

Introduction to the CSD-Enterprise Suite

Cambridge Crystallographic Data Centre (CCDC)

June 5th, 2018



The Cambridge Crystallographic Data Centre

The Cambridge Structural Database (CSD)

A database of organic and metal-organic
crystal structures

Established in 1965



~60 staff

Data curators

Software developers
Applications scientists

Cambridge UK • Piscataway NJ

International Data Repository

Archive of crystal structure data
High quality scientific database

Scientific Software Provider

Search/analysis/visualisation tools
Scientific applications

Collaborative Research Organisation

New methodologies
Fundamental research

Education and Outreach

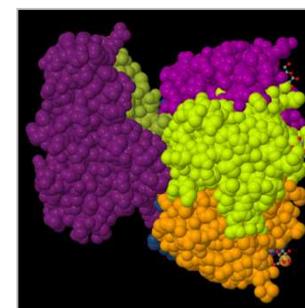
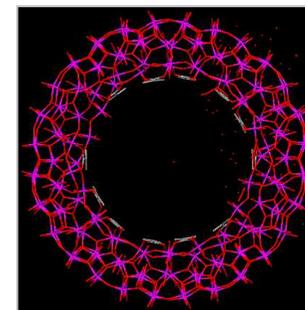
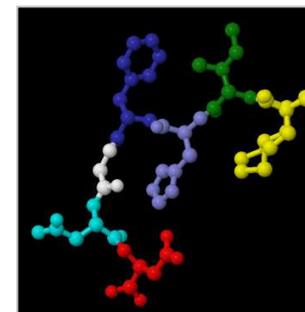
Conferences, Workshops,
Bespoke Training, Teaching Materials

Dedicated to the advancement of chemistry and crystallography for the public benefit through providing high quality information services and software.



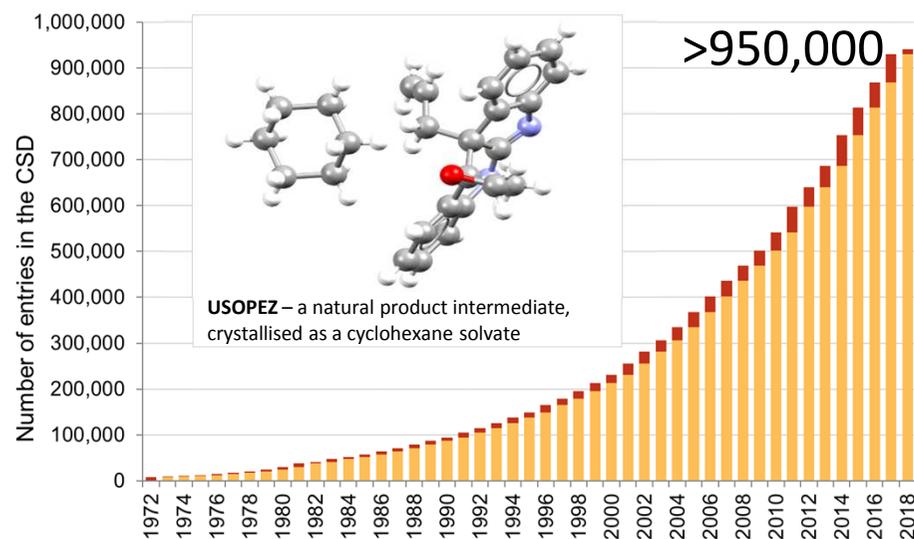
Crystal structure databases

- Cambridge Structural Database
 - organic and metal-organic compounds
 - >900,000 structures
 - Established in 1965
- Inorganic Crystal Structure Database
 - inorganic compounds
 - >185,000 structures
 - Established in 1978
- Protein Data Bank
 - biological macromolecules
 - >120,000 structures
 - Established in 1971





The Cambridge Structural Database



- ❑ 900,000+ small-molecule crystal structures
- ❑ Over 80,000 datasets deposited annually
- ❑ Structures available for anyone to download
- ❑ Links to over 1,000 journals
- ❑ Enriched and annotated by experts
- ❑ Access to data and knowledge

CSD Entry: AHEPUY

AHEPUY : N-(4-Hydroxyphenyl)acetamide morpholine
Space Group: P-1, Cell: a 8.710(4)Å b 9.920(5)Å c 12.385(5)Å, α 102.35(3) $^\circ$ β 108.33(2) $^\circ$ γ 96.68(3) $^\circ$

3D viewer

Chemical diagram

Additional CCDC details

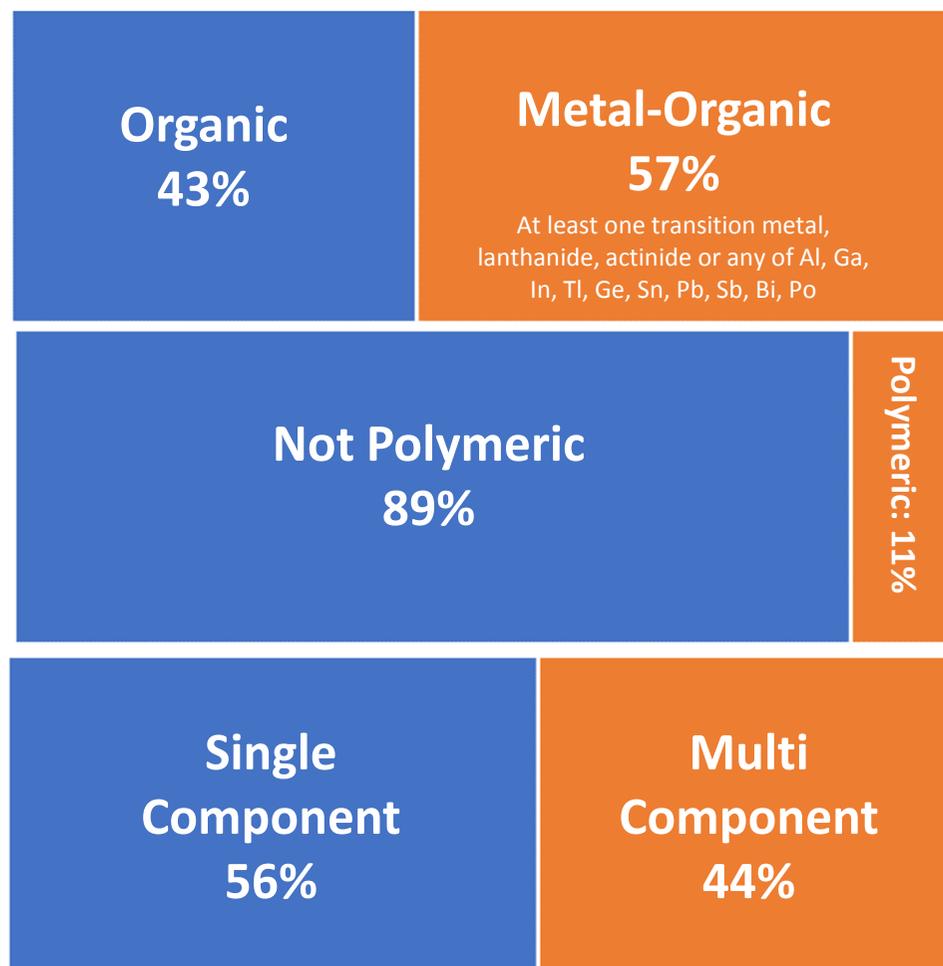
CCDC Number	200783
CCDC Citation	I.D.H.Oswald, W.D.S.Motherwell, S.Parsons, C.R.Pulham CCDC 200783: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/cc6qxp
Synonyms	Paracetamol morpholine, Acetaminophen morpholine
Deposited on	16/12/2002

Associated publications

I.D.H.Oswald, W.D.S.Motherwell, S.Parsons, C.R.Pulham, *Acta Crystallographica Section E: Structure Reports Online*, 2002, 58, 1290, DOI: 10.1107/S1600536802018111

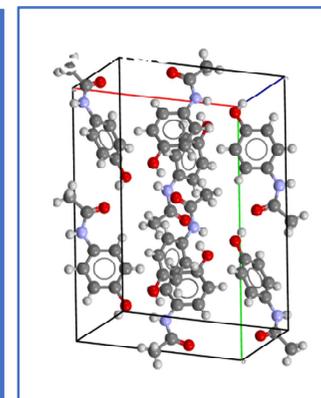


What's in the CSD?



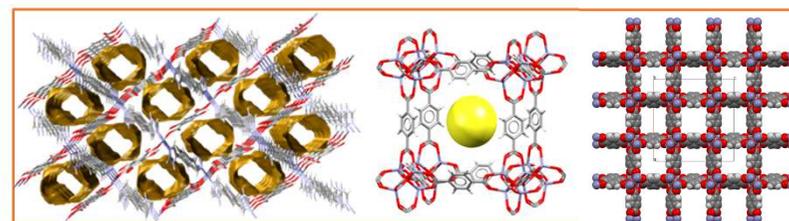
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



Metal-Organic

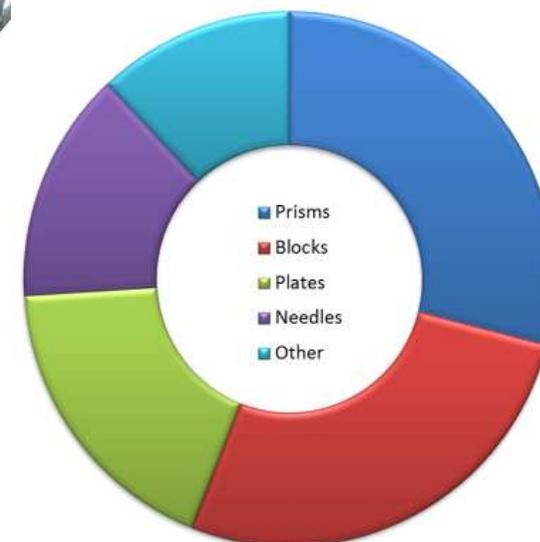
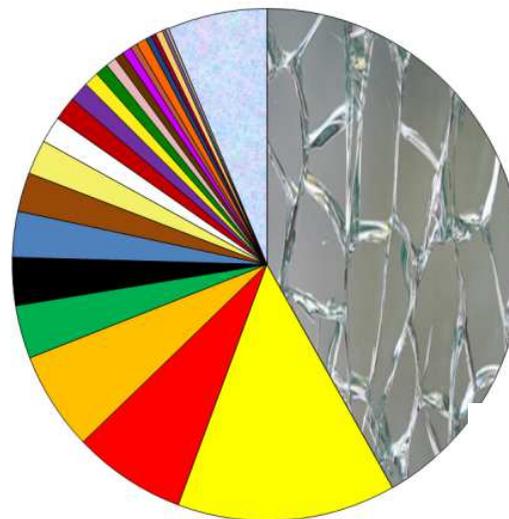
- Metal Organic Frameworks
- Models for new catalysts
- Porous frameworks for gas storage
- Fundamental chemical bonding





What other data is in the CSD?

- 9,959 polymorphic families
- 165,617 melting points
- 747,577 crystal colours
- 614,544 crystal shapes
- 21,939 bioactivity details
- 9,727 natural source data
- 46,238 entries classified
- > 250,000 oxidation states
 - Including 240,666 transition metal oxidation states





Accessing drugs and ligands in the CSD

CSD Entry: TOBBOB

Your query was: CCDC identifier(s): TOBBOB and the search returned 1 record

Refcode	CCDC Number
TOBBOB	150718

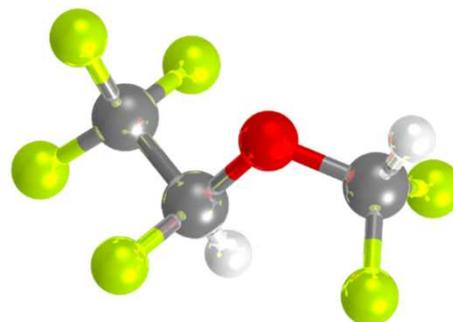
Download

3D viewer

Chemical diagram

Additional CCDC details

CCDC Number: 150718
CCDC Citation: V.Schurig, M.Juza, B.S.Oreen CCDC 150718: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/cds15aw
Synonyms: (+)-Desflurane, PDB Chemical Component code: DSF, DrugBank: DB01189
Deposited on: 09/10/2000



Refcode: TOBBOB

Formula: $C_3H_2F_6O$

Name: 2-Difluoromethoxy-1,1,1,2-tetrafluoroethane

Synonym: (+)-Desflurane; PDB Chemical Component code: DSF; DrugBank: DB01189

CCDC Class:

Source:

Melting Point: 147K

Colour: colorless

Extra Information: absolute configuration; inhalation anaesthetic activity

RCSB PDB Deposit Search Visualize Analyze Download Log Out MyPDB Login

Protein Data Bank in Europe

DSF

(2S)-2-(difluoromethoxy)-1,1,1,2-tetrafluoroethane

DSF is found in 1 [entry](#). Example includes 3P4W

Find related ligands: [Stereoisomers](#) [Similar ligands](#) [Chemical Structure Search](#)

View summary at Ligand Expo

Chemical Component Summary		Chemical Details	
Name	(2S)-2-(difluoromethoxy)-1,1,1,2-tetrafluoroethane	Formal Charge	0
Identifiers	(2S)-2-(difluoromethoxy)-1,1,1,2-tetrafluoro-ethane	Atom Count	12
Formula	$C_3H_2F_6O$	Chiral Atom Count	1
Molecular Weight	168.04 g/mol	Chiral Atoms	CAI
Type	NON-POLYMER	Bond Count	11
Isomeric SMILES	FC(F)C(C@H)(F)C(F)(F)F	Aromatic Bond Count	0
inChI	InChI=1S/C3H2F6O/C4=1(3(7,8)/10-2(5)6/11-2H11-m 1/s1	Leaving Atoms	n/a
inChIKey	DPYMFVXJLWUEU-PVJQKRUSA-N		

EMBL-EBI

Protein Data Bank in Europe

3p4w > DSF

(2S)-2-(difluoromethoxy)-1,1,1,2-tetrafluoroethane

Formula: $C_3H_2F_6O$

Molecular weight: 168 Da

SMILES: OpenEye GETScripts (1,7,9): [C@H](C(F)(F)F)OC(F)F

DSF 319 bound to chain A

Environment details

Quick links

- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

Downloads

EBI resources (DSF)

- DSF in PDBChem
- Binding site details
- Interaction statistics

DRUGBANK

Search Downloads About Help Contact Us

Targets (7) Carriers (1) Biointeractions (7) Show Drugs with Similar Structures

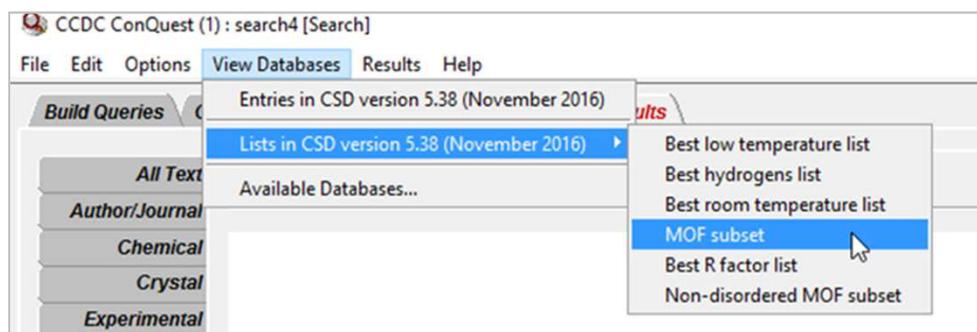
Get DrugBank to go! The DrugBank app for iOS and Android is coming soon. Sign up to get early access.

