

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

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

CSDEnterprise.

CSDMaterials.



DASH



Python API



Mercury



CSDDiscovery.



SuperStar



Python API



GOLD



CrossMiner



Mercury

CSDCore.



WebCSD



Mogul



MyStructures



ConQuest



CSD



IsoStar



Mercury



Hermes



Python API

CSDCommunity.



Mercury



enCIFer



Symmetry



Deposit



CellCheck



Educational



Access



MyStructures

Professional
Services

Research
& Knowledge
partnerships

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 code import io, diagram

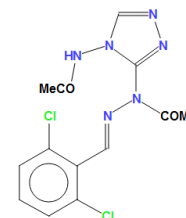
In [26]: import IPython.core.display
import StringIO

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'

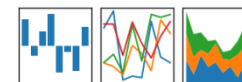
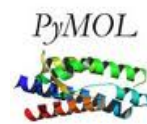
In [30]: # Generate a diagram for that CSD entry
diagram_generator = diagram.DiagramGenerator()
diagram_generator.settings.font_size = 12
img = diagram_generator.image(csd_entry)

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



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



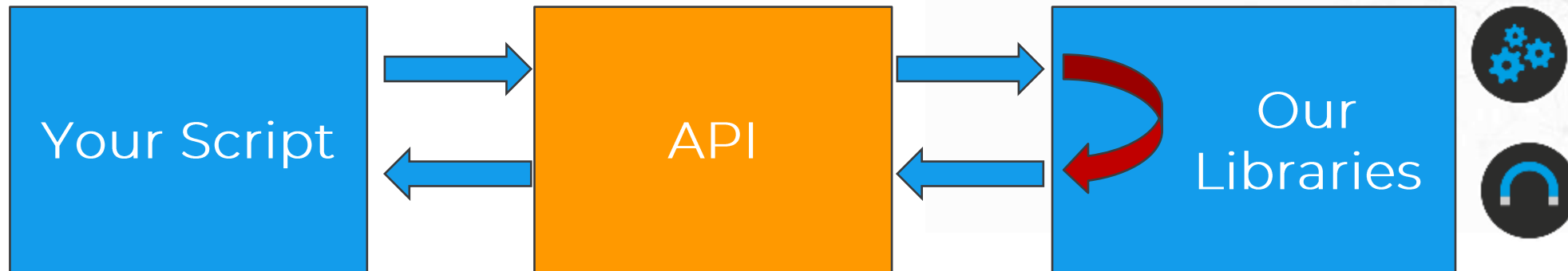
Benefits of scripting workflows



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

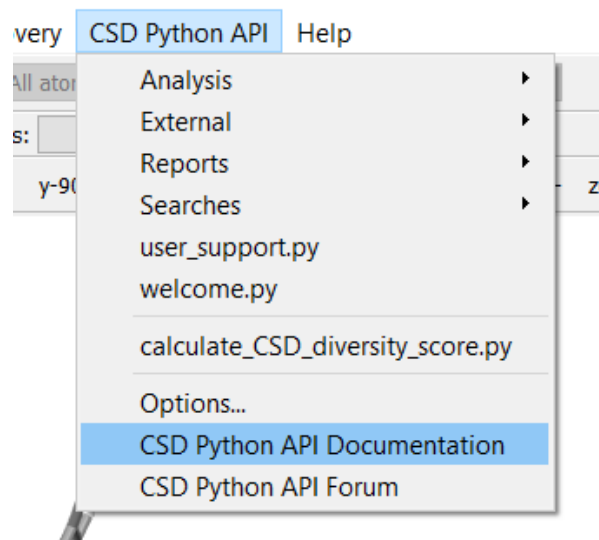
What is an API?

