REC

Introduction to Pharmacophore Searching
Using CSD-CrossMiner

CCDC Virtual Workshop

November 2024

CCDC
advancing structural science

Learning outcomes

- An understanding of the pharmacophore concept.
- Familiarity with the CSD-CrossMiner interface and how it can be used in your research.
- How to perform pharmacophore searches biologically relevant subsets of the CSD and PDB.
- How to set up an interactive pharmacophore query and save the results.
- How to analyse and interact with your results.

CCDC

We will make the recording available to you in the next few days.

Agenda

- *Show One:* Intro and CSD and CSD-CrossMiner overview
- *Show One:* PDB overview
- *Show One:* CSD-CrossMiner explained and demonstration
- *Try One:* Hands-on example
- *Explore More:* Case-studies of CSD-CrossMiner
- *Explore More:* Quiz and Summary
- *Extra time:* More time for hands on and Q&A

Special Guest

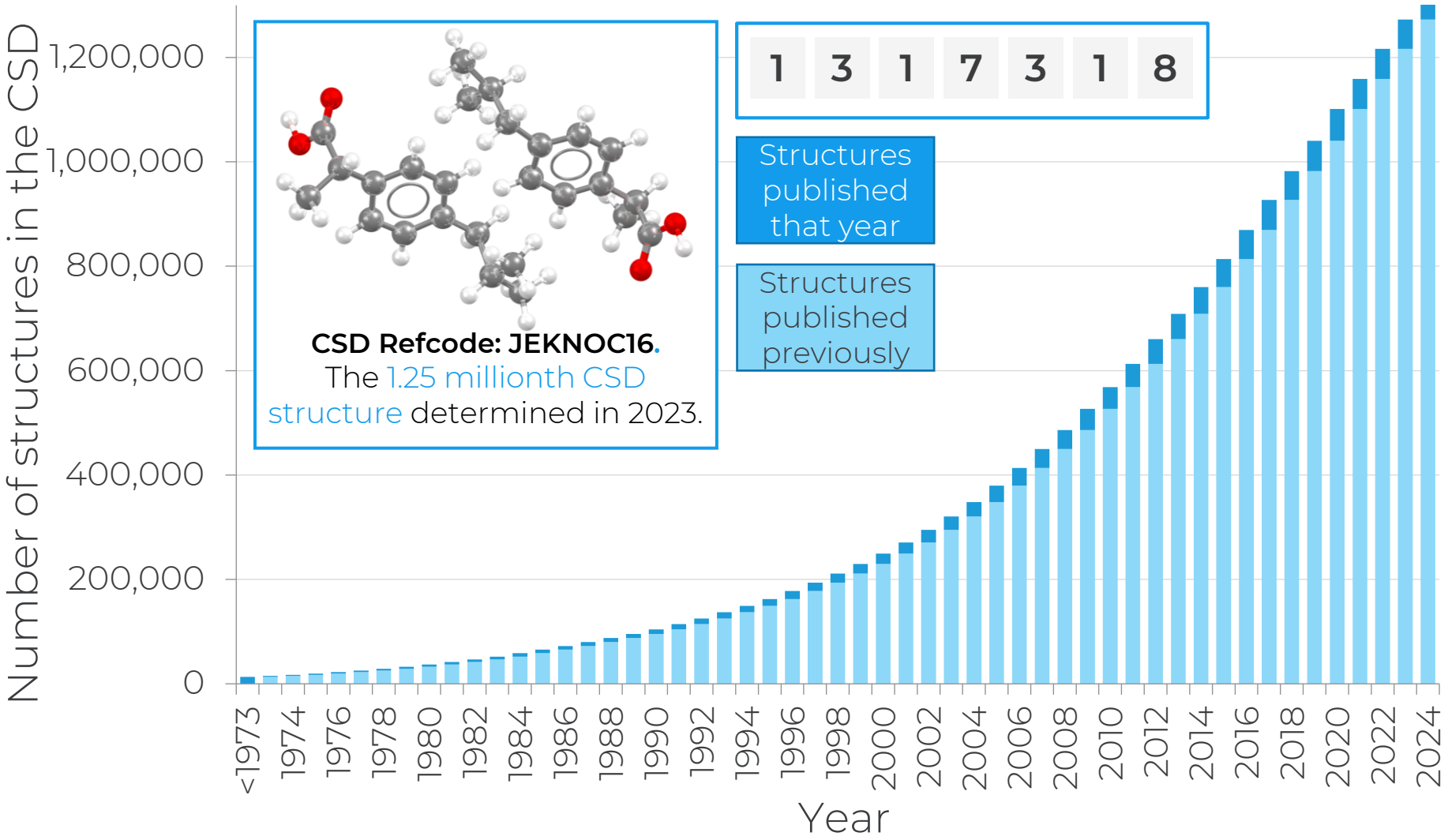
Deborah Harrus

PDBe Archive Project Leader

PDBe Protein Data Bank in Europe

CCDC

The Cambridge Structural Database



- Every published structure
  - Inc. ASAP & early view
  - *CSD Communications*
  - Patents
  - University repositories
  - Thesis
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for over 59 years
- A trusted CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

Small molecules, big impact

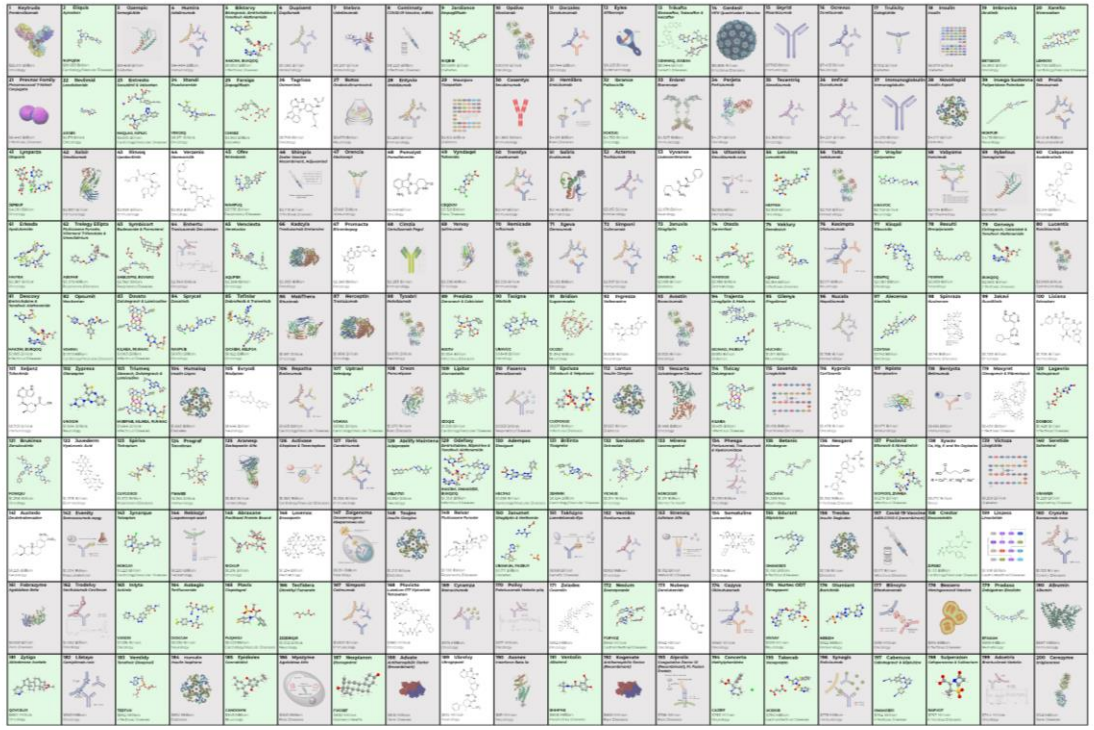
200 top drugs by retail sales in 2023

Small molecule structure in the CSD

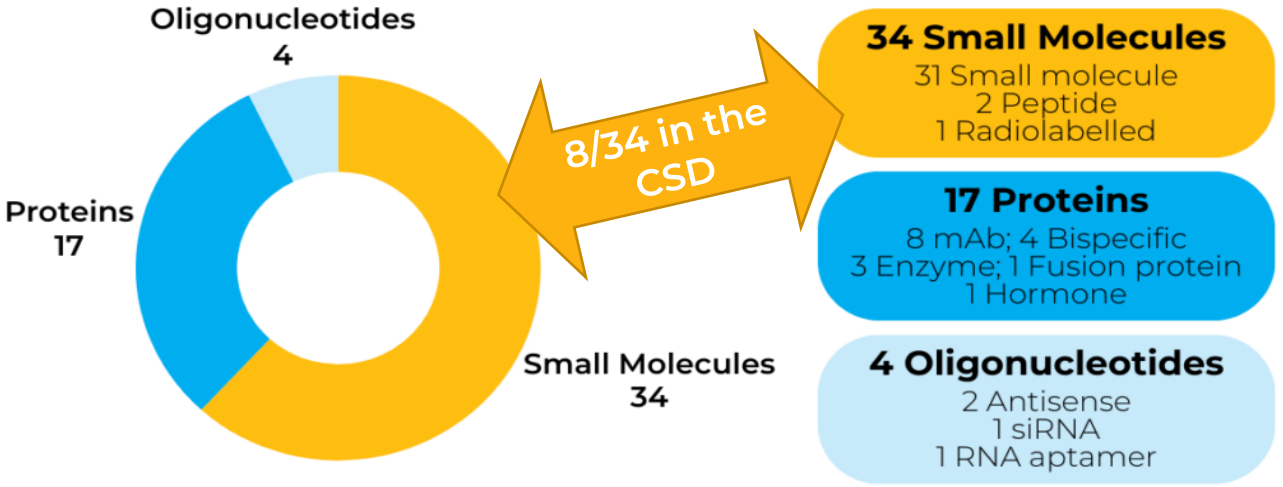
- Open-up new biological and therapeutic opportunities
- Compete against emergent modalities for rare diseases
- Are used to target RNA*

Small molecule structure CSD entry pending

Biological structure in the PDB

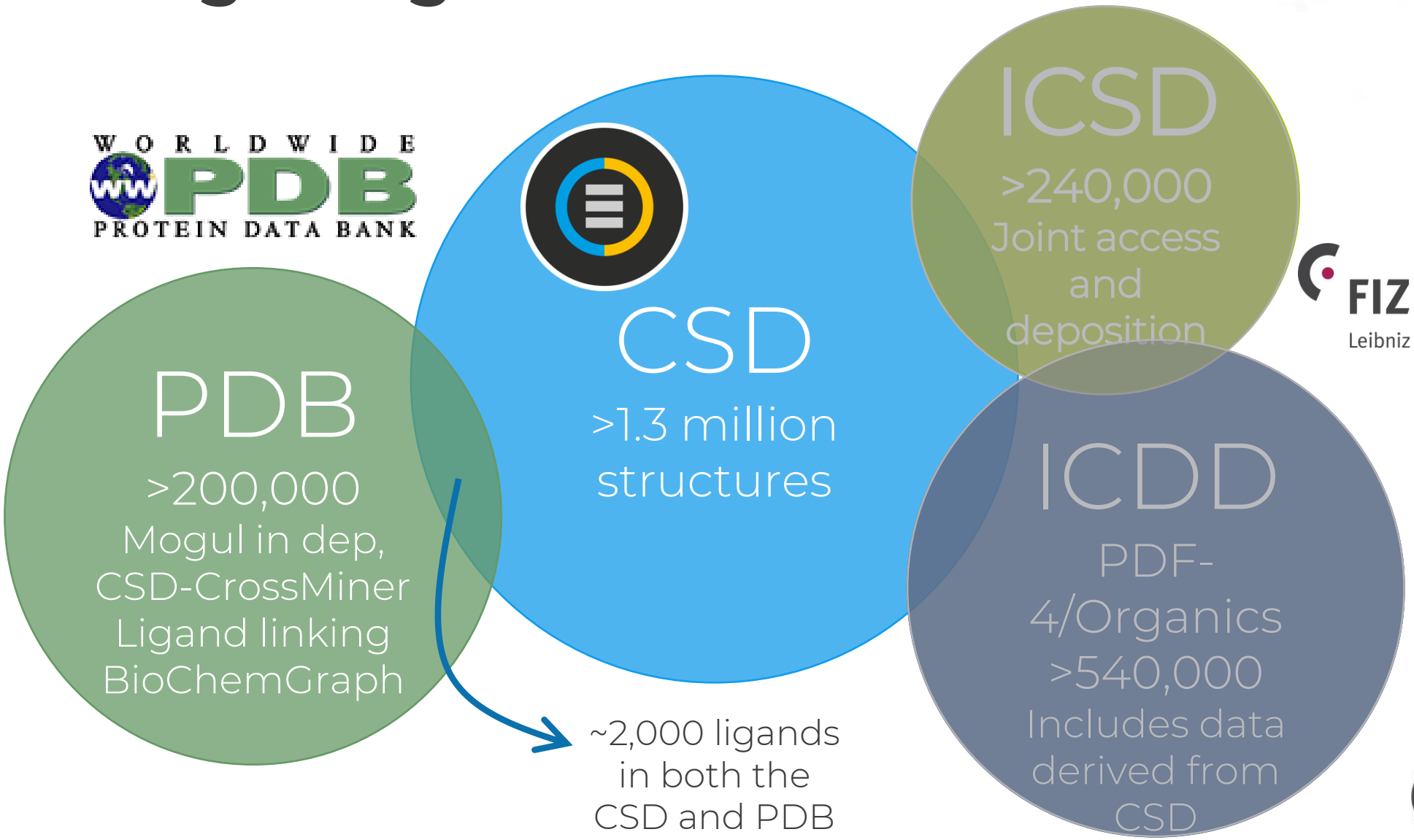


FDA novel drug approvals 2023



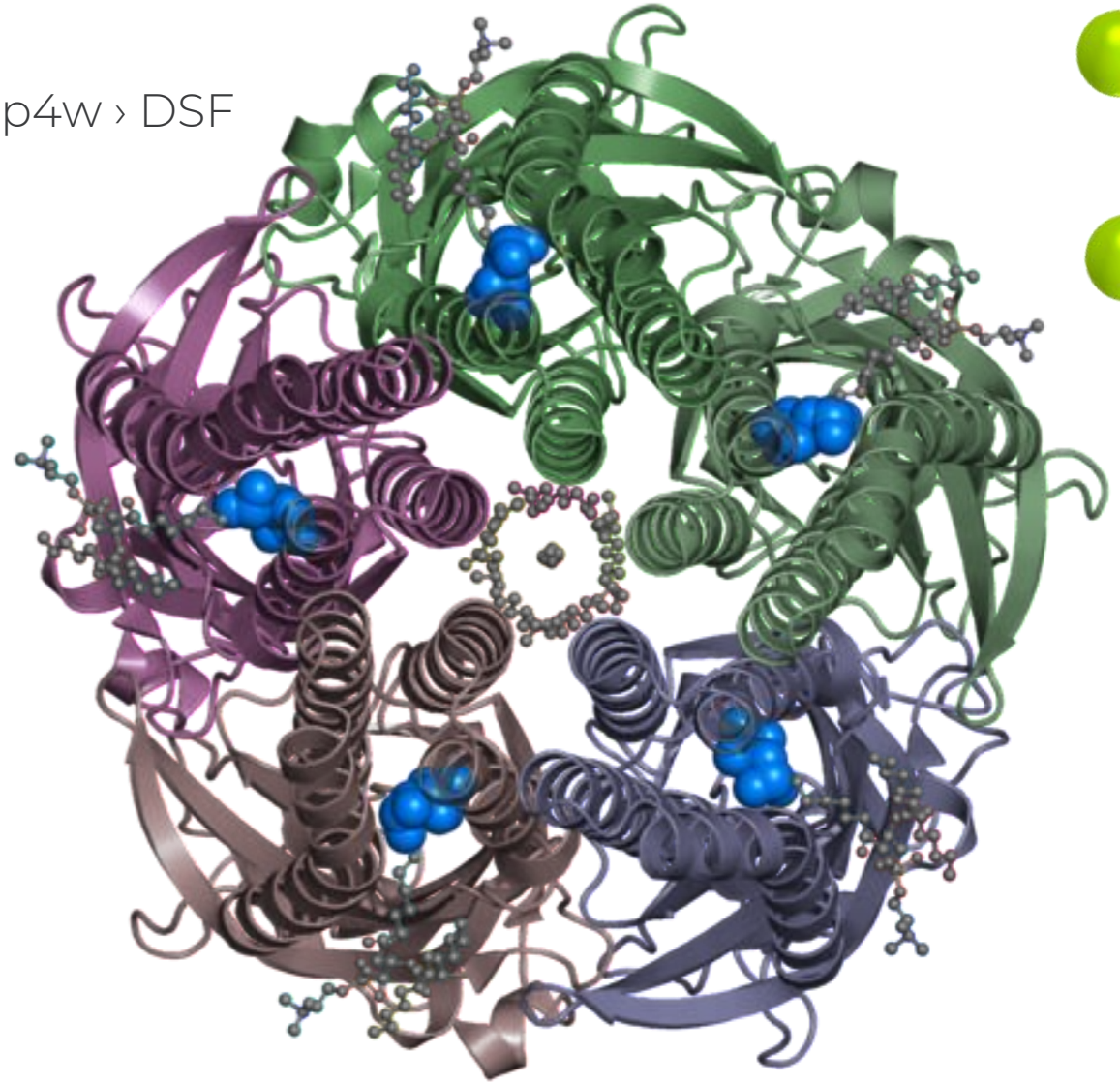
Adapted from poster compiled and produced by the Njardarson Group (The University of Arizona)
Nature Reviews Drug Discovery* **20, 85-90 (2021)