Identification	
Name	Desflurane
Accession Number	DB01189 (APR000907)
Type	Small Molecule
Groups	Approved
Description	Desflurane is a highly fluorinated methyl ethyl ether used for maintenance of general anaesthesia. Volatile agents such as desflurane may activate GABA channels and hyperpolarize cell membranes. In addition, they may inhibit certain calcium channels and therefore prevent release of neurotransmitters and inhibit glutamate channels. Volatile anaesthetics easily partition into cellular membranes and could expand the volume of the cell membrane and subsequently distort channels necessary for sodium ion flux and the development of action potentials necessary for synaptic transmission. Desflurane preconditions human myocardium against ischemia through activation of mitochondrial K(ATP) channels, adenosine A1 receptor, and alpha and beta adrenoceptors.
Structure	



Metal-Organic Frameworks

- Two CSD subsets identifying all metal-organic frameworks within the database
 1. MOF list >**75,000** entries intended to be as widely-applicable as possible
 2. Non-disordered list >**61,000** entries for high-throughput calculations
 - Both subsets updated quarterly
- Python script available to download



cm CHEMISTRY OF MATERIALS Perspective
pubs.acs.org/cm

Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future

Peyman Z. Moghadam,^{*,†,§} Aurelia Li,^{†,§} Seth B. Wiggin,^{†,§} Andi Tao,[†] Andrew G. P. Maloney,[†] Peter A. Wood,[†] Suzanna C. Ward,[†] and David Fairen-Jimenez^{*,†}

[†]Adsorption & Advanced Materials Laboratory (AAML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, United Kingdom
[‡]The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

Supporting Information

ABSTRACT: We report the generation and characterization of the most complete collection of metal–organic frameworks (MOFs) maintained and updated, for the first time, by the Cambridge Crystallographic Data Centre (CCDC). To set up this subset, we asked the question “what is a MOF?” and implemented a number of “look-for-MOF” criteria embedded within a bespoke Cambridge Structural Database (CSD) Python API workflow to identify and extract information on 69 666 MOF materials. The CSD MOF subset is updated regularly with subsequent MOF additions to the CSD, bringing a unique record for all researchers working in the area of porous materials around the world, whether to perform high-throughput computational screening for materials discovery or to have a global view over the existing structures in a single resource. Using this resource, we then developed and used an array of computational tools to remove residual solvent molecules from the framework pores of all the MOFs identified and went on to analyze geometrical and physical properties of nonordered structures.

INTRODUCTION
Metal–organic frameworks (MOFs)^{1–3} are one of the most exciting recent advances in porous materials science. MOFs are crystalline materials containing metal clusters connected by organic linkers (Figure 1, inset) and are characterized by their wide range of well-defined and, in principle, tailorable pore sizes (from micro- to mesoporosity), pore geometries, high void fractions, and large surface areas. MOFs can reach apparent surface areas as high as 10 000 m²/g vs ca. 1000 m²/g for zeolites and up to 3500 m²/g for activated carbons,^{4,5} and their large pore volumes have been unsurpassed by any other porous material to date. These unique properties have paved the way for MOF research to grow substantially, and applications are being considered in many areas including gas storage,^{6,7} separation,^{8–10} catalysis,^{11,12} and carbon capture,^{13–15} as well as biomolecule encapsulation,¹⁶ drug delivery,^{17–19} and imaging.²⁰

Because of the modular nature of MOFs, an almost unlimited number of structures can be envisioned. The Cambridge Crystallographic Data Centre (CCDC)—the world’s repository of small molecule crystal structures, which includes small organics as well as MOFs and other porous materials.²¹ As shown in Figure 1, the number of entries in the CSD has substantially increased over the last 44 years, reaching a milestone of 850 000+ entries in 2016. Among these, the number of MOFs has also increased dramatically in the past decade to an estimated number of ca. 70 000 materials, *vide infra*. This trend is only going to increase even further as the

Received: February 2, 2017
Revised: March 14, 2017
Published: March 14, 2017

ACS Publications © 2017 American Chemical Society 2618
DOI: 10.1021/acs.chemmater.7b00441
Chem. Mater. 2017, 29, 2618–2625



CSD-Community

- CSD-Community provides a selection of products and services free of charge for the benefit of the scientific community
- This includes:



enCIFer

CIF checking, editing and visualisation software from the CCDC



Relibase

Easy Searching for Protein-Ligand Complexes



CSDSymmetry

The most complete collation of observed molecular and crystallographic symmetry properties to date



Deposit Structures

Upload your data to the CCDC for inclusion in the Cambridge Structural Database



Mercury

Crystal Structure Visualisation and Exploration Made Easy



CSD Educational Collection

Free 500 structure subset of the CSD for teaching purposes



CellCheckCSD

A command-line tool for performing crystal structure reduced cell checks against the CSD



Access Structures

View and retrieve structures in the Cambridge Structural Database



CSD-Deposit

Deposit Structures
Upload your data to the CCDC for inclusion in the Cambridge Structural Database



CIF deposition and validation service

This web service enables you to submit CIF files and associated structure factor files to the CCDC and for your structures to be included in the Cambridge Structural Database.

Deposition allows you to correct syntax errors, check the integrity of your data and add additional data.

Please include structure factor data for all structures.

- Files should be in CIF, FCF or HKL format and may be included in a ZIP file
- At least one CIF file must be included in the submission
- All files submitted on one form should correspond to one publication only
- There is a limit of 50 MB per file and a limit of 100 MB for the total size of files uploaded
- For more information please see our Structure Deposition Information page

Your name:
 Your e-mail address:
 Institution (eg. University/Company):
 CCDC number(s) for resubmissions:
 CIF/FCF/HKL/ZIP files: Done

Remember my details
 I wish to run the IUCr checkCIF/PLATON service on my data

1 Upload **2 Check Syntax** **3 Validation** **4 Add Publication** **5 Enhance Data** **6 Review** **7 Submit**

No Structure Factor Data have been uploaded.
Structure Factor Data are an essential part of the deposition. You should click 'Go Back' to add Structure Factor Data to your deposition. If in exceptional circumstances you are unable to include Structure Factor data with your deposition you may continue by clicking 'Proceed to Next Step'.

Retrieve Deposited Files

1 Upload **2 Check Syntax** **3 Validation** **4 Add Publication** **5 Enhance Data** **6 Review** **7 Submit**

Check Syntax

The files highlighted in red in the left hand column contain errors that need fixing before proceeding.

Please click on any red file names in the left-hand column, make the appropriate edits and then click the 'Save & Recheck File' button before proceeding to the next step.

For more information on how to fix errors please see our [correcting CIFs](#) page.

Pick file to edit:

- invalid_2.cif
- MultipleCifs.cif

File contents invalid_2.cif

```

1 data_1
2 _chemical_formula_sum 'C12H14N2O4'
3 _chemical_formula_weight 282.26
4 _cell_length_a 10.1118(10)
5 _cell_length_b 10.1118(10)
6 _cell_length_c 10.1118(10)
7 _cell_angle_alpha 90.0000(10)
8 _cell_angle_beta 90.0000(10)
9 _cell_angle_gamma 90.0000(10)
10 _symmetry_space_group_name 'P 21 21 21'
11 _symmetry_international 'P212121'
12 _monoclinic_P21_21_21
13 _monoclinic_P21_21_21
14 _monoclinic_P21_21_21
15 _monoclinic_P21_21_21
16 _monoclinic_P21_21_21
17 loop
18 _atom_site_label
19 _atom_site_occupancy
20 _atom_site_fract_x
21 _atom_site_fract_y
22 _atom_site_fract_z
  
```

Please click on the error message to navigate to the location of the error in the CIF.

1 Upload **2 Check Syntax** **3 Validation** **4 Add Publication** **5 Enhance Data** **6 Review** **7 Submit**

Add Publication

Please check and add/update the publication details shown below.

If you don't know the full publication details then please provide the current list of authors for the data you are depositing.

Authors:
 Journal Name:
 Volume:
 Year:
 Page:
 Additional Information:

If you do not intend to publish the data in the scientific literature and would like to share the data immediately through the Cambridge Structural Database then please click the 'CSD Communication' button below.

1 Upload **2 Check Syntax** **3 Validation** **4 Add Publication** **5 Enhance Data** **6 Review** **7 Submit**

Validation

View reports on the consistency and integrity of your structures

Datablock:

Structure	IUCr checkCIF
invalid_2.cif	<input type="button" value="View Report"/> Enter Response
MultipleCifs.cif	<input type="button" value="View Report"/> Enter Response
data_I	<input type="button" value="View Report"/> No Response Requ
data_II	<input type="button" value="View Report"/> Enter Response
data_III	<input type="button" value="View Report"/> Enter Response

Home Deposit Structures Access Structures CCDC

1 Upload **2 Check Syntax** **3 Validation** **4 Add Publication** **5 Enhance Data** **6 Review** **7 Submit**

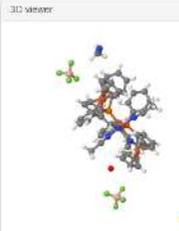
Enhance Data

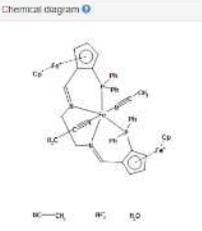
Please check the information below for each structure submitted and add as much additional information as possible. Update the CSD fields on the right hand side rather than the CIF directly. Any edits to the CSD fields will update the CIF automatically.

When you have checked each structure please proceed to the next step.

Pick a structure to edit:

- checkCIF_examples.cif
- data_sa2906c
- data_sa2906b
- data_sa2906a

3D viewer: 

Chemical diagram: 

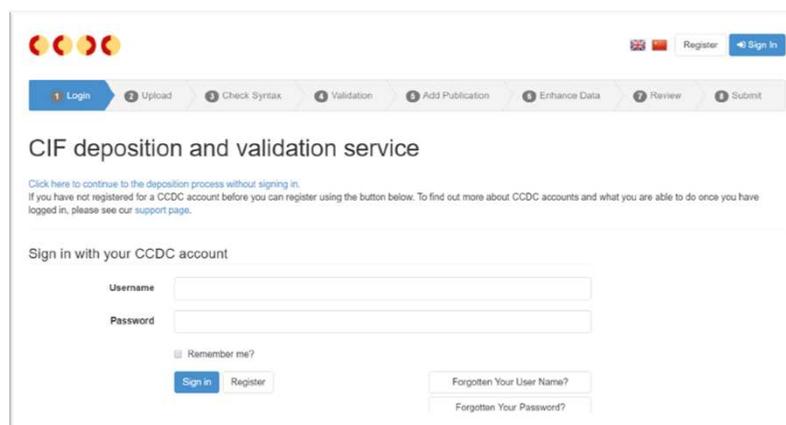
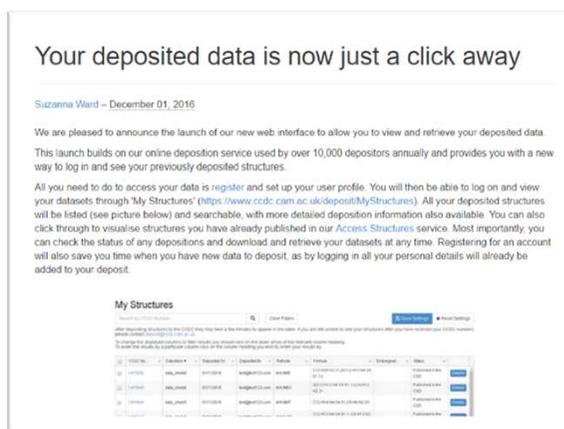
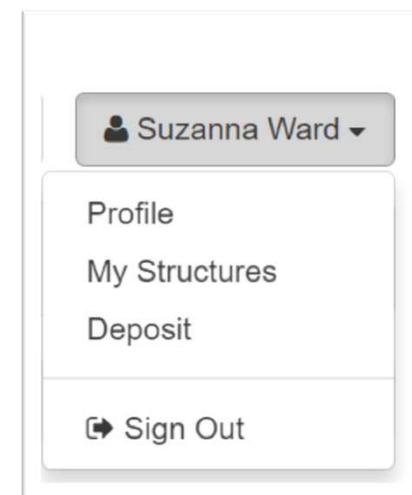
CSD Fields:

Compound name:
 Synonyms/other names:
 IUPAC name:
 Crystal colour:
 Crystal habit:



Increasing deposition functionality

- A deposition portal **launched December 1st 2016**
 - Ability to log on and view and retrieve depositions
 - Deposit and revise data
 - Edit and update basic information
 - Publish data directly through a database
 - Share data with co-workers
- Over 15,000 registered users





My Structure Details

[Revise Structure](#)[Back to My Structures](#)

Datablock: tBu10kbar

Space Group: I2/a, Cell: a 14.811Å b 6.456Å c 19.759Å, α 90° β 94.060(8)° γ 90°Formula: C₁₁ H₁₅ N₁ O₂, Temperature: 298 K

Additional CCDC Details

Deposition Number 1584792

Refcode

Compound Name 2-(4-butyl-6-((hydroxyimino)methyl)phenol

Deposited On 09/11/2017

Deposited By

Additional Depositors

Status Unpublished [Publish as a CSD](#)

Embargoed Date 05/03/2019

Associated Publications

Publication	Created on	Modified on
Suzanna Ward	09/11/2017	09/11/2017

[Add Publication](#)

Structure Shared With

Name	Edit Rights	Sharing Rights
------	-------------	----------------

[Share Structure](#)

<input type="checkbox"/>		825800	data_2007may0...	06/06/2011
--------------------------	--	--------	------------------	------------

<input type="checkbox"/>	Details	825789		
--------------------------	-------------------------	--------	--	--

2 20 items per page

[Download Selected 0](#) [View 0 in Access Structures](#) [Export](#) Clear selection

Chemical diagram

No Diagram Available



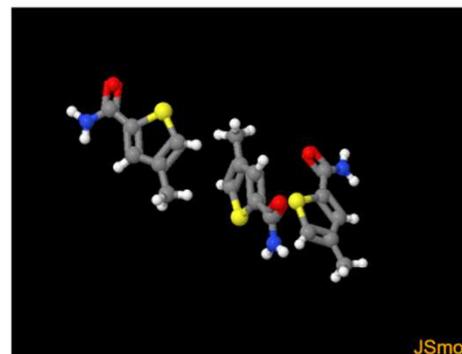
Suzanna Ward

[Save Settings](#)[Reset Settings](#)

EPAGAE : 4-Methylthiophene-2-carboxamide

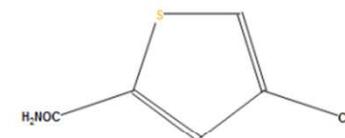
Space Group: P $\bar{1}$ (2), Cell: a 10.0846(3)Å b 10.4718(3)Å c 10.9149(3)Å, α 106.7180(10)° β 111.473(2)° γ 95.603(2)°

3D viewer

[H](#) [Disorder](#) [Reset](#) [Menu](#) [Open](#) [Share](#)

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

Chemical diagram

[View group symbols key](#)

Additional details

Deposition Number	825792
Data Citation	S.L.Huth, M.B.Hursthouse CCDC 825792: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/ccwq9gb
Deposited on	13/05/2011



CSD Communications

CSD COMMUNICATIONS

The Cambridge Crystallographic Data Centre (CCDC) allows you to publish data directly through the Cambridge Structural Database (CSD) as a *CSD Communication*. Over 5,000 structures were published in this way in 2016 making *CSD Communications* the number 1 place to publish crystal structures.

What are the benefits of a *CSD Communication*?

