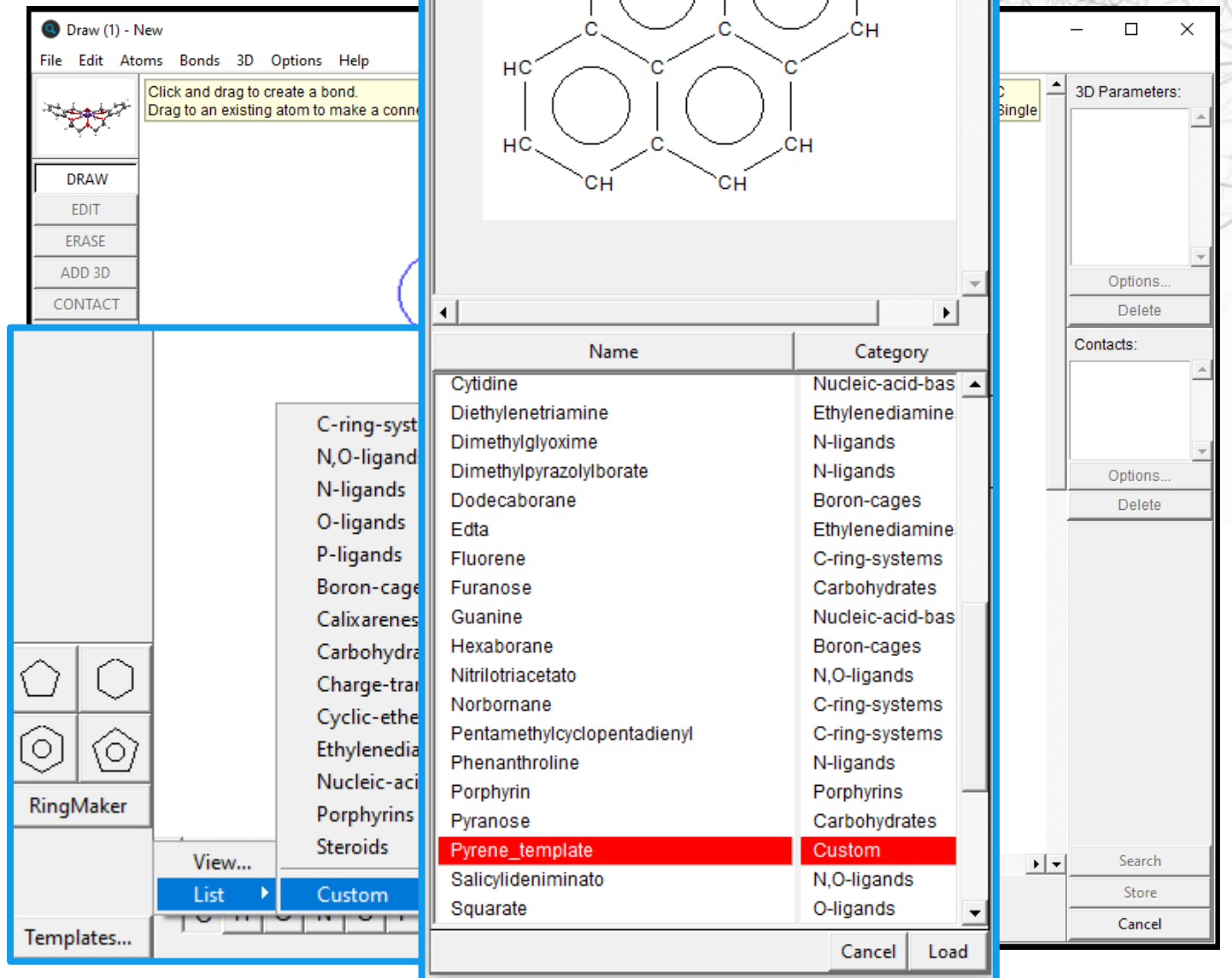
# Create your own Templates

1. Draw molecule
2. Save the Template
   - To access through the ConQuest custom template list, save the file in the csds_data\templates folder in your User Area.



# Create your own Templates

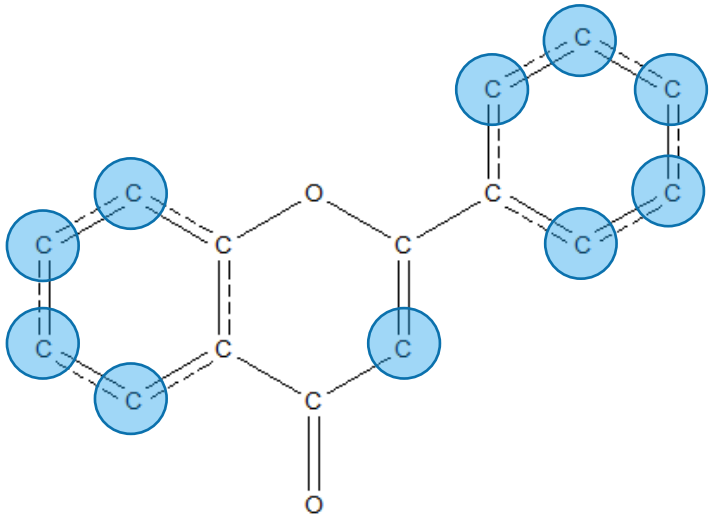
1. Draw molecule
2. Save the Template
   - To access through the ConQuest custom template list, save the file in the csds_data\templates folder in your User Area.
3. Template can now be accessed through Templates button in either ‘List’ or ‘View’

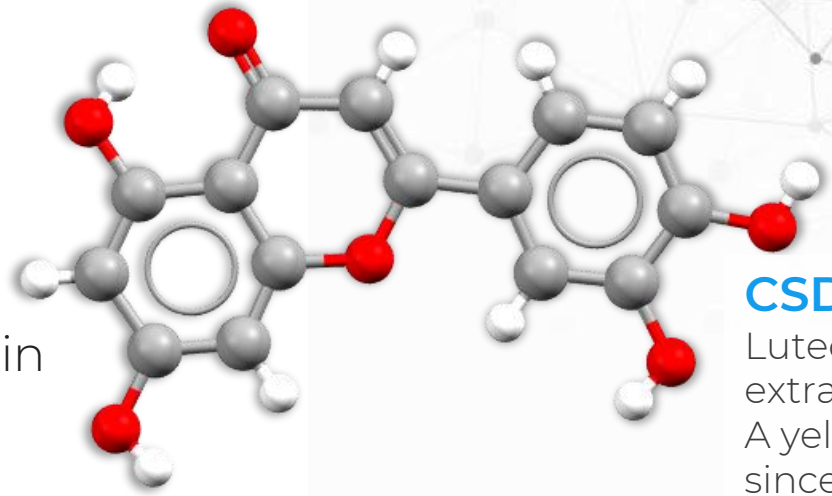


# Do you want to search for structures that can have functional groups in different places?

## Example: Flavones

- A class of flavonoids, having a 2-phenylchromen-4-one backbone with functional groups (e.g. OH, OMe).
- Many naturally occurring flavones found in plants and fungi (with hydroxy and/or methoxy substituents) – but synthetic flavones have also been created.



