CCDC

advancing structural science

Introduction to the Cambridge Structural Database

Making the most of structural data

Lee Daniels, PhD

Business Development Executive

May 20, 2020

CCDC

advancing structural science

About us

- **Origins:** Department of Chemistry, University of Cambridge - 1965. Established by Dr. Olga Kennard.
- Now a fully independent institution, non-profit organization, and registered charity.
- Strong track record in basic research through more than 750 peer-reviewed publications.

J.D. Bernal and research group including Olga Kennard at Stonehenge in 1948

CCDC

What we do

- The CCDC serves the global structural science community, helping address our customers' structural, physical, and chemical challenges through unique:
  - Data
  - Software
  - Knowledge

Charitable Objective:
Advancement of chemistry and crystallography for the public benefit

CCDC

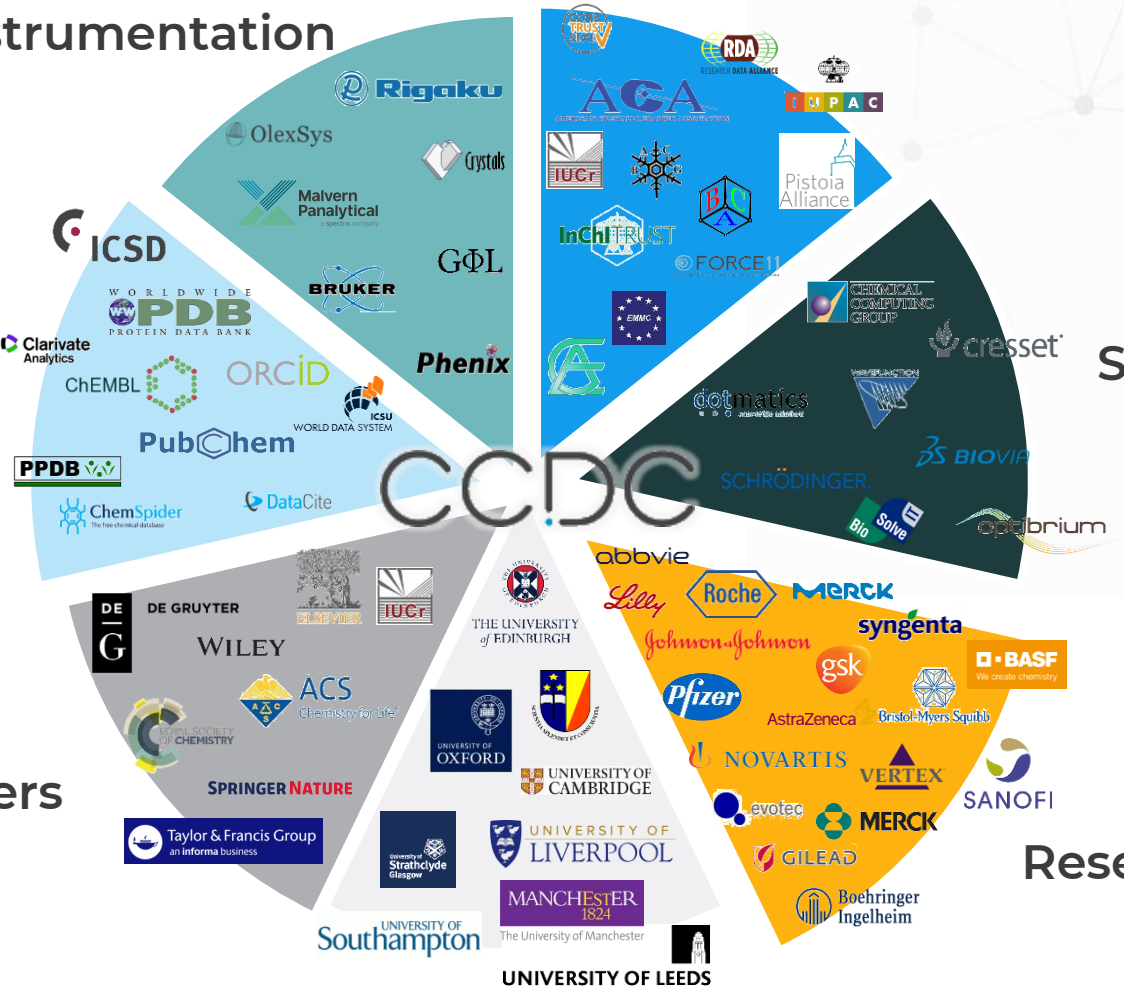
CCDC Community

Committees & Commissions

Instrumentation

Data Repositories

Software



Publishers

Research Partners

Academic Partners

CCDC

The Cambridge Structural Database (CSD)

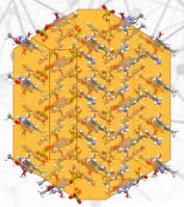
The world's only fully curated & enhanced database for small-molecule organic and metal-organic crystal structures.

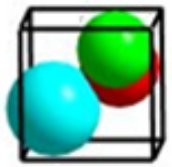
- Contains >1-million entries from X-ray and neutron diffraction
- Each structure undergoes validation by expert chemists and crystallographers
- Every entry enriched with bibliographic, chemical and physical property information
- Data made available instantly via Access Structures
- Contains data published directly through the CSD; not available anywhere else

CCDC

Crystal structure databases

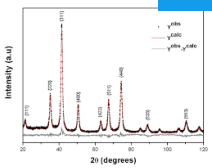
Database	Est.	Entries	Compounds
Cambridge Structural Database	1965	>1,000,000	organic and metal-organic
Inorganic Crystal Structure Database	1978	>200,000	inorganic
Crystallography Open Database	2003	>400,000	organic, inorganic, metal-organics and minerals
Protein Data Bank	1971	>120,000	biological macromolecules
International Centre for Diffraction Data	1941	>900,000	powder diffraction





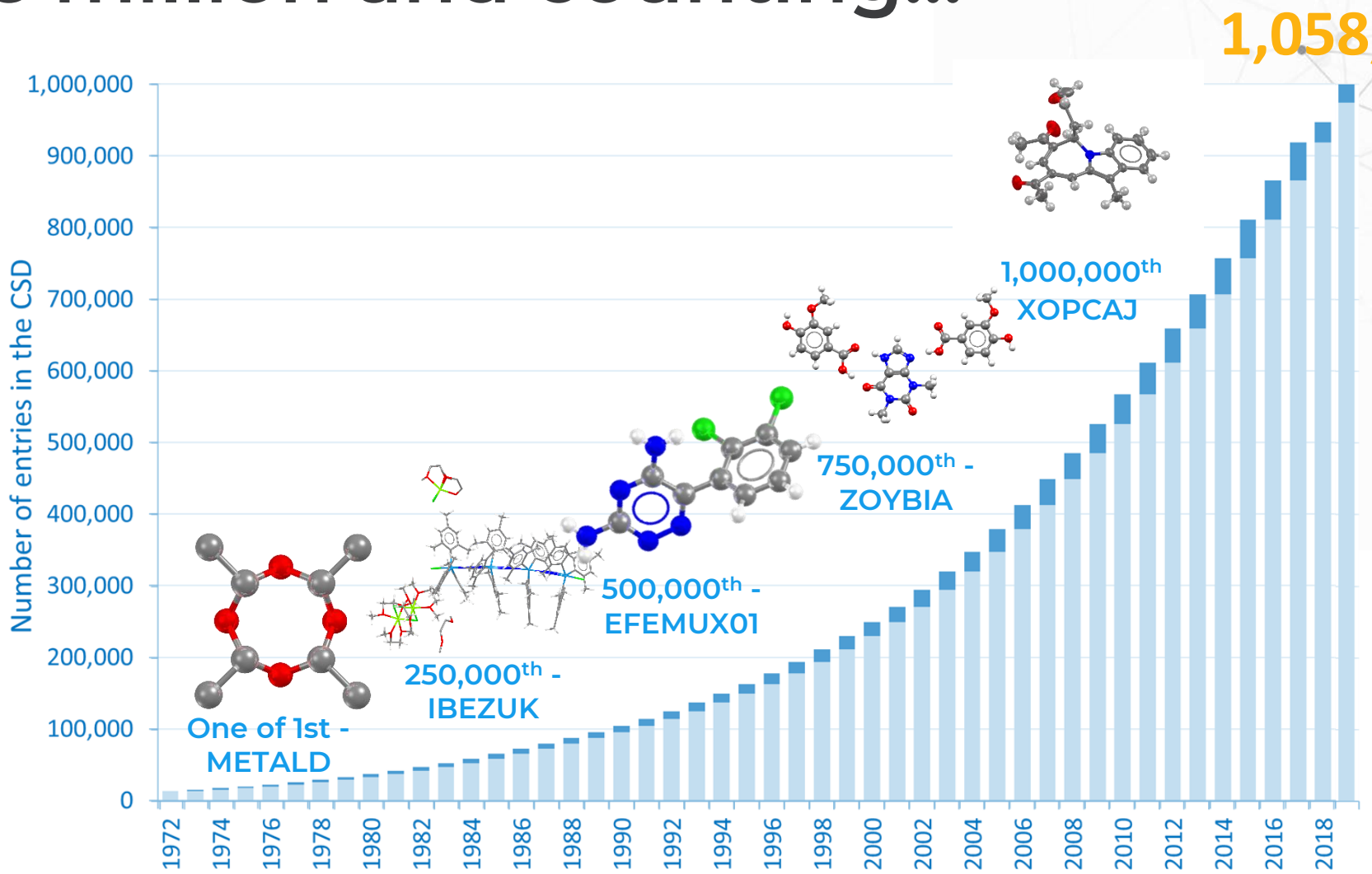
Joint access & deposition!

Links between ligands



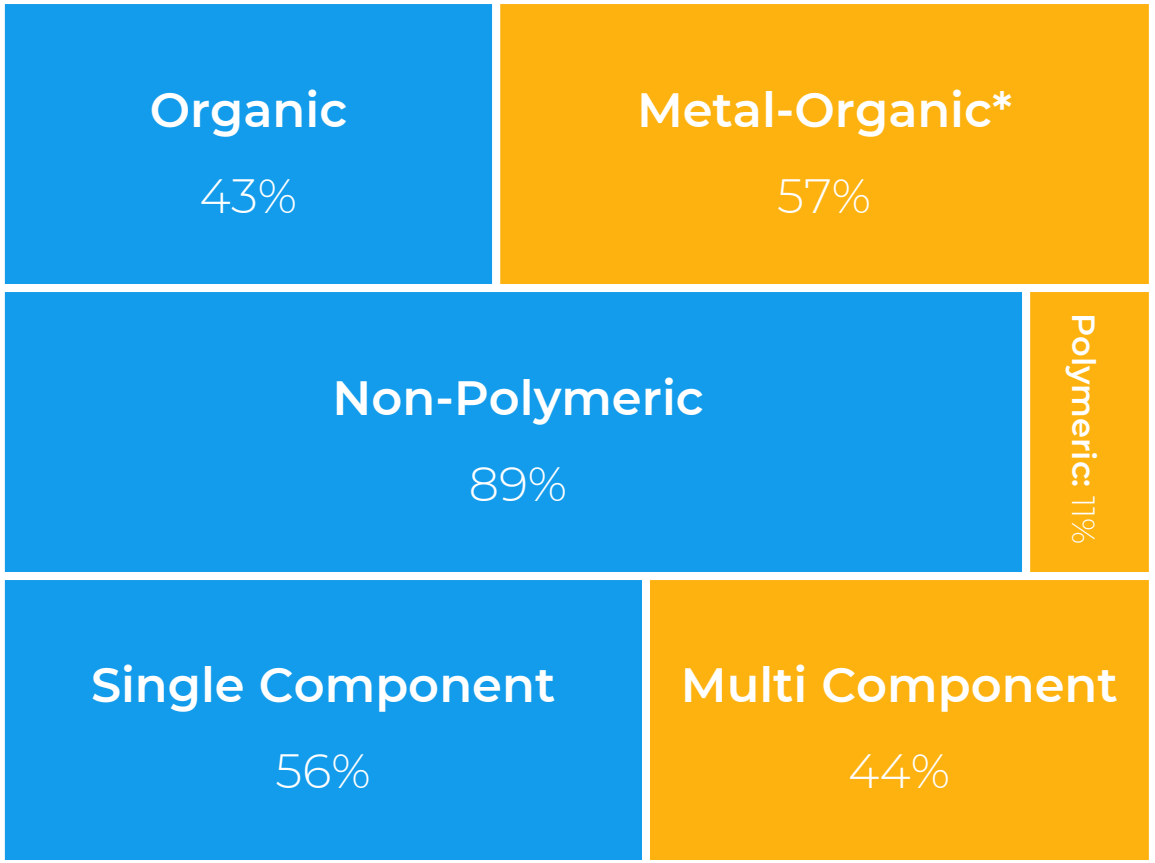
New data partnership as of 18 May!

One million and counting…

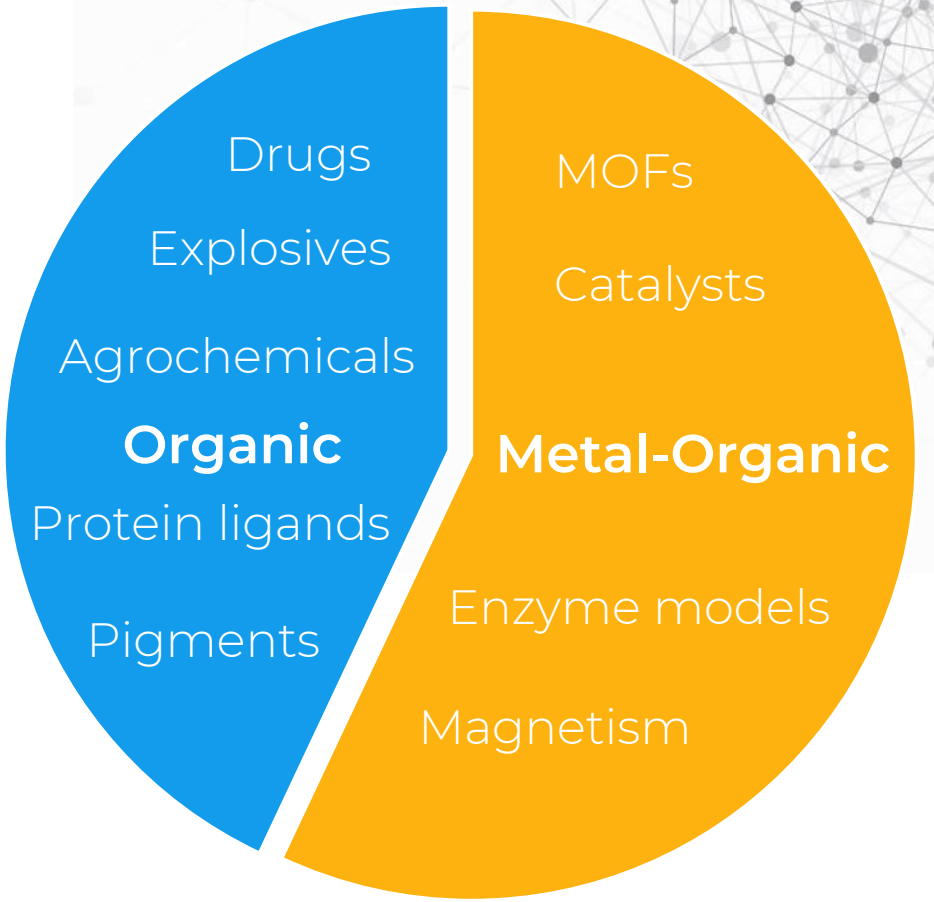


CCDC

What's in a million?