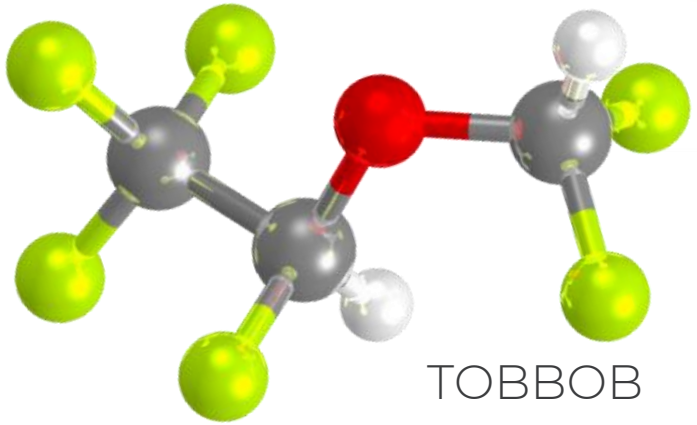
Using integrated structural databases



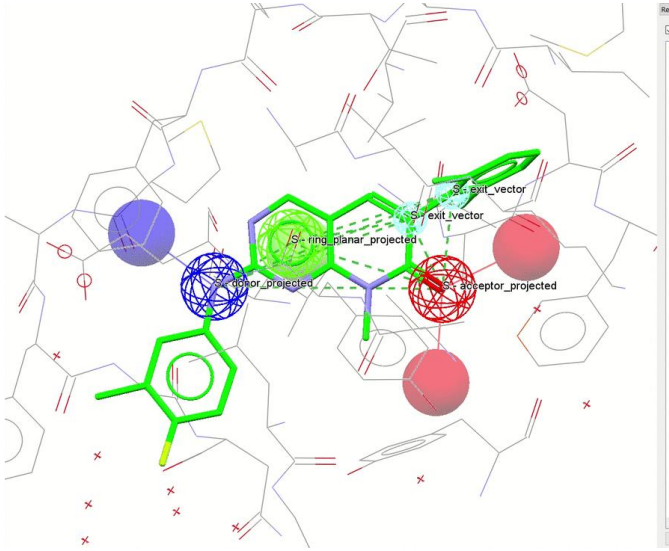
Connecting chemistry and biology

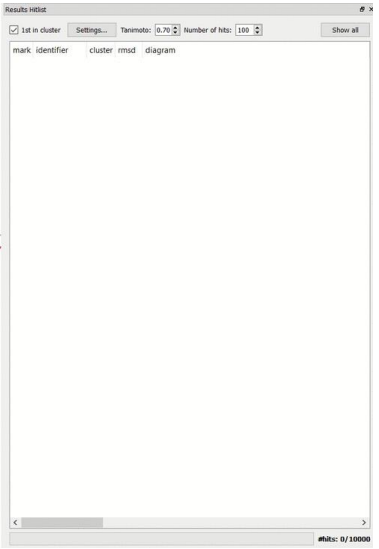
3p4w › DSF





TOBBOB





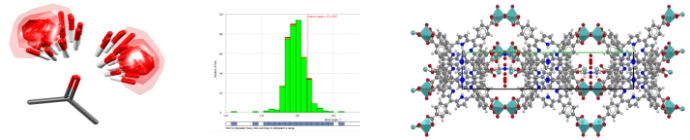
CCDC

The CSD Portfolio

CSDEnterprise.

CSDCore.

Search, visualise, analyse and communicate structural data
Insights into molecular and crystal shape and interactions

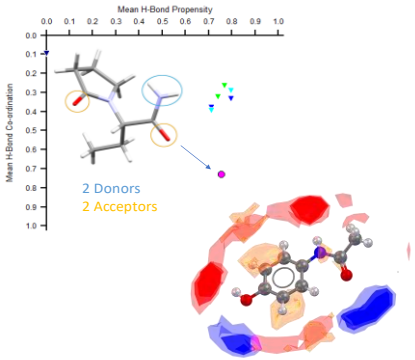


CSDDiscovery.

Design of new molecules

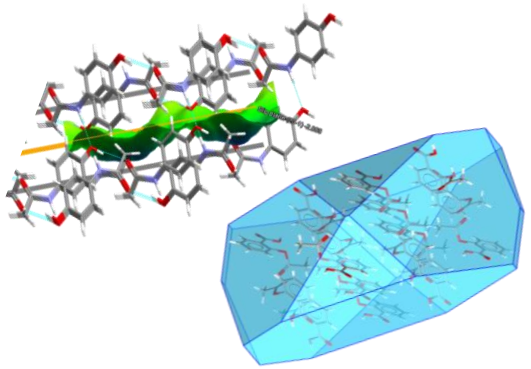
CSDMaterials.

Assessment of solid form stability and properties



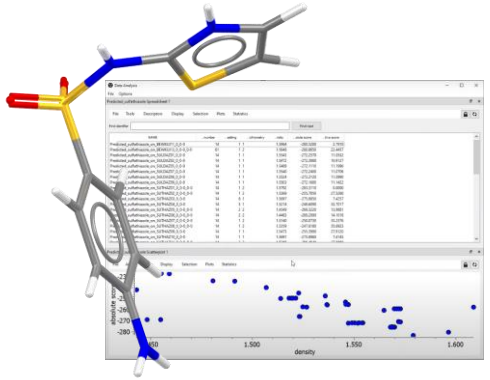
CSDParticle.

Anticipate particle properties and behaviour



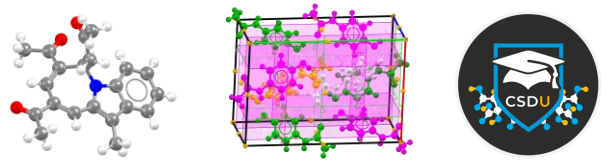
CSDTheory.

Generate solid form landscapes



CSDCommunity.

Deposit, publish, access and visualise structural data
Free functionality to share and learn from structures



Medicinal & Computational Chemists ♦ Crystallographers & Structural Biologists ♦ Solid Form & Crystallisation Scientists ♦ Functional Materials Scientists ♦ Educators ♦ Industry and Academia

CCDC

CSD-CrossMiner - origin

- Developed in collaboration with **Roche**.

*"We exemplify the utility of the approach by **showing applications relevant to real-world drug discovery** projects, including the identification of novel fragments for a specific protein environment or scaffold hopping."*

*"We believe that CSD-CrossMiner **closes an important gap in mining structural data** and will allow users to extract more value from the growing number of available crystal structures."*

Interactive and Versatile Navigation of Structural Databases

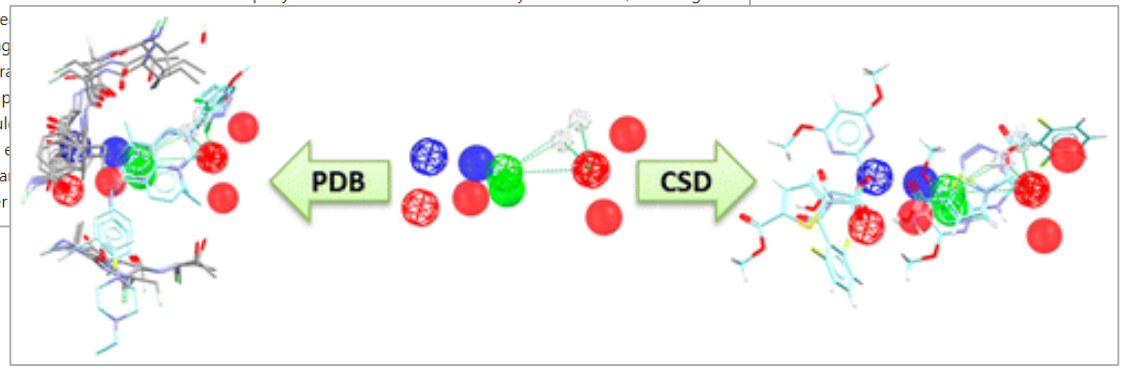
Oliver Korb [1], Bernd Kuhn [2], Jérôme Hert [2], Neil Taylor [3], Jason Cole [1], Colin Groom [1], Martin Stahl [2]

Affiliations + expand

PMID: 26745458 DOI: 10.1021/acs.jmedchem.5b01756

Abstract

We present CSD-CrossMiner, a novel tool for pharmacophore-based searches in crystal structure databases. Intuitive pharmacophore queries describing, among others, protein-ligand interaction patterns, ligand scaffolds, or protein environments can be built and modified interactively. Matching crystal structures are overlaid onto the query and visualized as soon as they are available, enabling the rese
showing
novel fra
search p
molecul
further e
importa
number

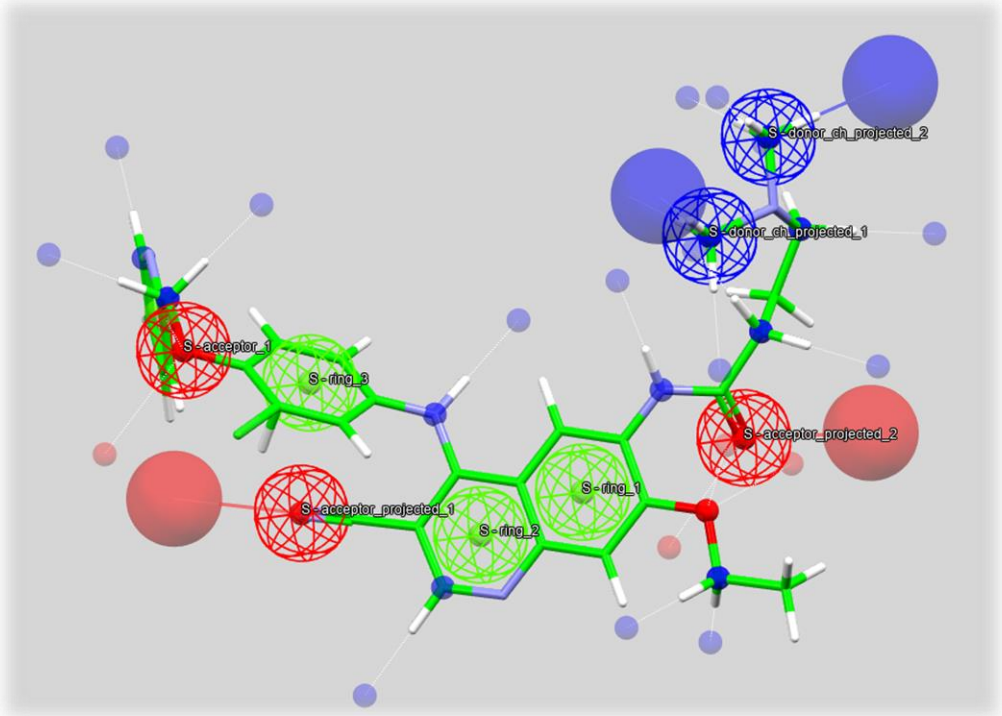


Korb et al,. J. Med. Chem. ***2016,*** *59, 4257-4266. DOI: 10.1021/acs.jmedchem.5b01756*

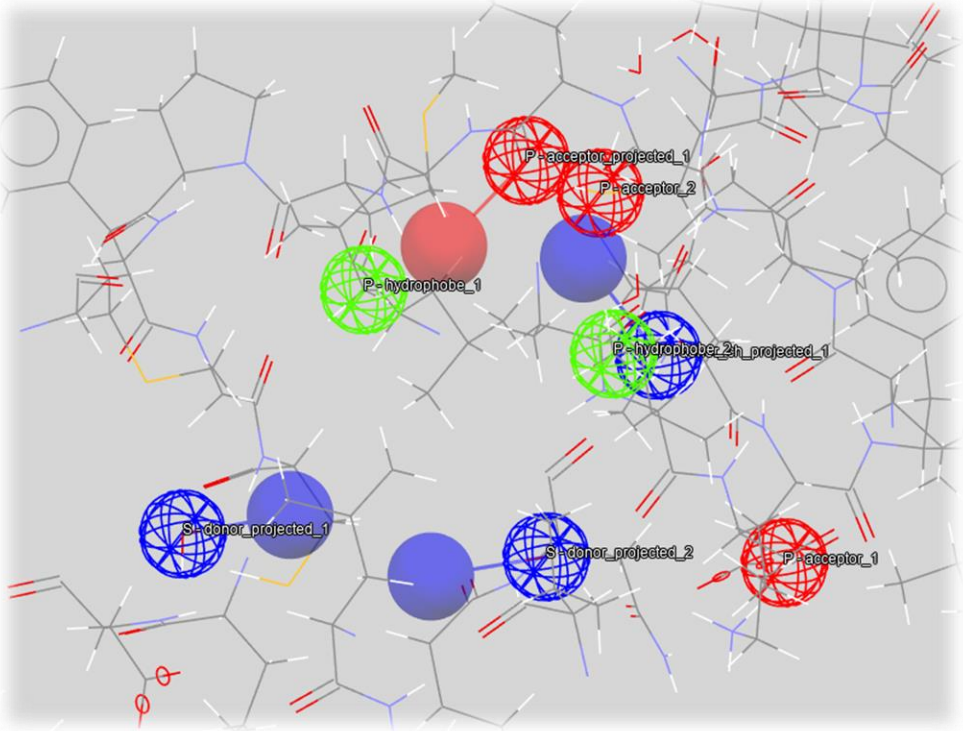
CCDC

CSD-CrossMiner: Why you may need it

You have a known ligand and want to find a new one with better properties but the same target interactions.



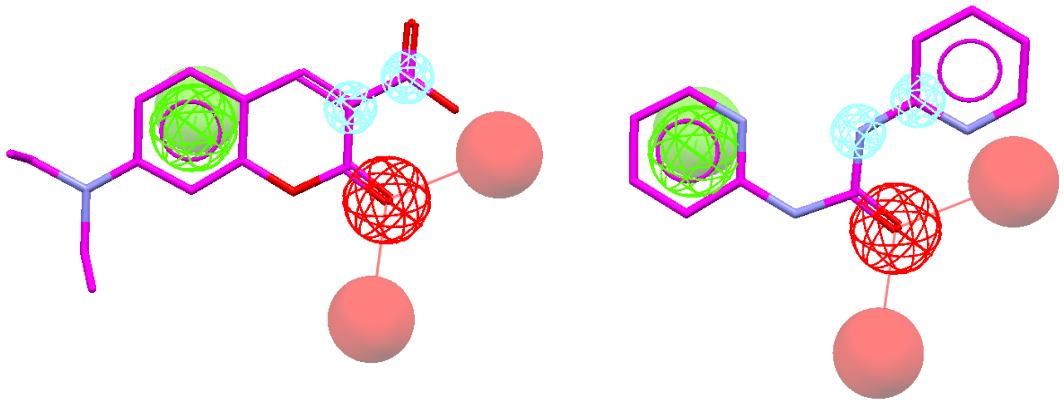
You have a known target with a known binding pocket and want to find compounds which can bind to the pocket.