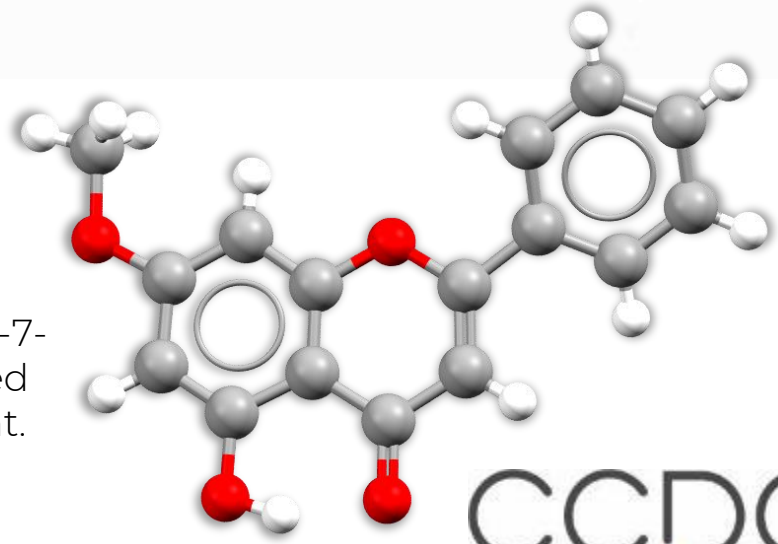



**CSD-OJEQUP**

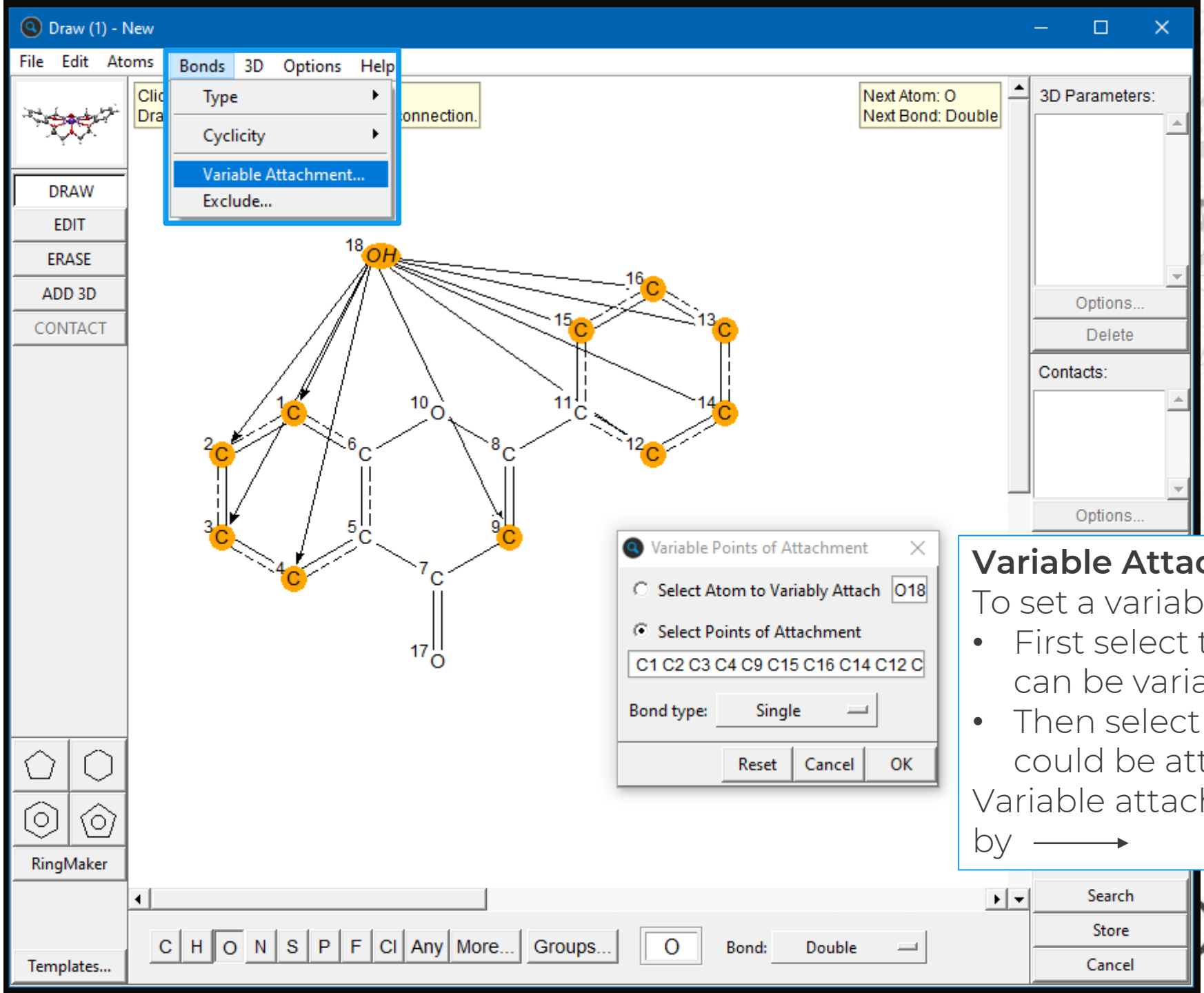
Luteolin (tetra-hydroxy flavone), extracted from *Reseda luteola*. A yellow dye that has been used since 1st millennium BC.

**CSD-BAZGOB**

Techtochrysin (5-hydroxy-7-methoxyflavone) extracted from the sour cherry plant.



CCDC

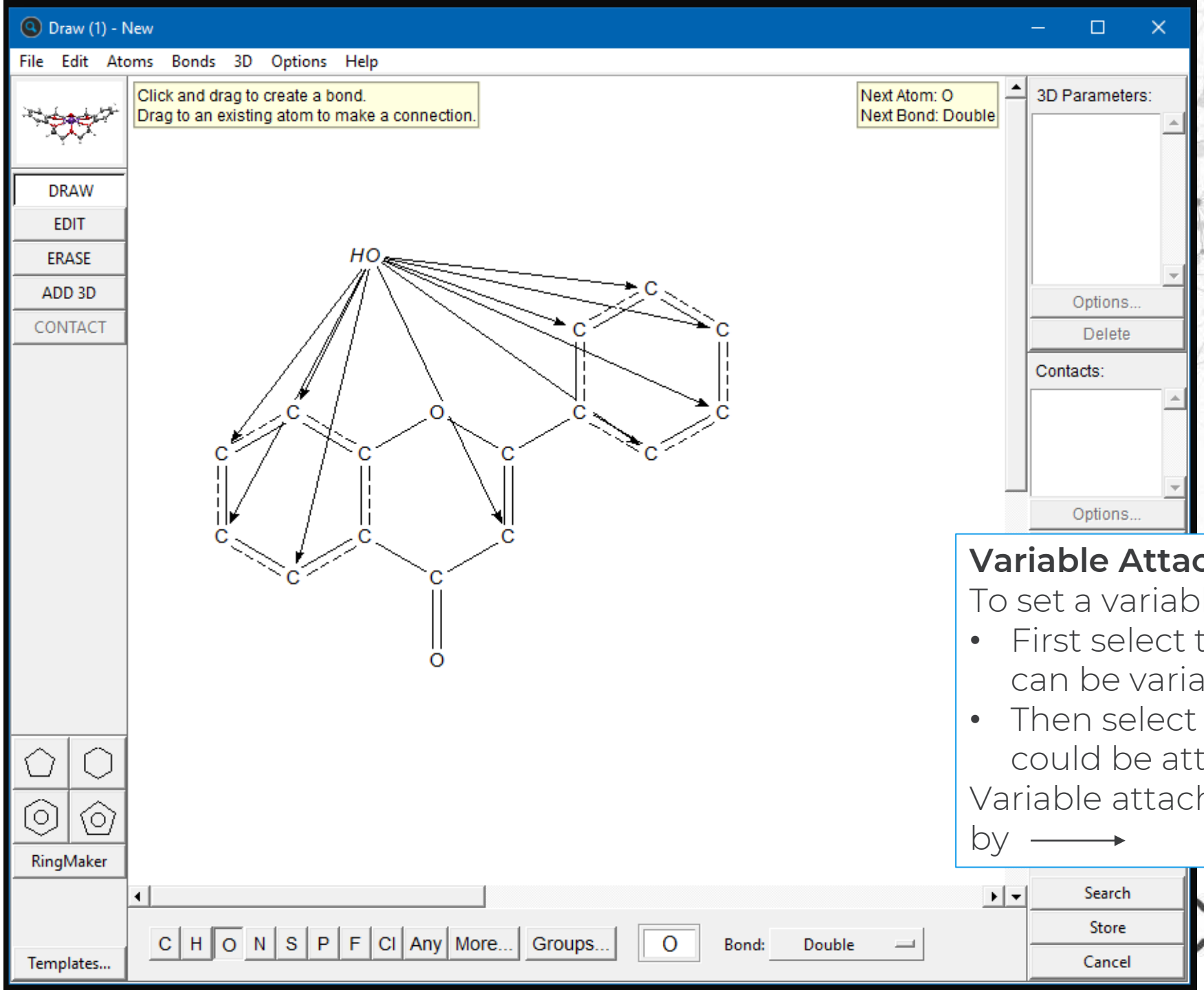



## Variable Attachments

To set a variable attachment:

- First select the atom that can be variably attached.
- Then select the places it could be attached to.

Variable attachments shown by ⟶




## Variable Attachments

To set a variable attachment:

- First select the atom that can be variably attached.
- Then select the places it could be attached to.

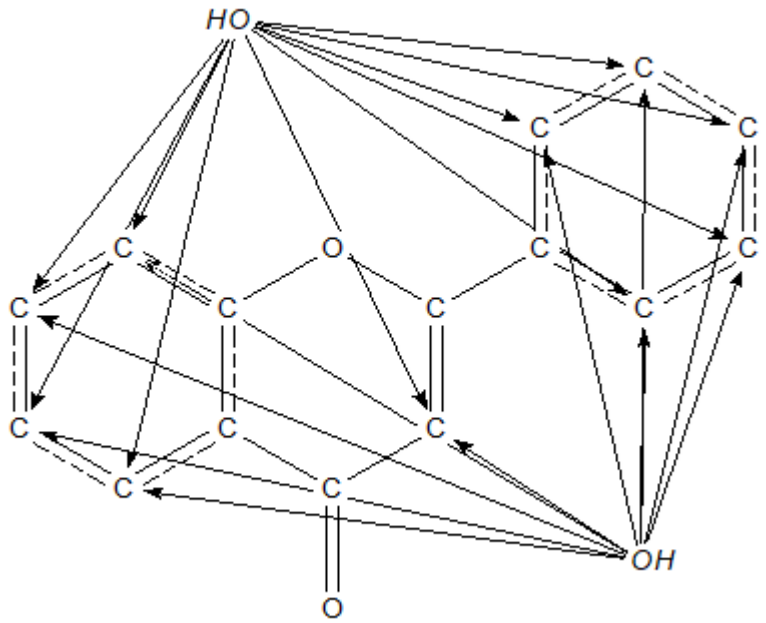
Variable attachments shown by ⟶

Draw (1) - New

File Edit Atoms Bonds 3D Options Help

Click and drag to create a bond.
Drag to an existing atom to make a connection.

Next Atom: O
Next Bond: Double




DRAW
EDIT
ERASE
ADD 3D
CONTACT

RingMaker

Templates...