*Contains Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po, Transition Metal, or Actinide

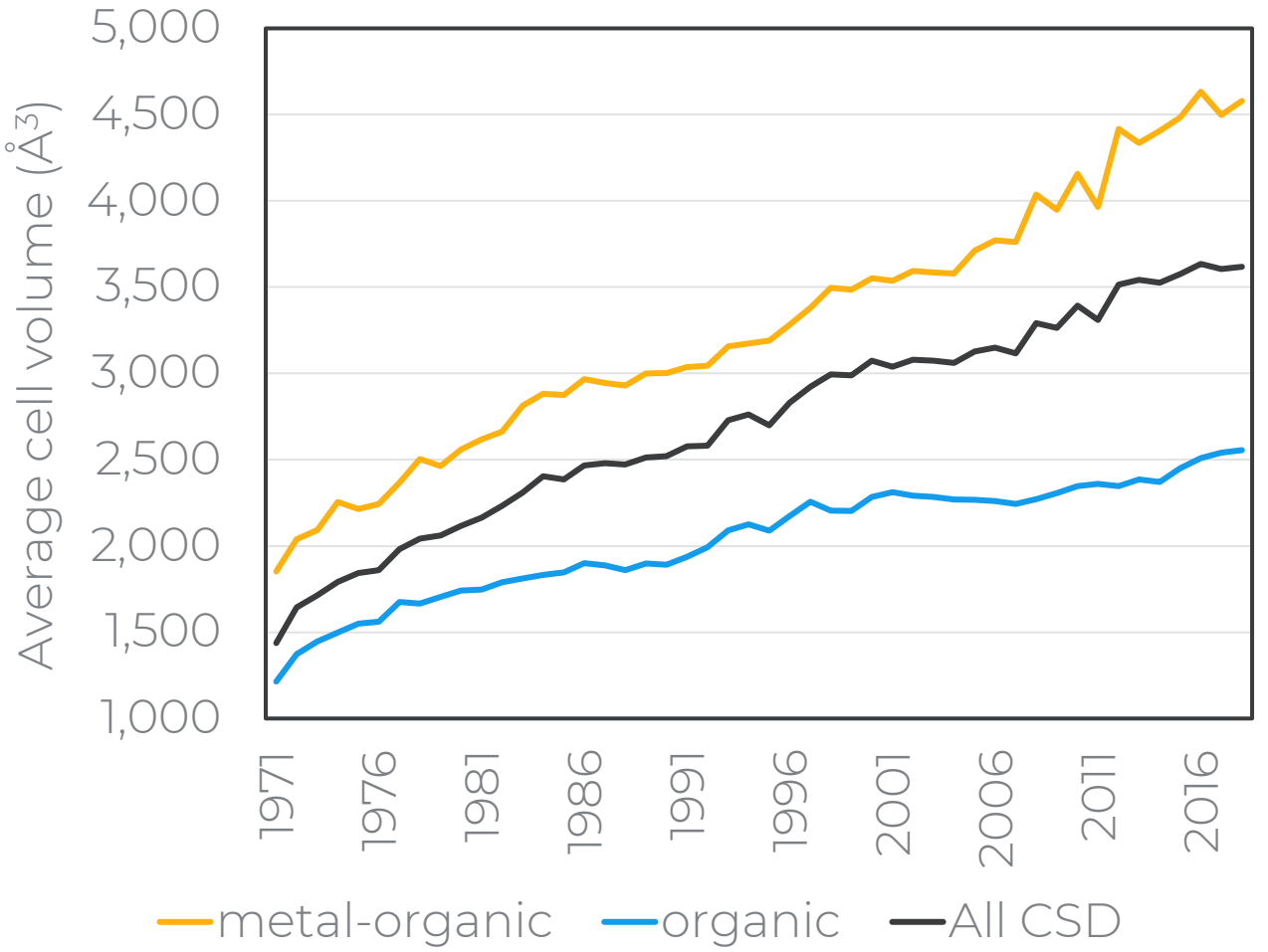


CCDC

Increasing complexity

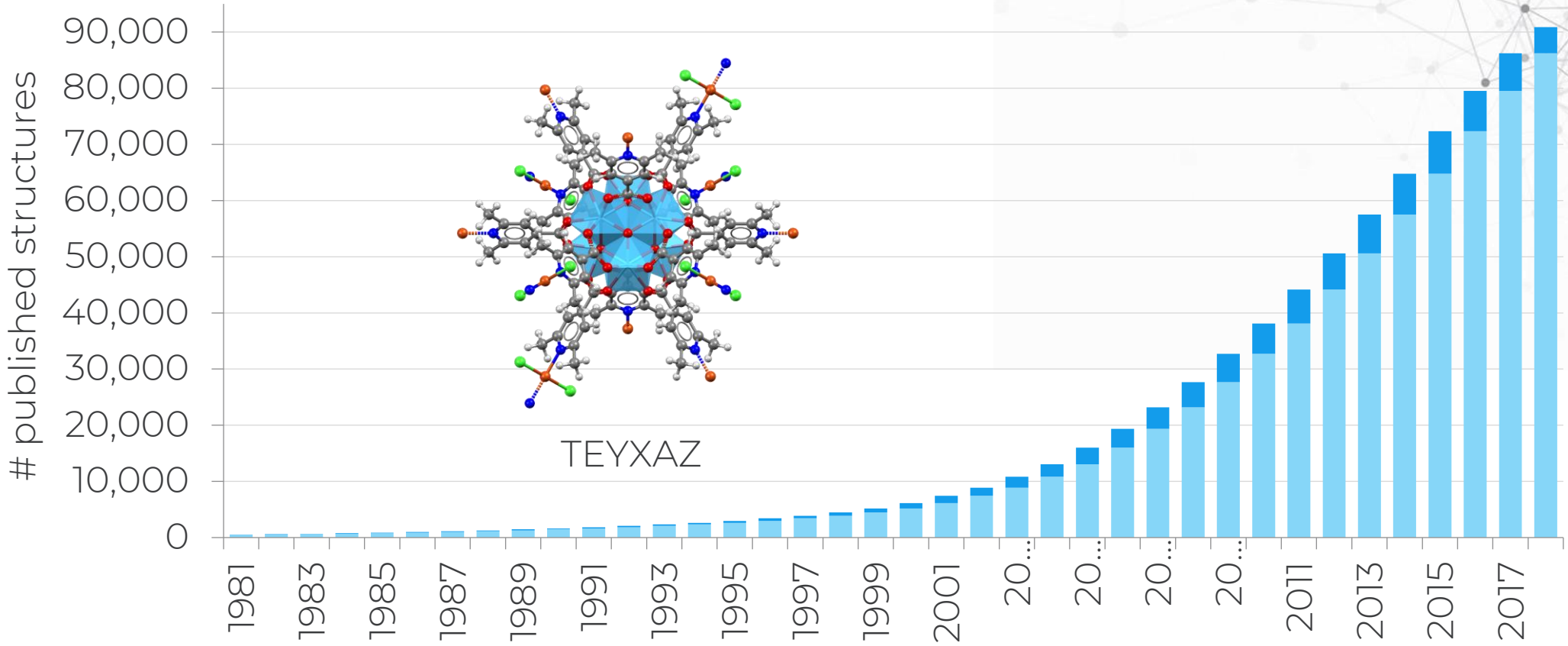
Increasing:

- Formula weights
- Unit cells
- Number of elements



CCDC

The rise of MOFs



Patrick F. Muldoon, Chong Liu, Carson C. Miller, Samuel Benjamin Koby, Michael O'Keeffe, Tian-Yi Luo, Nathaniel L Rosi, Sunil Saxena, Austin Gamble Jarvi, *Journal of the American Chemical Society*, 2018, 140, 6194, DOI: 10.1021/jacs.8b02192

CCDC

CSD pop quiz!

What format should you deposit your data in?

.xtal | .csd | .cif | .res

How many datasets are deposited annually?

100 | 500 - 1000 | 10,000 - 70,000 | 80,000 – 100,000

What is the most commonly deposited structure?

Ferrocene | Caffeine | Glycine | Sulfathiazole

What country deposits the most structures?

USA | China | India | UK

What percentage of depositions are hydrates?

13% | 74% | 2% | 28%

What stage should you deposit your data?

After publication | After refinement | During collection | Never

What types of structures can you deposit with the CCDC?

Organic | Inorganic | Metal-organic | All three

CCDC

CCDC Solutions

CSD-Enterprise
All CCDC application software

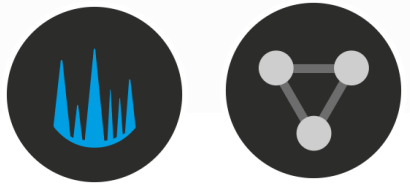
CSD-Discovery

Discover new molecules with pharmaceutical applications

CSD-System

Search, visualise, analyse and communicate structural data

CSD-Materials



Understand and predict solid form stability and properties

CCDC

CSD-System

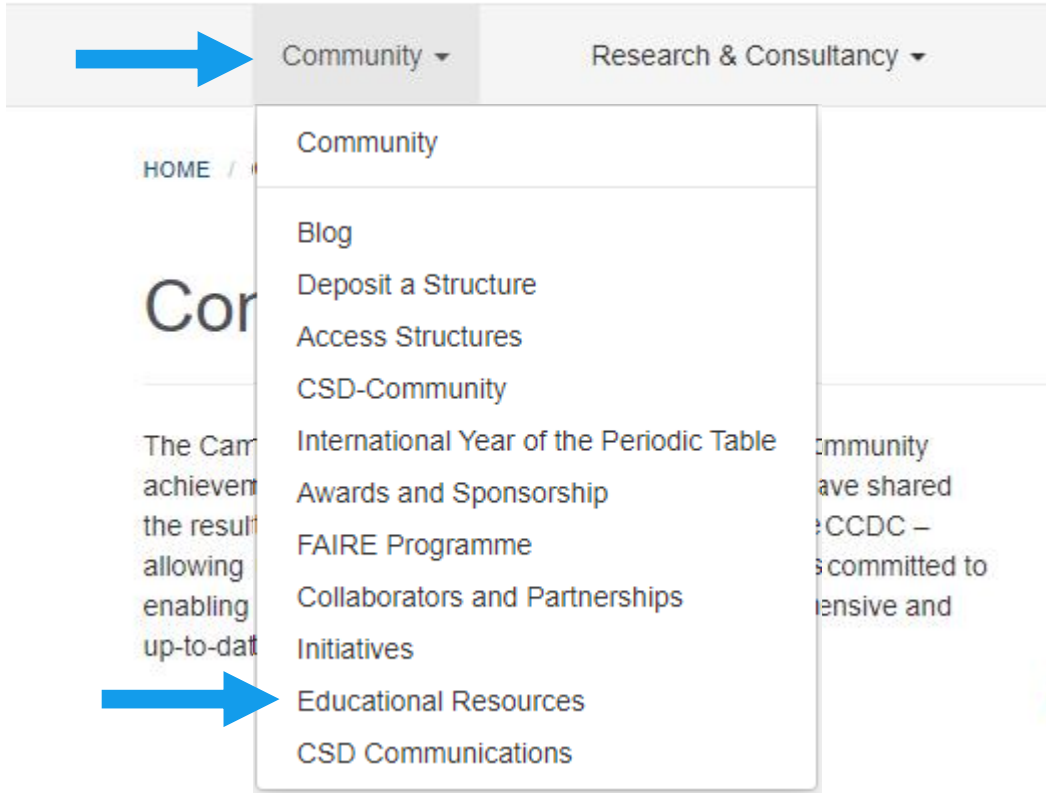
Essential search, visualisation and analysis features to deliver knowledge from the **CSD**

- **Mercury:** Visualisation
- **WebCSD** & **ConQuest:** Advanced search
- **Mogul:** Molecular conformations
- **IsoStar:** Intermolecular interactions
- **CSD Python API:** Structural property insights

CCDC

Educational Resources

CCDC



- Self-guided tutorials and workshops
- Teaching modules & data subset
- Periodic table in crystal structures (fun, community project!)