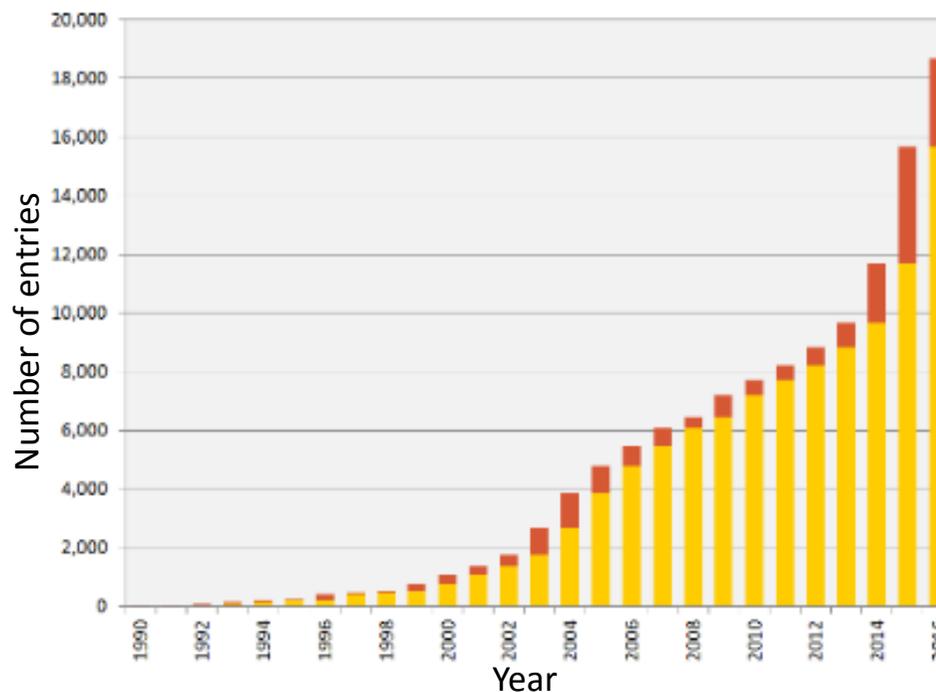
 RECOGNITION A citable DOI allows you to receive credit for your structures	 DISCOVERABILITY Automatic linking from third-party repositories	 EASE Publish data through the deposition process and portal
 INTEGRITY Resulting CSD entry curated and enhanced by our scientific editors	 COST-FREE Both free to create and free to access	 SPEED Your data made instantly available to the world via Access Structures
 COMPLIANCE Meets funder data policies	 SECURITY Free long-term data preservation in a trusted repository	 IMPACT Your data can be re-used to design drugs, materials, etc

From deposition to access through *CSD Communications*
For more information see <https://www.ccdc.cam.ac.uk/Community/depositstructure/CSDCommunications/>

www.ccdc.cam.ac.uk UK: +44 1223 336408 US: +1 (848) 445 4893

© Andy Brunning 2017 - www.compoundchem.com/chemunicate
Produced for the Cambridge Crystallographic Data Centre. www.ccdc.cam.ac.uk

- Allow data to be published directly through the CSD
 - During deposition
 - After Deposition via My Structures
- > 23,000 *CSD Communications*





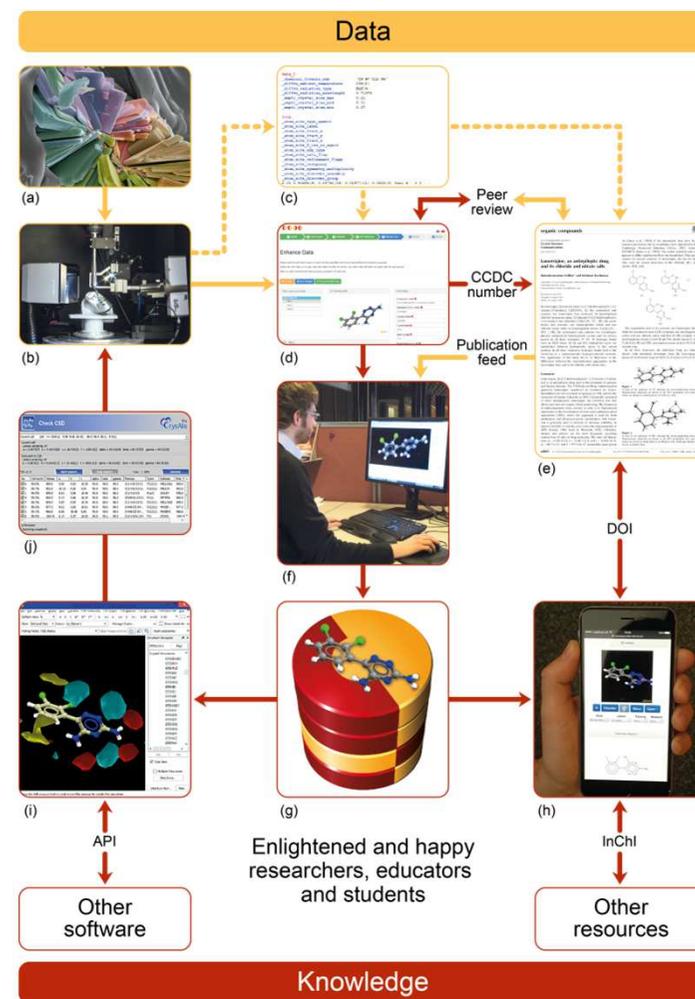
Deposition guidelines

- Data deposited pre-publication
 - Deposition when crystallographer most engaged
 - Enables links to data sets at the point of publication
 - Tailored deposition service

Small molecule single crystal data

Authors should present their crystal data in a CIF (Crystallographic Information File) format and deposit any organic or organometallic structural information with the Cambridge Crystallographic Data Centre (CCDC) before they submit their manuscript to us. Data will be held in the CCDC's confidential archive until publication of the article, when data for organic and organometallic compounds will be entered into the Cambridge Structural Database. Authors are encouraged to deposit inorganic crystal structures with the ICSD, hosted by FIZ Karlsruhe.

During submission of a manuscript to the Royal Society of Chemistry using our online submission system, authors will be asked to provide CCDC reference numbers. CIFs should not be submitted with the manuscript (these should have already been deposited with the CCDC/ICSD, see above). Any revised CIFs obtained subsequently should be deposited directly with the CCDC before the revised manuscript is submitted to us. CCDC or ICSD numbers should be included in the manuscript prior to submission.





Data management

Microsoft Dynamics CRM

Deposits: **Awaiting DC Intervention - External**

Name	Status Reason
1283576-DEP	Awaiting CCDC Numbers
1283574-DEP	Awaiting Correction
1283565-DEP	Awaiting CCDC Numbers
1283562-DEP	Awaiting CCDC Numbers
1283561-DEP	Awaiting Correction
1283560-DEP	In progress of assigning ...
1283556-DEP	Awaiting Correction
1283526-DEP	Awaiting Correction

Publication
Eur.J.Inorg.Chem. (2013) ,3316
Name Eur.J.Inorg.Chem. (2013) ,3316

Curated Data: Curated Data Associated View

Name	Deposit	CCDC Number	Source File (External Format)	Status Reason
1643223-C...	1000218-DEP	929,193	structure_1999264251.txt	Added to
1762050-C...	1274796-DEP	929,192	ccdc.cif	Added to
1762051-C...	1274796-DEP	929,191	dvshsun11vls_mach.cif	Added to
1643224-C...	1000218-DEP	929,192	structure_1999264250.txt	Inactive
1643225-C...	1000218-DEP	929,191	structure_1999264248.txt	Inactive

CSD Editor - 1660398

Chemical Diagram: OC1=CC=C(C=C1)C#CC#CC2=CC=CC=C2

Number Of Coordinates: 53

Compound Name: 1,1,6,6-Tetraphenyl-2,4-hexadiyne-1,6-diol bis(6-methyl-2(1H)-pyridinone)

Formula: C₃₀H₂₂O₂(C₆H₅N O)

Cell Lengths: a 8.4750(5) b 10.8530(6) c 11.3600(10)

Cell Angles: α 64.798(3) β 81.182(2) γ 69.185(7)

Space Group: P -1

Temperature (K): 393

Polymorph: P -1

Cross-References: Room Temp.(283-303)

Dashboard: Deposit 1

Status of Deposits: Awaiting DC Intervention - Status of Deposits

Useful Queues: Automatic Queue, Manual Queue

Status of Deposits: Awaiting DC Intervention - Internal

Activities: My Tasks

CSD Editor - 1660398

Chemical Diagram: OC1=CC=C(C=C1)C#CC#CC2=CC=CC=C2

Number Of Coordinates: 53

Compound Name: 1,1,6,6-Tetraphenyl-2,4-hexadiyne-1,6-diol bis(6-methyl-2(1H)-pyridinone)

Formula: C₃₀H₂₂O₂(C₆H₅N O)

Cell Lengths: a 8.4750(5) b 10.8530(6) c 11.3600(10)

Cell Angles: α 64.798(3) β 81.182(2) γ 69.185(7)

Space Group: P -1

Temperature (K): 393

Polymorph: P -1

Cross-References: Room Temp.(283-303)

Publication
Eur.J.Inorg.Chem. (2013) ,3316
Name Eur.J.Inorg.Chem. (2013) ,3316

Curated Data: Curated Data Associated View

Name	Deposit	CCDC Number	Source File (External Format)	Status Reason
1643223-C...	1000218-DEP	929,193	structure_1999264251.txt	Added to
1762050-C...	1274796-DEP	929,192	ccdc.cif	Added to
1762051-C...	1274796-DEP	929,191	dvshsun11vls_mach.cif	Added to
1643224-C...	1000218-DEP	929,192	structure_1999264250.txt	Inactive
1643225-C...	1000218-DEP	929,191	structure_1999264248.txt	Inactive

CCDC Number CCDC 910817

Identifier/CSD Refcode 1660398/REWJAG

Previous Refcode

Space Group P -1

Z, Z' Z: 1 Z': 0.5

Temperature (K) 393

Temperature (text) 393 K

Formula weight (CCDC) 632.745

Density (CCDC) 1.18895

Published Formula Weight 632.73

Density (author) 1.189

Powder Study unknown

Radiation Probe unknown

Color colorless

Habit Blocks

Disorder C16A,C18A,C20A,O2A and C16B,C18B,C20B disordered over two sites with occupancies 0.627:0.373.

Literature Reference Tian, S.; Wei, Y.; Deng, K.; Cao, T.; Lv, M.; Botoshansky, M.; Kalfort, *Cryst. Growth Des.* (2013), **13**, 936, doi:10.1021/cp3016707

Compound Name 1,1,6,6-Tetraphenyl-2,4-hexadiyne-1,6-diol bis(6-methyl-2(1H)-pyridinone)

Formula C₃₀H₂₂O₂(C₆H₅N O)

Cell Lengths a 8.4750(5) b 10.8530(6) c 11.3600(10)

Cell Angles α 64.798(3) β 81.182(2) γ 69.185(7)

Disorder C16A,C18A,C20A,O2A and C16B,C18B,C20B disordered over two sites with occupancies 0.627:0.373.

Errors Total of 1
CSD Editor unusually high temperature (above 320K/47deg.C)

Warnings Total of 3
Line95 CIFDatabase Data value should be a number ; diffn_standards_number

Cell contents C30 H22 O2(4) C6 H7 N1 O1; Formula unit: C30 H22 O2(4) C6 H7 N1 O1; Z= 1, Z'= 0.50; perfect match

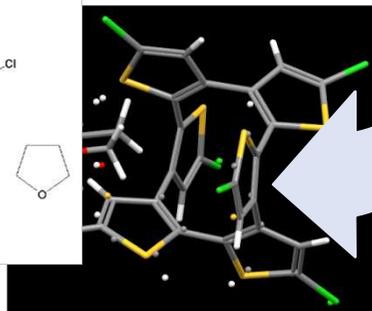
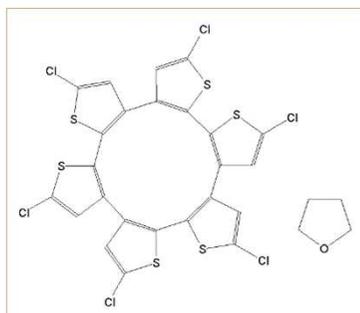
Discovery and re-use - Assignment of chemistry

Assignment of chemistry is important for discovery, data re-use, mining, analysis and interoperability

```

loop_
  _atom_site_label
  _atom_site_type_symbol
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_U_iso_or_equiv
  _atom_site_adp_type
  _atom_site_occupancy
  _atom_site_symmetry_multiplicity
  _atom_site_calc_flag
  _atom_site_refinement_flags
  _atom_site_disorder_assembly
  _atom_site_disorder_group
C11 Cl 0.5993(2) 1.0007(7) 0.8131(17) 0.044(3) Uani 0.50 1 d PDU A 1
S1 S 0.5321(3) 0.8260(6) 0.9322(3) 0.0327(11) Uani 0.50 1 d PDU A 1
C2 C 0.5529(4) 0.8802(9) 0.8184(9) 0.029(4) Uani 0.50 1 d PDU A 1
C3 C 0.5286(7) 0.8174(18) 0.7440(7) 0.031(4) Uani 0.50 1 d PDU A 1
H3A H 0.5350 0.8343 0.6771 0.037 Uiso 0.50 1 calc PR A 1
C4 C 0.4918(8) 0.7220(19) 0.7783(8) 0.027(4) Uani 0.50 1 d PDU A 1
C5 C 0.4900(6) 0.7171(14) 0.8779(9) 0.029(4) Uani 0.50 1 d PDU A 1
Cl2 Cl 0.3202(2) 0.4982(6) 1.0830(5) 0.0586(15) Uani 0.50 1 d PDU A 1
S2 S 0.38755(19) 0.6658(5) 0.9578(5) 0.0400(10) Uani 0.50 1 d PDU A 1
  
```

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$



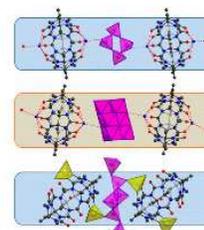
Sandwich-type Inorganic–Organic Hybrid Solids of Iso-polyvanadate Clusters and Decamethylcurcubit[5] uril

Li-Wei Han, Jing-Xiang Lin, Qi Yin, Bahar Karadeniz, Hong-Fang Li, Jian Lü, and Rong Cao

Publication Date (Web): January 15, 2016 (Article)

DOI: 10.1021/acs.cgd.5b01176

CRYSTAL
GROWTH
& DESIGN

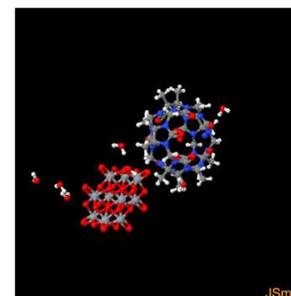


Accession Codes

CCDC: 1032491
CCDC: 1032492
CCDC: 1032493

NAJYU · tetrakis(ammonium) bis(decamethylcurcubit[5]uril) bis(μ-hydroxy)-octa(oxo)-octadeca(μ-oxido)-deca-vanadium octadecahydrate
Space Group: P-1, Cell: a 14.9979(6)Å b 16.5740(11)Å c 16.6113(6)Å, α 63.456(6)° β 71.772(4)° γ 66.779(5)°

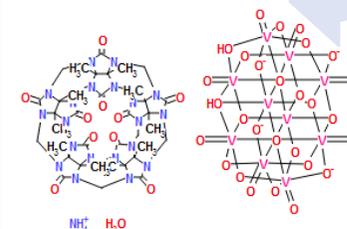
3D viewer



H Disorder Menu Open

Style Labels Packing Measure
Ball and stick No Labels None None

Chemical diagram



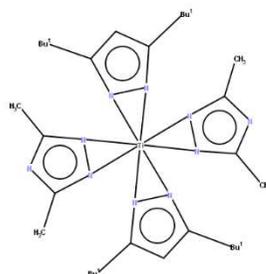
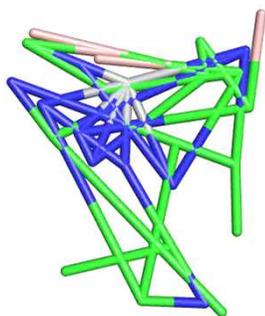
View group symbols key



Enrichment and curation

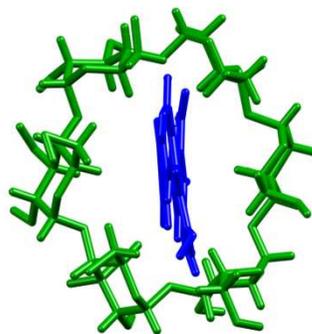
2D chemical diagrams:

Generated from existing curated CSD entries, CCDC diagram generation software, integrated third party software, manual curation



Classes:

ARACEB : (2)-(1,4-bis(2-(3,5-Dicarboxyphenyl)vinyl)benzene)-(alpha-cyclodextrin)-**rotaxane** dodecahydrate



Compound names:



Based on IUPAC conventions using ACDname and manual curation



Annual targeted database improvements

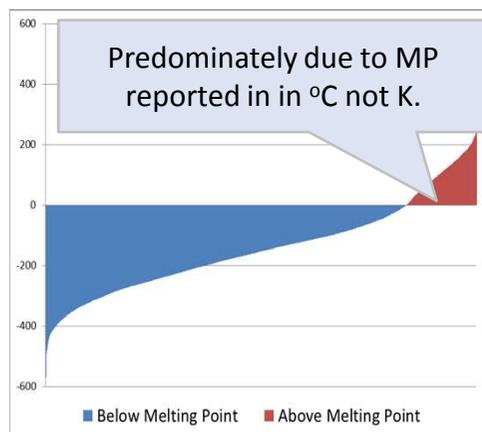


- Greater efficiency allows targeted improvements to further increase integrity, consistency, discoverability and usage of data
 - 50,000 existing entries improved in 2015
 - 74,000 existing entries improved in 2016
 - 62,00 existing entries improved in 2017

