

A faint, light gray network pattern of interconnected lines and nodes is visible in the background, resembling a molecular or structural diagram.

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



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

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

International
Repository of
Curated 3D
Structures

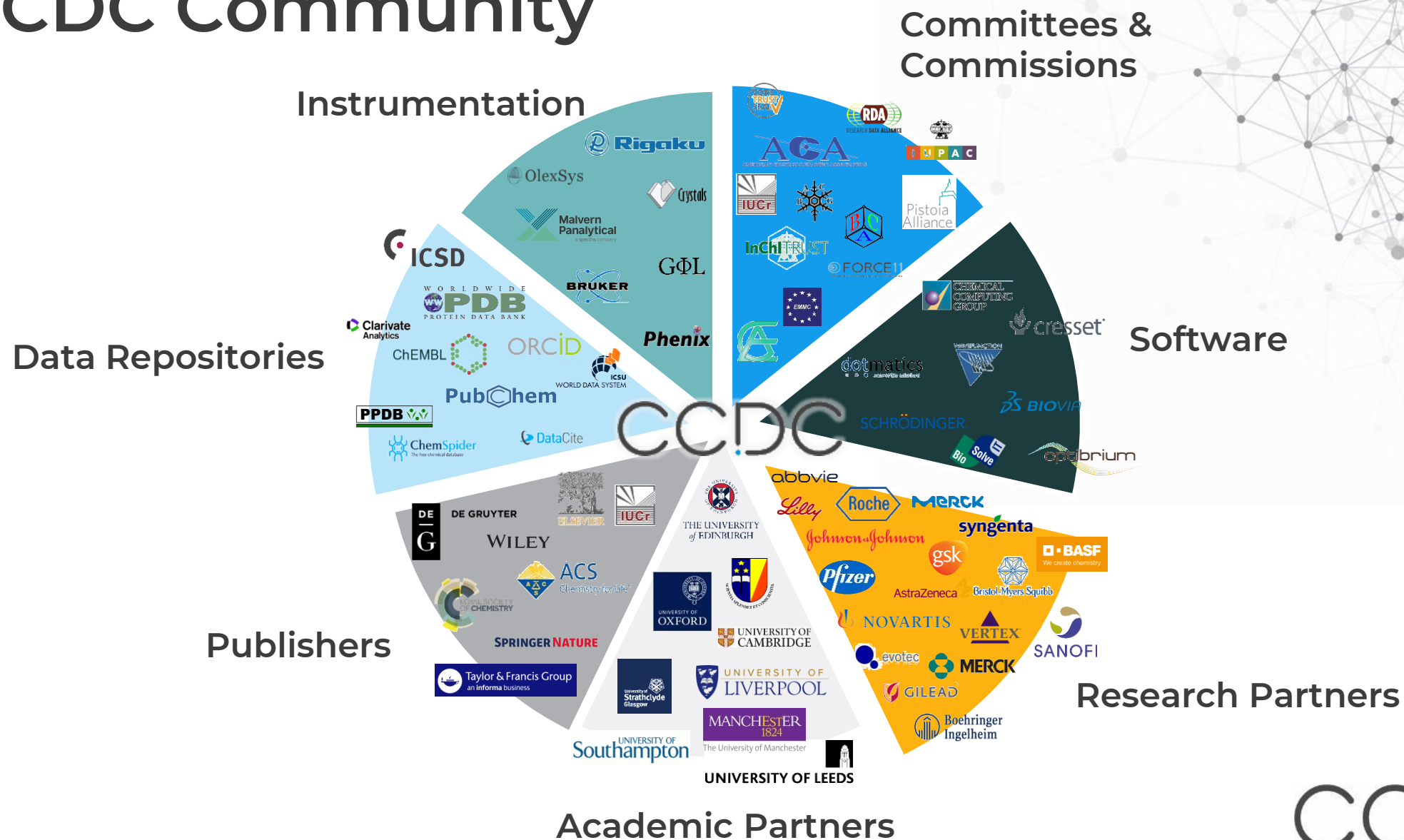
Collaborative
Research &
Knowledge-Based
Services

Scientific
Software

Education &
Outreach

CCDC

CCDC Community








CCDC

The Cambridge Structural Database (CSD)

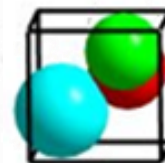
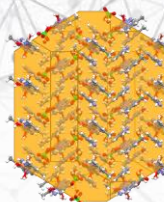
The world's only fully curated & enhanced database for small-molecule organic and metal-organic crystal structures.

- 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

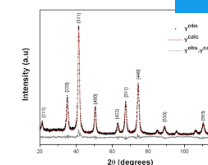
Crystal structure databases

	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!