3D Parameters:
Options...
Delete

Contacts:
Options...
Delete

Search
Store
Cancel

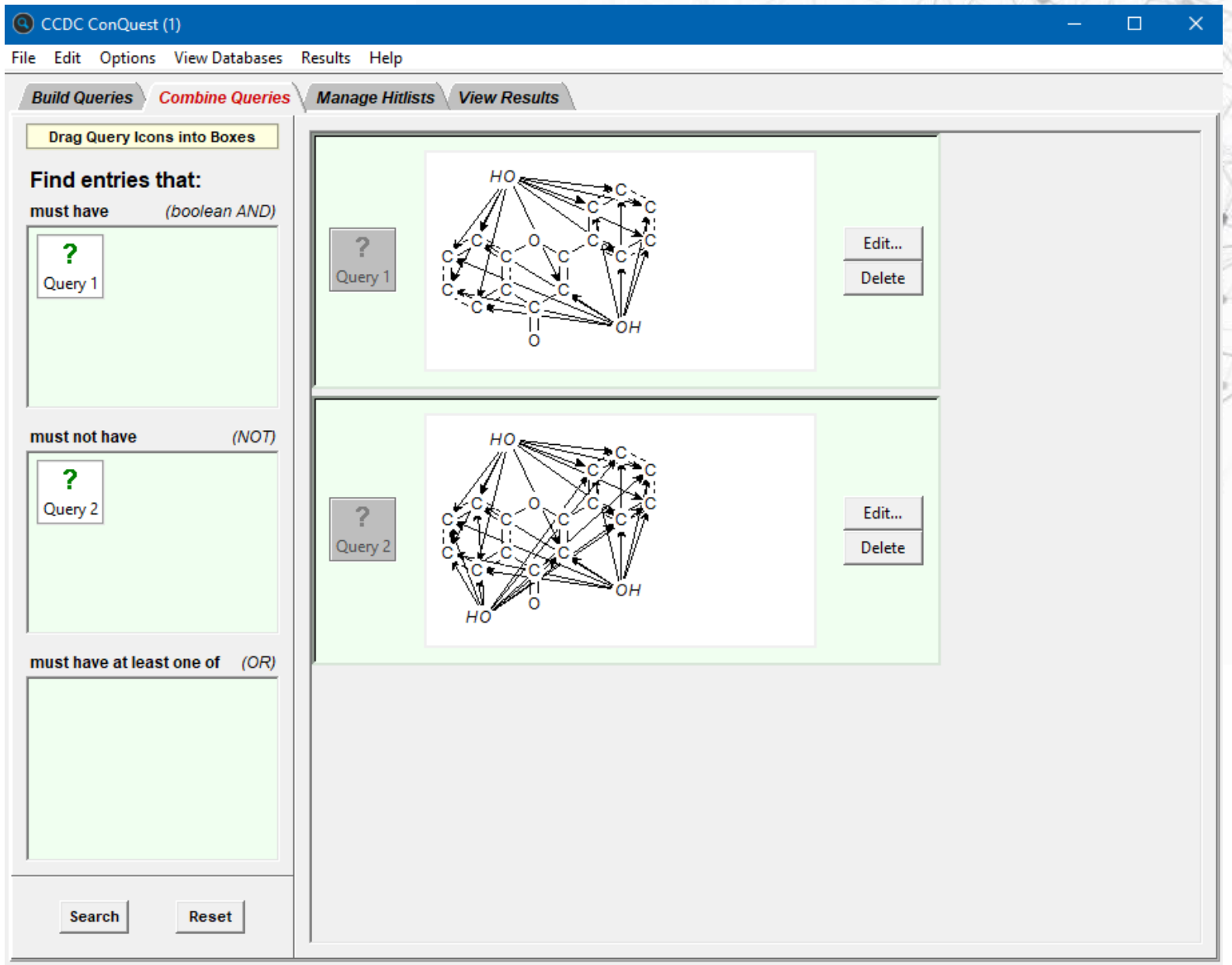
C H O N S P F Cl Any More... Groups... O Bond: Double

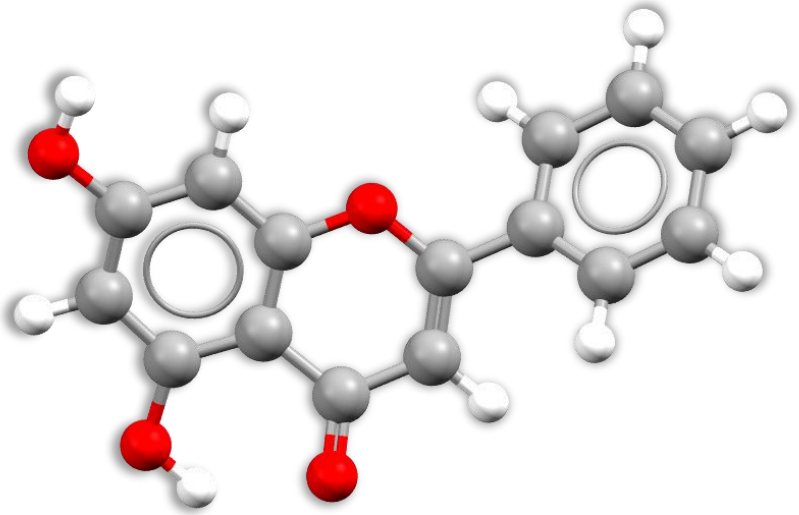
- How do I **only** find di-hydroxy flavones?

## Combine Queries

Query 1: 2 variably attached OH
Query 2: 3 variably attached OH

Must have: Query 1
Must not have: Query 2



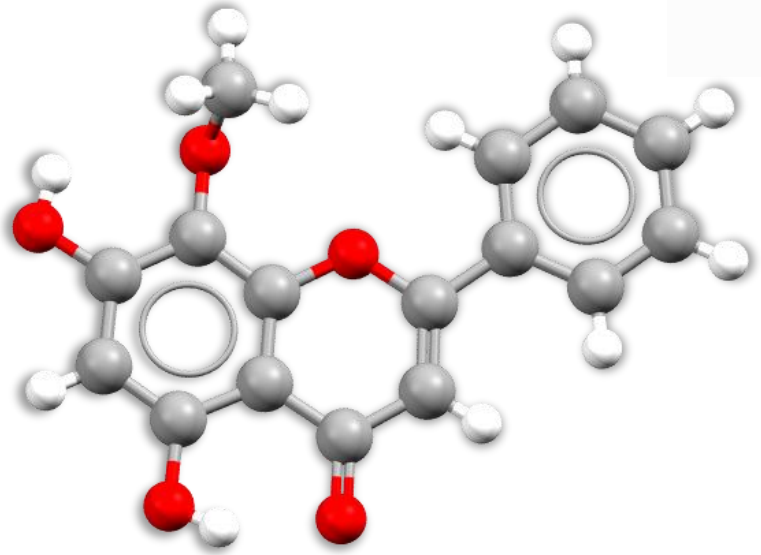


**CSD-IYIWEU**

Chrysin, 5-7-dihydroxyflavone, is found in honey, propolis and passionflowers.

**CSD-AFOCUV**

5,6-Dihydroxy-7,8-dimethoxyflavone, Isolated from a rare traditional Chinese medicinal herb, *Saussurea involucrate*

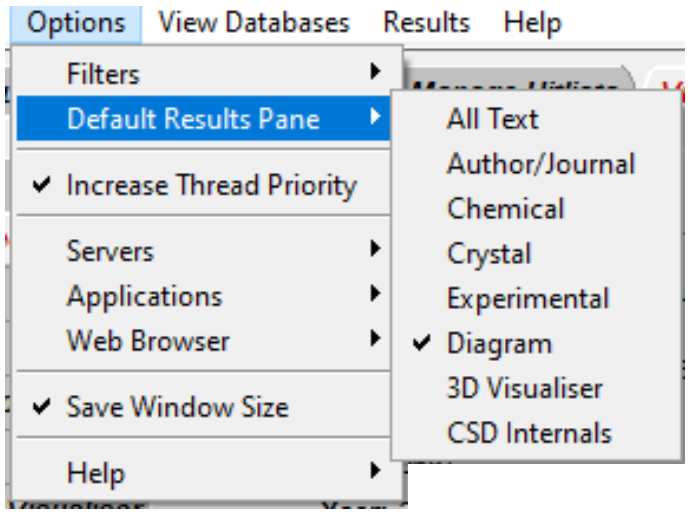


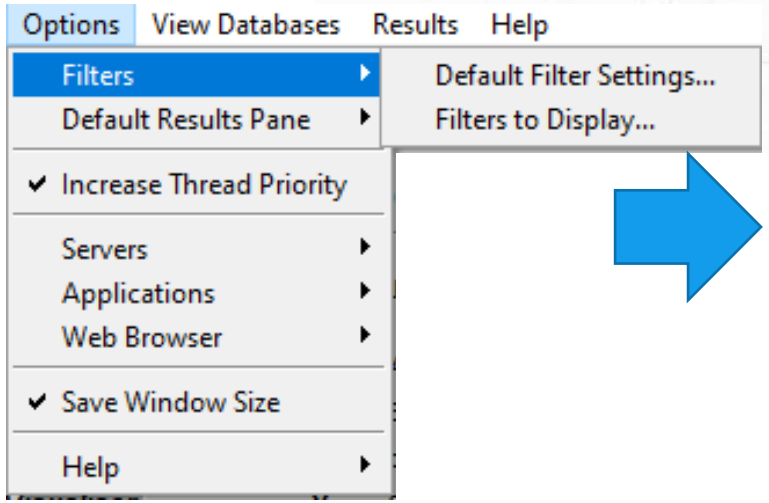
**CSD-PUZLII**

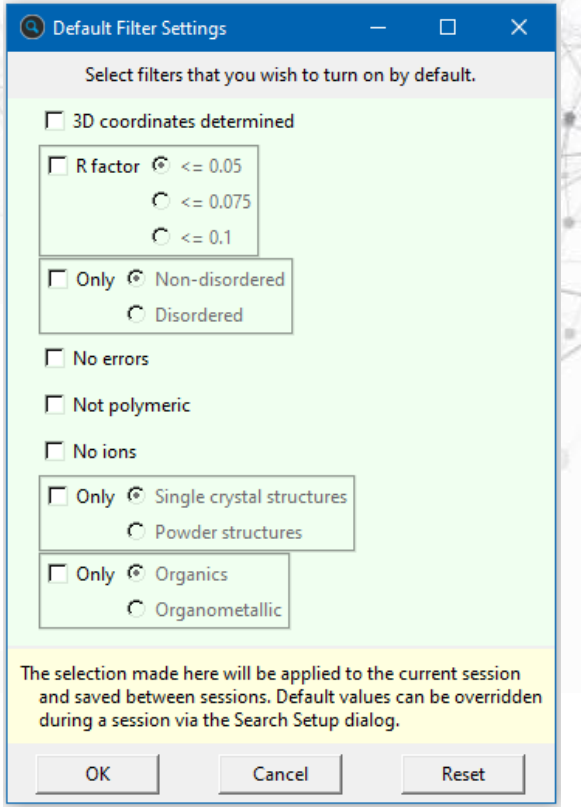
Wogonin, 5,7-Dihydroxy-8-methoxyflavone, extracted from *Scutellaria baicalensis*. It has been investigated for many different biological activities.

