CCDC

advancing structural science

CSD-Discovery

Knowledge-based computational approaches for drug discovery

Rupesh Chikhale
MPharm., PhD., MRSC

Research and Application Scientist

CCDC
advancing structural science

Agenda

- CCDC Introduction
- Knowledge-based methods overview
  - Hermes
  - GOLD
  - Superstar
  - CrossMiner
  - CSD-Python API
- Q&A

CCDC

The Cambridge Crystallographic Data Centre

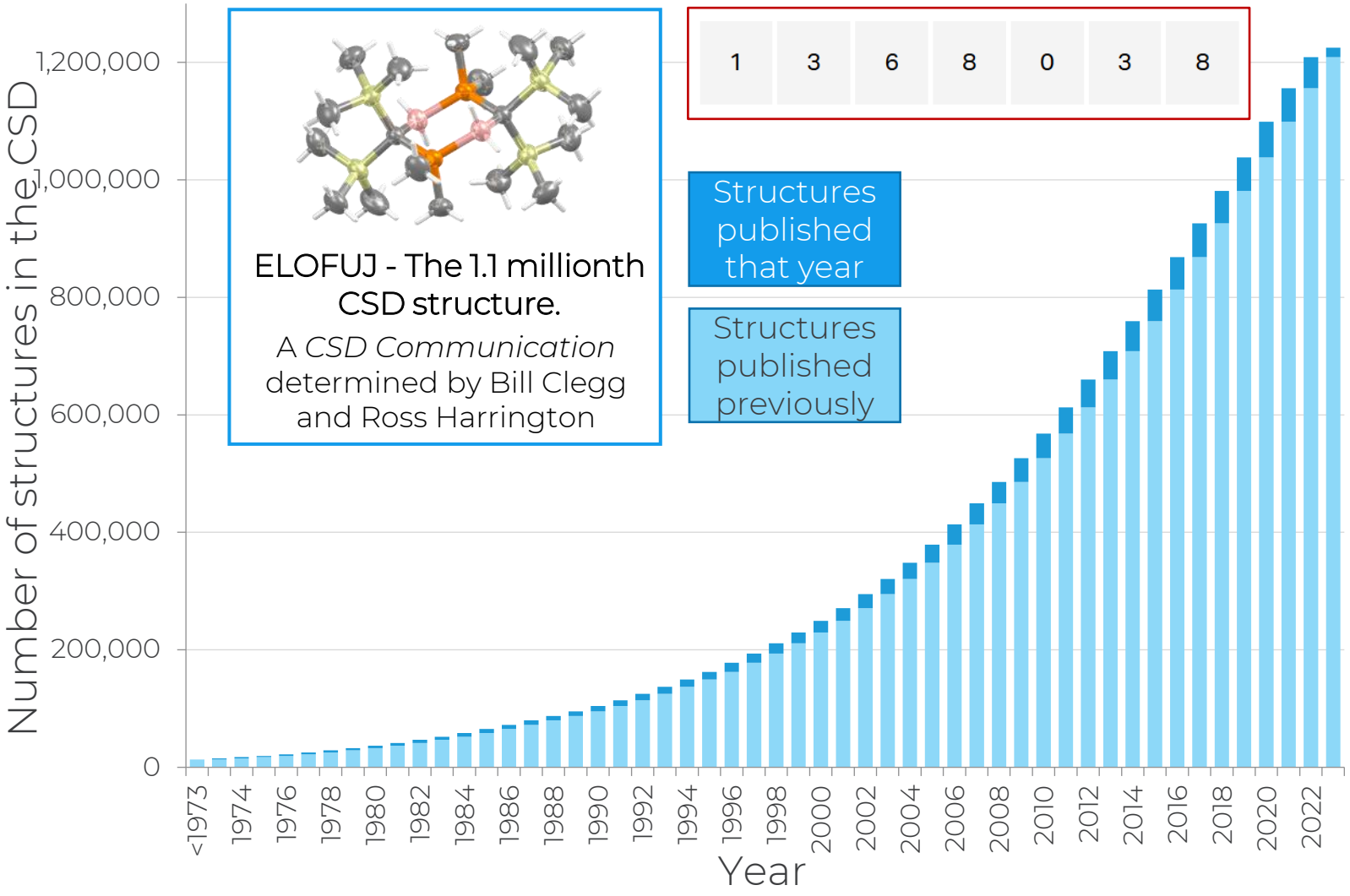
Established in 1965, Olga Kennard
Department of Chemistry, University of Cambridge

What we do
Charitable Objective:
Advancement of chemistry and crystallography for the public benefit

- International repository of 3D curated structures
- Collaborative research and knowledge-based services
- Scientific software
- Education and outreach

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 57 years
- A trusted CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

Our knowledge-based software solutions

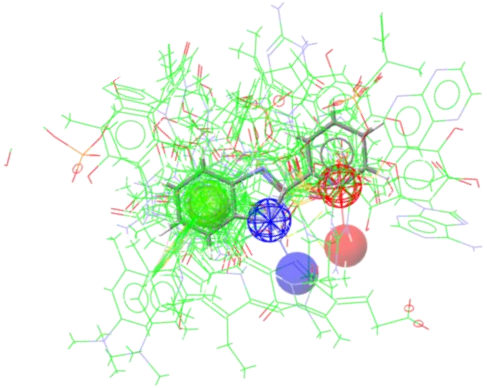
CSDCore.

Search, visualise, analyse and communicate structural data

Insights into molecular and crystal shape and interactions

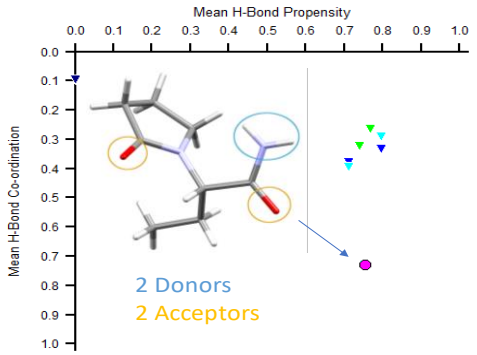
CSDDiscovery.

Design of new molecules



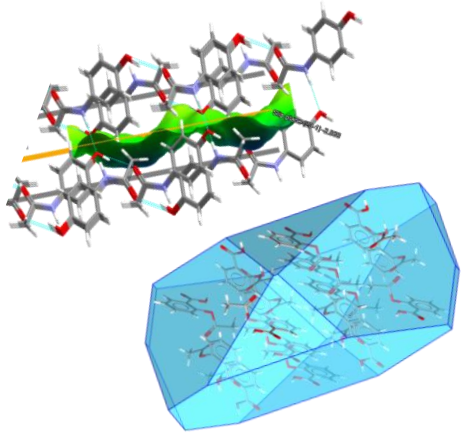
CSDMaterials.

Assessment of solid form stability and properties



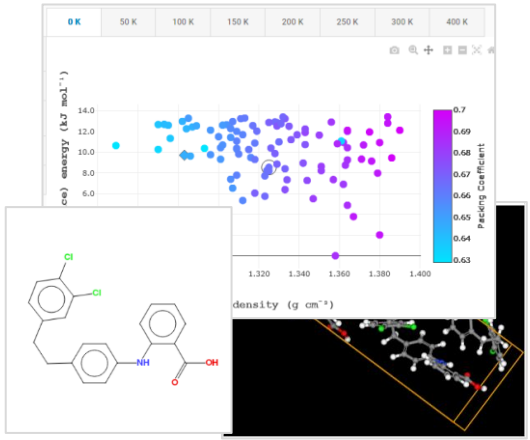
CSDParticle.

Anticipate particle properties and behaviour



CSDTheory.

