First steps in protein-ligand docking with GOLD

CCDC Virtual Workshop

Rupesh Chikhale, Francesca Stanzione, Ilaria Gimondi, Suzanna Ward, Eva Myers

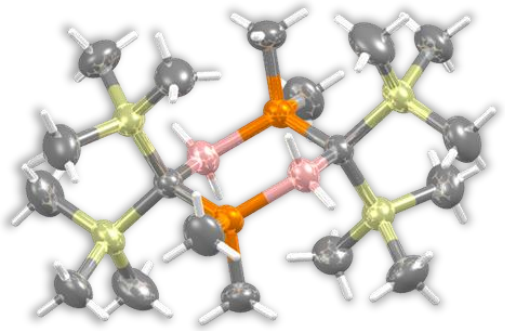
May 2022

CCDC
advancing structural science

Learning outcomes for today

- The basics of the Hermes interface.
- Step-by-step basics of GOLD.
- Where to get started with your docking simulation and how to run a standard protein-ligand dock with GOLD.
- How to identify the correct binding modes reliably and with confidence.

The Cambridge Structural Database



CSD Refcode ELOFUJ
The 1.1 millionth CSD structure
A *CSD Communication*
determined by Bill Clegg

>1,170,000

Structures published that year

Structures published previously

Number of Structures in the CSD

1,000,000
800,000
600,000
400,000
200,000
0

< 1973 1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019 2021

Year

- Every published structure
  - ASAP & early view
  - CSD Communications
  - Patents
  - University repositories
  - Theses
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for over 56 years
- A CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

CCDC

Inside the Cambridge Structural Database

The CSD is a database of all the published organic and metal-organic experimental crystal structures

Organic
43%

Metal-Organic
57%

At least one transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

Not Polymeric
89%

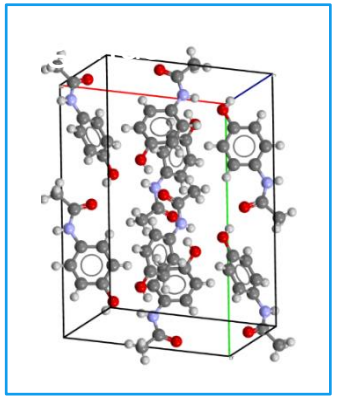
Polymeric: 11%

Single Component
56%

Multi Component
44%

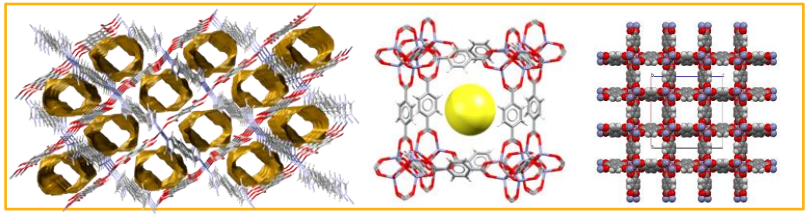
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



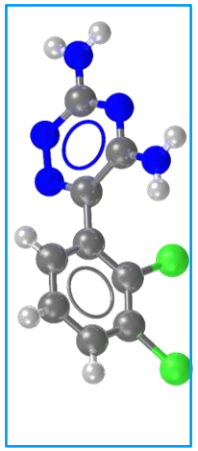
Metal-Organic

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



Additional data

- 11,525 polymorph families
- 171,683 melting points
- 909,992 crystal colours
- 778,663 crystal shapes
- 24,916 bioactivity details
- 11,387 natural source data
- > 250,000 oxidation states



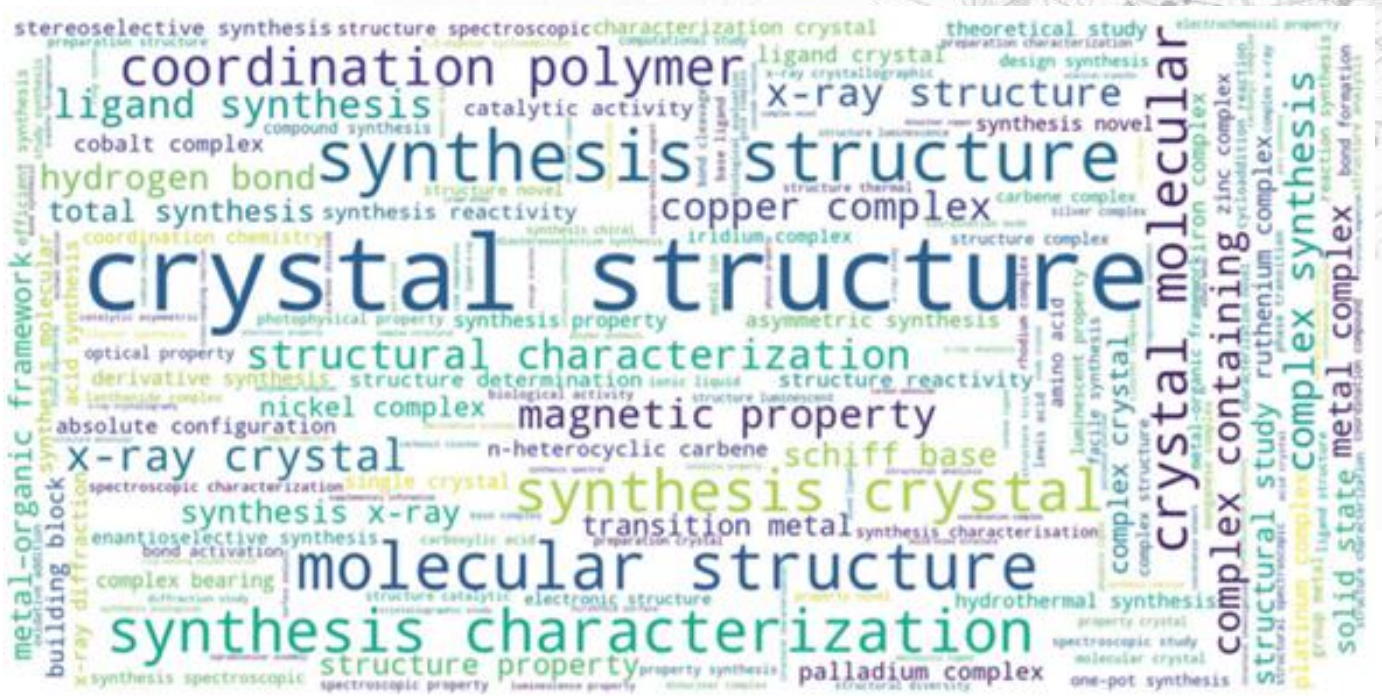
Links and subsets

- DrugBank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticide PDB

The CSD in research

The *Web of Science* subject categories that cite the 2006-2016 CCDC standard references most frequently

Subject category	Citations
Crystallography	5256
Chemistry Multidisciplinary	3275
Chemistry Inorganic Nuclear	1931
Chemistry Physical	1148
Materials Science Multidisciplinary	864
Chemistry Organic	568
Biochemistry Molecular Biology	307
Physics Atomic Molecular Chemical	203
Chemistry Medicinal	185
Spectroscopy	181



A word cloud of common bigrams in the titles of publications containing CSD-compliant crystal structures.

Willett P. et al, *CrystEngComm*, 2020,22, 7233-7241 DOI:10.1039/D0CE00045K

Top 200 drugs

- By retail sale

Small molecule structure in the CSD

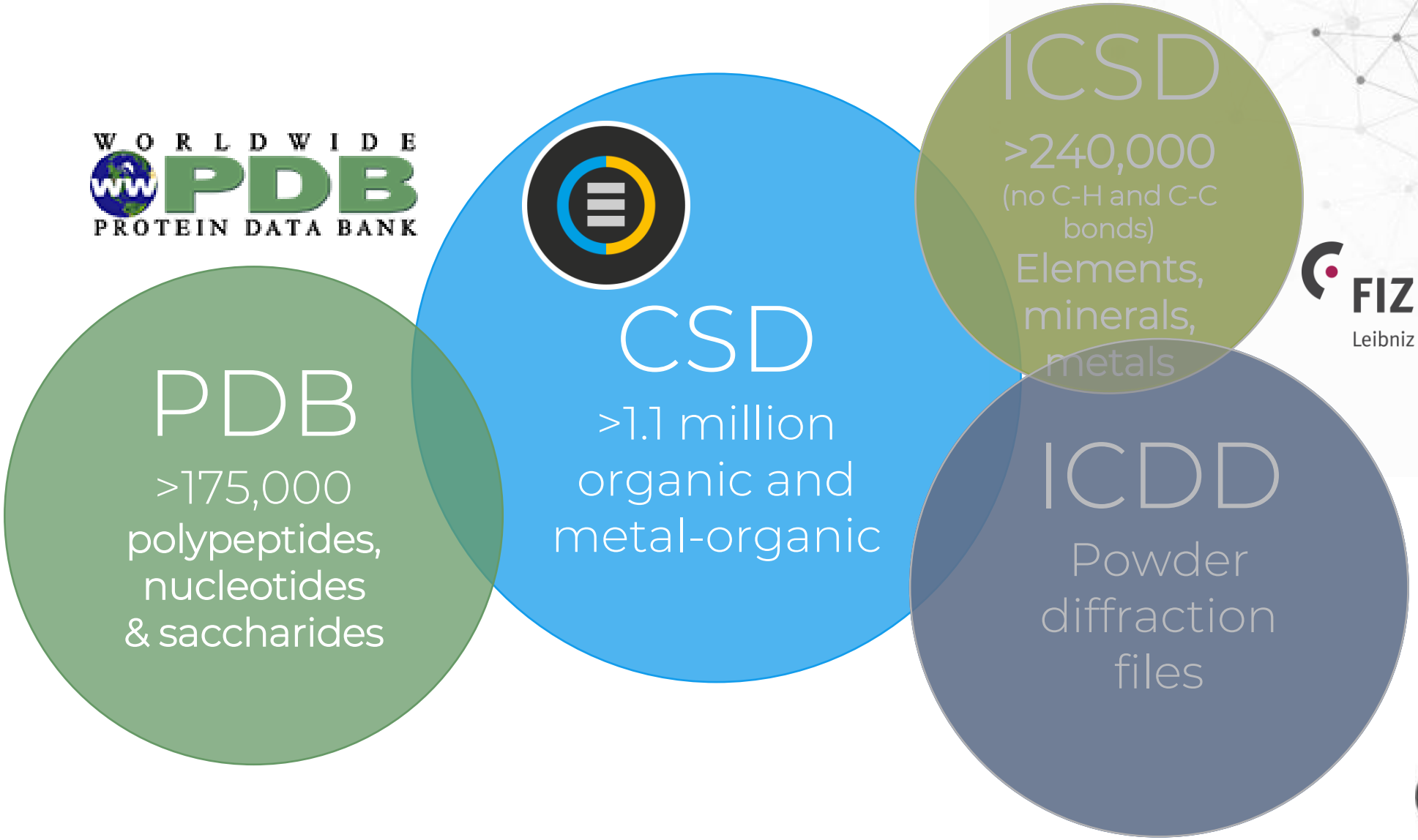
Biological structure in the PDB

Small molecule structure CSD entry pending

- Adapted from poster compiled and produced by the Njardarson Group at the University of Arizona

Adapted from the poster compiled and produced by the Njardarson Group (The University of Arizona), J. Chem. Ed. 2010, 87, 1348

