

Searching the CSD using ConQuest

CCDC Virtual Workshop 2020 – Session 1

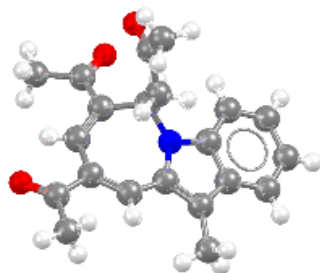
Ilaria Gimondi, Suzanna Ward, Yinka Olatunji-Ojo, Natalie Johnson, Jeff Lengyel

November 2020

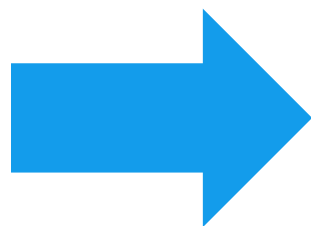
Learning outcomes for today

- How to search over a million published crystal structures using our desktop software
- How to refine your results to structures in targeted subsets allowing you to explore drugs, pesticides and MOFs in more detail

CSD Refcodes



CSD Refcode -
XOPCAJ



What is XOPCAJ?

- 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

The CSD software

CSDEnterprise.

CSDMaterials.



DASH



CSD
Python
API



Mercury

CSDDiscovery.



SuperStar



GOLD



CSD-CrossMiner



Hermes



CSD
Python API



Mercury



CSDSystem.



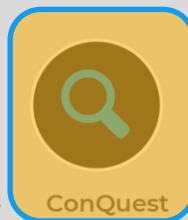
WebCSD



Mogul



My Structures



ConQuest



CSD
Python API



Mercury



IsoStar



CSD

CSDCommunity.



Mercury



enCIFer



CSDSymmetry



Deposit
Structures



CellCheckCSD



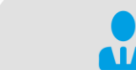
Educational
Collection



Access
Structures



My Structures



Professional
Services



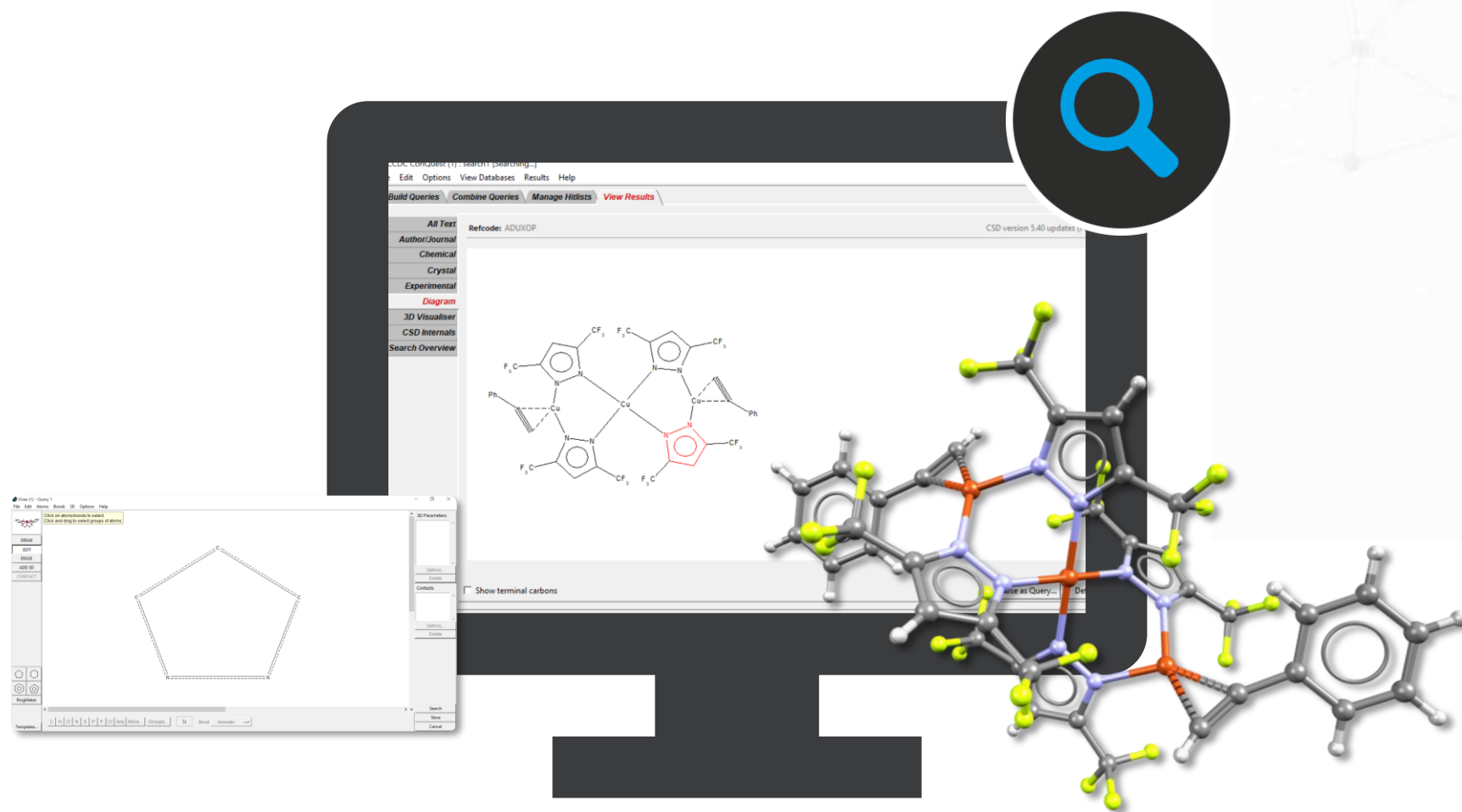
Research
& Knowledge
partnerships

CCDC

CSD-System – 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 the most 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