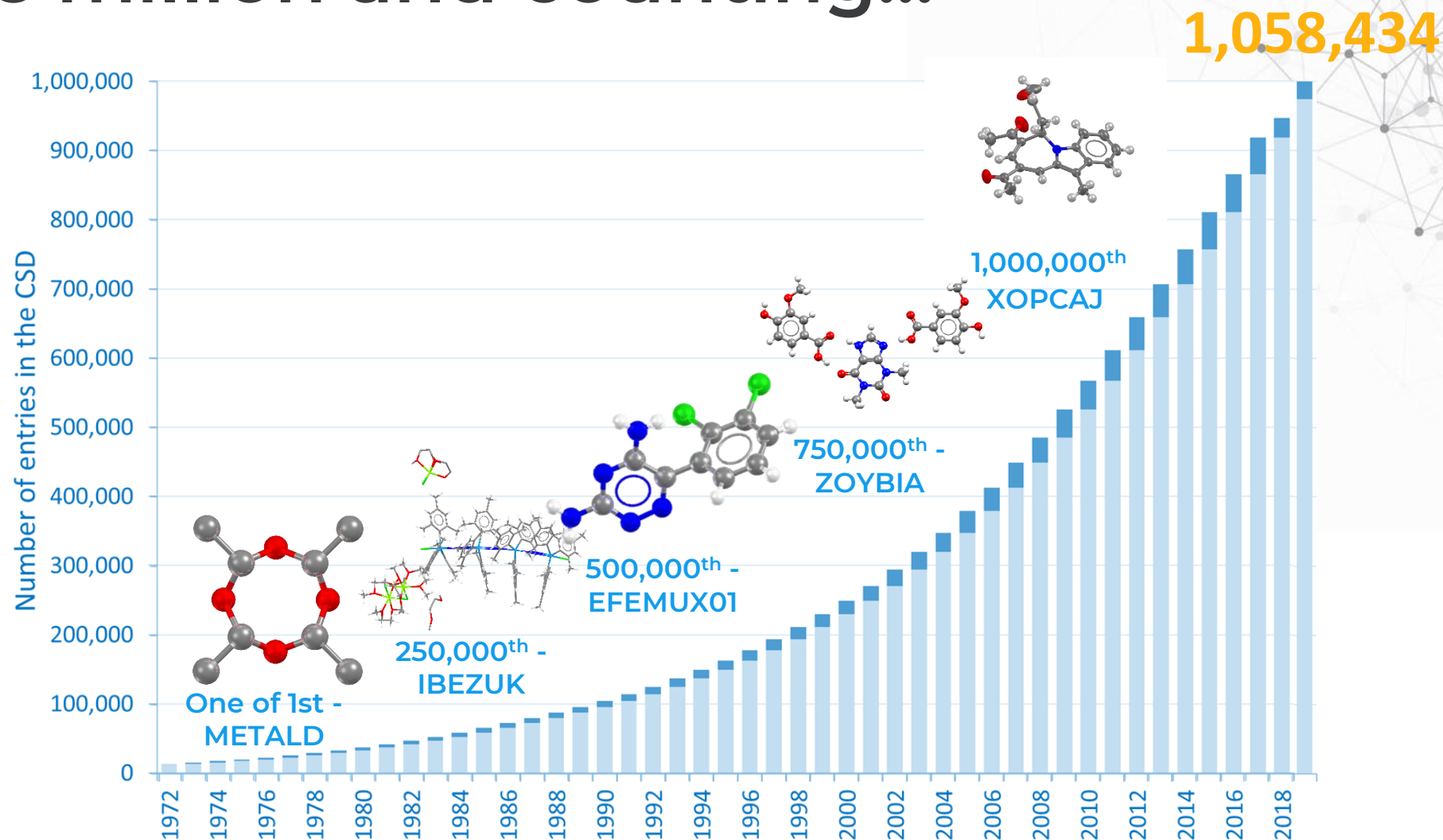
Links between ligands



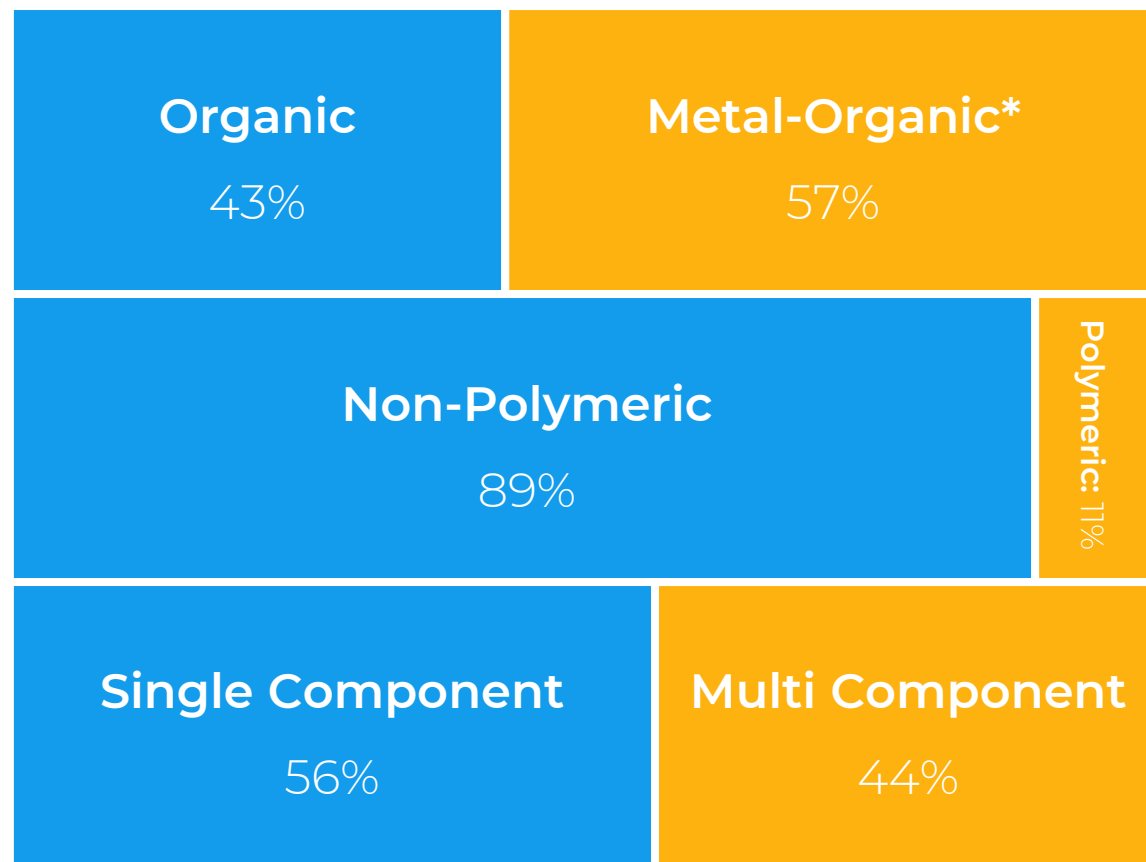
New data partnership as of 18 May!

CCDC

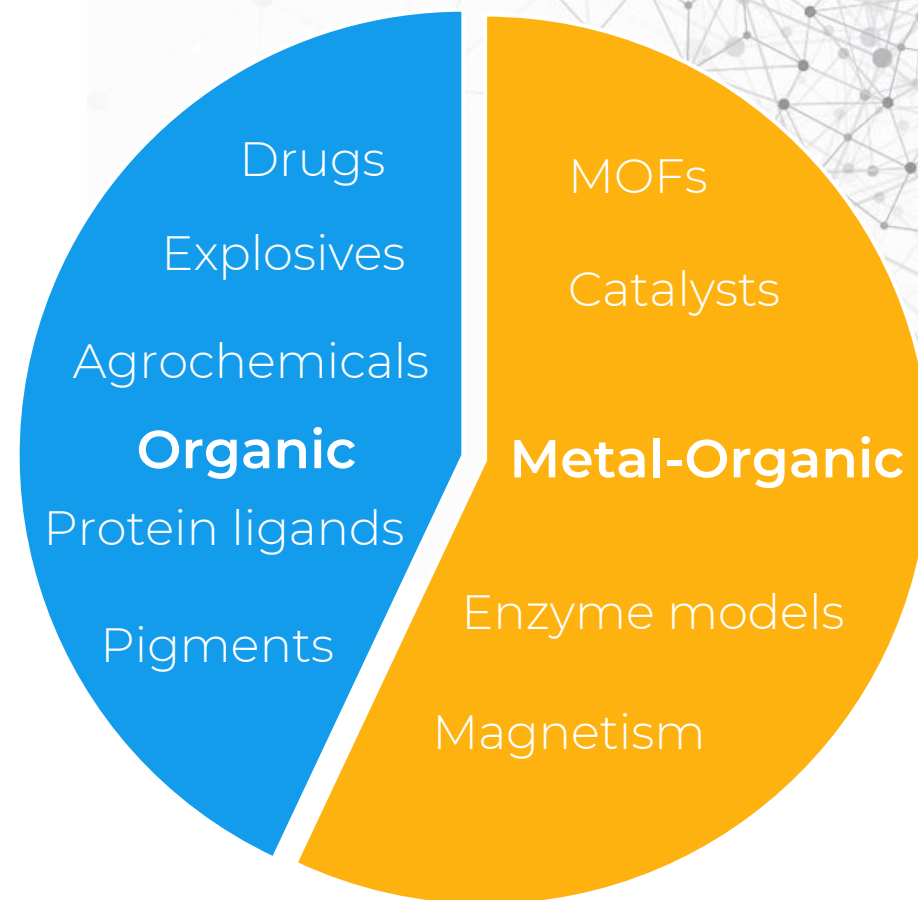
One million and counting...



What's in a million?



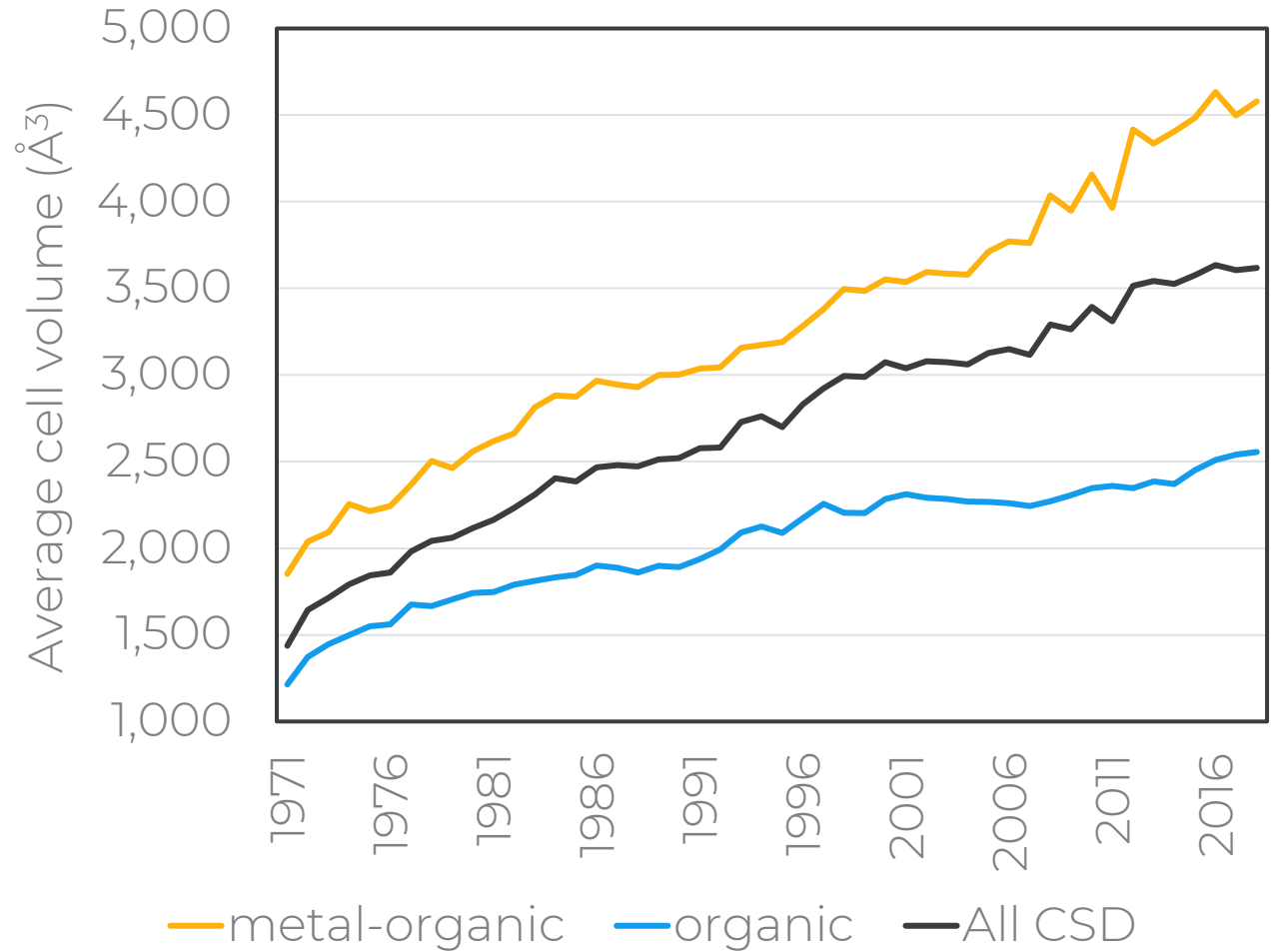
*Contains Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po, Transition Metal, or Actinide