CCDC

CSD-CrossMiner - applications

- Advance **hit-to-lead**
- Lead **optimisation**
- Find **scaffold hops** to improve pharmacokinetics



- Identify **off-target** effects
- Understand which **modifications** are tolerated in the binding site
- Generate **new ideas** in drug discovery projects

CCDC

Pharmacophore search applications

- **Design novel motifs** that mimic established ligands to improve physicochemical properties or solve patent issues.
- **Scaffold-hopping**: retrieve a diversity of ligand topologies that can be used as scaffolds for lead optimization.
- Look at chemistries that interact with **unexplored parts** of the binding site to improve binding properties of a ligand.

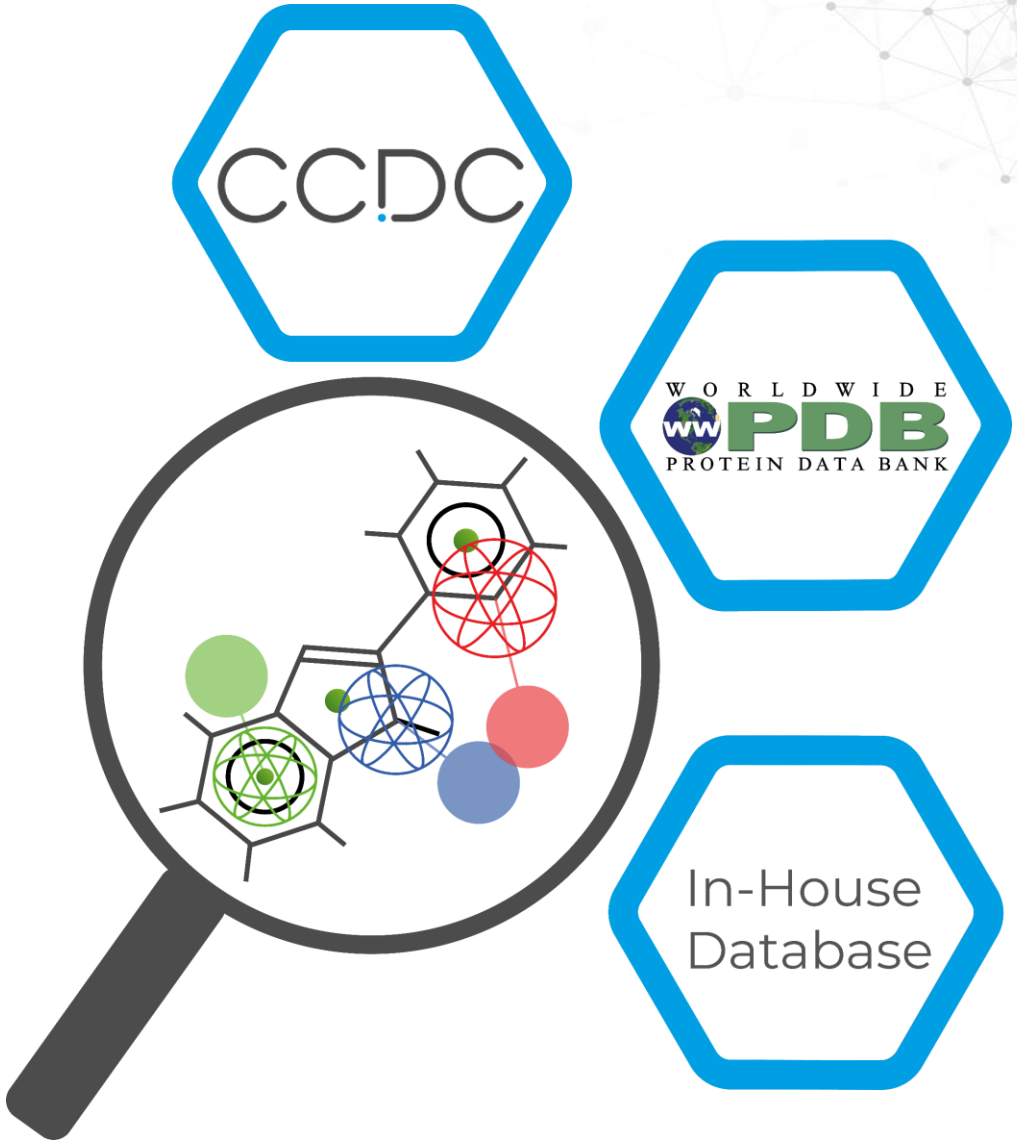
CSD-CrossMiner specific

- Determine **common protein binding sites** to predict selectivity.
- Shed light into **cross-pharmacology** between protein targets.
- Determine **structural motifs** that bind in **similar environments** to find insights for drug design and optimization.

CCDC

CSD-CrossMiner

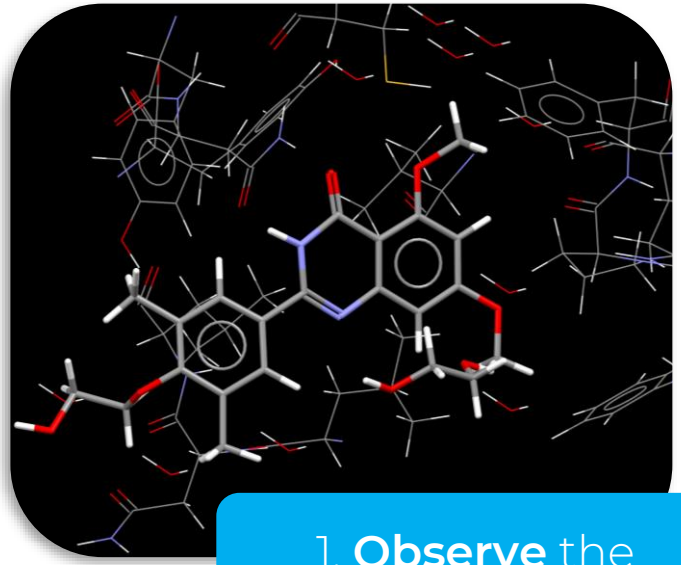
- Tool for building pharmacophore queries.
- Tool for searching structural databases by pharmacophore.
- Simultaneously search the PDB, CSD and your in-house database.
- Designed for speed – modify hypotheses on the fly.
- Structures are annotated for easy filtering of hits.



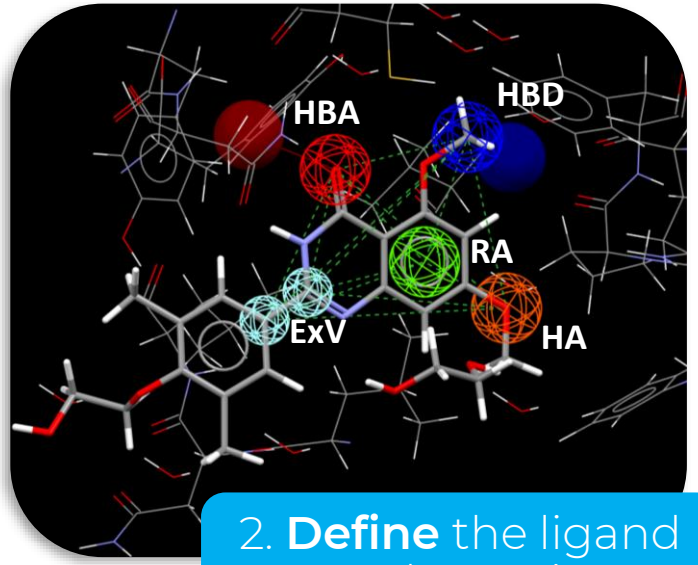
CCDC

Korb *et al,. J. Med. Chem.* **2016,** 59, 4257-4266. DOI: 10.1021/acs.jmedchem.5b01756

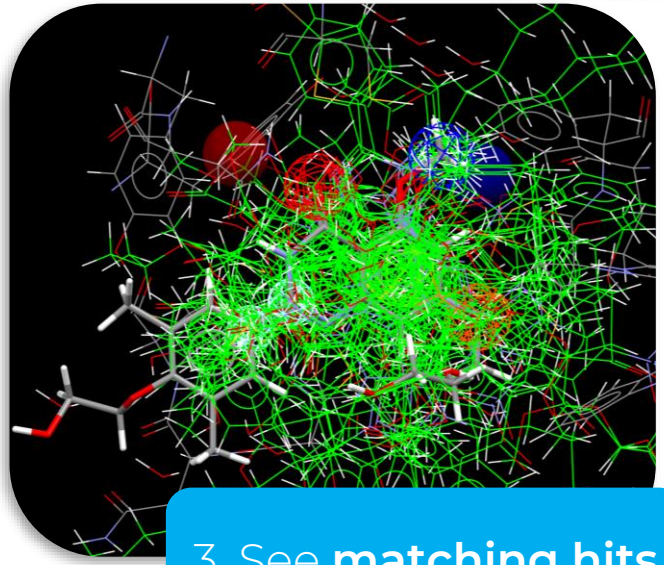
CSD-CrossMiner - workflow



1. **Observe** the protein-ligand complex.



2. **Define** the ligand and protein pharmacophore, and any constraints

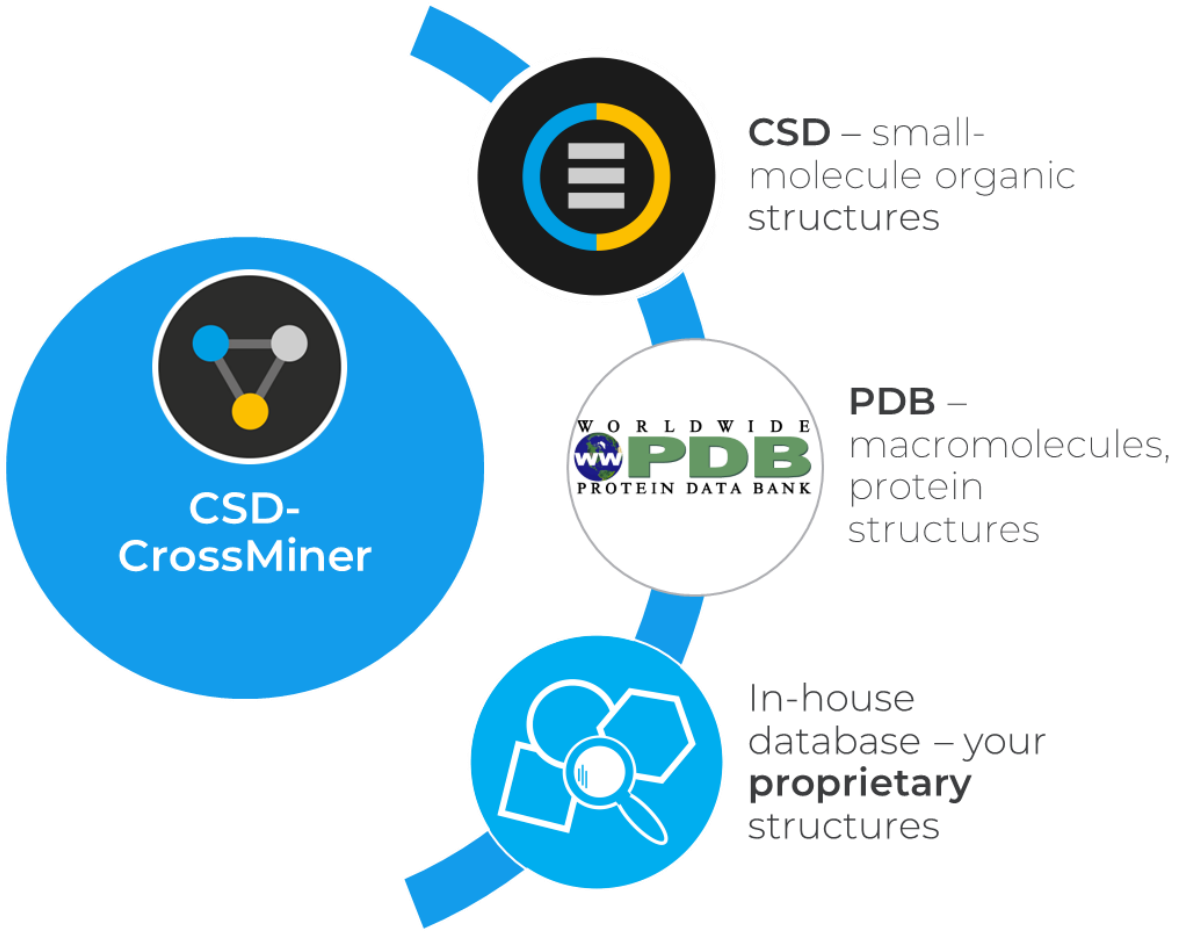


3. See **matching hits** overlaid – **explore** each in the browser.

HBD – Hydrogen Bond Donor; HBA – Hydrogen Bond Acceptor; HA – Heavy Atom;
RA – Ring Aromatic; ExV – Exit Vector

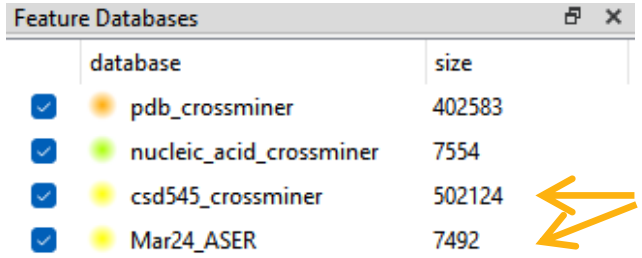
CCDC

CSD-CrossMiner: Feature Databases



https://www.ccdc.cam.ac.uk/solutions/software/csd-crossminer/

- **Simultaneously mine data sources.**
- Public and proprietary.
- Find matches by pharmacophore.
- These are **annotated versions** of the original databases, to include pharmacophore features information.



- Building of in-house databases and their integration into CSD-CrossMiner is possible.

CCDC

We will make the recording available to you in the next few days.

Agenda

- *Show One:* Intro and CSD and CSD-CrossMiner overview
- *Show One:* PDB overview
- *Show One:* CSD-CrossMiner explained and demonstration
- *Try One:* Hands-on example
- *Explore More:* Case-studies of CSD-CrossMiner
- *Explore More:* Quiz and Summary
- *Extra time:* More time for hands on and Q&A

Special Guest

Deborah Harrus

PDBe Archive Project Leader

CCDC

Agenda

We will make the recording available to you in the next few days.

- *Show One:* Intro and CSD and CSD-CrossMiner overview
- *Show One:* PDB overview
- *Show One:* CSD-CrossMiner explained and demonstration
- *Try One:* Hands-on example
- *Explore More:* Case-studies of CSD-CrossMiner
- *Explore More:* Quiz and Summary
- *Extra time:* More time for hands on and Q&A

CCDC

Show One: The basics – A pharmacophore

"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interaction with a specific biological target structure and to trigger (or block) its biological response."