CSD-System: Search



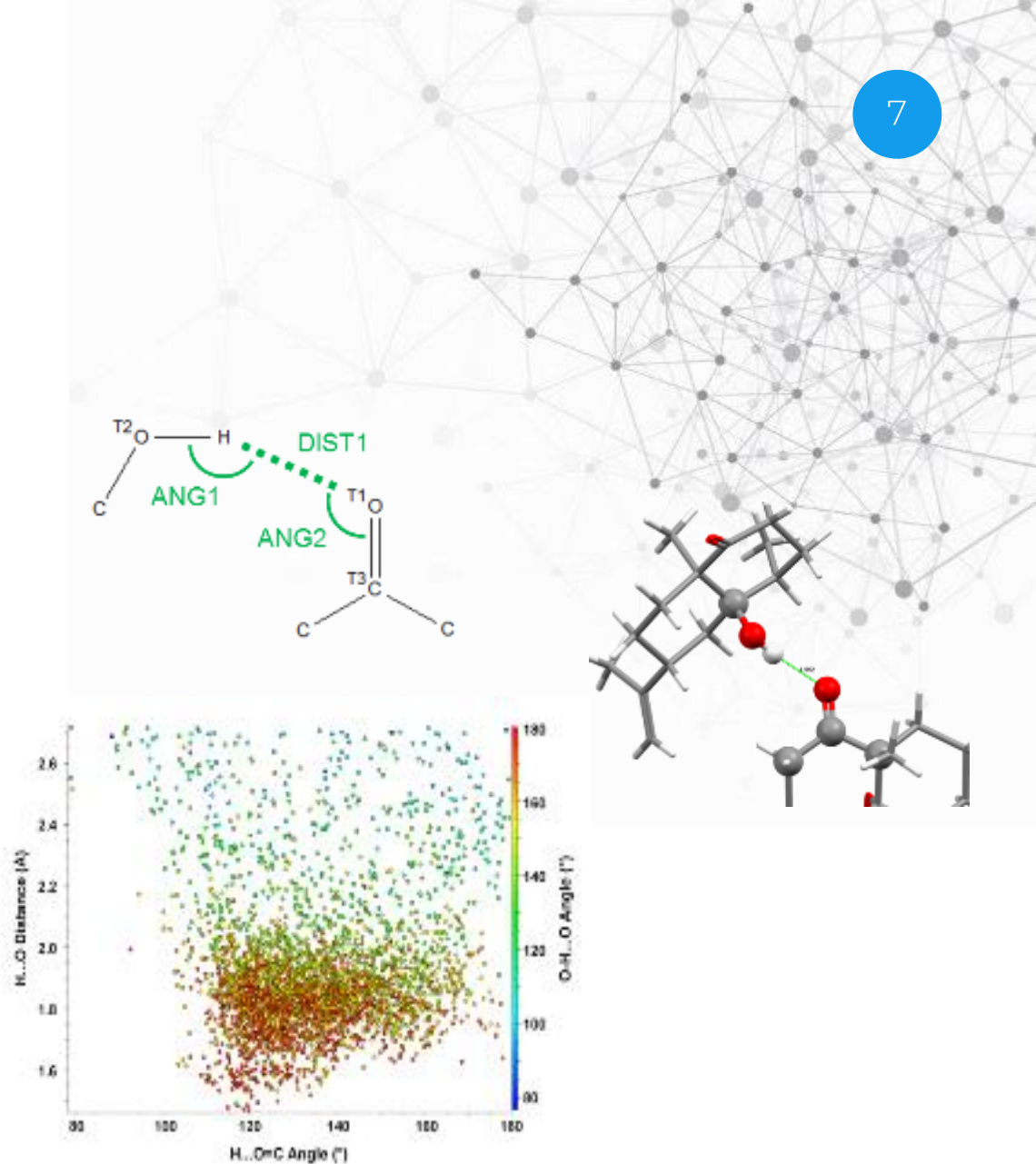
On your desktop: ConQuest



On a browser: WebCSD

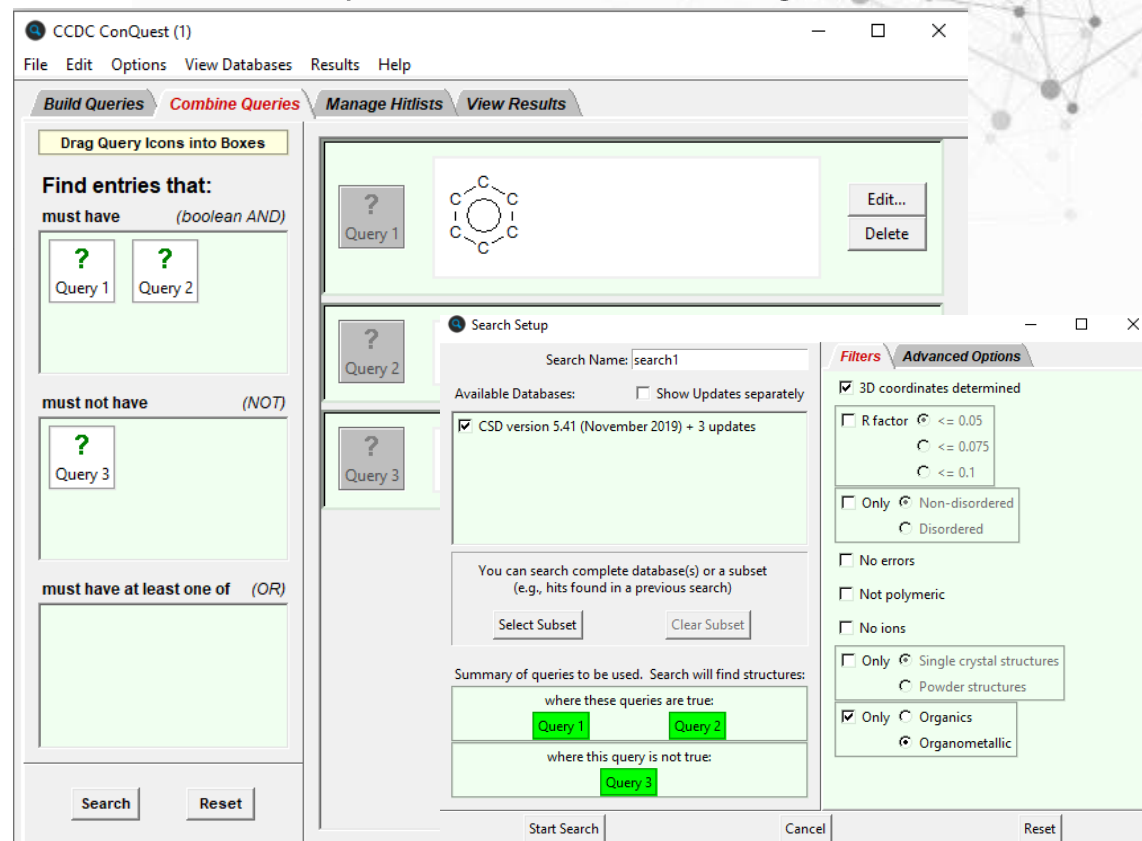
What is ConQuest?

- 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](#)



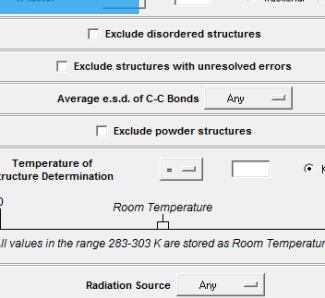
ConQuest

- 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 visualisation and advanced numerical analysis & plotting



ConQuest features

Search for information relating to the structure determination



ation

R-factor Any 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 Any K °C

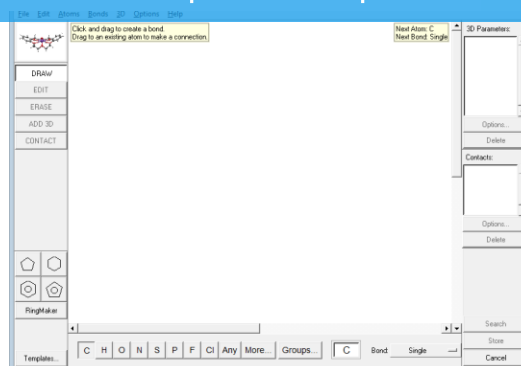
0 Room Temperature 610K

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

Radiation Source Any

Search Store Cancel Reset

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



Bibliographic search

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]
ADCG Chem.Res.Comm. [2005]
ADCG Chem.Res.Comm. [2001-2008]
ADH-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.)