Insights from predicted structure landscapes



Medicinal & Computational Chemists ♦ Crystallographers & Structural Biologists ♦ Solid Form & Crystallisation Scientists

CCDC

CSD-Discovery overview

CSD-Discovery.

GOLD: Protein-ligand docking and virtual screening

SuperStar: Analyse, predict and understand protein and ligand interactions

CSD-CrossMiner: Interrogate the CSD and the PDB for common interaction patterns

Ligand-based **virtual screening** workflow to find new hits.

CSD-Conformer Generator: Generation of molecular conformations.

CSD-Ligand Overlay: Flexible alignment of ligands.

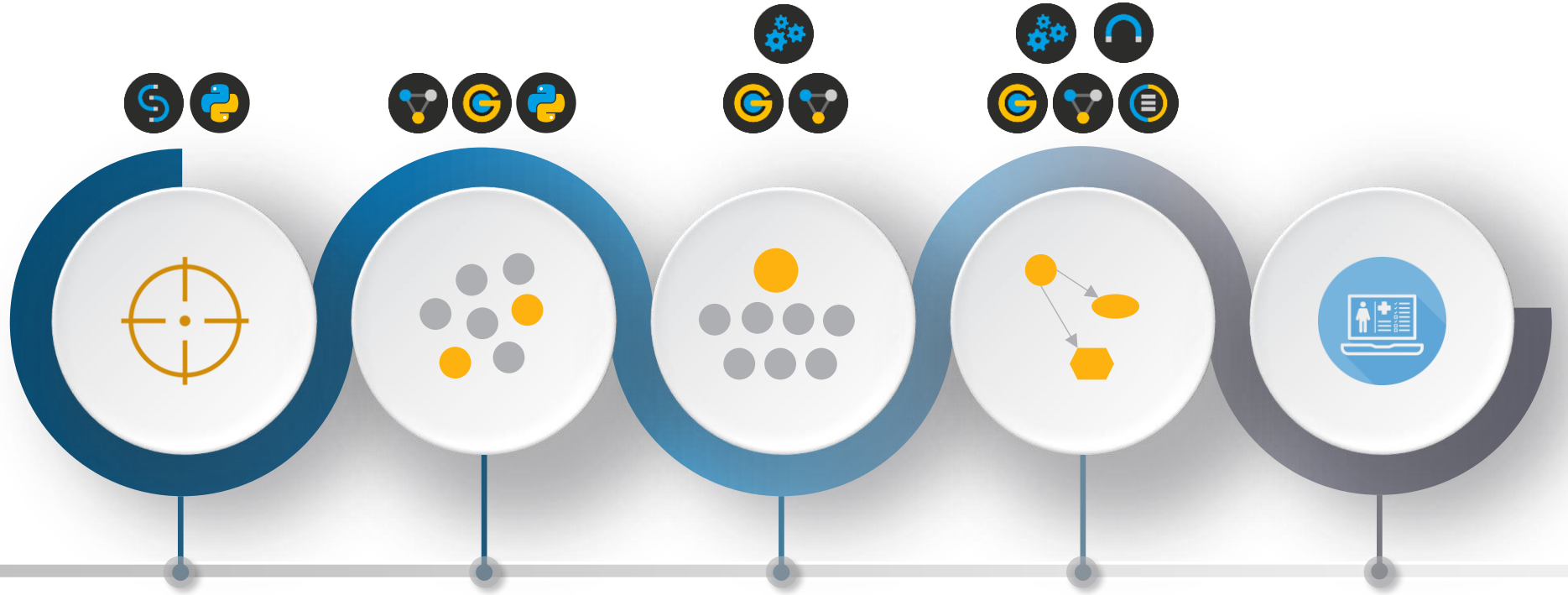
CSD Python API: Create CSD-driven analyses and workflows.

Script-based interfaces to the **Field-Based Ligand Screener**; protein **cavity and sub-pocket search** and comparison; **protein-ligand substructure & interaction pattern mining.**

CCDC

CSD-Discovery in drug discovery

Knowledge-based tools for drug design



TARGET SELECTION

HIT IDENTIFICATION

Structure- based virtual screening.

HIT TO LEAD

Assess how changes affect binding.
Optimize compound geometry.
Predict binding of small molecules to active pockets in proteins.

LEAD OPTIMISATION

Check the impact of changes with docking pose prediction.
Understand how changes affect conformations.

DRUG DEVELOPMENT

CCDC

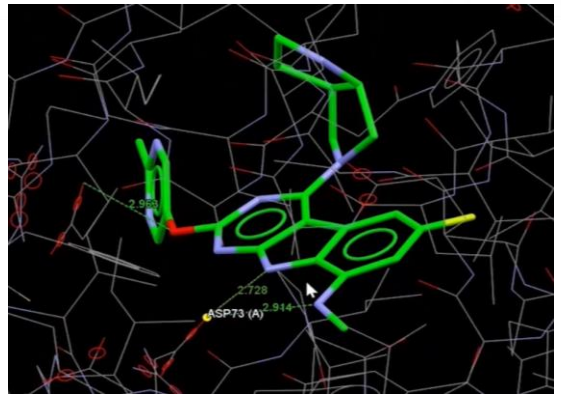
Hermes: Macromolecules visualizer and interface for most CSD-Discovery

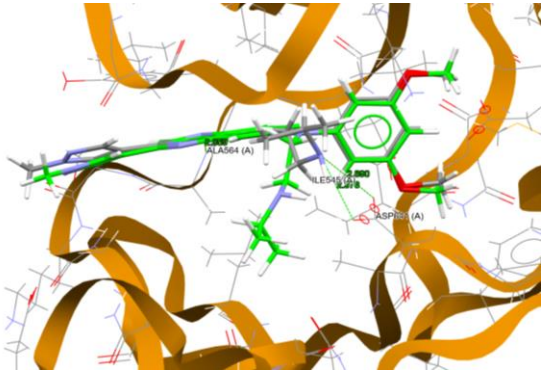
CCDC

Hermes – Structure visualisation

With Hermes you can:

- Visualise and edit macromolecules in 3D including proteins, nucleic acids, antibodies and ligands.
- Navigate around the structure and modify the appearance to display ribbons, etc.
- Interface to GOLD, Mogul, SuperStar, the CSD Ligand Overlay, and descriptors for GOLD docking poses.
- Generate publication quality images and share sessions for effective scientific communication.





Hermes is protein visualization software and is part of CSD-Core and CSD-Enterprise

CCDC

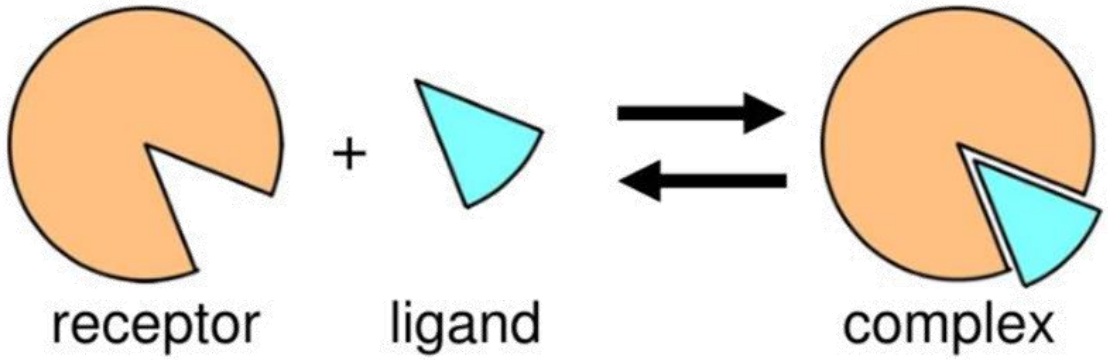
Hermes functionality

- 3D visualisation options:
  - Display styles, colours, labelling schemes, hide/unhide atoms, residues, ligands.
- Read in and prepare protein and ligand structures from external files:
  - Edit, prepare and add hydrogens, etc.
- The ability to:
  - Load and visualise contoured surfaces.
  - Overlay protein structures by least squares overlay or sequence alignment.
  - Measure and display distances, angles and torsion angles.
  - Find and display hydrogen bonds and nonbonded clashes, and customise how they are defined geometrically.
- Prepare and save publication quality displays

GOLD: Protein-ligand docking

CCDC

Docking



Docking studies are computational techniques for the exploitation of the possible binding modes of a substrate to a given receptor, enzyme or other binding site.