C.G.Wermuth, C.R.Ganellin, P.Lindberg, L.A. Mitscher (1998). "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)" DOI: 10.1351/pac199870051129

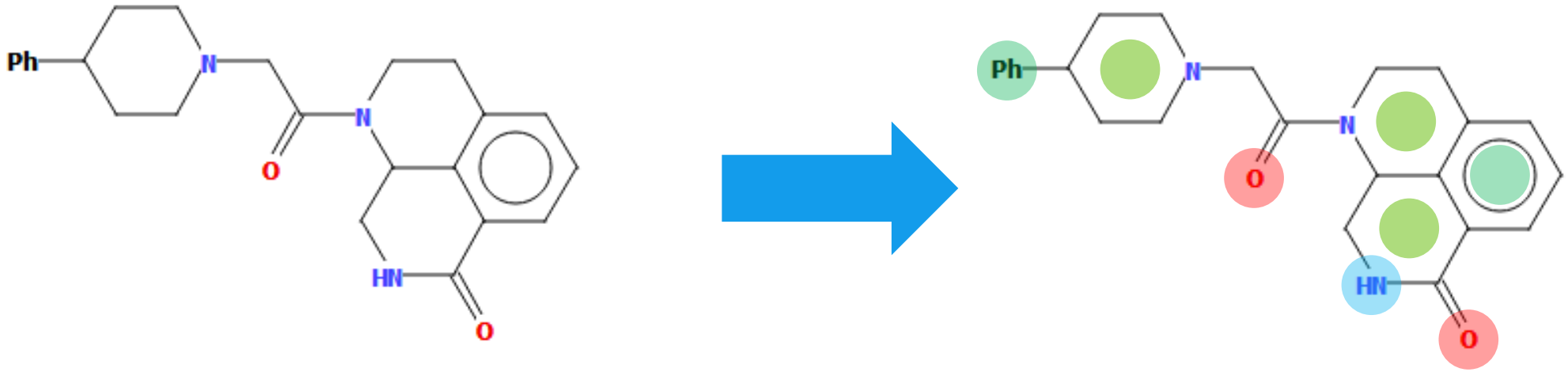
A concept is originated by Paul Ehrlich (1909) and developed by Schueler in his book "*Chemobiodynamics and Drug Design*" (1960)

Osman F. Güner and J. Phillip Bowen (2014). "Setting the Record Straight: The Origin of the Pharmacophore Concept" DOI: 10.1021/ci5000533

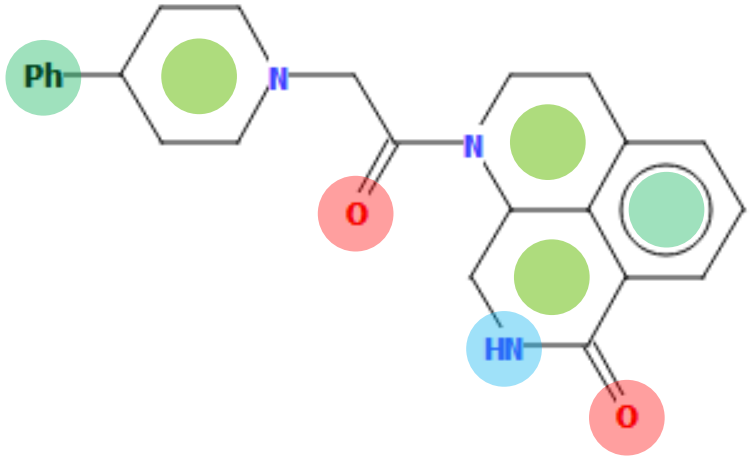
CCDC

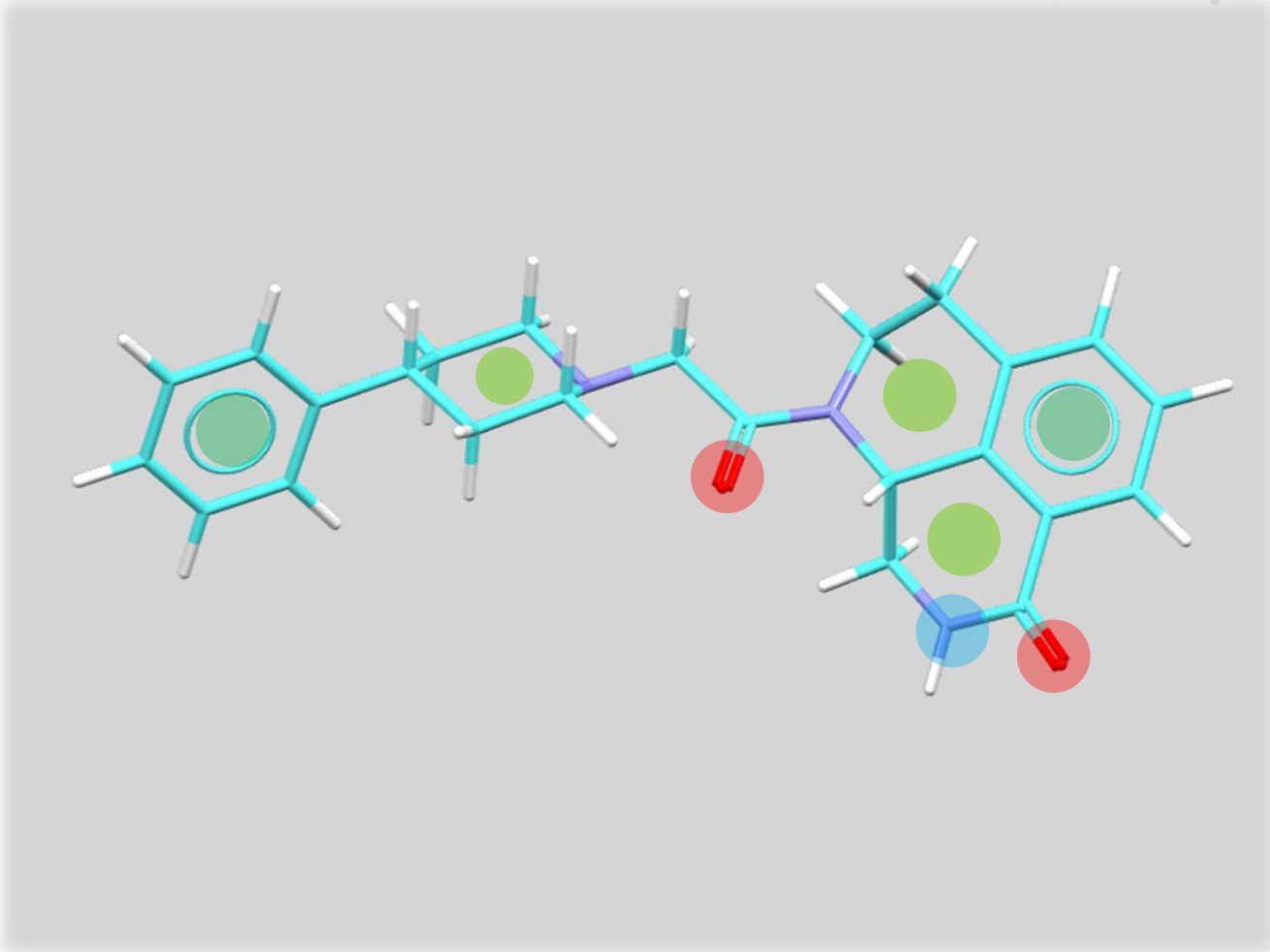
Pharmacophore feature

Pharmacophore feature – an atom or a group of atoms with sterical or electrostatic properties essential for the activity to a target protein.



Pharmacophore feature



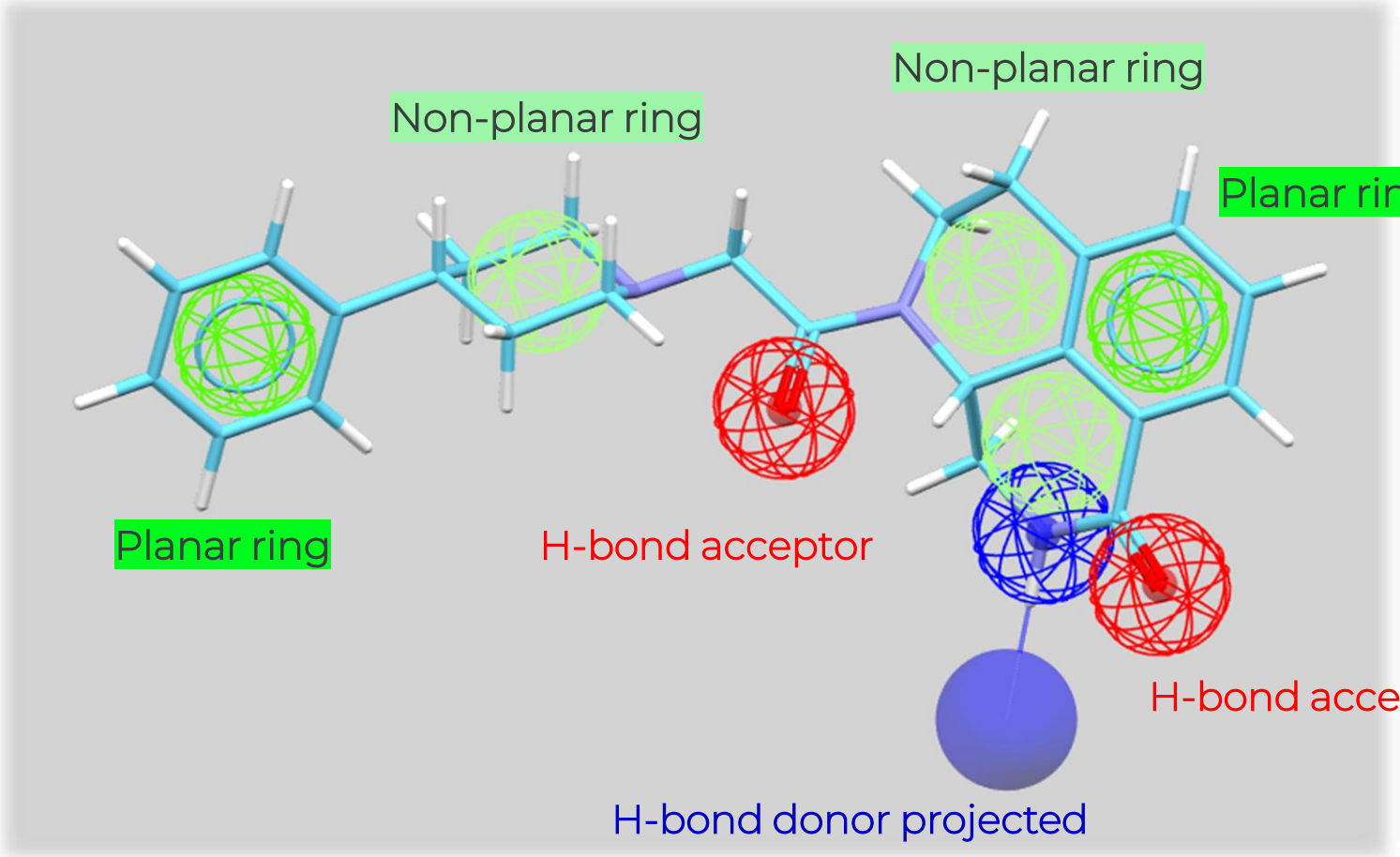


Pharmacophore features in CSD-CrossMiner

- acceptor
- acceptor_projected
- donor_ch_projected
- donor_projected
- heavy_atom
- hydrophobe
- ring
- ring_non_planar
- ring_planar_projected
- purine
- pyrimidine
- adenine
- cytosine
- guanine
- thymine
- uracil
- deoxyribose
- ribose
- exit_vector
- halogen
- bromine
- chlorine
- fluorine
- metal
- water

- ALA
- ARG
- ASN
- ASP
- CYS
- GLN
- GLU
- GLY
- HIS
- ILE
- LEU
- LYS
- MET
- PHE
- PRO
- SER
- THR
- TRP
- TYR
- VAL

- excluded_volume
- annotation_filter
- substructure_filter

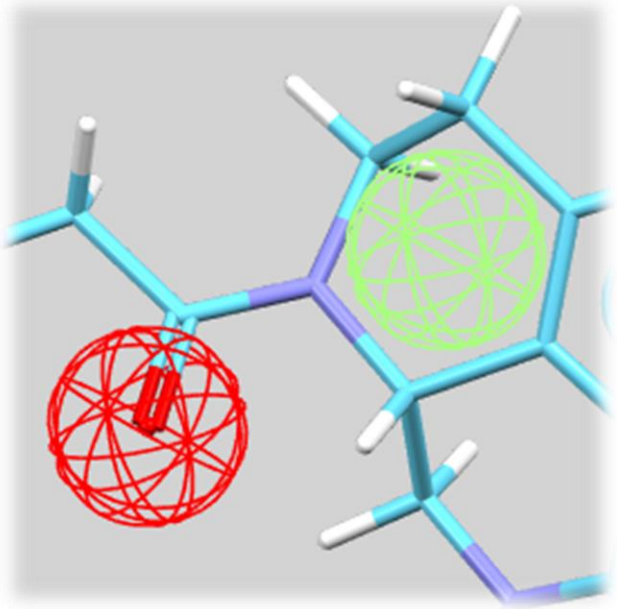


CCDC

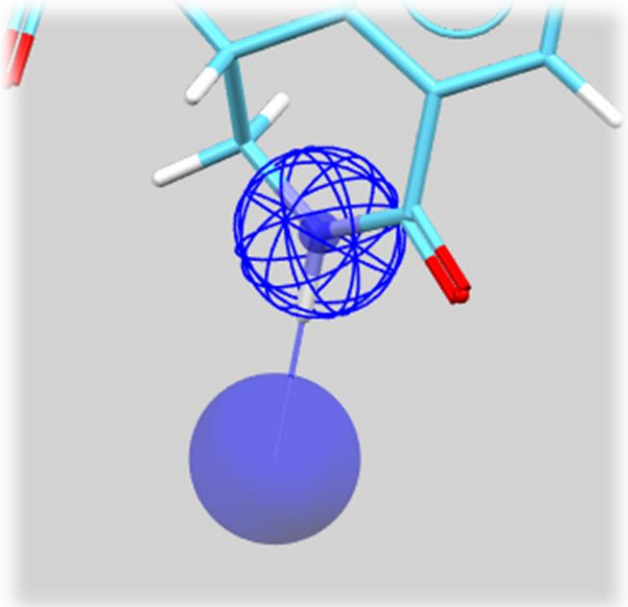
Pharmacophore features in CrossMiner

A pharmacophore point is a feature that has been selected to be a pharmacophore because its presence is necessary.

- Single point, acceptor

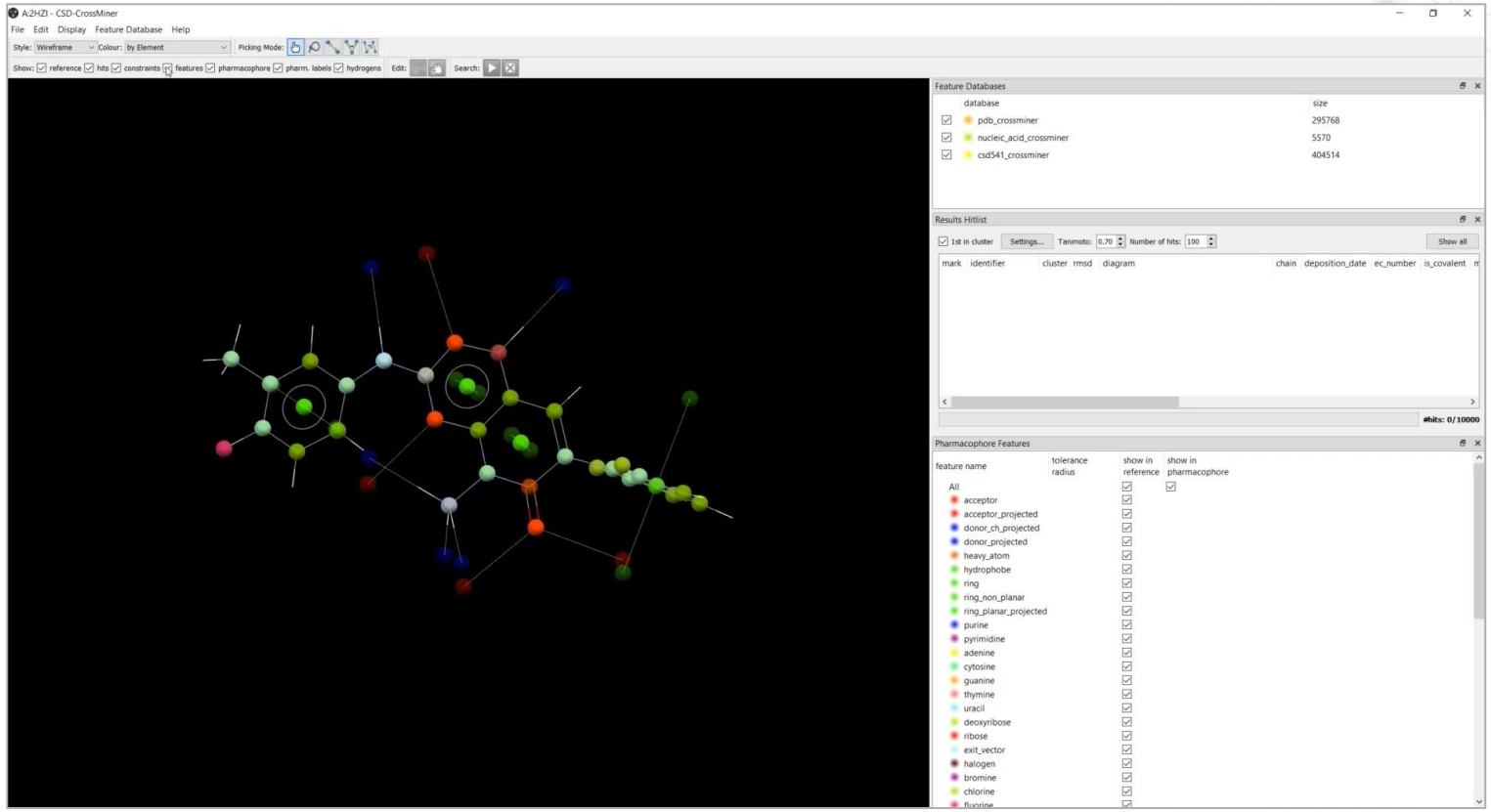


- Between two points, donor_projected



CSD-CrossMiner - features

- **Customise** feature definitions
  - Explore new chemistry
  - Improve search granularity
- **Excluded** volume
  - Cut noise by removing occupational volume



