CCDC

advancing structural science

What's Up

Customer Update Webinar

22nd July 2021

CCDC

advancing structural science

Today’s presenters

Natalie Johnson

Data Integrity Research Scientist

Oliver Anderson

Sales Operations Coordinator

Ilenia Giangreco

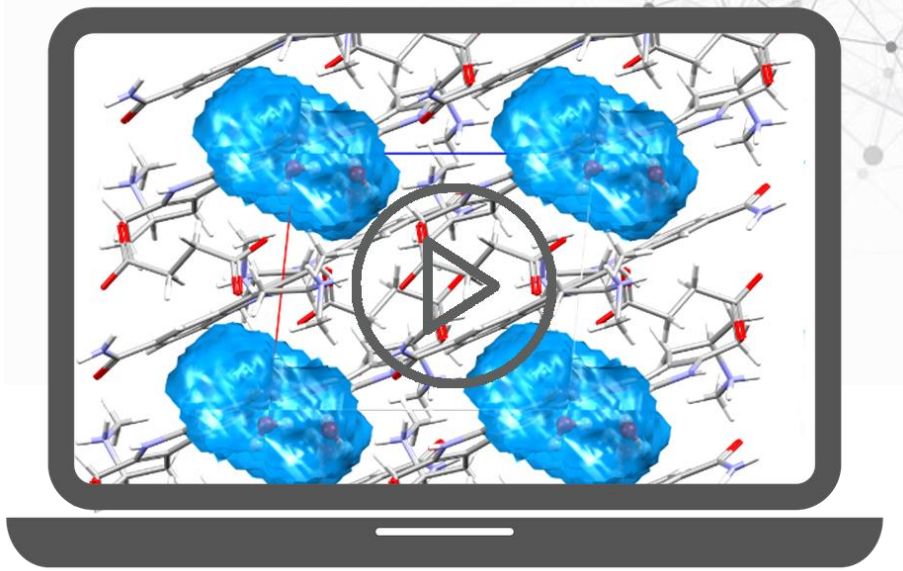
Discovery Science Team Leader

CCDC

Overview

In this webinar we will discuss:

- Latest updates and news
- Effective searching with WebCSD and Access Structures
- Docking with GOLD in the API – including new updates
- Q&A: the floor is yours



CCDC

Latest updates

- 2021.1 CSD Release is available – this release includes several new features in CSD-Discovery to improve ligand preparation and docking.
  - Find out how more at https://www.ccdc.cam.ac.uk/solutions/whats-new/

- CSP Blind Test: we released the 2D chemical structures of the target compound XXXII
  - Learn more at https://www.ccdc.cam.ac.uk/Community/blog/CSP-blind-test-reveal-target-XXXII/

CCDC

Latest news from CCDC

- CSD University launched: it's an online space to learn about the CSD and CCDC software.
  - First module: Visualisation structural chemistry data with Mercury
  - Start learning here https://www.ccdc.cam.ac.uk/Community/educationalresources/CSDU/

CCDC

Upcoming events from CCDC

- 28th July **CfC Workshop** – only for CfC members

If you want to know more about how to became a member email us at hello@ccdc.cam.ac.uk

- 30th July - **Structural Science Awakens: 71st Annual Meeting of the ACA**
- 12th – 22nd August: **IUCr Congress 2021**
- 22nd – 26th August: **ACS Symposium-Resilience in Chemistry**
- 29th August – 2nd September: **EFMC-ISMC International Symposium on Medical Chemistry 2021**
- 7th – 8th September: **Materials Science Meeting**

Register at www.ccdc.cam.ac.uk/News/Events

CCDC

WebCSD

Effective searching and Access Structures

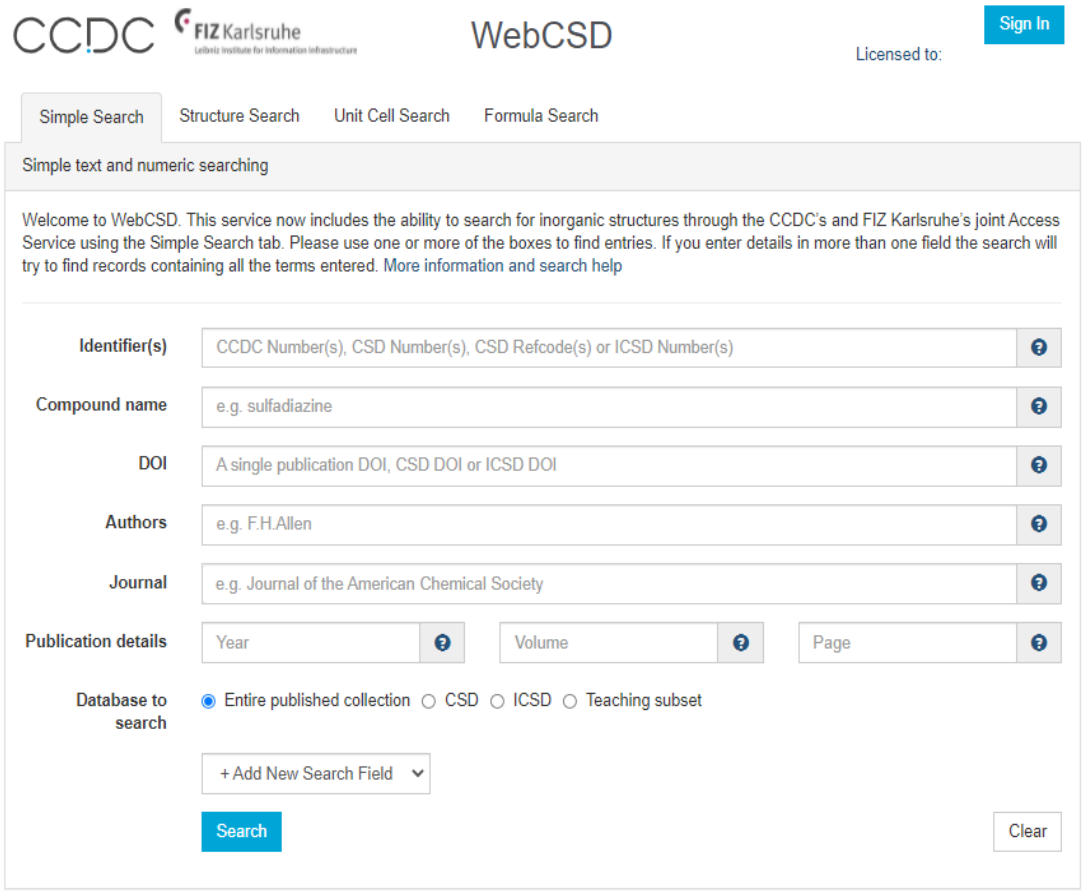
Dr Natalie Johnson

Data Integrity Research Scientist

CCDC

WebCSD

- An online portal to access the Cambridge Structural Database
- Accessed through any standard internet browser – no local installation of software required
- Offers a subset of tools for searching, browsing and viewing crystal structures