Glossary of Terms Used In Computational Drug Design, *Pure&Appl. Chem.*,Vol. 69, No. 5, pp. 1137-1152, 1997

CCDC

GOLD: Protein-Ligand Docking Software

- GOLD has proven success in virtual screening, lead optimisation, and identifying the correct binding mode of active molecules.

- Relied on by researchers in academia and industry worldwide.

- Reliable
- Flexible
- Configurable

CCDC

GOLD: Protein ligand docking

FEATURES

- Multiple scoring functions for scoring & rescoring.
- Ten constraints available including covalent docking, H-bonds constraint, scaffold constraint and pharmacophore constraint.
- Protein flexibility: ensemble docking (backbone), rotamers (side-chains), soft potentials (localised loop).
- Modified genetic algorithm can turn key water molecules on/off.
- Handles metals and automatically assigns a geometry during protein initialisation.

ACCESS

- User-friendly interface (Hermes).
- Available through the CSD Python API.
- GOLD on the cloud and GOLD HPC to screen hundreds of millions of compounds.

CCDC

GOLD: Protein-Ligand docking software

Reliable

- Make confident **binding mode predictions.**
- Achieve high **database enrichments.**
- Identify the **correct binding mode** reliably.

Flexible

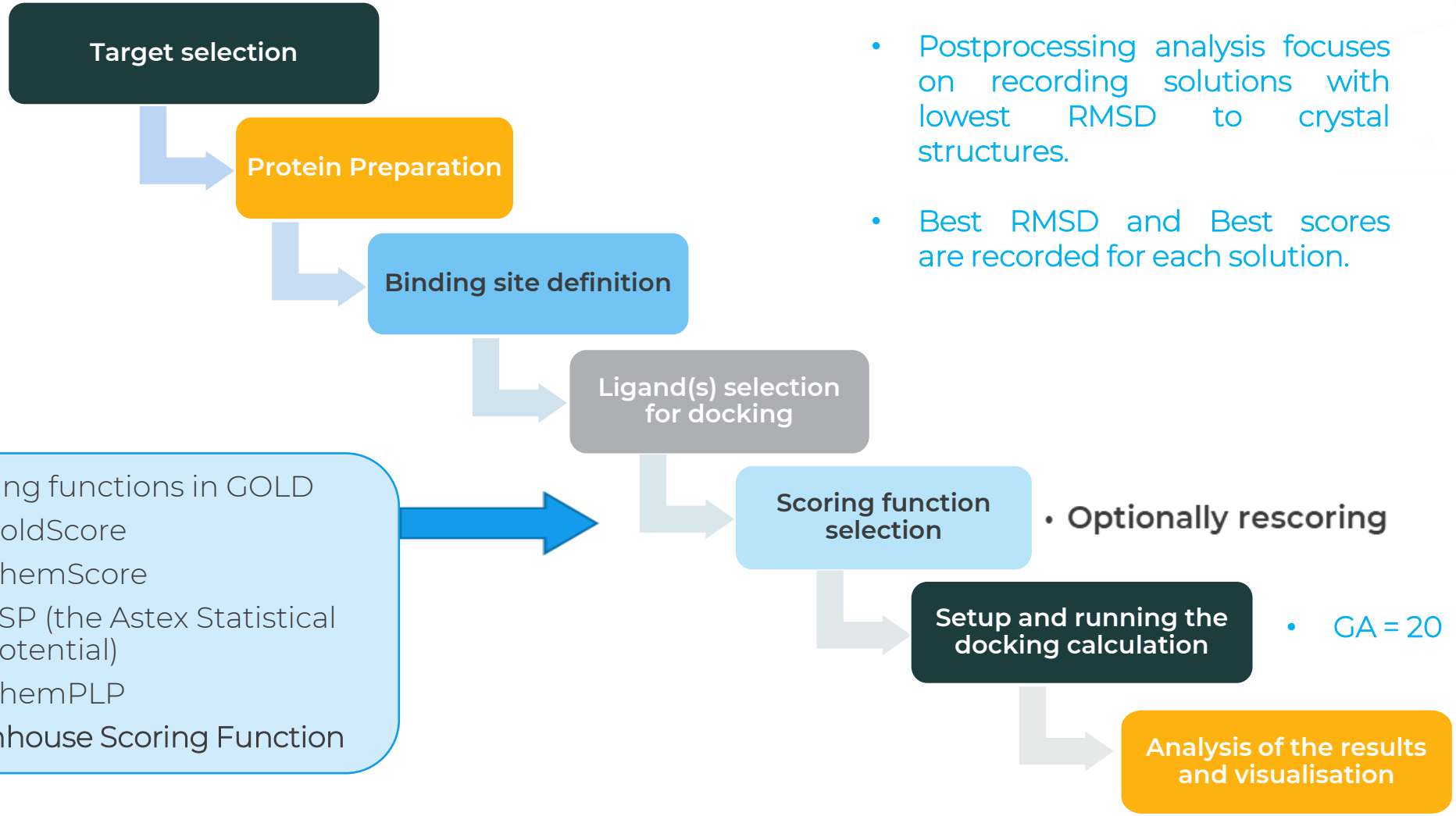
- Use ensemble docking to reflect **different receptor conformations.**
- Account for receptor flexibility through **side-chain flexibility.**
- Avoid computationally expensive **sequential docking.**

Configurable

- Adjust speed versus accuracy.
- Select from a wide range of scoring functions.
- Employ a wide range of constraints.
- Eliminate unfavourable ligand conformations.

CCDC

Methodology for GOLD receptor-ligand docking studies

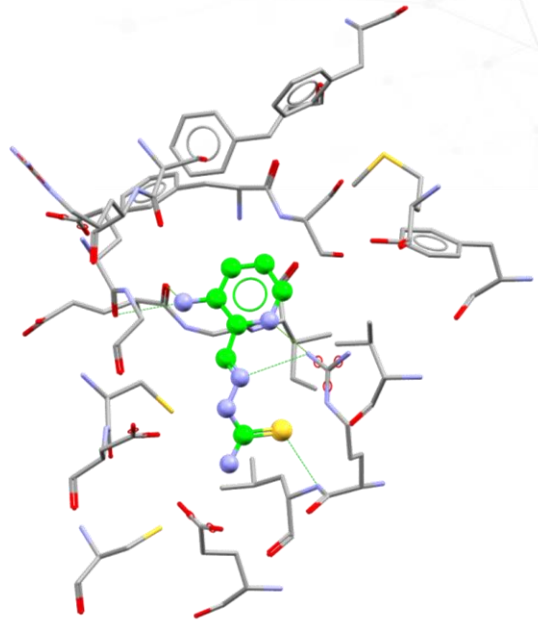


- Postprocessing analysis focuses on recording solutions with lowest RMSD to crystal structures.
- Best RMSD and Best scores are recorded for each solution.