CCDC Number

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

Search Store Cancel Reset

Generic text search

Text Search **Required Fields**

New Box

Either select from list or enter in box(es) below

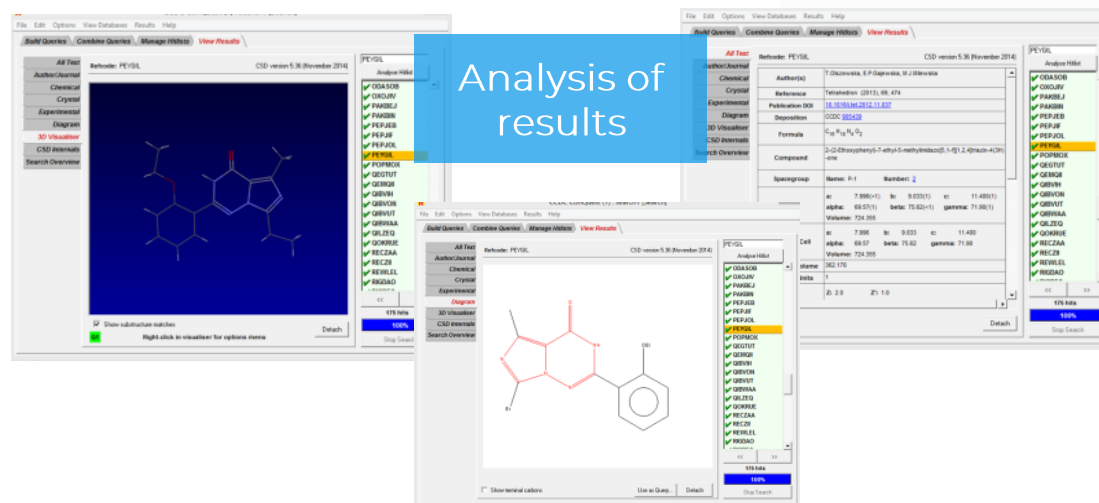
absolute configuration
acouster
activity
agent
ar-sensitive
bai
black
blade
block
blue
brown
colorless
column

Gen
text s

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.

Search Store Cancel Reset

Analysis of results



Elemental make-up

Elements Required to be Present

Type in elements, e.g. C H S

or

Elements must be in

☐ same molecule

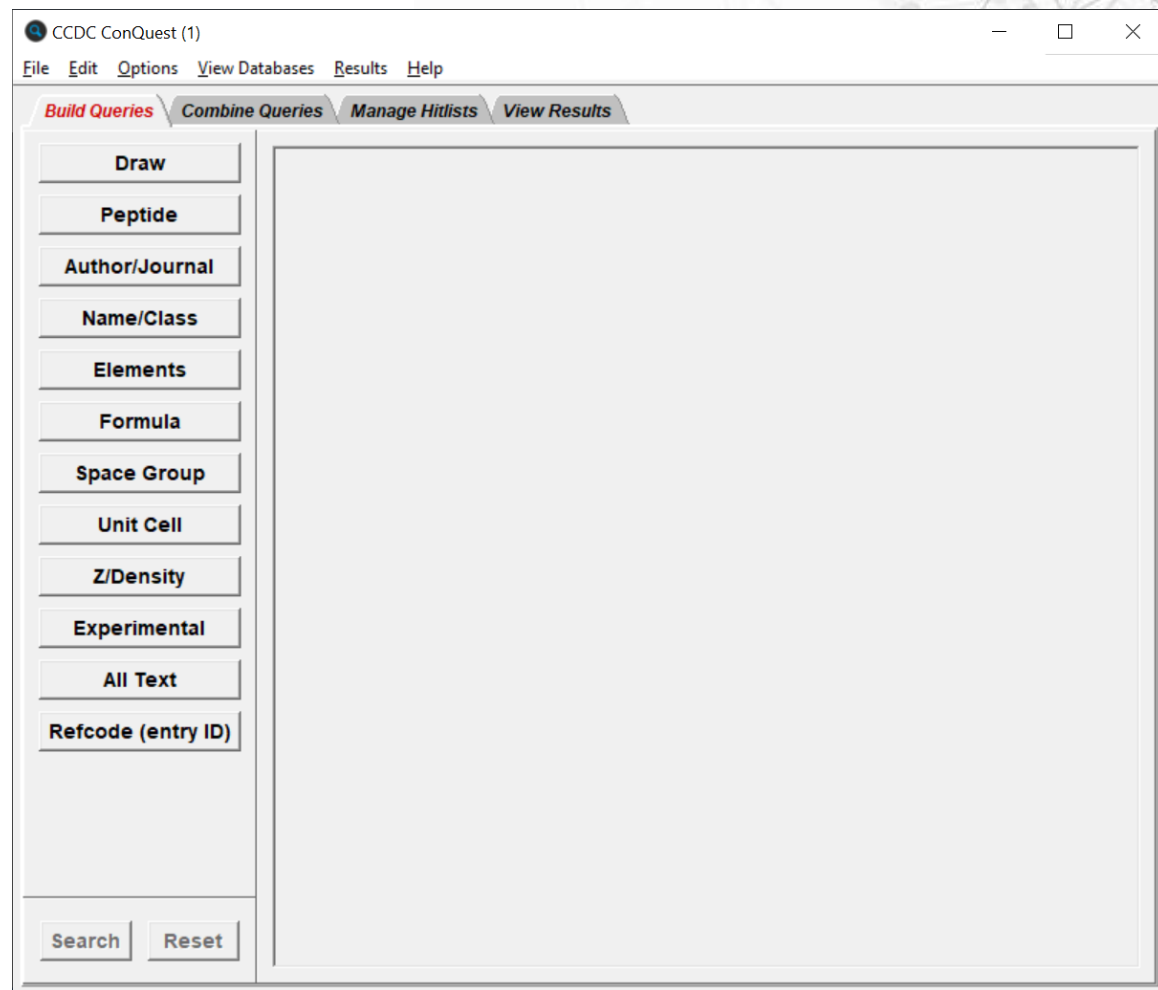
☒ same crystal structure

☒ Other elements allowed in molecule/structure

How many Permitted Element in Formula Unit

Show One: ConQuest – Opening and search options

10



ConQuest – Searching and filtering

Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)

Search Reset

Search Setup

Search Name: search1

Available Databases: ☐ Show Updates separately

☒ CSD version 5.41 (November 2019) + 3 updates

You can search complete database(s) or a subset (e.g., hits found in a previous search)

Select Subset Clear Subset

Single query being used. Search will find structures:
where this query is true:
Query 1

Filters **Advanced Options**

☐ 3D coordinates determined

☐ R factor ☒ ≤ 0.05
☐ ≤ 0.075
☐ ≤ 0.1

☐ Only ☒ Non-disordered
☐ Disordered

☐ No errors

☐ Not polymeric

☐ No ions

☐ Only ☒ Single crystal structures
☐ Powder structures

☐ Only ☒ Organics
☐ Organometallic

Start Search Cancel Reset

ConQuest – Author and journal 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) - Query 5