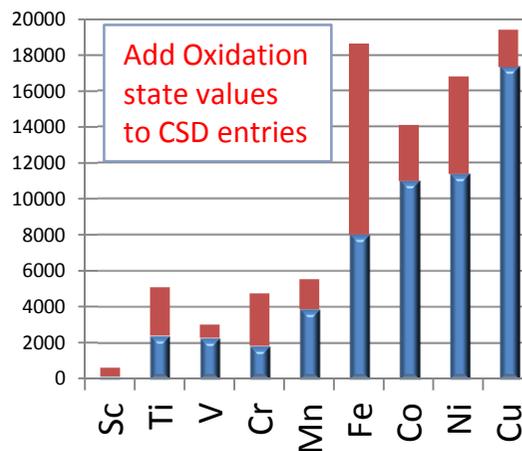


Targeted improvements to drug subset

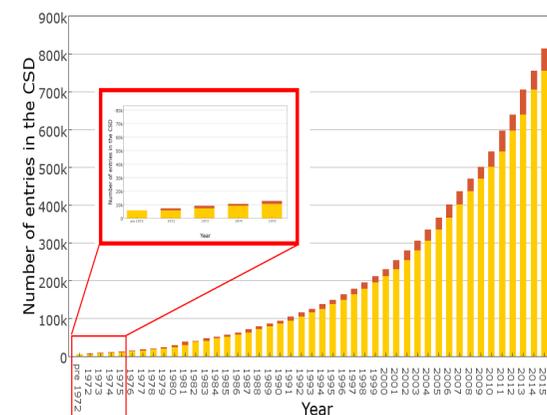
Study Temperature relative to Melting Point



Targeted improvements to metal-organics



Ensure consistency of early CSD entries





Accessing data

- Any

Your query was: Compound

Select all Download Selected

Entry search

Welcome to
Please use
containing a

CCDC identifier

Compound

Publication

- AHEPUY
- AMUBAM
- COKCEL
- COTZAN

CSD Entry: AHEPUY

Your query was: Compound name: paracetamol and the search returned more than 30 records. [Back to Search List](#) [New Search](#)

Results

Refcode	CCDC Number
<input checked="" type="checkbox"/> AHEPUY	200783
<input checked="" type="checkbox"/> AMUBAM	803736
<input checked="" type="checkbox"/> COKCEL	683518
<input checked="" type="checkbox"/> COTZAN	735836
<input checked="" type="checkbox"/> COTZAN01	735839
<input checked="" type="checkbox"/> COTZAN02	735844
<input checked="" type="checkbox"/> COTZAN03	735847
<input checked="" type="checkbox"/> COTZAN04	735850
<input checked="" type="checkbox"/> COTZAN05	735833
<input checked="" type="checkbox"/> CUQKAC	1412589
<input checked="" type="checkbox"/> CUZVEA	1438911
<input checked="" type="checkbox"/> HUMJEE	202320
<input checked="" type="checkbox"/> HXACAN	1178858
<input checked="" type="checkbox"/> HXACAN01	1178859
<input checked="" type="checkbox"/> HXACAN02	1178860
<input checked="" type="checkbox"/> HXACAN03	1178861
<input checked="" type="checkbox"/> HXACAN04	129925
<input checked="" type="checkbox"/> HXACAN05	1178862
<input checked="" type="checkbox"/> HXACAN06	1178863
<input checked="" type="checkbox"/> HXACAN07	135451
<input checked="" type="checkbox"/> HXACAN08	135452
<input checked="" type="checkbox"/> HXACAN09	144374

[Download](#)

AHEPUY : N-(4-Hydroxyphenyl)acetamide morpholine
Space Group: P-1, Cell: a 8.710(4)Å b 9.920(5)Å c 12.385(5)Å, α 102.35(3) $^\circ$ β 108.33(2) $^\circ$ γ 96.68(3) $^\circ$

3D viewer

Chemical diagram

Additional CCDC details

CCDC Number	200783
CCDC Citation	I.D.H.Oswald, W.D.S.Motherwell, S.Parsons, C.R.Pulham CCDC 200783: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/cc6qxwp
Synonyms	Paracetamol morpholine, Acetaminophen morpholine
Deposited on	16/12/2002

Download deposited CIF

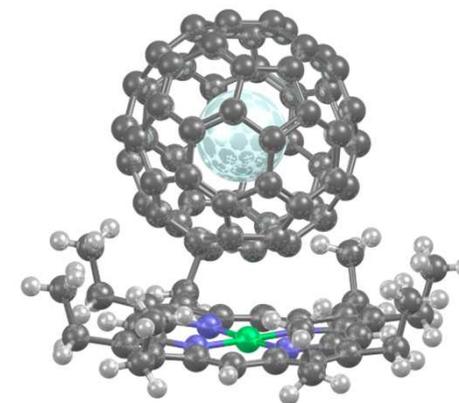
- Deposited CIF(s)
- Deposited CIF(s) without structure factor data
- Deposited file(s) with any available structure factor data and checkCIF reports included

Parsons, C.R.Pulham, Acta Crystallographica Section E: Structure Reports Online, 2002, 36802018111



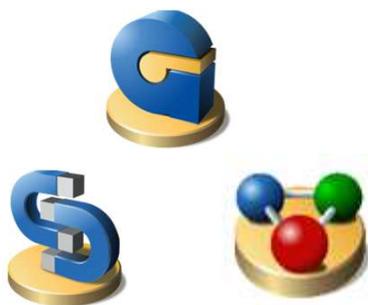
The CCDC Portfolio 2017-2018

- Alongside the CSD, we produce a suite of software enabling research across the whole breadth of structural science

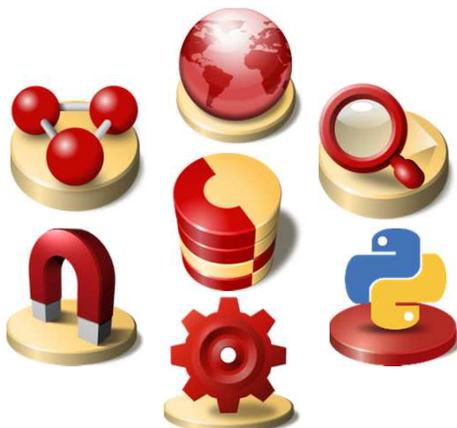


CSD-Enterprise

CSD-Discovery



CSD-System



CSD-Materials

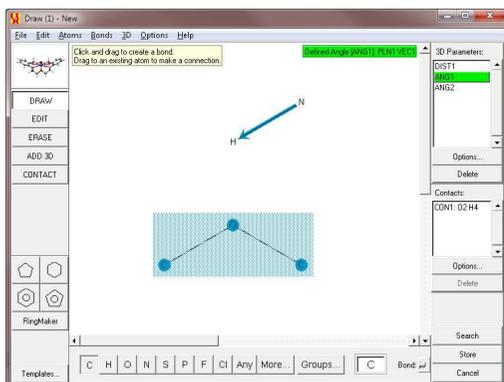




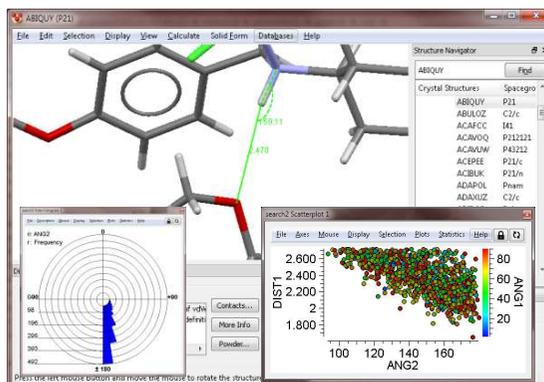
CSD-System



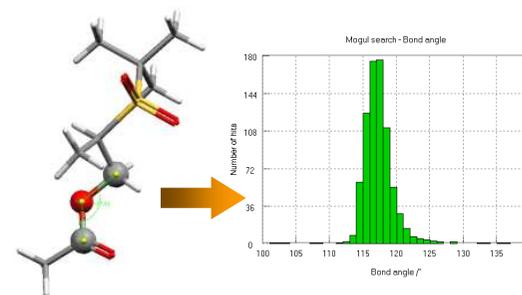
ConQuest: Advanced 3D searching



Mercury: Visualisation & data analysis



Mogul: Molecular geometry analysis



WebCSD: On-line portal to the CSD

Database Identifier	Deposition Number
ALH10	1459086
AMBLD	839979
	839980
	688078
	1024208
	1137604
	1052427
	1052428
	968930
	1421702



CSD Python API: Custom search and analysis

```
import sys
from ccdc import conformer
from ccdc import io

args = parser.parse_args()

mol_reader = io.MoleculeReader(args.inmolfn)
engine = conformer.GeometryAnalyser()

molecules = []
min_unusual_torsions = sys.maxint
for (idx, molecule) in enumerate(mol_reader):
    molecule.standardise_aromatic_bonds()
    molecule.standardise_delocalised_bonds()

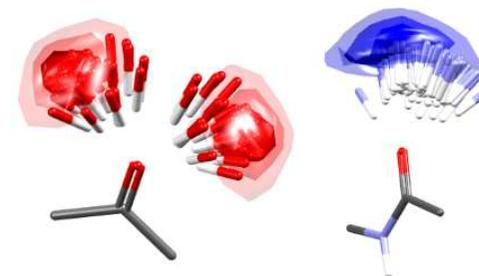
    # Do the analysis
    geometry_analysed_molecule = engine.analyse_molecule(molecule)

    # Count number of unusual torsions
    molecule.unusual_torsions = []
    for t in geometry_analysed_molecule.analysed_torsions:
        if t.unusual and t.enough_hits:
            num_unusual_torsions = len(molecule.unusual_torsions)
            molecule.num_unusual_torsions = num_unusual_torsions
            molecules.append(molecule)

    if num_unusual_torsions < min_unusual_torsions:
        min_unusual_torsions = num_unusual_torsions
```



IsoStar: Molecular interaction analysis





New WebCSD launched in 2017

- Unit cell search
- 2D structure search
 - exact
 - substructure
 - similarity
- Additional text/numeric searches
- Display of full CSD entry

The screenshot shows the 'Unit Cell Searching' section of the WebCSD interface. It features a 'Lattice centring' dropdown menu set to 'Primitive (P)'. Below this are input fields for lattice parameters: 'a', 'b', and 'c' (each with a value of 'e.g. 10.0') and angles 'α', 'β', and 'γ' (each with a value of 'e.g. 90.0' or 'e.g. 120.0'). A 'Sign In' button and 'Licensed to: CCDC Main Site' text are visible in the top right corner.

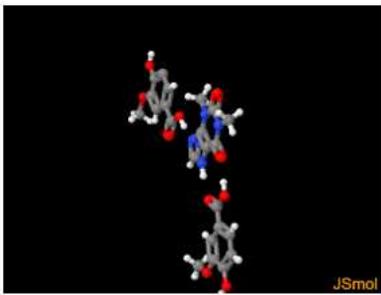
The screenshot shows the 'Chemical structure searching' section of the WebCSD interface. It includes a 'Go to WebCSD v1' checkbox and a drawing area with a toolbar and a vertical element list (C, N, O, S, H, F, Cl, Br, Y, I, Li, +, -). A chemical structure of a fused ring system with a methoxy group is drawn in the center. Below the drawing area, there are radio buttons for 'Match condition: Exact', 'Substructure', and 'Similarity', with 'Substructure' selected. A 'Search' button and a 'Clear' button are at the bottom left, and an 'Advanced' dropdown is at the bottom right. The 'dotmatics' logo is in the bottom right corner of the drawing area.



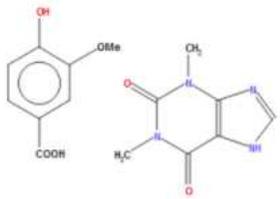
WebCSD - More information visible to users

ZOYBIA : bis(4-Hydroxy-3-methoxybenzoic acid) 1,3-dimethyl-3,7-dihydro-1H-purine-2,6-dione
Space Group: P21/c, Cell: a 11.2741(6)Å b 15.7712(10)Å c 13.3746(8)Å, α 90.00° β 90.572(2)° γ 90.00°

3D viewer



Chemical diagram



View group symbols key

Additional CCDC details

CCDC Number	1022107
CCDC Citation	Ayesha Jacobs, Françoise M. Amombo Noa CCDC 1022107: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/cc139166
Synonyms	bis(Vanillic acid) theophylline
Deposited on	30/08/2014

Associated publications

 Ayesha Jacobs, Françoise M. Amombo Noa, *CrystEngComm*, 2015, 17, 98, DOI: 10.1039/C4CE01795A

Chemical details

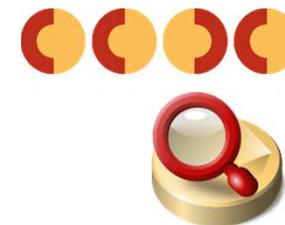
Formula	2(C ₈ H ₈ O ₄),C ₇ H ₈ N ₄ O ₂
---------	--

Crystal details

Space group	P21/c
Unit cell	a 11.2741(6)Å b 15.7712(10)Å c 13.3746(8)Å α 90.00° β 90.572(2)° γ 90.00°
Reduced cell	a 11.2741Å b 13.3746Å c 15.7712Å α 90.0000° β 90.0000° γ 90.5720°
Z, Z'	4.00, 1.00
Habit	rectangle
Recrystallisation solvent	water
Disorder	C7A,O4A and C9A,O5A disordered over two sites with occupancies 0.812:0.188.
Colour	colorless

Experimental details

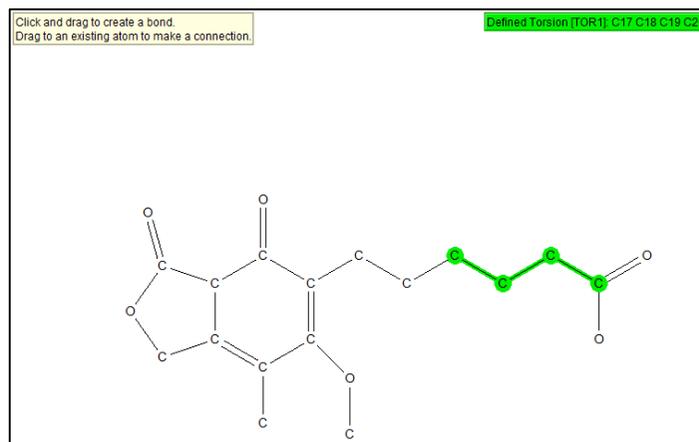
R-factor (%)	4.7
Temperature (K)	173.00
Density (CCDC)	1.44257



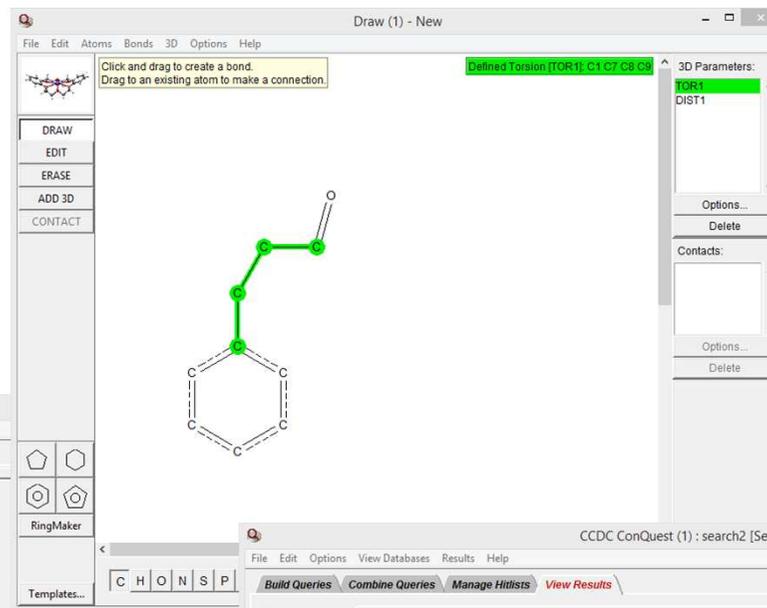
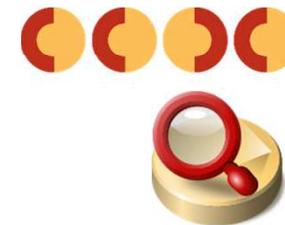
Questions

- *Is my molecule, or something like it, in the CSD?*
- *Are there any structures of aspirin in the CSD?*
- *What values are seen for this type of torsion angle?*
- *What is the typical intermolecular contact distance between these two groups?*

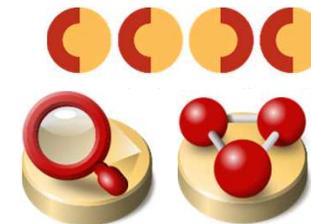
→ Use ConQuest (CSD-System)



ConQuest (CSD-System)



Hitlists can be exported in sdf, mol2, pdb, cif, plus additional formats.