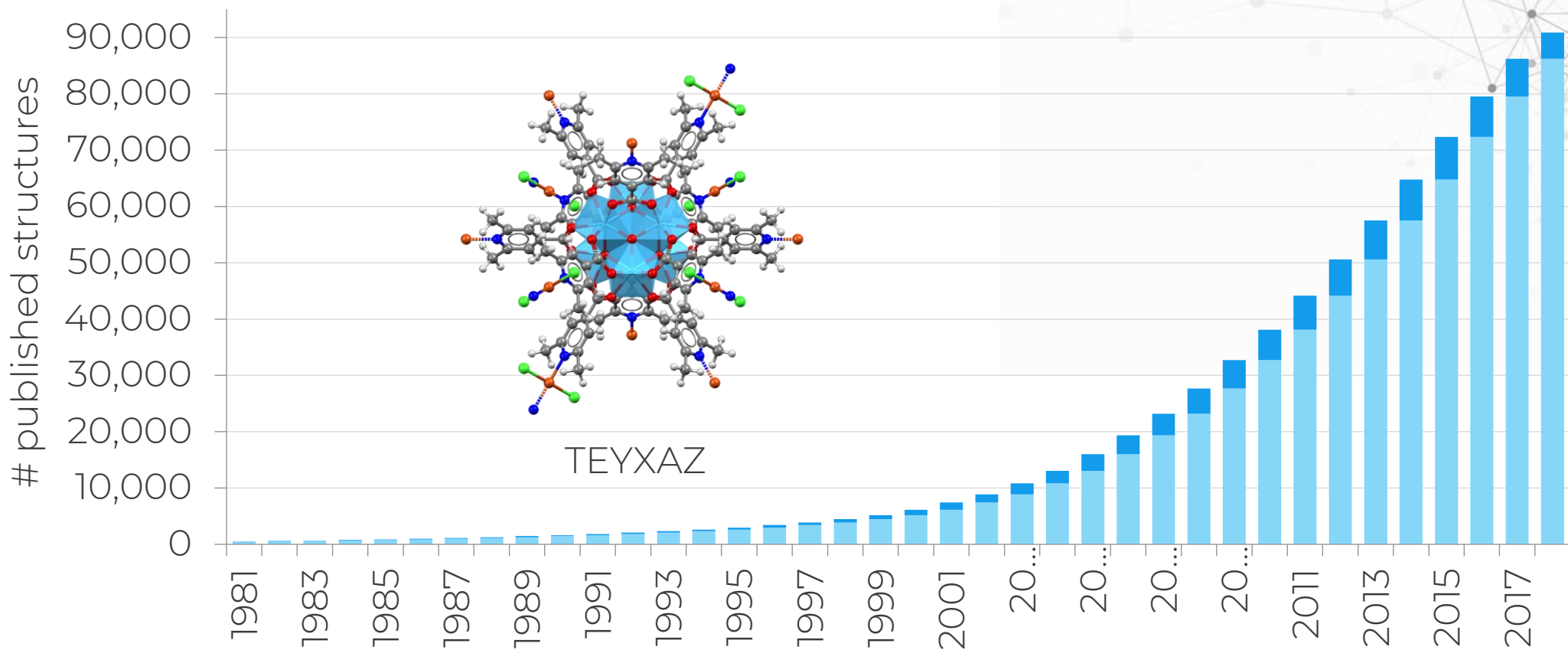
Increasing complexity

Increasing:

- Formula weights
- Unit cells
- Number of elements



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

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

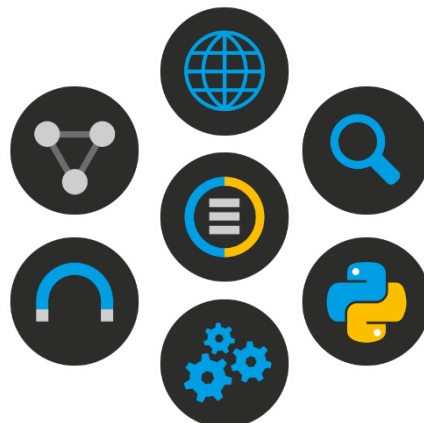
CSD-Enterprise
All CCDC application software

CSD-Discovery



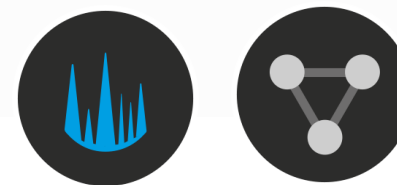
*Discover new molecules
with pharmaceutical
applications*

CSD-System



*Search, visualise, analyse
and communicate
structural data*

CSD-Materials



*Understand and predict
solid form stability and
properties*

CSD-System

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



Mercury: Visualisation



WebCSD & **ConQuest:** Advanced search



Mogul: Molecular conformations



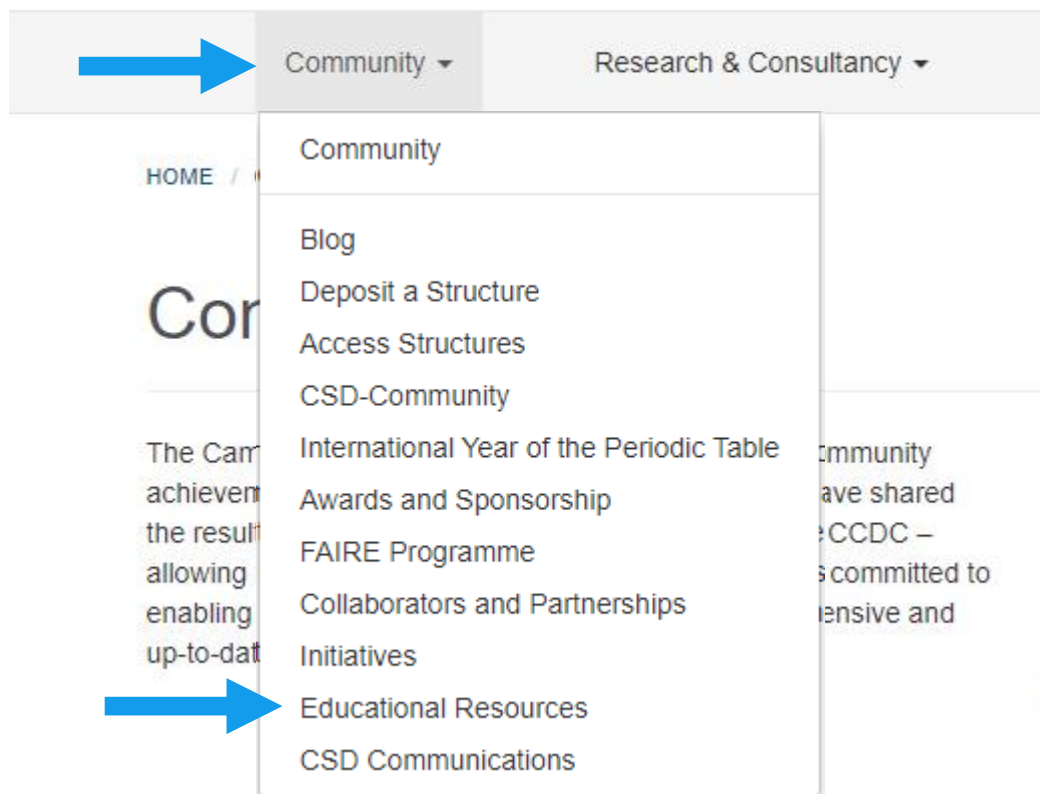
IsoStar: Intermolecular interactions



CSD Python API: Structural property insights

Educational Resources

CCDC



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

CCDC

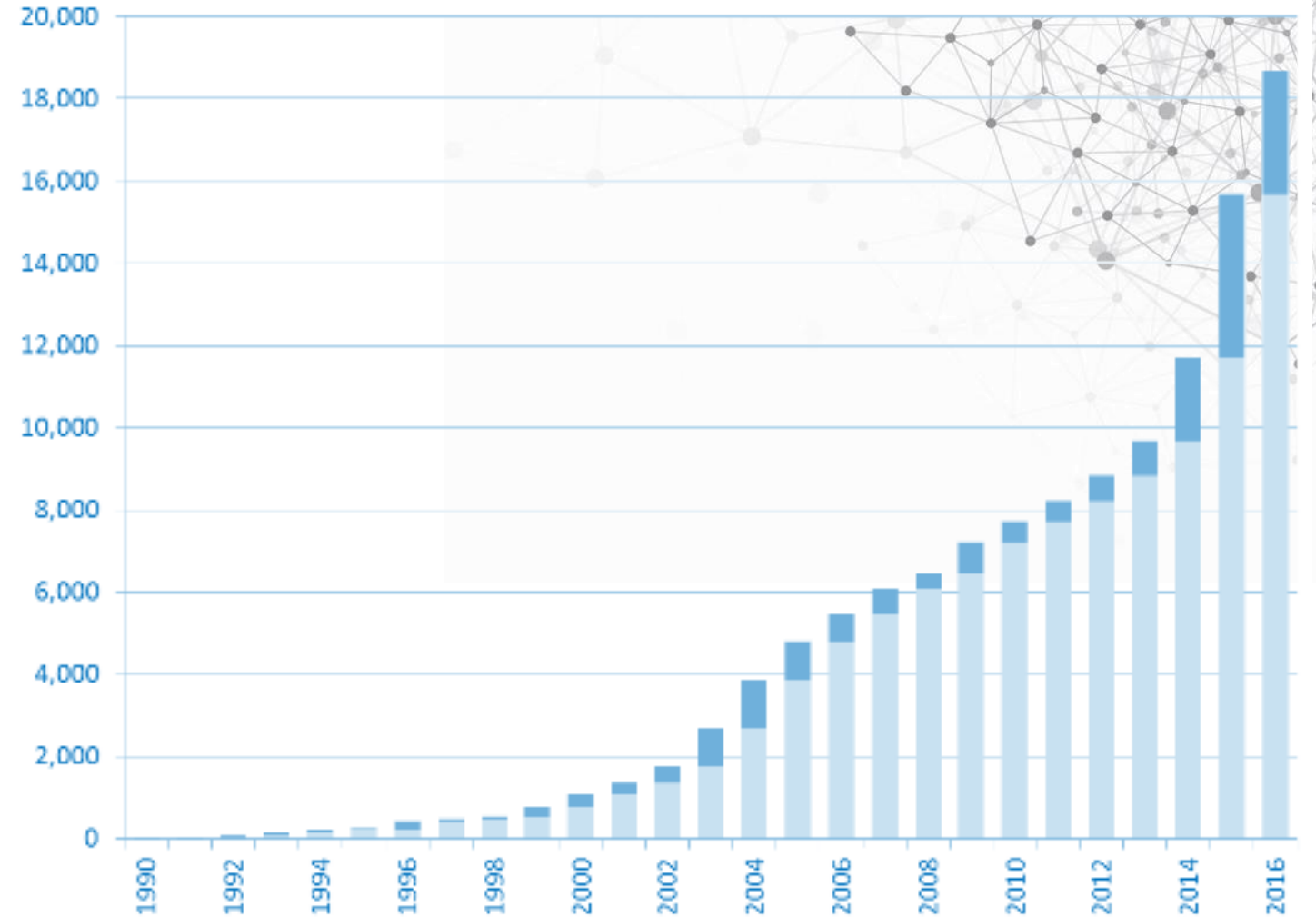
CSD Teaching Subset

- 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-ammon
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 _{∞h}	Molecules of Interest	Acetylene
ACIFEX		Optical Isomers (multi	(1S)-(-)-alpha-Pinene
ACIFIB		Optical Isomers (multi	(1S)-(-)-beta-Pinene