Authors' Names **New Box**

M.Zema ☐ 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

2D Mat. [2017]
Biotech [2015]
A.C.A.(Spring) [1974-1975]
A.C.S.Mtg.172,Inorg. [1976]
AAPS PharmSciTech [2004-2013]
ACA Abstr.Papers(Winter) [1967-1986]
ACA,Ser.2 [1977-1984]
ACGC Chem.Res.Comm. [2001-2009]
ACH-Models Chem. [1994-2000]
ACS Appl. Bio Mater. [2019]

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**

Search Setup

Search Name: search2

Available Databases: ☐ Show Updates separately

☒ CSD version 5.41 (November 2019) + 3 updates

You can search complete database(s) or a subset
(e.g., hits found in a previous search)

Select Subset **Clear Subset**

Single query being used. Search will find structures:
where this query is true:
Query 5

Start Search **Cancel** **Reset**

Filters **Advanced Options**

☐ 3D coordinates determined

☐ R factor ☒ <= 0.05
☐ <= 0.075
☐ <= 0.1

☐ Only ☒ Non-disordered
☐ Disordered

☐ No errors

☐ Not polymeric

☐ No ions

☐ Only ☒ Single crystal structures
☐ Powder structures

☐ Only ☒ Organics
☐ Organometallic

Query highlighted

Left clicking each tab will display different information

CCDC ConQuest (1) : search2 [Search]

File Edit Options View Databases Results Help

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

Refcode: AVETAW CSD version 5.41 (November 2019)

Author(s)	M.Boiocchi, M.Bonizzoni, L.Fabrizzi, F.Foti, M.Licchelli, A.Taglietti, M.Zema
Reference	Dalton Trans. (2004), , 653
Publication DOI	10.1039/b312980b
Deposition	CCDC 221987
Formula	$C_{10}H_{24}N_4Ni^{2+} \cdot 2(ClO_4^-)$
Compound	(3-(4-(3-Aminopropyl)piperazin-1-yl)propylamine)-nickel(ii) diperchlorate
Spacegroup	Name: Pbca Number: 61
Cell	a: 14.426(2) b: 15.390(2) c: 16.171(2) alpha: 90.00 beta: 90.00 gamma: 90.00 Volume: 3590.306
Reduced Cell	a: 14.426 b: 15.390 c: 16.171 alpha: 90.00 beta: 90.00 gamma: 90.00 Volume: 3590.306
Molecular Volume	448.788
Chemical Units	2
Z, Z'	Z: 8.0 Z': 1.0

Detach

AVETAW

Analyse Hitlist

- ✓ AVETAW
- ✓ AVETEA
- ✓ BAQFOO
- ✓ BETHEO02
- ✓ BOGKOX
- ✓ DUCWAA
- ✓ DUCWOO
- ✓ EBELON
- ✓ EBELUT
- ✓ FASQOI
- ✓ FASQUO
- ✓ GOLLUR
- ✓ HIGGAG
- ✓ HUDBEO
- ✓ IBIYOH
- ✓ IBIYUN
- ✓ IBIZAU
- ✓ IBIZEY
- ✓ IWAKEW
- ✓ IWAKIA
- ✓ KELZOQ

<< >>

49 hits

100%

Stop Search

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.

From data to publication

CCDC ConQuest (1): search2 [Search]

McKervey, A. R. Maguire, S. M. Tuladhar and M. Fiona Twohig, *J. Chem. Soc.* 1047–1054 DOI: [10.1039/P19900001047](https://doi.org/10.1039/P19900001047); (b) H. Duddeck, *J. Chem. Soc.*, 1055–1063 DOI: [10.1039/P19900001055](https://doi.org/10.1039/P19900001055); (c) P. Panne and J. M. Fox, *J. Appl. Cryst.* 1990, 23, 1055–1063. [External Links.](#)

Footnote

+ Electronic supplementary information (ESI) available: Experimental procedure for the synthesis of 15. For ESI and crystallographic data in CIF or other electronic format see DOI: [10.1039/b312980b](https://doi.org/10.1039/b312980b).

This journal is © The Royal Society of Chemistry 2009

ROYAL SOCIETY OF CHEMISTRY

Code: AVETAW CSD version 5.41 (November 2010)

Author(s)	M. Boiocchi, M. Bonizzoni, L. Fabbrizzi, F. Foti, M. Licchelli, A. Taglietti, M. Zema
Reference	DOI: 10.1039/b312980b , 653
Publication DOI	10.1039/b312980b
Chemical Formula	$C_{10}H_{24}N_4Ni^{2+} \cdot 2(ClO_4^-)$
Compound	(3-(4-(3-Aminopropyl)piperazin-1-yl)propylamine)-nickel(II) diperchlorate
Spacegroup	Name: Pbcn Number: 61
Unit Cell	a: 14.426(2) b: 15.390(2) c: 16.171(2) alpha: 90.00 beta: 90.00 gamma: 90.00 Volume: 3590.306
Reduced Cell	a: 14.426 b: 15.390 c: 16.171 alpha: 90.00 beta: 90.00 gamma: 90.00 Volume: 3590.306
Molecular Volume	448.788
Chemical Units	2

ACS Publications
Most Trusted. Most Cited. Most Read.

Polymorph α
Polymorph β

Synopsis
The pressure- and temperature-dependent transition between resorcinol polymorphs α and β has been rationalized in terms of transforming H-bonds and their networks.

Introduction
Resorcinol, an intermediate often used in chemical practice and a pharmaceutical agent, was one of the first organic compounds for which the phenomenon of polymorphism was described and the first organic compound for which the structures of both polymorphs were determined in 1938 by Robertson and Ubbelohde.^(1,2) Until today, the resorcinol crystals belong to the best known examples of polymorphs.⁽³⁾

Keywords: celecoxib; active pharmaceutical ingredient; API; solvate; crystal structure; isostructurality; disorder; PIXEL; anti-inflammatory.

CCDC references: 2011633; 2011634; 2011635; 2011636; 2011637; 2011638

Similar articles **PowerPoint slides**

References

audio. Accelrys Software Inc., San Diego, CA, USA. [Google Scholar](#)
Cryst. Growth Des. **12**, 2147–2152. [Web of Science](#) [CrossRef](#) [CAS](#) [Google Scholar](#)
gia, A. (2014). *CrystEngComm*, **16**, 24–27. [Web of Science](#) [CSD](#) [CrossRef](#) [CAS](#)
IUCrJ, **6**, 751–760. [CSD](#) [CrossRef](#) [CAS](#) [PubMed](#) [IUCr Journals](#) [Google Scholar](#)

Wiley Online Library

are two successive benzyne cycloadditions which is slightly higher than that obtained (steps). Bis-cycloadduct 15 was subjected to aromatization (TiCl₄, Zn, THF, RT, 1 h)[16] followed by hydrolysis of the silyl acetal (16) to give 17. The aromatization of 16 was carried out in a solvent mixture of CH₂Cl₂ and CH₃OH in 73% yield.[17]

16 M. A. Meador, H. Hart, *J. Org. Chem.* 1989, **54**, 2336–2341. [CrossRef](#) | [CAS](#) | [Web of Science®](#) Times Cited: 16 | [eJournals@cambridge - find full text'](#)