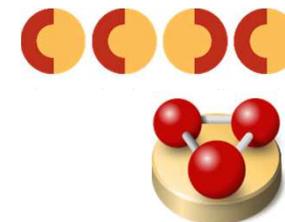
Question

- How can I analyse data produced by ConQuest searches?
→ Use Mercury (Data Analysis Module)

The screenshot displays the Mercury Data Analysis interface. On the left, a 3D ball-and-stick model of a chemical structure is shown with a green dashed line indicating a distance of 22.58. The central window shows a table of search results with columns for Identifier, NAME, Query, Fragment, DIST1, and TOR1. Below the table is a histogram of TOR1 values. On the right, a heatmap shows the relationship between DIST1 and TOR1 for various search results.

Identifier	NAME	Query	Fragment	DIST1	TOR1
search6(AHASOT)0	AHASOT	2	1	3.9780	-70.9520
search6(AHASOT)1	AHASOT	2	2	6.2310	178.7080
search6(AHATAG)2	AHATAG	2	1	5.8770	166.2020
search6(AHINI)3	AHINI	2	1	5.6440	-146.3230
search6(AJANAC)4	AJANAC	2	1	3.7030	-22.5800
search6(AJAWAL)5	AJAWAL	2	1	3.7560	-4.4100
search6(AJAWAL)6	AJAWAL	2	2	5.2330	113.8950
search6(AJAWAL)7	AJAWAL	2	3	4.5220	115.1910
search6(AJAWAL)8	AJAWAL	2	4	5.4350	-126.5050
search6(AJAWAL)9	AJAWAL	2	5	4.5220	-121.0880
search6(AJIHIM)10	AJIHIM	2	1	3.7490	6.2410
search6(AJIHIM)11	AJIHIM	2	2	4.2950	111.0160
search6(AJIHIM)12	AJIHIM	2	3	4.2950	-105.8090
search6(AJIKEL)13	AJIKEL	2	1	4.4270	-41.5040
search6(AJIKEL)14	AJIKEL	2	2	4.2950	-38.7090
search6(AJIKEL)15	AJIKEL	2	3	5.5700	122.0500

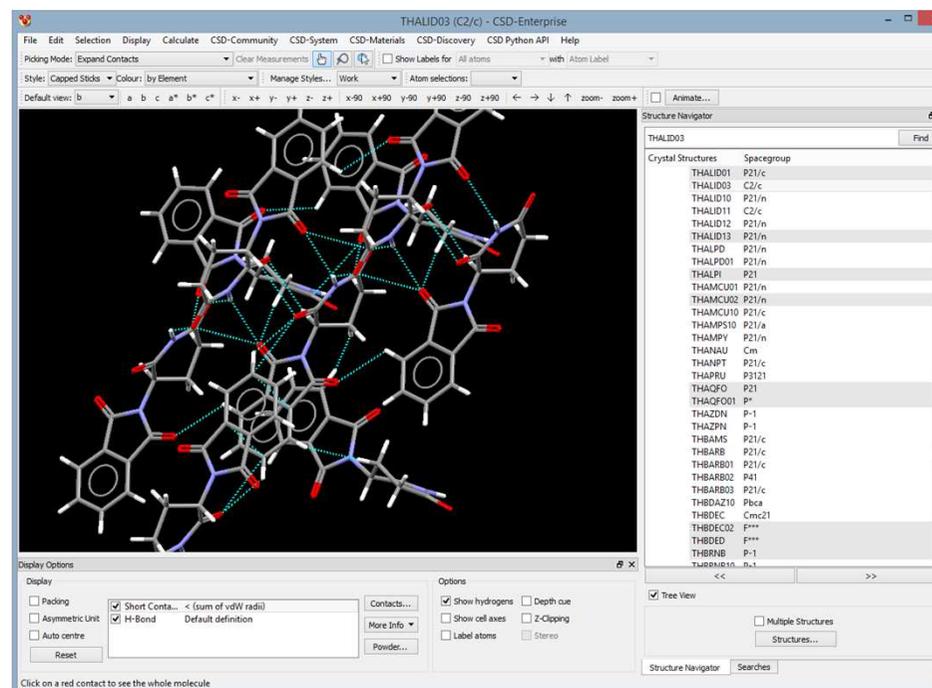
Can save the data to a csv file for analysis in other tools (Excel, Spotfire, StarDrop).

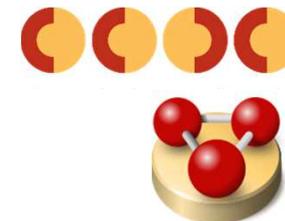


Questions

- *What is the 3D structure of my molecule?*
- *What are the crystal contacts of my molecule?*
- *What is the crystal packing of my molecule?*

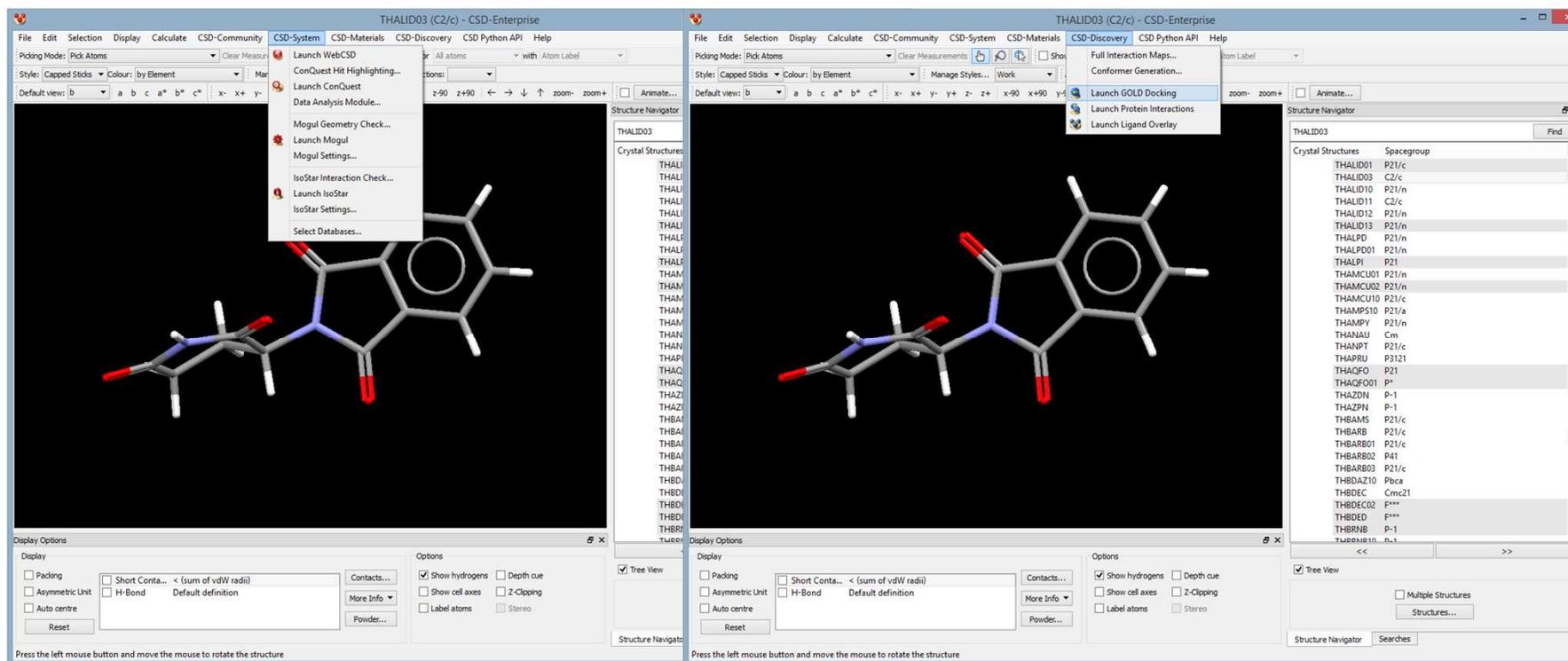
→ Use Mercury
(CSD-System)





More about Mercury

- Overall interface to small molecule visualisation & analysis
- Interface to other CCDC tools
- Can generate high-quality graphics (POV-Ray integration)



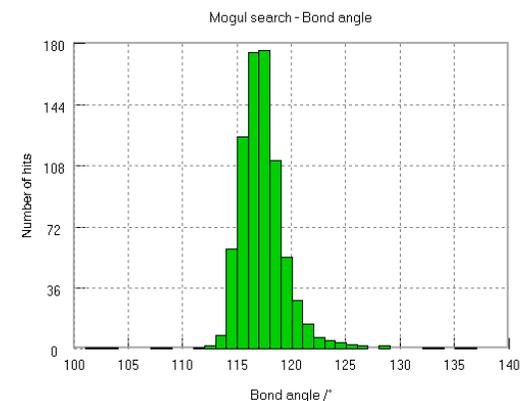
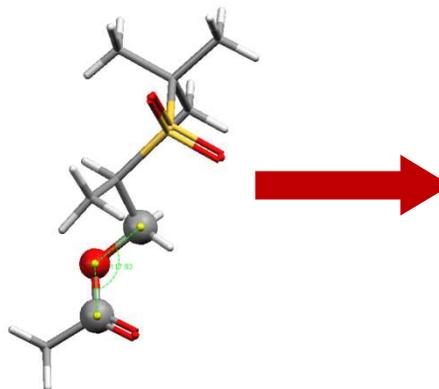


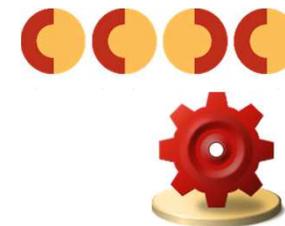
Questions

- *Is the geometry of my molecule typical of the 900,000+ structures in the CSD?*
- *Are the geometries of my modeled compounds realistic?*
- *What is the typical geometry of my molecular idea?*
- *Is the geometry of my docking pose reasonable?*

→ Use Mogul (CSD-System)

Pre-computed knowledge base of **intra**molecular geometry information in organics





Mogul (CSD-System)

Specify geometric parameter to search on, or validates complete geometry of molecule
Classifies each bond length, valence angle, torsion angle, and ring as usual or “unusual”

Blue = Not Unusual

Red = Unusual

Type	Molecule	Fragment	Classification	No. of hits	Query value	Mean	Std. dev.	z-score	x - mean	Minimum	Maximum	Median	d(min)	Local density
		C5 C6 N4 C7	Not unusual (enough hits)	560	178.175								0.000	0.939
		O4 C8 N5 C9	Not unusual (enough hits)	1411	-1.945								0.000	0.944
		C7 C8 N5 C9	Not unusual (enough hits)	1027	174.007								0.000	0.959
		O5 C10 N6 C11	Not unusual (enough hits)	1851	-7.290								0.000	0.985
		C9 C10 N6 C11	Not unusual (enough hits)	566	173.419								0.000	0.988
		O1 C2 C1 N1	Not unusual (enough hits)	1277	149.101								0.000	0.351
		N1 C1 C2 N2	Not unusual (enough hits)	1274	-32.546								0.000	0.356
		O1 C2 C1 C13	Not unusual (enough hits)	1288	-87.103								0.000	0.500
		C13 C1 C2 N2	Not unusual (enough hits)	1285	91.249								0.000	0.500
		C4 C3 N2 C2	Not unusual (enough hits)	1223	-95.442								0.000	0.116
		C14 C3 N2 C2	Not unusual (enough hits)	1775	141.096								0.000	0.109
		O2 C4 C3 N2	Not unusual (enough hits)	1277	-166.125								0.000	0.208
		N2 C3 C4 N3	Not unusual (enough hits)	1274	13.675								0.000	0.177
		C6 C5 N3 C4	Not unusual (enough hits)	862	105.689								0.000	0.114
		O3 C6 C5 N3	Not unusual (enough hits)	842	2.909								0.000	0.154
		N4 C6 C5 N3	Not unusual (enough hits)	832	179.131								0.000	0.136
		C8 C7 N4 C6	Not unusual (enough hits)	1223	53.790								0.000	0.249
		C15 C7 N4 C6	Not unusual (enough hits)	1775	-73.685								0.000	0.083
		O4 C8 C7 N4	Not unusual (enough hits)	1277	-146.303								0.000	0.366
		N4 C7 C8 N5	Not unusual (enough hits)	1274	37.667								0.000	0.364
		C10 C9 N5 C8	Not unusual (enough hits)	862	90.867								0.000	0.280
		O5 C10 C9 N5	Not unusual (enough hits)	842	174.340								0.000	0.287
		N6 C10 C9 N5	Not unusual (enough hits)	832	-6.319								0.000	0.276
		C12 C11 N6 C10	Not unusual (enough hits)	862	-118.969								0.000	0.080
		O6 C12 N1 C1	Not unusual (enough hits)	1574	4.555								0.000	0.985
		C11 C12 N1 C1	Not unusual (enough hits)	560	-174.111								0.000	0.986
		O6 C12 C11 N6	Not unusual (enough hits)	842	14.101								0.000	0.227
		N1 C12 C11 N6	Not unusual (enough hits)	832	-167.244								0.000	0.243
		O2 C4 C3 C14	Unusual (enough hits)	1288	-43.029								0.000	0.046
		C14 C3 C4 N3	Unusual (enough hits)	1285	136.772								0.000	0.044
		O4 C8 C7 C15	Unusual (enough hits)	1288	-18.794								0.000	0.028
		C15 C7 C8 N5	Unusual (enough hits)	1285	165.177								0.000	0.032
ring	AAGAGG10_1	N1 C1 C2 N2 C3 C4 N3 C5 C6 N4 C7 C8 N5 C9 C10 N6 C11 C12	Not unusual (enough hits)	15						0.000	68.884			0.133

Double-click unusual torsion to examine in detail



Mogul (CSD-System)

CCDC Mogul 1.7.3: AAGAGG10 (P212121) - CSD-Enterprise

File Searches Databases Help

Build query Results and analysis View structures

Results Navigator

All hits: 1239
Accepted hits: 1239
R-factor: Any Heaviest Element: Any
Exclude: None

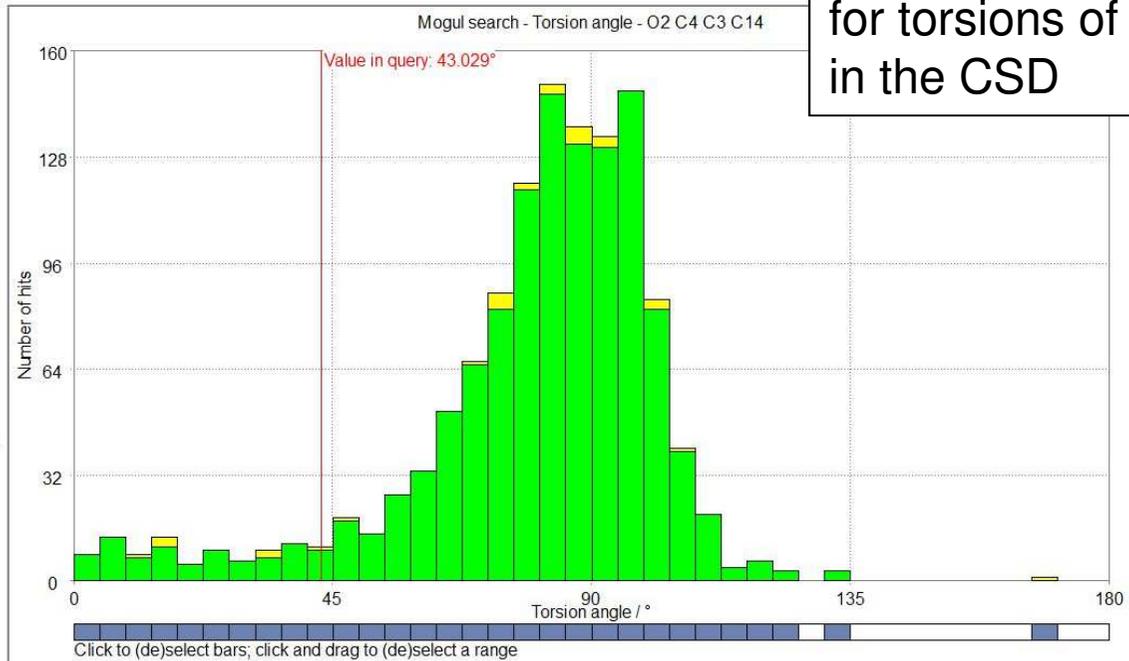
Relevance	Number	Contribution
> <input checked="" type="checkbox"/> 1.00	1239	100.0%

View diagrams...