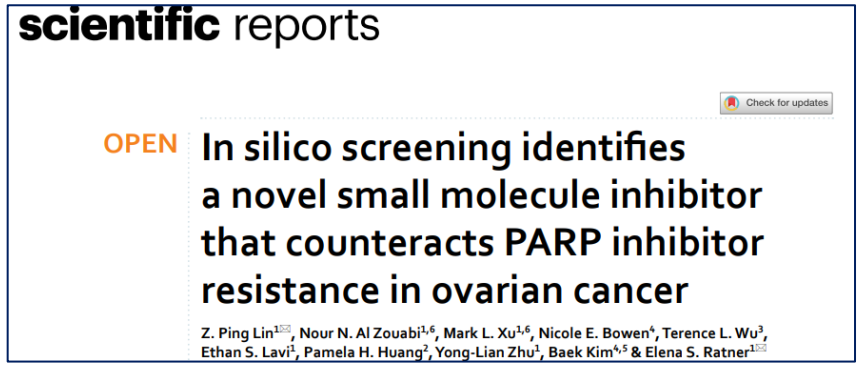
CCDC

GOLD: Application in drug discovery

- Several PARP inhibitors - for epithelial ovarian cancer (EOC) with BRCA gene mutations or defective homologous recombination (HR) repair.
- BRCA mutation reversal and restoration of HR repair has resulted in the PARP inhibitors being less effective in their treatment.
- Targeting HR repair would sensitize repair-proficient EOC to PARP inhibitors.
- Lin, et al., used virtual screening to investigate protein-ligand docking of the triapine-binding pocket on RNR to identify a next generation inhibitors for the treatment of ovarian cancer.
- Virtual screening using GOLD of over 200,000 compounds revealed hydrogen bonds and hydrophobic interactions with the triapine binding pocket. Hit-clustering was performed on the most promising 200 compounds, narrowing the candidate field down to 25 candidates.
- Experimental validation and further physicochemical analysis identified a potential candidate.



Docking post of triapine in the cavity was modelled with GOLD



scientific reports

Check for updates

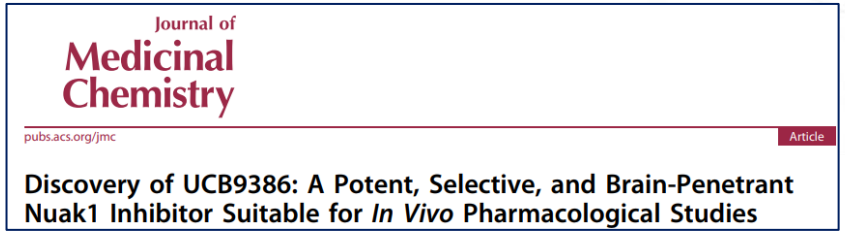
OPEN

In silico screening identifies a novel small molecule inhibitor that counteracts PARP inhibitor resistance in ovarian cancer

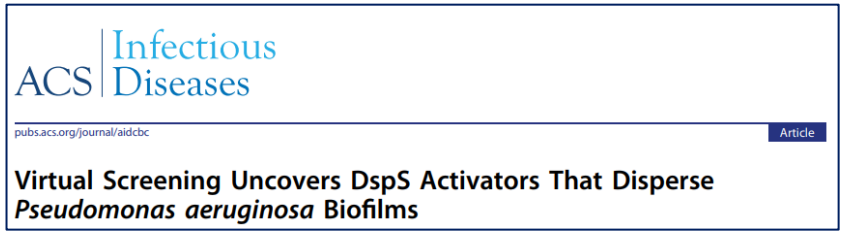
Z. Ping Lin[1✉], Nour N. Al Zouabi[1,6], Mark L. Xu[1,6], Nicole E. Bowen[4], Terence L. Wu[3], Ethan S. Lavi[1], Pamela H. Huang[2], Yong-Lian Zhu[1], Baek Kim[4,5] & Elena S. Ratner[1✉]

GOLD: Application in various studies

In medicinal chemistry to design of novel drug molecules for Central Nervous System Disorders

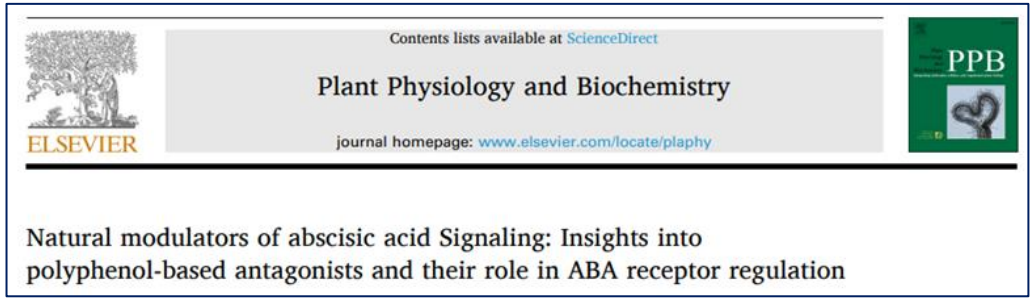


Virtual Screening to identify novel of drug molecules against infectious disease



Docking studies and pose prediction to discover therapeutic modulators for Alzheimer's disease

In Agrochemical Development – For the development of drought resistant crop variety



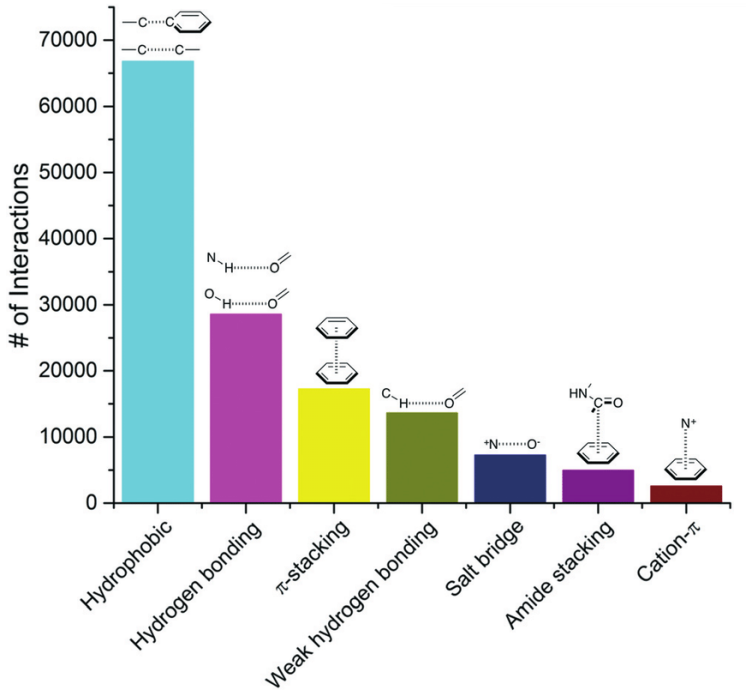
Superstar: Analysing interaction preferences

CCDC

SuperStar: non-covalent interactions

Knowledge-based software

- Critical driving force for the structure, stability, and function of proteins.
- Important for drug discovery as their interplay is responsible for binding affinity within proteins and ligands.



Most common non-covalent interactions observed in protein–ligands extracted from the PDB

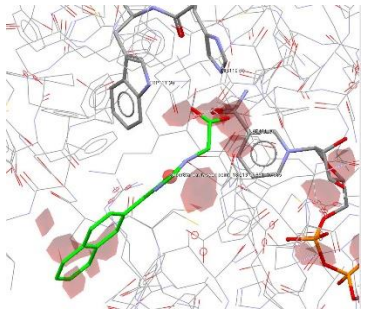
Freitas & Schapira. *Med. Chem. Commun.*, 2017, 8, 1970–1981.