CSD Communications

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



Find us on social media!



facebook.com/ccdc.cambridge



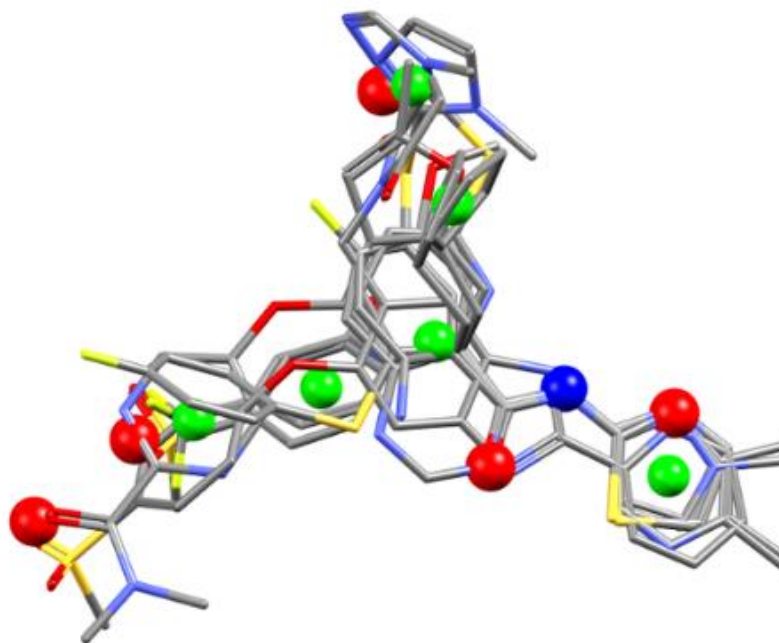
@ccdc_cambridge



CCDC Cambridge
@ccdc_cambridge

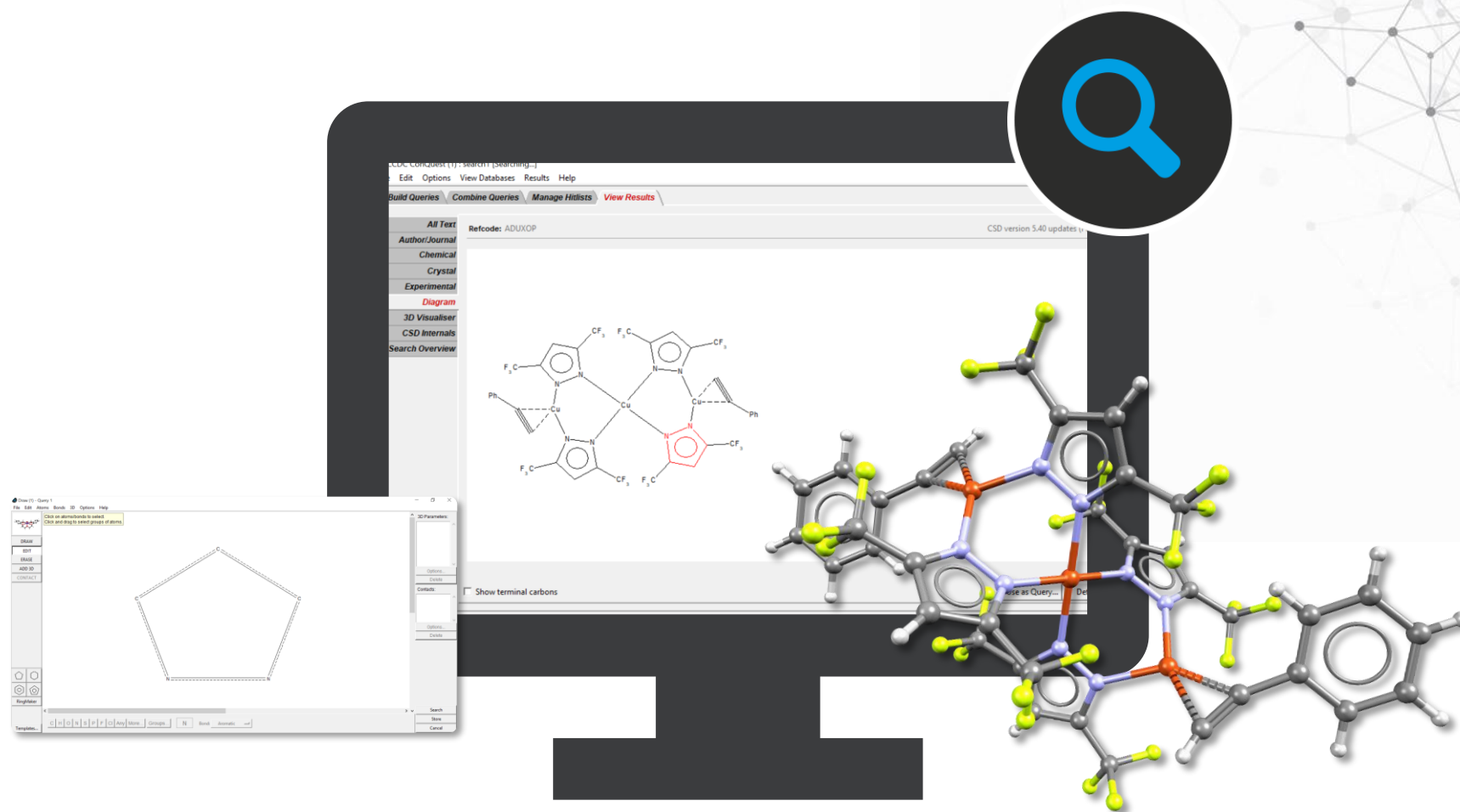
Join us on the online workshop: structure-based drug design with CSD-Discovery on Wednesday 27th May, at 3pm (BST). Learn more about Mogul, SuperStar, CSD-CrossMiner and GOLD. Register your place here hubs.ly/H0qswHC0