17 CCDC 1543805 (16) contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre.

18 Nitrile oxide 18 could be stored at –18 °C for at least one month.

Outline

Abstract

Graphical abstract

1. Introduction

2. Results and discussion

3. Conclusions

4. Experimental

Acknowledgements

Supplementary data

Research Data

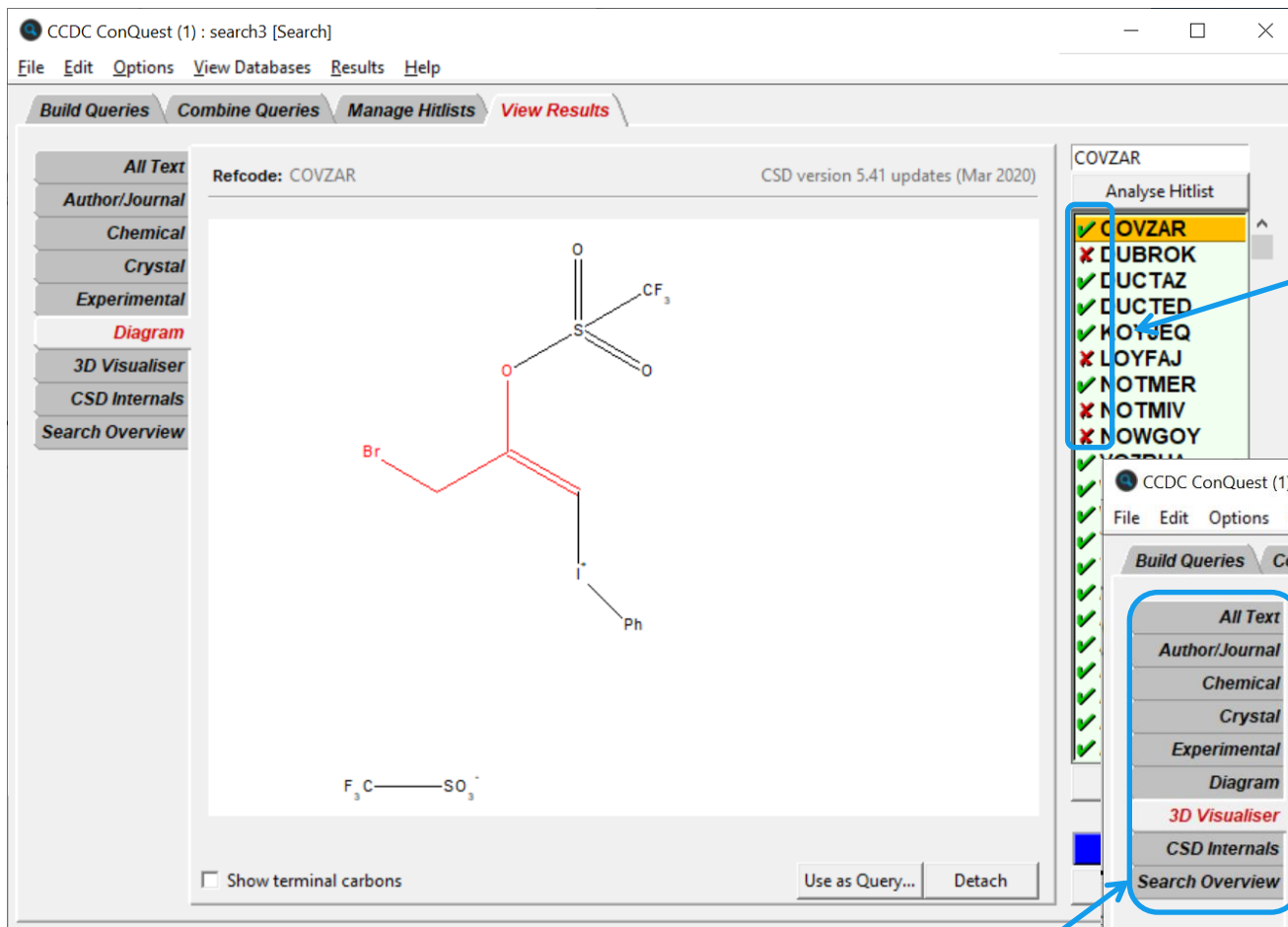
References and notes

ELSEVIER **SCHOLIX**

Research data for this article

Cambridge Crystallographic Data Center
Crystallographic data
Data associated with the article:
CCDC 689113: Experimental Crystal Structure Determination [↗](#)

CCDC



15

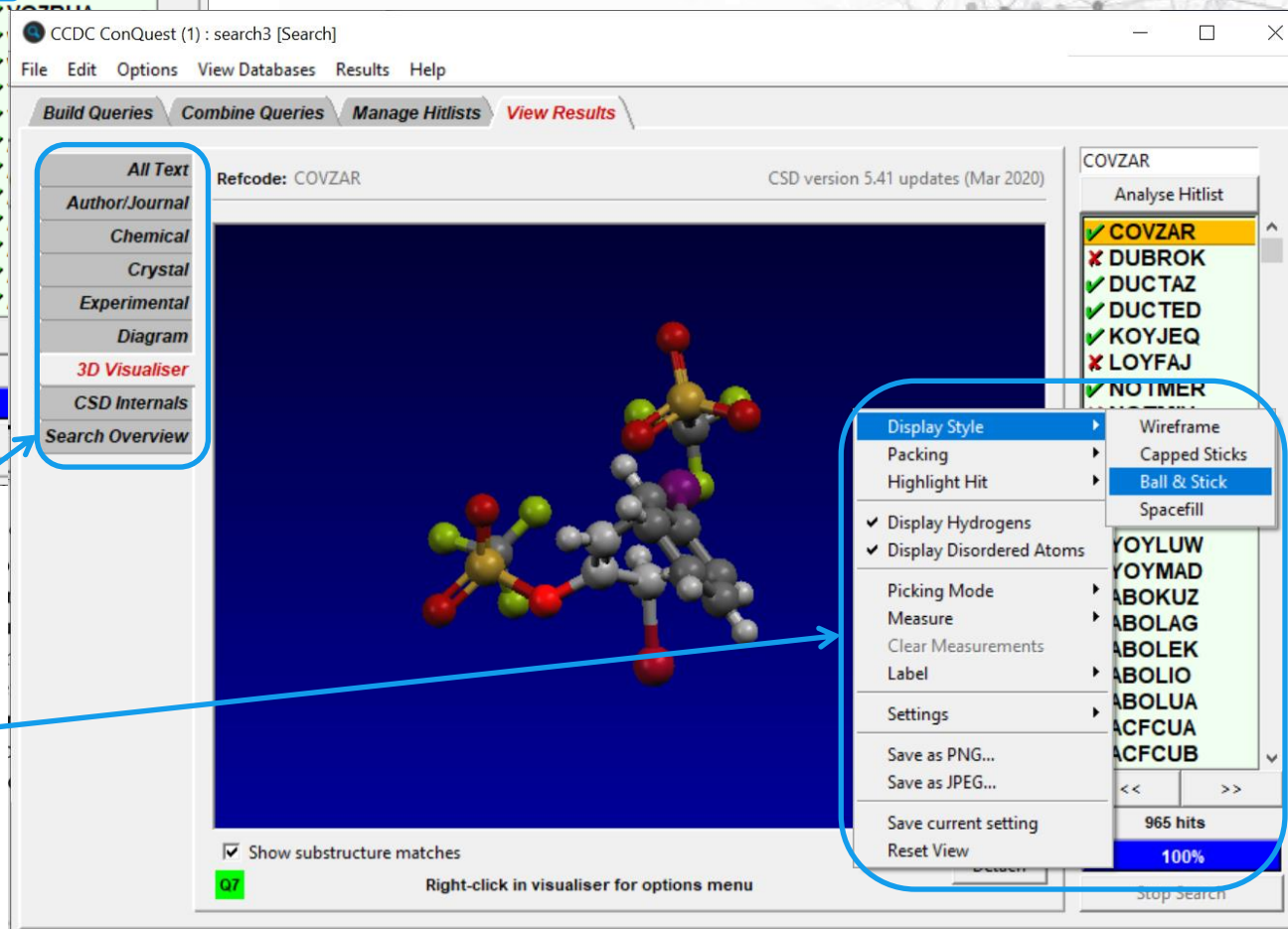
A hit list will be returned within the “results” tab of ConQuest. Left clicking on each REFCODE will toggle on and off that hit for future manipulation.