CCDC

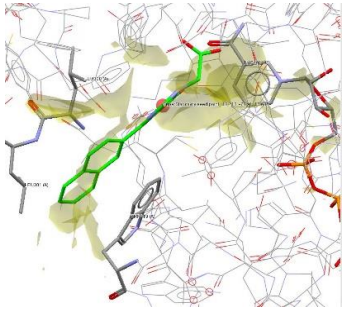
Target selection and validation

Given a whole protein and no prior knowledge, Fragment Hotspot Maps identify the location and quality of binding sites.

H-bond acceptor propensity map

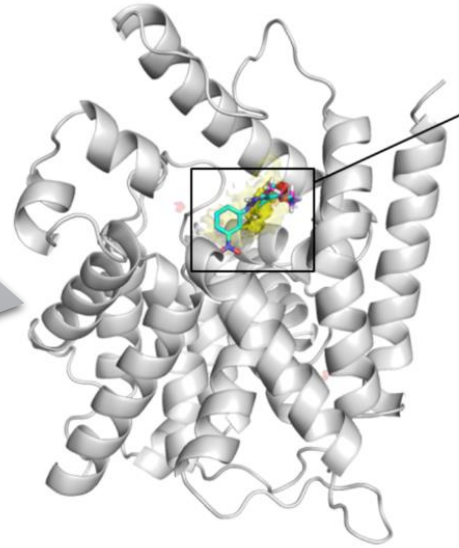


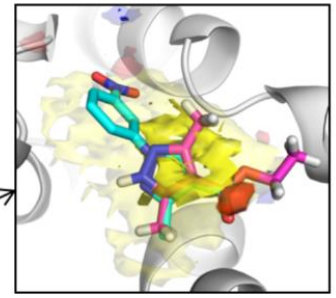
hydrophobic propensity map



SuperStar.

Identifying binding hot-spots in protein.





IsoStar.

A knowledge-based library of intermolecular interactions derived from the CSD.

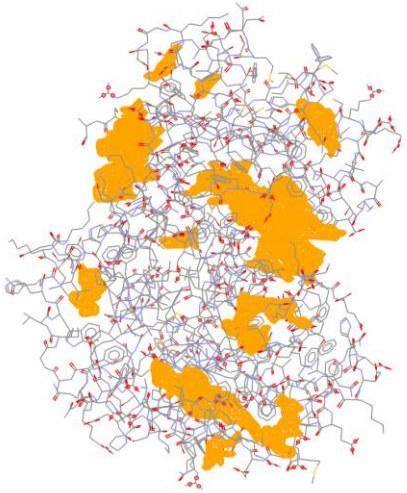
SuperStar uses crystallographic information about non-bonded interactions to generate interaction maps within protein binding sites:

- Interaction preferences visualised simply in 3D
- Assess pocket druggability

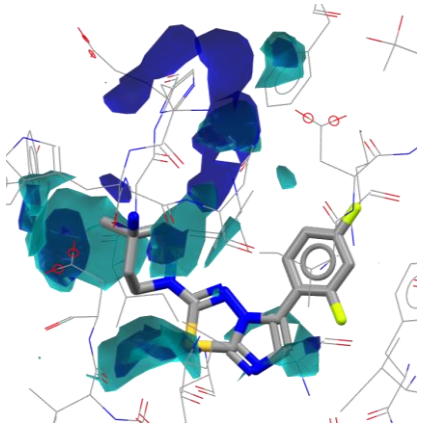
CCDC

SuperStar : applications

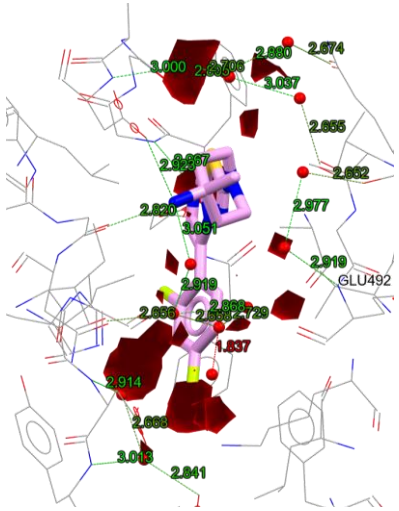
Knowledge-based software



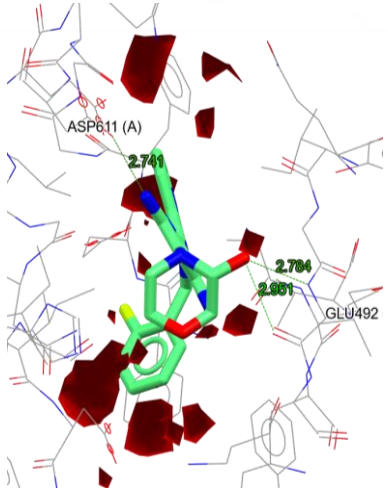
Cavity Detection



Hot-spots in proteins and ligands.



Identification of structural waters



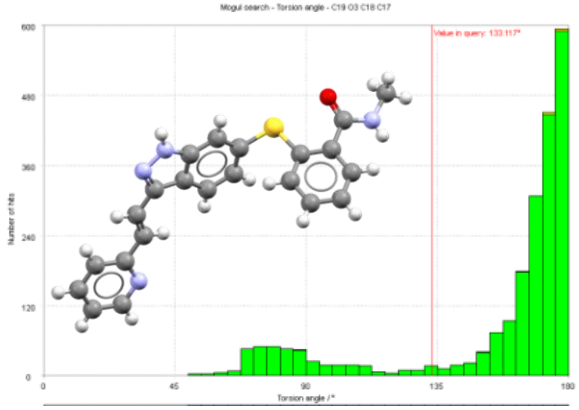
Validation of docking poses

"A useful guide to recognize possible interaction geometries while designing ligands"
(Journal of Computer-Aided Molecular Design, 14: 787–803, 2000)

CCDC

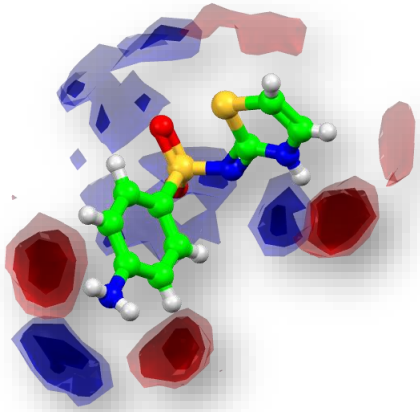
Hit to Lead Back to Optimization

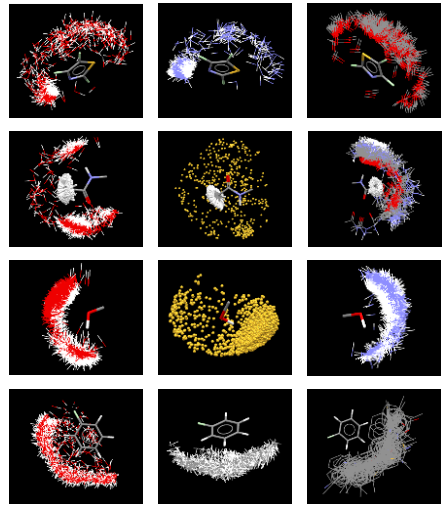
- Validate the complete geometry of a given query structure and identify any unusual features with Mogul.



- Optimize compound geometry by assessing how changes affect binding.
  - Selection of linking groups.
  - Ring-substituent geometry.
  - Intramolecular H-bonding.
  - Torsional barriers.

- Gain insights into the conformational and interaction preferences of pharmaceuticals and design molecules with desired shapes.
- Check the impact of changes with docking pose prediction.
- Find isosteric replacements or scaffold hops.





