Programmatic search and analysis using the CSD Python API

CCDC Virtual Workshop Spring 2021 – Session 3

Alex Moldovan, Natalie Johnson, Yinka Olatunji-Ojo, Ilaria Gimondi, Suzanna Ward

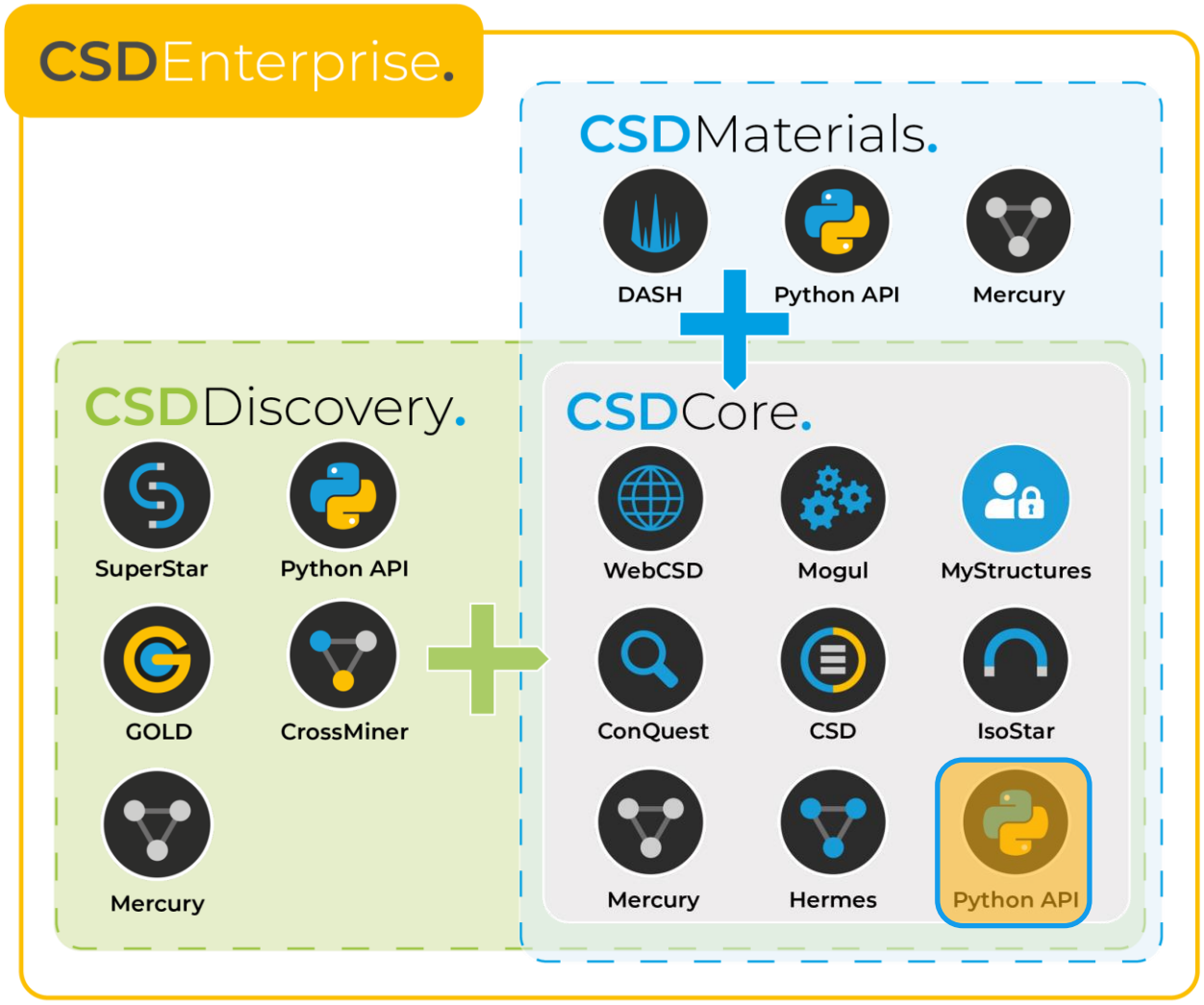
April 2021

CCDC
advancing structural science

Learning outcomes for today

- Familiarise yourself with how CSD entries are represented in the CSD Python API.
- Learn how to access CSD entries.
- Learn how to read different file formats.
- Learn how to run a search and output results.