CCDC

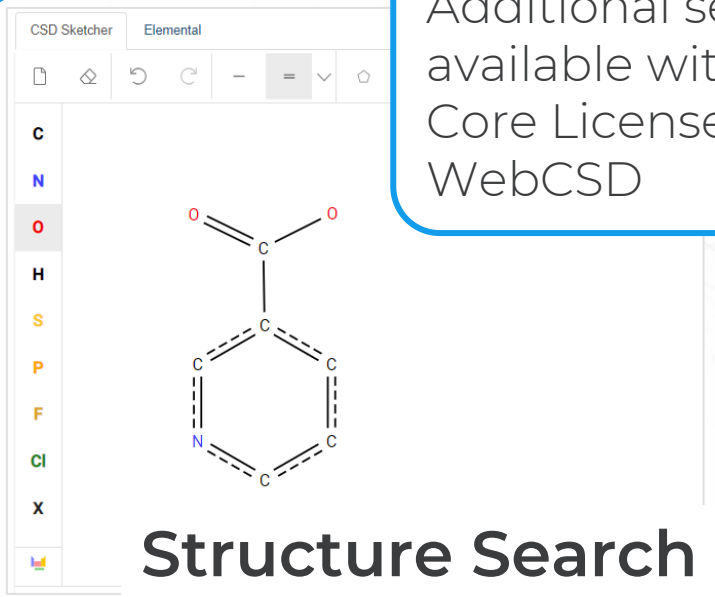
Searching

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

Access Structures/WebCSD

Identifier(s) — CCDC Number(s), CSD Number(s), CSD Refcode(s) or ICSD Number(s)
Compound name — e.g. sulfadiazine
DOI — A single publication DOI, CSD DOI or ICSD DOI
Authors — e.g. F.H.Allen
Journal — e.g. Journal of the American Chemical Society
Publication details — Year | Volume | Page
Database to search — Entire published collection | CSD | ICSD | Teaching subset
+ Add New Search Field
Search

Simple Search



Structure Search

Additional searching available with CSD-Core License in WebCSD

Unit Cell Search

Lattice centring — Primitive (P)

a	e.g. 10.0	α	e.g. 90.0
b	e.g. 10.0	β	e.g. 90.0
c	e.g. 10.0	γ	e.g. 120.0

Formula Search

e.g. C8 H9 N1 O2

CCDC

Search results

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

Your query was: Compound name: paracetamol and the search returned more than 30 records.

☐ Select all | Download Selected | View Selected

AHEPUY

Deposition Number(s): 200783
Space Group: P $\bar{1}$ (2)
Cell: *a* 8.710(4)Å *b* 9.920(5)Å *c* 12.385(5)Å, *α* 102.35(3)° *β* 108.33(2
Compound Name: N-(4-Hydroxyphenyl)acetamide morpholine
Synonyms: Paracetamol morpholine, Acetaminophen morpholine

AMUBAM

Deposition Number(s): 803736
Space
Cell: *a*
Comp
Synon

COKCEL

Depos
Space
Cell: *a*
Comp
Synon

COTYOA02

Dep
Spac
Cell:
Compound Name: Acetaminophen
Synonyms: paracetamol, DrugBank: DB00316

COTYOA03

Deposition Number(s): 735856
Space Group: P $2_1/a$ (14)
Cell: *a* 12.864(8)Å *b* 9.354(5)Å *c* 7.104(6)Å, *α* 90° *β* 115.83(5)° *γ* 90°
Compound Name: Acetaminophen
Synonyms: paracetamol, DrugBank: DB00316

Results

Database Identifier	Deposition Number
AHEPUY	200783

Next

Download

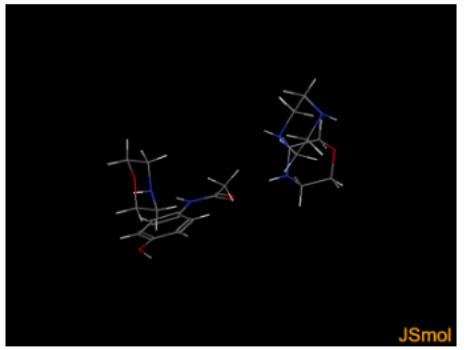
Download deposited CIF

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

☐ Include checkCIF reports as a PDF when available

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

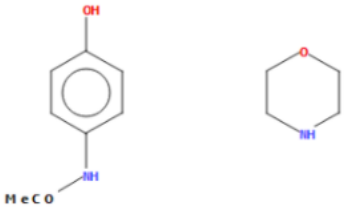
3D viewer



H | Disorder | Menu | Open

Style: Wireframe | Labels: All Metals | Packing: None | Measure: None

Chemical diagram



View group symbols key

Additional details

Deposition Number	200783
Data Citation	I.D.H.Oswald, W.D.S.Motherwell, S.Parsons, C.R.Pulham CCDC 200783: Experimental Crystal Structure Determination, 2003, DOI: 10.5517/cc6qxwp
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