CCDC

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

CSD Teaching Subset

- 750+ Structures for educational purposes
  - Drug molecules
  - Fundamental Chemistry
  - Metal-Organic Frameworks
- Available on-line or in free Mercury
- Curated list of teaching materials and other useful details

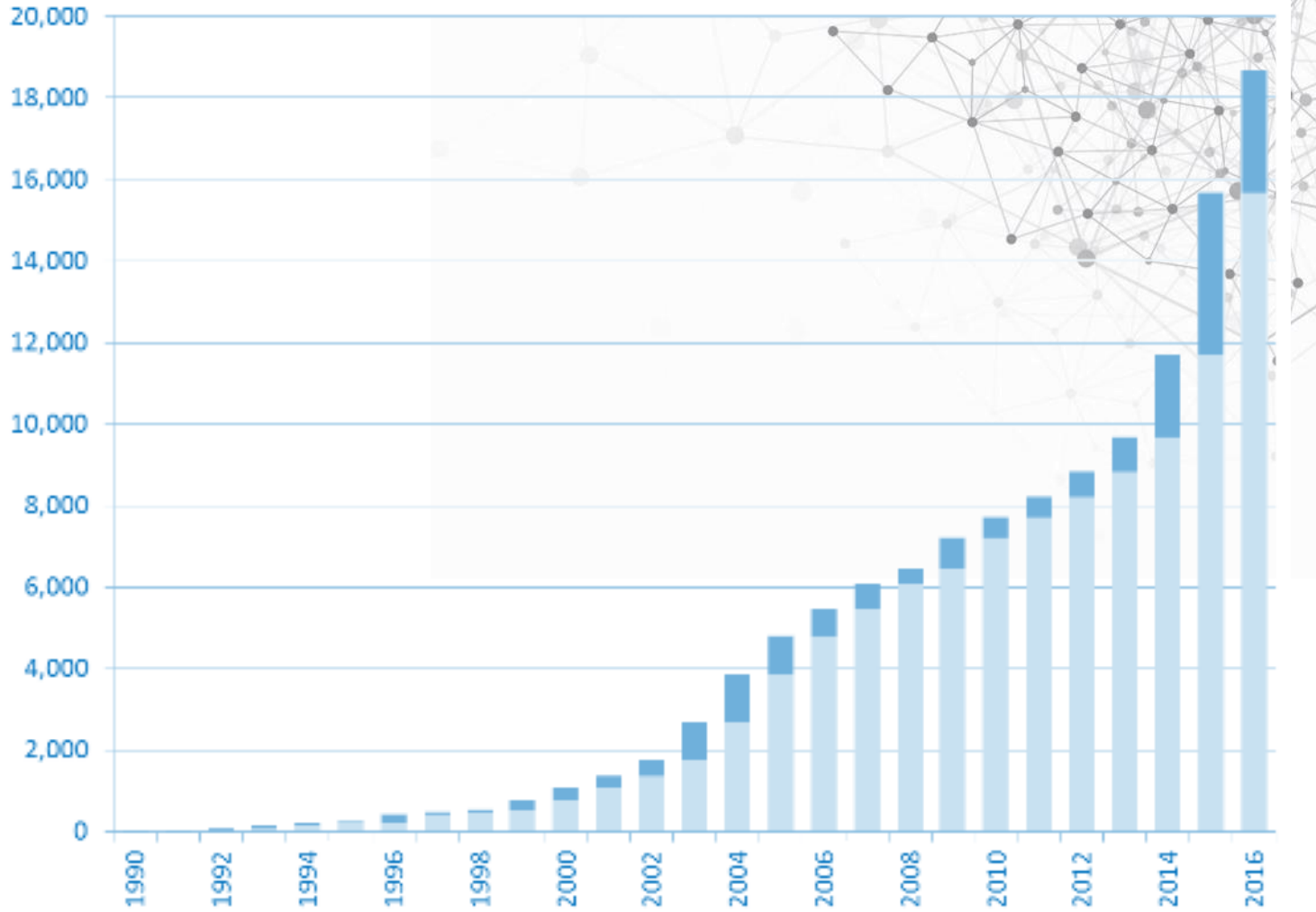
A	F	G	
NAME	Point Group	Teaching Sheet	Compound Name
ACABRH02	C_{2v}	Metal Coordination	Acetylacetonato-dicar
ACAJIX	C_3	Transition Metals (oct	tris(2-(4-fluoromethyl
ACALDA	C_{3v}		Acetaldehyde-ammor
ACANIL01			Acetanilide
ACAQUR		Metal Coordination	bis(2-(2-Methoxypher
ACARBM01		VSEPR Shapes (tetrah	Ammonium carbamat
ACASED		Metal Coordination	bis(Hydrogen tris(3,5-
ACAZEK		Metal Coordination	catena-[bis(mu!2$-Br
ACCAAH	C_i		Acetylenedicarboxylic
ACCTHP			3-Carboxy-2-acetyl-th
ACENYL01		Geometric Isomers (ci	Acenaphthylene
ACEPOO	C_{2v}	Functional Groups (ar	5-Bromo-1,3-dichloro-
ACEQII		Transition Metals (oth	bis(1,2-bis(Dimethoxy
ACETAC07		Molecules of Interest	Ethanoic acid
ACETPH		Functional Groups (ke	Acetophenone
ACETYL03	$D_{\infty h}$	Molecules of Interest	Acetylene
ACIFEX		Optical Isomers (multi	(1S)-(-)-alpha-Pinene
ACIFIB		Optical Isomers (multi	(1S)-(-)-beta-Pinene

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

CCDC

CSD Communications

- *CSD Communications* allow you to publish data directly through the CSD
- More than 5,000 CSD Communications in 2016



CCDC

Find us on social media!

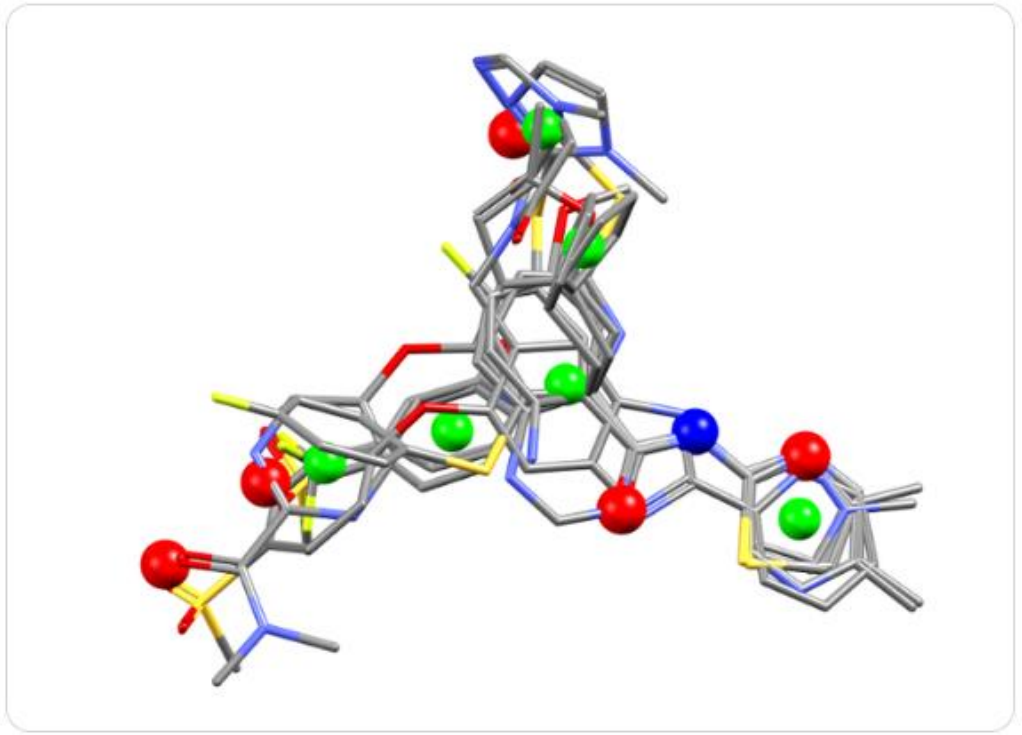
facebook.com/ccdc.cambridge

@ccdc_cambridge

CCDC Cambridge
@ccdc_cambridge

Join us on the online workshop: structure-based drug design with CSD-Discovery on Wednesday 27th May, at 3pm (BST). Learn more about Mogul, SuperStar, CSD-CrossMiner and GOLD. Register your place here hubs.ly/H0qswHC0

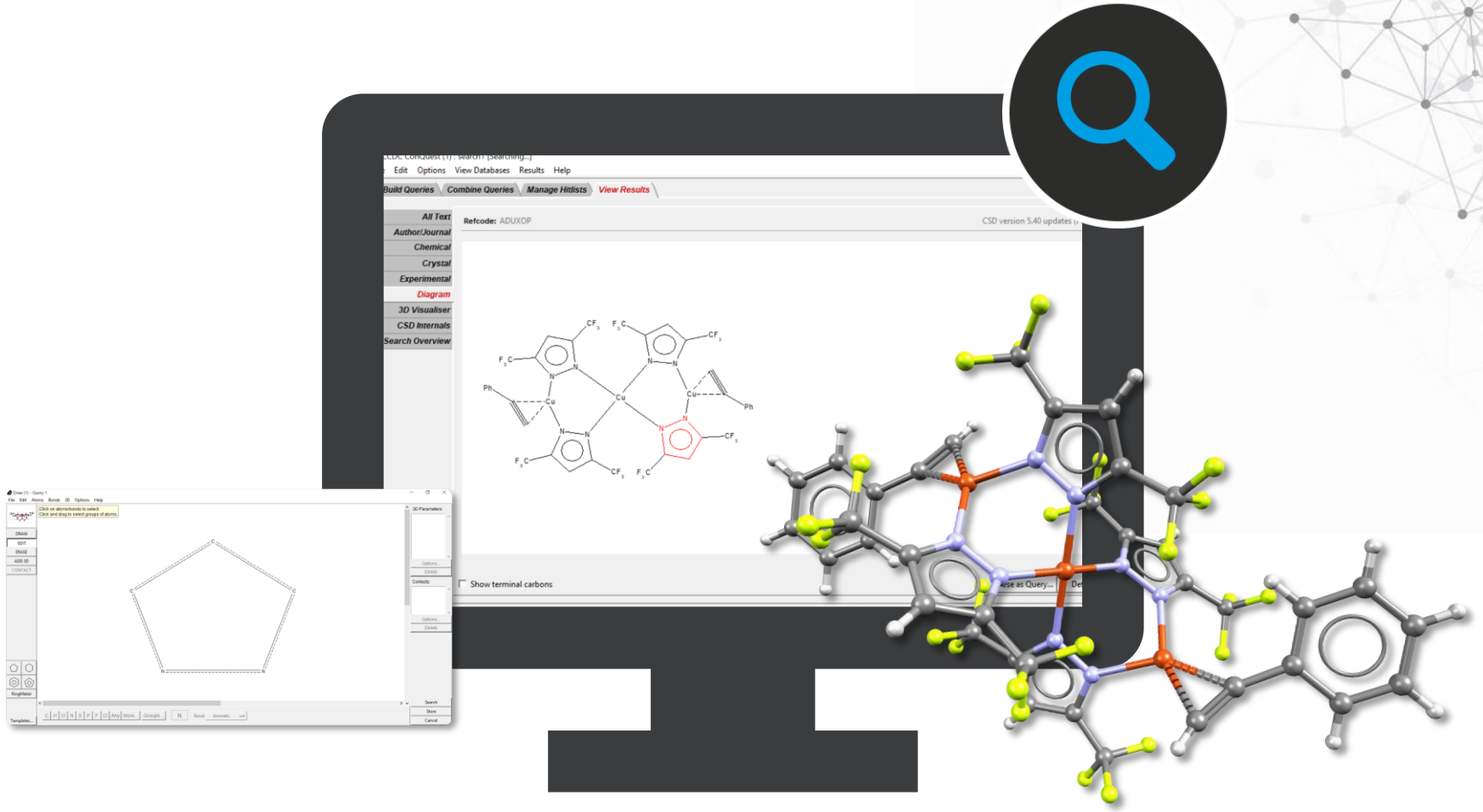
#DrugDiscovery #CSDDiscovery #remotelearning



10:54 AM · May 18, 2020 · HubSpot

CCDC

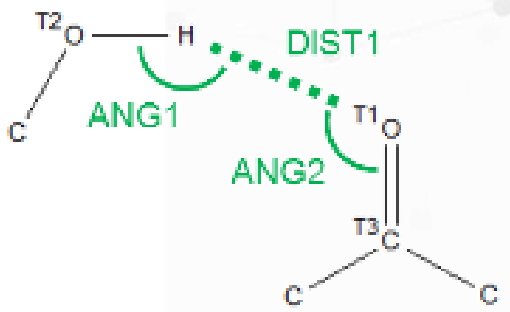
CSD-System: Search

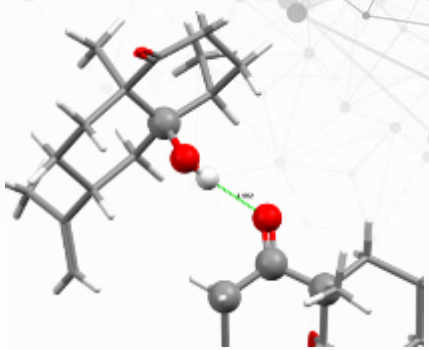


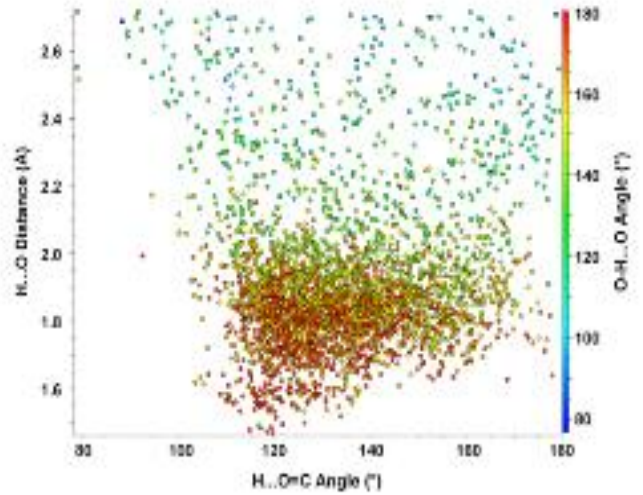
CCDC

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 substructure searching
  - 3D Geometrical searching
  - Intermolecular non-bonded contact searching