More hits...

Statistics

Total : 1239
Selected : 1239
| d(min) | : 0.000°



Distributions of angles for torsions of this type in the CSD

All fragments...

View query...

Histogram display

Displayed hits: 1239

Selected hits: 1239

Select all hits in histogram

Deselect all hits in histogram

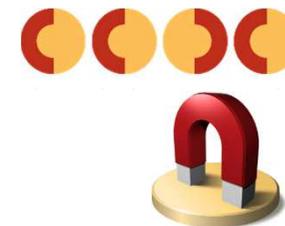
Data libraries

CSD 5.39

CSD Nov17 update

Filters...

Cluster

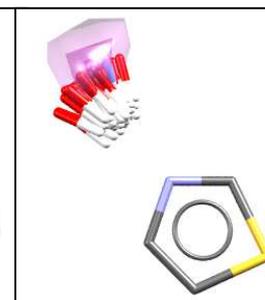
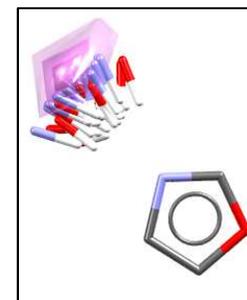
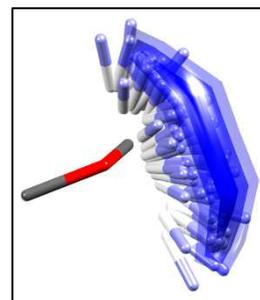


Questions

- How do **particular** groups typically interact in structures in the CSD? (e.g. is the N or the O of oxazole more likely to H-bond to an OH group?)
- How do **particular** ligand groups interact with particular protein residues? (e.g. how can we bind to a tryptophan?)
- What particular group can be a bioisostere for another particular group? (e.g. do thiazole and oxazole form similar H-bond interactions?)

→ Use IsoStar (CSD-System)

Pre-computed knowledge base of intermolecular interactions





IsoStar (CSD-System)

Client-Server system: <http://isostar.ccdc.cam.ac.uk>

Select Central Group (Phenyl)

Select Contact Group (R-NH₃⁺)

Select CSD or PDB data



View Scatter plot or Contour plot

- In both cases, adjusted for symmetry

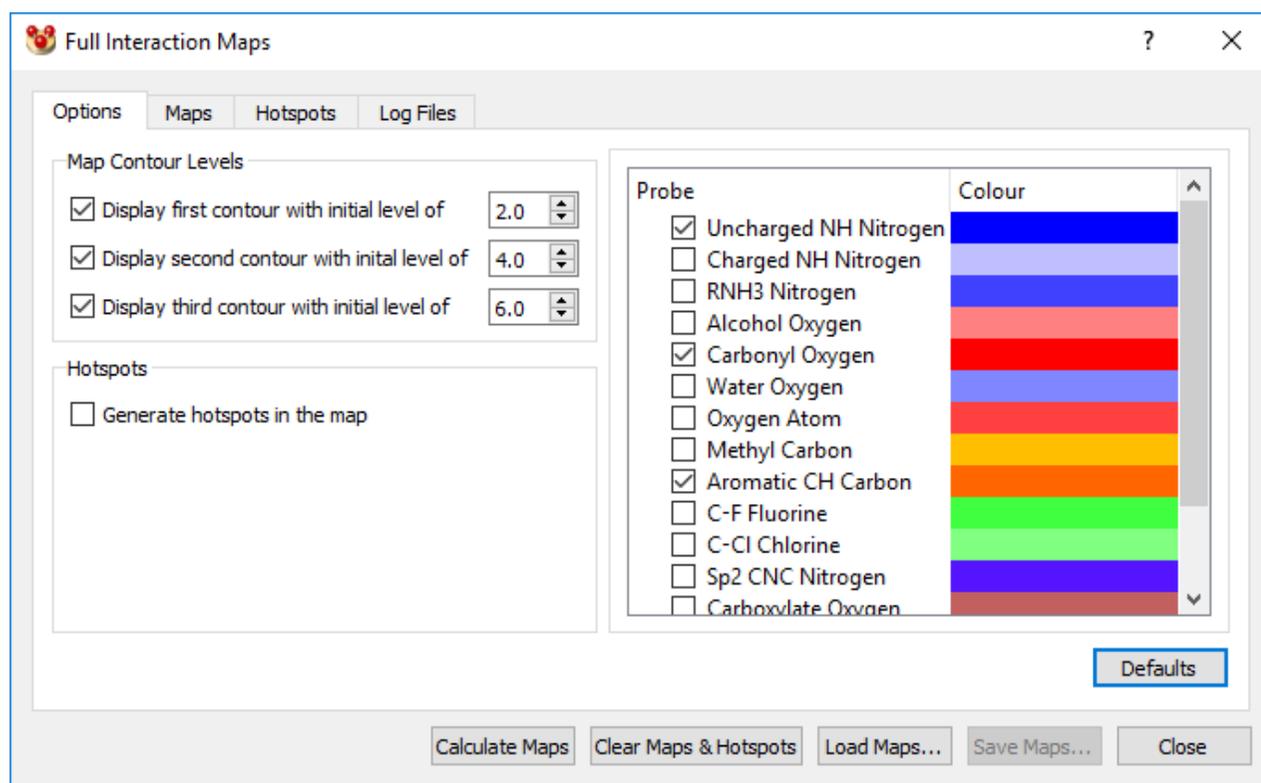
Also: 1,550 theoretical energy minima (IMPT); IsoGen to create custom plots



Question

- What are the common interactions around my molecule taken as a whole?

Full Interaction Maps (CSD-Discovery/Materials)



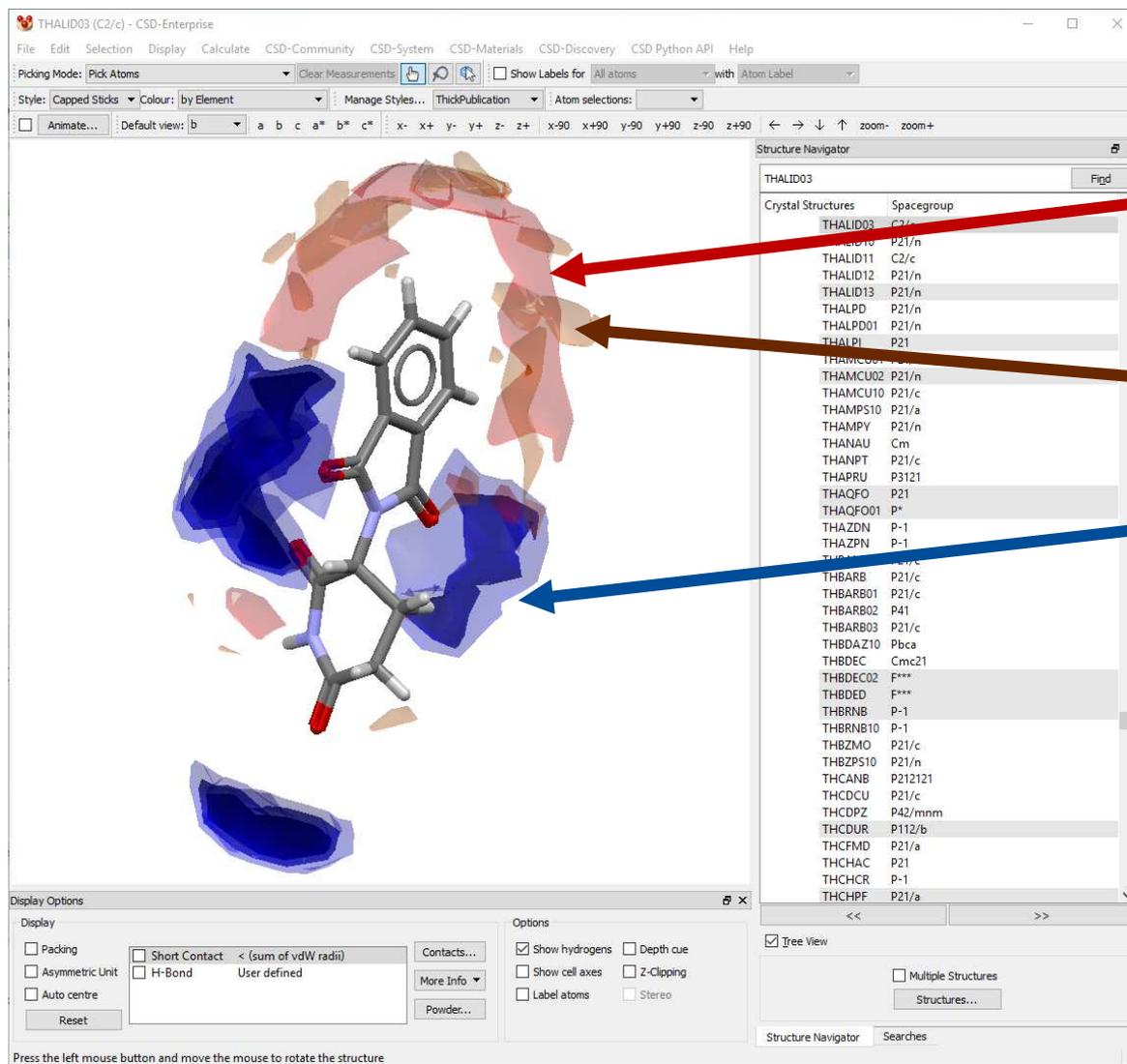
Uses CSD data from IsoStar Accessed in Mercury.

For currently loaded entry:

- Select Full Interaction Maps in CSD-Discovery or CSD-Materials menu.
- Retain defaults
- Click Calculate Maps.



Full Interaction Maps (CSD-Discovery/Materials)



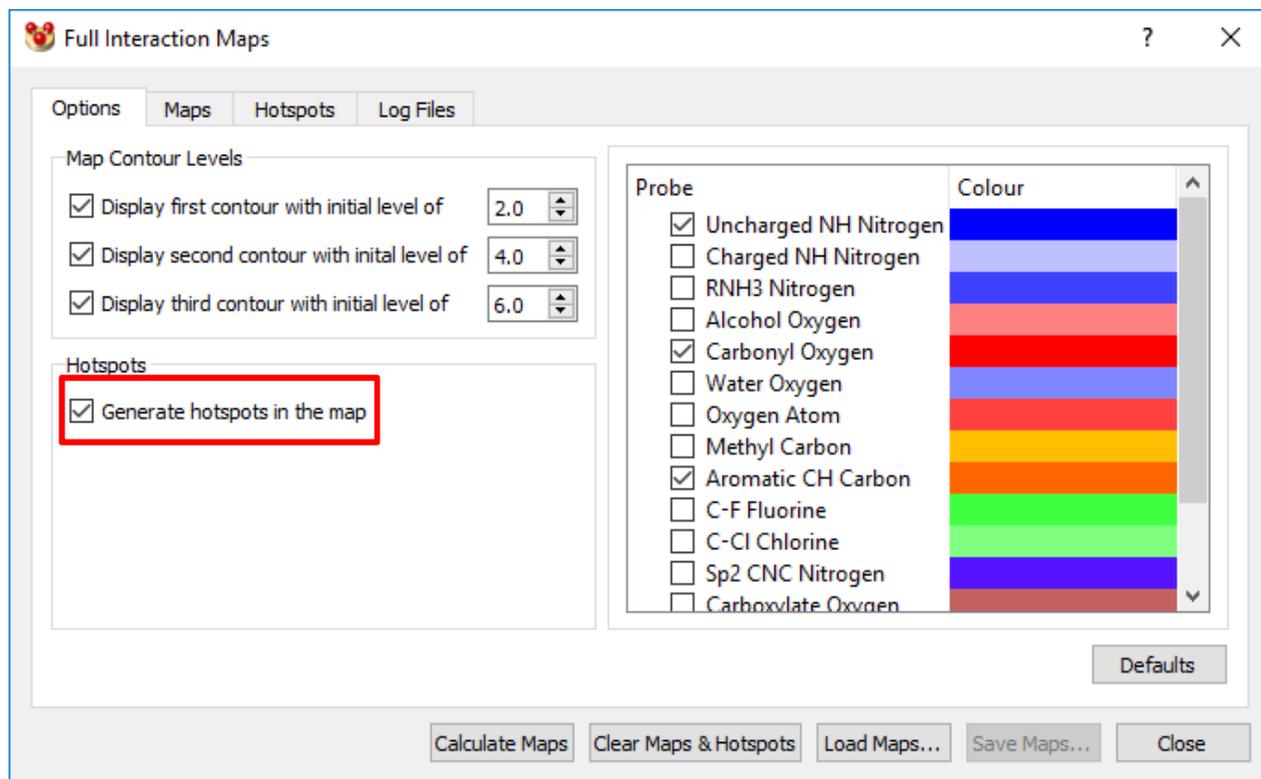
Red = frequently observed H-bonding acceptors

Brown = frequently observed hydrophobic (bit hard to see)

Blue = frequently observed H-bonding donors



Full Interaction Maps (CSD-Discovery/Materials)



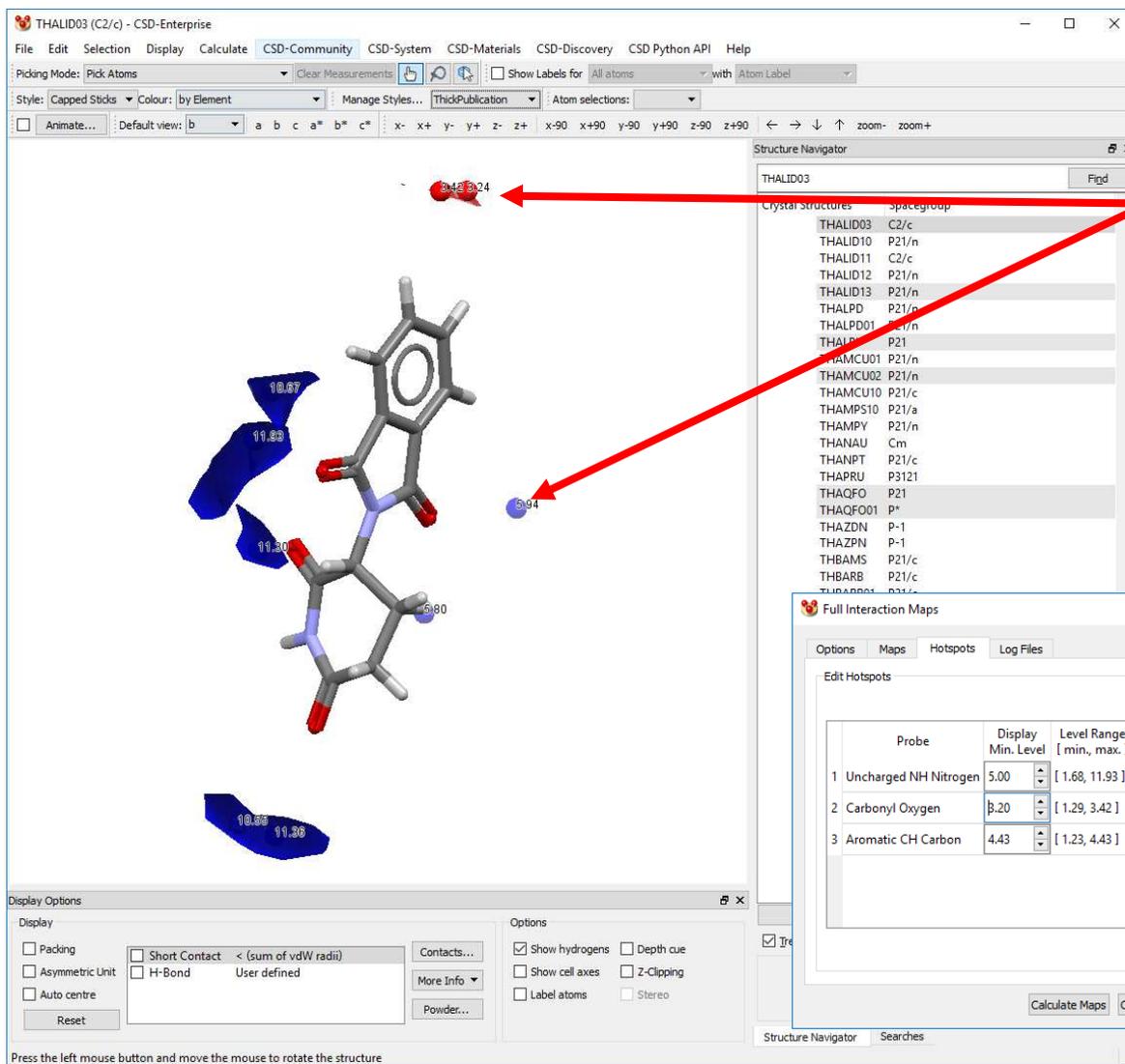
Uses CSD data from IsoStar.