Structural databases

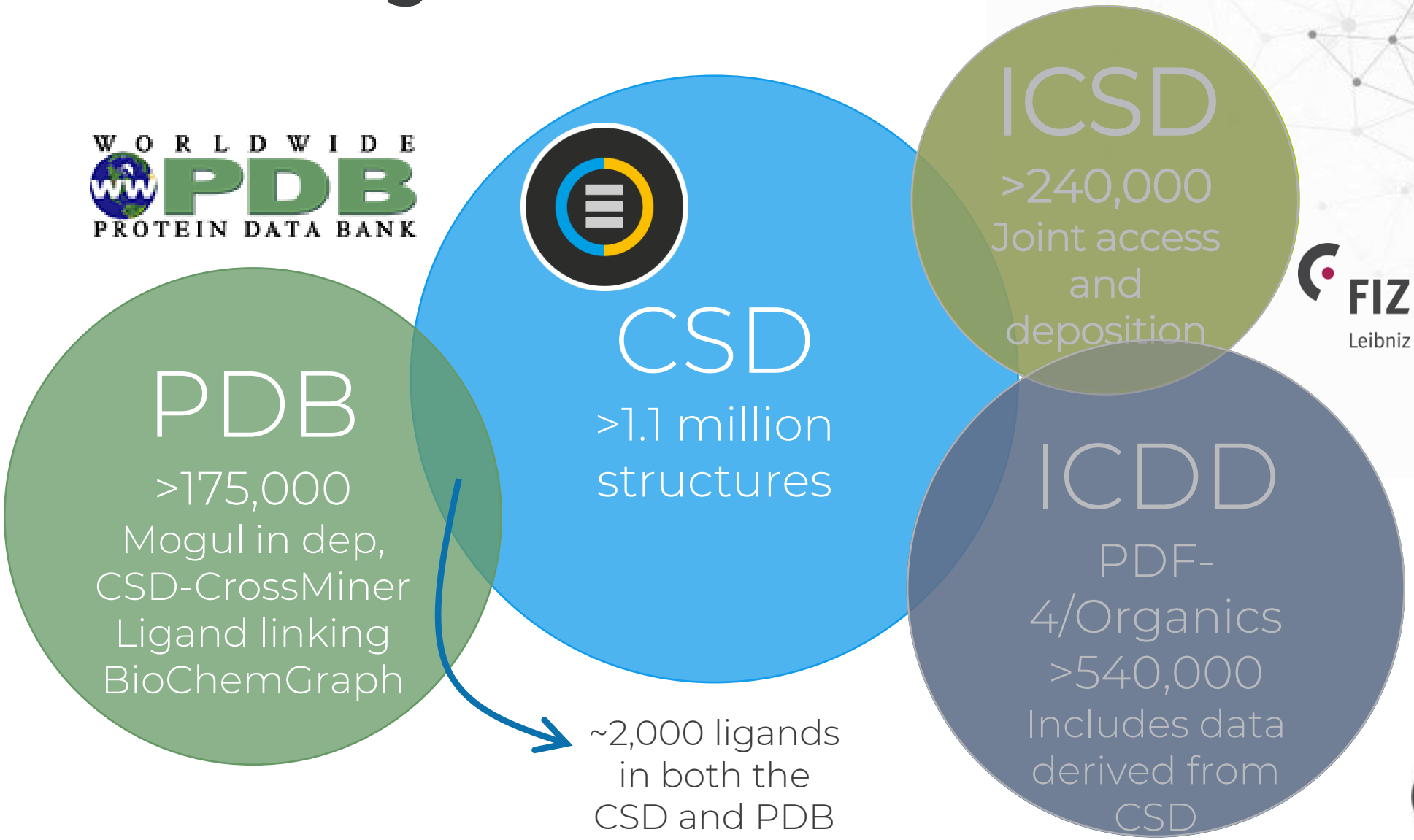


FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

ICDD

CCDC

More integrated structural databases



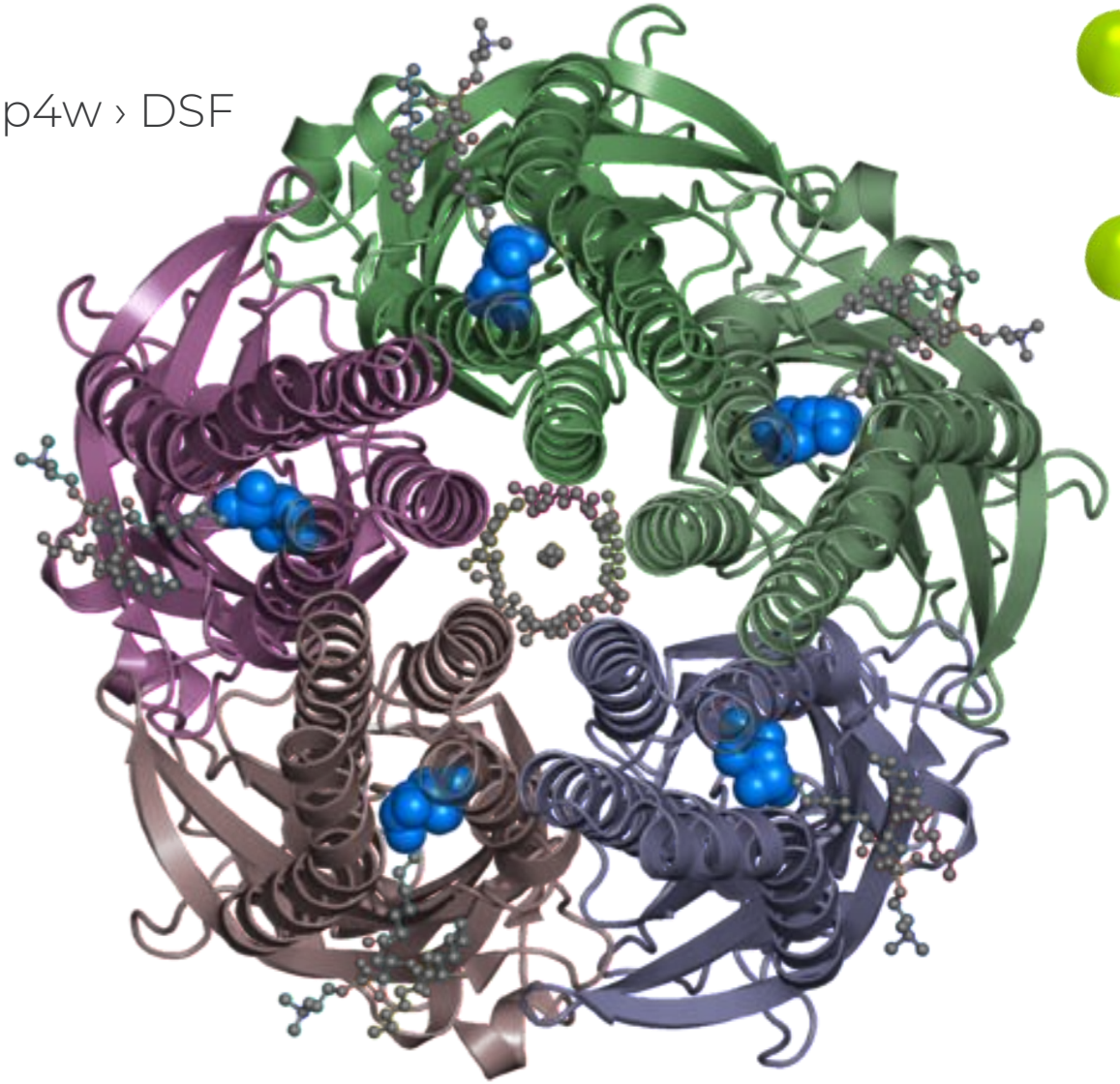
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

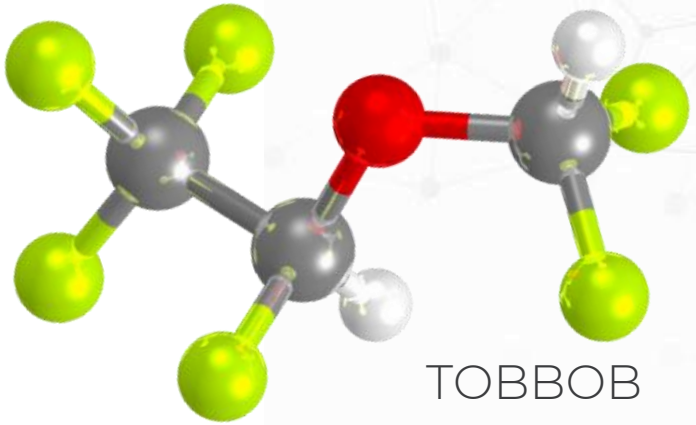
ICDD

CCDC

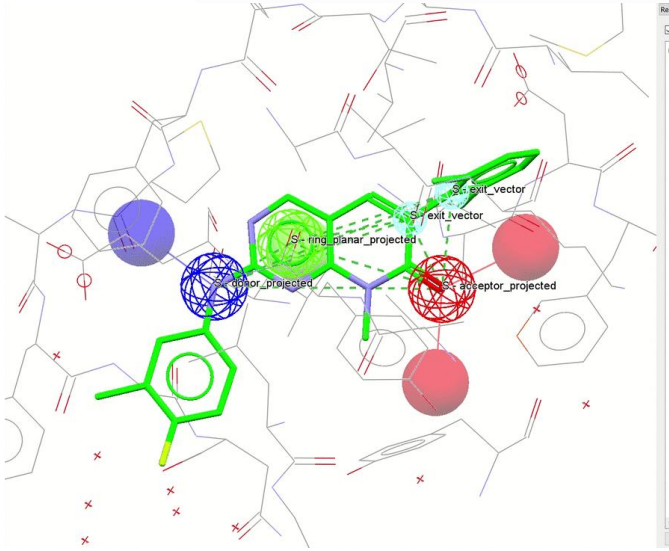
Connecting chemistry and biology

3p4w › DSF



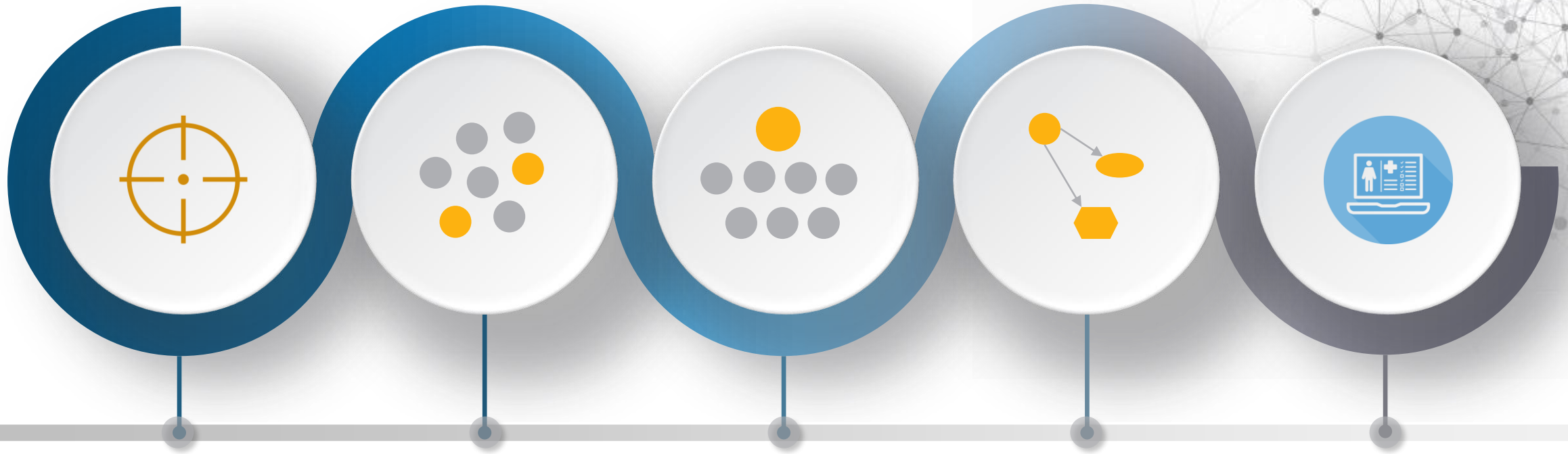


TOBBOB



CCDC

Drug Discovery Pipeline



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.

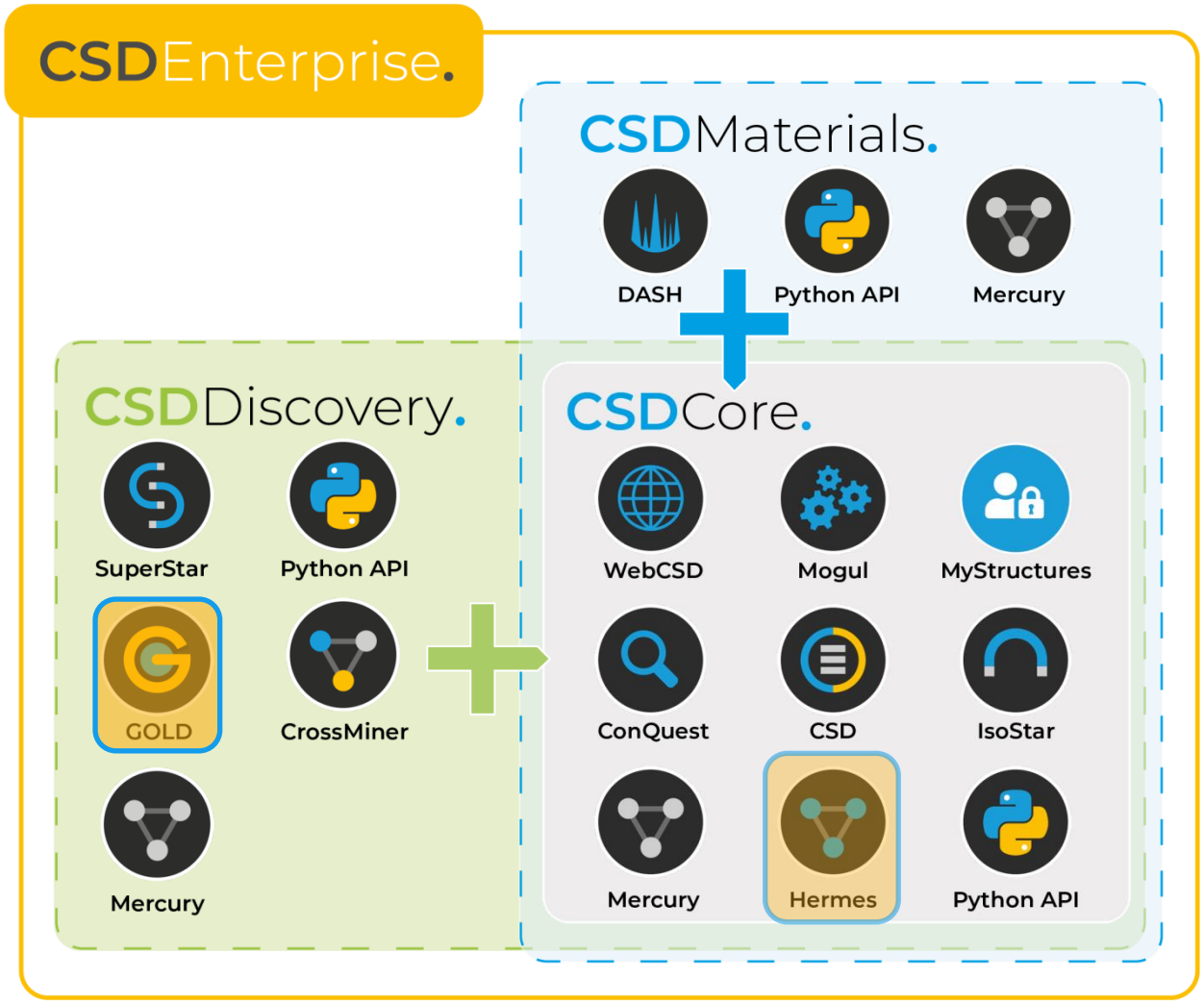
LEAD OPTIMISATION

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

DRUG DEVELOPMENT

CCDC

The CSD software



Professional Services

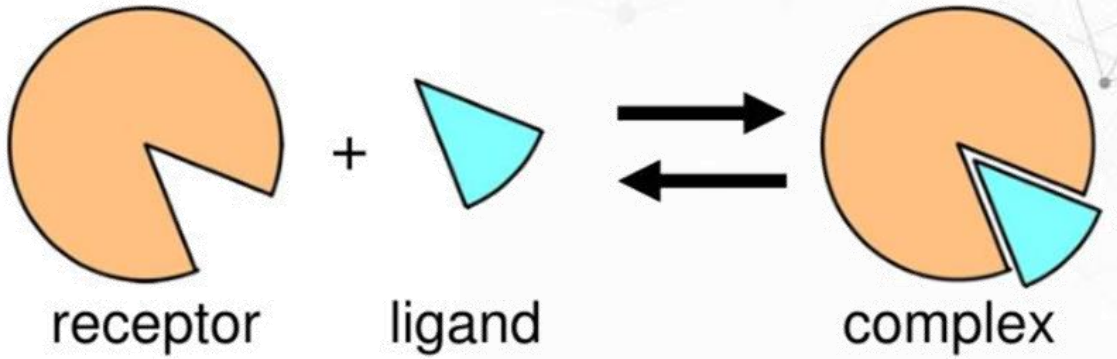
Research & Knowledge partnerships

CSDCommunity.

Mercury
enCIFer
Symmetry
Deposit
CellCheck
Educational
Access
MyStructures

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

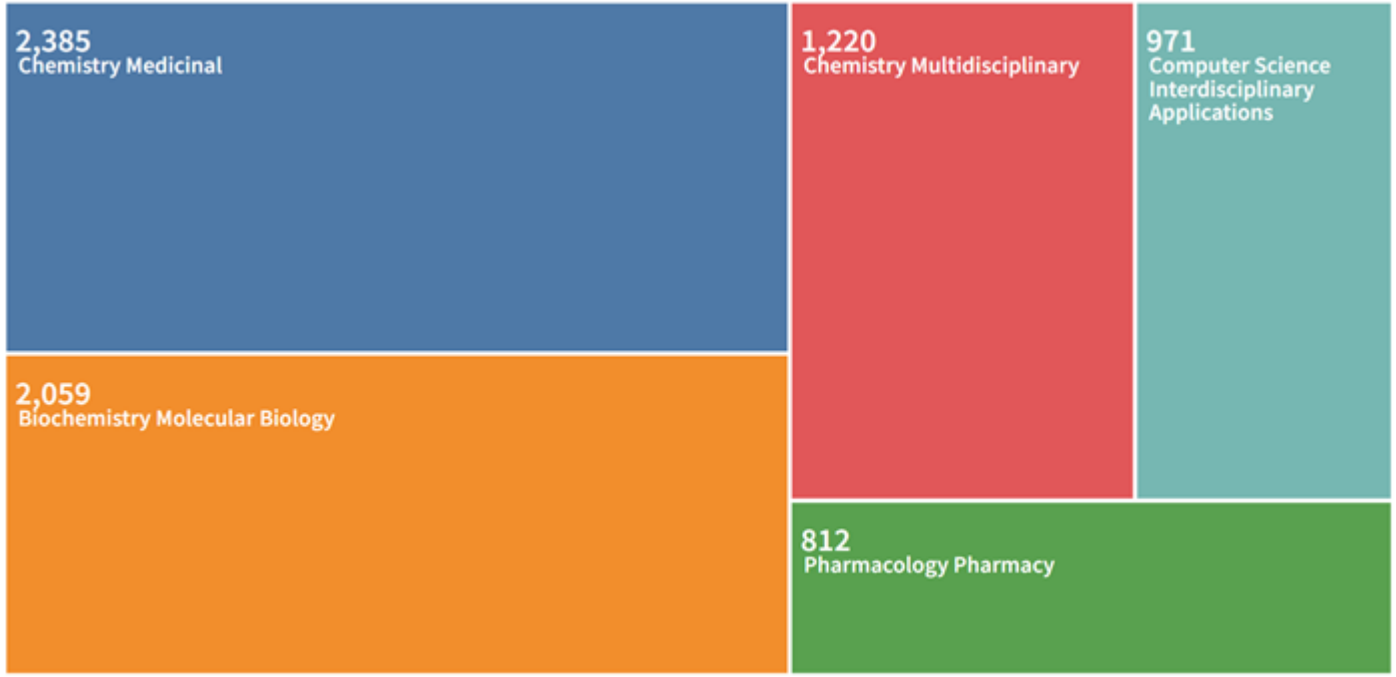
- GOLD has proven success in virtual screening, lead optimisation, and identifying the correct binding mode of active molecules.