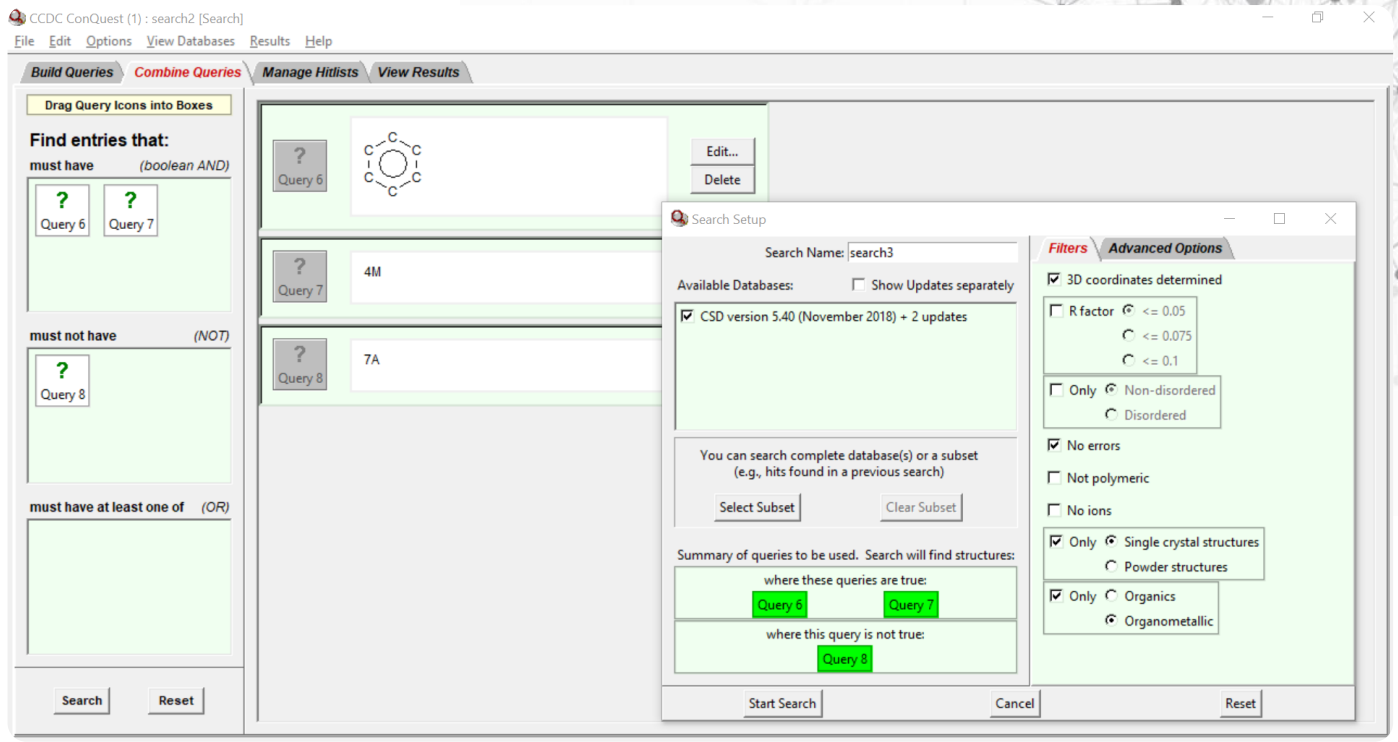




CCDC

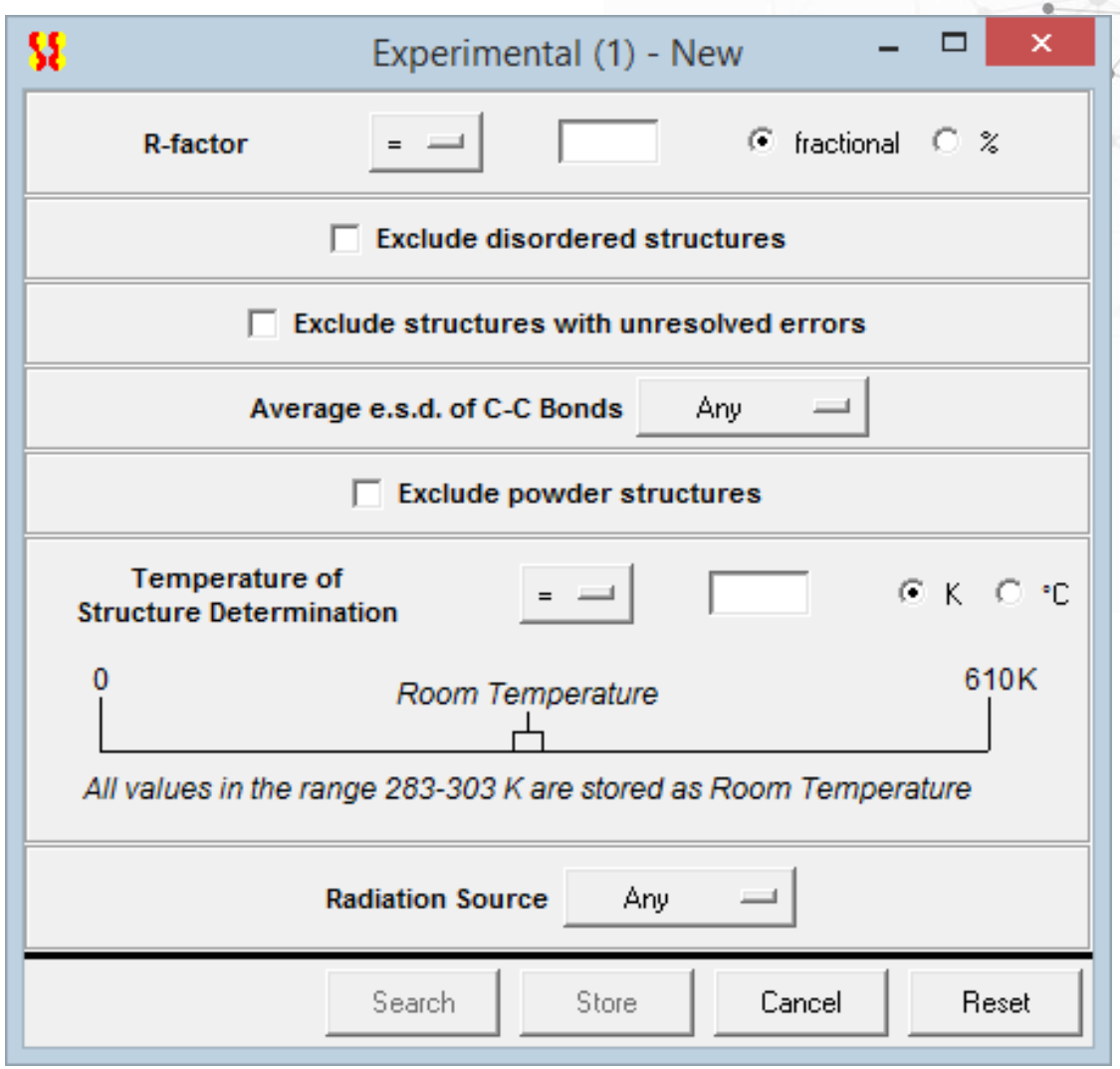
Searching with ConQuest

- Large range of filtering options based on experimental or chemical considerations (e.g. R-factor, temperature, only organics)
- Combine searches using various Boolean options and manage hitlists post-search
- Export results to Mercury for visualization and advanced numerical analysis & plotting



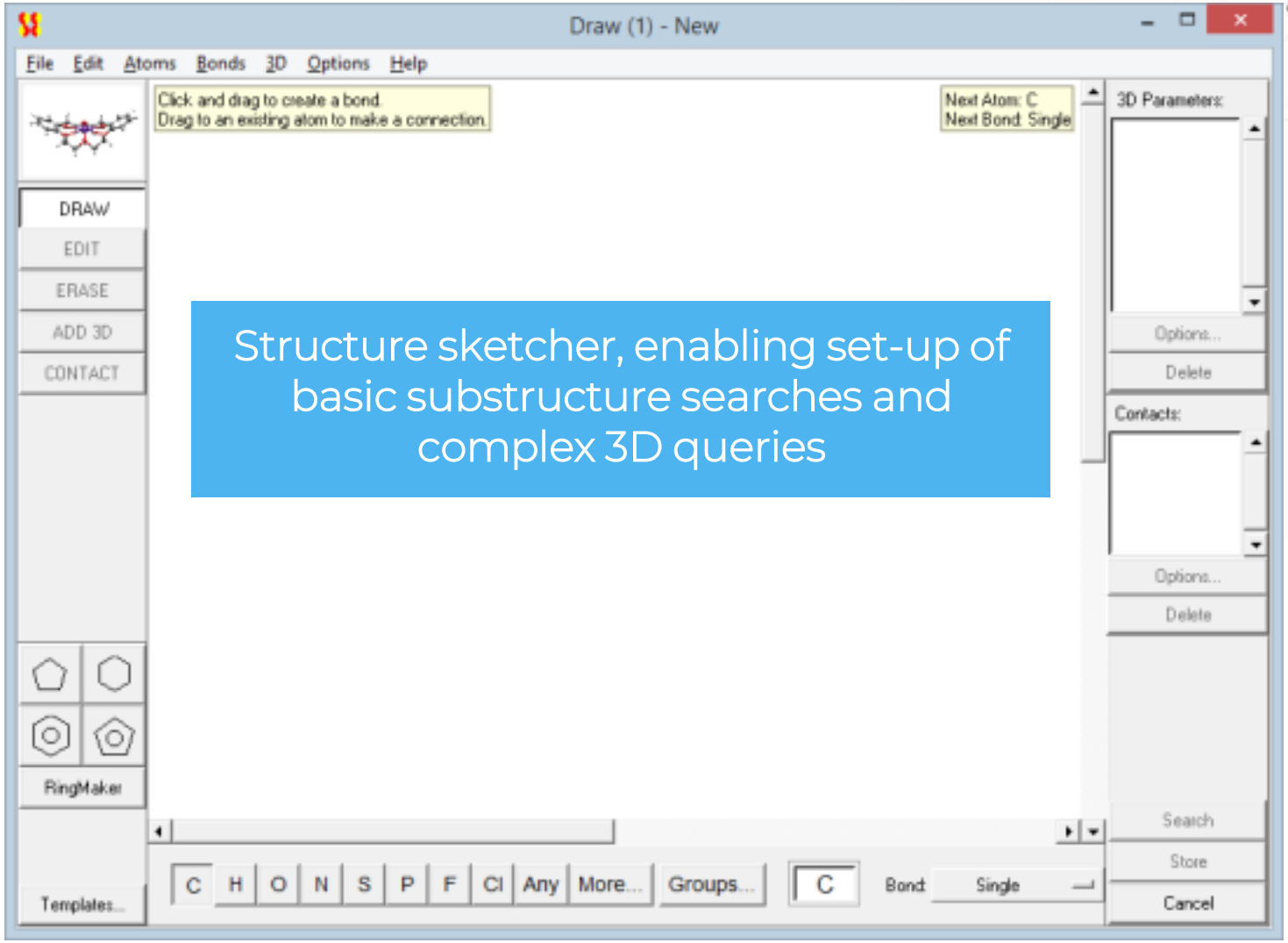
ConQuest features

Search for information relating to the structure determination



CCDC

ConQuest features

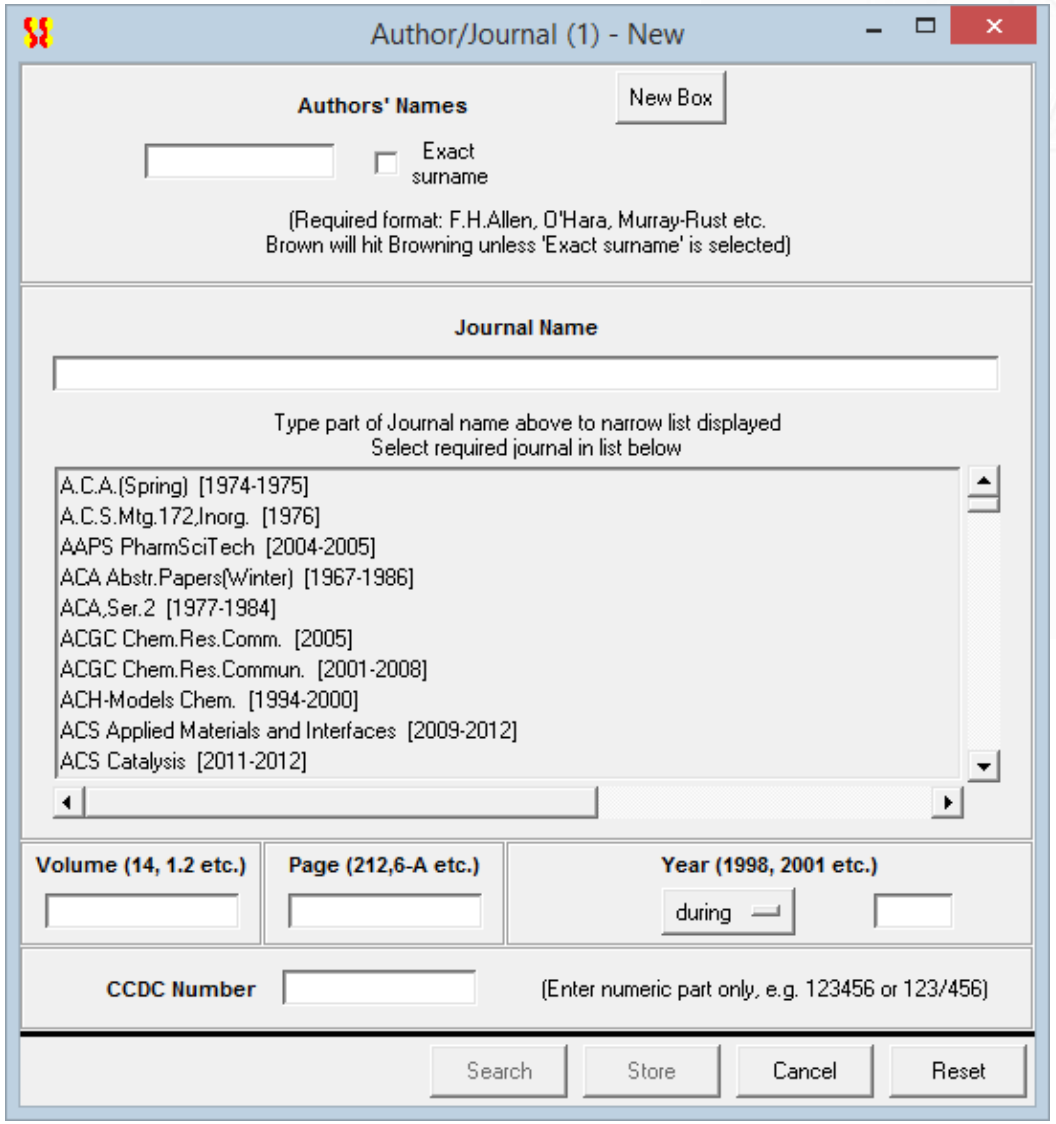


Structure sketcher, enabling set-up of basic substructure searches and complex 3D queries