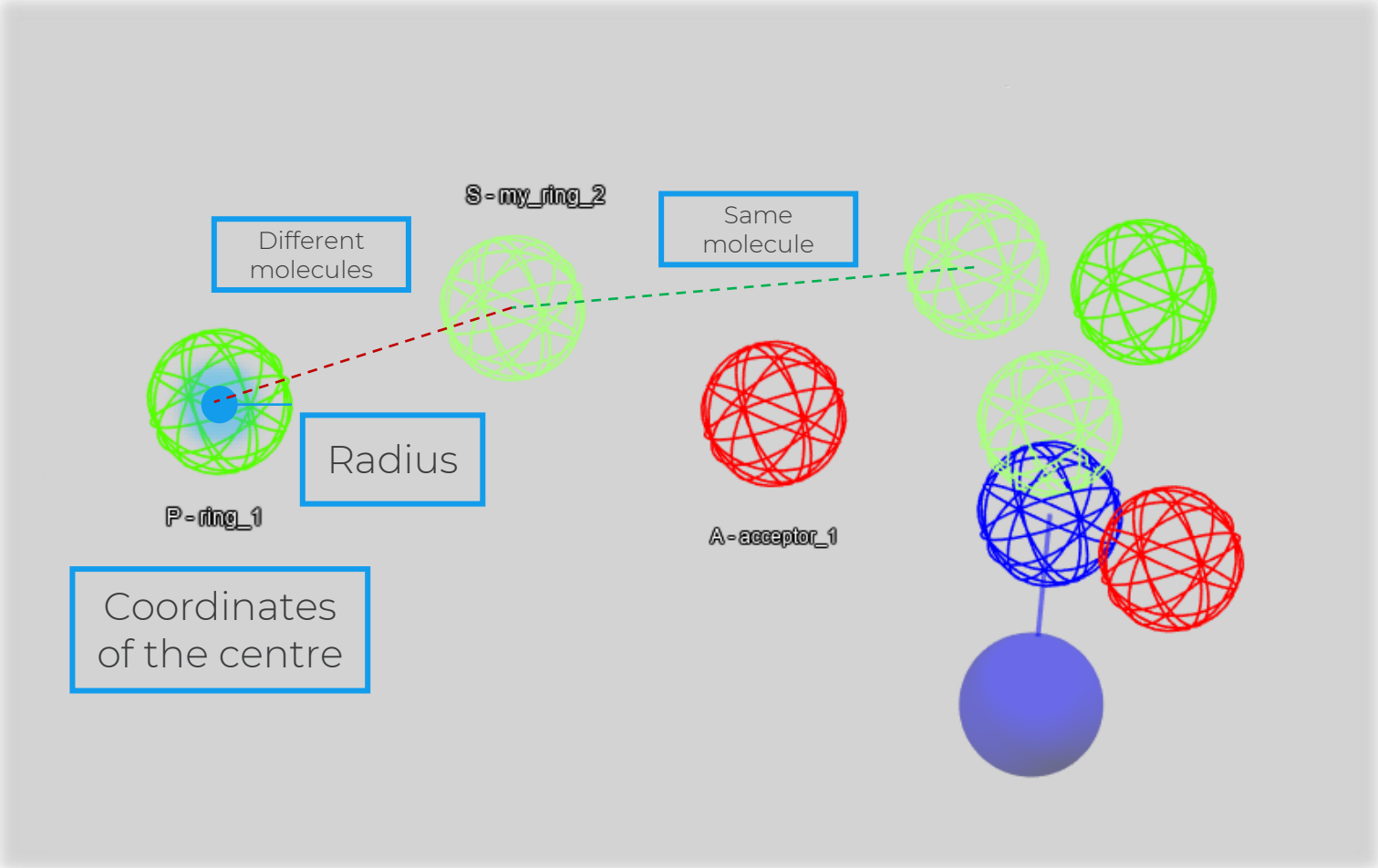
Pharmacophore query

Pharmacophore query (or just pharmacophore) is ensemble of steric and electronic features that characterise a protein and/or a small molecule.

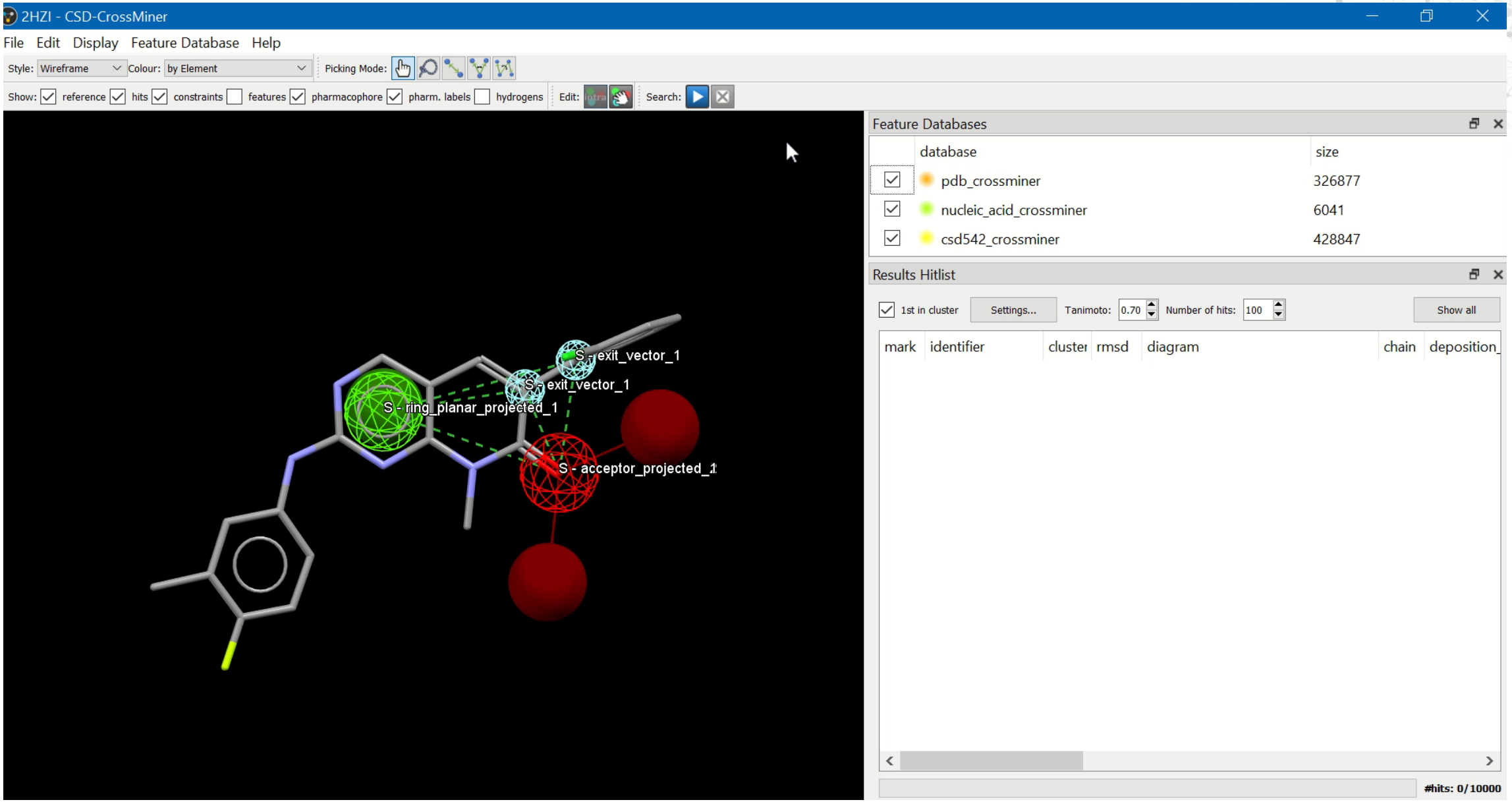
- What interactions are present and most common in protein-ligand complexes you have?
- Which pharmacophore features are the most common among known ligands binding to a same target?
- Are there any unsatisfied interactions in the target's pocket?
- Are there any interactions you'd like to eliminate?

CCDC

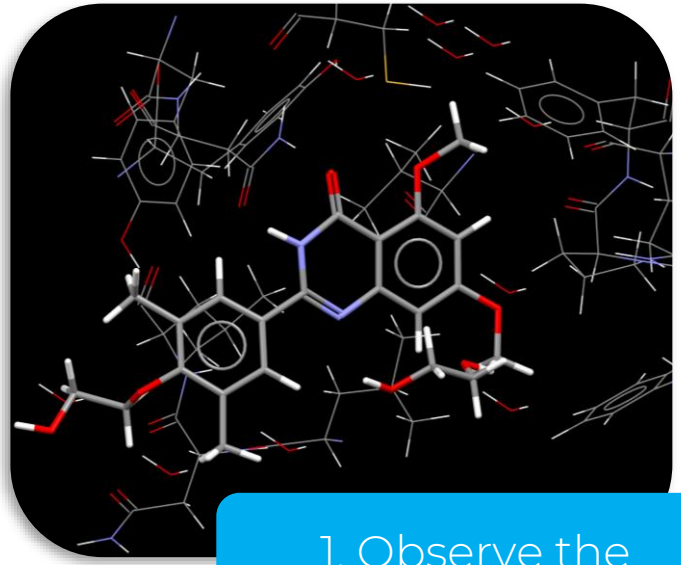
Pharmacophore query



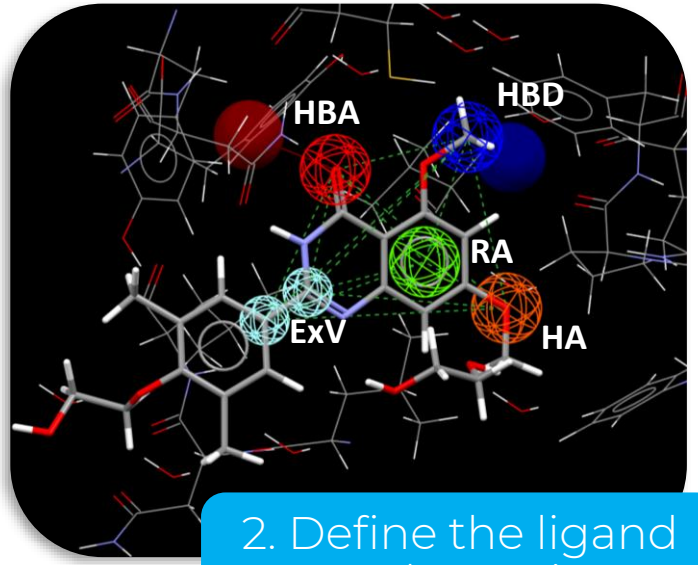
Run the search



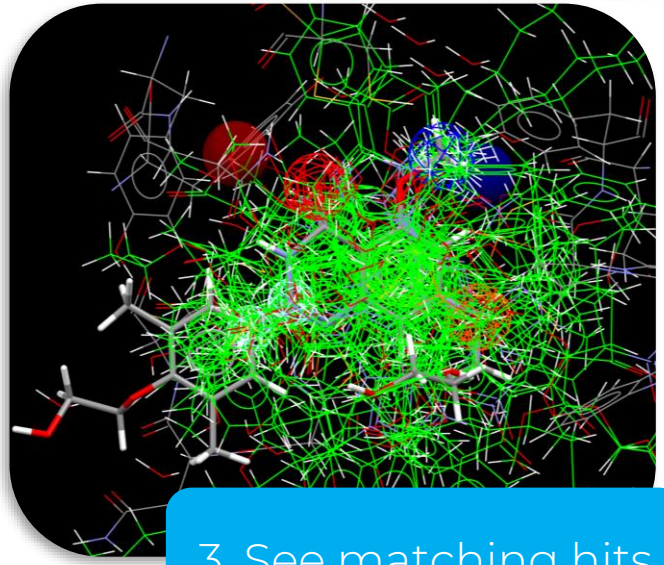
CSD-CrossMiner - workflow



1. Observe the protein-ligand complex.



2. Define the ligand and protein pharmacophore, and any constraints

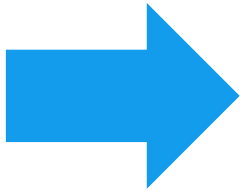


3. See matching hits overlaid – explore each in the browser.

HBD – Hydrogen Bond Donor; HBA – Hydrogen Bond Acceptor; HA – Heavy Atom; RA – Ring Aromatic; ExV – Exit Vector

CCDC

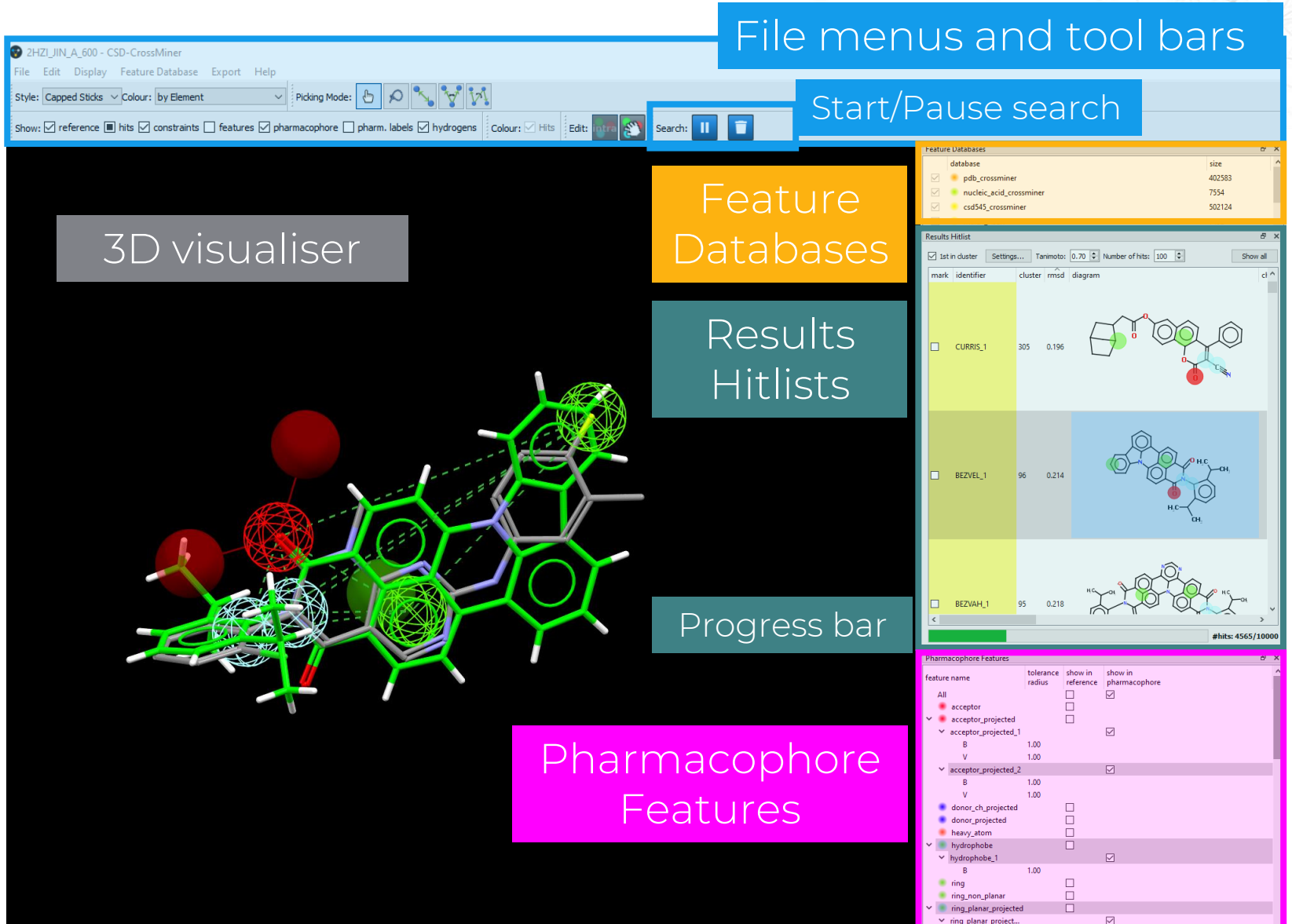
Show One: CSD-CrossMiner Demonstration



In this demo we will see how to:

- Navigate CSD-CrossMiner
- Perform a Pharmacophore Search

This is a good time to open CSD-CrossMiner!