CCDC

# Customise your searching



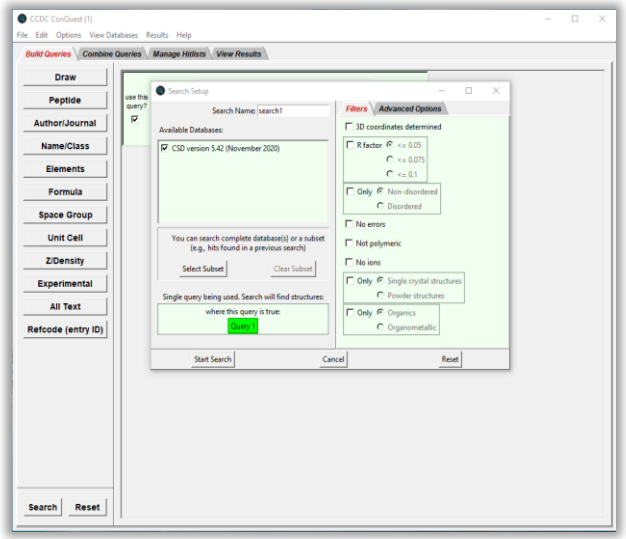

Have a results panel you prefer to see first in ConQuest?
Select it in:
Options>Default Results Pane

Always using the same search filter settings?
Change your default settings in: Options> Filters> Default Filter Settings...

CCDC

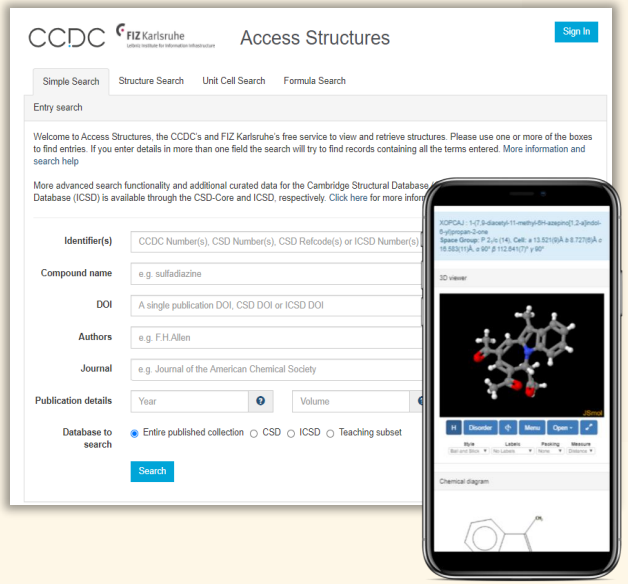
# Finding fantastic structures

## Desktop



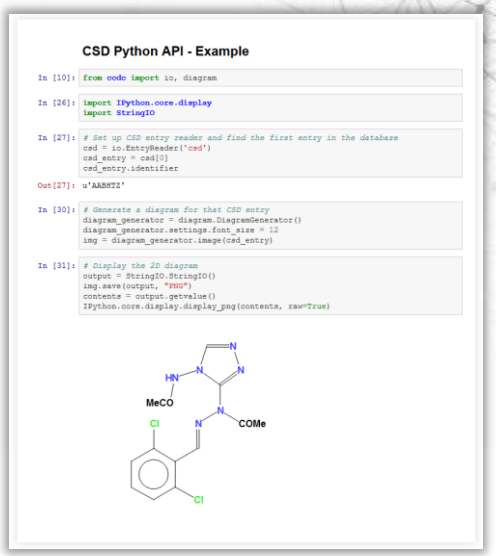

ConQuest

## Web Browser




WebCSD / Access Structures

## Programmatic

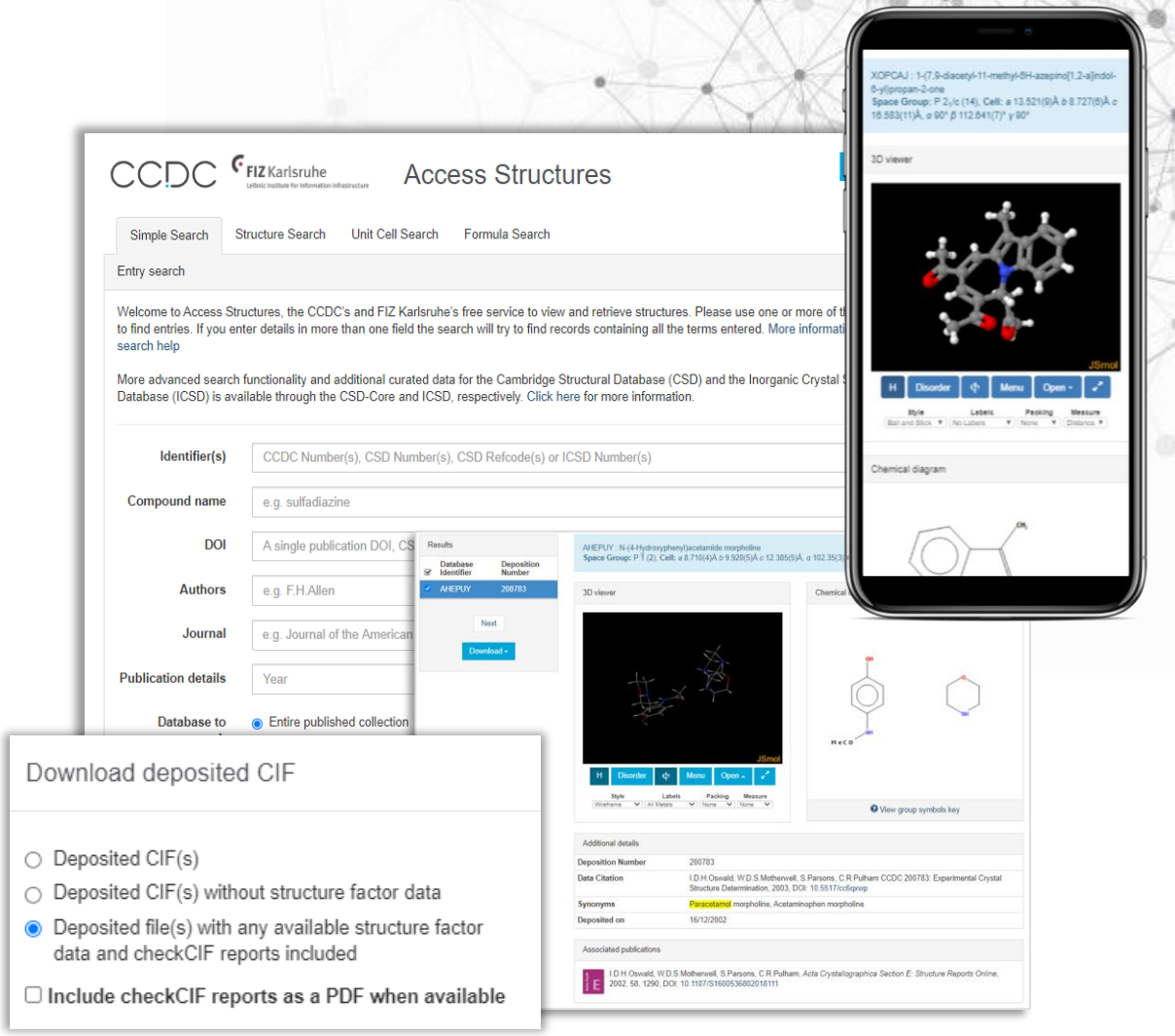



CSD Python API

CCDC

# Access Structures

- Online portal to access the CSD and ICSD
- Free to view and retrieve individual structures
- Basic searching
- Accessed through any standard internet browser – no local installation of software required