CCDC

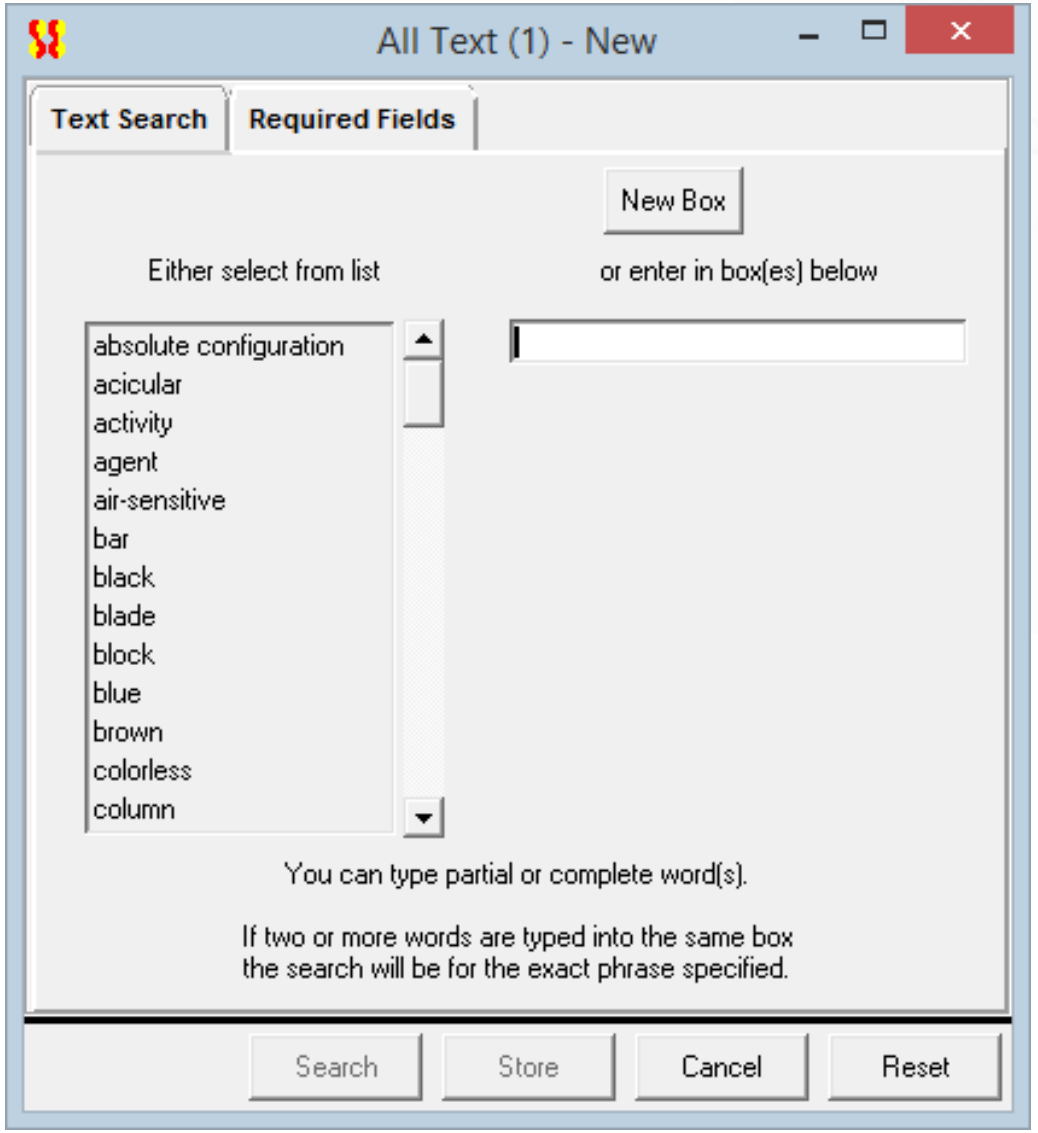
ConQuest features

Bibliographic search



ConQuest features

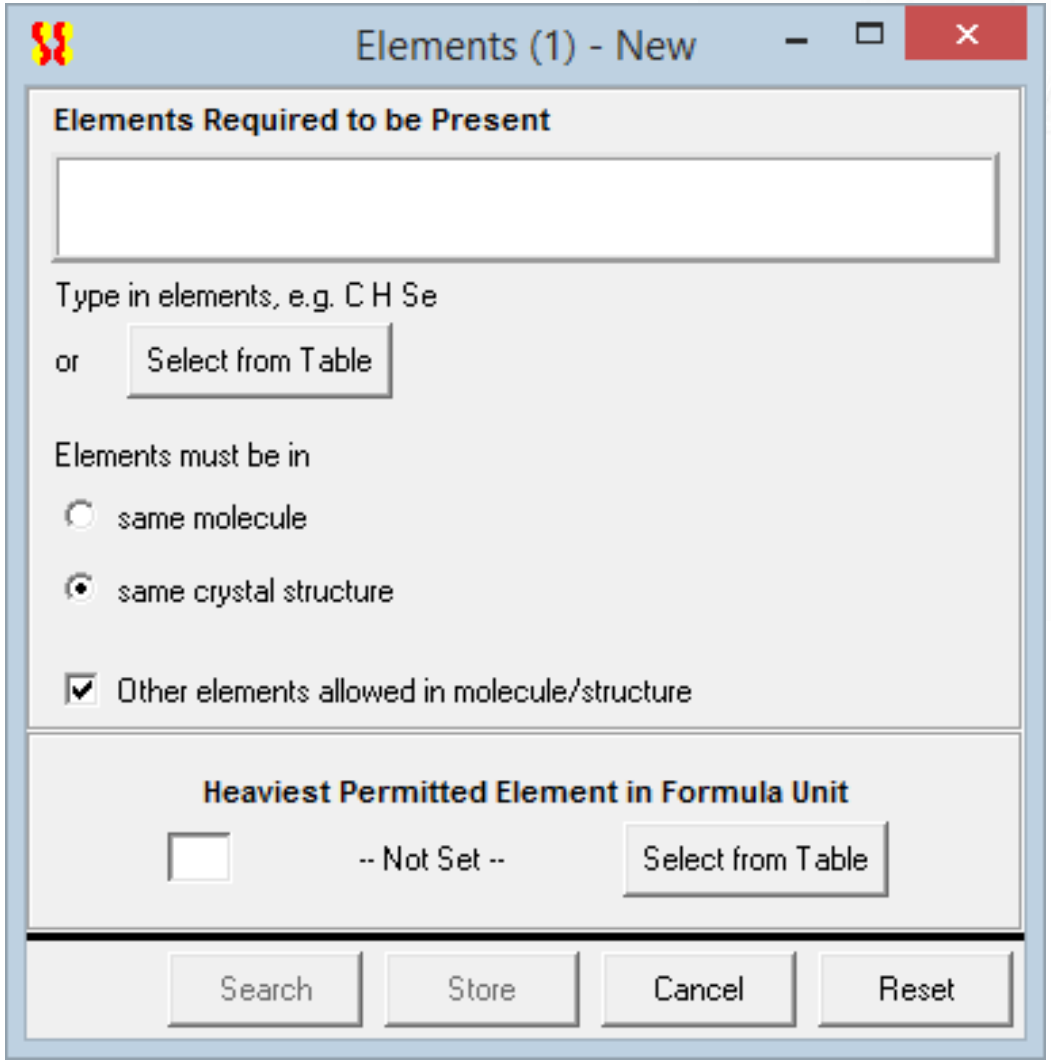
Generic text search



CCDC

ConQuest features

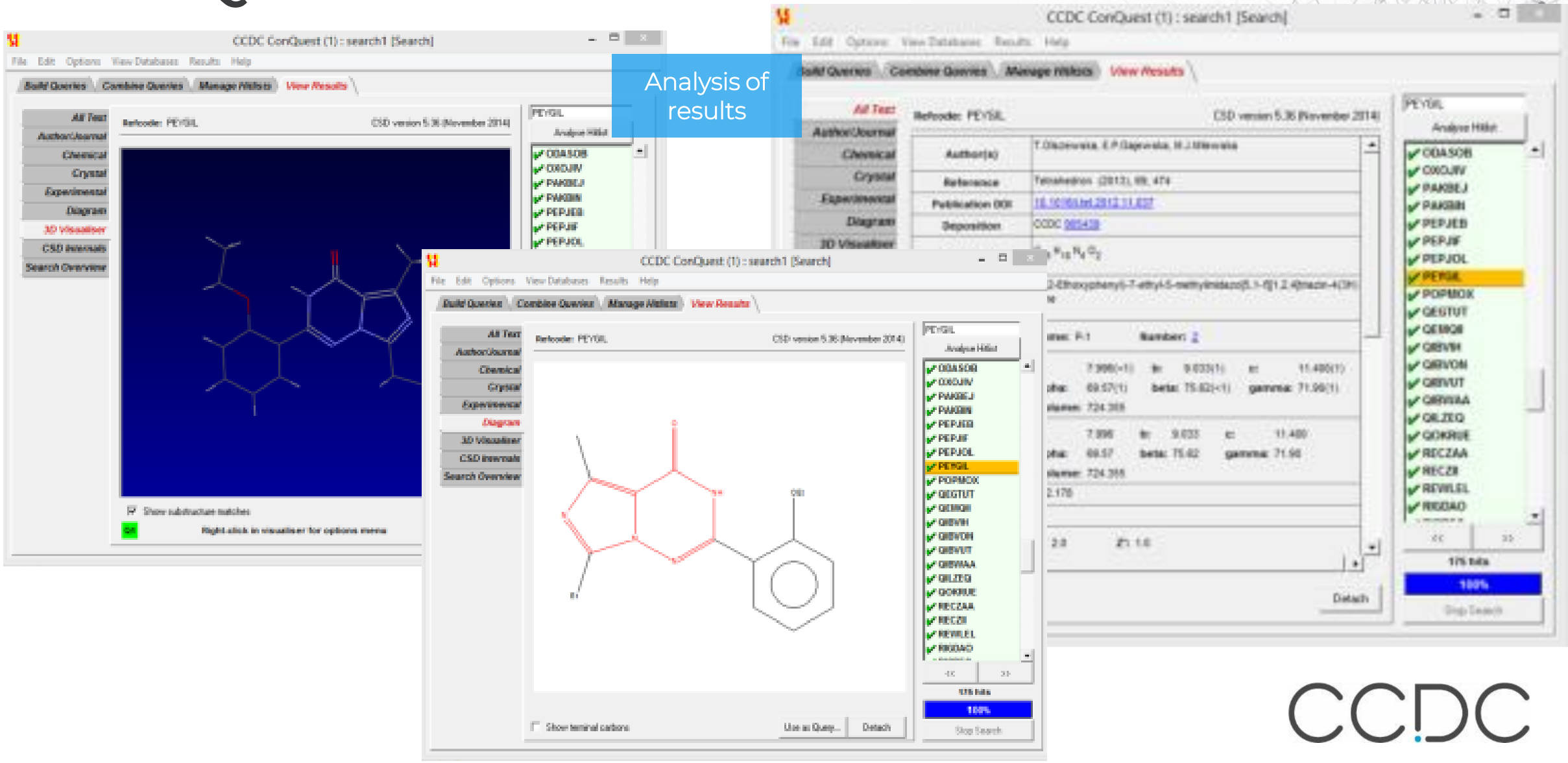
Elemental make-up



CCDC

ConQuest features

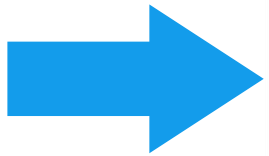
Analysis of results

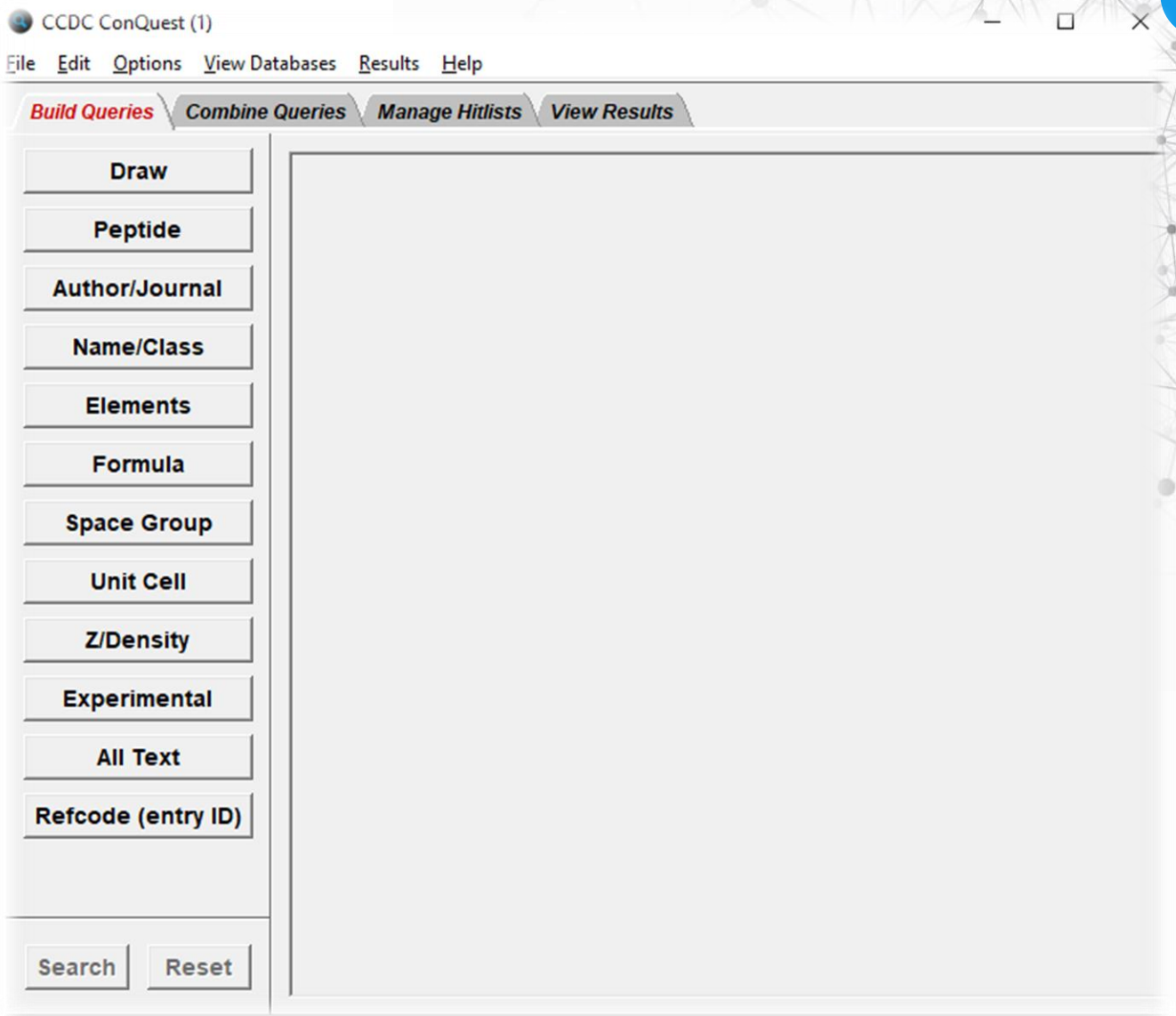


CCDC

ConQuest – Basics

1. Open ConQuest



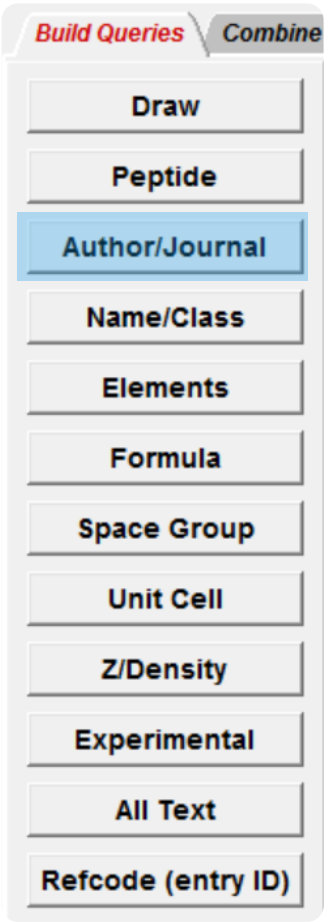