CCDC

Live Demo

- Searching using WebCSD

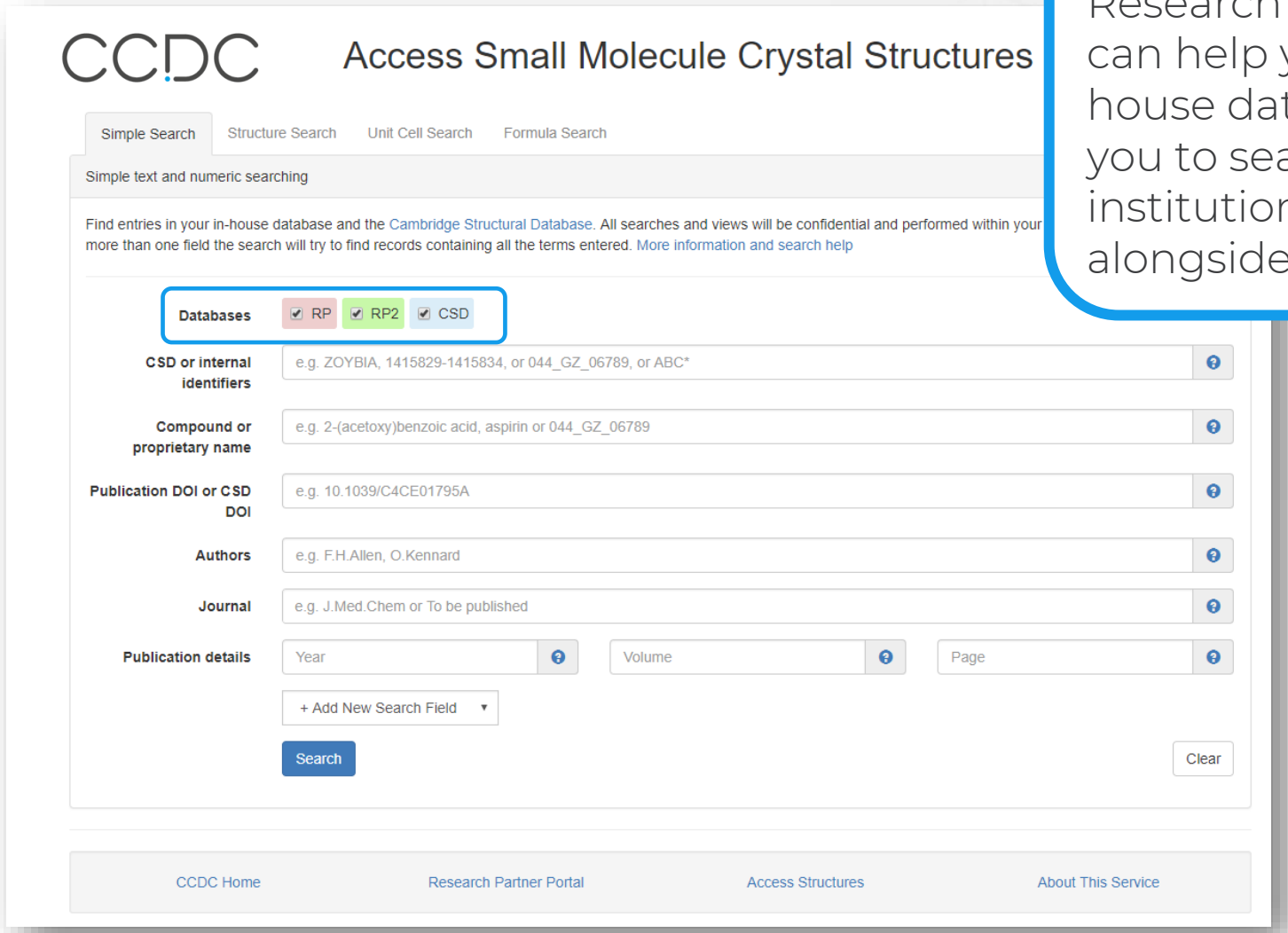
- Accessing data from search results

https://www.ccdc.cam.ac.uk/structures/

CCDC

In house databases

If your institution is a Research Partner, the CCDC can help you set up an in-house database allowing you to search your institutions proprietary data alongside the CSD.



CCDC

CSD Discovery

Docking with GOLD in the API

Dr Ilenia Giangreco

Discovery Science Team Leader

CCDC

Why?

- More advanced settings are now available when docking with GOLD programmatically through the CSD Python API
- Previously some advanced features in GOLD were only accessible via Hermes, and users were asked to configure their parameter file in the GUI before they could run it via the CSD Python API

- The docking module in the CSD Python API is accessible to CSD-Discovery and CSD-Enterprise users as well as RPs.

What?

To computational chemists in academia and in the pharmaceutical industry, this latest version of the docking module in the CSD Python API will allow

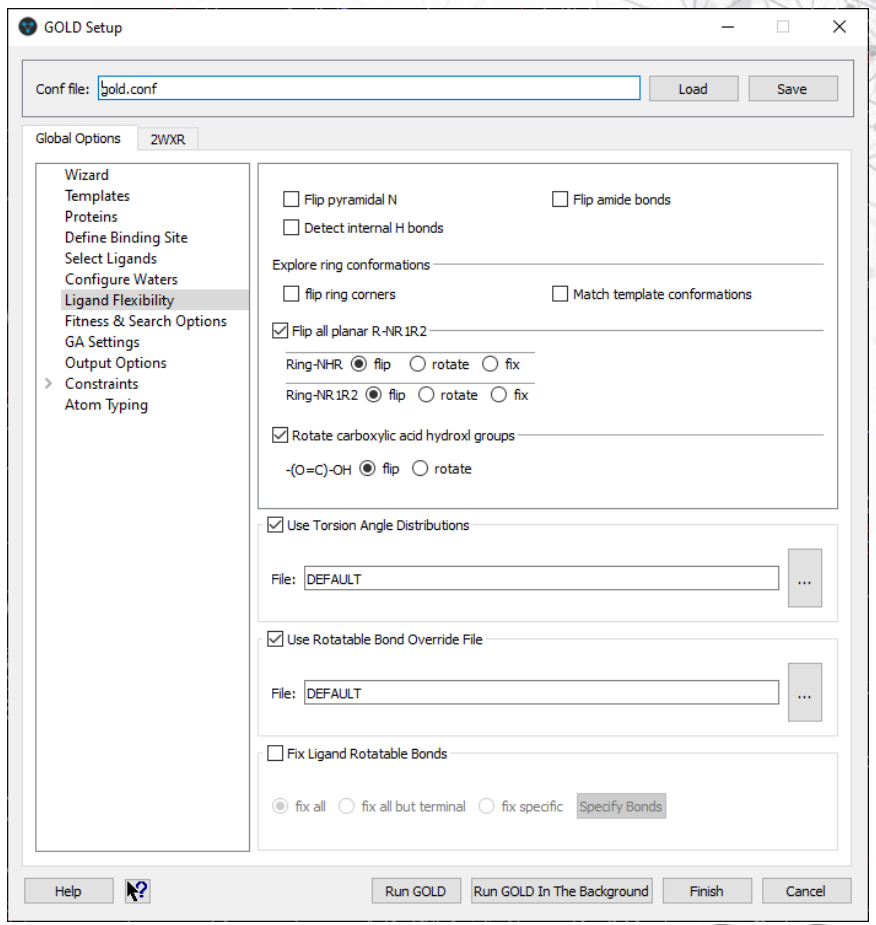
- water handling
- side chain flexibility
- pharmacophore constraints

providing a much more complete range of docking capabilities to be configured programmatically.