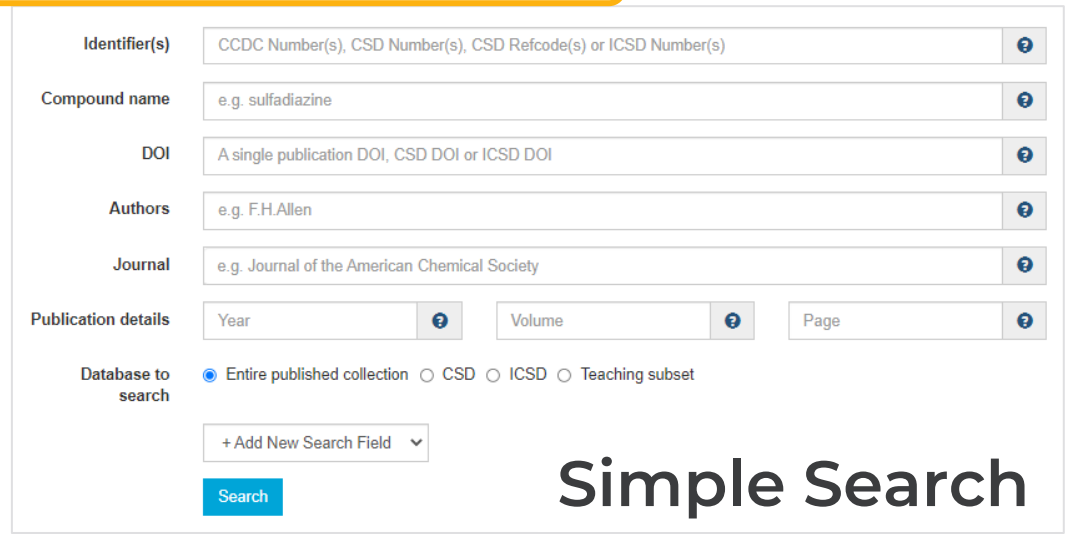

www.ccdc.cam.ac.uk/structures

CCDC

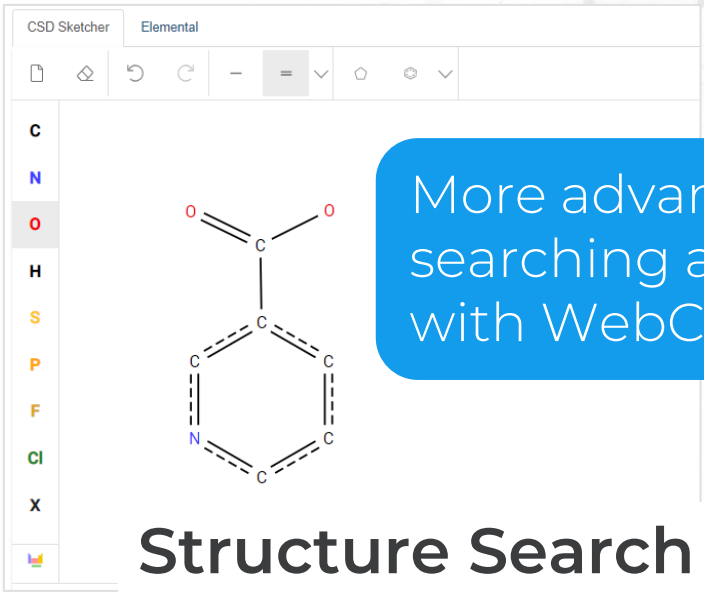
# WebCSD

Simple Search | Structure Search | Unit Cell Search | Formula Search

## Access Structures and WebCSD



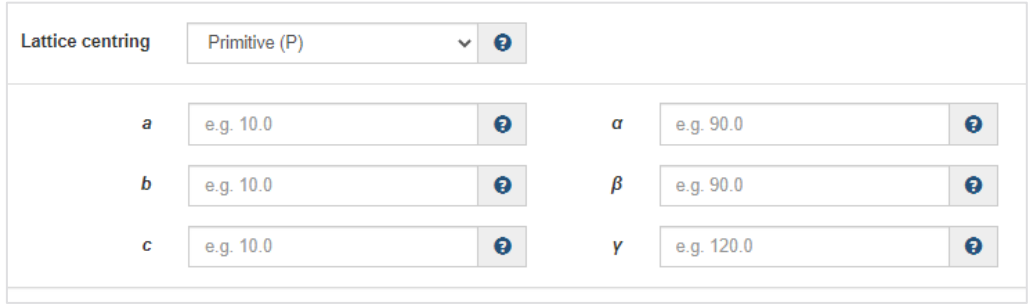

### Simple Search




More advanced searching available with WebCSD

### Structure Search

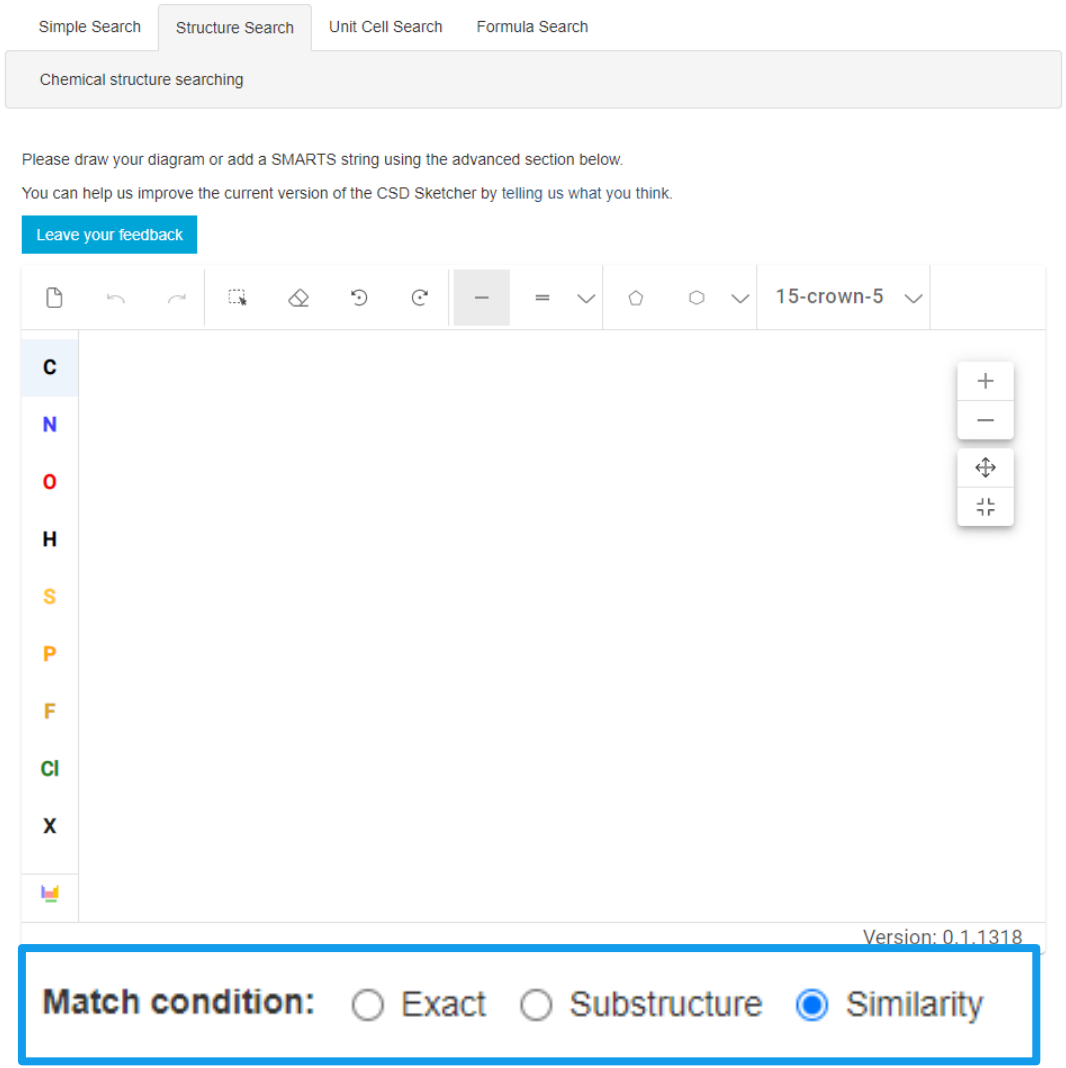
### Unit Cell Search




### Formula Search


CCDC

# Structure Search




## Exact

Returns structures that contain the exact molecule as it is drawn.

## Substructure

Returns hits where the draw query is a part of any molecule (similar to conquest).

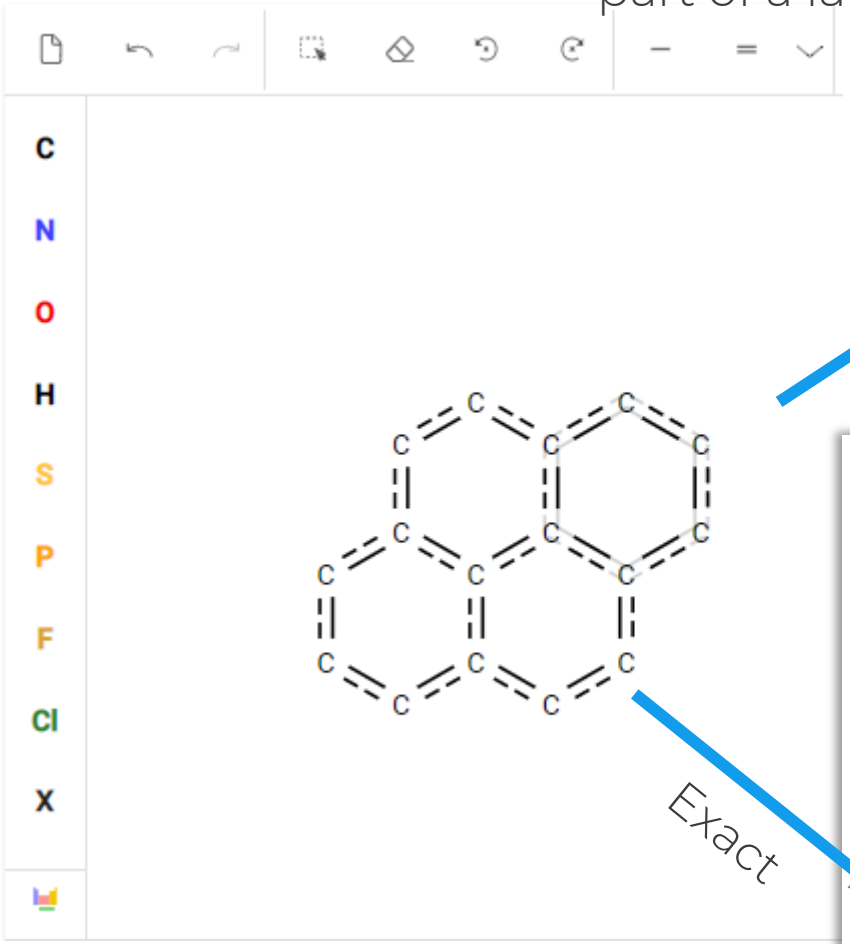
## Similarity

Calculates a molecular fingerprints using Tanimoto for the drawn molecule and compares that to pre-calculated fingerprints for structures in the CSD.