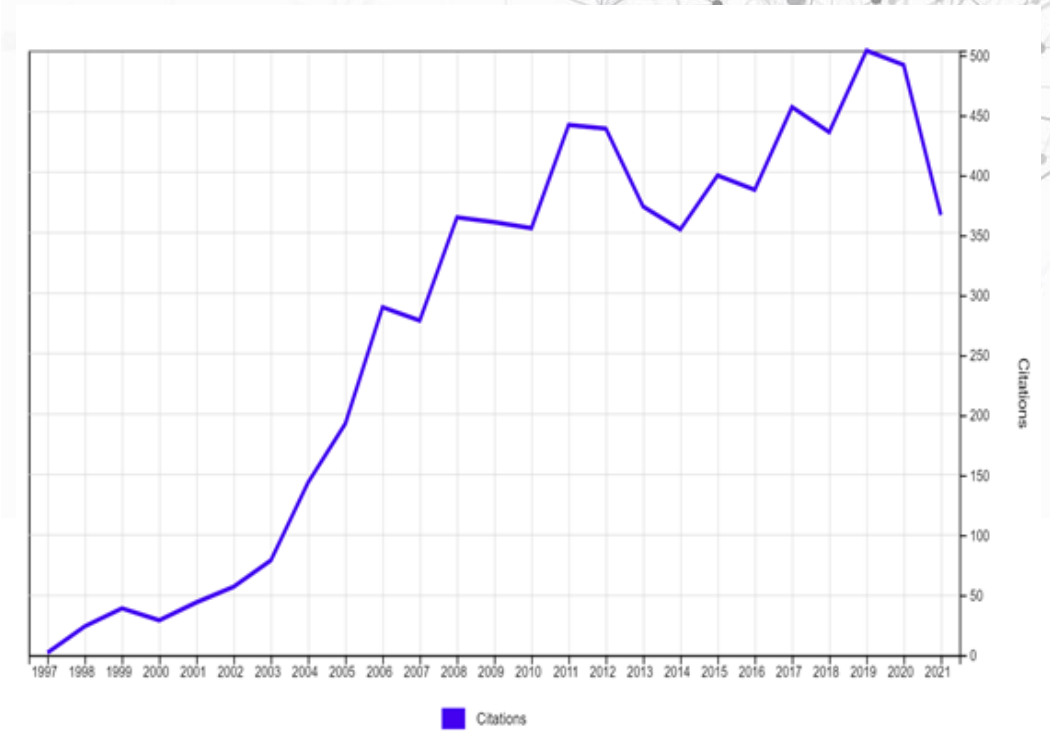
- Relied on by researchers in academia and industry worldwide.

- Reliable
- Flexible
- Configurable

CCDC

GOLD docking in research

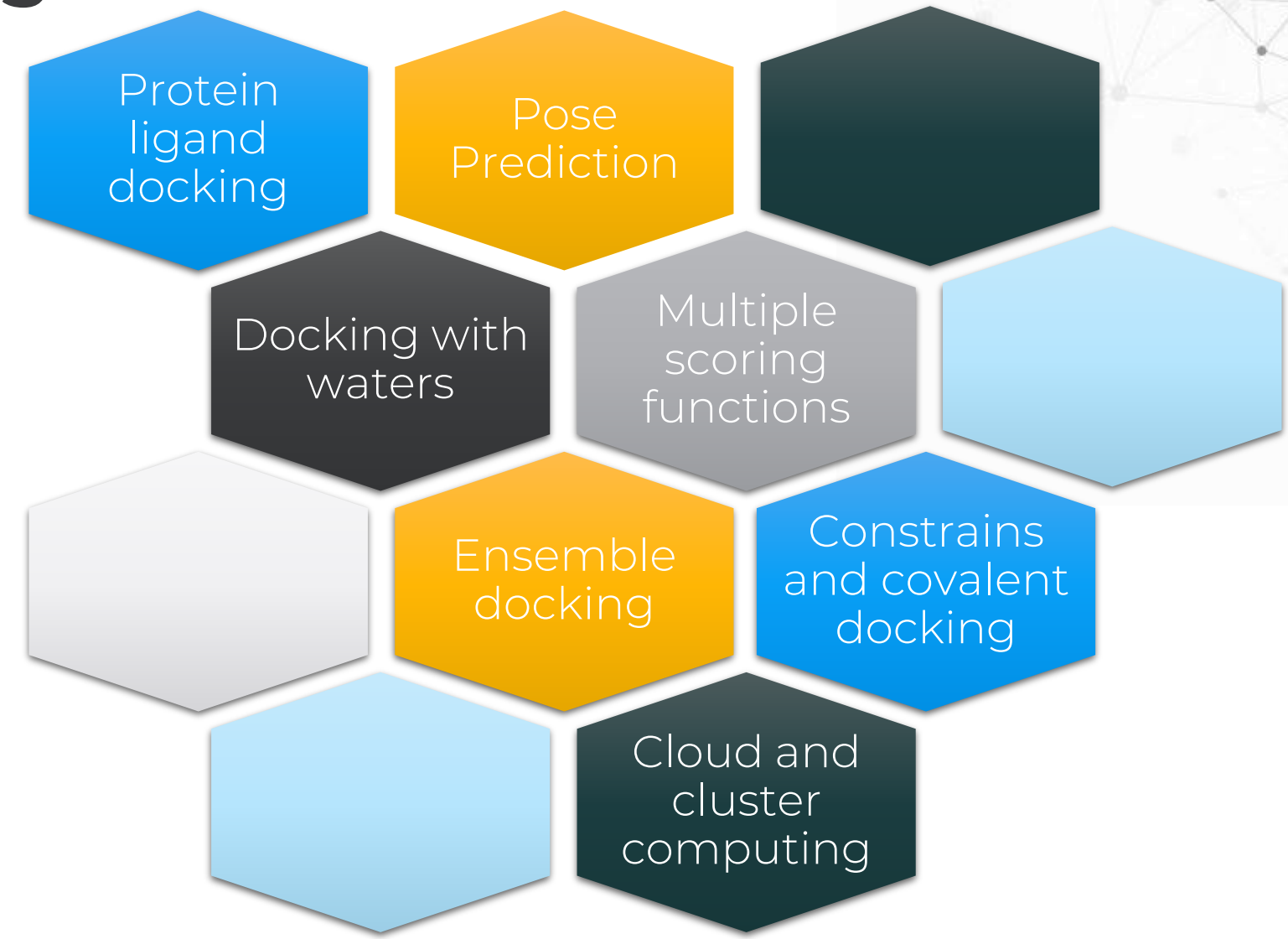




Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol., Vol. 267, pp. 727-748., 1997.
DOI: https://doi.org/10.1006/jmbi.1996.0897

CCDC

GOLD: The all in one molecular docking package



CCDC

What are we going to learn today?

- Protein-ligand docking of a Kinase inhibitor
  - We would learn how to…
    - Import a protein co-crystal from the Protein Data Bank (PDB).
    - Prepare the protein crystal structure for docking.
    - Perform molecular docking experiment in GOLD.
    - Analyse the results obtained from docking experiment.

Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation, J. Mol. Biol., Vol. 245, pp. 43-53, 1995.

CCDC

Steps in molecular docking

Target selection

- Co-crystal structure of protein-ligand complex
- from PDB

Protein Preparation

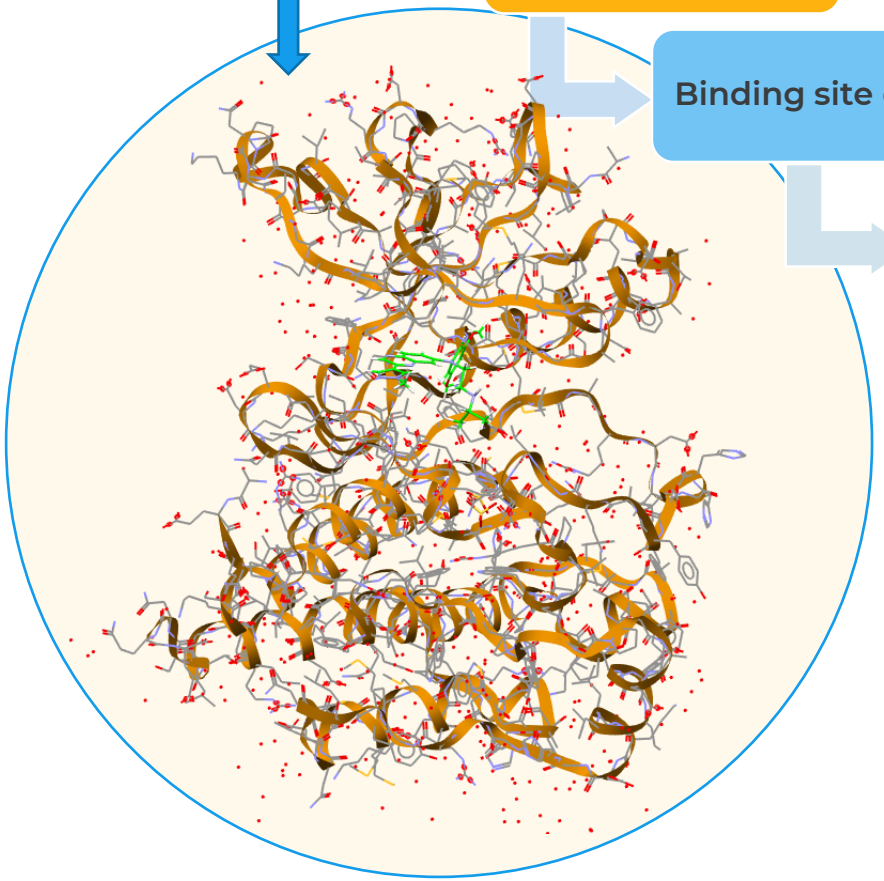
Binding site definition

Ligand(s) selection for docking

Scoring function selection

Setup and running the docking calculation

Analysis of the results and visualisation



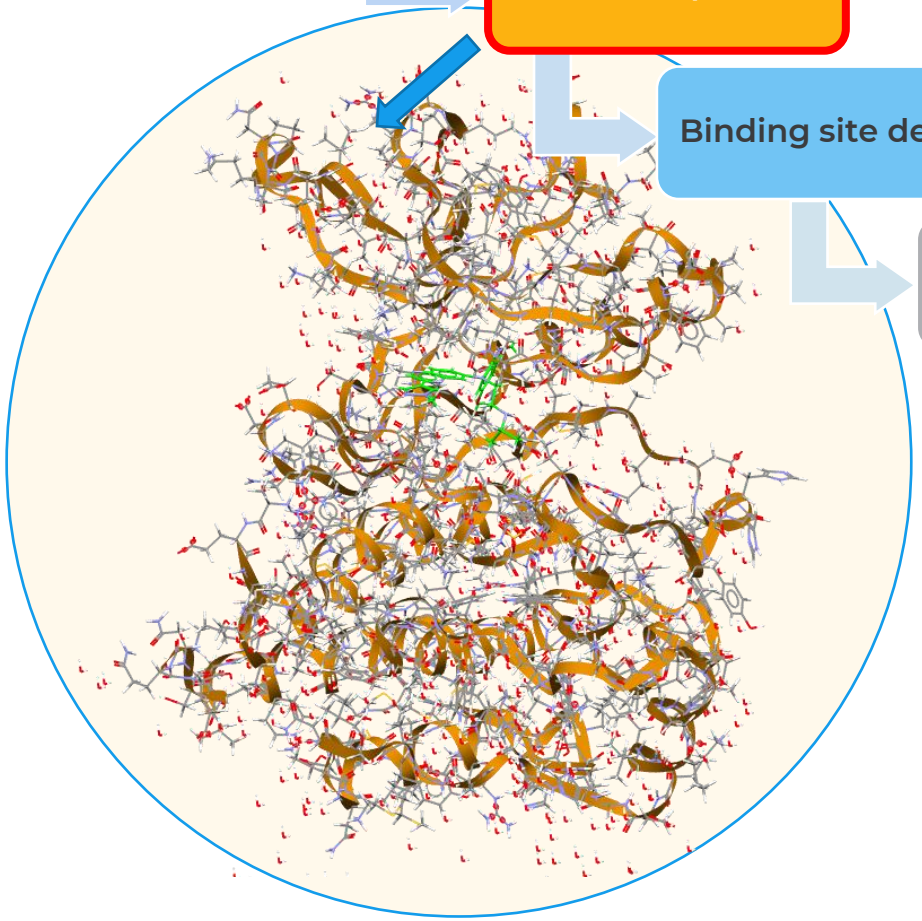
CCDC

Steps in molecular docking

Target selection

Protein Preparation

- Add hydrogens



Binding site definition

Ligand(s) selection for docking

Scoring function selection

Setup and running the docking calculation

Analysis of the results and visualisation

CCDC

Steps in molecular docking

Target selection

Protein Preparation

- Add hydrogens
- Check protonation of His, Gln, Asn

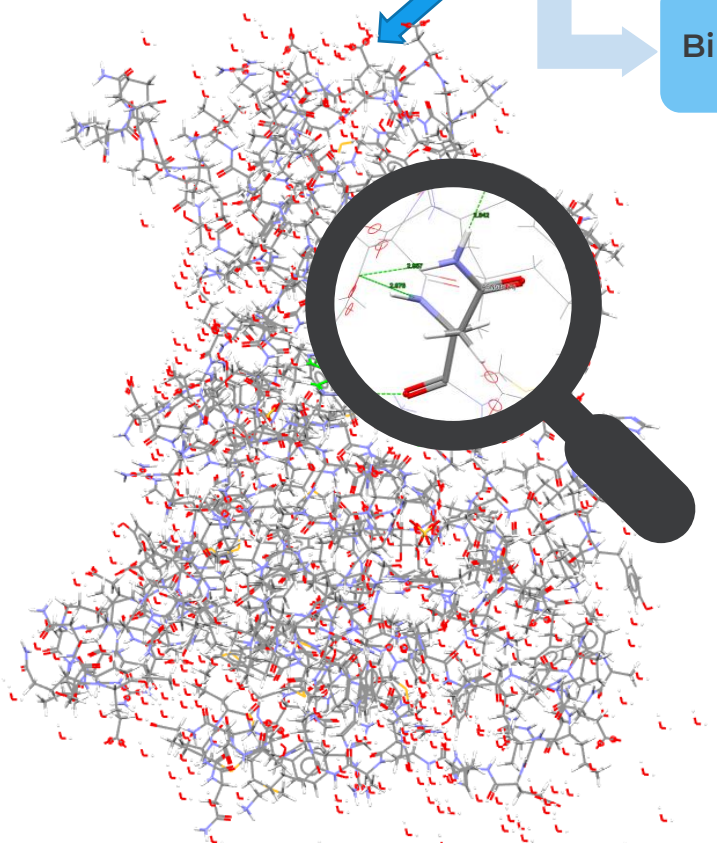
Binding site definition

Ligand(s) selection for docking

Scoring function selection

Setup and running the docking calculation

Analysis of the results and visualisation



CCDC

Steps in molecular docking

Target selection

Protein Preparation

- Add hydrogens
- Check protonation of His, Gln, Asn.
- Remove waters or retain required ones