Additional new features

Additional GOLD Docking configuration flags have been exposed:

- flip_planar_nitrogen
- flip_free_corners
- flip_amide_bonds
- flip_pyramidal_nitrogen
- save_lone_pairs
- match_template_conformations
- rotate_carboxylic_hydroxyl_groups
- use_torsion_angle_distributions
- fix_ligand_rotatable_bonds
- rotatable_bond_override_file
- fix_all_protein_rotatable_bonds
- solvate_all
- use_internal_ligand_energy_offset

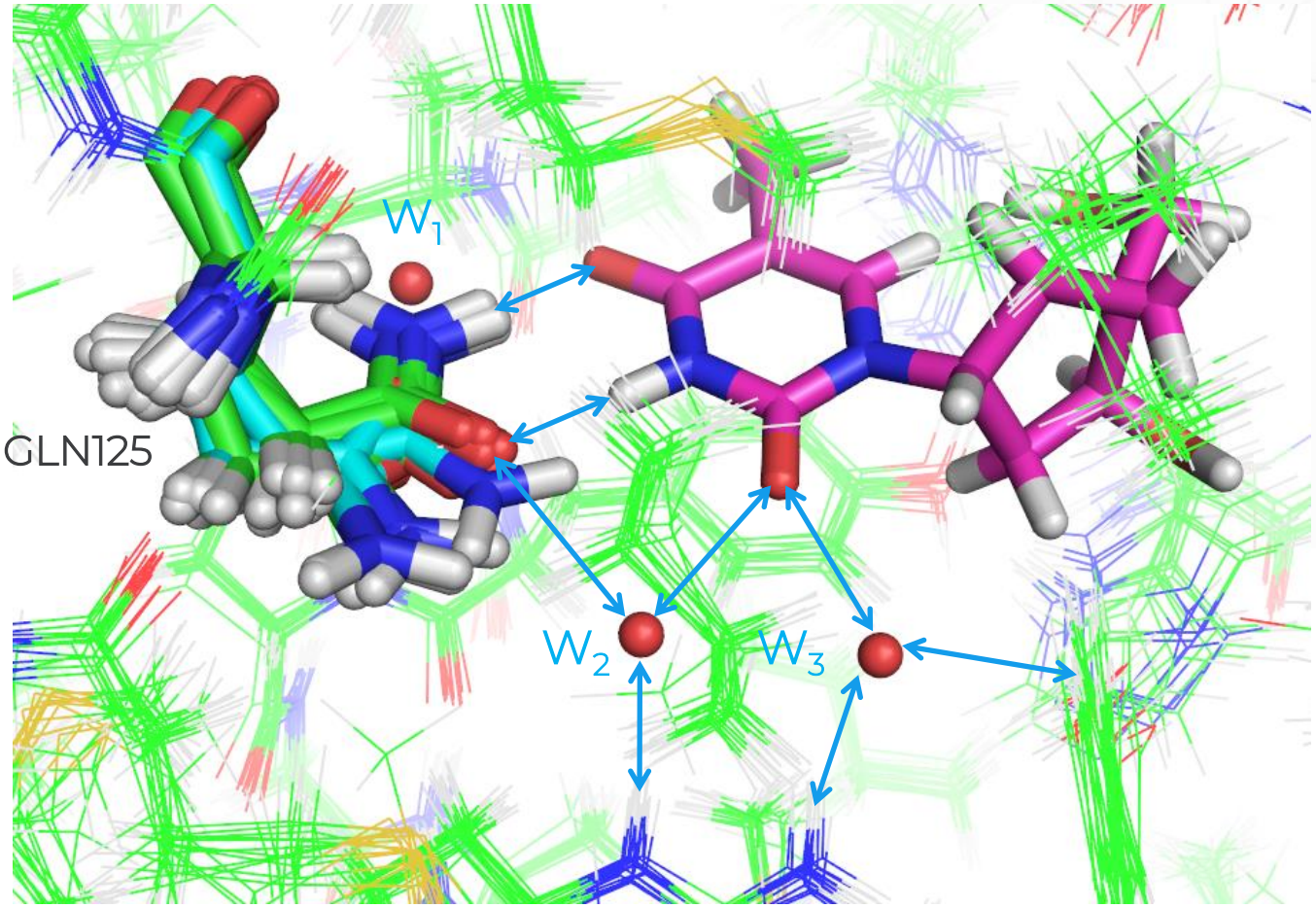


What’s still missing?

- Metal coordination
- Interaction motif
- Per atom scores
- Soft potential
- Covalent docking