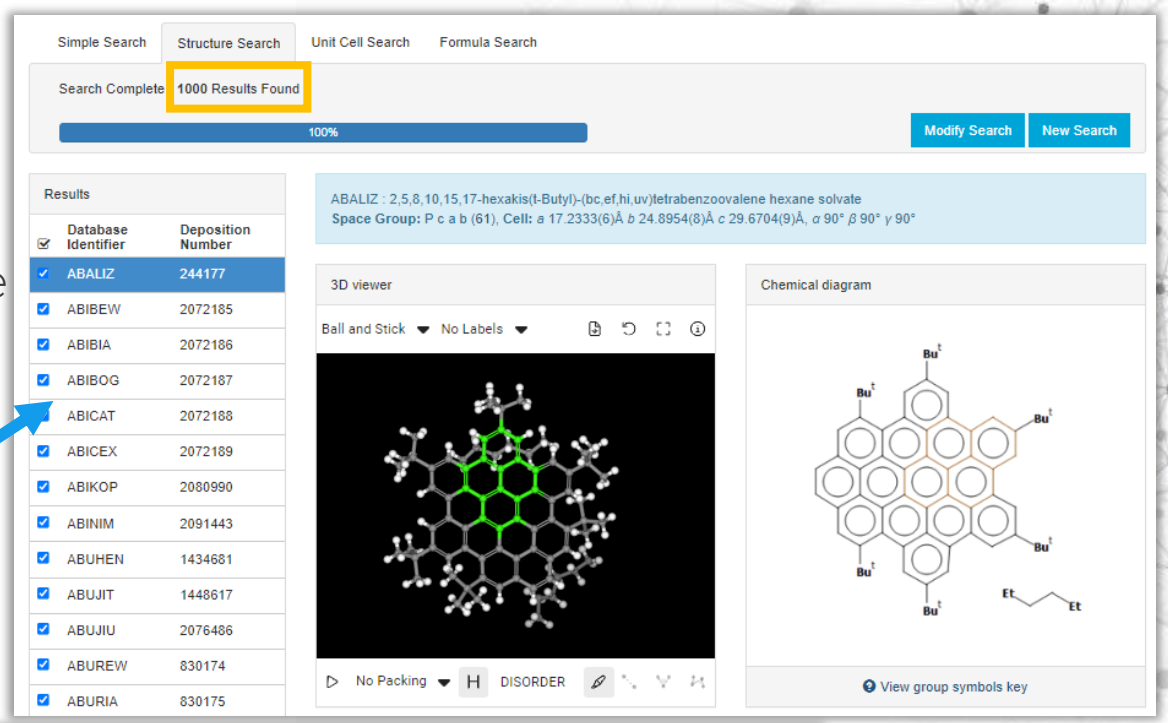
Required: Full molecule drawn (like exact search)

Returns: Similar structures and a similarity score (where 1 is an identical molecule).

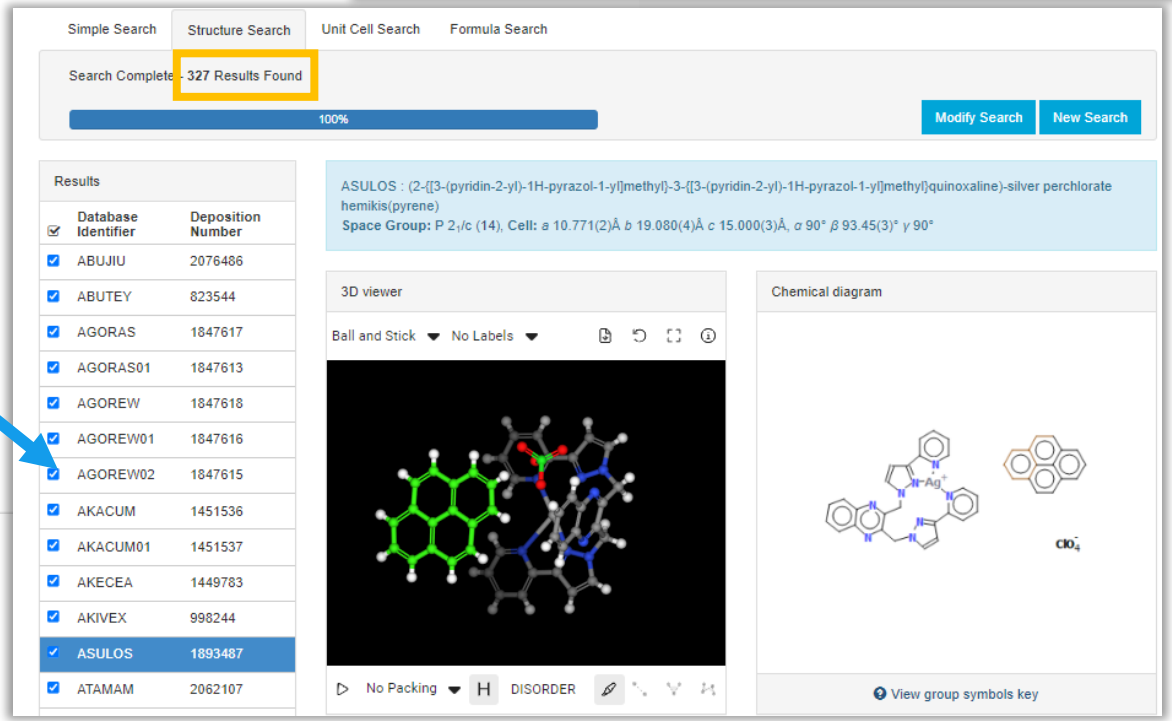
>1,000 hits
Includes structures where yrene structure is part of a larger molecule




Substructure

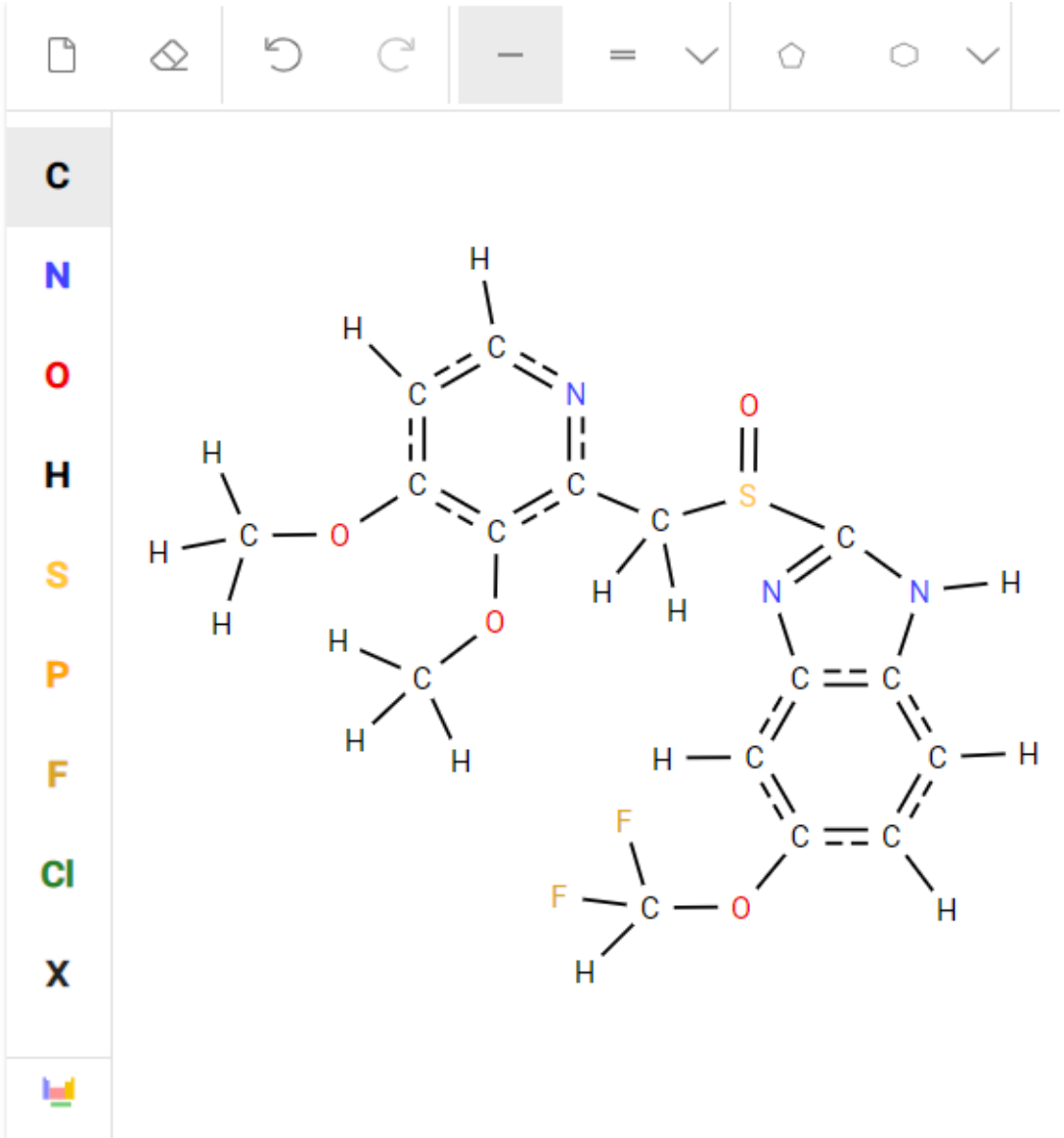



Exact

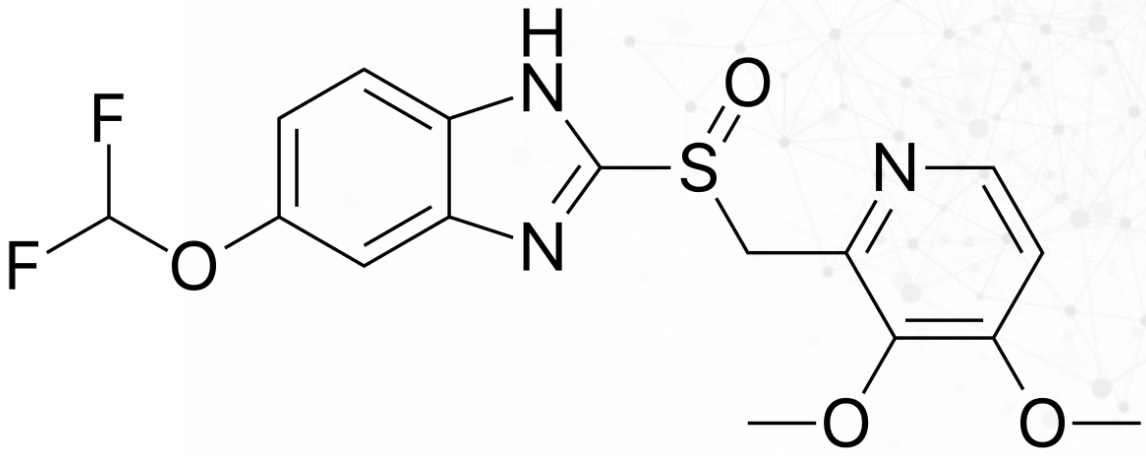



327 hits
Only structures with pyrene on its own

CCDC




Similarity



**Example: Pantoprazole**, a medication to treat stomach ulcers
Not currently in the CSD

CCDC

Results

| | Database Identifier | Deposition Number | Similarity Score More Info |
|---|---|---|---|
| ☑ | BOMQOK | 1113696 | 0.998 |
| ☑ | EYEJUP | 1511140 | 0.775 |
| ☑ | WETBIJ | 1563921 | 0.763 |
| ☑ | CENBIJ | 1556811 | 0.745 |
| ☑ | FOPYOZ | 692400 | 0.745 |
| ☑ | UDAVIF | 633383 | 0.745 |
| ☑ | UDAVIF01 | 633384 | 0.745 |
| ☑ | UDAVIF02 | 633385 | 0.745 |
| ☑ | UDAVIF03 | 633386 | 0.745 |
| ☑ | VAYXOI | 1280849 | 0.745 |
| ☑ | VAYXOI01 | 1280850 | 0.745 |
| ☑ | VAYXOI02 | 159050 | 0.745 |
| ☑ | VAYXOI03 | 633382 | 0.745 |