The CSD software



Professional Services

Research & Knowledge partnerships

CSDCommunity.

Mercury
enCIFer
Symmetry
Deposit
CellCheck
Educational
Access
MyStructures

CCDC

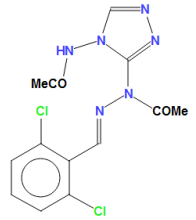
CSD Python API

The CSD-Python API
(Application Programming Interface)

- Enables you to use many capabilities of the CSD-Core without being bound by graphical interfaces
- You can readily create CSD-driven analyses and workflows tailored to your needs
- Our Mercury/Hermes interface allows you to easily access your scripts and run analysis

CSD Python API - Example

```
In [10]: from ccdc import io, diagram

In [26]: import IPython.core.display
         import StringIO

In [27]: # Set up CSD entry reader and find the first entry in the database
         csd = io.EntryReader('csd')
         csd_entry = csd[0]
         csd_entry.identifier

Out[27]: u'AABHTZ'

In [30]: # Generate a diagram for that CSD entry
         diagram_generator = diagram.DiagramGenerator()
         diagram_generator.settings.font_size = 12
         img = diagram_generator.image(csd_entry)

In [31]: # Display the 2D diagram
         output = StringIO.StringIO()
         img.save(output, "PNG")
         contents = output.getvalue()
         IPython.core.display.display_png(contents, raw=True)
```



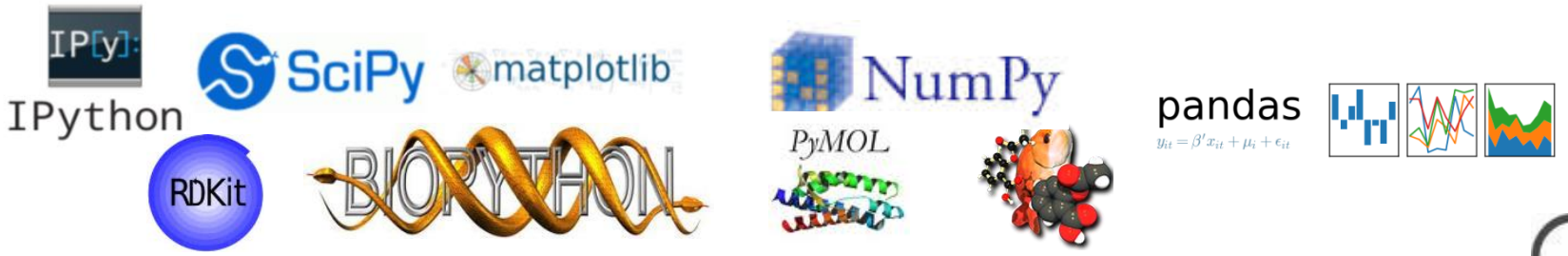
python™

CCDC

The CSD Python API – Why?