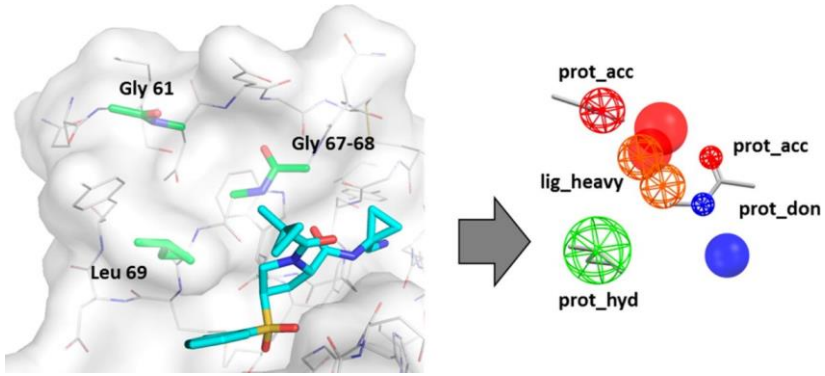
CCDC

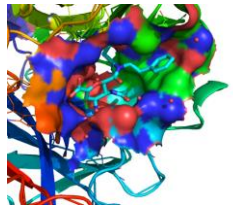
CSD-CrossMiner: Pharmacophore search

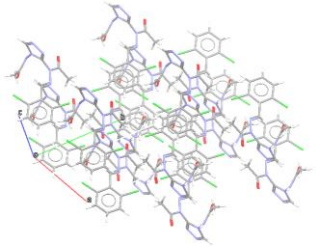
CCDC

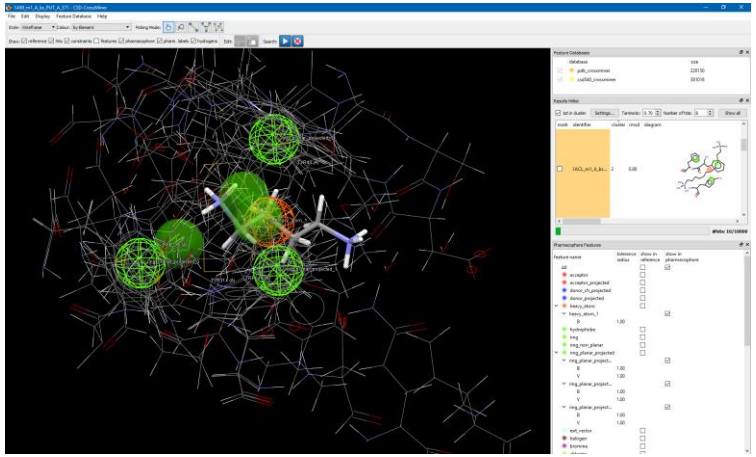
CSD-CrossMiner overview

- Pharmacophore-based searches of structural databases (CSD & PDB & any in-house databases, simultaneously).
- Modify a hypothesis/results on the fly: interactive tool.
- Annotated for easy filtering of hits.



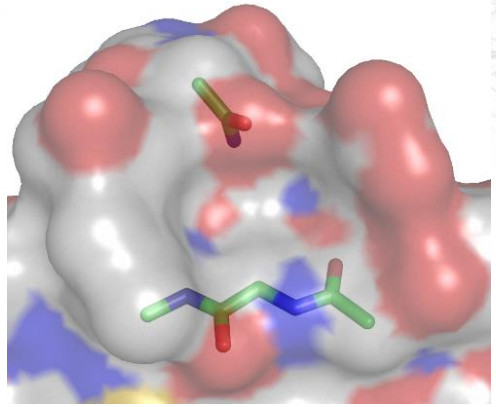






Interactive and versatile navigation of structural databases
Korb O., *et al. J. Med. Chem.* (2016) 59, 4257-4266.

CSD-CrossMiner applications



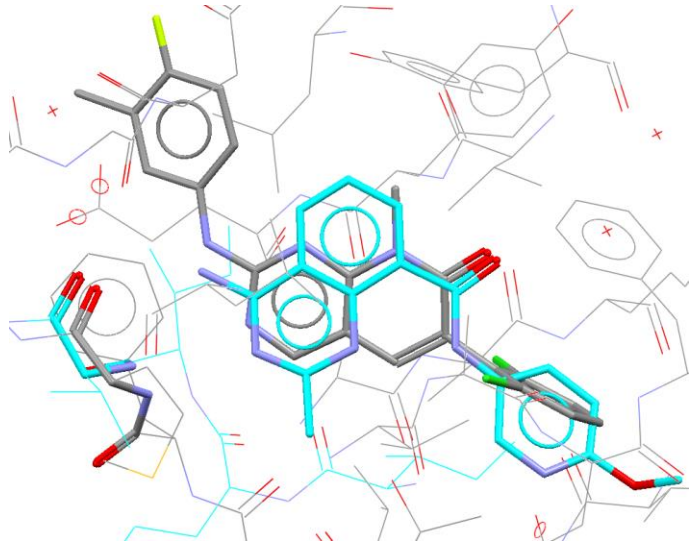
- Determine common protein binding sites in PDB structures.
- Determine structural motifs that bind in similar environments.
- Inform cross-pharmacology between protein targets.
- Generate new ideas:
  - Design novel motifs that mimic extablished ligands

    ➔ improve molecular properties; solve patent issues
  - Scaffold-hopping: retrieve a diversity of ligand topologies that can be used as scaffolds

    ➔ quickly advance a project with known ligands; optimise leads

  - Growing into other regions of a binding site

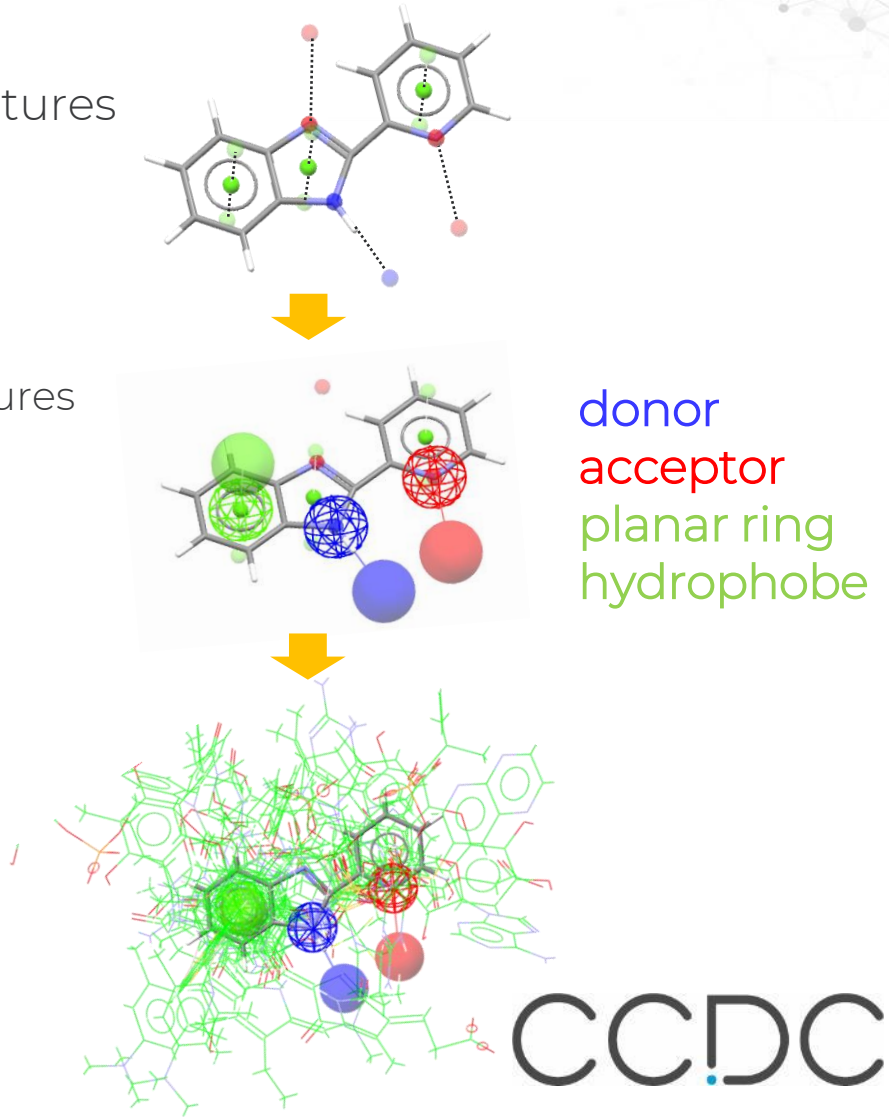
    ➔ improve selectivity; improve bioactivity; reduce cross-reactivity