For currently loaded entry:

- Select Full Interaction Maps in CSD-Discovery or CSD-Materials menu.
- Turn on “Generate hotspots”.
- Click Calculate Maps.



Full Interaction Maps (CSD-Discovery/Materials)



Points are:

- Peaks in contours
- Fold over random (e.g., 10 = 10 times more likely than random)

Full Interaction Maps

Options Maps Hotspots Log Files

Edit Hotspots

Probe	Display Min. Level	Level Range [min, max.]	Color	Visible	Label	Use Opacity
1 Uncharged NH Nitrogen	5.00	[1.68, 11.93]	Blue	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2 Carbonyl Oxygen	3.20	[1.29, 3.42]	Red	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3 Aromatic CH Carbon	4.43	[1.23, 4.43]	Custom...	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

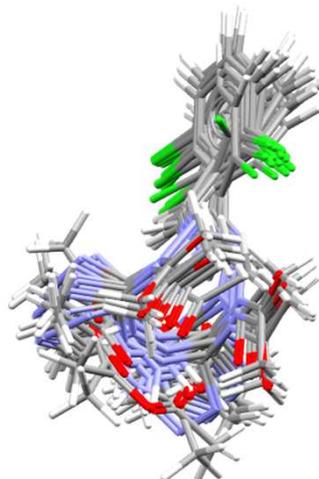
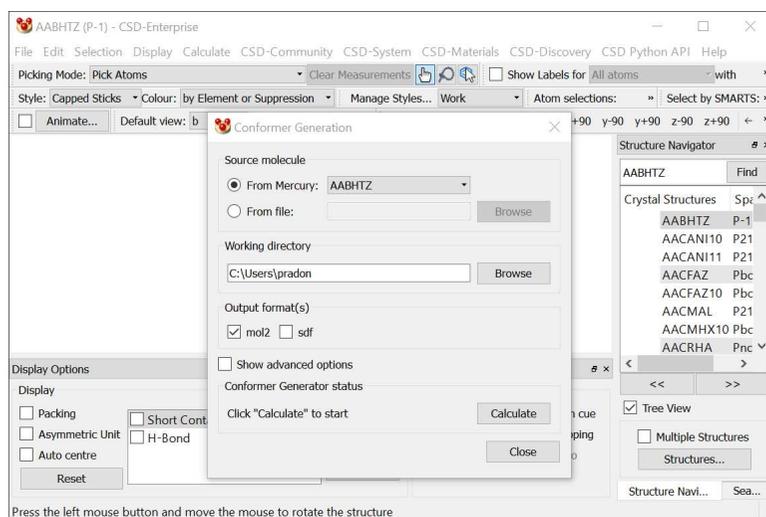
Interactions



Question

- *What are the most likely diverse conformations of my molecule based on the CSD?*

CSD-Conformer Generator (CSD-Discovery/Materials)



Uses CSD data from Mogul Accessed in Mercury or Hermes, or through the command line, or through the CSD Python API.

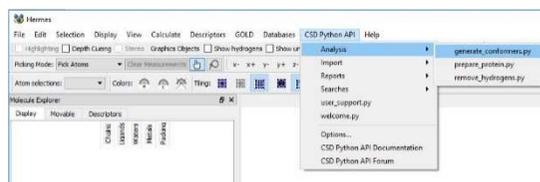
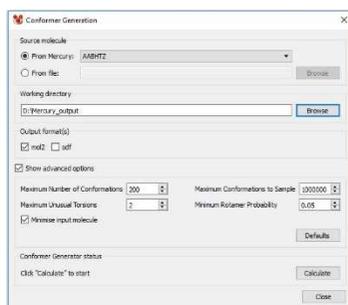
In Mercury, for currently loaded entry:

- Select Conformer Generation in CSD-Discovery or CSD-Materials menu.
- Retain defaults
- Click Calculate.



CSD-Conformer Generator

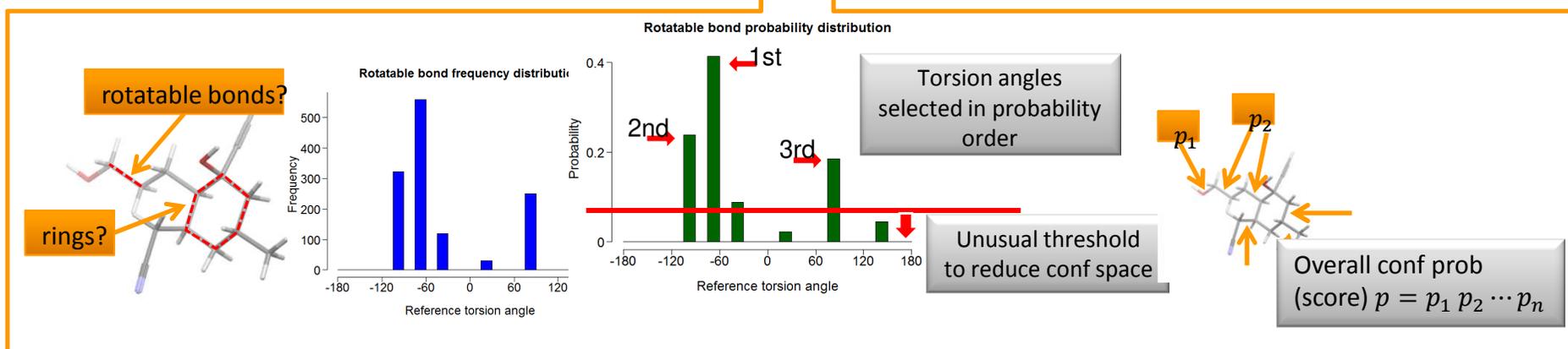
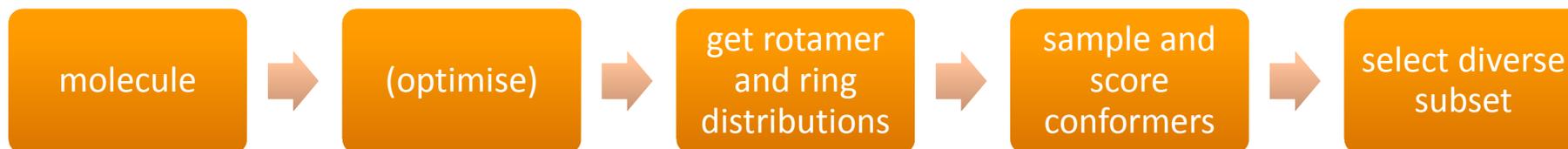
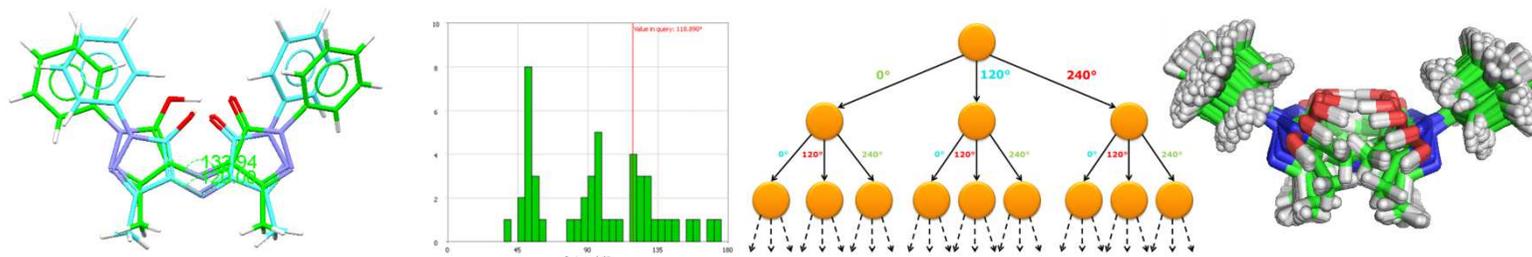
- Part of CSD-Discovery, CSD-Materials and CSD-Enterprise suites



- Program based entirely on Mogul data for conformer generation
- New version: improved handling of flexible ring systems
 - using Mogul templates for isolated, fused, spiro-linked and bridged rings
 - if no template available, then generated on the fly with cyclic bond rotamer distributions
- Fast & chemically plausible conformations
- Initial optimisation step also useful to prepare compound libraries before virtual screening



CSD-Conformer Generator: Methodology

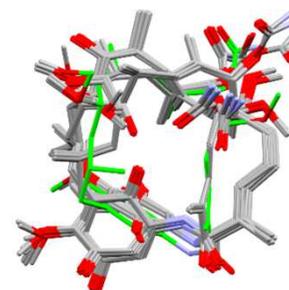
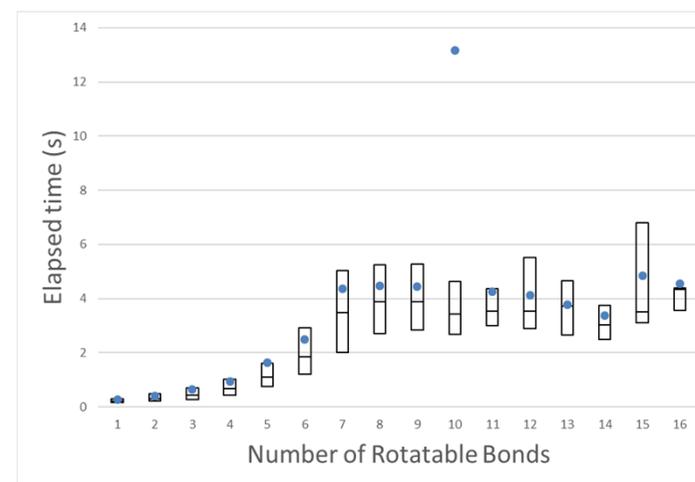
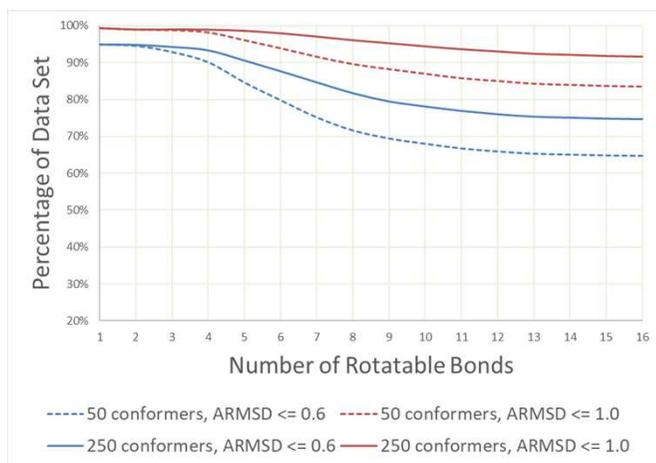




CSD-Conformer Generator: Latest validation

- Performance evaluated against the Platinum² diverse data set → get fast & chemically plausible conformations

	n	Top n conformers		
		1	50	250
least ARMSD < 0.5 Å (%)		18	57	66
least ARMSD < 1.0 Å (%)		39	84	92
mean ARMSD			0.59	0.48



Geldanamycin, 19 atoms in ring (PDB 3C11): top 50 conformers, least ARMSD = 0.95 Å



CSD Python API: Contents

- CSD Python API components:
 - Reading and writing of molecules, crystals and entries
 - Entry, crystal and molecule functions (incl. PXRD output & comparison)
 - Search capabilities (substructure, similarity, text/numeric, reduced cell)
 - Molecular geometry analysis
 - Intermolecular interaction analysis
 - Geometric, molecular & crystallographic descriptor calculation
 - Crystal packing similarity calculation
 - 2D diagram generation
 - 3D conformer generation
 - ... much more to come, exposing CSD-Materials and CSD-Discovery features



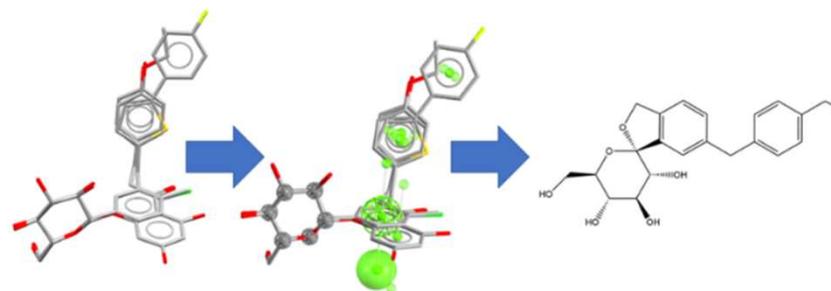


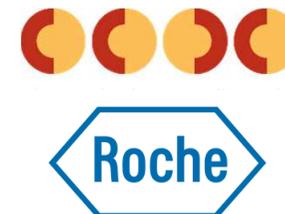
Questions

- *What other molecules exhibit a similar short-contact or hydrogen-bond pattern?* (easy non-bonded contact search in the CSD: how can we use pharmacophore features to search for a given small molecule-small molecule interaction pattern?)
- *What other scaffolds would fit in my molecule?* (easy scaffold hopping in the CSD: how can we use exit vectors to retrieve diverse ligand topologies that could be used as scaffold?)

→ Use CSD-CrossMiner (CSD-Discovery)

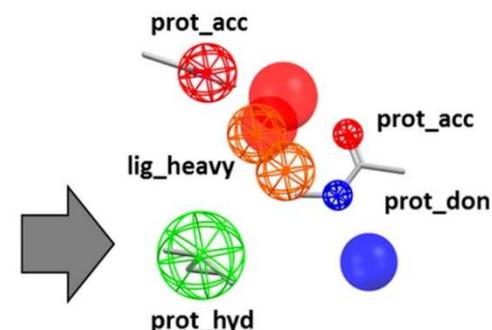
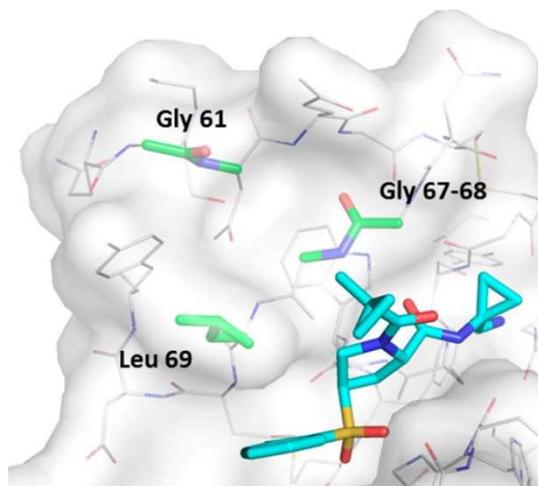
Search structural databases (i.e. the CSD and/or the PDB), concurrently and interactively, in terms of pharmacophore queries





CSD-CrossMiner overview

- Pharmacophore-based searches of structural databases (CSD & PDB, simultaneously)
- Interactive tool: modify a hypothesis/results on the fly
- Applications in drug design:
 - Which structural motifs bind in a similar environment?
 - Which ligand motifs have similar protein interaction patterns?