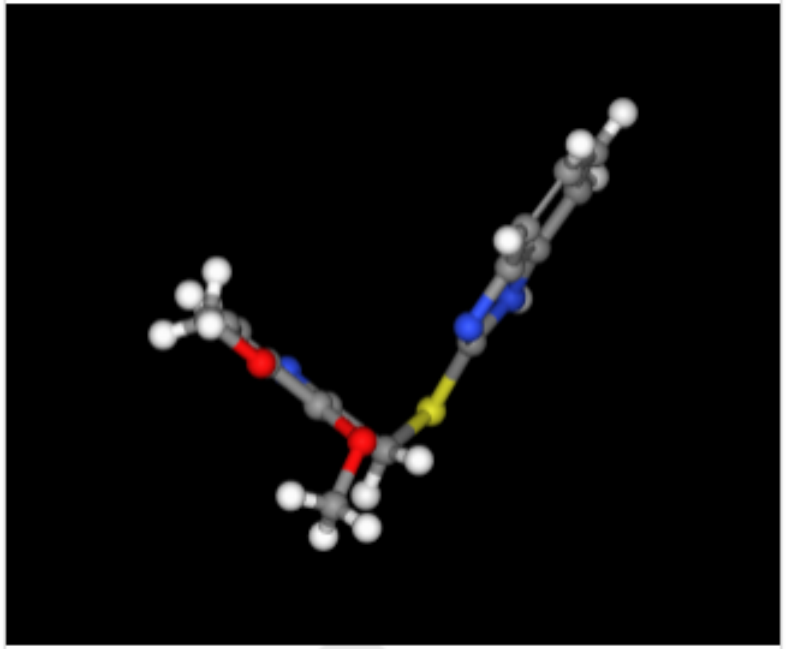
1 - 13 of 13 items

Download

WETBIJ : 2-{[(3,4-dimethoxypyridin-2-yl)methyl]sulfanyl}-1H-benzimidazole
**Space Group:** P $2_1$ $2_1$ $2_1$ (19), **Cell:** *a* 9.1828(16)Å *b* 11.625(2)Å *c* 13.463(2)Å, *α* 90° *β* 90° *γ* 90°

3D viewer

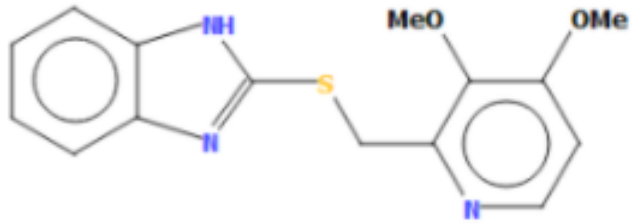
Ball and Stick | No Labels

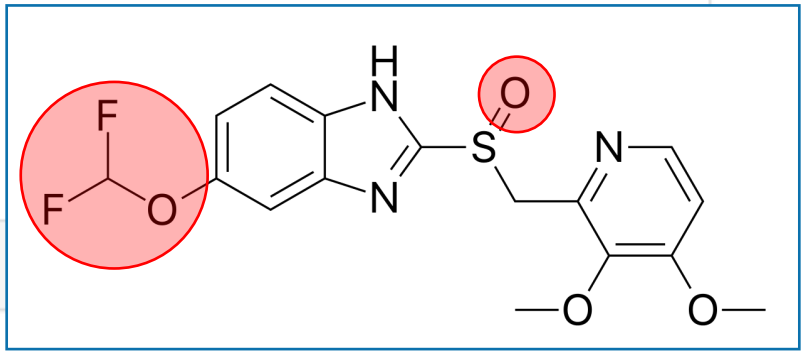


No Packing | H

DISORDER

Chemical diagram




CCDC

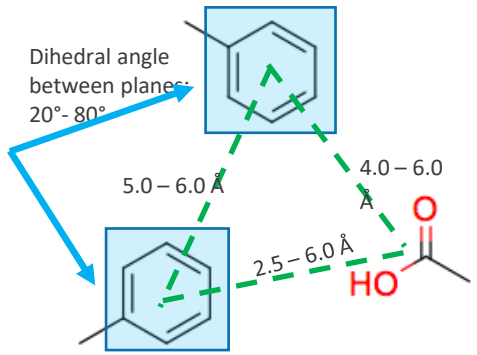
# What have we learnt?

- How to search over a million published crystal structures using our desktop software
  - Building queries from the Draw Window
  - Different ways to build queries (Name, Journal, …)
  - Edit queries
  - View hitlists
- How to refine your results to structures in targeted subsets allowing you to explore drugs, pesticides and MOFs in more detail
  - Restrict search to a subset
  - Bonus: Load and combine subsets