Left clicking each tab will display different information



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



The screenshot displays the CCDC ConQuest (1) : search3 [Search] application window. The 'File' menu is open, highlighting 'Export Entries as...'. The main window shows a 3D molecular model of a complex organic structure. The 'Export Entries: search3' dialog box is open, showing the following options:

- Select file type:** TAB: Tab separated list
- Select what to export:** ☒ All selected entries
- Select options:**
 - ☒ Bibliographic
 - ☒ Chemical
 - ☒ Crystallographic
 - ☒ Experimental
 - ☐ CSD internals
- Either:** Edit Filename and Save (Path: sers\ward\csds_data\search3.tab) **Save**
- Or:** Save via **File Popup**
- Progress bar:** 0% **Cancel**
- ☐ Keep window open when finished

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

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 Replace Delete

Chemical Class

CCDC Chemical Classifications are a set of categories of chemical compounds. They are used to group entries in the CCDC Chemical Classifications database. They are not intended to be used as a classification system for chemical compounds. They are not intended to be used as a classification system for chemical compounds.

Note that the CCDC Chemical Classifications are not a comprehensive set of categories. They are a subset of the categories used in the CCDC Chemical Classifications database. They are not intended to be used as a classification system for chemical compounds.

Find entries classified as: ----not defined----

and: ----not defined----

Search Store Cancel Reset

ConQuest – Space Group and cell search

Build Queries **Combine**

Draw

Peptide

Author/Journal

Name/Class

Elements

Formula

Space Group

Unit Cell

Z/Density

Experimental

All Text

Refcode (entry ID)



Space Group (1) - New

Space Group

Enter full or partial space group symbol or number ...

... or pick from list. Click on symbol to select

A1
B1
C1
F1
I1
P1
A-1

Current Space Group (with alternative settings)

☒ Include alternative settings of selected space group

Spacegroup Symmetry -----not defined-----

Crystal System -----not defined-----

Search Store Cancel Reset

Unit Cell (1) - New

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 1.5 % of longest cell dimension