#DrugDiscovery #CSDDiscovery #remotelearning



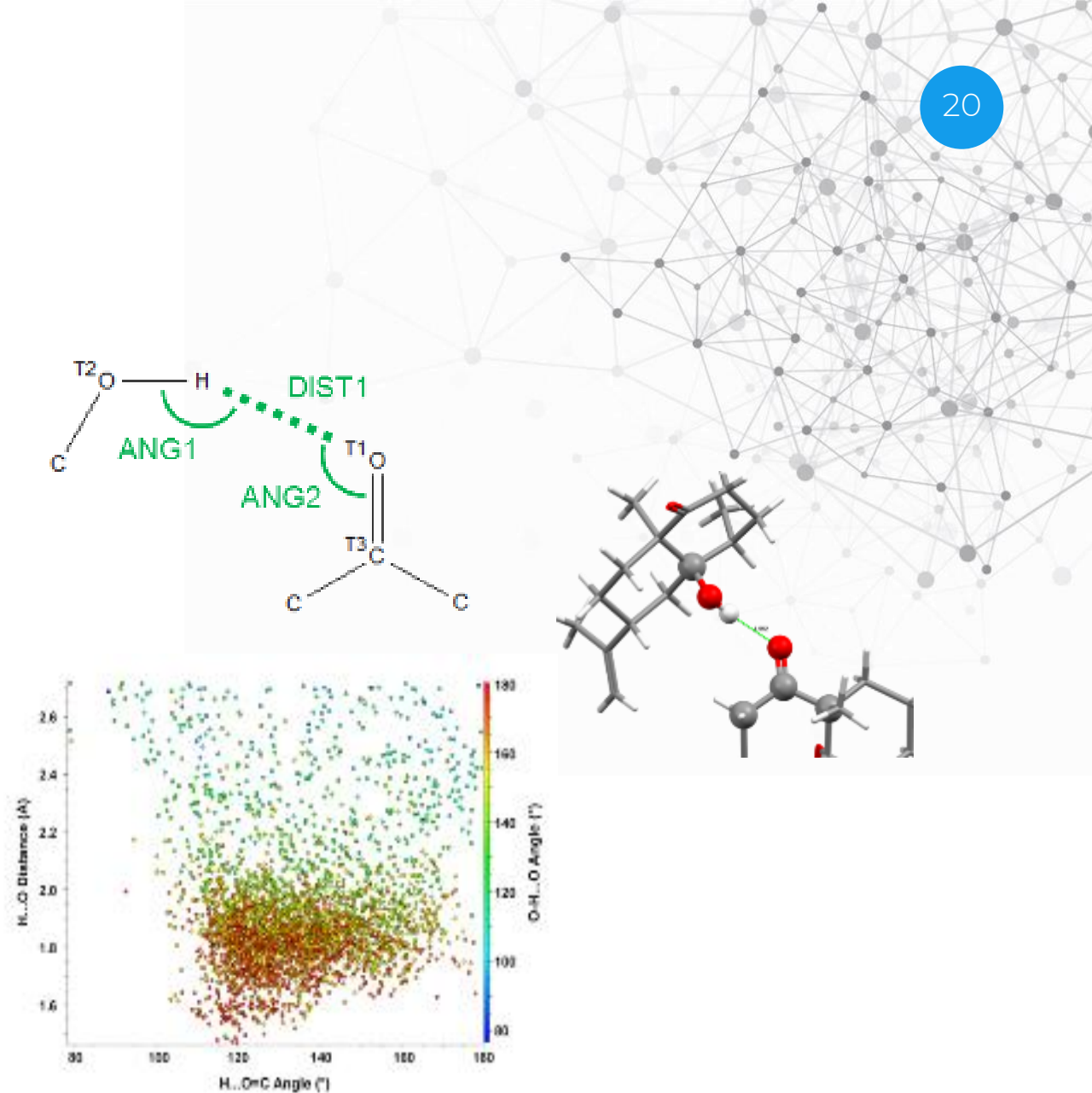
10:54 AM · May 18, 2020 · HubSpot

CSD-System: Search



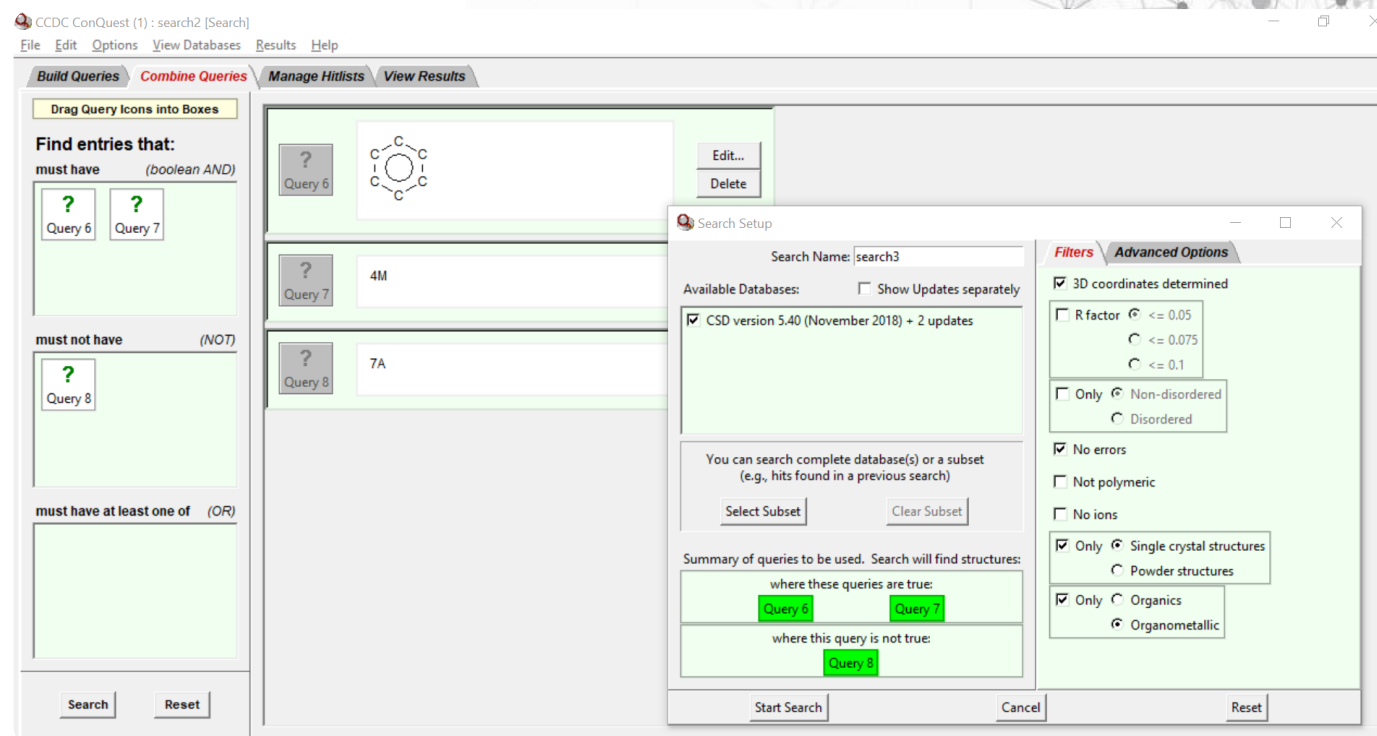
What is ConQuest?

- 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



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

Experimental (1) - New

R-factor = ☒ fractional ☐ %

☐ Exclude disordered structures

☐ Exclude structures with unresolved errors

Average e.s.d. of C-C Bonds Any

☐ Exclude powder structures

Temperature of Structure Determination = ☒ K ☐ °C

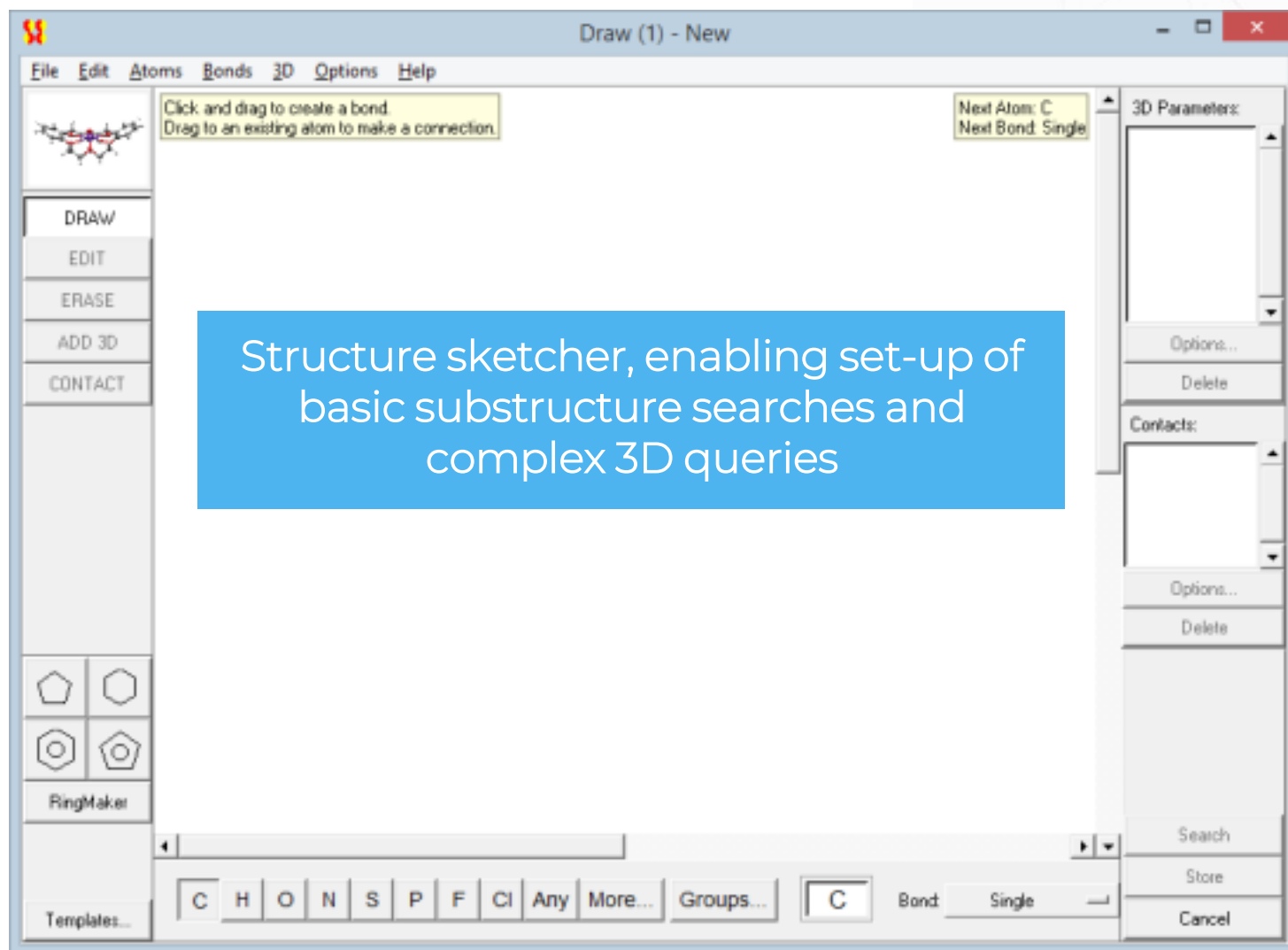
0 Room Temperature 610K

All values in the range 283-303 K are stored as Room Temperature

Radiation Source Any

Search Store Cancel Reset

ConQuest features



ConQuest features

Bibliographic
search

Author/Journal (1) - New

Authors' Names New Box

☐ Exact surname

(Required format: F.H.Allen, O'Hara, Murray-Rust etc.
Brown will hit Browning unless 'Exact surname' is selected)

Journal Name

Type part of Journal name above to narrow list displayed
Select required journal in list below

- A.C.A.(Spring) [1974-1975]
- A.C.S.Mtg.172.Inorg. [1976]
- AAPS PharmSciTech [2004-2005]
- ACA Abstr.Papers(Winter) [1967-1986]
- ACA,Ser.2 [1977-1984]
- ACGC Chem.Res.Comm. [2005]
- ACGC Chem.Res.Comm. [2001-2008]
- ACH-Models Chem. [1994-2000]
- ACS Applied Materials and Interfaces [2009-2012]
- ACS Catalysis [2011-2012]

Volume (14, 1.2 etc.) **Page (212,6-A etc.)** **Year (1998, 2001 etc.)** during

CCDC Number (Enter numeric part only, e.g. 123456 or 123/456)

ConQuest features

Generic text search

The screenshot shows a software window titled "All Text (1) - New" with a standard Windows-style title bar (minimize, maximize, close buttons). Inside the window, there are two tabs: "Text Search" (which is active) and "Required Fields".

Under the "Text Search" tab, there is a "New Box" button at the top right. Below it, the interface is split into two sections:

- On the left, under the heading "Either select from list", there is a scrollable list box containing the following terms: absolute configuration, acicular, activity, agent, air-sensitive, bar, black, blade, block, blue, brown, colorless, and column. To the right of this list is a vertical scrollbar.
- On the right, under the heading "or enter in box(es) below", there is a single-line text input field.