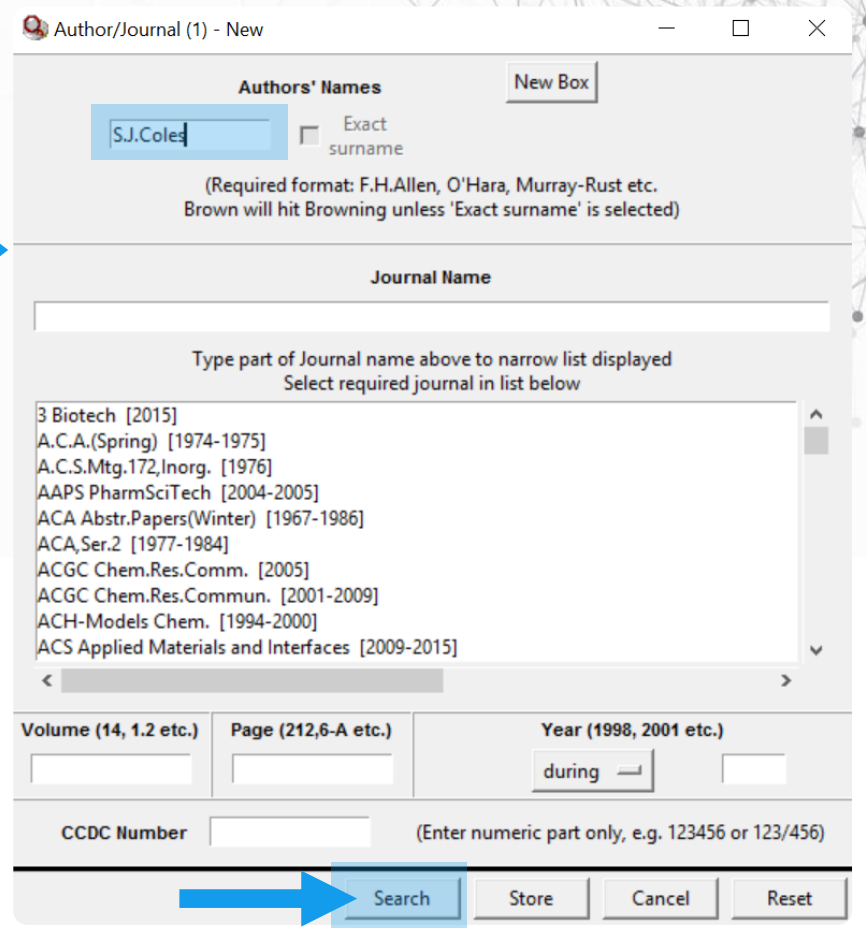
CCDC

ConQuest – Basics

2. Build Queries:

Author Search

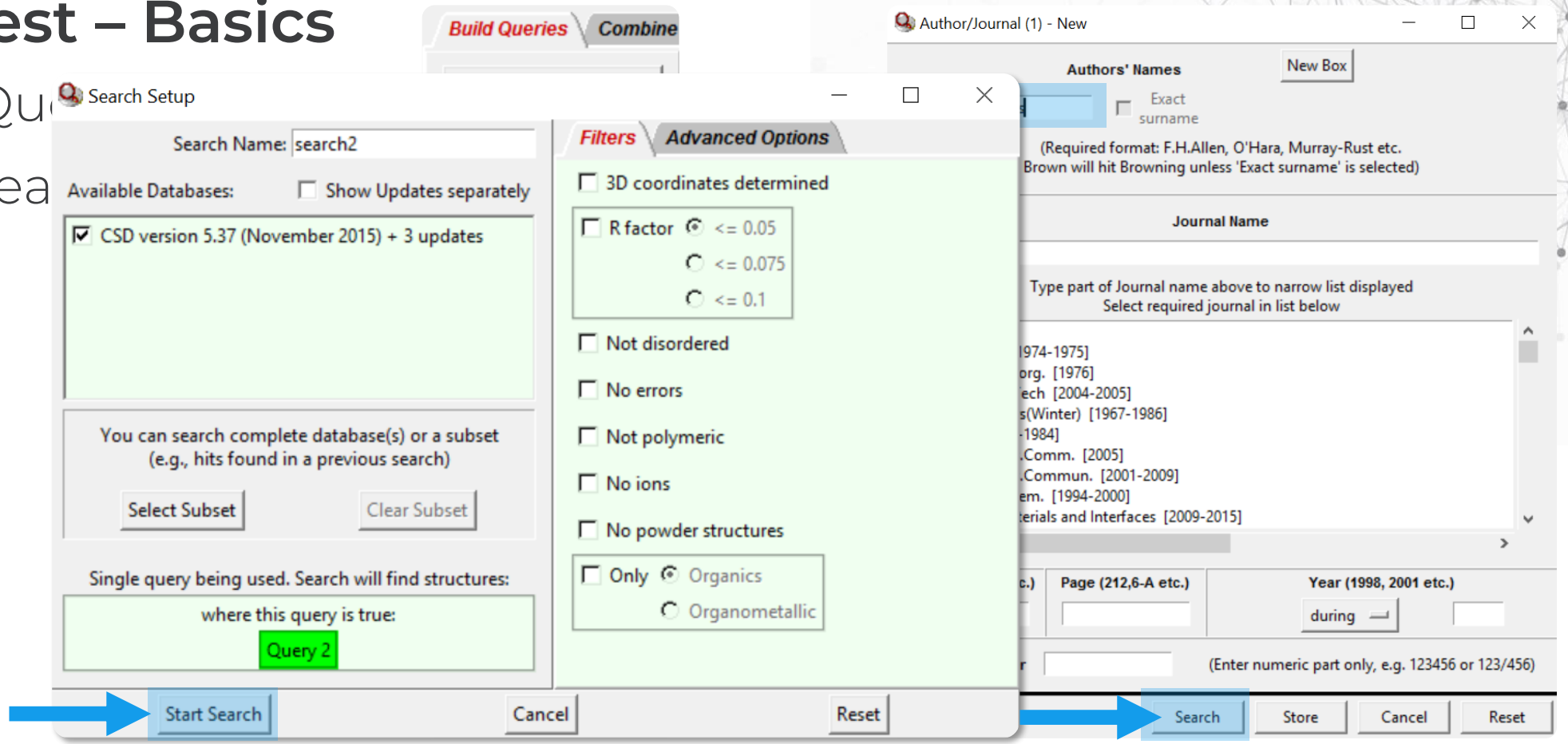




ConQuest – Basics

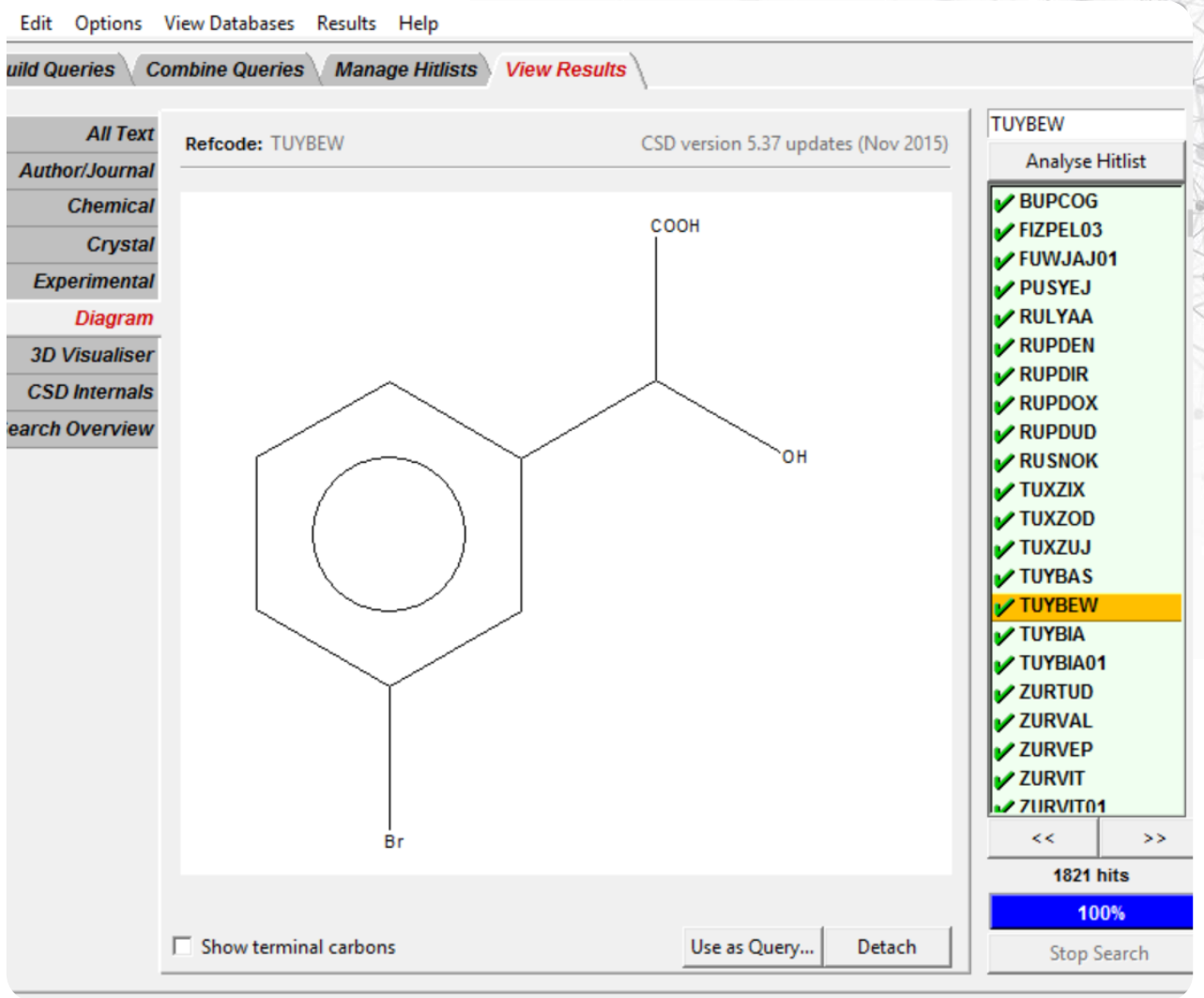
2. Build Qu

Author Sea

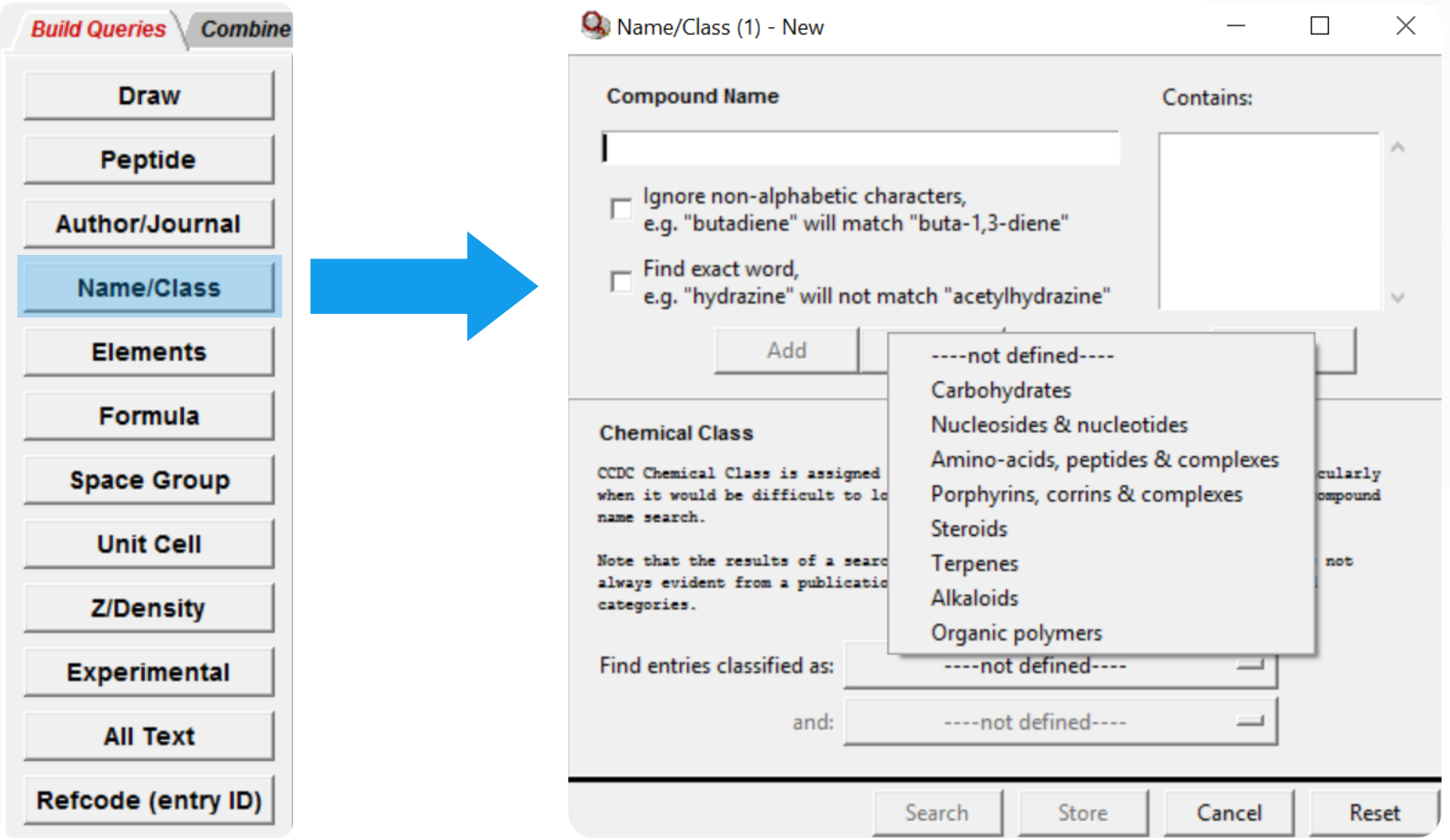


ConQuest – Basics

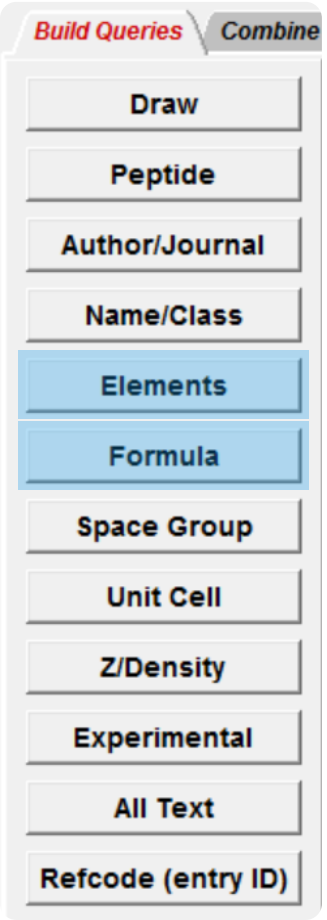
3. View results

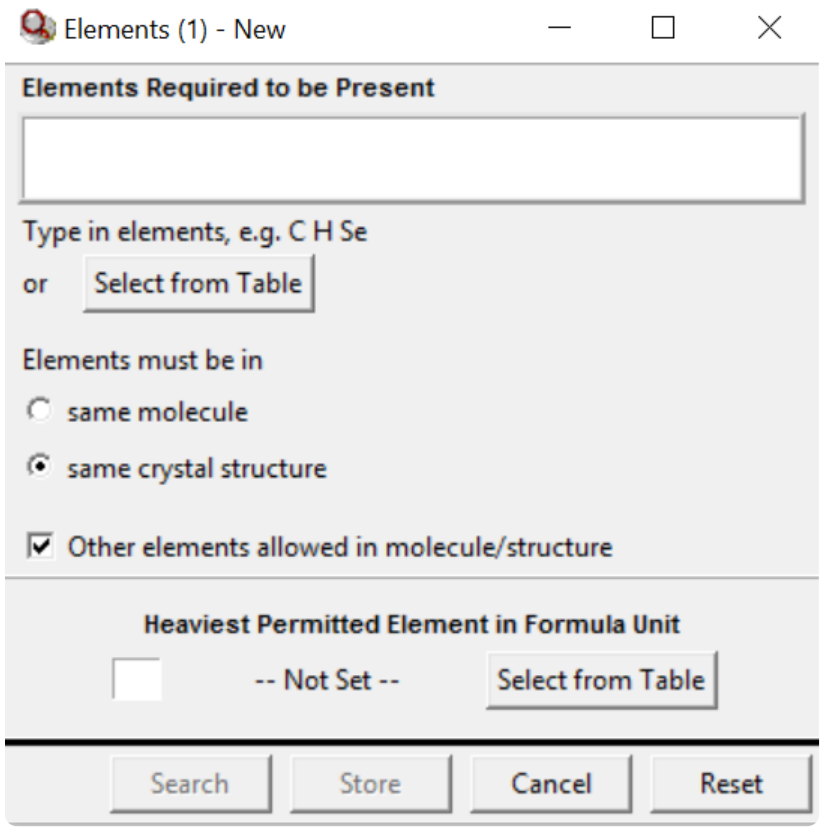


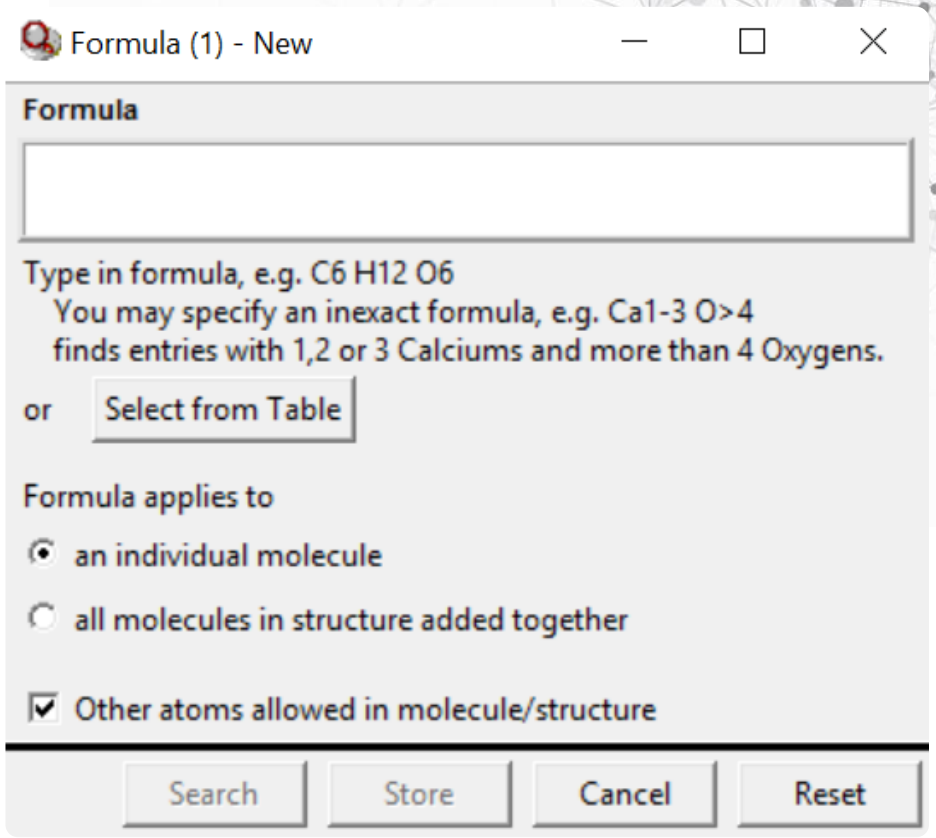
ConQuest – Compound name search