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

Cell Parameters

a (Å)	=		alpha (°)	=	
b (Å)	=		beta (°)	=	
c (Å)	=		gamma (°)	=	

Search Store Cancel Reset

ConQuest – Elements and 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

☐ Exclude powder structures

Temperature of Structure Determination = ☒ K ☐ °C

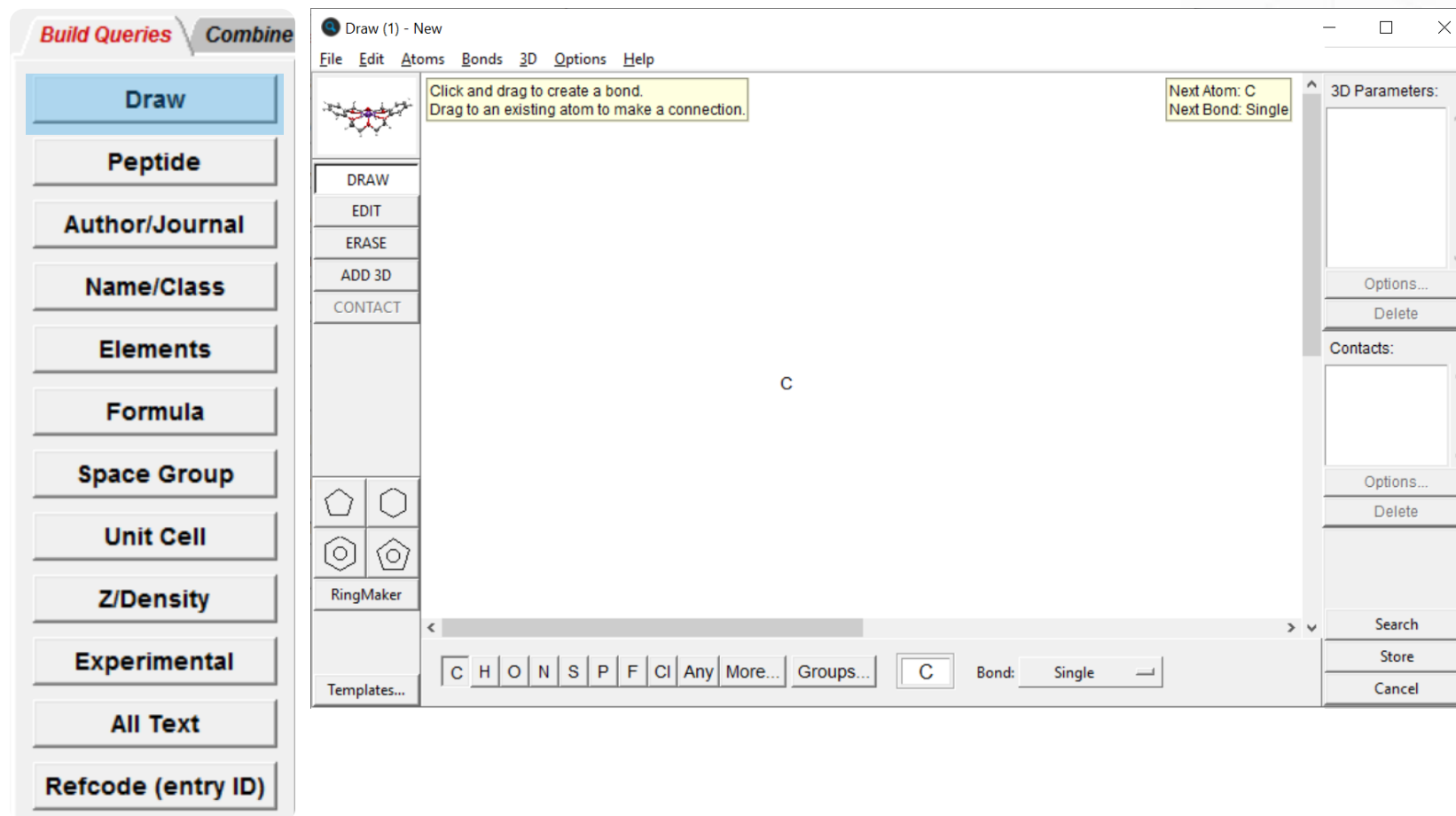
0 Room Temperature 610K

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

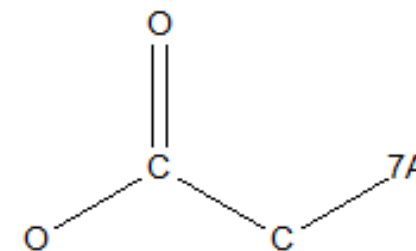
Radiation Source

Search Store Cancel Reset

ConQuest – Draw/Structure search

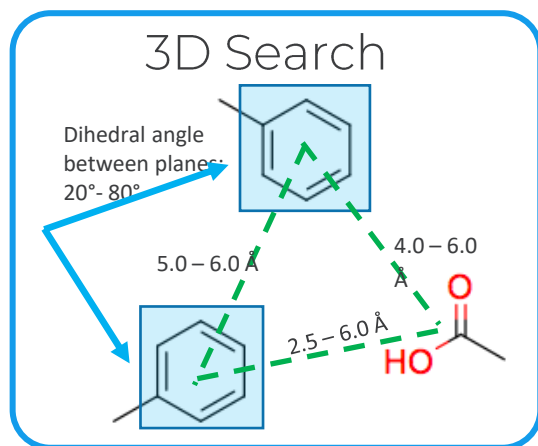


In the demo we will search for:

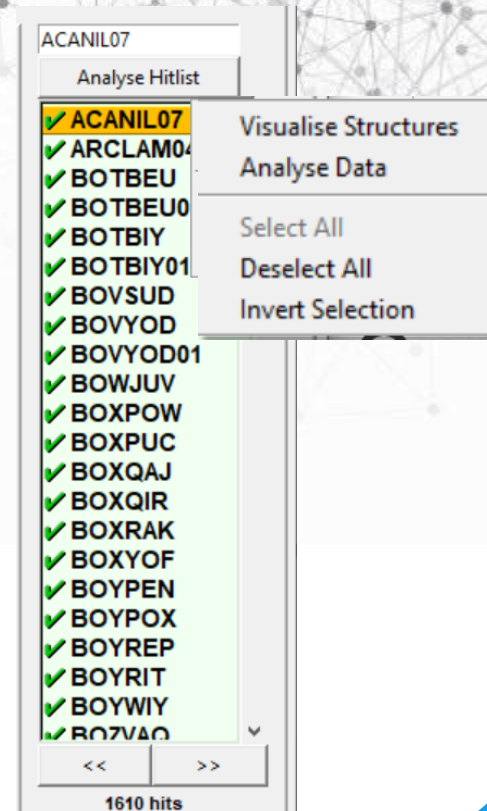
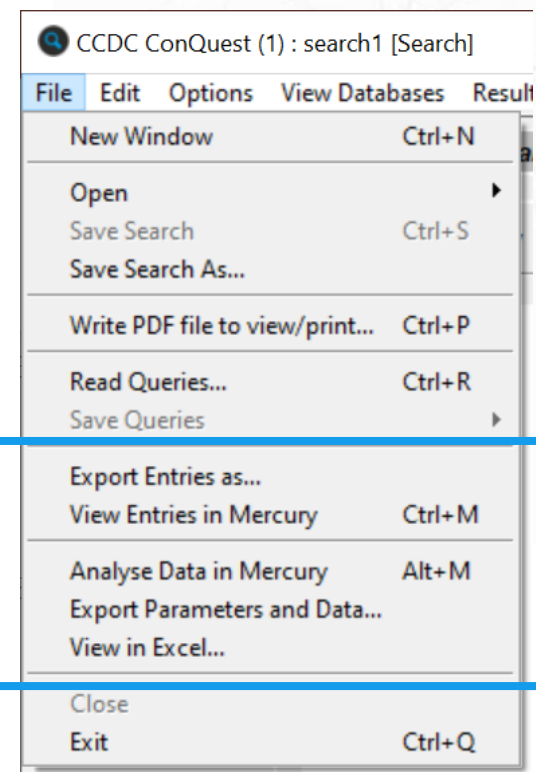


where 7A represents any halogen

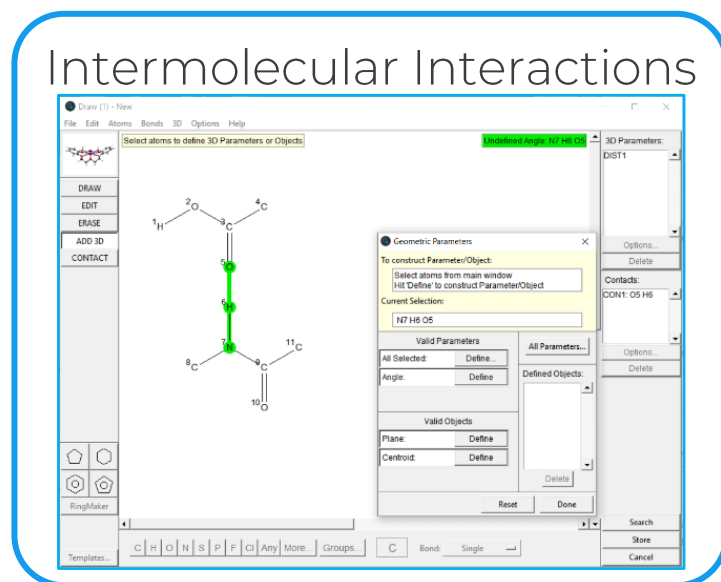
What else can you explore in ConQuest?



Export & Analyse Data



Search
In-House
Databases



... and more!

CCDC

Data analysis in Mercury

CCDC ConQuest (1) : search3 [Search]

File Edit Options View Databases Results

New Window Ctrl+N

Open

Save Search Ctrl+S

Save Search As...

Write PDF file to view/print... Ctrl+P

Read Queries... Ctrl+R

Save Queries

Export Entries as...

View Entries in Mercury Ctrl+M

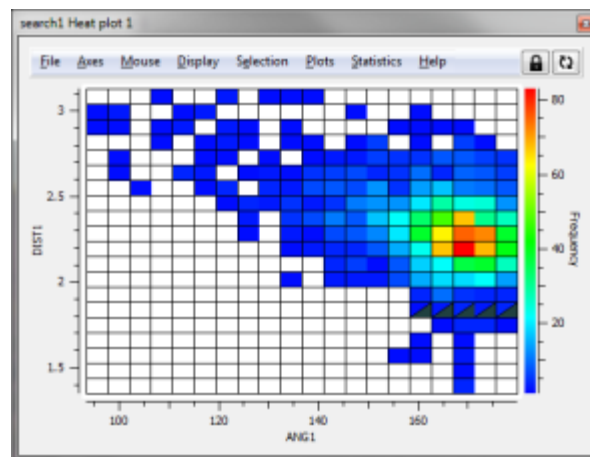
Analyse Data in Mercury Alt+M

Export Parameters and Data...