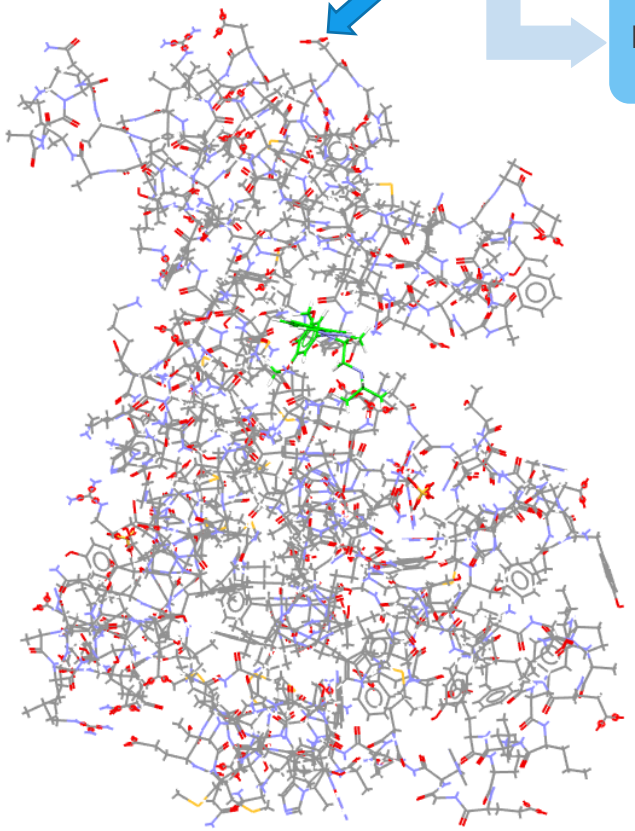
Binding site definition

Ligand(s) selection for docking

Scoring function selection

Setup and running the docking calculation

Analysis of the results and visualisation



CCDC

Steps in molecular docking

Target selection

Protein Preparation

- Add hydrogens
- Check protonation of His, Gln, Asn.
- Remove waters or retain required ones
- Extract ligand and save

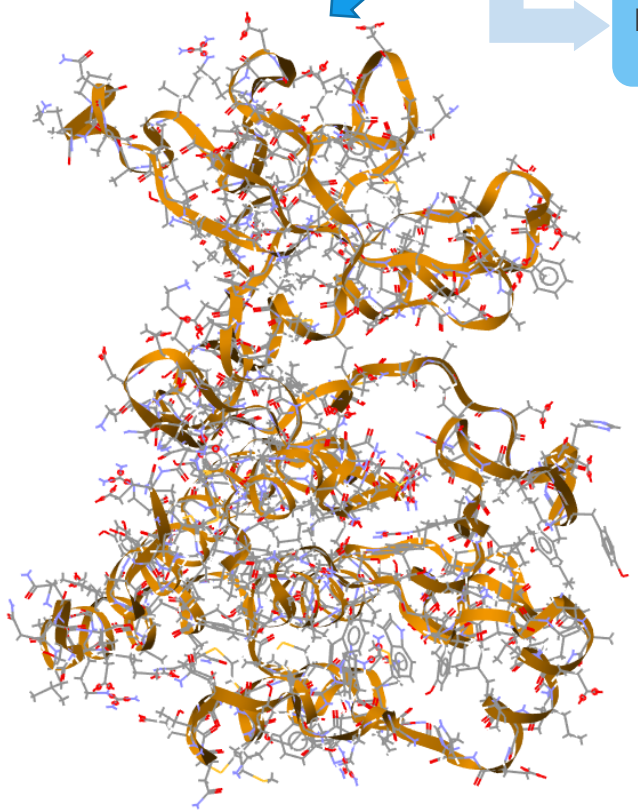
Binding site definition

Ligand(s) selection for docking

Scoring function selection

Setup and running the docking calculation

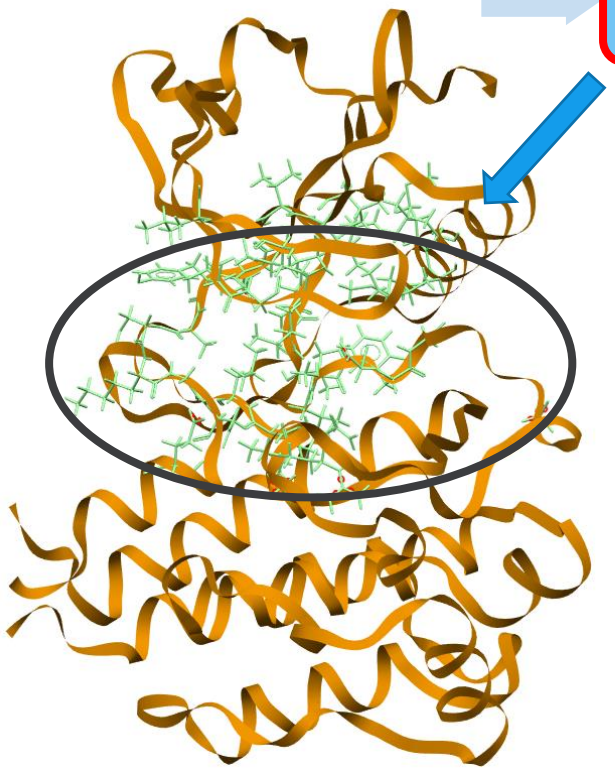
Analysis of the results and visualisation



CCDC

Steps in molecular docking

- Target selection
- Protein Preparation
- **Binding site definition**
  - Select the binding site for docking
- Ligand(s) selection for docking
- Scoring function selection
- Setup and running the docking calculation
- Analysis of the results and visualisation



Binding site: A specific region (or atom) in a molecular entity that is capable of entering into a stabilizing interaction with another molecular entity.

Read more in the Glossary in the handout .

CCDC

Steps in molecular docking

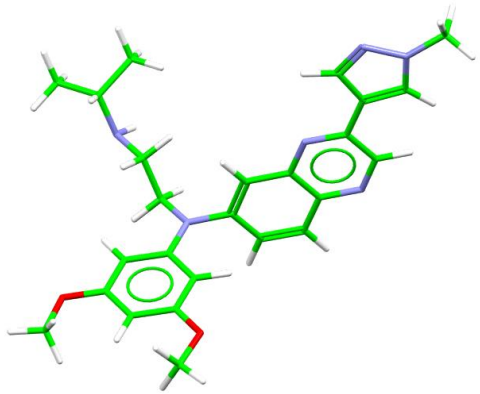
Target selection

Protein Preparation

Binding site definition

Ligand(s) selection for docking

- SDF, MOL2



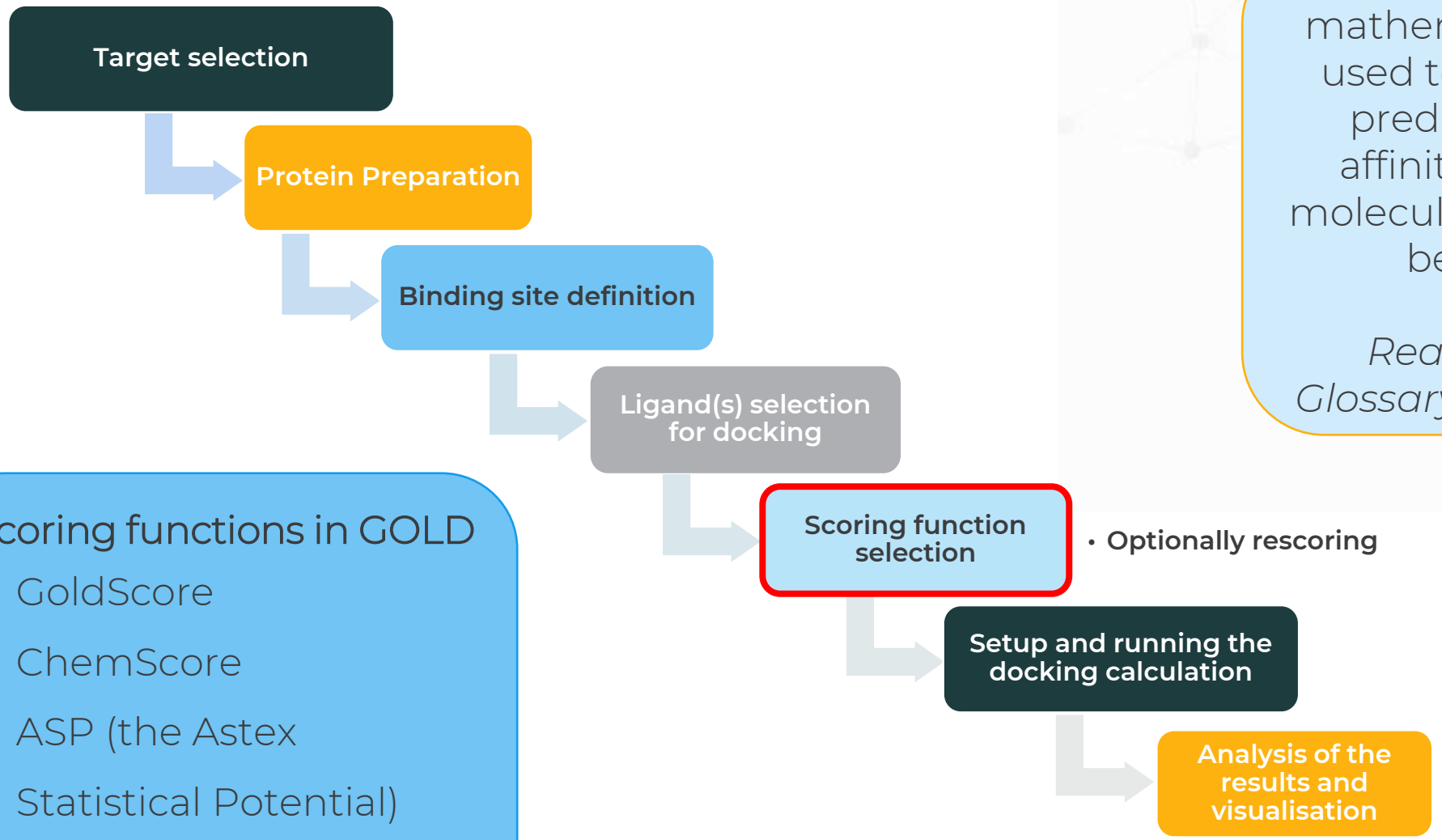
Scoring function selection

Setup and running the docking calculation

Analysis of the results and visualisation

CCDC

Steps in molecular docking



Scoring functions:
mathematical functions used to approximately predict the binding affinity between two molecules after they have been docked.

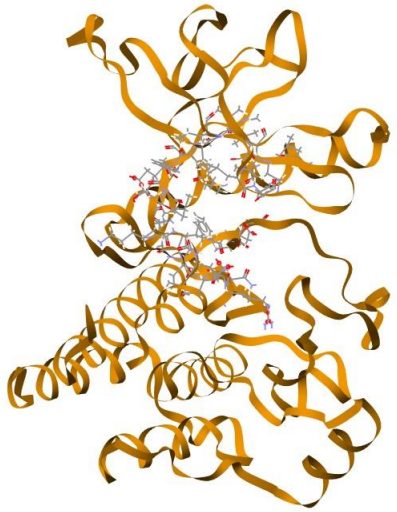
Read more in the Glossary in the handout .

Scoring functions in GOLD

- GoldScore
- ChemScore
- ASP (the Astex Statistical Potential)
- ChemPLP

CCDC

Steps in molecular docking

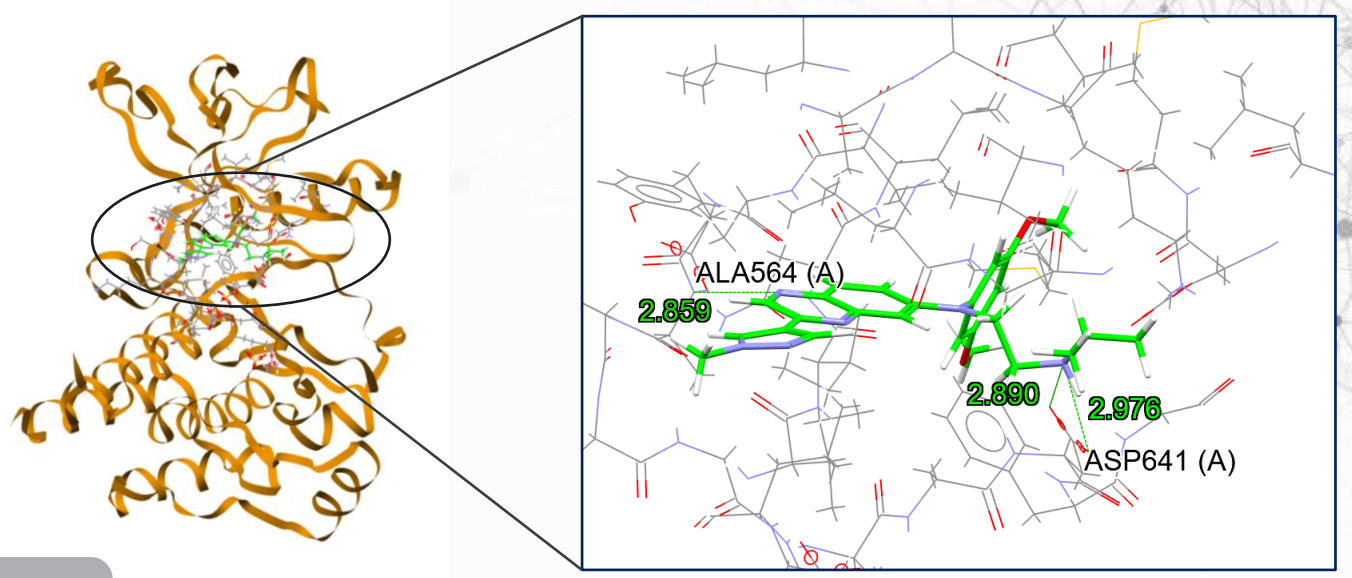


- Target selection
- Protein Preparation
- Binding site definition
- Ligand(s) selection for docking
- Scoring function selection
- Setup and running the docking calculation
- Analysis of the results and visualisation

CCDC

Steps in molecular docking

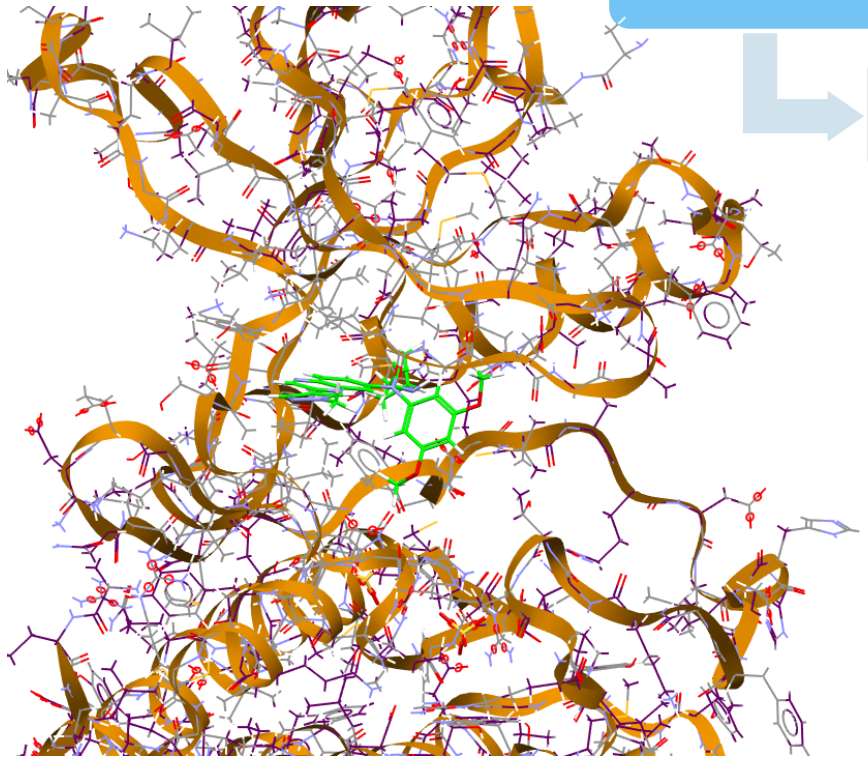
- Target selection
- Protein Preparation
- Binding site definition
- Ligand(s) selection for docking
- Scoring function selection
- Setup and running the docking calculation
- Analysis of the results and visualisation