- In CSD-Core:
  - Many things easy to do for one entry/structure, hard/tedious for 100s or 1000s
  - Custom analyses
  - Quick/simple tools and tweaks

- Use with other tools & packages



CCDC

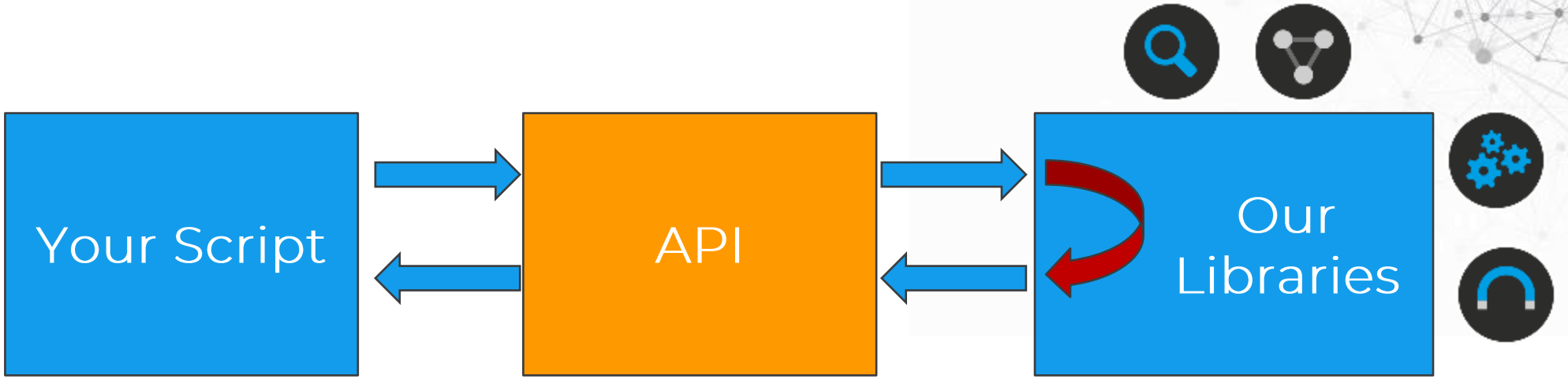
Benefits of scripting workflows

- Automating reporting and analyses saves time
- Results will be consistent and predictable
- Having a scripted workflow makes it easy to share knowledge with colleagues

CCDC

What is an API?