- Application Programming Interface



CSD Python API: Documentation

- Documentation is easy to navigate and highly descriptive



Next topic
Conditions of Use
Quick search

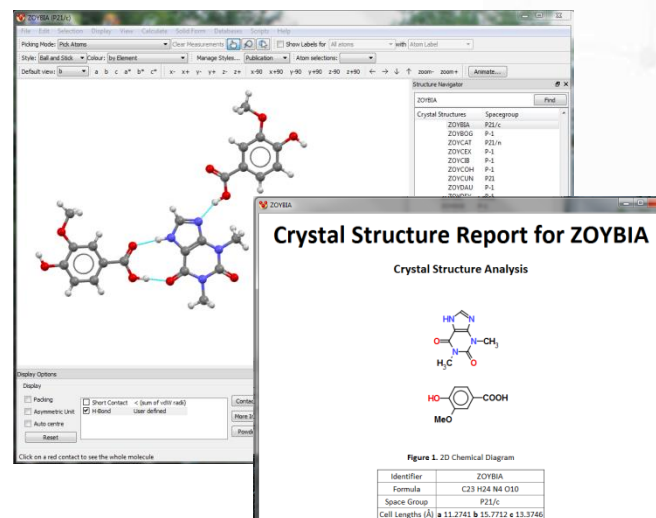
The CSD Python API

- Conditions of Use
- Release notes
 - Overview
 - Citing the CSD Python API
 - Licensed Features
 - Change Log

- Descriptive documentation
 - Quick primer to using the CSD Python API
 - Using the CSD Python API with Mercury and Hermes
 - Reading and writing molecules and crystals
 - Working with entries
 - Working with crystals
 - Working with molecules, atoms and bonds
 - Editing molecules
 - Search philosophy
 - Text-numeric searching
 - Substructure searching
 - Similarity searching
 - Reduced cell searching
 - Combined searches
 - Conformer generation and molecular minimisation
 - Field-based virtual screening
 - Working with proteins
 - Working with cavities
 - Docking and scoring
 - Pharmacophore searching
 - Protein-ligand searching
 - SMARTS implementation
 - Molecular geometry analysis
 - Analysing molecular interactions preferences
 - Interaction Maps
 - Crystal packing similarity
 - Generating 2D diagrams of molecules
 - Descriptors
 - Graph Sets
 - HBond Propensities
 - HBond Coordination
 - Morphology
 - Utilities