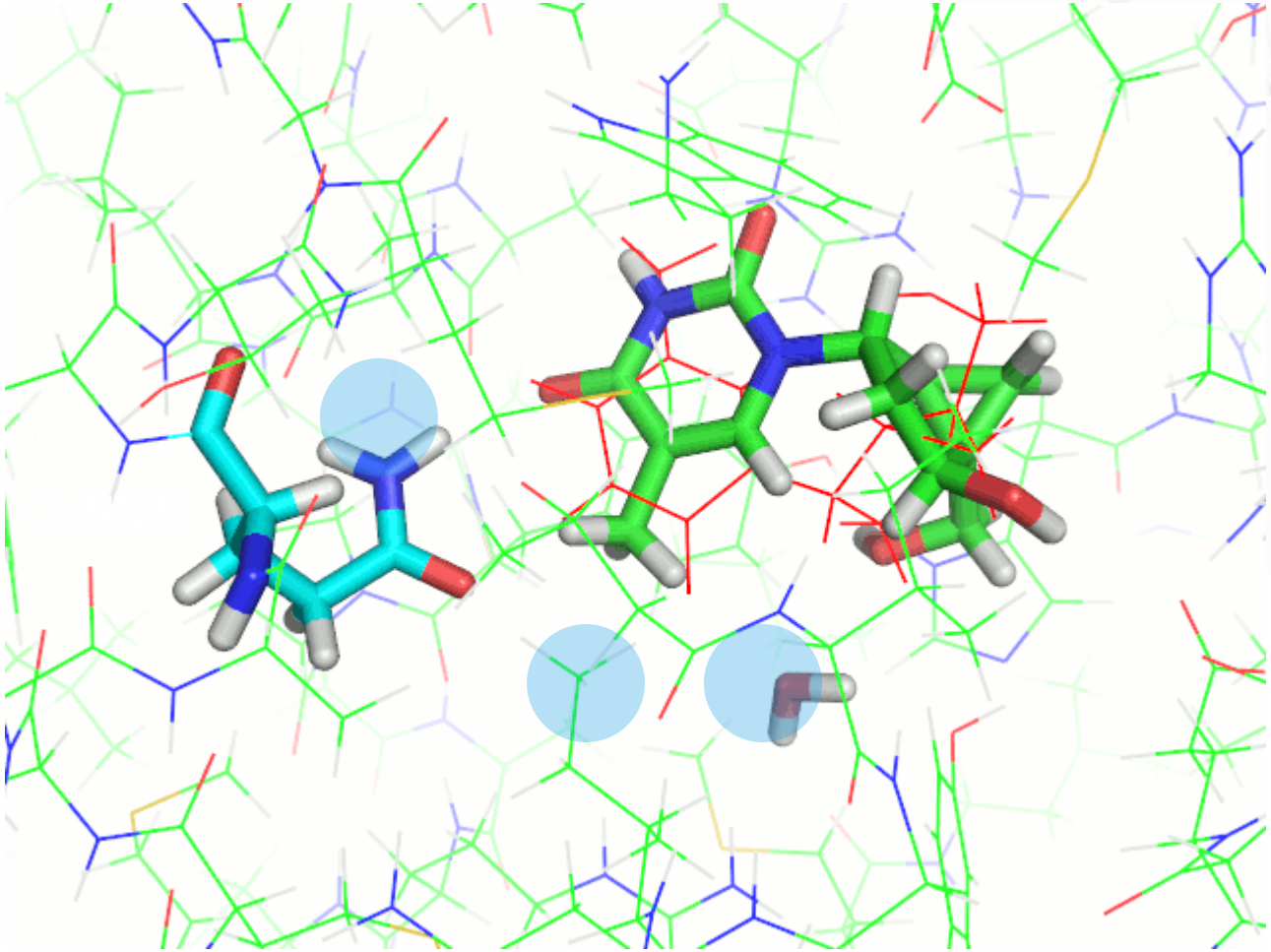
Docking with waters

Ensemble docking with waters



CCDC

How GOLD works



Reproducing with the CSD Python API (1)

```
from ccdc.docking import Docker
from ccdc.io import MoleculeReader, EntryReader
from ccdc.molecule import Molecule
from ccdc.protein import Protein
import os

docker = Docker()
settings = docker.settings

proteins = ['1E2H_protein.mol2', '1E2I_protein.mol2', '1OF1_protein.mol2', '4IVQ_protein.mol2']
merged_protein = Molecule()

# creating a merged protein from multiple proteins and adding proteins individually to the ensemble
for protein in proteins:
    settings.add_protein_file(os.path.abspath(protein))
    protein = Protein.from_file(os.path.abspath(protein))
    merged_protein.add_molecule(protein)


# defining the binding site
ligand = MoleculeReader('1E2K_ligand.mol2')[0]
settings.binding_site = settings.BindingSiteFromLigand(merged_protein, ligand, 8, whole_residues=True)

# defining the ligand
settings.add_ligand_file('1E2K_ligand.mol2', 10)
```

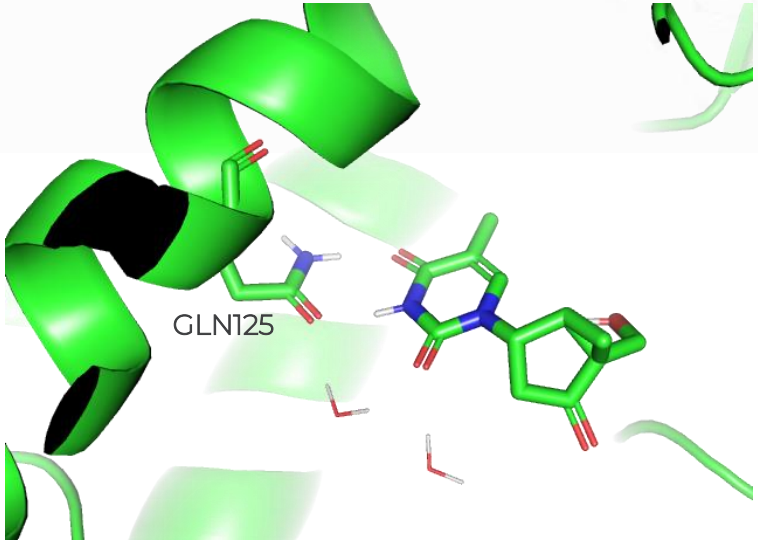
https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD-tutorial-EnsembleDocking.pdf

CCDC

Reproducing with the CSD Python API (2)

```
# additional settings
settings.fitness_function = 'plp'
settings.output_directory = 'results_api'
settings.early_termination = False

# setting water molecules to spin and translate up to 1Å
settings.add_water_file(os.path.abspath('water_1.mol2'), toggle_state='toggle', spin_state='trans_spin', movable_distance=1.0)
settings.add_water_file(os.path.abspath('water_2.mol2'), toggle_state='toggle', spin_state='trans_spin', movable_distance=1.0)
settings.add_water_file(os.path.abspath('water_3.mol2'), toggle_state='toggle', spin_state='trans_spin', movable_distance=1.0)

# inspecting the results
results = docker.dock()
docked_ligands = Docker.Results(settings).ligands
for l in docked_ligands:
    print(l.fitness(), len(l.docked_waters))

scores = [l.fitness() for l in docked_ligands]
i = scores.index(max(scores))
top_ranked = docked_ligands[i]
print(("Fitness: {}, EnsembleProteinID: {}, Number of docked waters:
{}.".format(top_ranked.fitness(),top_ranked.scoring_term('ensemble', 'id'),
(top_ranked.docked_waters)))
```

Fitness: 84.7412, EnsembleProteinID: 3.0, Number of docked waters: 2.