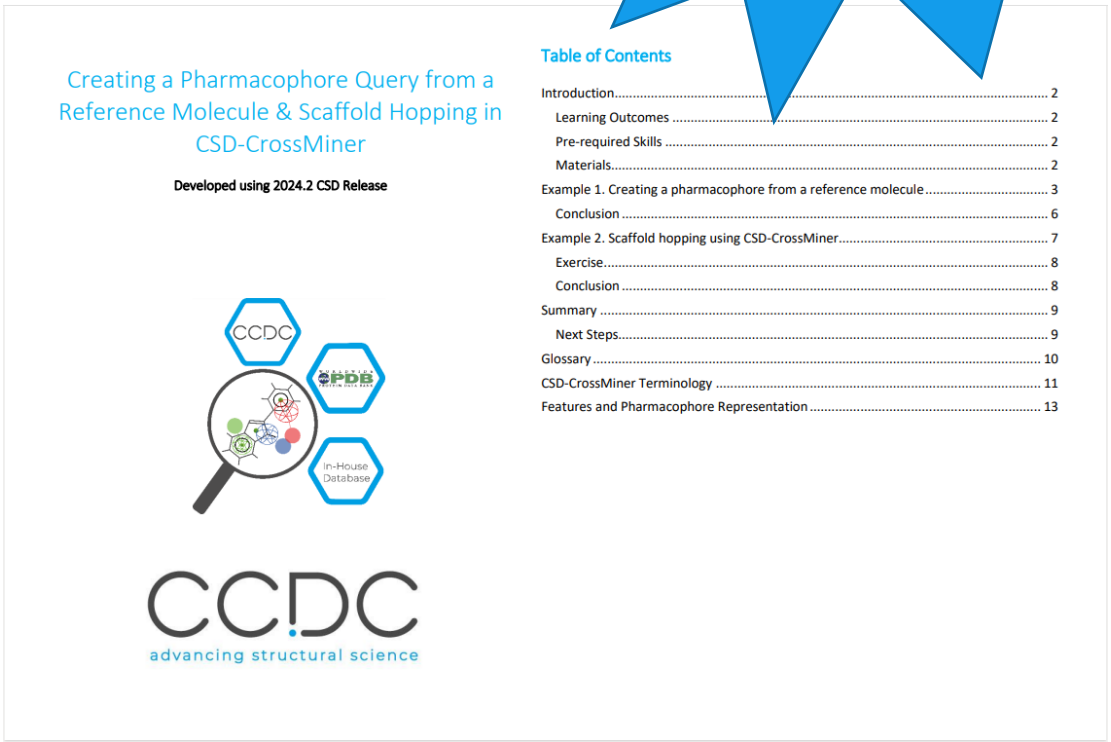


*Try One: H*ands-on exercise

We will make the recording available to you in the next few days.

It's your turn!

- Try the case study from the handout.
- Your tutors are on hand to help you!
- To ask questions during this time type a message in the chat box.
- If you finish early, ask us for more challenging examples.



Creating a Pharmacophore Query from a Reference Molecule & Scaffold Hopping in CSD-CrossMiner

Developed using 2024.2 CSD Release

CCDC

advancing structural science

Table of Contents

https://info.ccdc.cam.ac.uk/2024-autumn-virtual-workshop

CCDC

Agenda

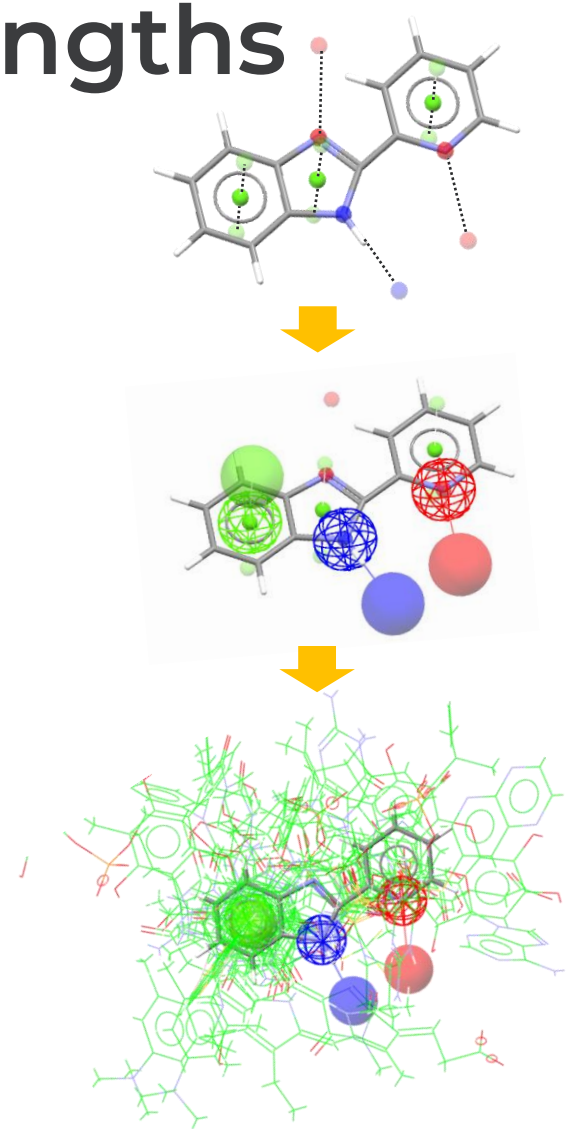
We will make the recording available to you in the next few days.

- *Show One:* Intro and CSD and CSD-CrossMiner overview
- *Show One:* PDB overview
- *Show One:* CSD-CrossMiner explained and demonstration
- *Try One:* Hands-on example
- *Explore More:* Case-studies of CSD-CrossMiner
- *Explore More:* Quiz and Summary
- *Extra time:* More time for hands on and Q&A

CCDC

CSD-CrossMiner - Strengths

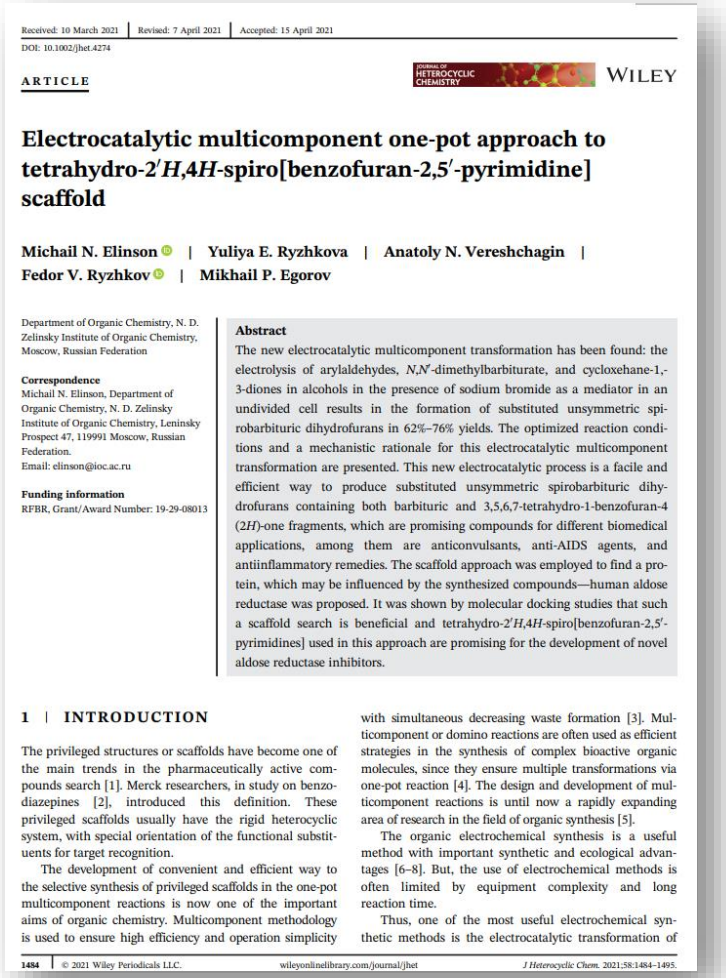
- **CSD** and **PDB** databases included
  - With optional **proprietary** data
- Intuitive **pharmacophore** searching
  - Build and modify queries around key pharmacophore features.
- **Interactive** – change your hypothesis on the fly
  - See results update instantly.



donor
acceptor
planar ring
hydrophobe

CCDC

Use-case: Developing novel inhibitors



Received: 10 March 2021 | Revised: 7 April 2021 | Accepted: 15 April 2021
DOI: 10.1002/jhet.4274

ARTICLE

JOURNAL OF HETEROCYCLIC CHEMISTRY WILEY

Electrocatalytic multicomponent one-pot approach to tetrahydro-2′*H*,4*H*-spiro[benzofuran-2,5′-pyrimidine] scaffold

Michail N. Elinson | **Yuliya E. Ryzhkova** | **Anatoly N. Vereshchagin** | **Fedor V. Ryzhkov** | **Mikhail P. Egorov**

Department of Organic Chemistry, N. D. Zelinsky Institute of Organic Chemistry, Moscow, Russian Federation

Correspondence
Michail N. Elinson, Department of Organic Chemistry, N. D. Zelinsky Institute of Organic Chemistry, Leninsky Prospect 47, 119991 Moscow, Russian Federation.
Email: elinson@ioc.ac.ru

Funding information
RFBR, Grant/Award Number: 19-29-08013

Abstract
The new electrocatalytic multicomponent transformation has been found: the electrolysis of arylaldehydes, *N*,*N*′-dimethylbarbiturate, and cycloxehane-1,-3-diones in alcohols in the presence of sodium bromide as a mediator in an undivided cell results in the formation of substituted unsymmetric spirobarbituric dihydrofurans in 62%–76% yields. The optimized reaction conditions and a mechanistic rationale for this electrocatalytic multicomponent transformation are presented. This new electrocatalytic process is a facile and efficient way to produce substituted unsymmetric spirobarbituric dihydrofurans containing both barbituric and 3,5,6,7-tetrahydro-1-benzofuran-4(2*H*)-one fragments, which are promising compounds for different biomedical applications, among them are anticonvulsants, anti-AIDS agents, and antiinflammatory remedies. The scaffold approach was employed to find a protein, which may be influenced by the synthesized compounds—human aldose reductase was proposed. It was shown by molecular docking studies that such a scaffold search is beneficial and tetrahydro-2′*H*,4*H*-spiro[benzofuran-2,5′-pyrimidines] used in this approach are promising for the development of novel aldose reductase inhibitors.

1 | INTRODUCTION

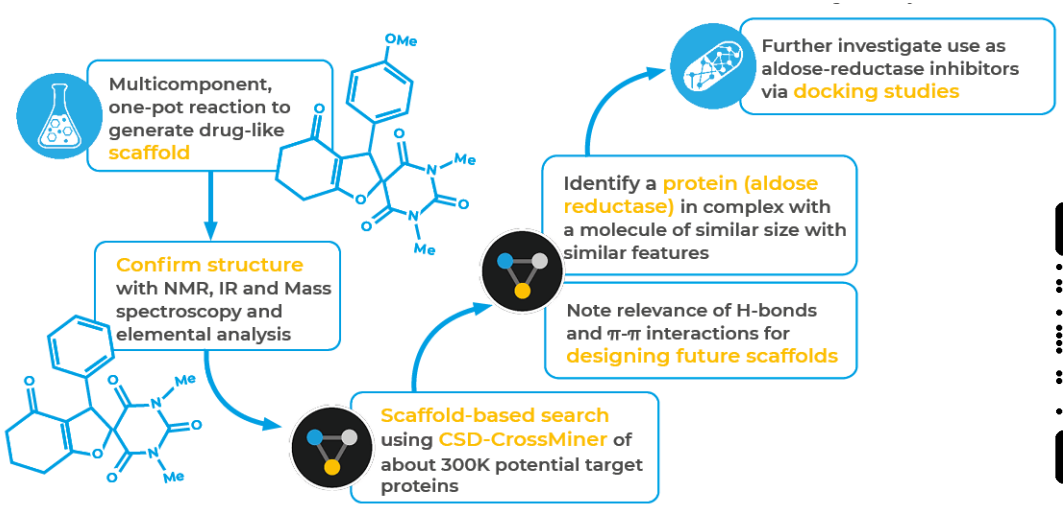
The privileged structures or scaffolds have become one of the main trends in the pharmaceutically active compounds search [1]. Merck researchers, in study on benzodiazepines [2], introduced this definition. These privileged scaffolds usually have the rigid heterocyclic system, with special orientation of the functional substituents for target recognition.

The development of convenient and efficient way to the selective synthesis of privileged scaffolds in the one-pot multicomponent reactions is now one of the important aims of organic chemistry. Multicomponent methodology is used to ensure high efficiency and operation simplicity with simultaneous decreasing waste formation [3]. Multicomponent or domino reactions are often used as efficient strategies in the synthesis of complex bioactive organic molecules, since they ensure multiple transformations via one-pot reaction [4]. The design and development of multicomponent reactions is until now a rapidly expanding area of research in the field of organic synthesis [5].

The organic electrochemical synthesis is a useful method with important synthetic and ecological advantages [6–8]. But, the use of electrochemical methods is often limited by equipment complexity and long reaction time.

Thus, one of the most useful electrochemical synthetic methods is the electrocatalytic transformation of

- Used a scaffold search in CSD-CrossMiner
- Searched for a target protein for spirobarbituric dihydrofuran compounds.
- Revealed two compounds that exhibited properties similar to the ones of known inhibitors of aldose reductase.

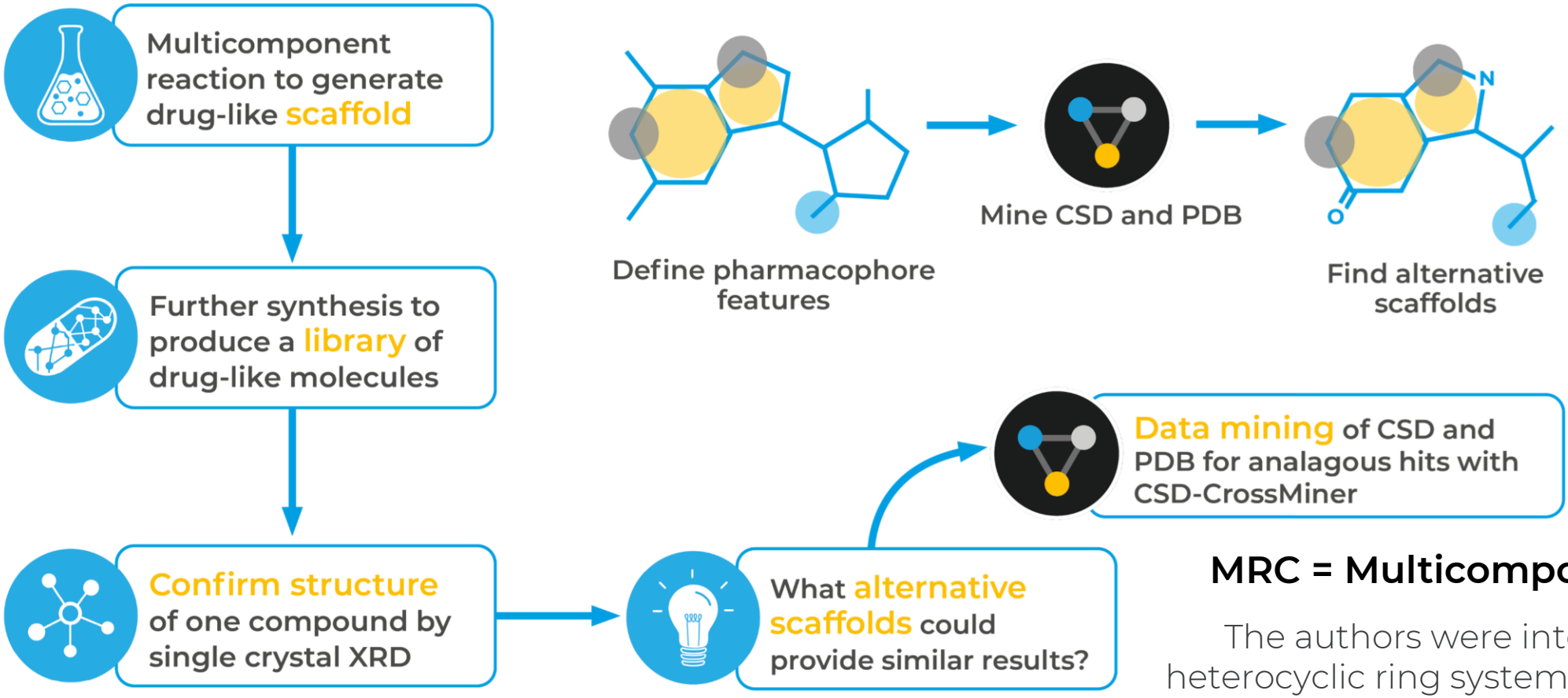


Michail N. Elinson, Yuliya E. Ryzhkova, Anatoly N. Vereshchagin, Fedor V. Ryzhkov, Mikhail P. Egorov, *J Heterocyclic Chem.* 2021, 58, 1484–1495. DOI: 10.1002/jhet.4274.