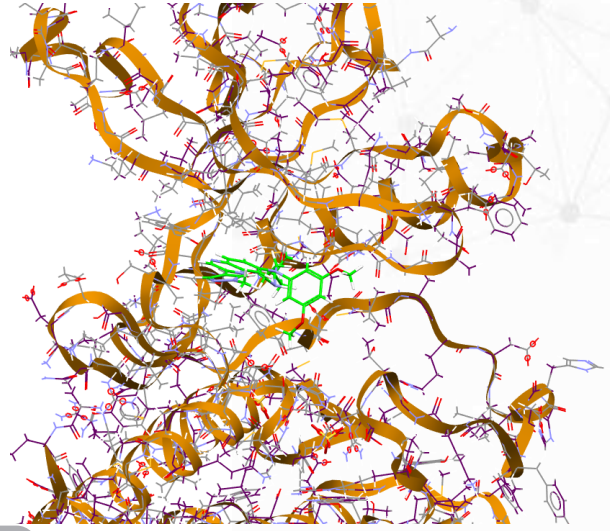


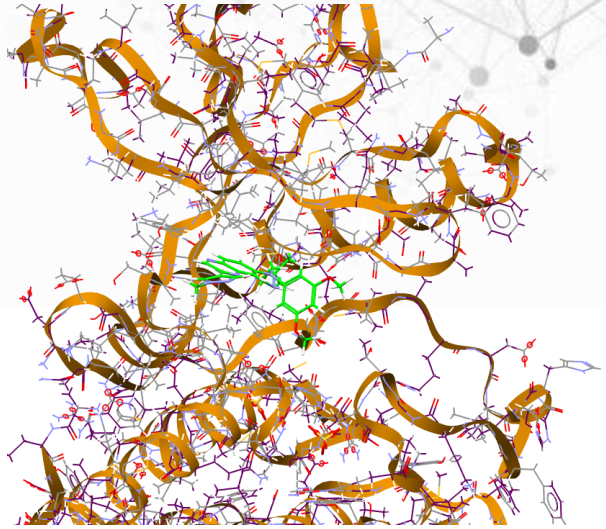
CCDC

Steps in molecular docking

- Target selection
- Protein Preparation
- Binding site definition
- Ligand(s) selection for docking
- Scoring function selection
- Setup and running the docking calculation
- Analysis of the results and visualisation



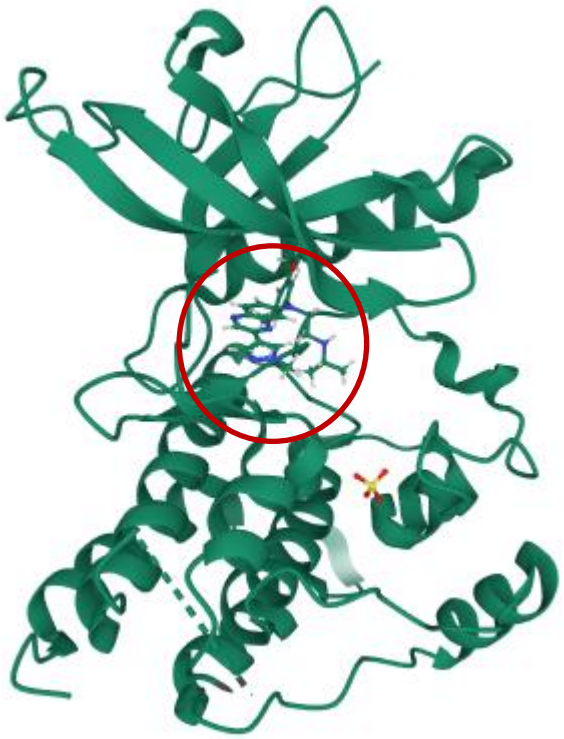


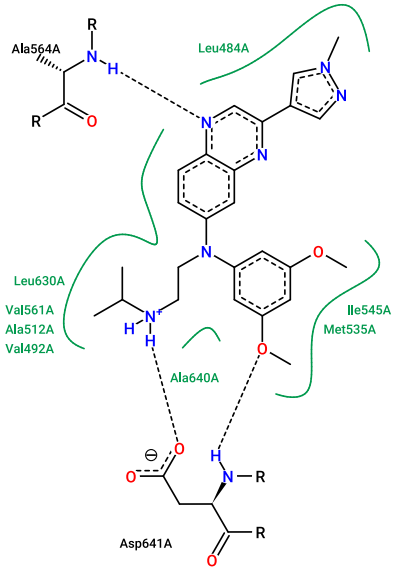


CCDC

Docking with GOLD: Case Study

PDB: 5EW8





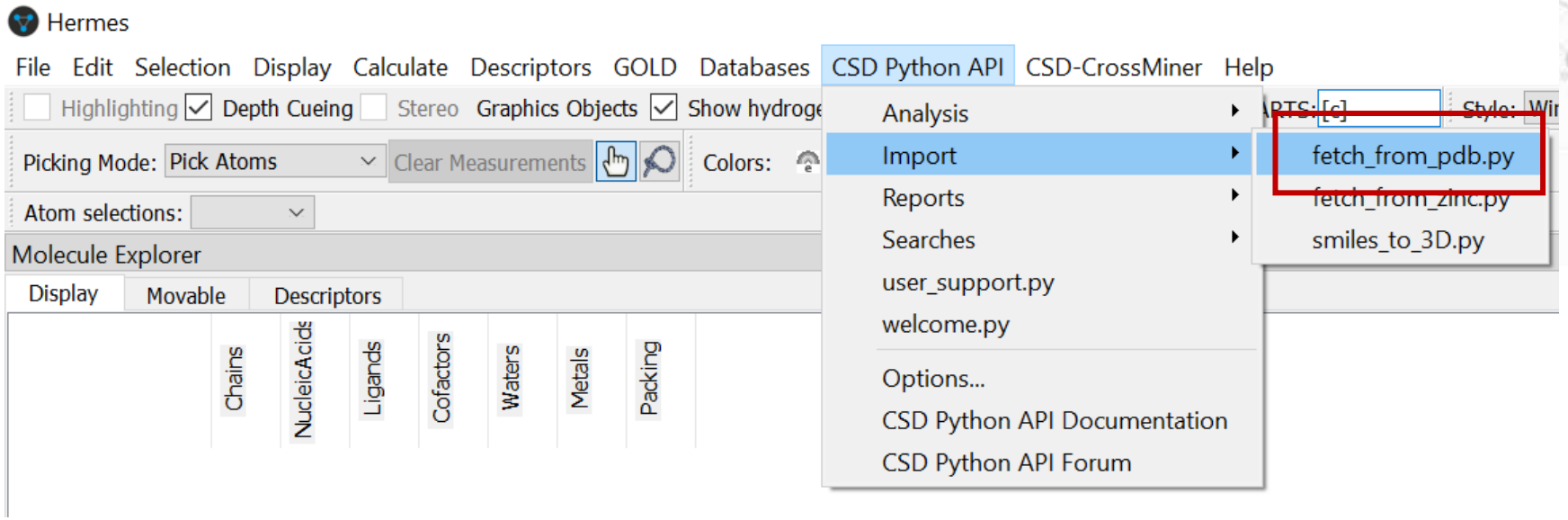
- Fibroblast growth factor receptors (FGFRs) are a family of receptor **tyrosine kinases** expressed on the cell membrane that play crucial roles in both developmental and adult cells.
- Our **ligand** is (3,5-dimethoxyphenyl)-~{N}'-[3-(1-methylpyrazol-4-yl)quinoxalin-6-yl]-~{N}-propan-2-yl-ethane-1,2-diamine), aka Erdafitinib.
- It is the first-ever fibroblast growth factor receptor (FGFR) kinase inhibitor indicated for patients with locally advanced or metastatic urothelial carcinoma.

Patani H., et al., Landscape of activating cancer mutations in FGFR kinases and their differential responses to inhibitors in clinical use. Oncotarget. 2016; 7: 24252-24268.

CCDC

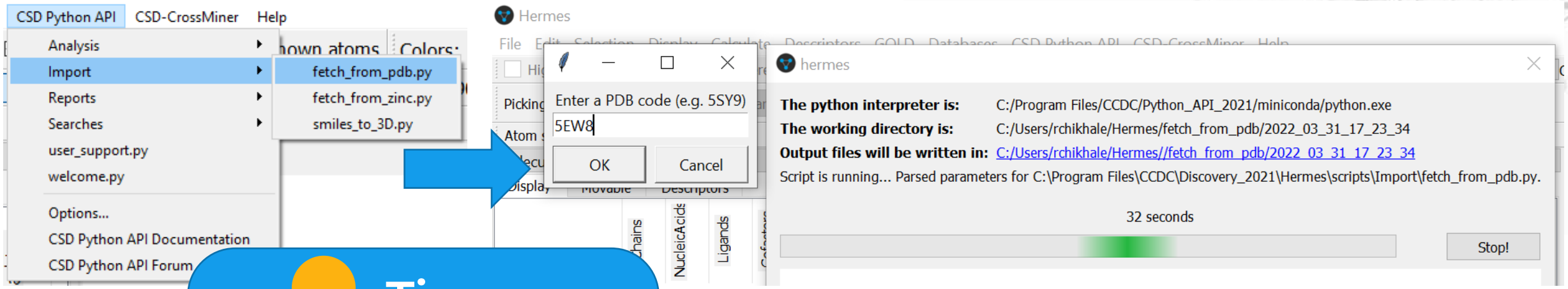
Docking with GOLD: Importing Protein

- Open the Hermes Interface and import the protein crystal structure from the PDB using the ‘fetch_from_pdb.py’ function.



Docking with GOLD: Importing Protein

- Provide with a PDB code in the '**fetch_from_pdb.py**' function search dialogue box.

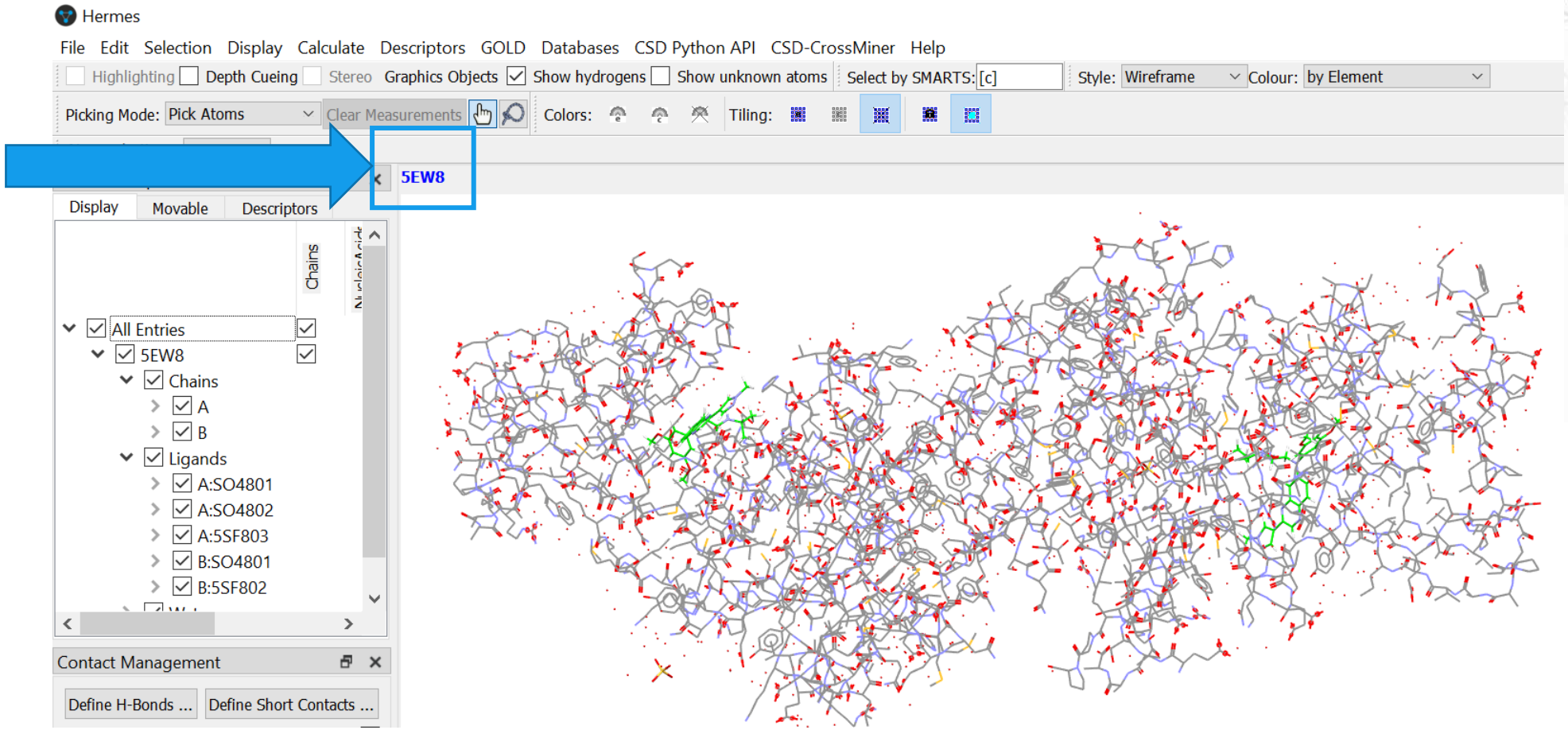


Tip

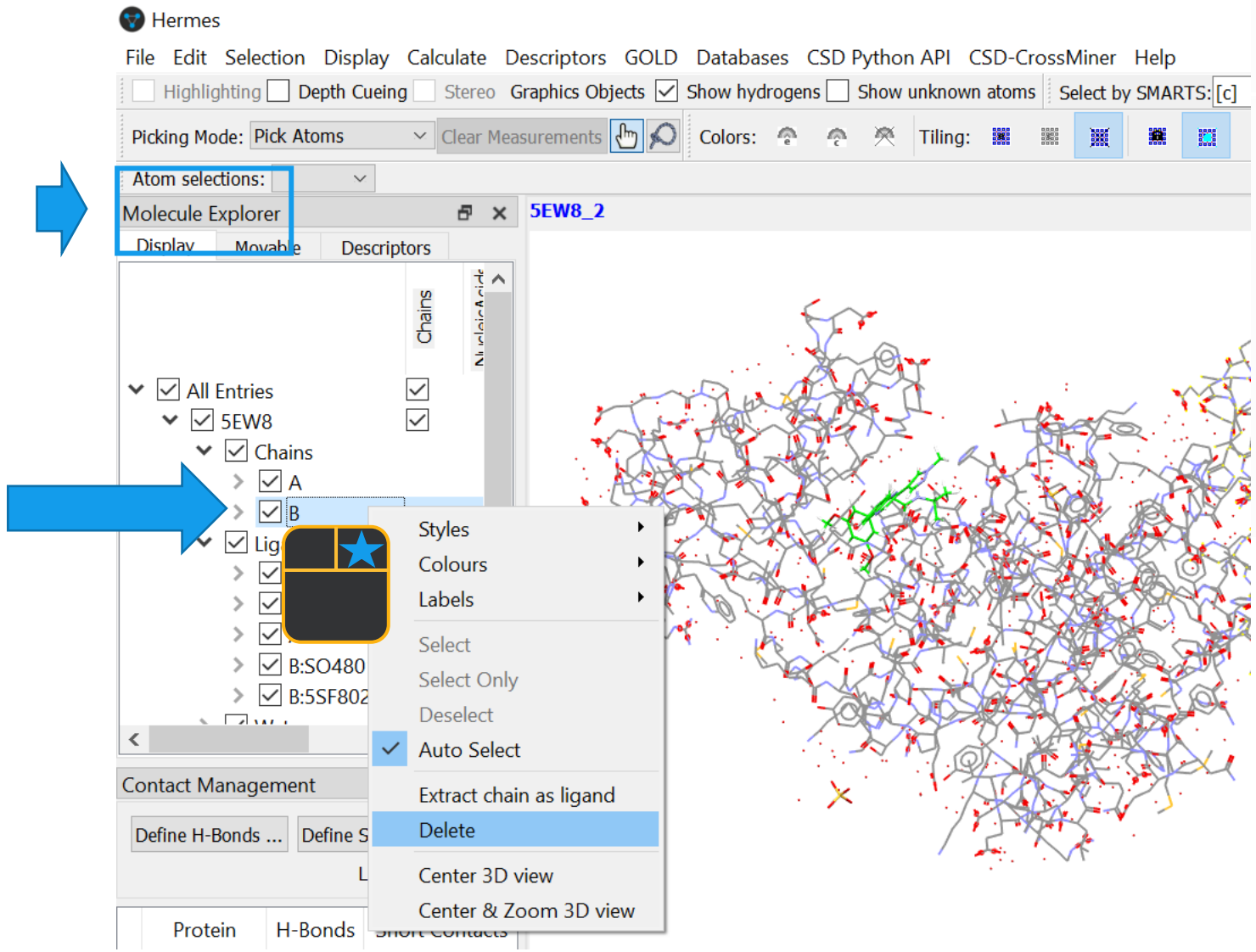
Be aware that the window for entering the PDB code might pop up in a corner of your screen.