At the bottom of the dialog, there are four buttons: "Search", "Store", "Cancel", and "Reset".

Below the list box, there is instructional text: "You can type partial or complete word(s).
If two or more words are typed into the same box
the search will be for the exact phrase specified."

ConQuest features

Elemental
make-up

The screenshot shows a dialog box titled "Elements (1) - New" with a standard Windows-style title bar (minimize, maximize, close buttons). The dialog is divided into several sections. The top section is titled "Elements Required to be Present" and contains a large empty text box. Below this, it says "Type in elements, e.g. C H Se" followed by "or" and a button labeled "Select from Table". The next section is titled "Elements must be in" and contains two radio button options: "same molecule" (which is unselected) and "same crystal structure" (which is selected). Below these is a checked checkbox labeled "Other elements allowed in molecule/structure". The bottom section is titled "Heaviest Permitted Element in Formula Unit" and contains a small empty text box, the text "-- Not Set --", and a button labeled "Select from Table". At the very bottom of the dialog are four buttons: "Search", "Store", "Cancel", and "Reset".

ConQuest features

Analysis of results

The image displays three overlapping screenshots of the CCDC ConQuest (1) : search1 [Search] interface, illustrating various features and search results.

Top Left Screenshot: Shows the main interface with the '3D Visualiser' tab selected. A chemical structure is displayed in the central area. The left sidebar lists search categories: All Text, Author/Journal, Chemical, Crystal, Experimental, Diagram, 3D Visualiser, CSD intervals, and Search Overview. The right sidebar shows a list of search results, including CSDASOB, CSDJUV, PAKBEJ, PAKGBI, PEPJEB, PEPJIF, PEPJOL, PEPJIL, POPMOX, GEGTUT, GEMQI, GEMVH, GEMVON, GEMVUT, GEMVAA, GILZEG, GOKRUE, RECZAA, RECZBI, REWLEL, and REWDAO.

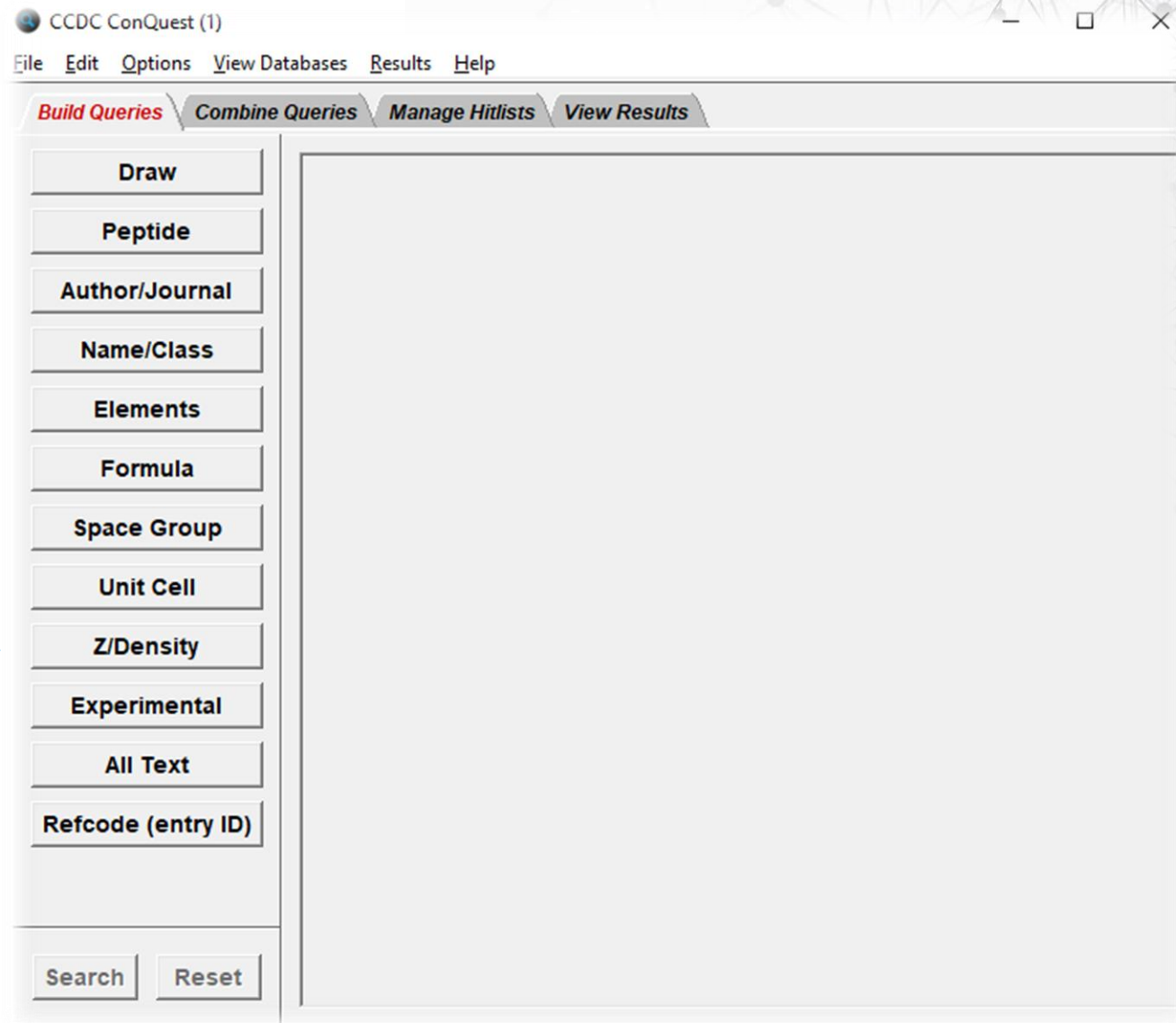
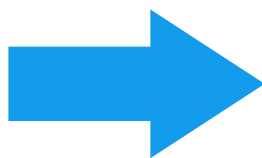
Top Right Screenshot: Shows the 'All Text' tab selected. It displays search results for the query 'PEYJIL'. The results include the author(s) T. Okawara, E. P. (Sapirata, H. J. (Mitsunaka), the reference Tetrahedron (2012), 69, 474, the publication DOI 10.1016/j.tet.2012.11.027, and the deposition CCDC 905439. The right sidebar shows a list of search results, including CSDASOB, CSDJUV, PAKBEJ, PAKGBI, PEPJEB, PEPJIF, PEPJOL, PEPJIL, POPMOX, GEGTUT, GEMQI, GEMVH, GEMVON, GEMVUT, GEMVAA, GILZEG, GOKRUE, RECZAA, RECZBI, REWLEL, and REWDAO.

Bottom Screenshot: Shows the '3D Visualiser' tab selected. It displays a chemical structure of a substituted benzene ring. The right sidebar shows a list of search results, including CSDASOB, CSDJUV, PAKBEJ, PAKGBI, PEPJEB, PEPJIF, PEPJOL, PEPJIL, POPMOX, GEGTUT, GEMQI, GEMVH, GEMVON, GEMVUT, GEMVAA, GILZEG, GOKRUE, RECZAA, RECZBI, REWLEL, and REWDAO.

CCDC

ConQuest – Basics

1. Open ConQuest



ConQuest – Basics

2. Build Queries: Author Search

Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)



Author/Journal (1) - New

Authors' Names **New Box**

☐ Exact surname

(Required format: F.H.Allen, O'Hara, Murray-Rust etc.
Brown will hit Browning unless 'Exact surname' is selected)

Journal Name

Type part of Journal name above to narrow list displayed
Select required journal in list below

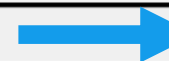
3 Biotech [2015]
A.C.A.(Spring) [1974-1975]
A.C.S.Mtg.172,Inorg. [1976]
AAPS PharmSciTech [2004-2005]
ACA Abstr.Papers(Winter) [1967-1986]
ACA,Ser.2 [1977-1984]
ACGC Chem.Res.Comm. [2005]
ACGC Chem.Res.Comm. [2001-2009]
ACH-Models Chem. [1994-2000]
ACS Applied Materials and Interfaces [2009-2015]

Volume (14, 1.2 etc.) Page (212,6-A etc.) Year (1998, 2001 etc.)

during

CCDC Number (Enter numeric part only, e.g. 123456 or 123/456)

Search Store Cancel Reset



ConQuest – Basics

2. Build Query

Author Search

The screenshot displays the ConQuest software interface. The main window is titled "Search Setup" and contains the following elements:

- Search Name:** search2
- Available Databases:** ☒ CSD version 5.37 (November 2015) + 3 updates
- Show Updates separately:** ☐
- You can search complete database(s) or a subset (e.g., hits found in a previous search):**
 -
 -
- Single query being used. Search will find structures:**
 - where this query is true: **Query 2**
- Buttons:** (highlighted with a blue arrow), ,

The "Filters/Advanced Options" window is also visible, showing the following options:

- ☐ 3D coordinates determined
- ☐ R factor ☒ ≤ 0.05 , ☐ ≤ 0.075 , ☐ ≤ 0.1
- ☐ Not disordered
- ☐ No errors
- ☐ Not polymeric
- ☐ No ions
- ☐ No powder structures
- ☐ Only ☒ Organics, ☐ Organometallic

In the background, the "Author/Journal (1) - New" window is partially visible, showing the "Authors' Names" section with a "New Box" button and a list of authors/journals.