https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD-tutorial-EnsembleDocking.pdf

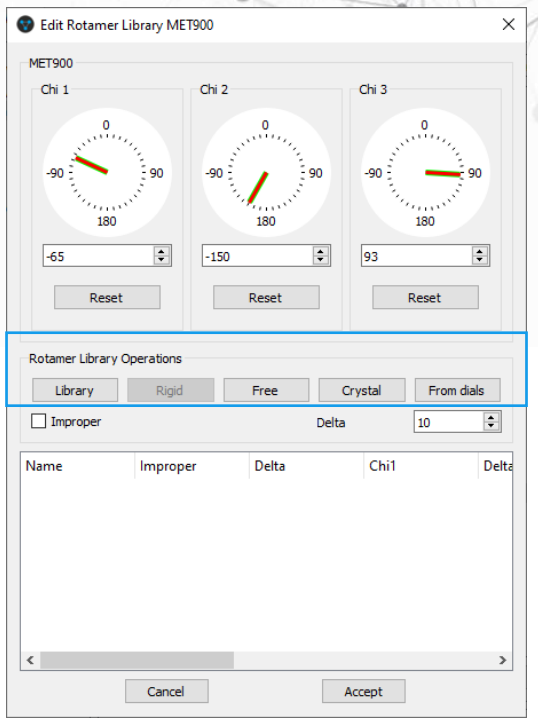
CCDC

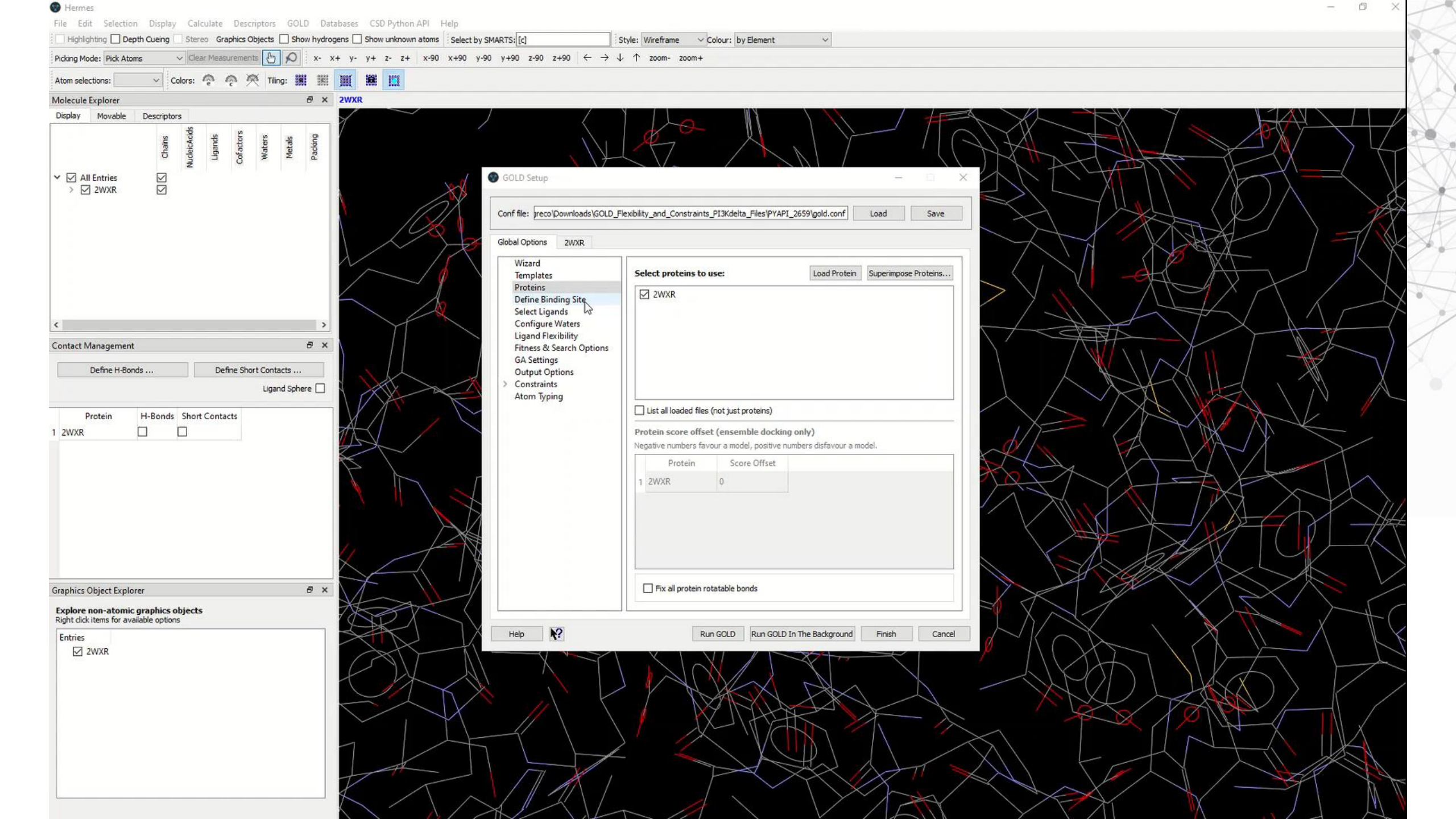
Setting up flexible side chains

Flexible side chains in GOLD

- Up to 10 protein side chains can be treated as flexible during docking
  - Each flexible side chain will be allowed to undergo torsional rotation around one or more of its acyclic bonds
    - This can make docking more difficult as it increases the search space and the chance of false positives
    - A side chain should be made flexible only if there is a good reason to believe (e.g. from X-ray data) that it is likely to move in response to ligand binding

- To make a side chain flexible one or more allowed rotamers need to be defined
  - Each rotamer specifies the torsion angles that are permitted to vary, and the allowed values or ranges of values for those torsion angles
  - Rotamers can be specified in different ways:
    - From a pre-canned rotamer library read by a default parameter file (Library)
    - From the protein input file (Crystal)
    - From user-defined angles and ranges (From Dials)
    - Fully flexible over the range -180 to +180 (Free)