Docking with GOLD: Importing Protein

- Once imported, the crystal structure will look like this in Hermes.



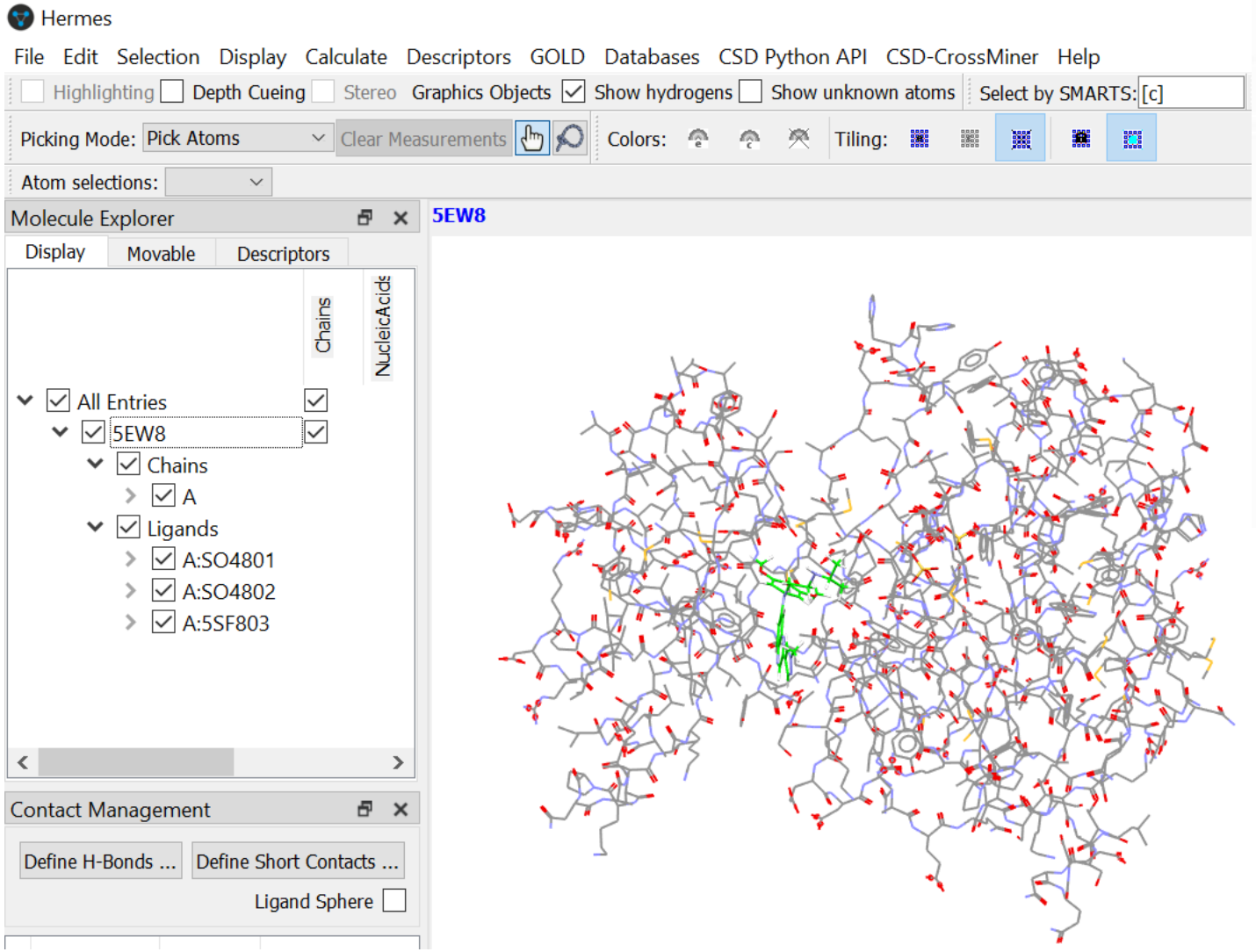
Docking with GOLD: Protein preparation



- In the *Molecule Explorer* window selected chain of the protein can be removed.
- Other components like cofactors, water molecules can be removed as per requirement.

CCDC

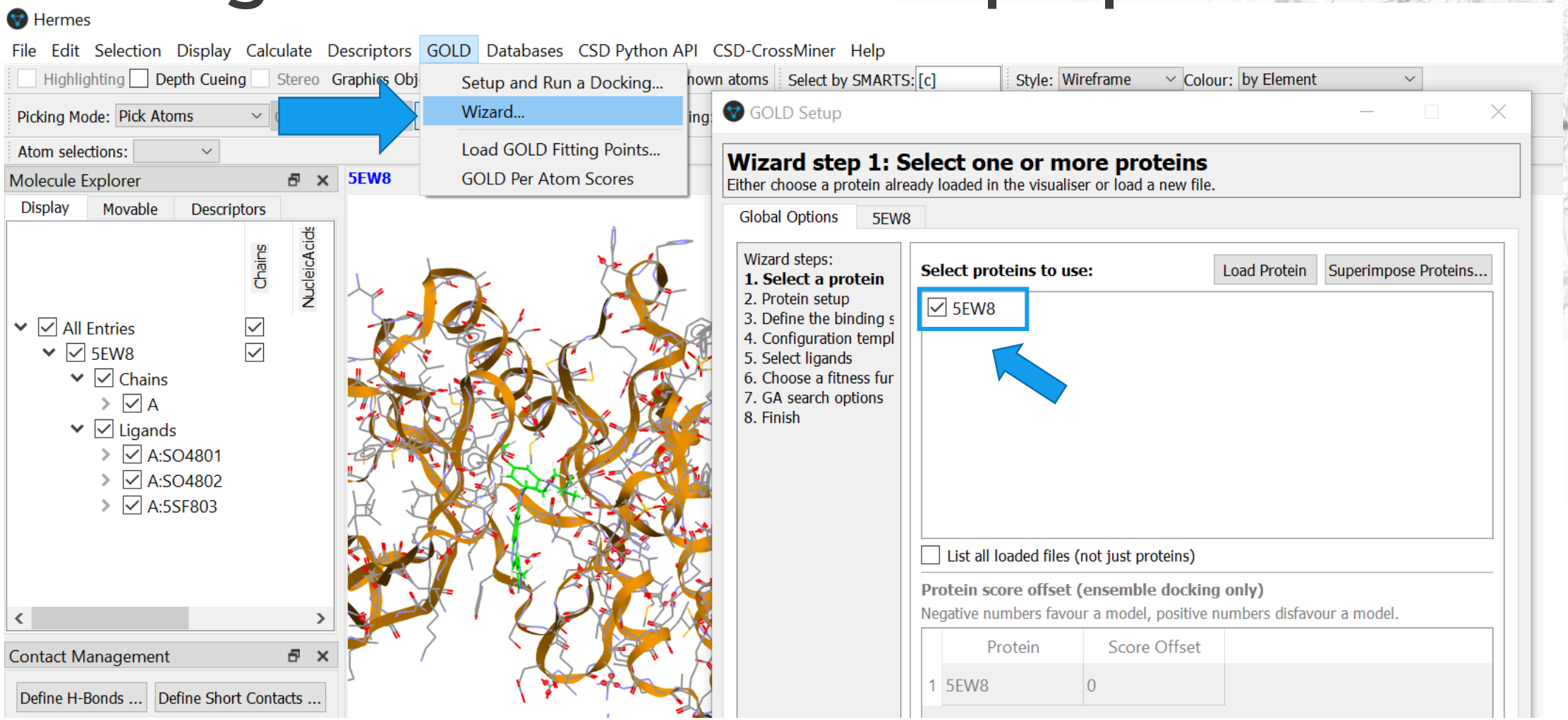
Docking with GOLD: Protein preparation



- Now we are ready for the next stage.

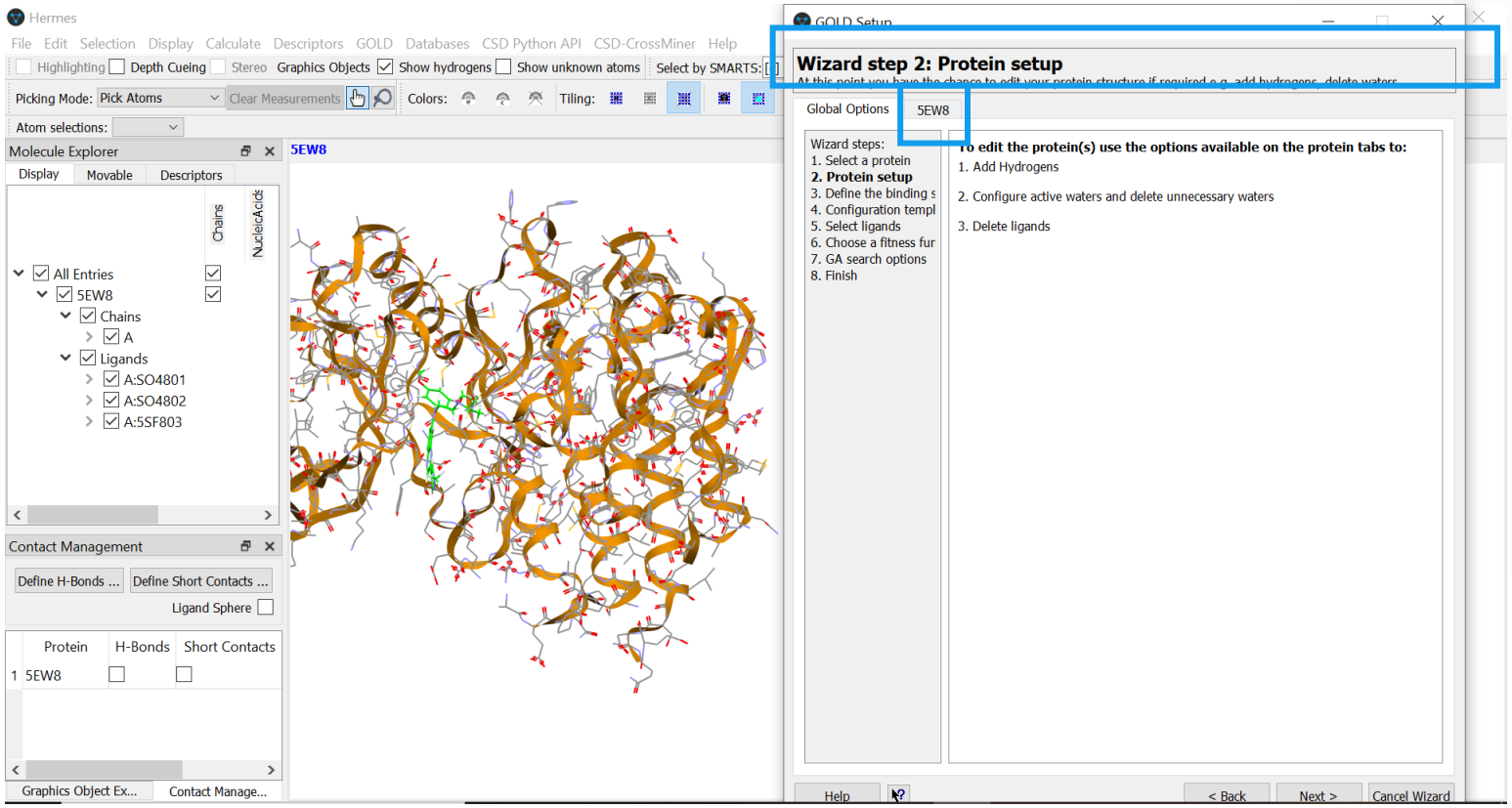
CCDC

Docking with GOLD: Protein preparation

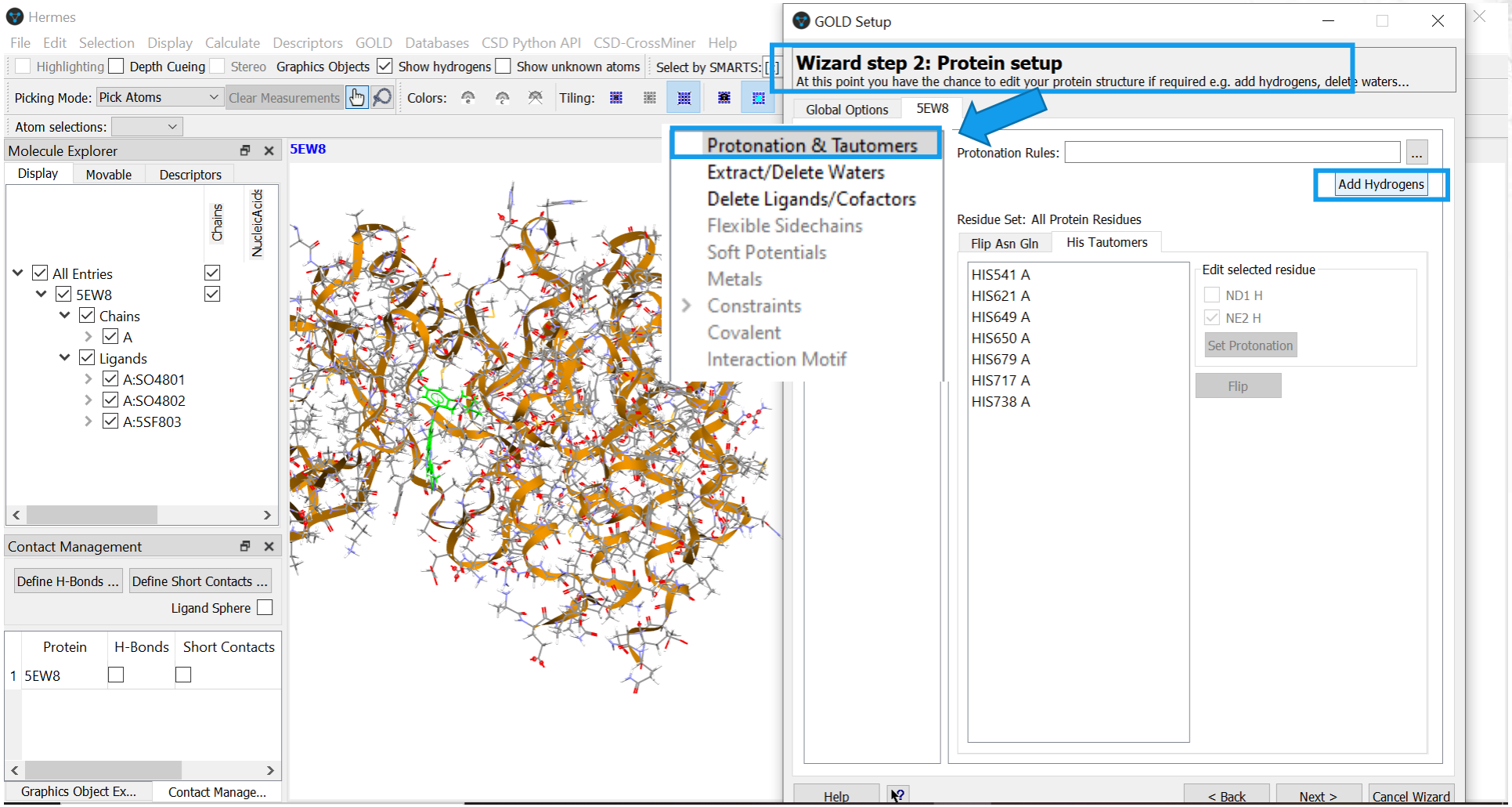


- Launch the GOLD wizard as shown above.
- Then, select the protein to investigate.