CCDC

From features to pharmacophores

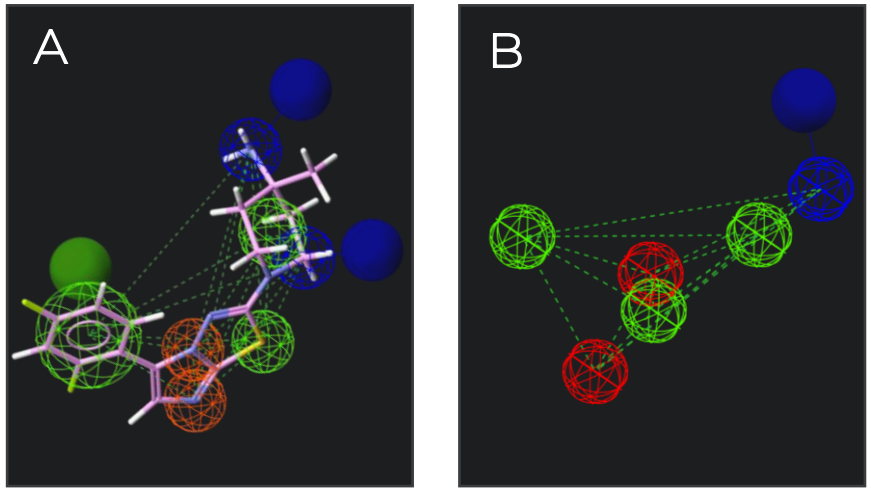
- Molecular structures are annotated with (customisable) features
  - Based on SMARTS patterns
  - Stored in feature database
- Pharmacophore query is based on tolerance spheres
  - Sphere radii reflect uncertainty in the position of the features
  - Large sphere → less strict
- Pharmacophore points can be
  - Single point
  - Directional (two points)
- Customisation of pharmacophore points
  - Molecule class: protein vs. small molecule vs. any
  - Constraint: intramolecular vs. intermolecular vs. none

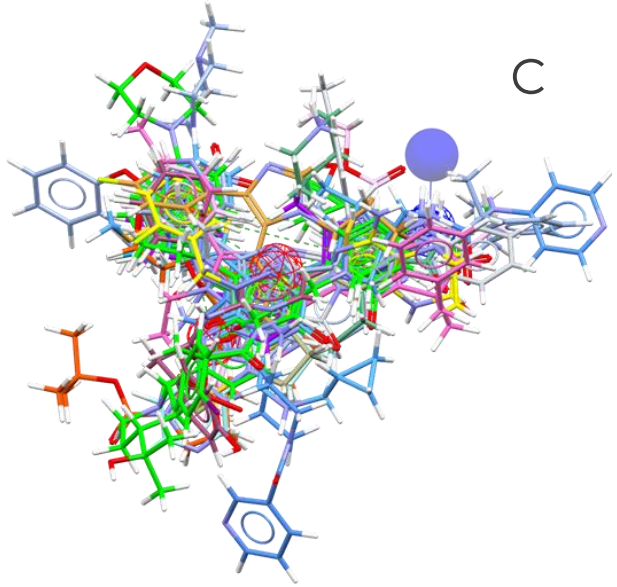


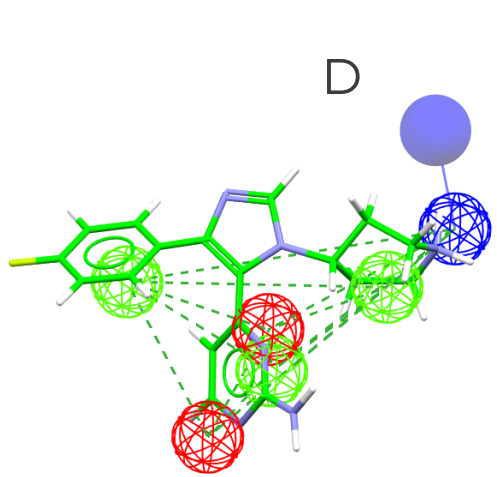
CCDC

Ligand-Based Virtual Screening

CSD –CrossMiner Pharmacophore



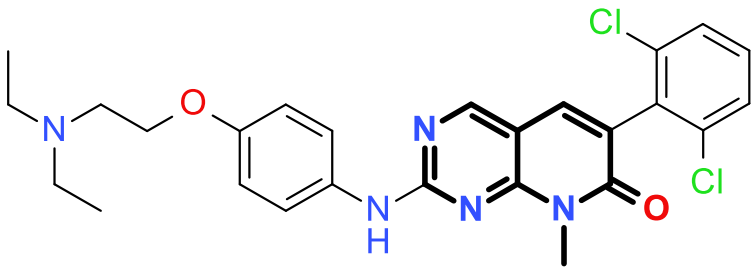




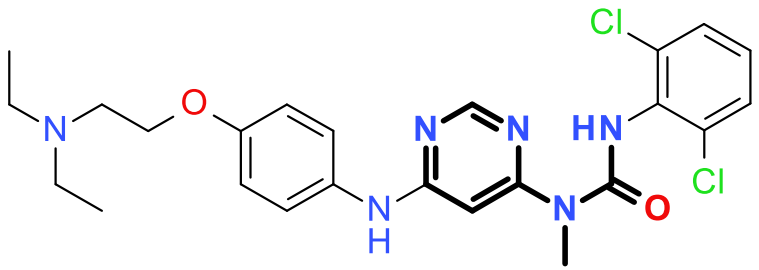
- ✓ Heavy atoms features replaced by H-bond acceptors.
- ✓ Hydrophobic features volume reduced.
- ✓ Not essential H-bond donor feature removed.

CCDC

Kinase inhibitor scaffold hop



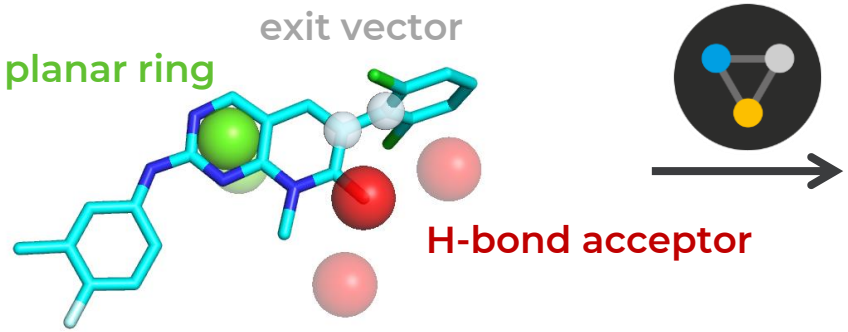
PD166285 tyr-kinase inhibitor



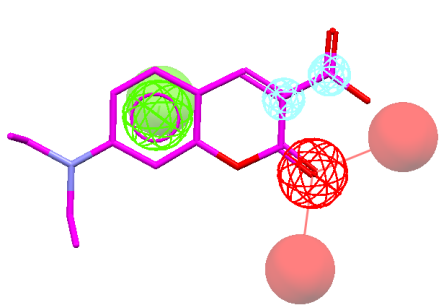
Prototype compound 1

IC_{50} < 1 uM against different kinases

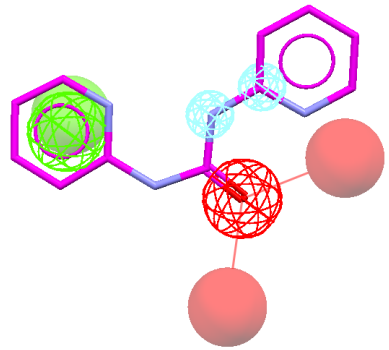
c-Src	0.066
EGF-R	0.38
c-Abl	0.25
FGFR-1	0.57
c-Kit	0.93
KDR	0.96
Tie-2	0.30
p38	0.35
EphB4	0.43