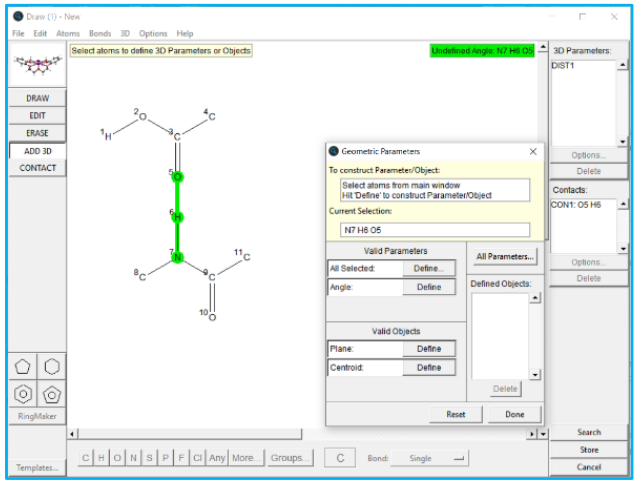
CCDC

# What else can you explore in ConQuest?

## 3D Search

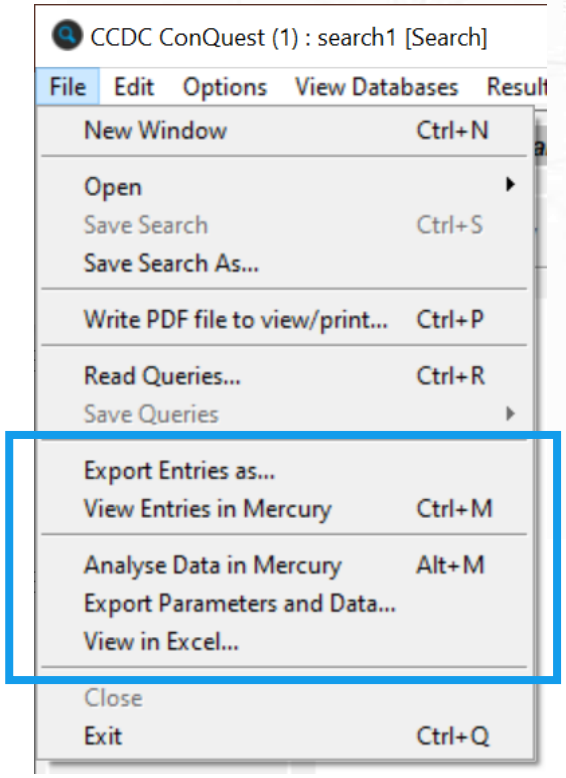


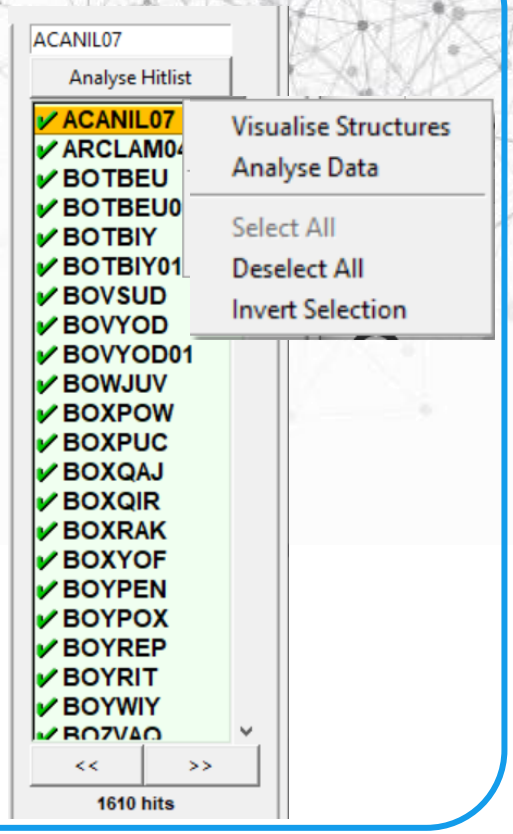

## Intermolecular Interactions



Search In-House Databases

## Export & Analyse Data

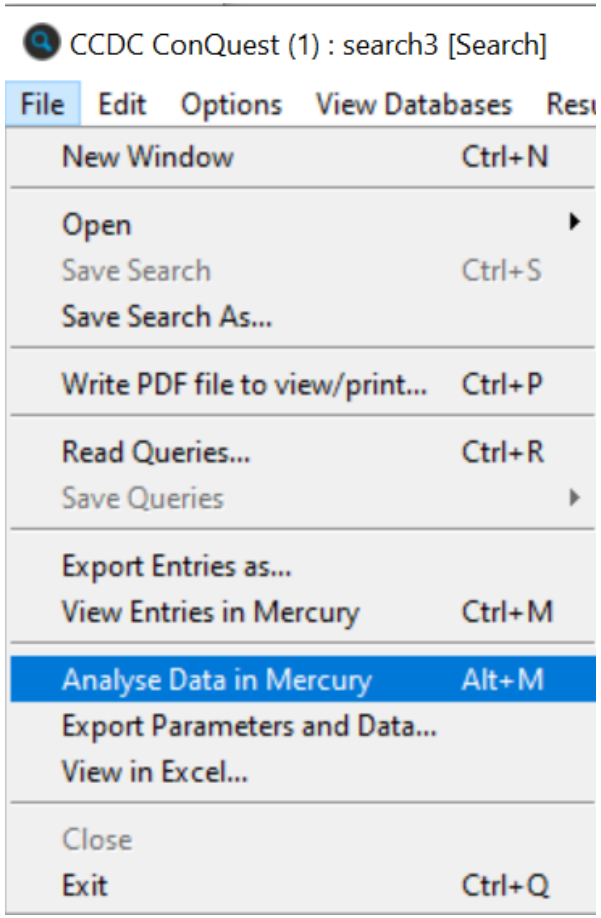


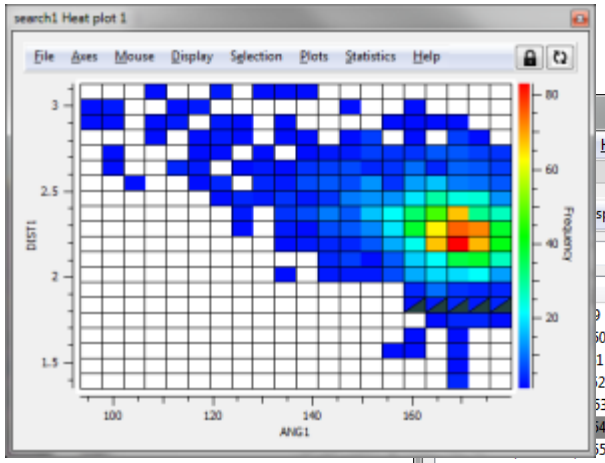



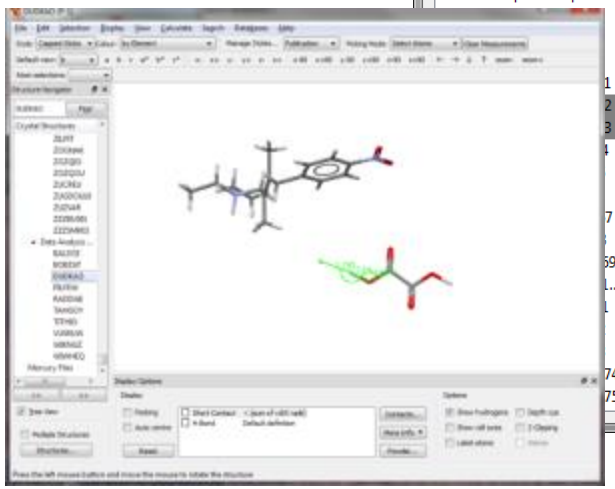

... and more!

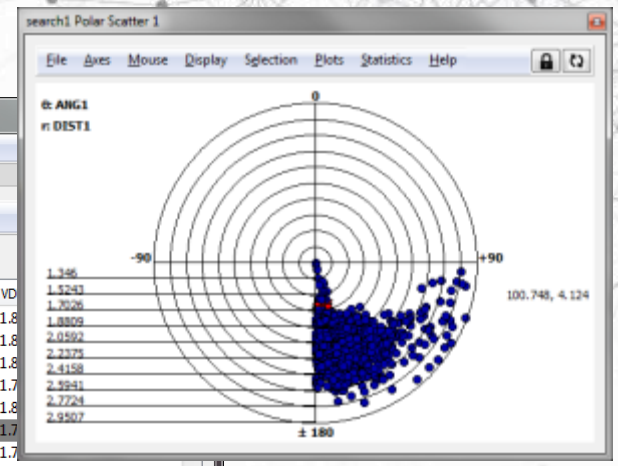
CCDC

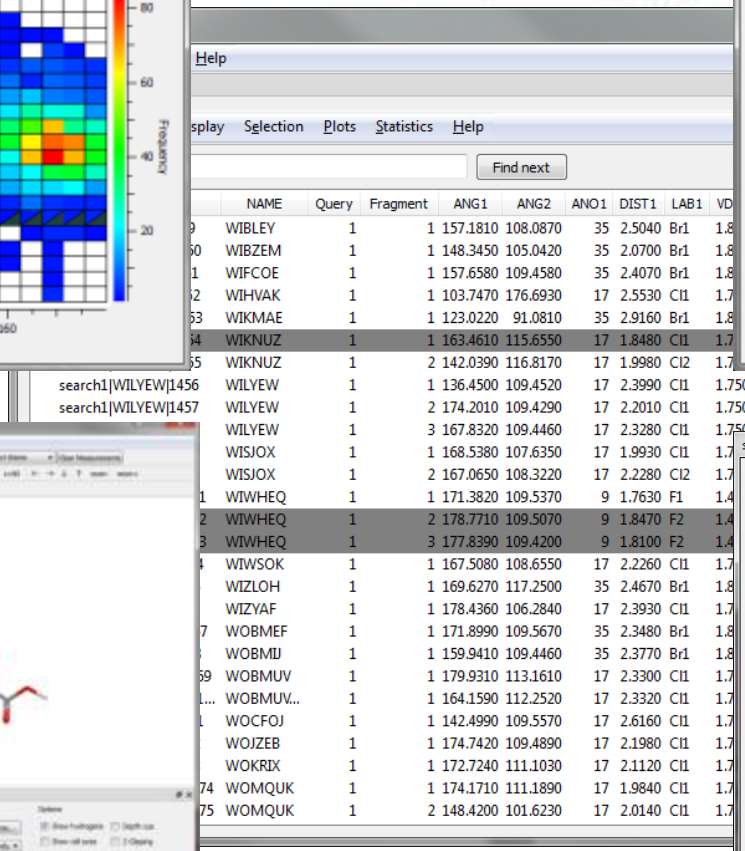
# Data analysis in Mercury

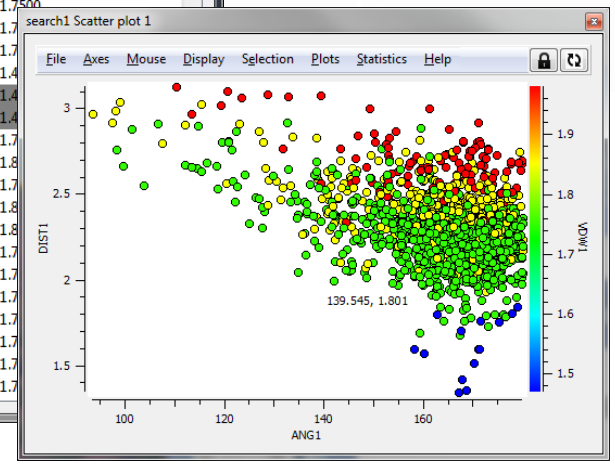





CCDC

https://www.ccdc.cam.ac.uk/Community/educationalresources/

# Want to explore more?

## Training and Educational Resources

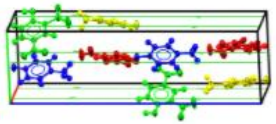
The wealth of information contained within the Cambridge Structural Database (CSD) extends far beyond a collection of crystal structures. Knowledge derived from these materials informs much of chemistry, biochemistry, and biology. Chemical and structural concepts are often difficult to grasp without real world, interactive examples for students to explore.

The CCDC and our colleagues continually produce educational materials for use in classroom and computer lab settings, or as independent study modules. Many of these materials make use of the Teaching Subset - a freely available set of over 750 structures that can be investigated with the free version of our Mercury visualisation and analysis program. Of course, our database of over one million entries are available for free through our Access Structures portal.

If you are an educator looking for supplementary teaching materials, find out more about the Teaching Database **here**. If you have developed your own modules using the CSD and would like to share them with the broader community, please contact us at education@ccdc.cam.ac.uk.

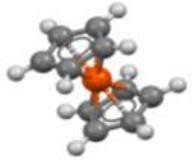
To keep up to date with the latest news from education and outreach at the CCDC, sign up for the Education and Outreach Newsletter **here**.

Register for E&O newsletter

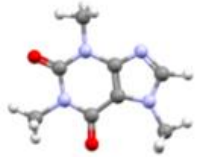


**CSD-Core**

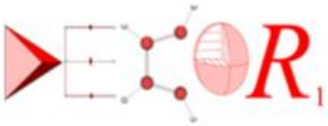
Essential crystallographic and structural chemistry capabilities.



Information on the Teaching Subset



Access a series of teaching modules for use in the classroom



DECOR: Educational Resources for Teaching Crystallography

Self-guided workshops

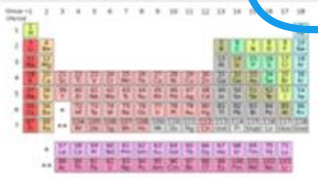
Download a series of self-guided workshop materials for CCDC tools and features

Watch software training and support videos


CSDU modules

On-demand modules with completion certificate


Access fun science activities for kids through the CCDC Home learning page



Explore the Periodic Table through Crystal Structures

YouTube and LabTube channels

CCDC

# Want to explore more?

## More Virtual Workshops in the coming weeks

https://www.ccdc.cam.ac.uk/News/Events/

Stability and design of solid forms

May 10th
9:30 BST

Stability and design of solid forms

May 12th
21:30 BST

First steps in protein ligand docking

May 18th
11:00 BST

REGISTER HERE

CCDC