Docking with GOLD: Protein preparation



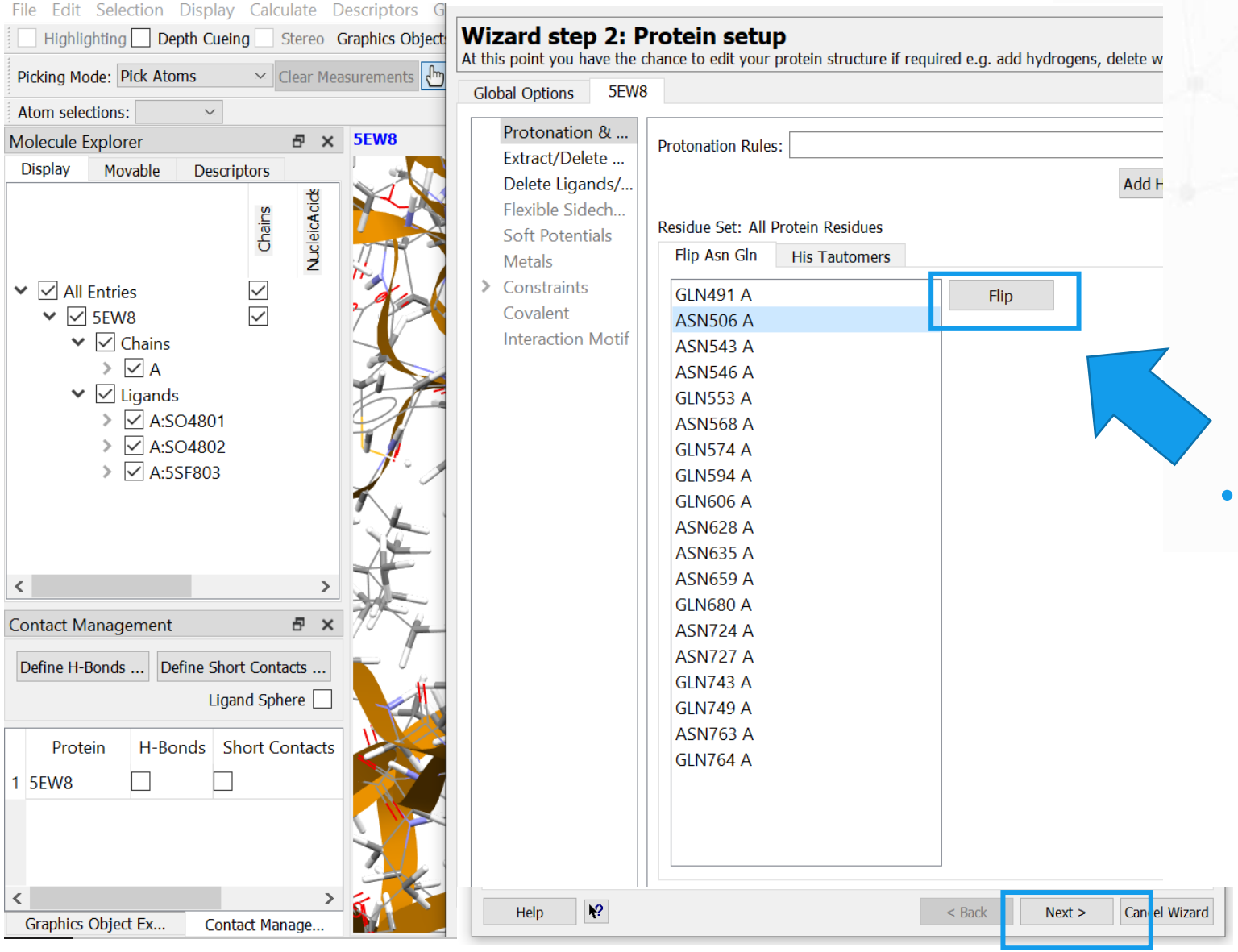
Docking with GOLD: Protein preparation



- Add missing hydrogens.

CCDC

Docking with GOLD: Protein preparation



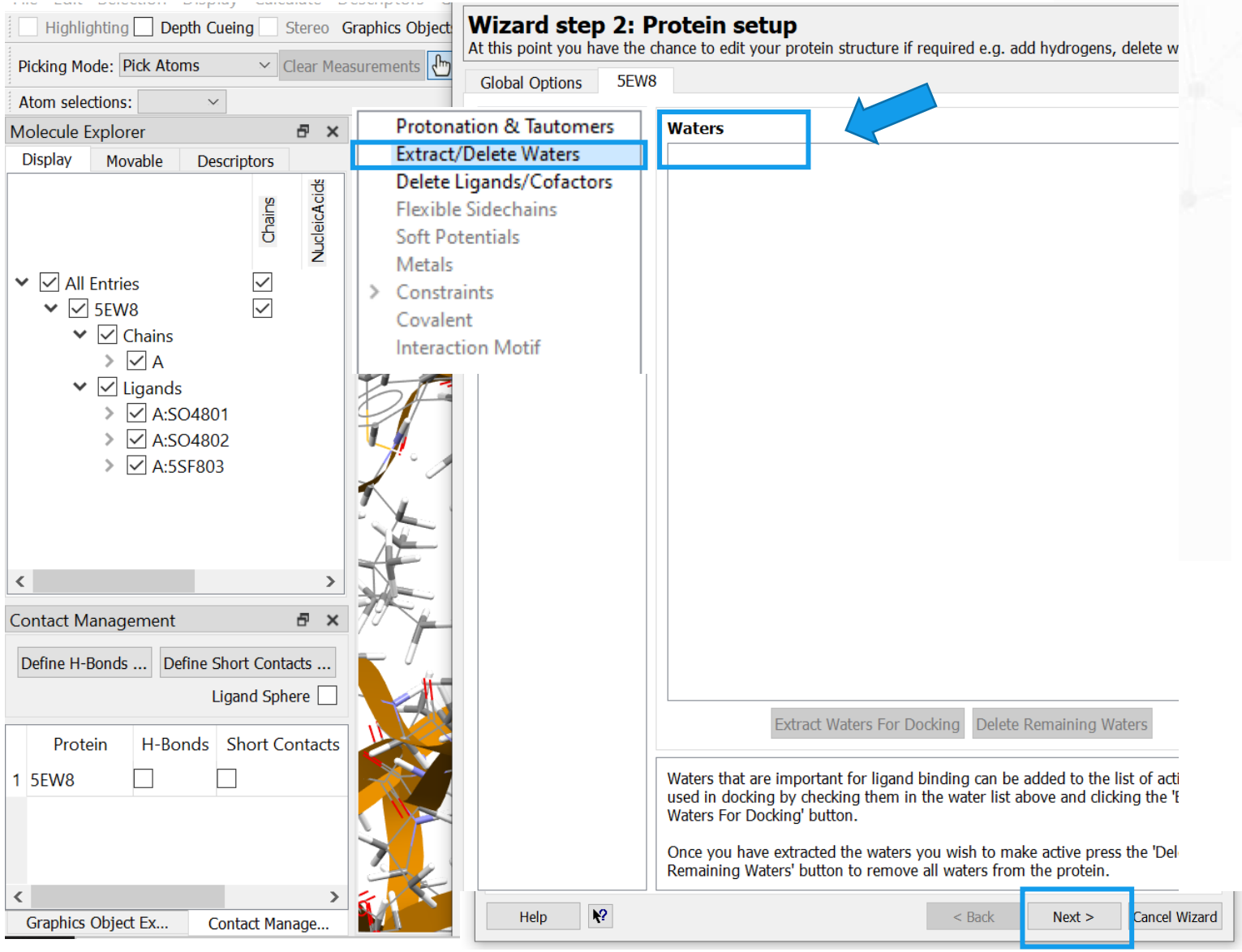
Tip

Always check the protonation around the binding site.

- You can flip the GLN and ASN based on the observation of these residues.

CCDC

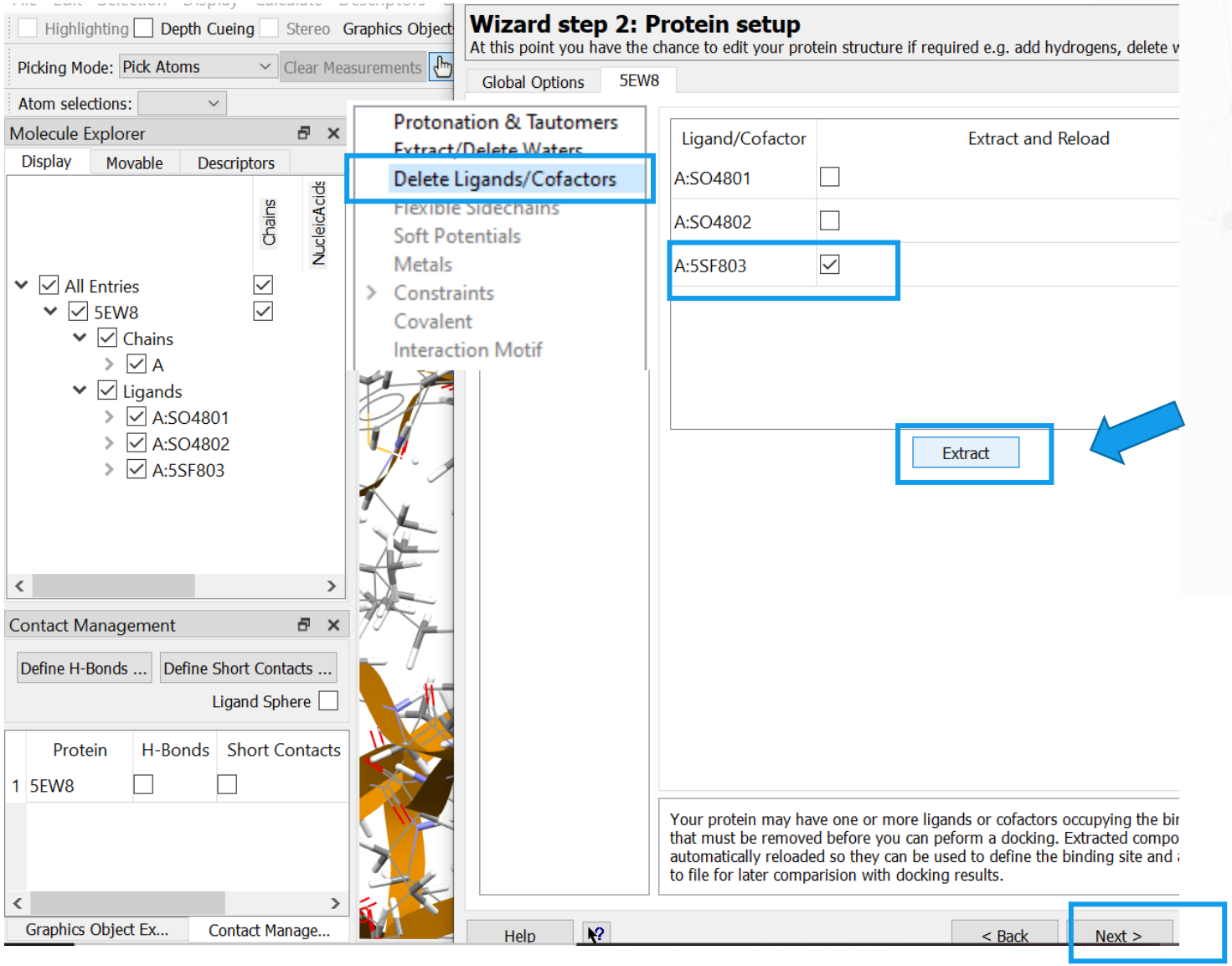
Docking with GOLD: Protein preparation



- You can extract and delete water molecules.
- In the example shown there are no water molecules to remove as we deleted them after loading the protein (before opening the GOLD Wizard).

CCDC

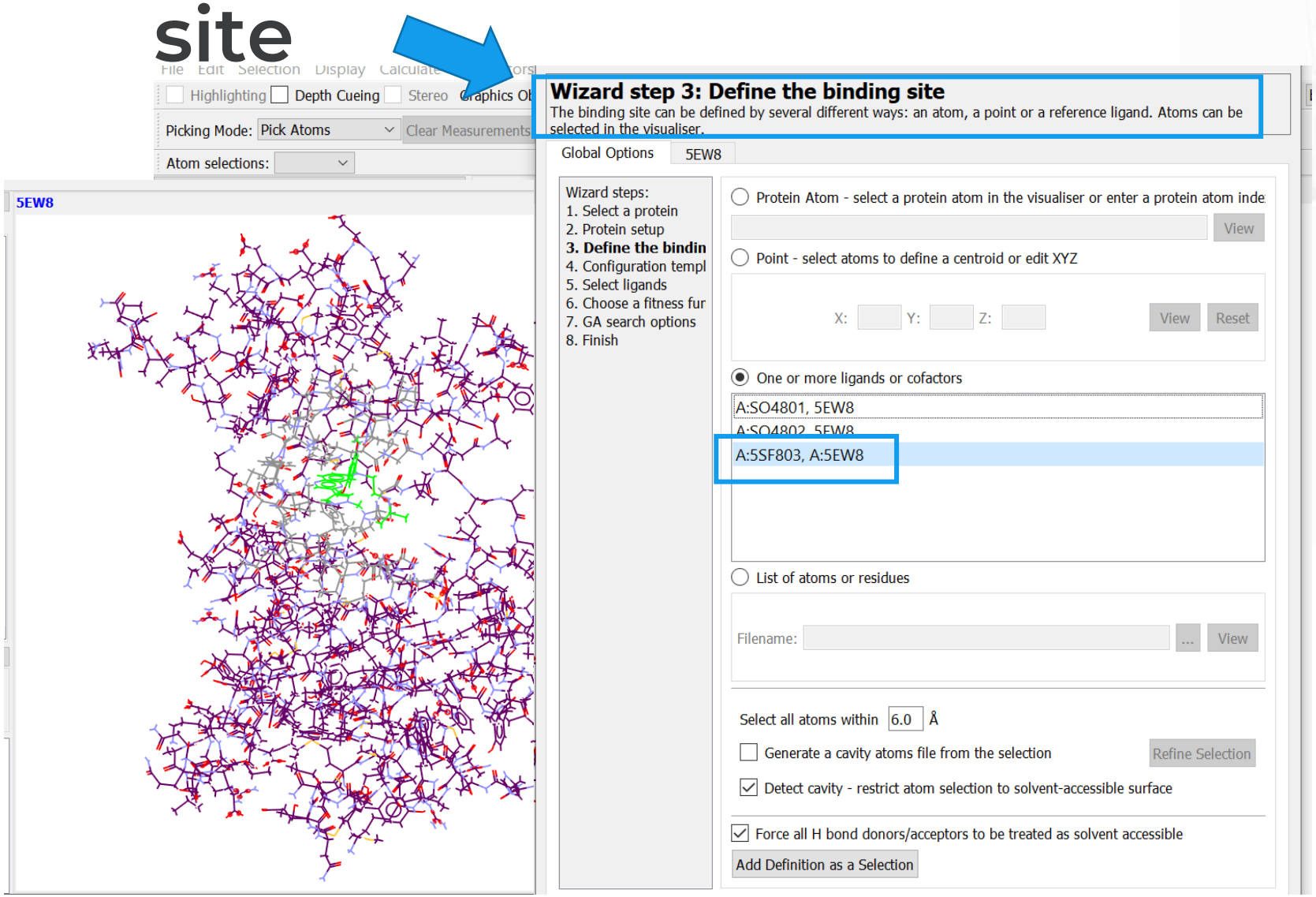
Docking with GOLD: Protein preparation



- You can extract and save the bound ligands.