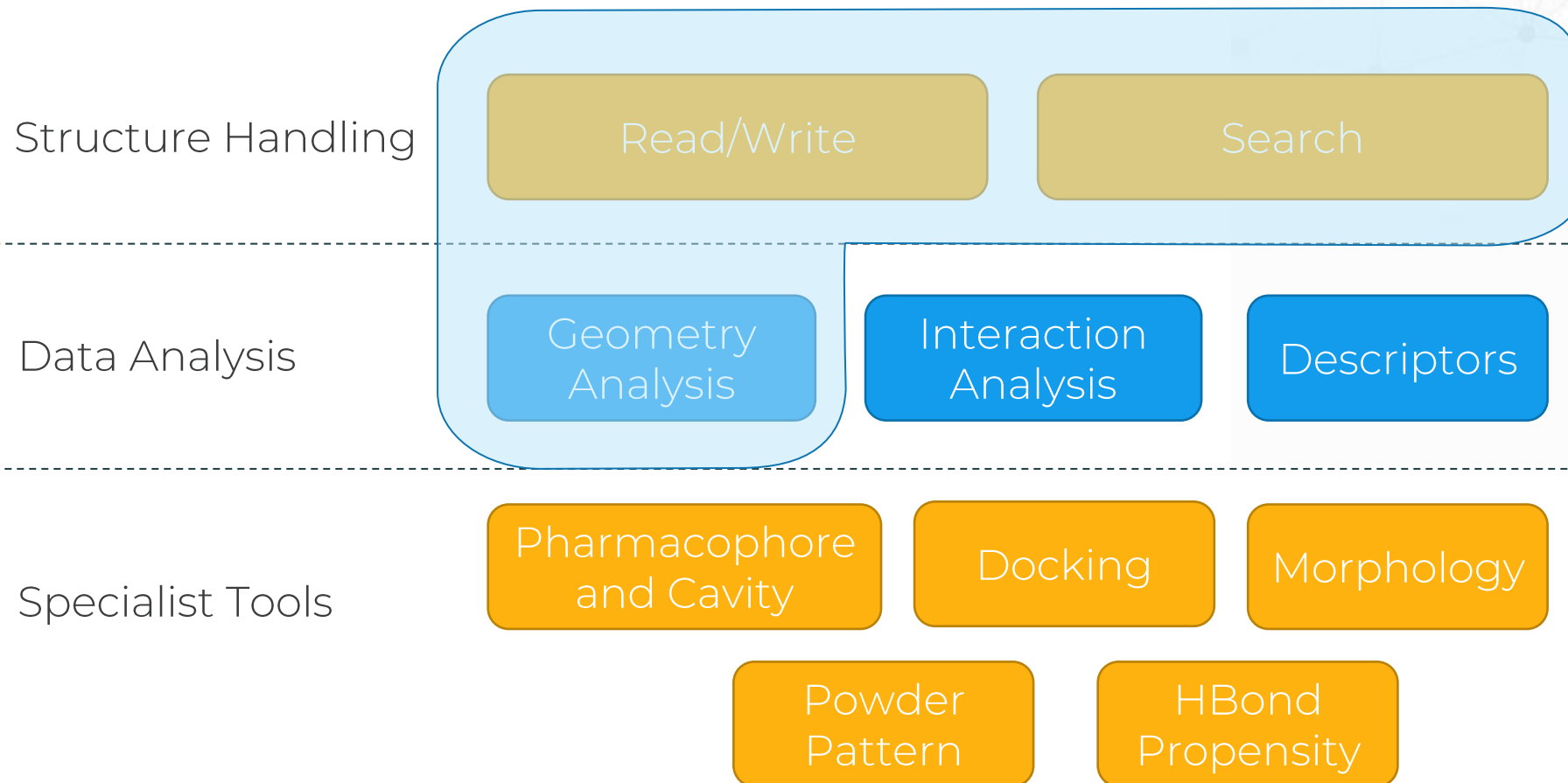
CSD Python API: Interface

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

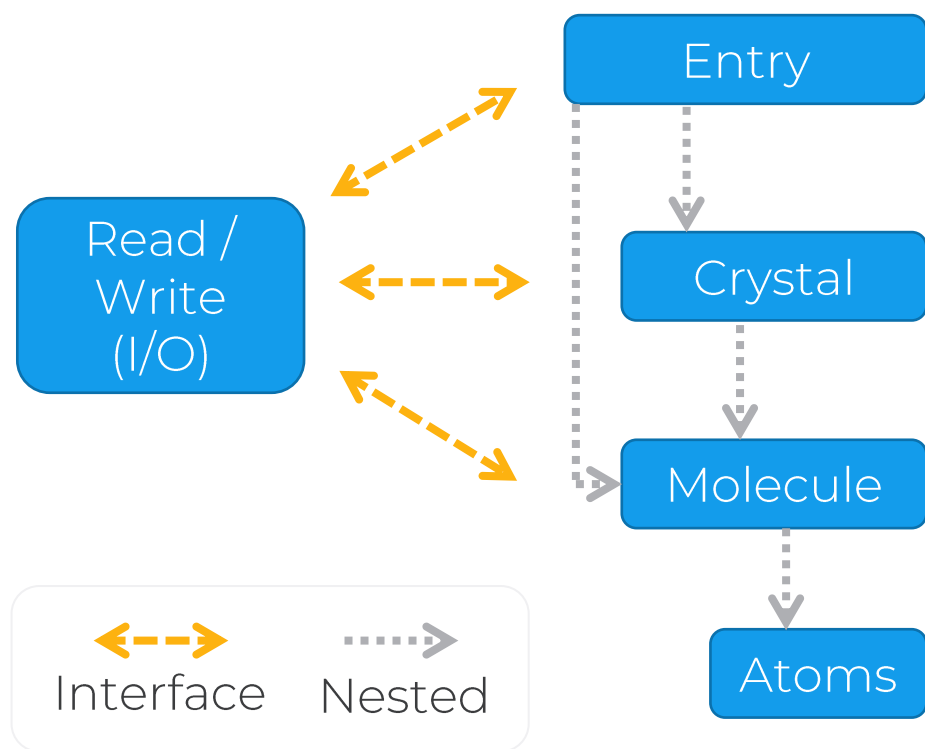


- Allows direct running of built-in or user-generated Scripts from the Mercury graphical user interface
- Scripts menu is designed to work from the currently loaded structure in the visualiser (e.g. CSD entry, CIF or MOL2)