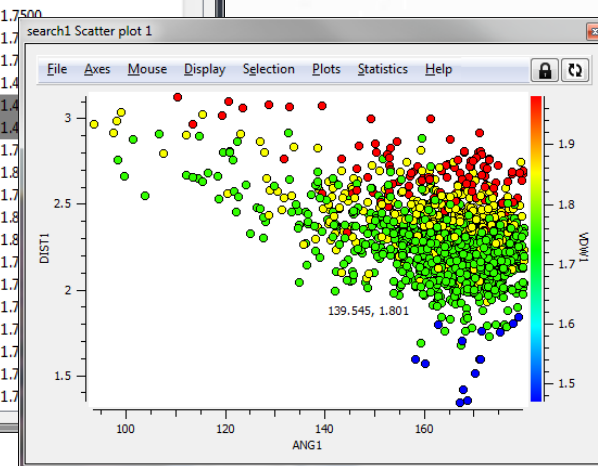
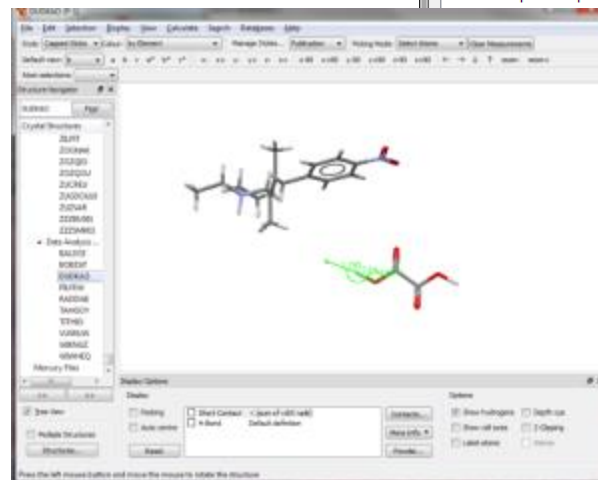
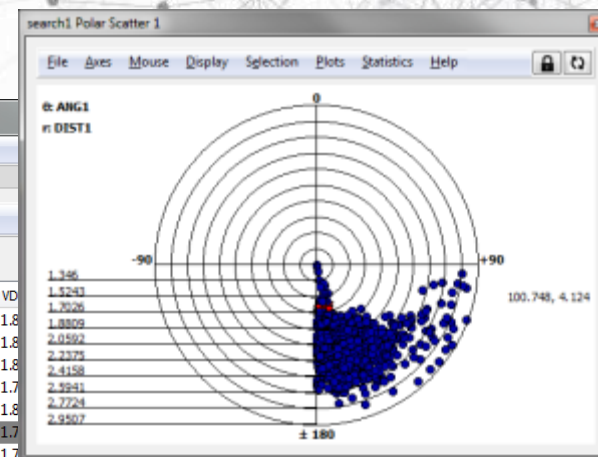
View in Excel...

Close

Exit Ctrl+Q

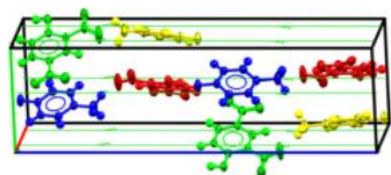


	NAME	Query	Fragment	ANG1	ANG2	ANO1	DIST1	LAB1	VD
9	WIBLEY	1	1	157.1810	108.0870	35	2.5040	Br1	1.8
10	WIBZEM	1	1	148.3450	105.0420	35	2.0700	Br1	1.8
1	WIFCOE	1	1	157.6580	109.4580	35	2.4070	Br1	1.8
12	WIHVAK	1	1	103.7470	176.6930	17	2.5530	Cl1	1.7
33	WIKMAE	1	1	123.0220	91.0810	35	2.9160	Br1	1.8
34	WIKNUZ	1	1	163.4610	115.6550	17	1.8480	Cl1	1.7
35	WIKNUZ	1	2	142.0390	116.8170	17	1.9980	Cl1	1.7
	WILYEW	1	1	136.4500	109.4520	17	2.3990	Cl1	1.7500
	WILYEW	1	2	174.2010	109.4290	17	2.2010	Cl1	1.7500
	WILYEW	1	3	167.8320	109.4460	17	2.3280	Cl1	1.7500
	WISJOX	1	1	168.5380	107.6350	17	1.9930	Cl1	1.7
	WISJOX	1	2	167.0650	108.3220	17	2.2280	Cl1	1.7
1	WIWHEQ	1	1	171.3820	109.5370	9	1.7630	F1	1.4
2	WIWHEQ	1	2	178.7710	109.5070	9	1.8470	F2	1.4
3	WIWHEQ	1	3	177.8390	109.4200	9	1.8100	F2	1.4
4	WIWSOK	1	1	167.5080	108.6550	17	2.2260	Cl1	1.7
	WIZLOH	1	1	169.6270	117.2500	35	2.4670	Br1	1.8
	WIZYAF	1	1	178.4360	106.2840	17	2.3930	Cl1	1.7
7	WOBMEF	1	1	171.8990	109.5670	35	2.3480	Br1	1.8
	WOBMUJ	1	1	159.9410	109.4460	35	2.3770	Br1	1.8
69	WOBMUV	1	1	179.9310	113.1610	17	2.3300	Cl1	1.7
...	WOBMUV...	1	1	164.1590	112.2520	17	2.3320	Cl1	1.7
1	WOCFOJ	1	1	142.4990	109.5570	17	2.6160	Cl1	1.7
	WOJZEB	1	1	174.7420	109.4890	17	2.1980	Cl1	1.7
	WOKRIX	1	1	172.7240	111.1030	17	2.1120	Cl1	1.7
74	WOMQUK	1	1	174.1710	111.1890	17	1.9840	Cl1	1.7
75	WOMQUK	1	2	148.4200	101.6230	17	2.0140	Cl1	1.7



Want to explore more?

Educational Resources



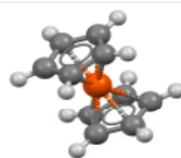
CSD-System

Essential crystallographic and structural chemistry capabilities.

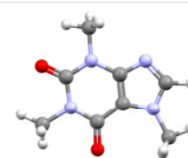
The wealth of information contained within the Cambridge Structural Database (CSD) extends far beyond a collection of crystal structures. Knowledge derived from these materials informs much of chemistry, biochemistry, and biology. Chemical and structural concepts are often difficult to grasp without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

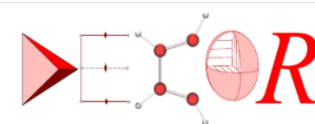
If you are an educator looking for supplementary teaching materials, find out more about the Teaching Database [here](#). If you have developed your own modules using the CSD and would like to share them with the broader community, please contact us at education@ccdc.cam.ac.uk.



Information on the Teaching Subset



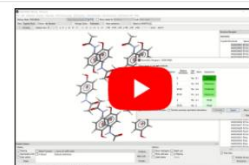
Access a series of teaching modules for use in the classroom



DECOR: Educational Resources for Teaching Crystallography



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



Watch software training and support videos

Explore the Periodic Table through Crystal Structures



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

Self-guided workshops about ConQuest in <https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-system-workshops/>

YouTube and LabTube channels links from <https://www.ccdc.cam.ac.uk/Community/educationalresources/ccdc-videos/>