- Application Programming Interface

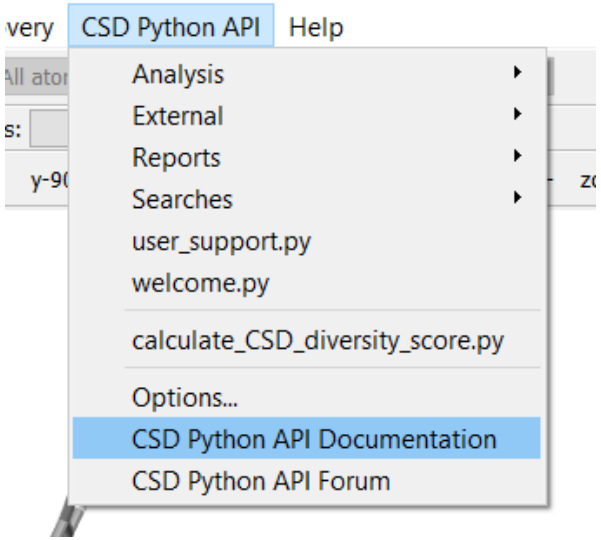


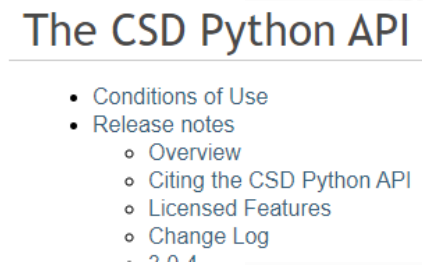
python™

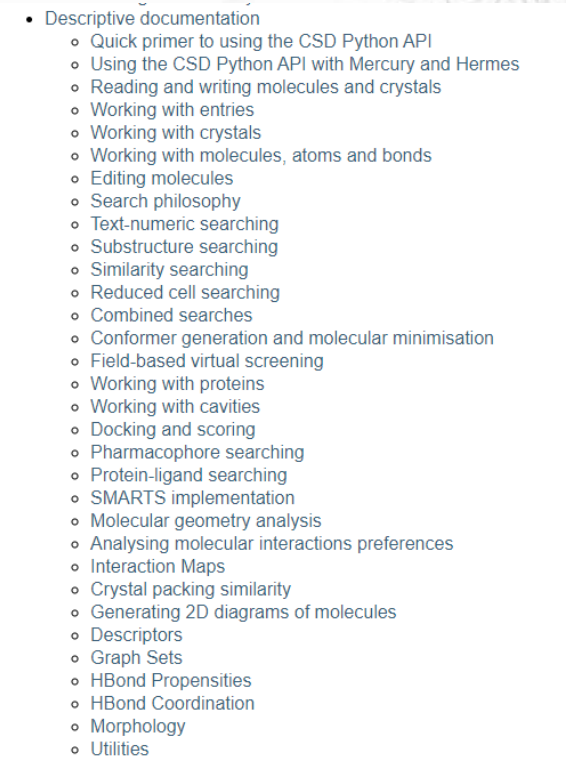
CCDC

CSD Python API: Documentation

- Documentation is easy to navigate and highly descriptive





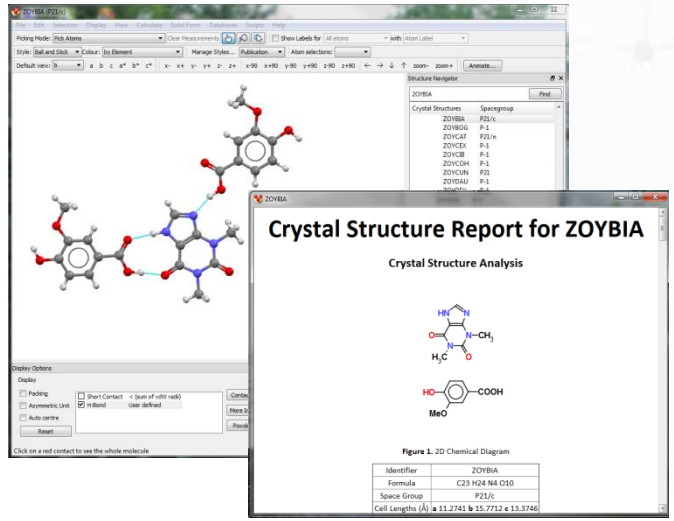


ccdc.cam.ac.uk/docs/csd_python_api/