ConQuest – Basics

3. View results

Edit Options View Databases Results Help

Build Queries Combine Queries Manage Hitlists **View Results**

Refcode: TUYBEW CSD version 5.37 updates (Nov 2015)

Chemical Structure: O=C(O)C(O)c1ccccc1Br

Analyse Hitlist:

- ✓ BUPCOG
- ✓ FIZPEL03
- ✓ FUWJAJ01
- ✓ PUSYEJ
- ✓ RULYAA
- ✓ RUPDEN
- ✓ RUPDIR
- ✓ RUPDOX
- ✓ RUPDUD
- ✓ RUSNOK
- ✓ TUXZIX
- ✓ TUXZOD
- ✓ TUXZUJ
- ✓ TUYBAS
- ✓ TUYBEW**
- ✓ TUYBIA
- ✓ TUYBIA01
- ✓ ZURTUD
- ✓ ZURVAL
- ✓ ZURVEP
- ✓ ZURVIT
- ✓ ZURVIT01

1821 hits

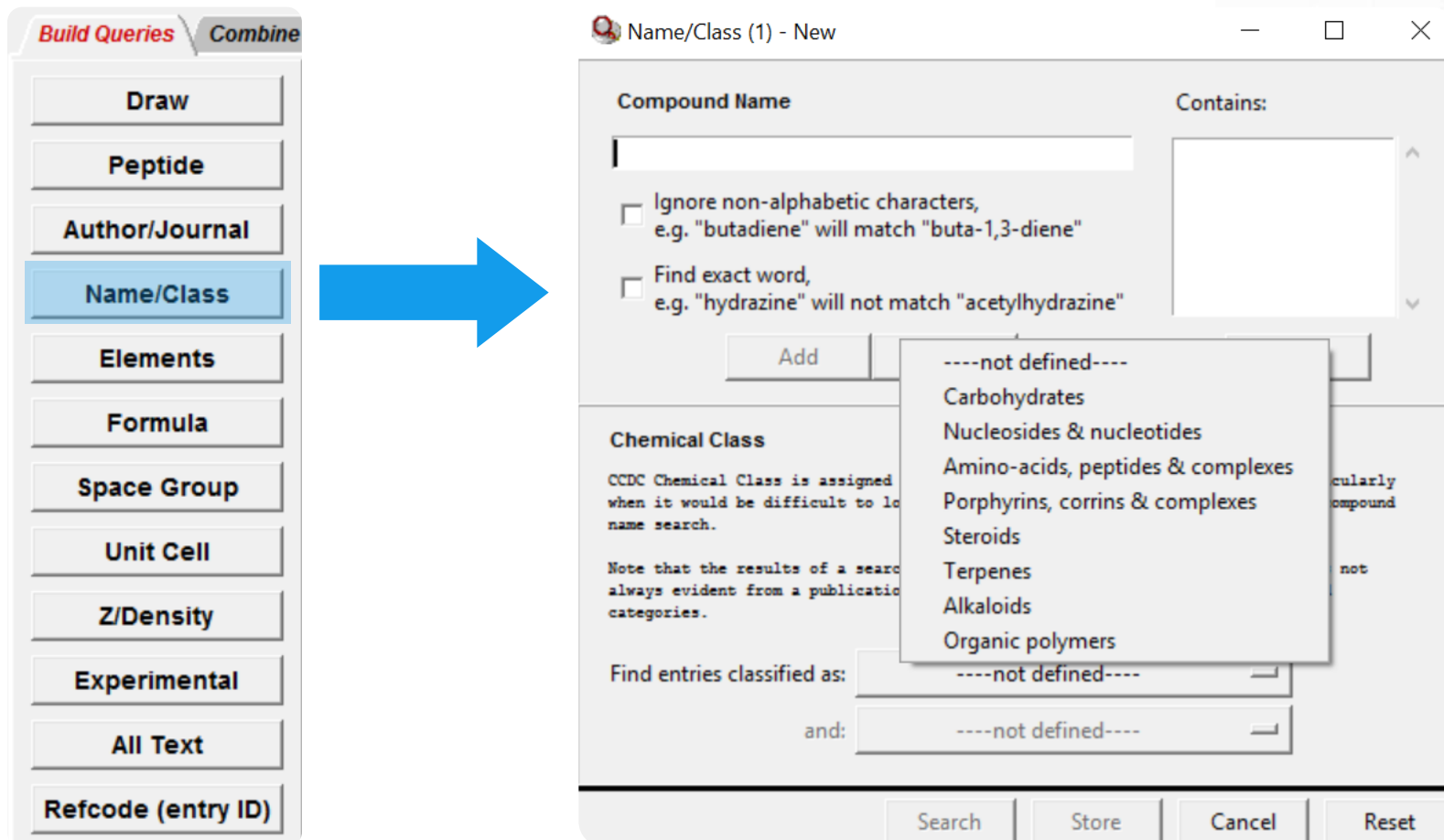
100%

Stop Search

☐ Show terminal carbons

Use as Query... Detach

ConQuest – Compound name search



Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)

Name/Class (1) - New

Compound Name

Contains:

☐ Ignore non-alphabetic characters,
e.g. "butadiene" will match "buta-1,3-diene"

☐ Find exact word,
e.g. "hydrazine" will not match "acetylhydrazine"

Add

Chemical Class

CCDC Chemical Class is assigned when it would be difficult to do a name search.

Note that the results of a search are always evident from a publication categories.

Find entries classified as:

and:

----not defined----

Carbohydrates

Nucleosides & nucleotides

Amino-acids, peptides & complexes

Porphyrins, corrins & complexes

Steroids

Terpenes

Alkaloids

Organic polymers

Search Store Cancel Reset

ConQuest – Elements & formula search

Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)

Elements (1) - New

Elements Required to be Present

Type in elements, e.g. C H Se
or **Select from Table**

Elements must be in

☐ same molecule

☒ same crystal structure

☒ Other elements allowed in molecule/structure

Heaviest Permitted Element in Formula Unit

-- Not Set -- **Select from Table**

Search **Store** **Cancel** **Reset**

Formula (1) - New

Formula

Type in formula, e.g. C6 H12 O6
You may specify an inexact formula, e.g. Ca1-3 O>4
finds entries with 1,2 or 3 Calciums and more than 4 Oxygens.

or **Select from Table**

Formula applies to

☒ an individual molecule

☐ all molecules in structure added together

☒ Other atoms allowed in molecule/structure

Search **Store** **Cancel** **Reset**

ConQuest – Z/Density and experimental info search

Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)

Z/Density (1) - New

No. of "Molecules" per Unit Cell (Z)	=	
No. of "Molecules" per Asymmetric Unit (Z')	=	
No. of Atoms with 3D Coordinates	=	
Calculated Density (g/cm ³)	=	
No. of Chemical Units (molecules, ions, etc.) in Entry	=	

Search Store Cancel Reset

Experimental (1) - New

R-factor = ☒ fractional ☐ %

☐ Exclude disordered structures

☐ Exclude structures with unresolved errors

Average e.s.d. of C-C Bonds Any

☐ Exclude powder structures

Temperature of Structure Determination = ☒ K ☐ °C

0 Room Temperature 610K

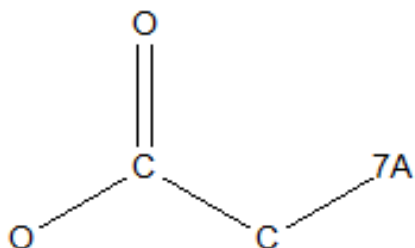
All values in the range 283-303 K are stored as Room Temperature

Radiation Source Any

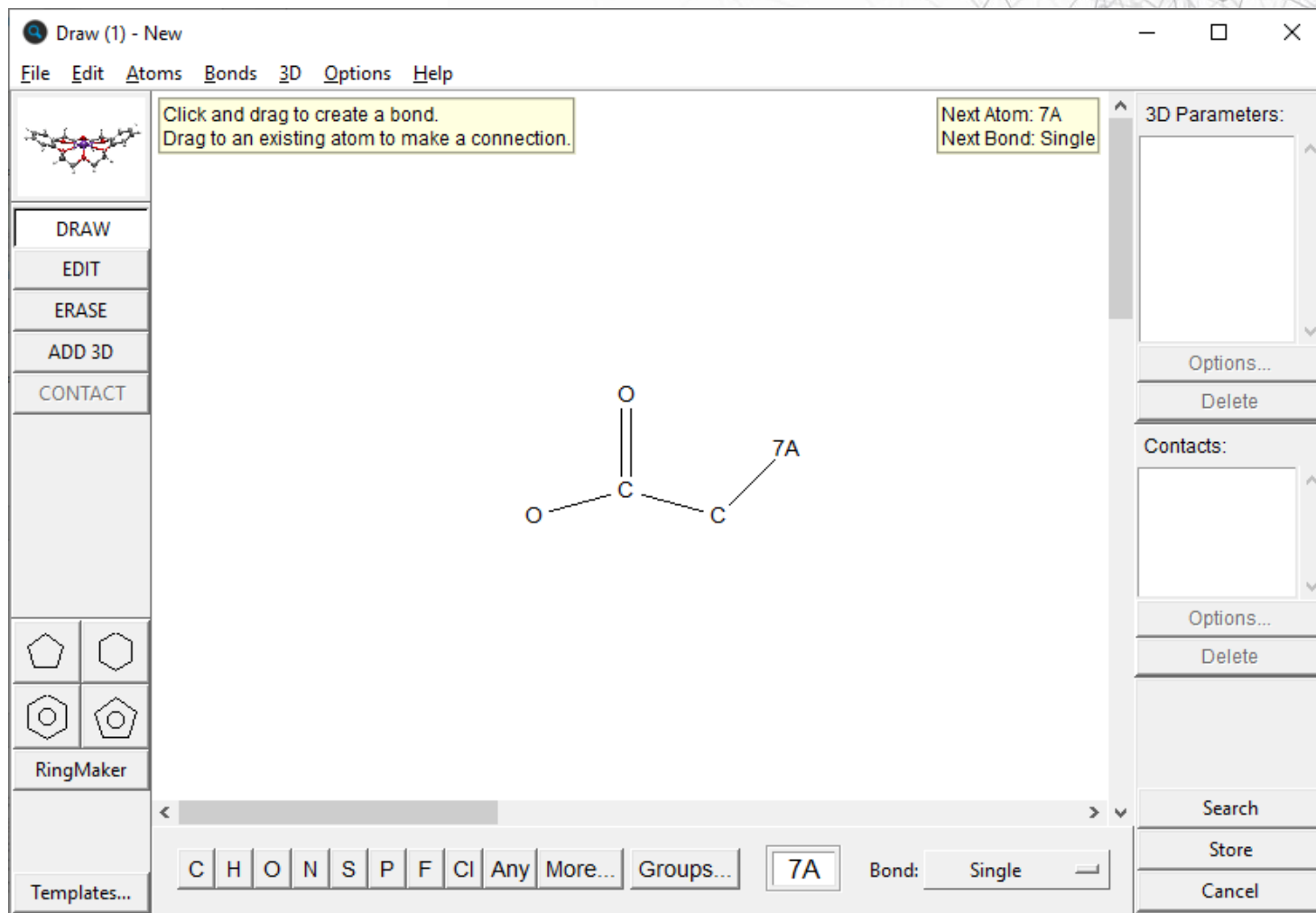
Search Store Cancel Reset

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

Navigation bar: Community ▾ Research & Consultancy ▾ Solutions ▾ News & Events ▾ Support