Python API Structure



Structure Handling



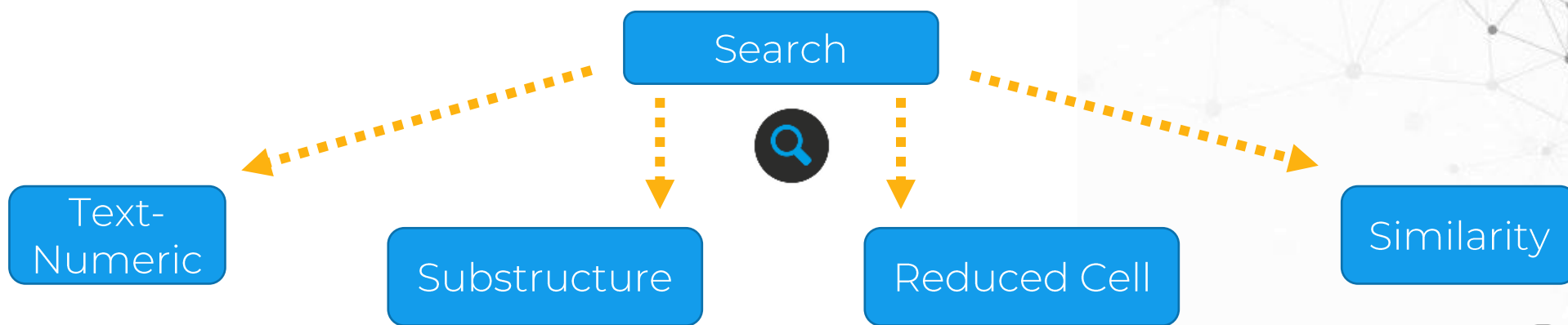
Submitted Properties – Habit, Colour, etc.
Journal Source Info

Crystallographic Data – Full Structure

Molecular Component of Crystal

Individual Atoms and Their Properties

Structure Handling



Identifier(s)
CCDC Number(s), CSD Number(s), CSD Refcode(s) or ICSD Number(s) ⓘ

Compound name
e.g. sulfadiazine ⓘ

DOI
A single publication DOI, CSD DOI or ICSD DOI ⓘ

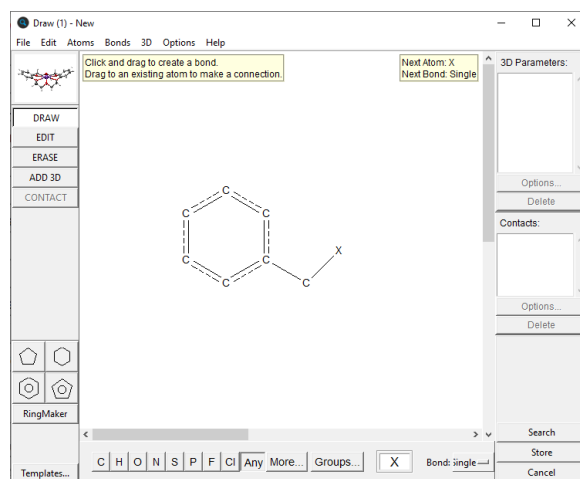
Authors
e.g. F.H.Allen ⓘ

Journal
e.g. Journal of the American Chemical Society ⓘ

Publication details
Year ⓘ
Volume ⓘ
Page ⓘ

Database to search
☒ Entire published collection ☐ CSD ☐ ICSD ☐ Teaching subset

Search



Unit Cell

Do you want to search on the reduced cell?
You should search on reduced cell if you want to find structures which match a particular set of cell dimensions (a,b,c,alpha,beta,gamma)

☐ Yes, do a reduced cell search ☐ No, do not do a reduced cell search

Tolerance % of longest cell dimension

Lattice Type ----not defined----

Cell Parameters

a (Å)	=	<input type="text"/>	alpha (°)	=	<input type="text"/>
b (Å)	=	<input type="text"/>	beta (°)	=	<input type="text"/>
c (Å)	=	<input type="text"/>	gamma (°)	=	<input type="text"/>

Search
Store
Cancel