CCDC

Use-case: Exploring molecular scaffolds for a MCR library

Dr. Markella Konstantinidou *et al.*
EurJOC, 2020, 34: 5601-5605



https://www.ccdc.cam.ac.uk/discover/blog/crossminer-in-action-scaffolds/

MRC = Multicomponent reaction

The authors were interested in the heterocyclic ring system produced by the three component Groebke-Blackburn-Bienaymé reaction as a possible pharmaceutical scaffold.

Use-case: Catalyst ligand design

Catalysis Science & Technology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Catal. Sci. Technol.*, 2023, **13**, 2407

High-throughput computational workflow for ligand discovery in catalysis with the CSD†

Marc A. S. Short, [a] Clare A. Tovee, [c] Charlotte E. Willans [b] and Bao N. Nguyen [*b]

A novel semi-automated, high-throughput computational workflow for ligand/catalyst discovery based on the Cambridge Structural Database is reported. Two potential transition states of the Ullmann–Goldberg reaction were identified and used as a template for a ligand search within the CSD, leading to >32 000 potential ligands. The $\Delta G^{\ddagger}$ for catalysts using these ligands were calculated using B97-3c//GFN2-xTB with high success rates and good correlation compared to DLPNO-CCSD(T)/def2-TZVPP. Furthermore, machine learning models were developed based on the generated data, leading to accurate predictions of $\Delta G^{\ddagger}$, with 70.6–81.5% of predictions falling within ± 4 kcal mol^{-1} of the calculated $\Delta G^{\ddagger}$, without the need for the costly calculation of the transition state. This accuracy of machine learning models was improved to 75.4–87.8% using descriptors derived from TPSS/def2-TZVP//GFN2-xTB calculations with a minimal increase in computational time. This new workflow offers significant advantages over currently used methods due to its faster speed and lower computational cost, coupled with excellent accuracy compared to higher-level methods.

Received 16th January 2023,
Accepted 20th March 2023

DOI: 10.1039/d3cy00083d

rsc.li/catalysis

1 Introduction

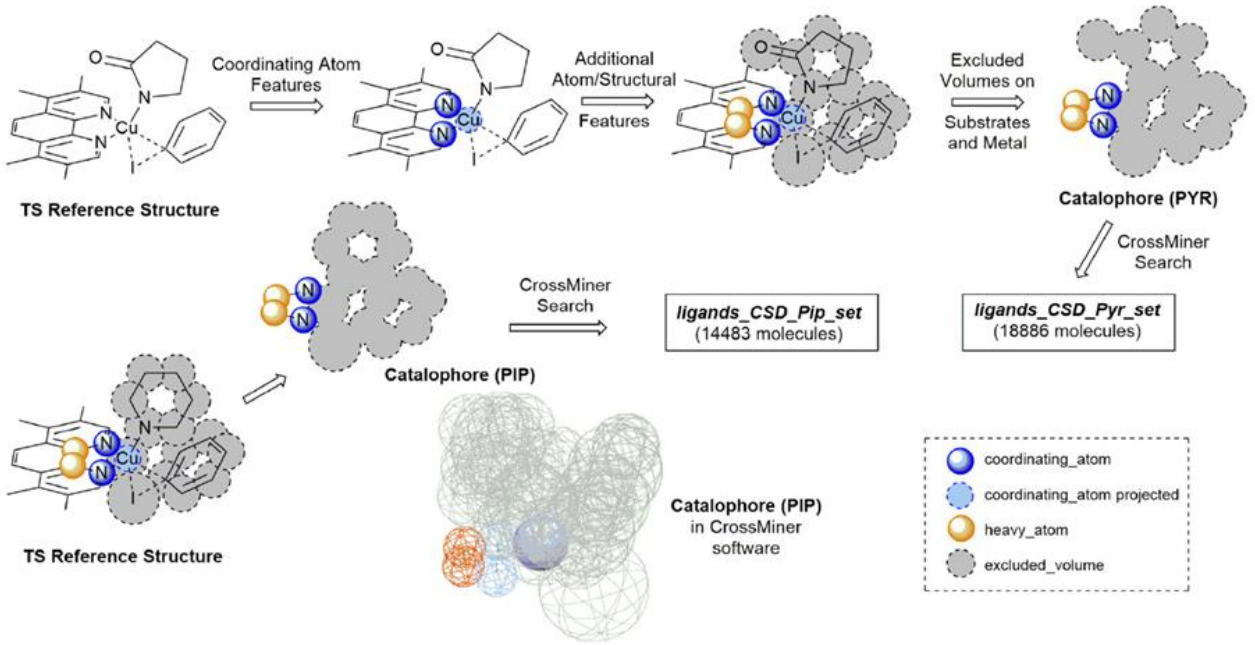
The development of organometallic catalysts, and suitable ligands, is a key challenge in the area of catalysis. While the process for traditional precious metals, such as Pd, Ru and Rh, is well established based on extensive mechanistic understanding and data-based approaches,[1–7] ligand design for base metal catalysts is still a nascent area of research and needs to balance many more catalytic and catalyst decomposition pathways.[8–10] Properties such as activity, selectivity and stability are the most common criteria when selecting a ligand, but solubility, toxicity and cost are also important properties to consider.[11] Recent applications of data science to catalysis have highlighted the computer-guided search for optimal ligands and reaction conditions as a major technology which can significantly progress this field of research.[12,13]

While high-throughput experimental approaches have proven effective at finding suitable ligands from libraries and optimising reaction conditions,[14–17] these are limited by the available ligand libraries. *In silico* ligand exploration allows faster access to the entire chemical space and can lead to the discovery of unexpected ligands. In addition, new developments in high-throughput computational techniques,[18–21] and cheminformatics tools can underpin additional filters such as ligand cost/complexity, toxicity and availability for a variety of applications in different chemical sectors.[22–24] However, research in this field has been hampered by a lack of suitable tools for the automated exploration of ligand space, while taking into account synthetic feasibility of the ligands.[13,25–27] In this paper, we report an alternative approach which leverages the extensive Cambridge Structural Database (CSD) and its tools to explore ligand space in a relevant catalytic reaction. This has the benefit of avoiding the synthetic feasibility challenge completely, while still maintaining a very wide chemical space coverage.

The approach was demonstrated with the copper(I)-catalysed Ullmann–Goldberg reaction, an important C–N cross-coupling reaction which has been highlighted by pharmaceutical companies as a desirable synthetic tool in the near future due to its mild conditions compared to the palladium-catalysed counterpart and sustainability credentials.[28] Despite this level of interest, the Pd-catalysed Buchwald–Hartwig coupling reaction is still preferred due to its reliability and better-developed ligands. Several different reaction mechanisms have been

[a] School of Chemical and Process Engineering, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK
[b] School of Chemistry, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK. E-mail: b.nguyen@leeds.ac.uk
[c] The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK. E-mail: tovee@ccdc.cam.ac.uk

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3cy00083d

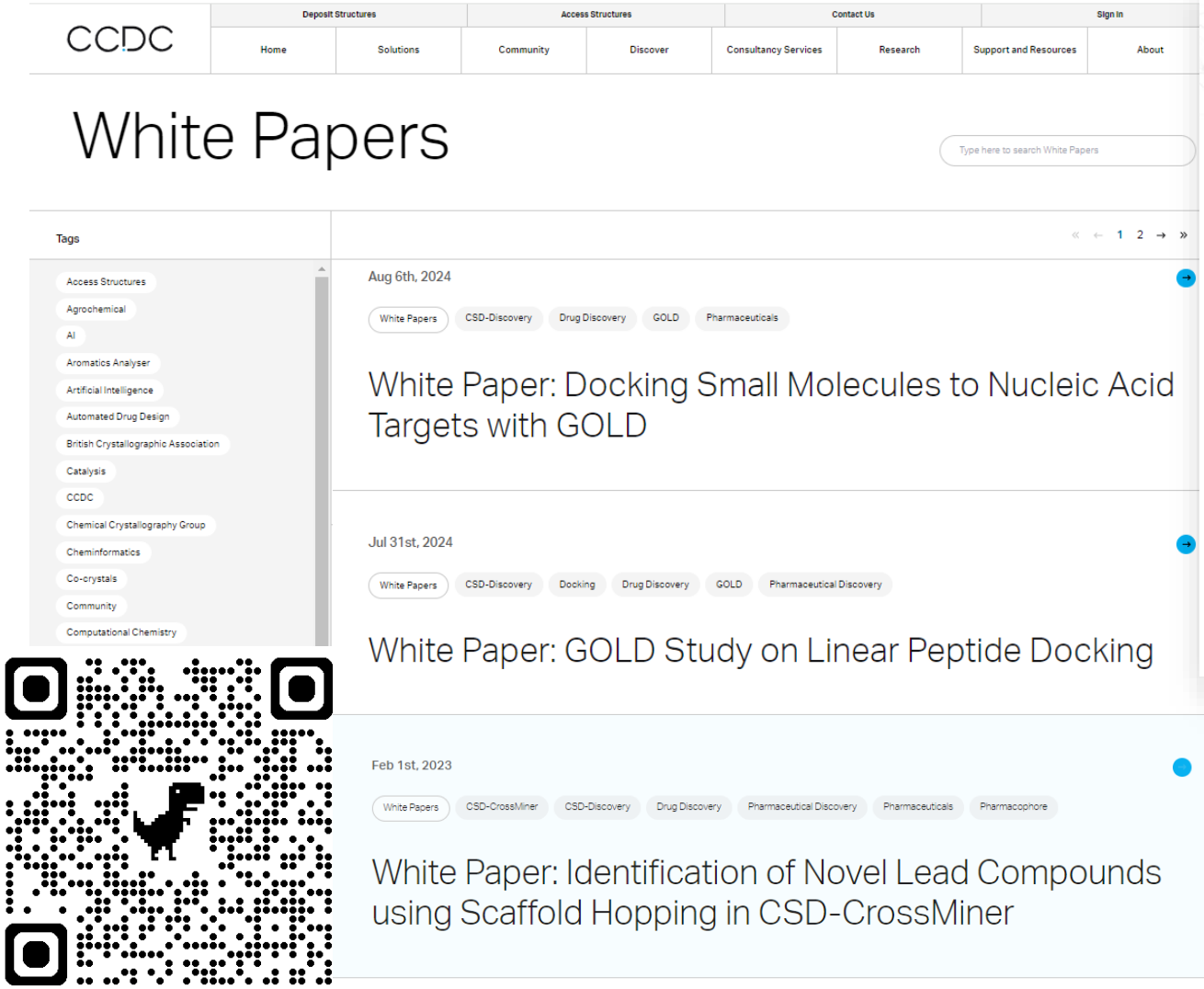


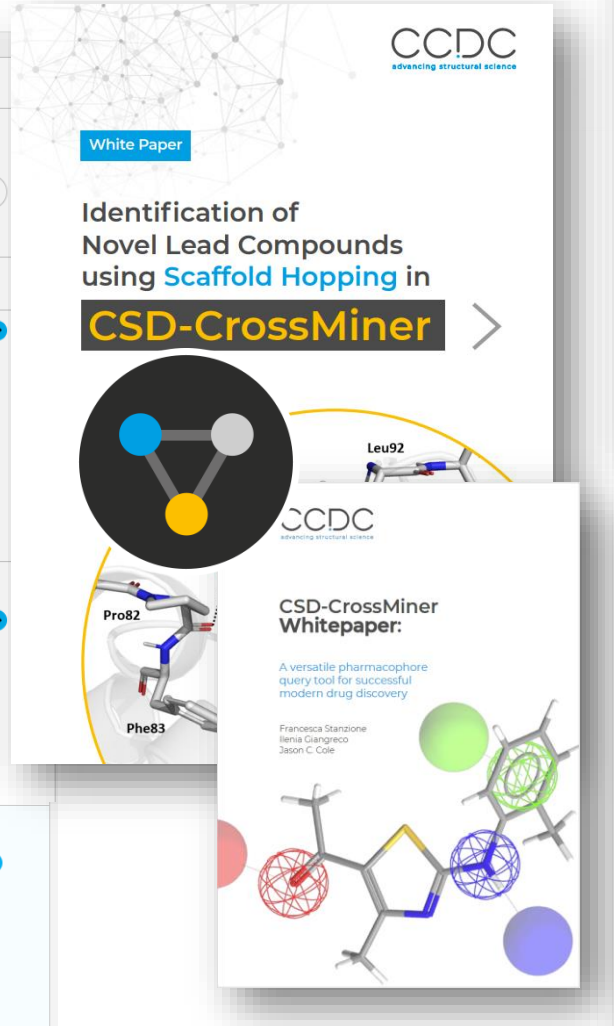
Workflow for the identification of ligands present in the CSD and for the generation of a catalophore from a transition state reference structure.

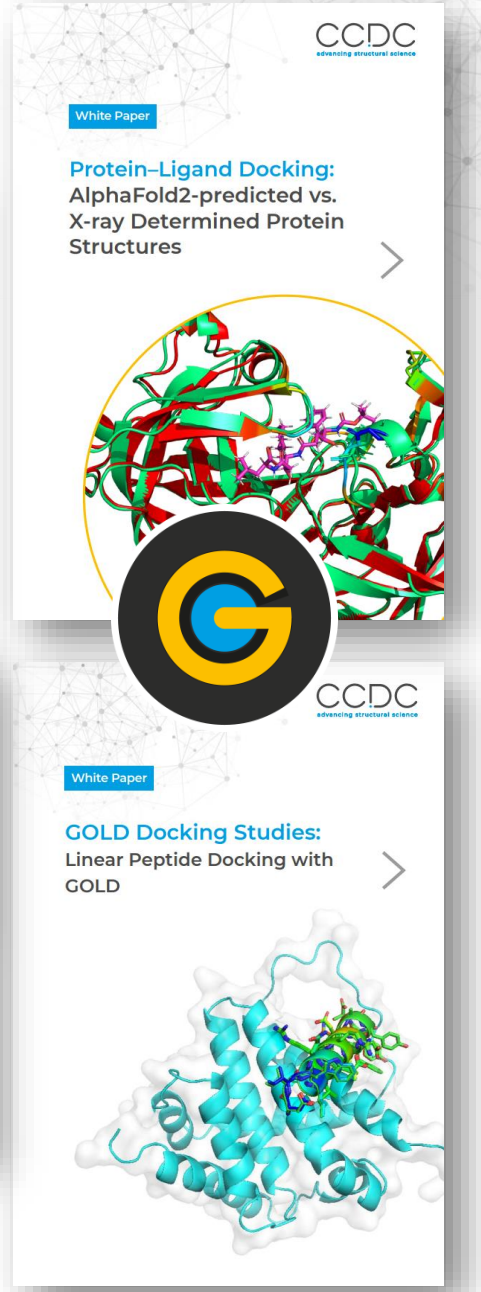
Marc A. S. Short, Clare A. Tovee, Charlotte E. Willans and Bao N. Nguyen, Catal. Sci. Technol., 2023,13, 2407-2420. DOI: 10.1039/d3cy00083d