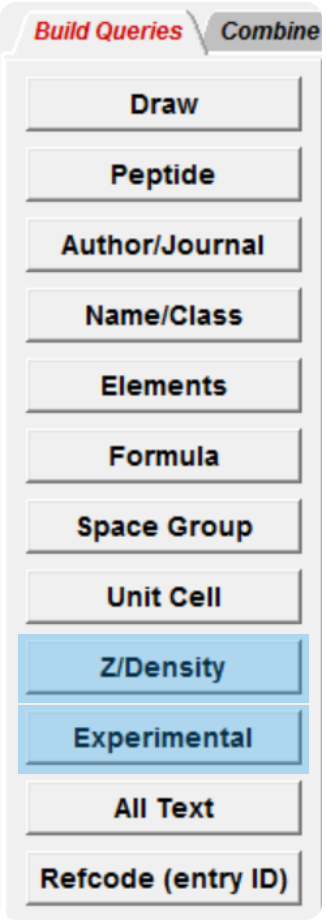
ConQuest – Elements & formula search

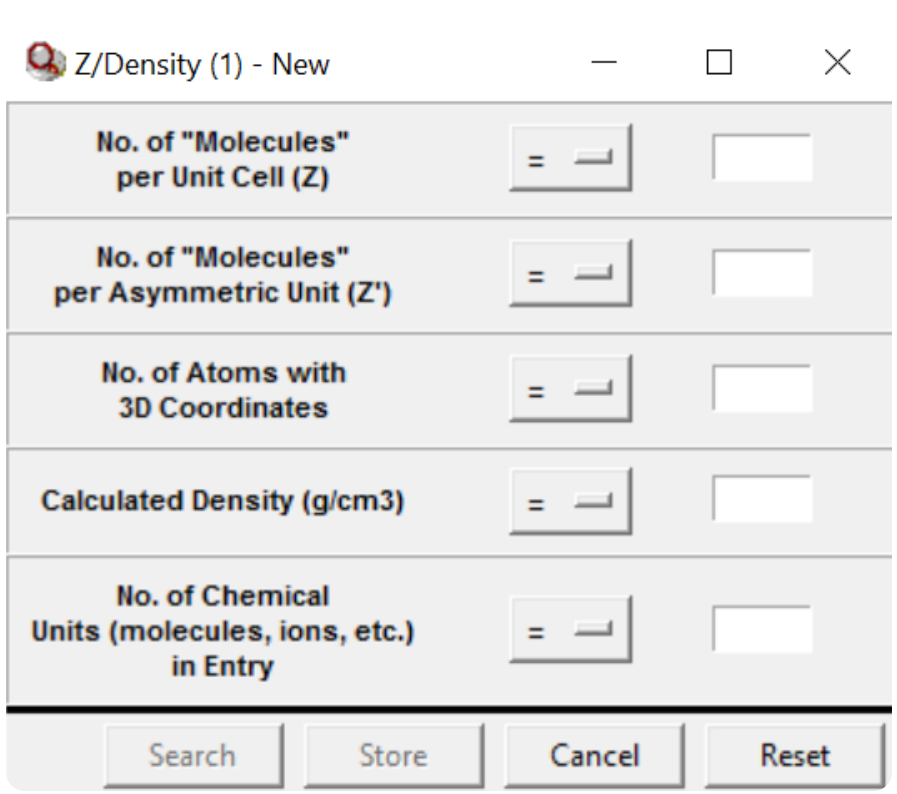


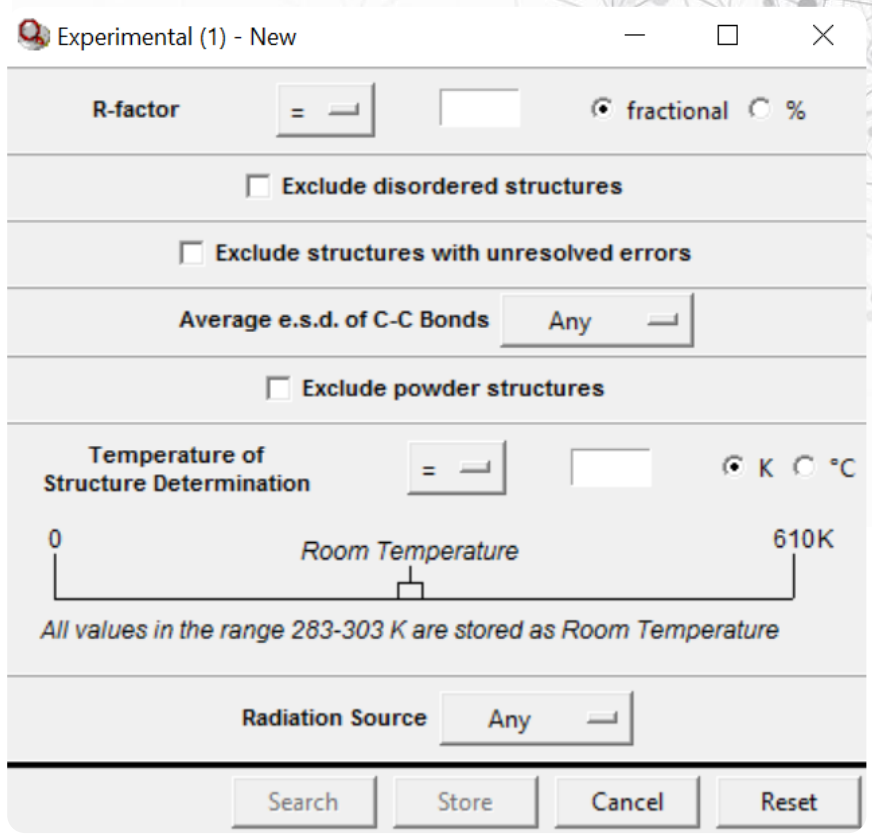




ConQuest – Z/Density and experimental info search

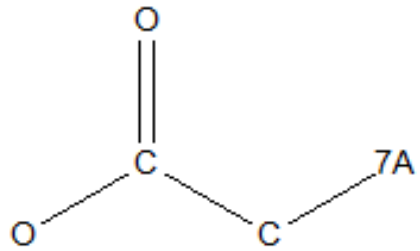




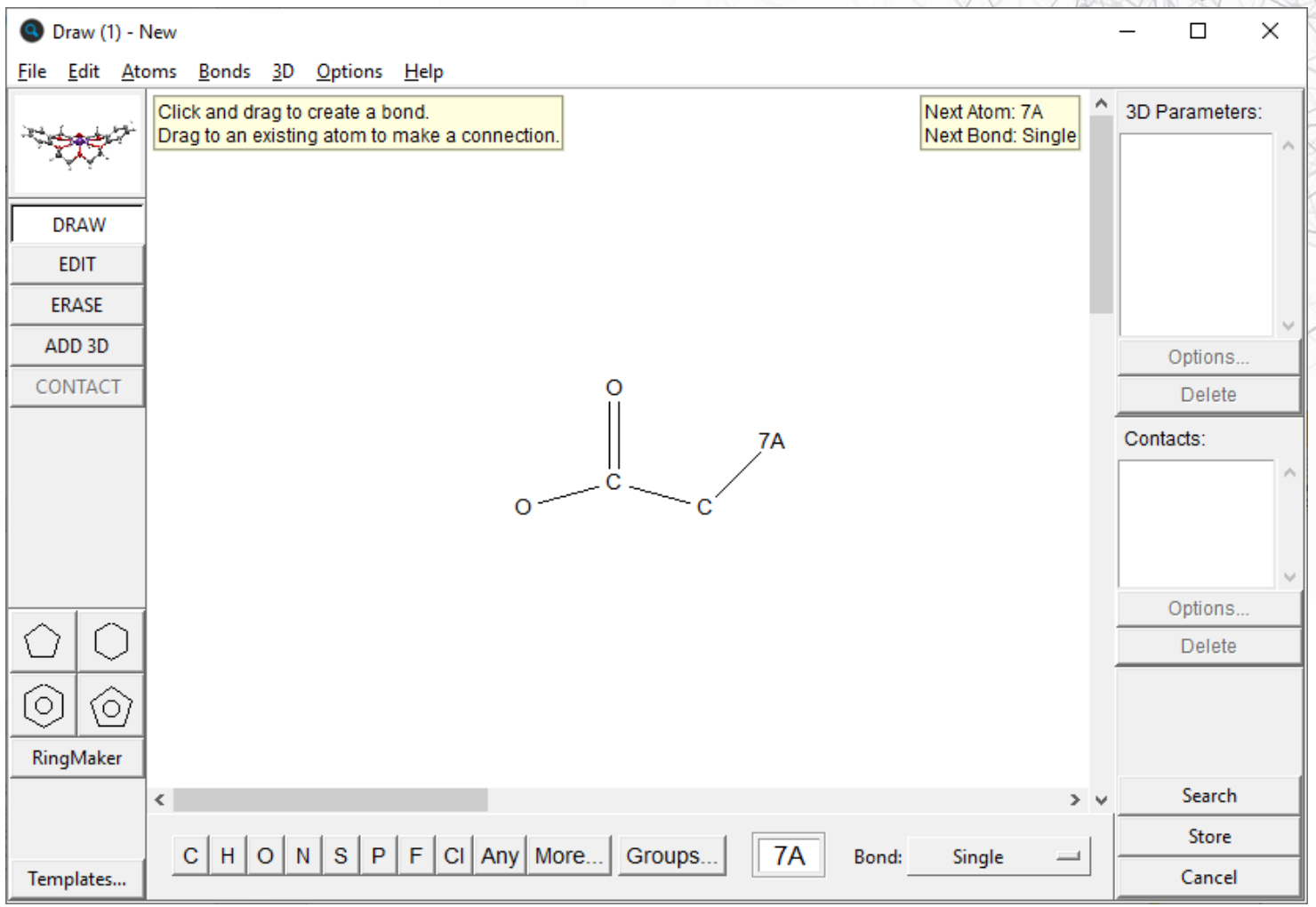


ConQuest substructure search

Search for:



Where 7A represents any halogen



What if you don't have a copy of ConQuest?

WebCSD to the rescue!