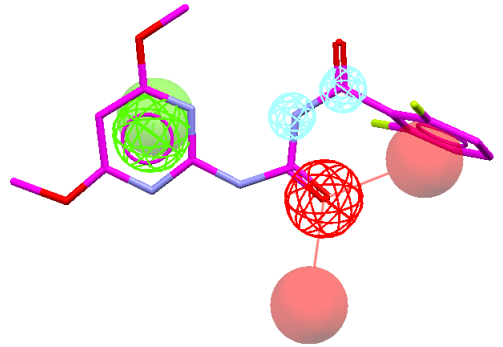
PDB:2hzi (Abl kinase)



IDUQOO



ALOJIW



IFAFEB

Entry into a new class of protein kinase inhibitors by pseudo ring design.
Furet P., *et al. Bioorg. Med. Chem. Lett.* (2008) 18, 897-900 (Novartis)

CCDC

Ligand-Based Virtual Screening Workflow

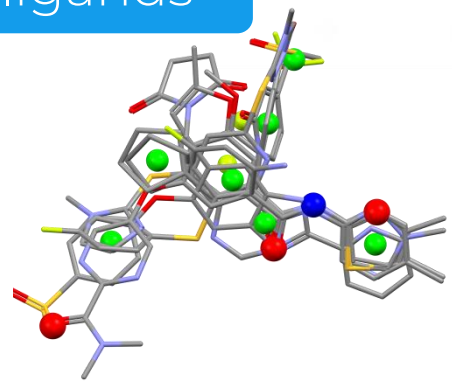
CCDC

Ligand-Based Virtual Screening Framework

Set of known active ligands

CSD-CONFORMER GENERATOR & CSD-LIGAND OVERLAY

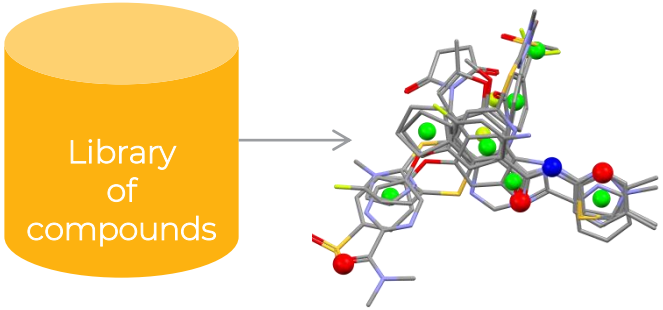
Pharmacophore modelling



Applications

FIELD-BASED LIGAND SCREENER

Virtual screening



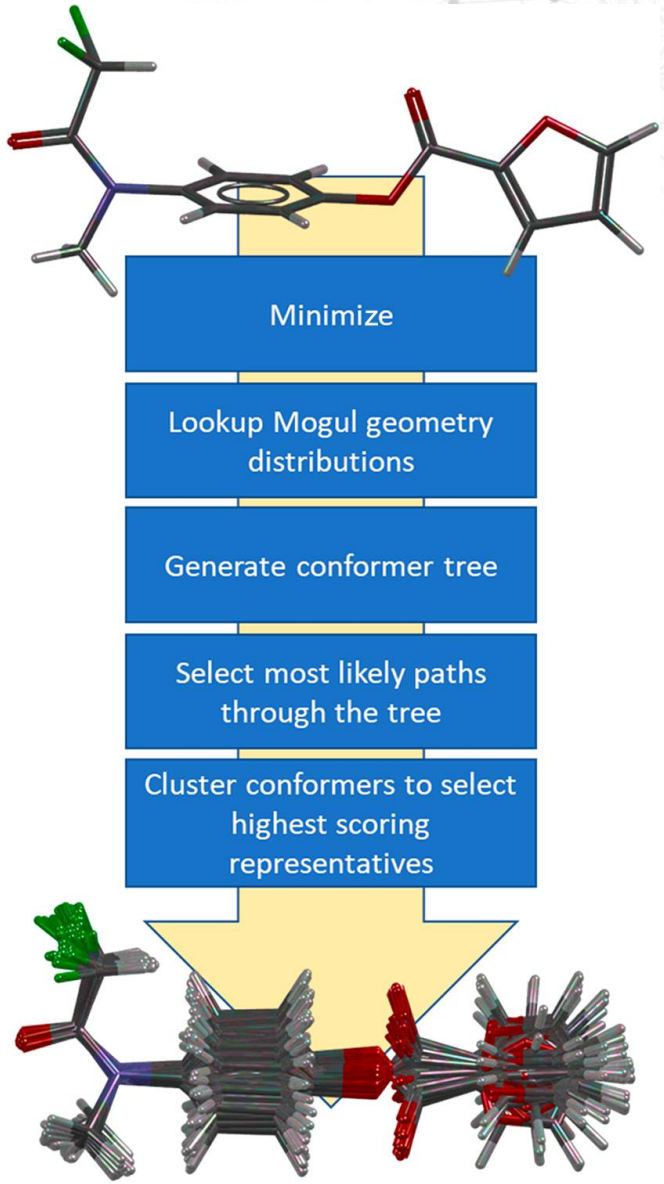
Others:
De novo design
Lead optimisation
Activity profiling
...

CCDC

CSD-Conformer Generator

- Part of CSD-Discovery, CSD-Materials and CSD-Enterprise suites
- The starting point are Mogul data libraries with some changes and improvements to tailor them to conformer generation
- 2022.1 release of Conformer Generator 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

Workflow of conformer generation



Knowledge-Based Conformer Generation Using the Cambridge Structural Database. Cole J. C., *et al. J. Chem. Inf. Model.* (2018), 58, 615-629.

CSD-Python API:

Knowledge-based software

- Programmatic access to the CSD software and extended functionalities
- Flexible, versatile and continuously evolving
- Integration with 3rd party software
- User-designed drug discovery workflows and jupyter notebooks
- Cloud infrastructures (AWS, Azure)
- Open source CSD GitHub repository: https://github.com/ccdc-opensource

CSD Python API 3.0.16 documentation » The CSD Python API

- API documentation
  - IO API
  - Entry API
  - Crystal API
  - Molecule API
  - Search API
  - Conformer API
  - Protein API
  - Cavity API
  - Docking API
  - Pharmacophore API
  - Screening API
  - Particle API
  - Interaction API
  - Descriptors API
  - Morphology API
  - Diagram API
  - Utilities API
  - CSD Landscape Generator
  - CSP Database API
  - CSP Prediction API

scikit learn · TensorFlow · PubChem · ChEMBL · UniChem 2.0 · PyTorch · biopython · docker · mongoDB · PostgreSQL · Open for Innovation KNIME · kubernetes · Microsoft Azure · amazon web services · CCDC

Q&A

CCDC

Thank you

CCDC