CCDC

CSD Python API: Interface

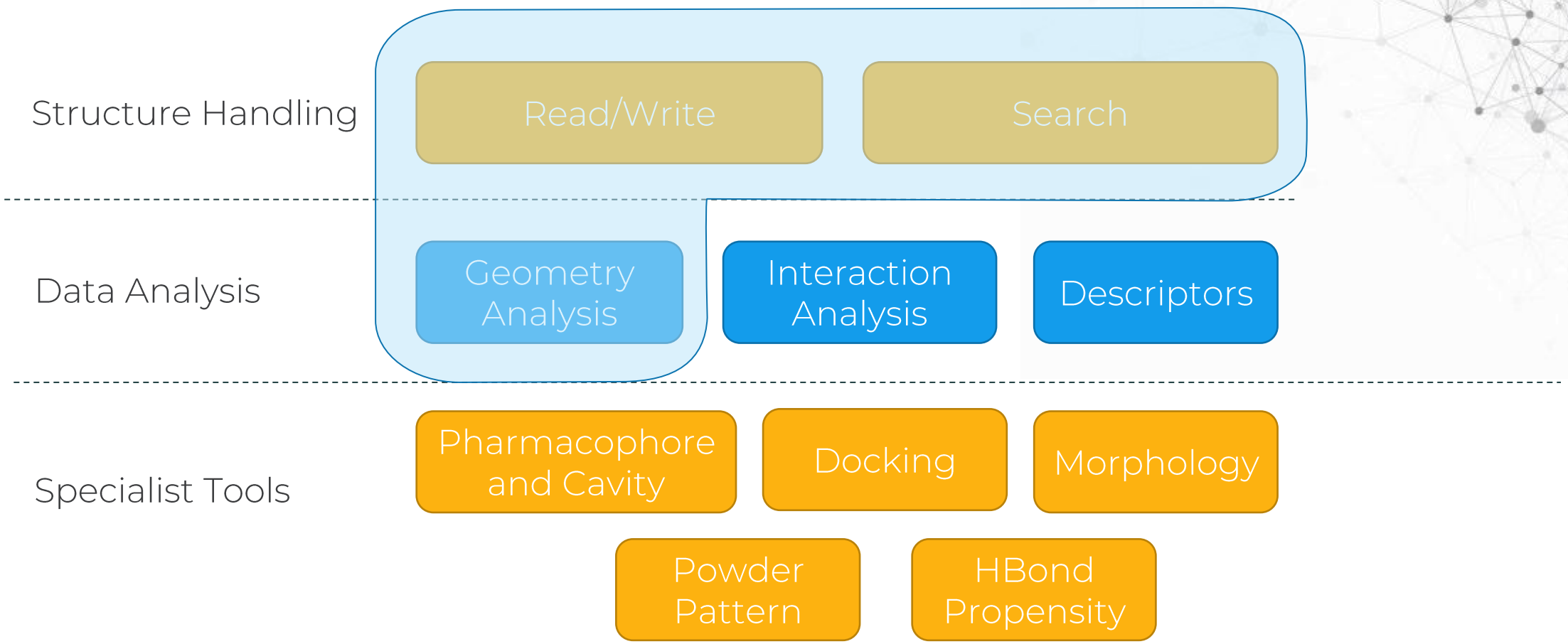
- Mercury/ Hermes includes a “CSD Python API” menu



- Allows direct running of built-in or user-generated Scripts from the Mercury graphical user interface
- Scripts menu is designed to work from the currently loaded structure in the visualiser (e.g. CSD entry, CIF or MOL2)