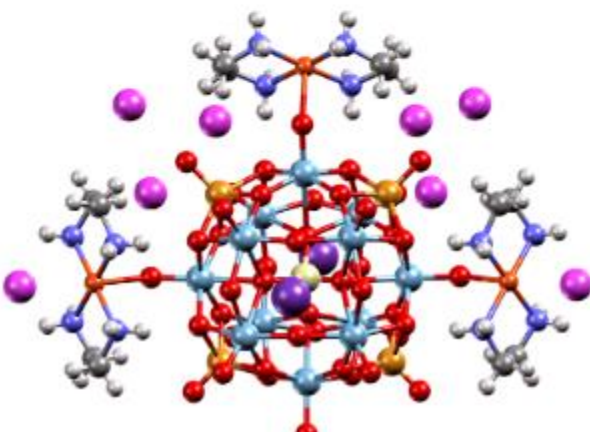
HOME / Community

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

The Cambridge Crystallographic Centre has achieved the result of allowing up-to-date

Community have shared the CCDC – is committed to intensive and



Awar
Depo
CSD-
Colla

CCDC

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

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.

The drawing interface includes a horizontal toolbar at the top with icons for drawing lines, polygons, rings, and other structures. On the left is a vertical palette with buttons for elements (C, N, O, S, H, F, Cl, Br), functional groups (alcohol, amine, etc.), and other symbols like brackets, a plus sign, and a minus sign.

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

Simple Search

Structure Search

Unit Cell Search

Formula Search

Search Complete - 219 Results Found

100%

Modify Search

New Search

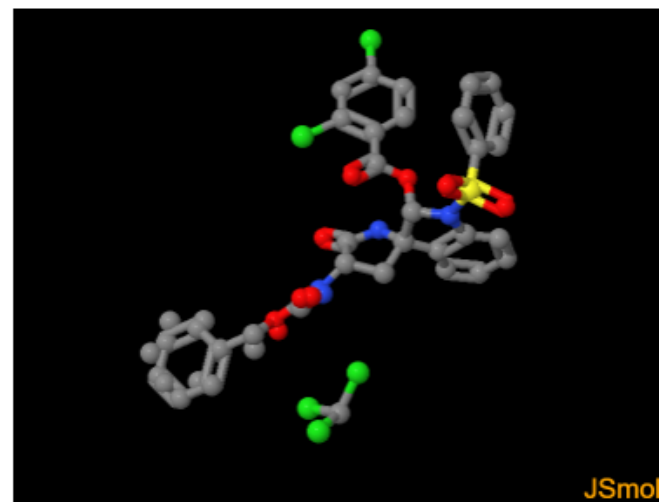
Results

<input type="checkbox"/>	Database Identifier	Deposition Number
<input type="checkbox"/>	ABECEP	1100088
<input checked="" type="checkbox"/>	ACEYAJ	255902
<input type="checkbox"/>	ADEPAB	613666
<input type="checkbox"/>	ADEPEF	613667
<input type="checkbox"/>	AGAGEW	933112
<input checked="" type="checkbox"/>	AGAGIA	962016
<input type="checkbox"/>	BAFWOU	1105049
<input type="checkbox"/>	BAMBOJ	1528977
<input type="checkbox"/>	BEKLOV	1569815
<input type="checkbox"/>	BEXHOE	1822669
<input type="checkbox"/>	BOBZEA	1829117
<input type="checkbox"/>	BOMAIU10	1113591
<input type="checkbox"/>	BOXRIR	1029163
<input type="checkbox"/>	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

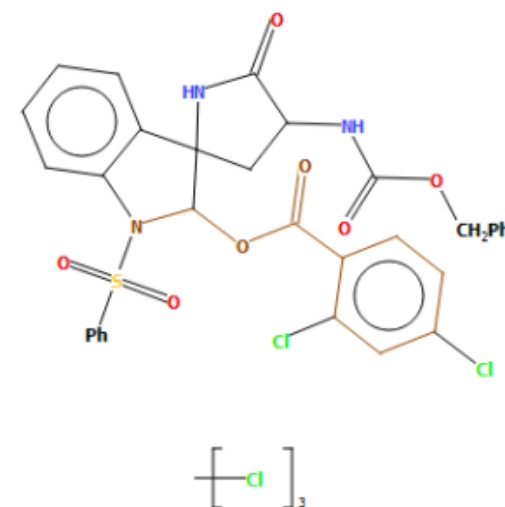
Space Group: P 2₁ 2₁ 2₁ (19), **Cell:** a 9.6220(13)Å b 18.082(2)Å c 19.975(3)Å, α 90° β 90° γ 90°

3D viewer



Style: Ball and Stick ▾ Labels: No Labels ▾ Packing: None ▾ Measure: None ▾

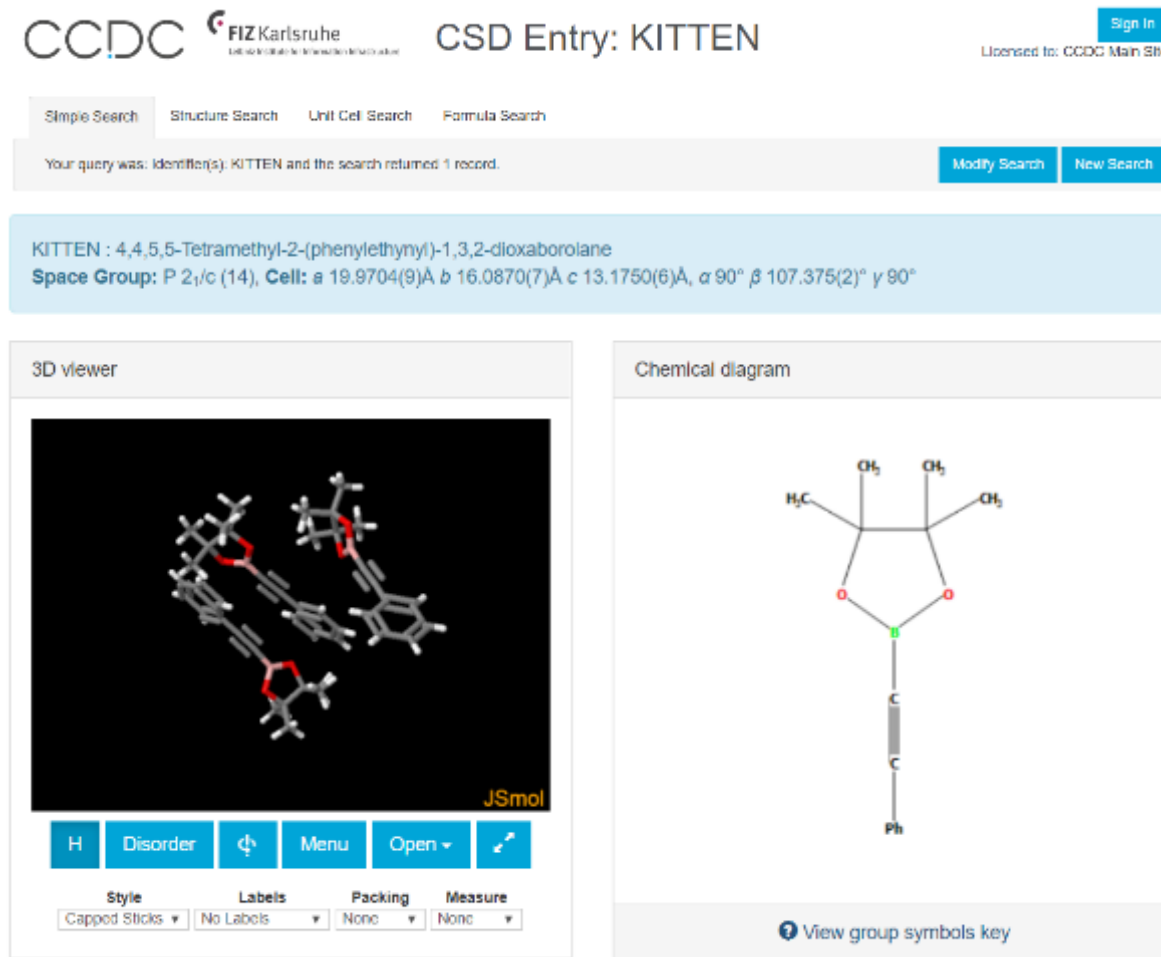
Chemical diagram



View group symbols key

WebCSD

- Search from your browser
- Accessible on mobile devices!



Next time:

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

For now: let's practice searching...

A faint, light gray network pattern of interconnected lines and nodes is visible in the background, resembling a molecular or structural diagram.

ccDC

advancing structural science