Hermes
File Edit Selection Display Calculate Descriptors GOLD Databases CSD Python API Help
Highlighting Depth Cueing Stereo Graphics Objects Show hydrogens Show unknown atoms
Select by SMARTS: [c]
Style: Wireframe
Colour: by Element
Picking Mode: Pick Atoms
Clear Measurements
x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90 zoom- zoom+
Atom selections:
Colors:
Tiling:
Molecule Explorer
Display Movable Descriptors
Chains NucleicAcids Ligands Cofactors Waters Metals Packing
All Entries
2WXR
2WXR
Contact Management
Define H-Bonds ...
Define Short Contacts ...
Ligand Sphere
Protein H-Bonds Short Contacts
1 2WXR
Graphics Object Explorer
Explore non-atomic graphics objects
Right click items for available options
Entries
2WXR
GOLD Setup
Conf file: preco\Downloads\GOLD_Flexibility_and_Constraints_PI3Kdelta_Files\PYAPI_2659\gold.conf
Load
Save
Global Options
2WXR
Wizard
Templates
Proteins
Define Binding Site
Select Ligands
Configure Waters
Ligand Flexibility
Fitness & Search Options
GA Settings
Output Options
Constraints
Atom Typing
Select proteins to use:
Load Protein
Superimpose Proteins...
2WXR
List all loaded files (not just proteins)
Protein score offset (ensemble docking only)
Negative numbers favour a model, positive numbers disfavour a model.
Protein Score Offset
1 2WXR 0
Fix all protein rotatable bonds
Help
Run GOLD
Run GOLD In The Background
Finish
Cancel

Flexible side chains in the CSD Python API

```
from ccdc.docking import Docker
from ccdc.io import MoleculeReader, EntryReader

docker = Docker()
settings = docker.settings
settings.add_protein_file('2wxr_catalytic_domain.mol2')

# adding the three ligands to be docked
settings.add_ligand_file('IC8.mol2', 10)
settings.add_ligand_file('40L_t1.mol2', 10)
settings.add_ligand_file('40L_t2.mol2', 10)

# defining the binding site
protein = settings.proteins[0]
val828 = [a for a in protein["A:VAL828"].atoms if a.label=="N"][0]
settings.binding_site = settings.BindingSiteFromAtom(protein, val828, 10.0)

# other settings
settings.fitness_function = 'goldscore'
settings.output_directory = 'output_api'
settings.early_termination = False

# Setting a protein hydrogen bond constraint
backbone_nh_val828 = [a for a in protein["A:VAL828"].atoms if a.label=="H"]
settings.add_constraint(settings.ProteinHBondConstraint(backbone_nh_val828))
```

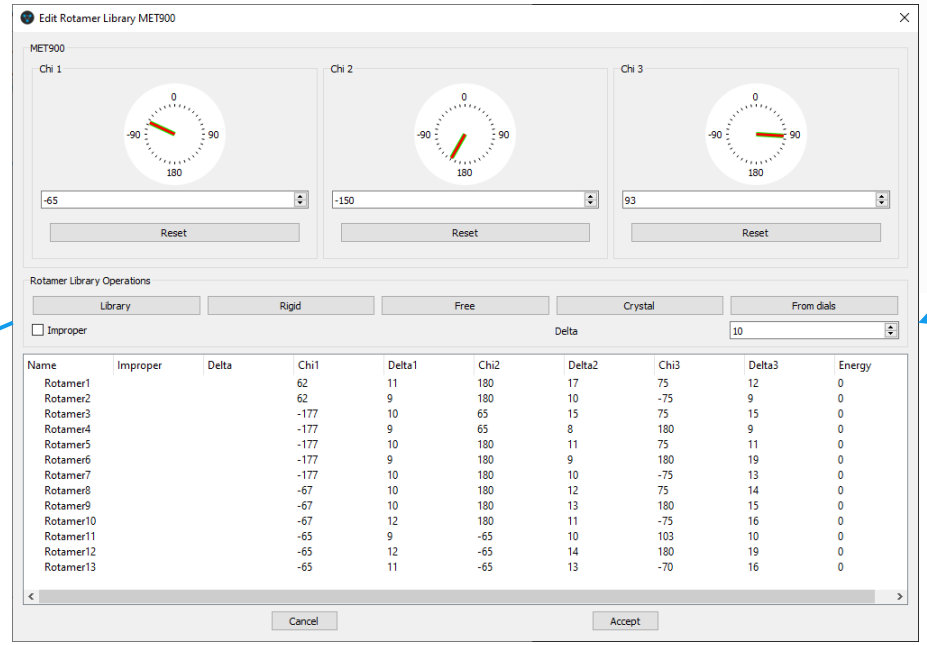
https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD-tutorial-Flexibility-Constraints.pdf

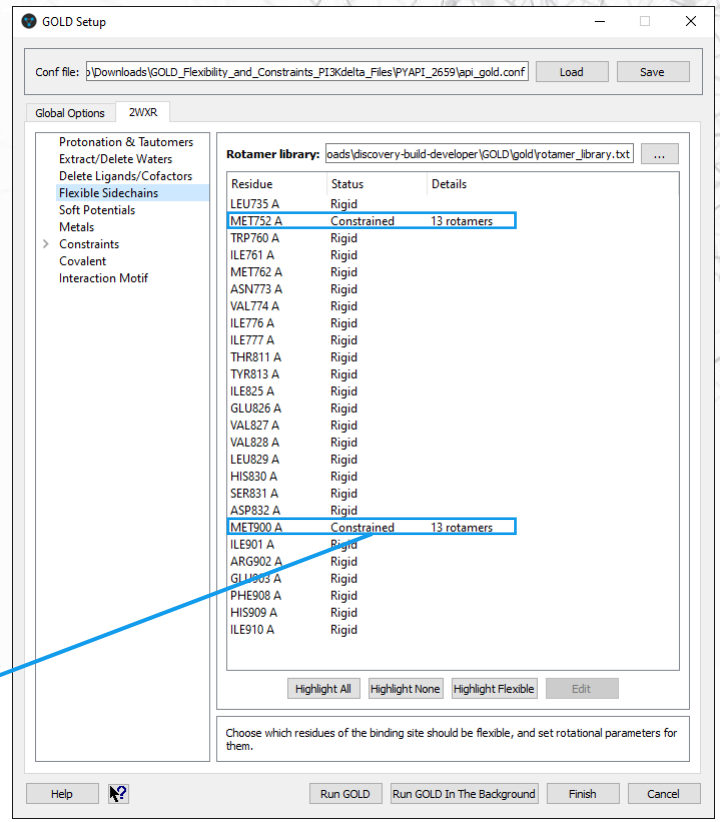
CCDC

```
# Setting up flexible side chains
met900 = [r for r in settings.binding_site.residues if r.identifier=="A:MET900"][0]
met900_rotamer_library = settings.RotamerLibrary(settings.protein_files[0], met900)
met900_rotamer_library.add_default_rotamers()
settings.add_rotamer_library(protein, met900_rotamer_library)

met752 = [r for r in settings.binding_site.residues if r.identifier=="A:MET752"][0]
met752_rotamer_library = settings.RotamerLibrary(settings.protein_files[0], met752)
met752_rotamer_library.add_default_rotamers()
settings.add_rotamer_library(protein, met752_rotamer_library)

results = docker.dock()
```