CCDC

A collection of white papers



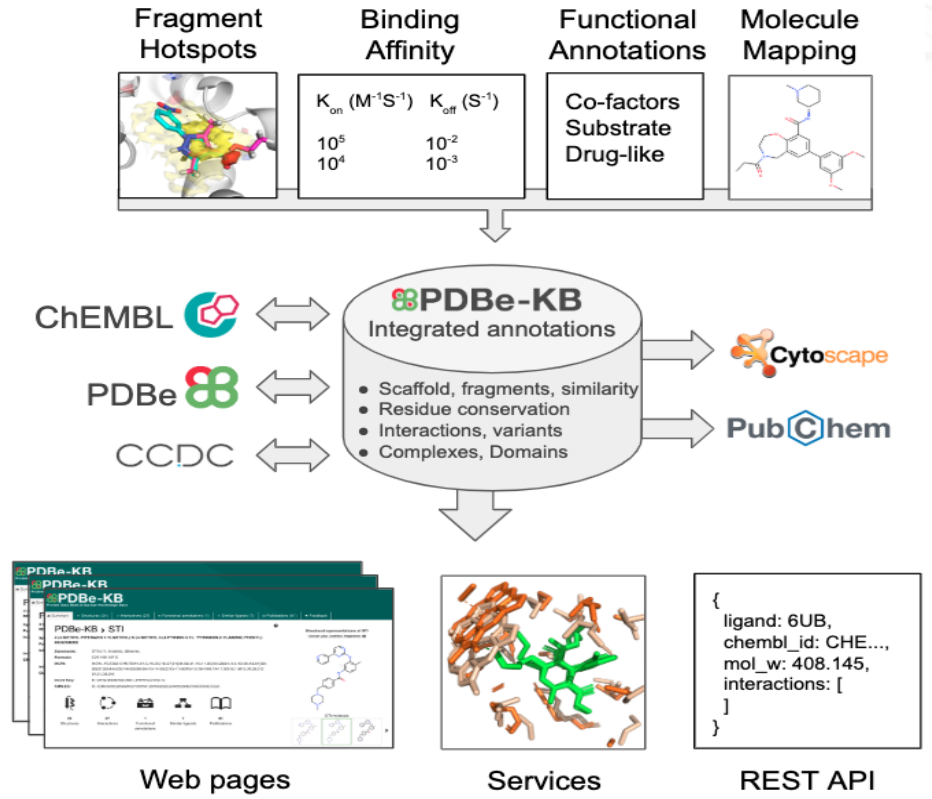




https://www.ccdc.cam.ac.uk/discover/whitepapers/

New release

- A collaboration between the CCDC, PDBe and ChEMBL.
- The release integrates structural, functional, and biochemical data.
  - Links over **235,000 CSD entries** with ChEMBL, PDBe, and other resources via UniChem.
- Accelerates research in drug discovery, drug repurposing, target validation, and understanding of drug mechanism.



https://www.ebi.ac.uk/about/news/updates-from-data-resources/biochemgraph-data/

What have we learnt

- ✓An understanding of the pharmacophore concept.
- ✓Familiarity with the CSD-CrossMiner interface and how it can be used in your research.
- ✓How to perform pharmacophore searches biologically relevant subsets of the CSD and PDB.
- ✓How to set up an interactive pharmacophore query and save the results.
- ✓How to analyse and interact with your results.

Get a certificate and help us improve

Do you want to earn a completion certificate for this session?

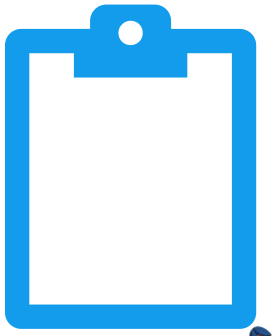
1. Complete the test
2. Score 10/10
3. Receive your certificate within a couple of hours from passing the test.

- Test available for **48h** (until **Thursday at 12:15 pm GMT)!**

How did it go today?

What do you want to learn next?

- Let us know in the polls, in the chat or via email!



If the polls do not pop-up on your screen, you can find it in the chat box. To vote, select your answer(s) and then click Submit.

Find the link for the test in the chat box or in our next email

CCDC

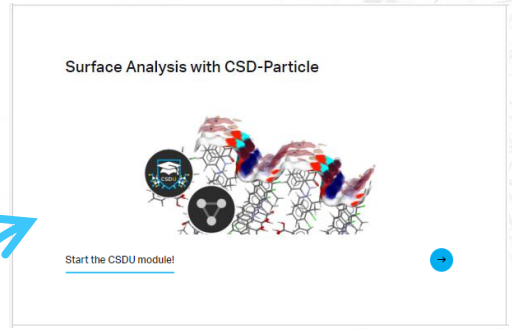
Want to explore more?

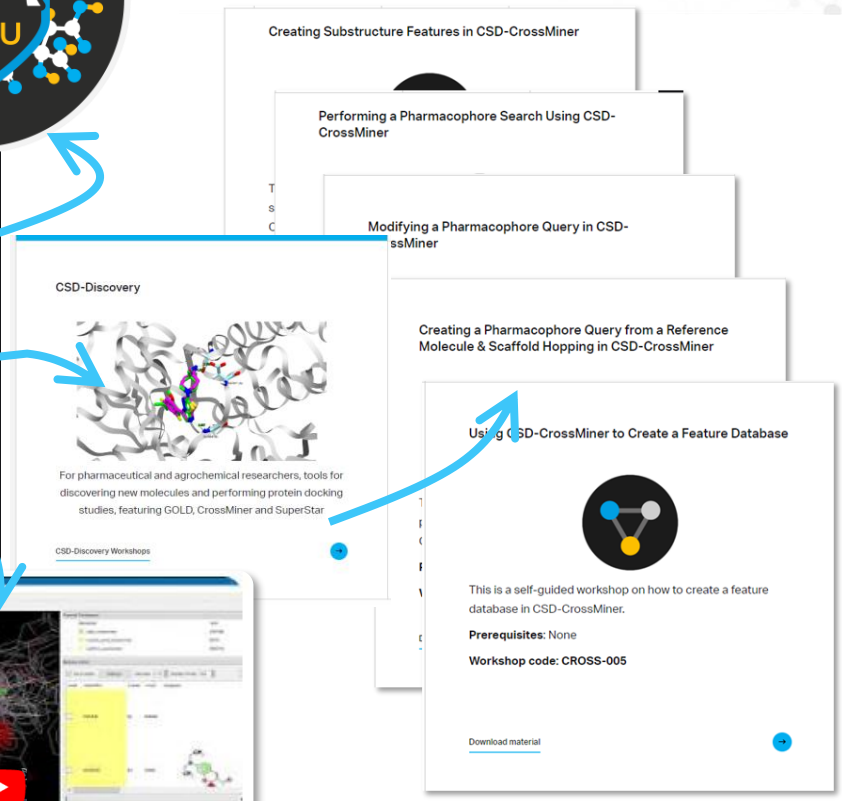
On-demand training resources

Solutions | Community | Discover | Consultancy

- Access/Deposit Structures
- CCDC for the Community
- Education and Outreach
- Events
- Free Products
- Training and Learning

- Free Online Courses
- Self-Guided Workshops
- Training and Support Videos





CCDC

Free online training courses

With completion certificates!

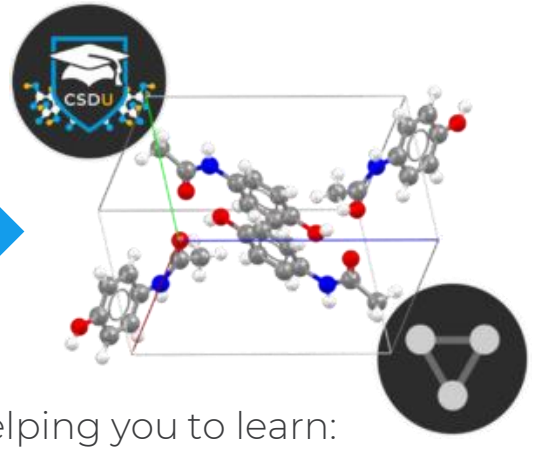
CSDU

On-demand modules to learn how to use the CSD software at your own pace.

UWatch · UTry · UTest

- Mercury
- Python API
- WebCSD
- GOLD Docking
- Mogul
- FIMs
- Particle
- Coming soon! Cross Miner

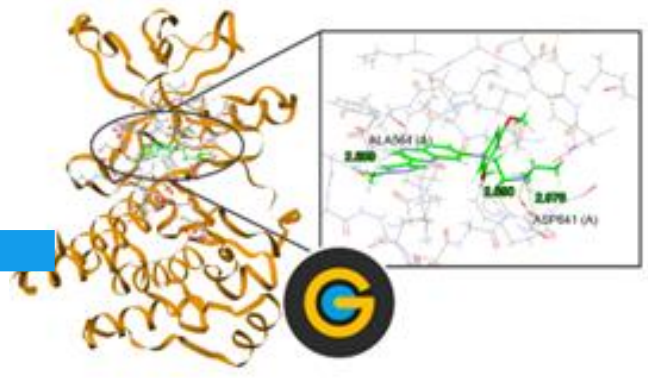
Visualization 101



Helping you to learn:

- The basics of Mercury software.
- How to explore and pack structures.
- How to create high resolution images.

GOLD Docking

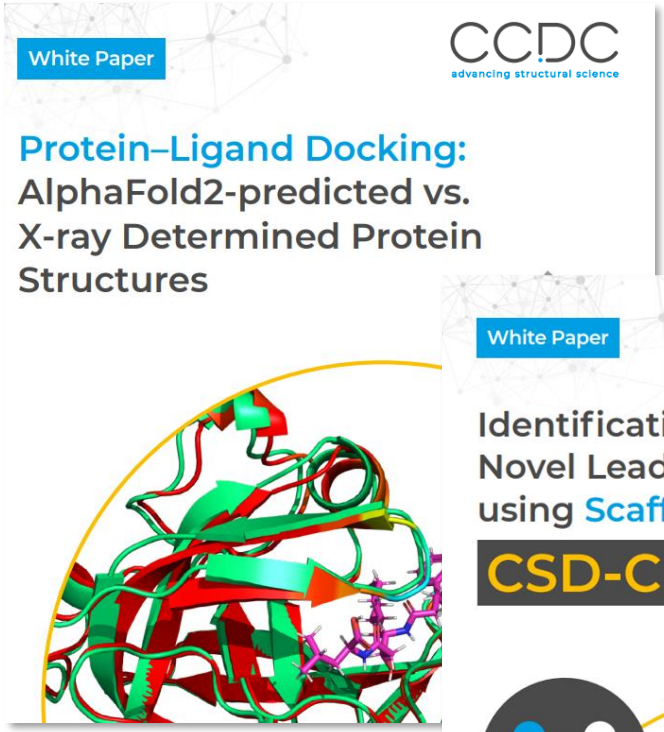


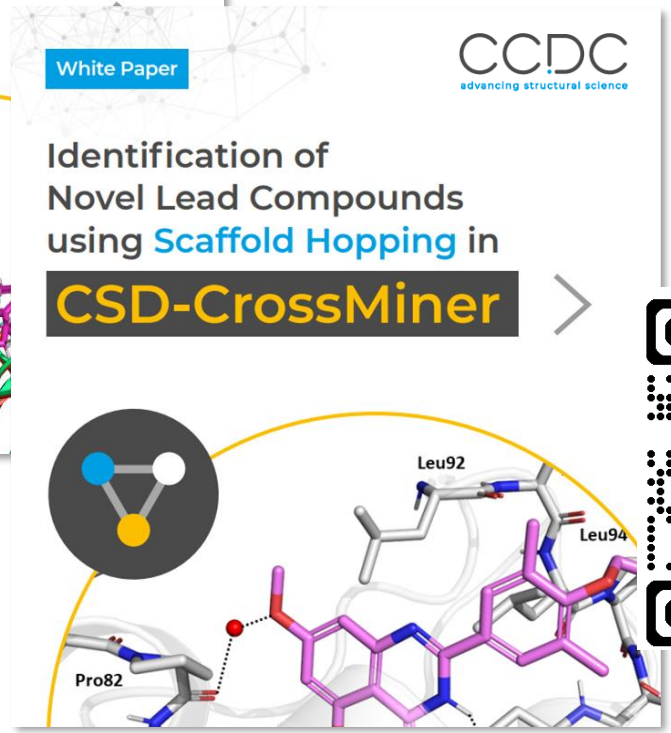
Helping you to learn:

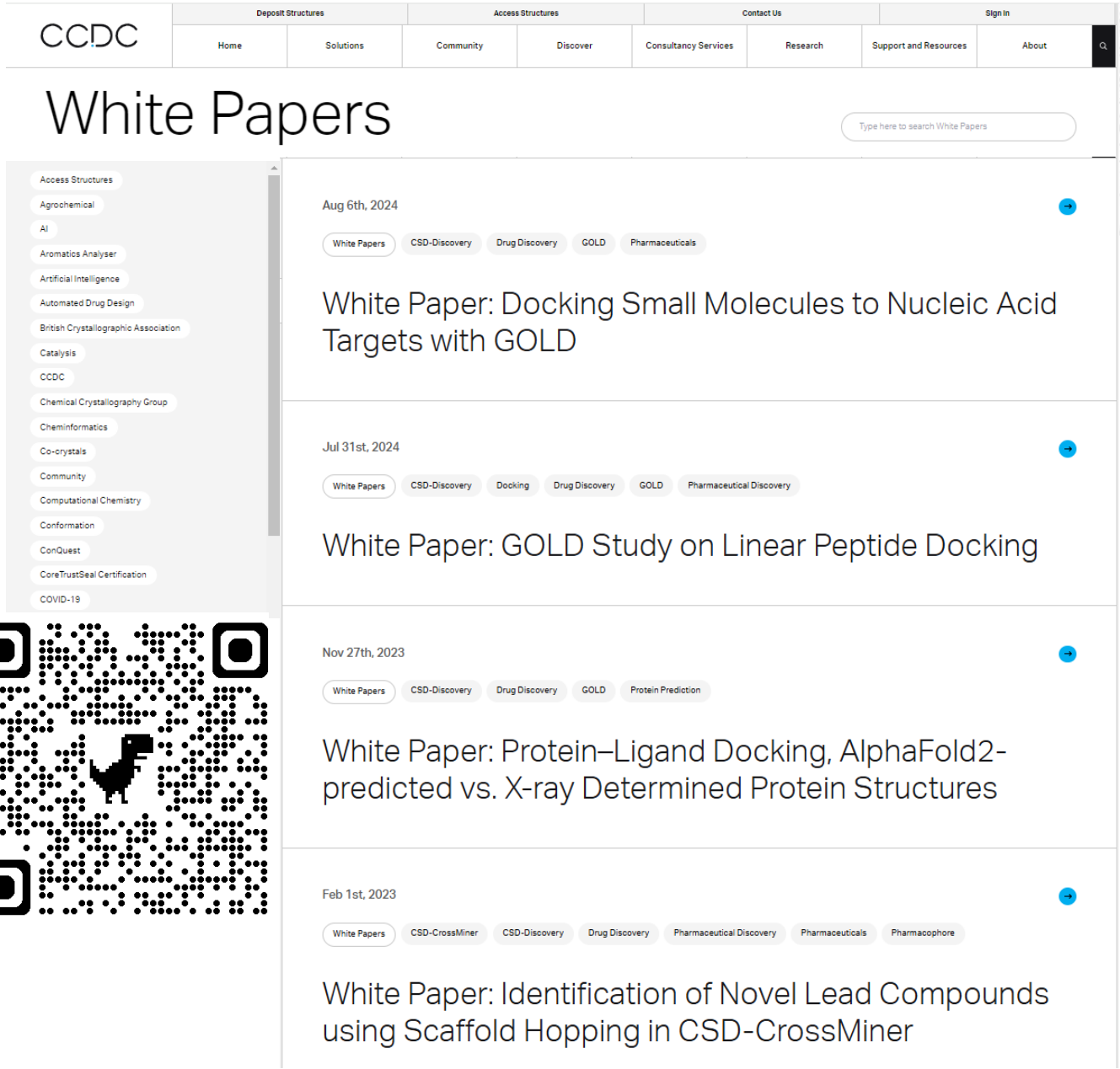
- The basics of Hermes.
- The basics of protein-ligand docking.
- Step-by-step guide on how to use GOLD to perform protein ligand docking.
- Where to get started with your docking simulation.

https://www.ccdc.cam.ac.uk/community/training-and-learning/csdu-modules/

A collection of white papers







https://www.ccdc.cam.ac.uk/discover/whitepapers/

More learning events

CHEMAI Virtual Satellite Event

- **27th Nov** Unlocking **CSD data** for Functional Materials innovation.

CCDC Webinars

- **21st Nov** How To Use **New Semiconductor Data** in the CSD for Research and Design.

CCDC Virtual Workshops

- Coming back next year! Sign-up to our newsletter and follow us on social media.

Find all CCDC events and workshops on our website!

https://www.ccdc.cam.ac.uk/community/events/

CCDC