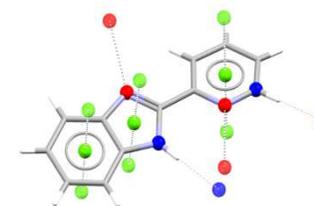
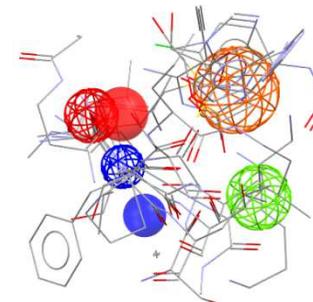
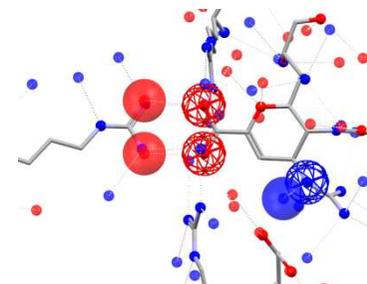


feature name	tolerance	radius	show in reference	show in pharmacophore
ring_planar_projected			<input type="checkbox"/>	<input checked="" type="checkbox"/>
ring_planar_projected...				
B	1.00			
V	1.00			
ring_planar_project...				
B	1.00			
V	1.00			
ring_planar_project...				



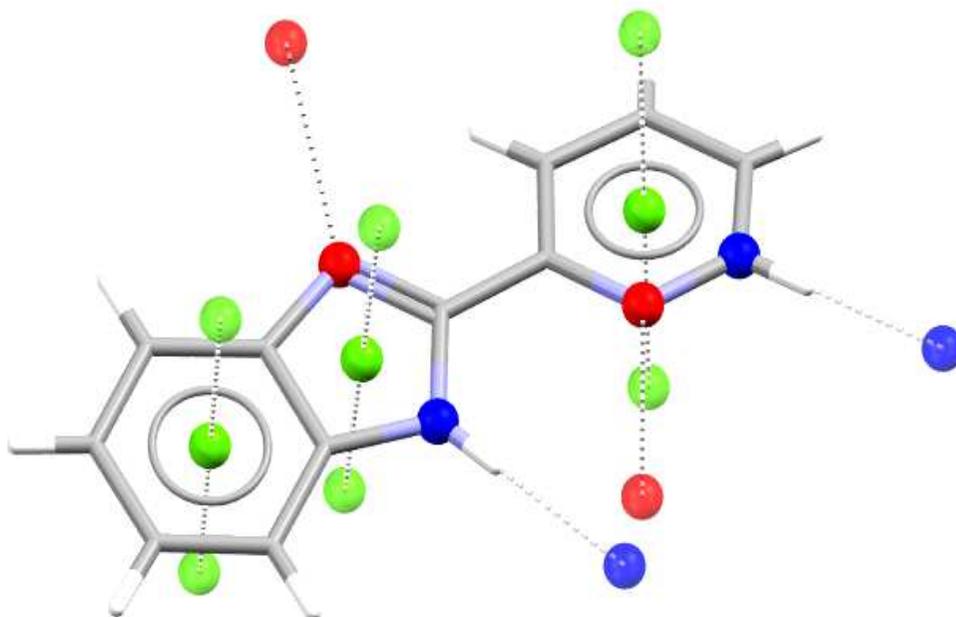
Our motivation for using pharmacophores

- Intuitive as the functional groups can be specified relative to a 3D structure
 - Sphere positions and radii can easily be refined
- “What you see is what you get”
 - Hits will be overlaid onto the query and can be analysed in the same context
- Concept well-understood by medicinal chemists
- Flexible feature definitions
 - User can control the level of chemical abstraction





Feature assignment



planar ring

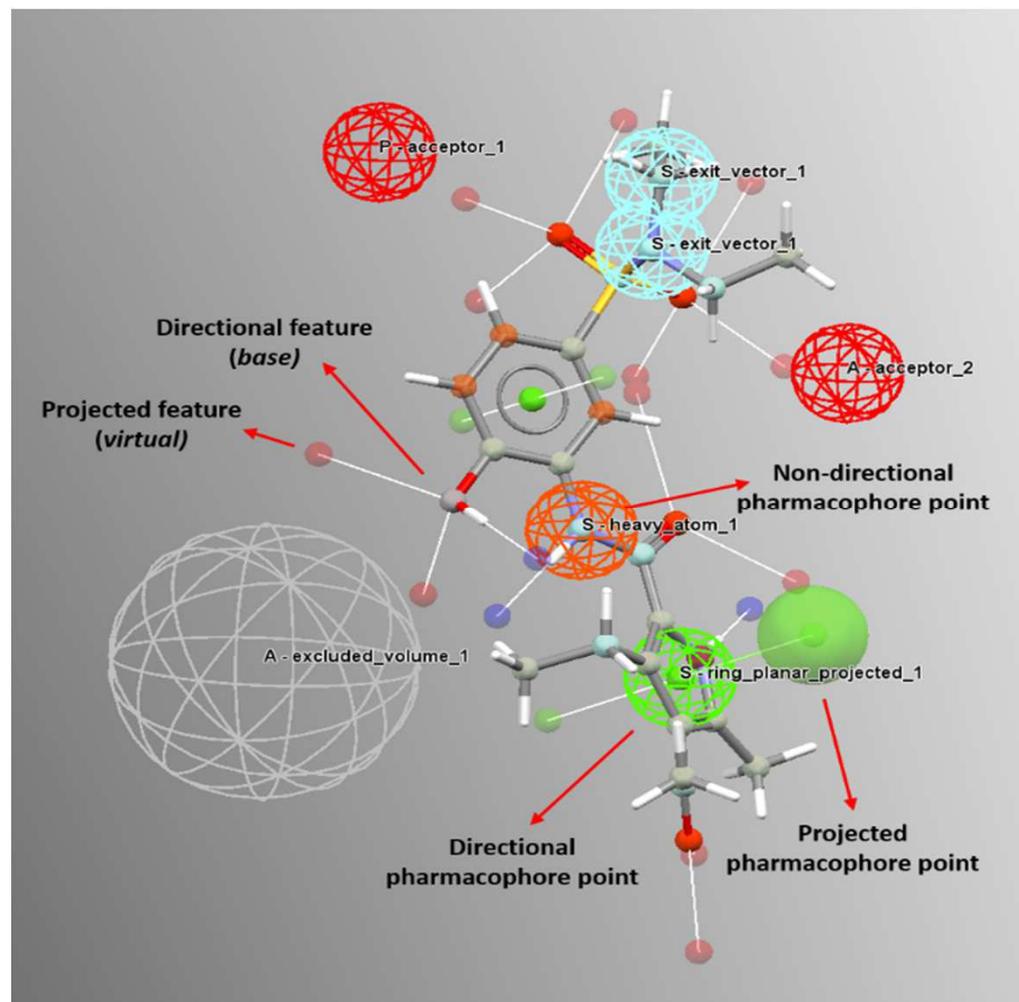
donor

acceptor

- Features are derived from the substructures by applying a point generation rules.
- Features are based on a series of modifiable SMARTS patterns.
- Each entry in a structural database is indexed with a set of feature definitions which are used to perform the actual 3D search.

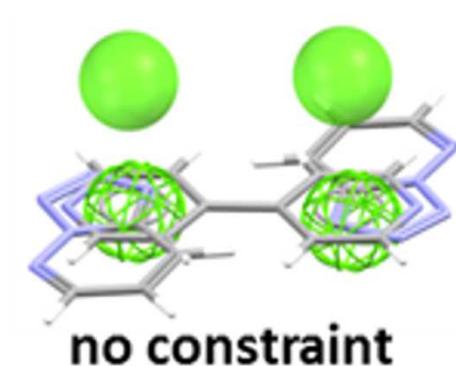
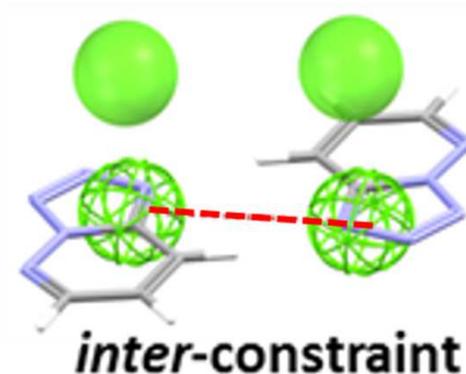
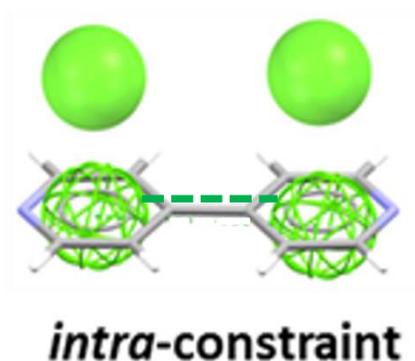
Pharmacophores in CSD-CrossMiner

- Molecular structures are annotated with *features*
- Pharmacophore query is based on *tolerance spheres*
 - Sphere radii reflect uncertainty in the position of the features
- Pharmacophore points can be set to be protein (**P**), small molecule (**S**) or any molecule (**A**)





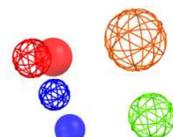
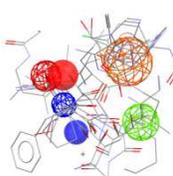
Constraints



- ***Intra-constraint*** → pharmacophore points located in the same structure
- ***Inter-constraint*** → pharmacophore points located in different structures
- ***No constraint*** → pharmacophore points located either in the same or in different structures



Pharmacophore search

-  1. generate initial query
 -  2. run search
 -  3. analyse results
 -  4. modify query
-  can be time-consuming!

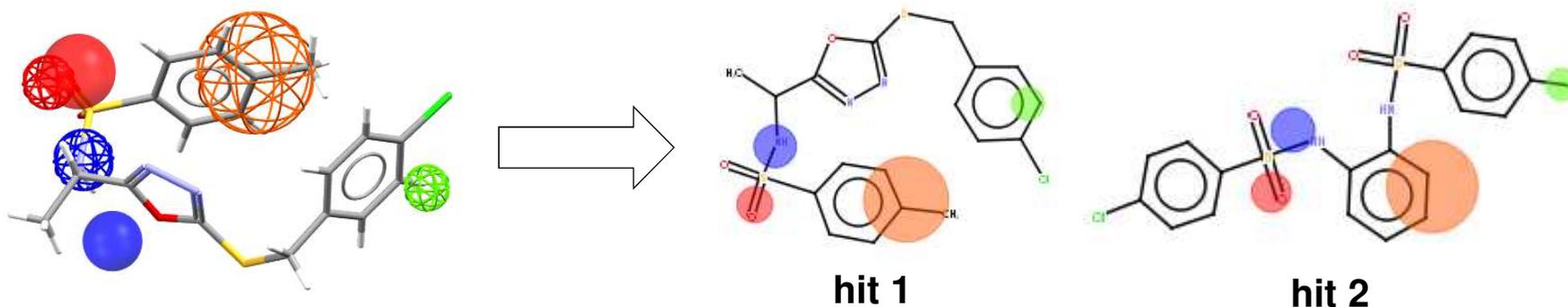
Wouldn't it be nice to *interactively* modify the query and analyse the results?





Hit processing

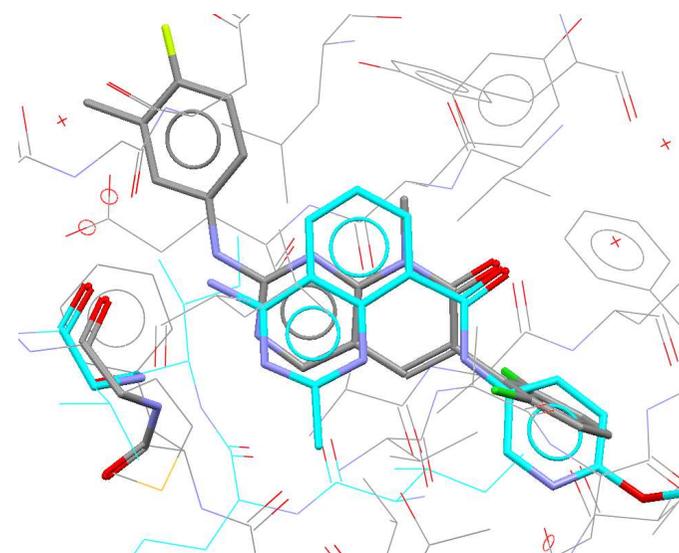
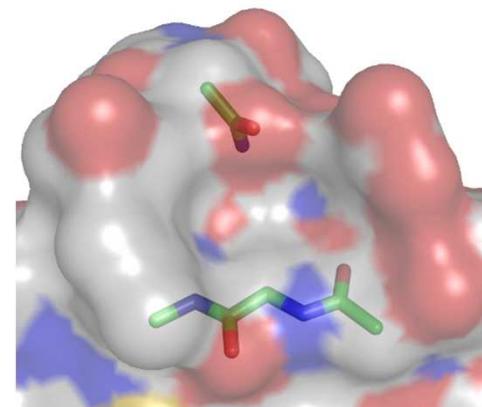
- Rank hits according to Kabsch¹ overlay *rmsd*
 - Overlay feature point to centre of the query sphere
- Hits are clustered according to molecular similarity
 - Only display diverse set of hits with respect to *Tanimoto* threshold (ligand, protein, or ligand+protein)
- 2D diagrams
 - 3D pharmacophore is projected into a 2D representation





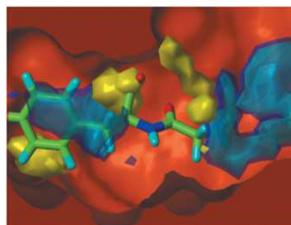
Applications of CSD-CrossMiner

- Determine common protein binding sites in PDB structures
- Determine structural motifs that bind in similar environments
- Inform cross-pharmacology between protein targets
- Find new ideas:
 - Design novel motifs that molecularly mimic established ligands
 - Scaffold-hopping: retrieve a diversity of ligand topologies that can be used as scaffolds

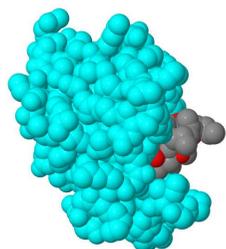




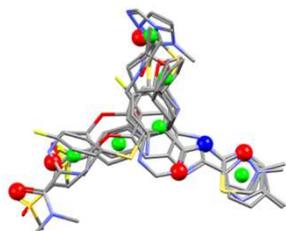
CSD-Discovery: Other applications not covered here



SuperStar: Knowledge-based prediction of intermolecular interactions in a protein binding site based on data from CSD or PDB crystal structures



GOLD Suite: Protein-ligand docking - virtual screening, lead optimisation and binding mode prediction, with the **Hermes** protein visualiser

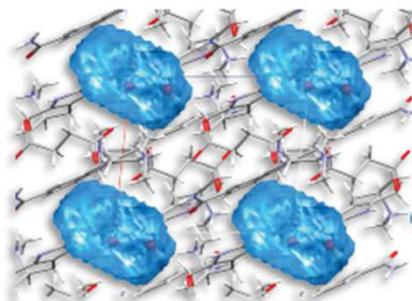


Ligand-Based Virtual Screening Workflow: identify plausible common binding modes of structurally diverse ligands using the **Ligand Overlay Program**, and virtually screen molecules against such a binding mode hypothesis using the **Field-Based Ligand Screener**



CSD-Materials: Other applications not covered here

Informatics-based solutions that aid in the understanding and prediction of solid form stability and properties



- Interpret crystal packing and compare with CSD data using powerful **Packing Feature**, **Similarity** and **Motif searches**, and **Hydrogen Bond propensity Analysis**
- Understand the effects of hydration on your lattice with the **Hydrate Analyser**
- Explore the structures of potential co-crystals using the **Molecular Complementarity Tool**
- Use CSD data to help solve crystal structures from powder diffraction data using **DASH**



CSD-Enterprise: Summary of what was covered here

- **ConQuest**: CSD search & geometric data generation
- **Mercury**: Visualisation, analysis, overall small molecule interface
- **Mogul**: Geometric analysis
- **IsoStar**: Precomputed functional group-to- functional group interactions
- **Full Interaction Maps**: Interactive whole molecule interactions
- **CSD-Conformer Generator**: Quick generation of chemically plausible conformations
- **CSD Python API**: Scripting, handling of many molecules, integration, conformer generator and much more
- **CSD-CrossMiner**: interactive and simultaneous search of CSD & PDB in terms of pharmacophore queries