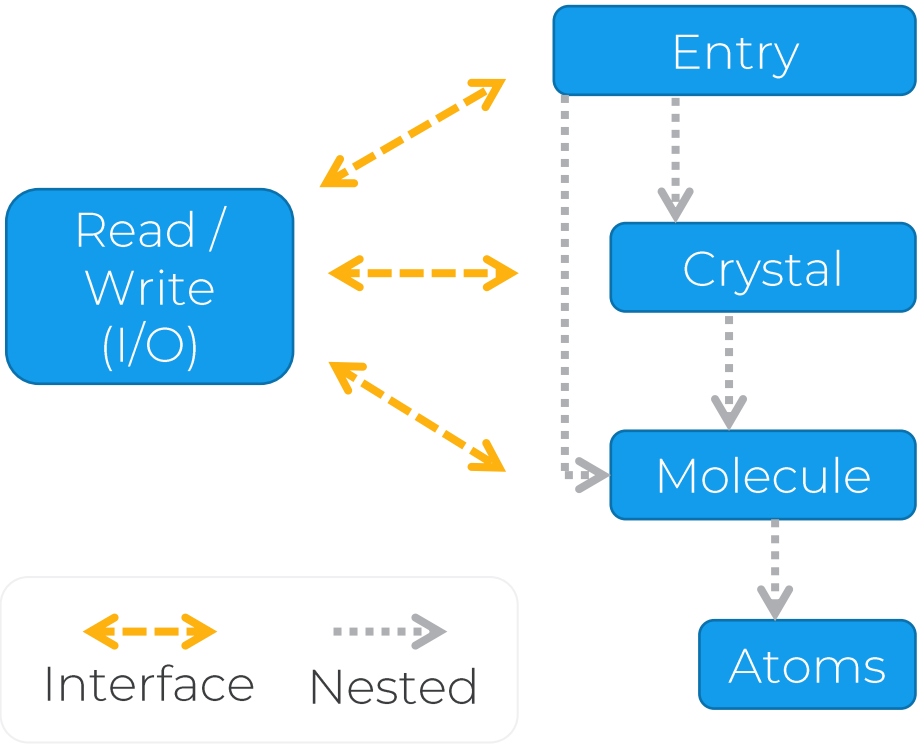
CCDC

Python API Structure



CCDC

Structure Handling



Submitted Properties – Habit, Colour, etc.
Journal Source Info

Crystallographic Data – Full Structure

Molecular Component of Crystal

Individual Atoms and Their Properties

CCDC

Structure Handling

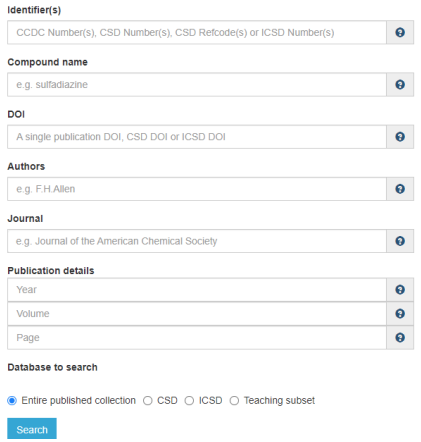
Search

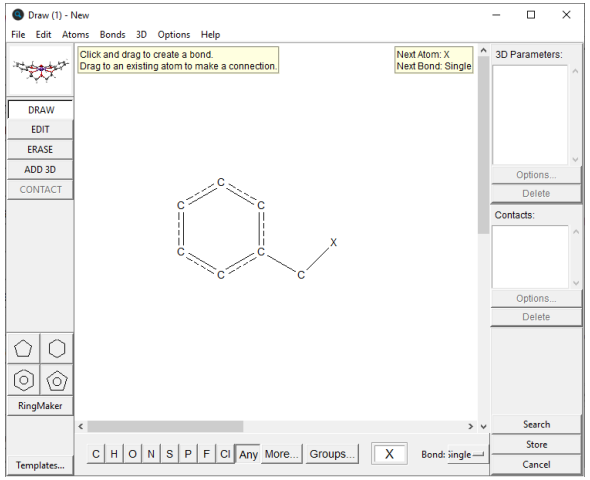
Text-Numeric

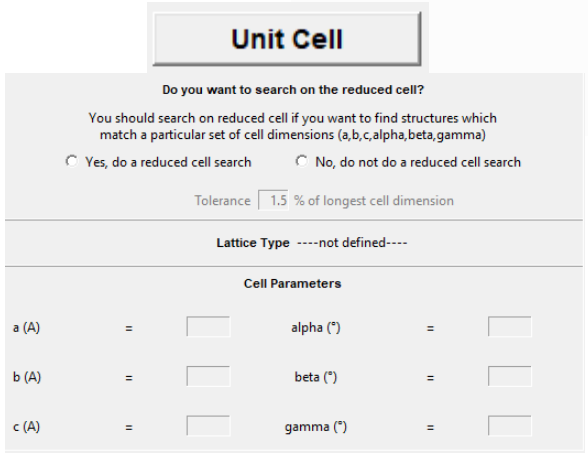
Substructure

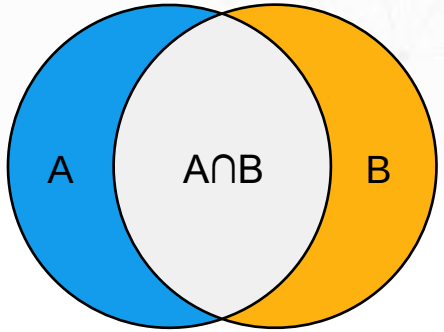
Reduced Cell

Similarity









CCDC