WebCSD (“Access Structures”)

CCDC

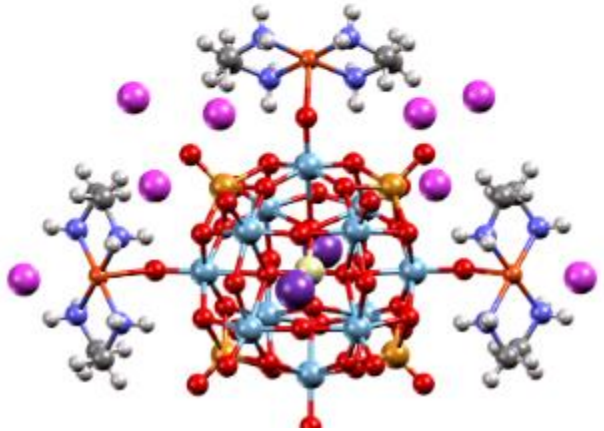
Community ▾ | Research & Consultancy ▾ | Solutions ▾ | News & Events ▾ | Suppor

- Community
- Blog
- Deposit a Structure
- Access Structures
- CSD-Community
- International Year of the Periodic Table
- Awards and Sponsorship
- FAIRE Programme
- Collaborators and Partnerships
- Initiatives
- Educational Resources
- CSD Communications

HOME / (

Cor

The Cam… mmunity
achievem… ave shared
the result… CCDC –
allowing … committed to
enabling … ensive and
up-to-dat…



Awar…
Depo…
CSD-…
Colla…

CCDC

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

CCDC FIZ Karlsruhe Leibniz Institute for Information Infrastructure

WebCSD

Sign In

Licensed to: CCDC Main Site

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

Simple text and numeric searching

Welcome to WebCSD. This service now includes the ability to search for inorganic structures through the CCDC's and FIZ Karlsruhe's joint Access Service using the Simple Search tab. Please use one or more of the boxes to find entries. If you enter details in more than one field the search will try to find records containing all the terms entered. More information and search help

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

+ Add New Search Field

Search

Clear

CCDC

WebCSD

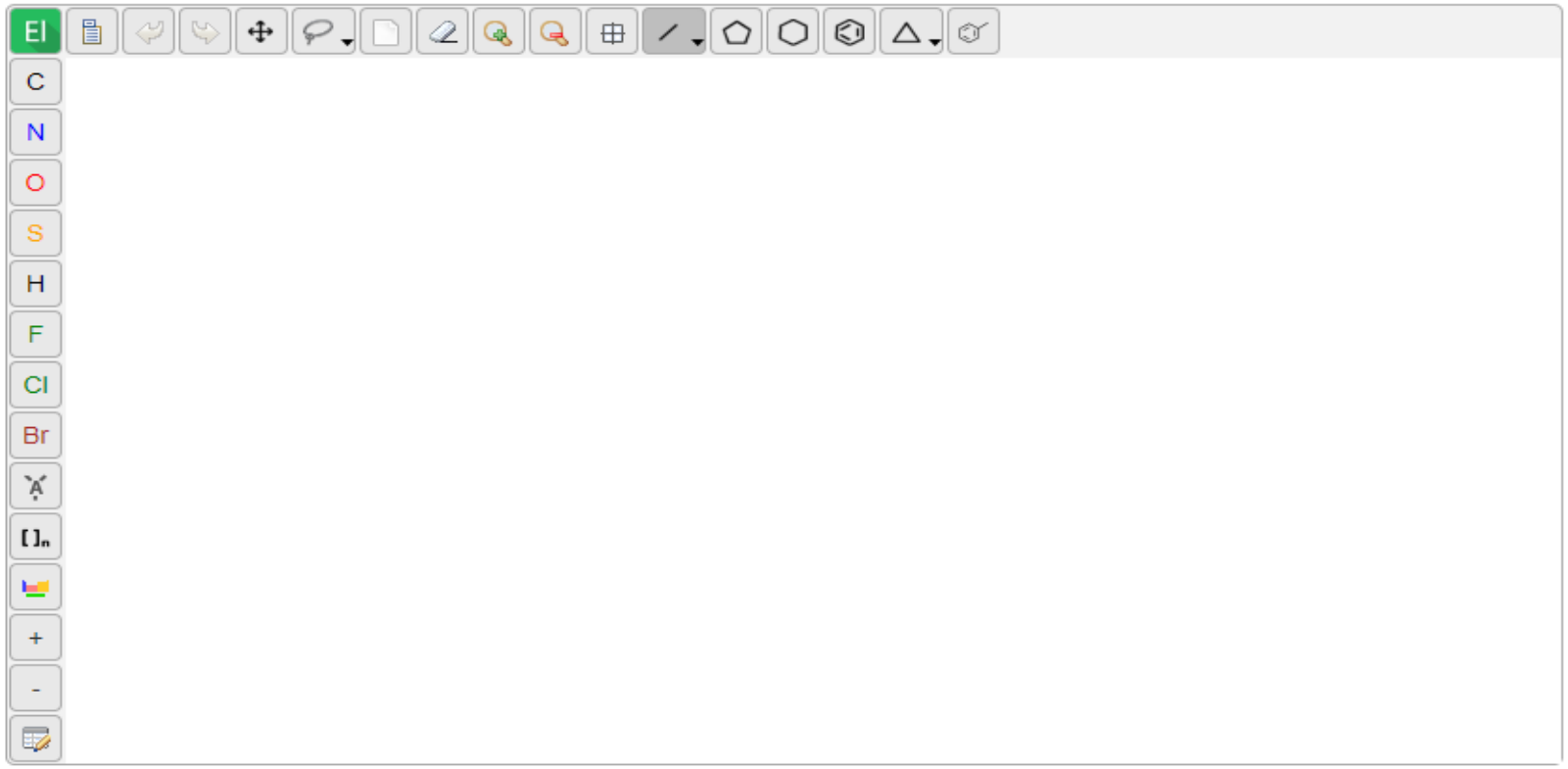
Sign In

Licensed to: CCDC Main Site

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

Chemical structure searching

Please draw your diagram or add a SMARTS string in the 'advanced' section below.



Match condition: Exact | Substructure | Similarity

Advanced

Search | Clear

Help

Keyboard shortcuts

- Copy: **Ctrl-C**
- Delete: **Ctrl-X**
- Paste: **Ctrl-V**
- Undo: **Ctrl-Z**
- Redo: **Ctrl-Y**
- Select all: **Ctrl-A**

Query features

Query features describe how an atom or bond should behave in substructure searches. To add a feature:

1. Right click on atom or bond
2. Hover over 'query features' (atom only)
3. Hover over a feature type (e.g. H-count, type)
4. Select one of the options

More Information

CCDC

CSD Entry: AGAGIA

Sign In

Licensed to: CCDC Main Site

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

Search Complete - **219** Results Found

100%

Modify Search | New Search

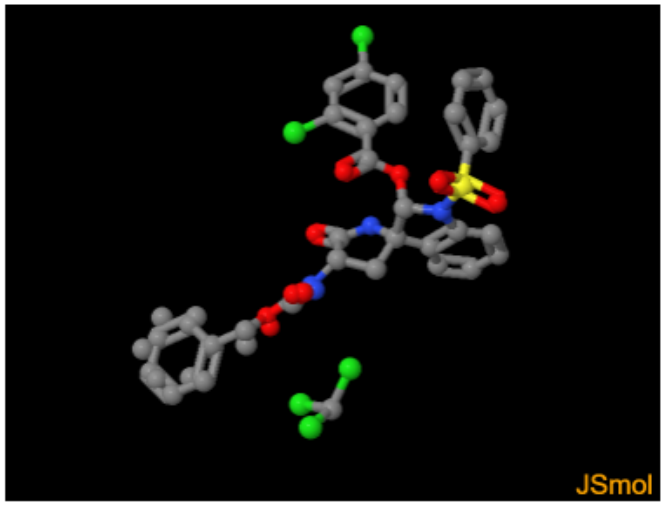
Results

	Database Identifier	Deposition Number
☐	ABECEP	1100088
☑	ACEYAJ	255902
☐	ADEPAB	613666
☐	ADEPEF	613667
☐	AGAGEW	933112
☐	AGAGIA	962016
☐	BAFWOU	1105049
☐	BAMBOJ	1528977
☐	BEKLOV	1569815
☐	BEXHOE	1822669
☐	BOBZEA	1829117
☐	BOMAIU10	1113591
☐	BOXRIR	1029163
☐	BUFFIS	750799

AGAGIA : 4'-(((Benzyloxy)carbonyl)amino)-5'-oxo-1-(phenylsulfonyl)-1,2-dihydrospiro[indole-3,2'-pyrrolidin]-2-yl 2,4-dichlorobenzoate chloroform solvate

Space Group: P 2_1 2_1 2_1 (19), **Cell:** *a* 9.6220(13)Å *b* 18.082(2)Å *c* 19.975(3)Å, *α* 90° *β* 90° *γ* 90°

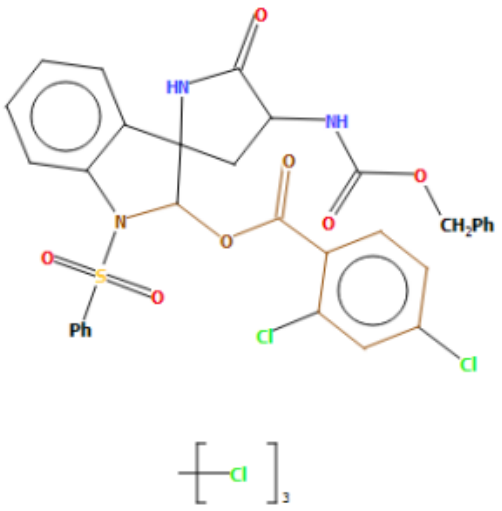
3D viewer



H | Disorder | Menu | Open

Style	Labels	Packing	Measure
Ball and Stick	No Labels	None	None

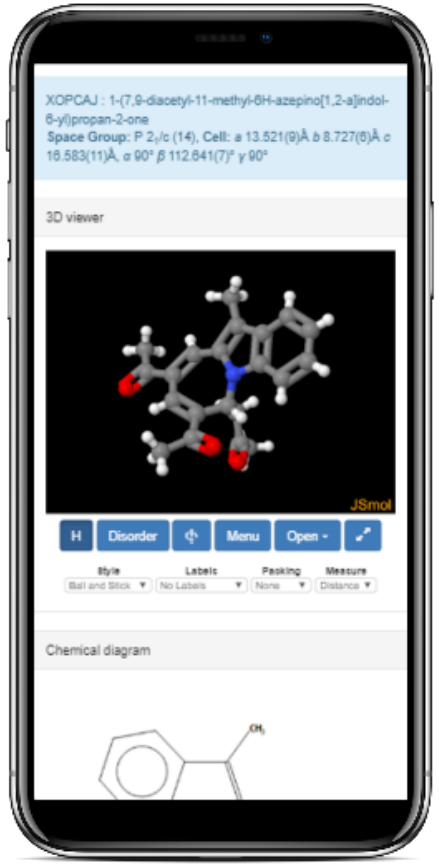
Chemical diagram

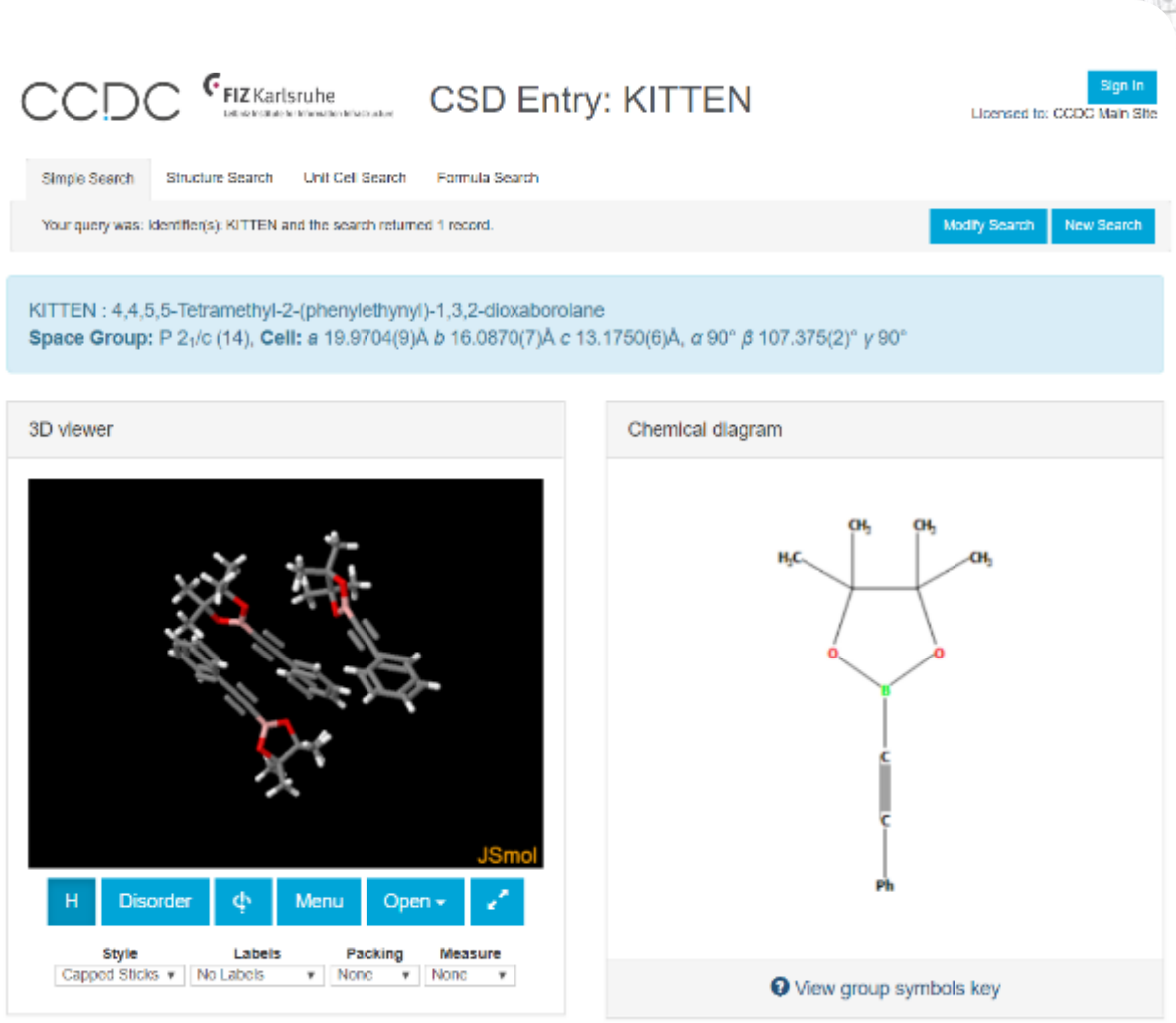


View group symbols key

WebCSD

- Search from your browser
- Accessible on mobile devices!





CCDC

Next time:

- Mercury for visualization and analysis
- MOGUL for geometry checks
- Full Interaction Maps
- CellCheckCSD
- Depositing structures

For now: let’s practice searching…

CCDC

CCDC

advancing structural science