api_gold.conf

```
rotamer_lib
  name MET900
 chi1 3031 3032 3035 3036
 chi2 3032 3035 3036 3037
 chi3 3035 3036 3037 3038
 rotamer 62 (11) 180 (17) 75 (12)
 rotamer 62 (9) 180 (10) -75 (9)
 rotamer -177 (10) 65 (15) 75 (15)
 rotamer -177 (9) 65 (8) 180 (9)
 rotamer -177 (10) 180 (11) 75 (11)
 rotamer -177 (9) 180 (9) 180 (19)
 rotamer -177 (10) 180 (10) -75 (13)
 rotamer -67 (10) 180 (12) 75 (14)
 rotamer -67 (10) 180 (13) 180 (15)
 rotamer -67 (12) 180 (11) -75 (16)
 rotamer -65 (9) -65 (10) 103 (10)
 rotamer -65 (12) -65 (14) 180 (19)
 rotamer -65 (11) -65 (13) -70 (16)
end_rotamer_lib
```



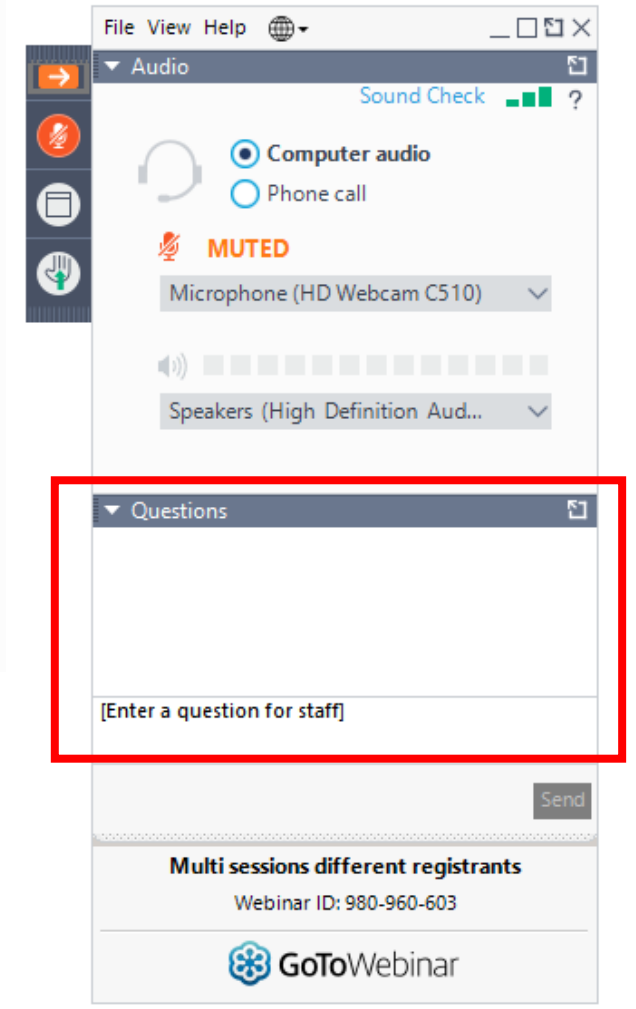


https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD-tutorial-Flexibility-Constraints.pdf

CCDC

Q&A

- Type your questions in the box as shown

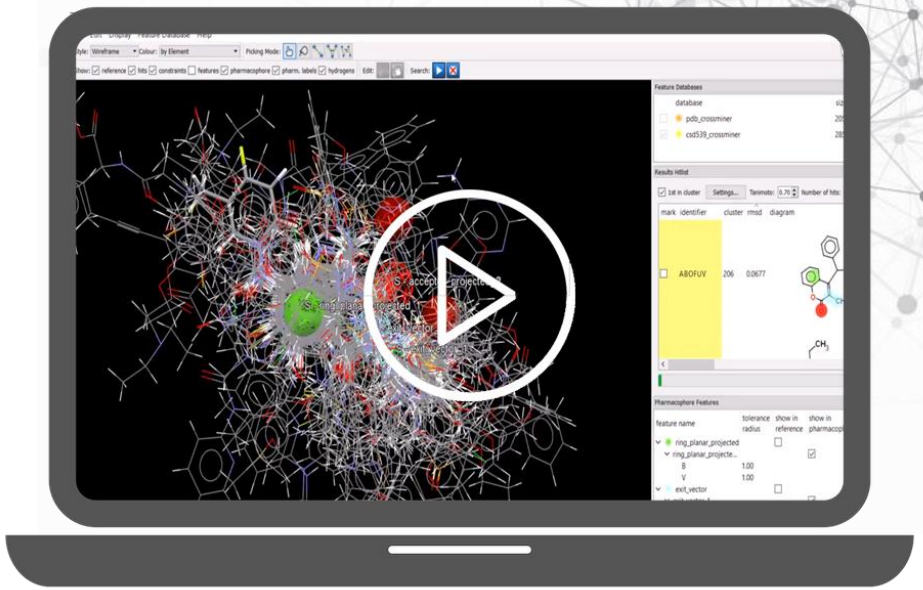


CCDC

Next What's Up Webinar

- Next webinar: **September 23rd**
  - SMILES to 3D structures generation
  - How to enhance your CSD depositions
- Follow us on social media
- Send us your ideas and news

  hello@ccdc.cam.ac.uk



CCDC

Thank you

hello@ccdc.cam.ac.uk

The Cambridge Crystallographic Data Centre
12 Union Road, Cambridge CB2 1EZ, United Kingdom
Registered Charity No. 800579

CCDC