CCDC

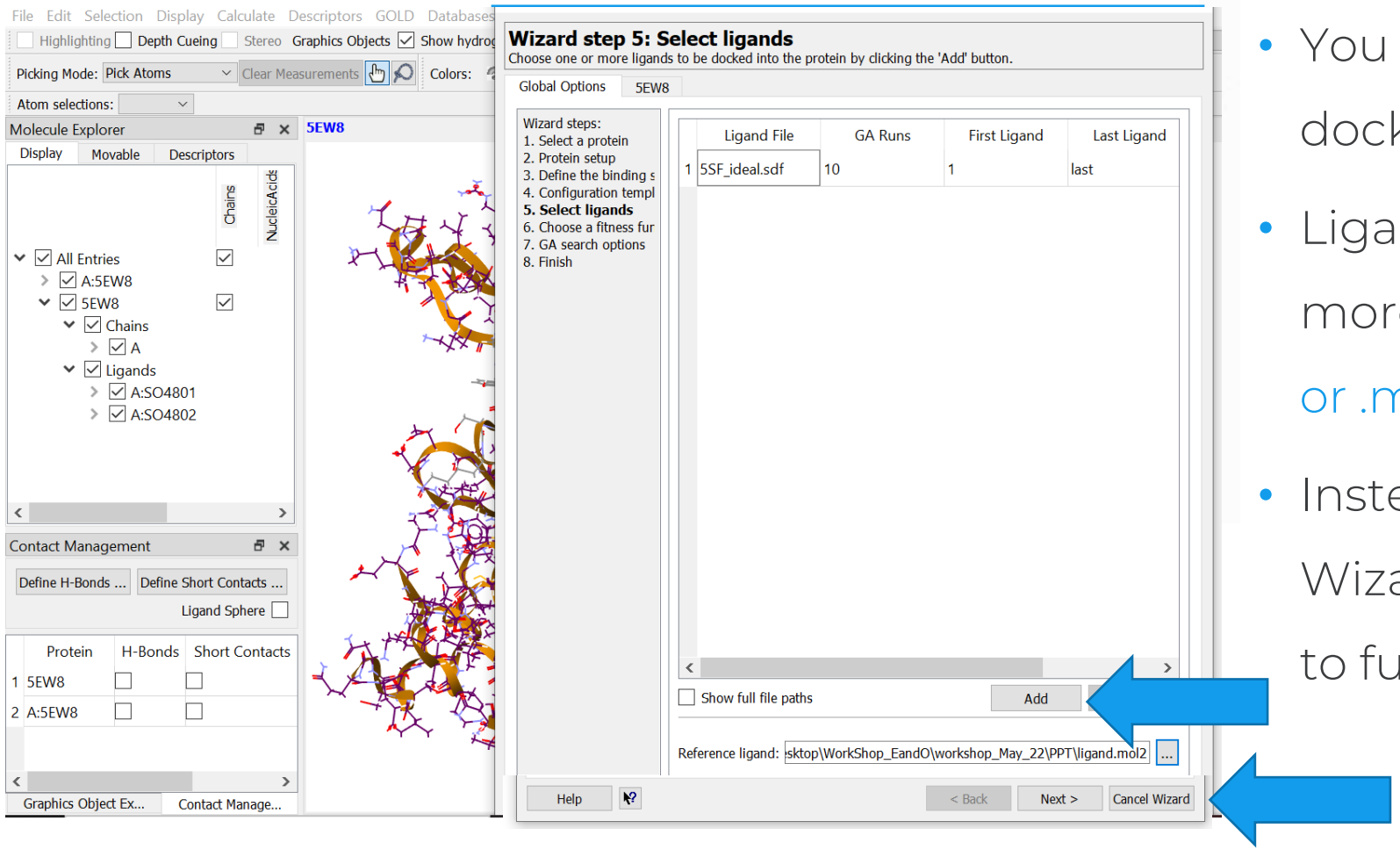
Docking with GOLD: Defining the binding site



- There are various ways in which you can define the binding site.
- Decide and select the one depending on the target protein or specific needs.
- You will see the binding site highlighted in the 3D visualizer.

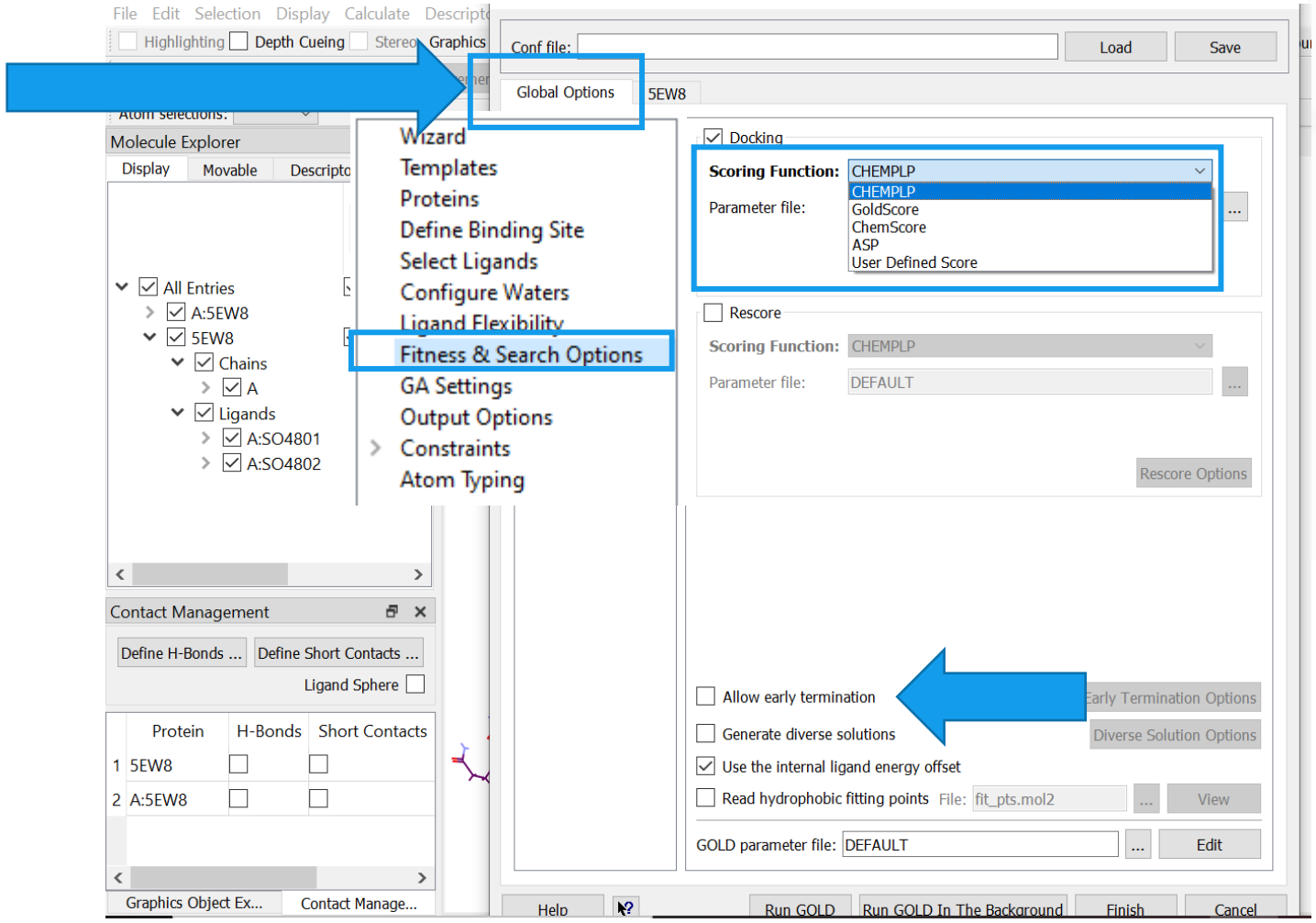
CCDC

Docking with GOLD: Select ligand/s



- You can select the ligands to be docked.
- Ligand file could contain one or more than one molecule in .sdf or .mol format.
- Instead of continuing with the Wizard, cancel the wizard now to further configure.

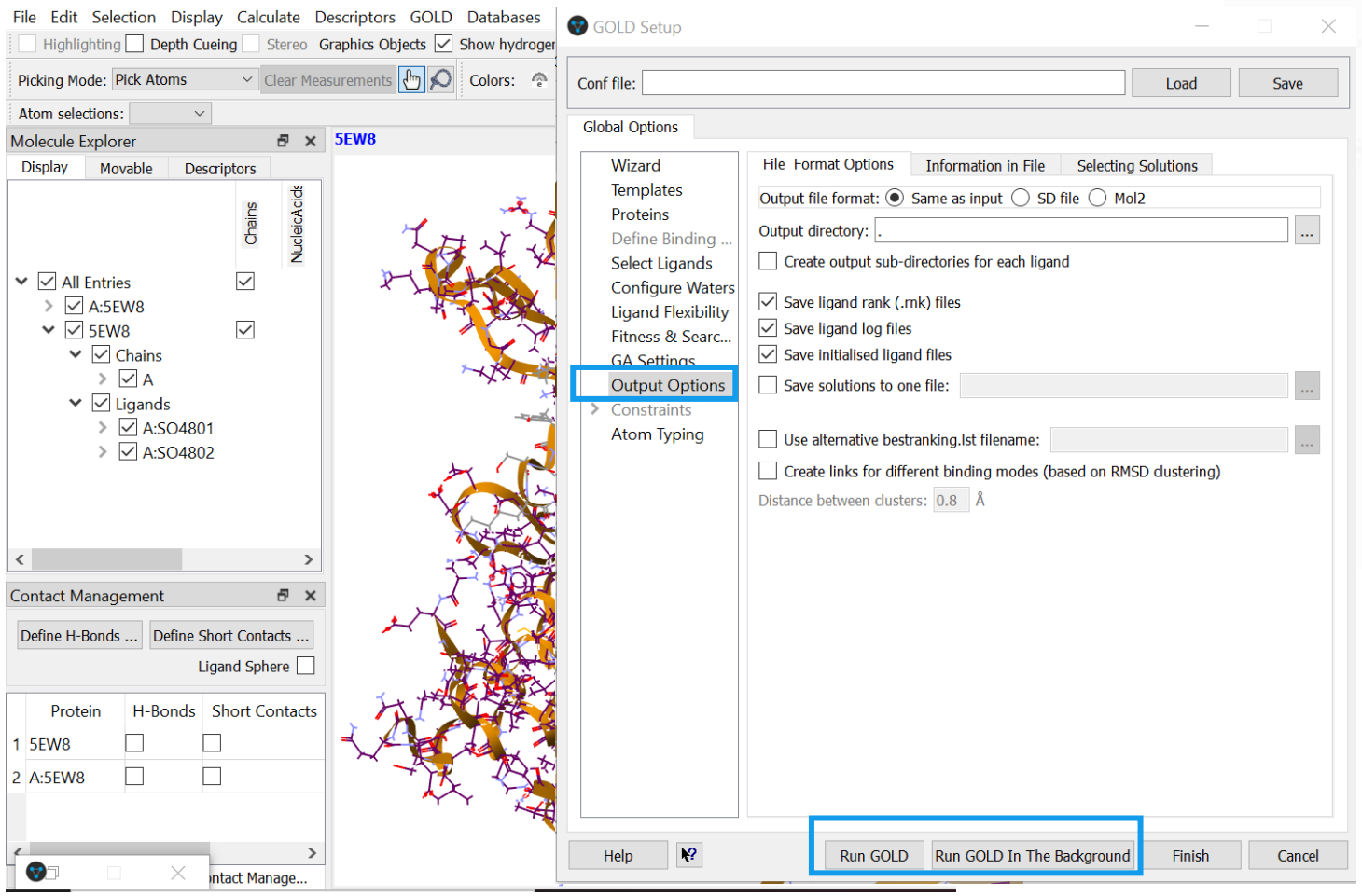
Docking with GOLD: Scoring function



- Under the Global Options, Fitness and Search Options you can select the Scoring Function.
- Here rescoring function is available for generating consensus.

- Untick *Allow early termination*

Docking with GOLD: Run the setup



In Output Options

- Select output destination/directory.
- Format of output you require.

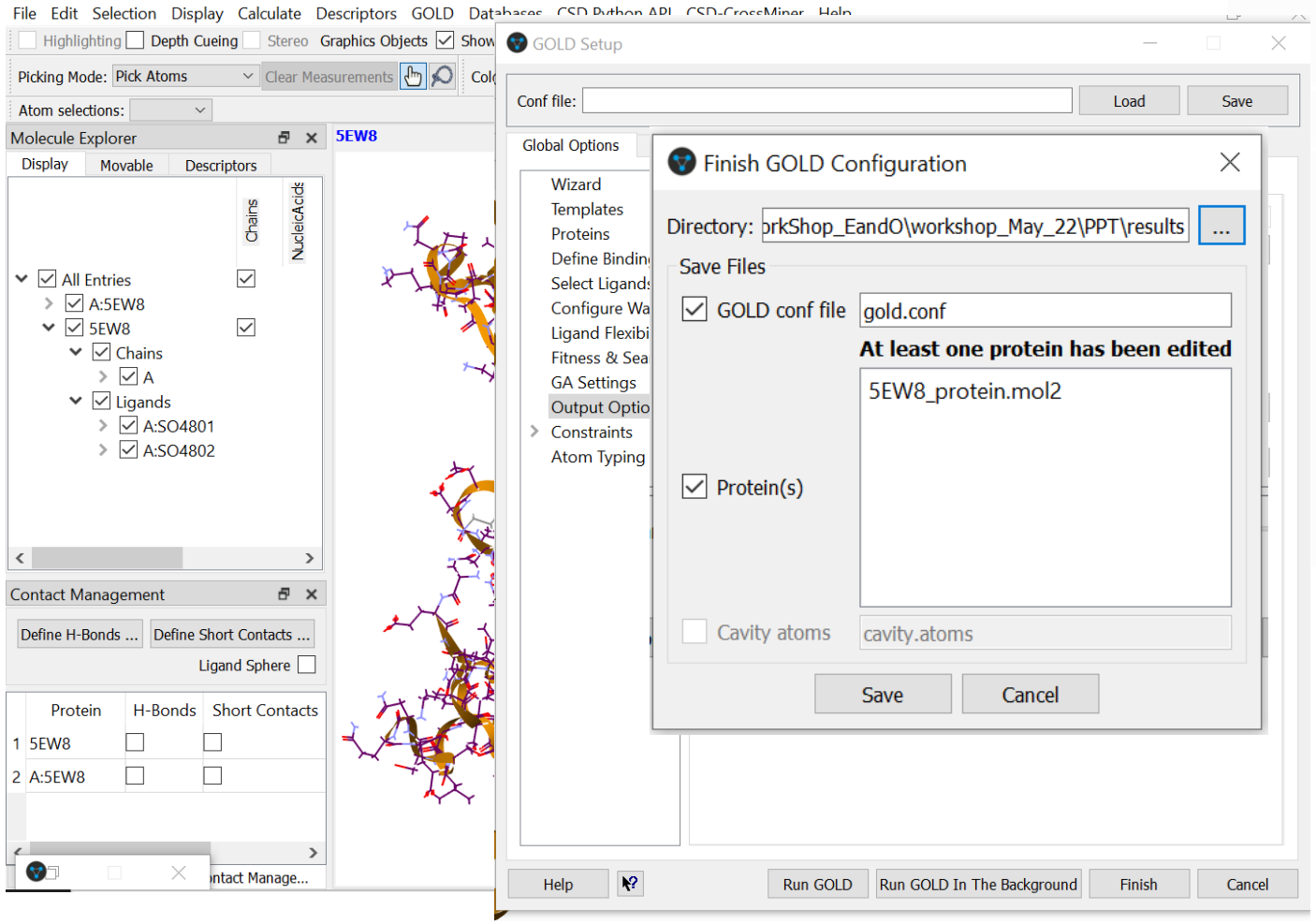
Run the GOLD calculations.

- Two options are available;

1. Run Gold: Interactive mode.

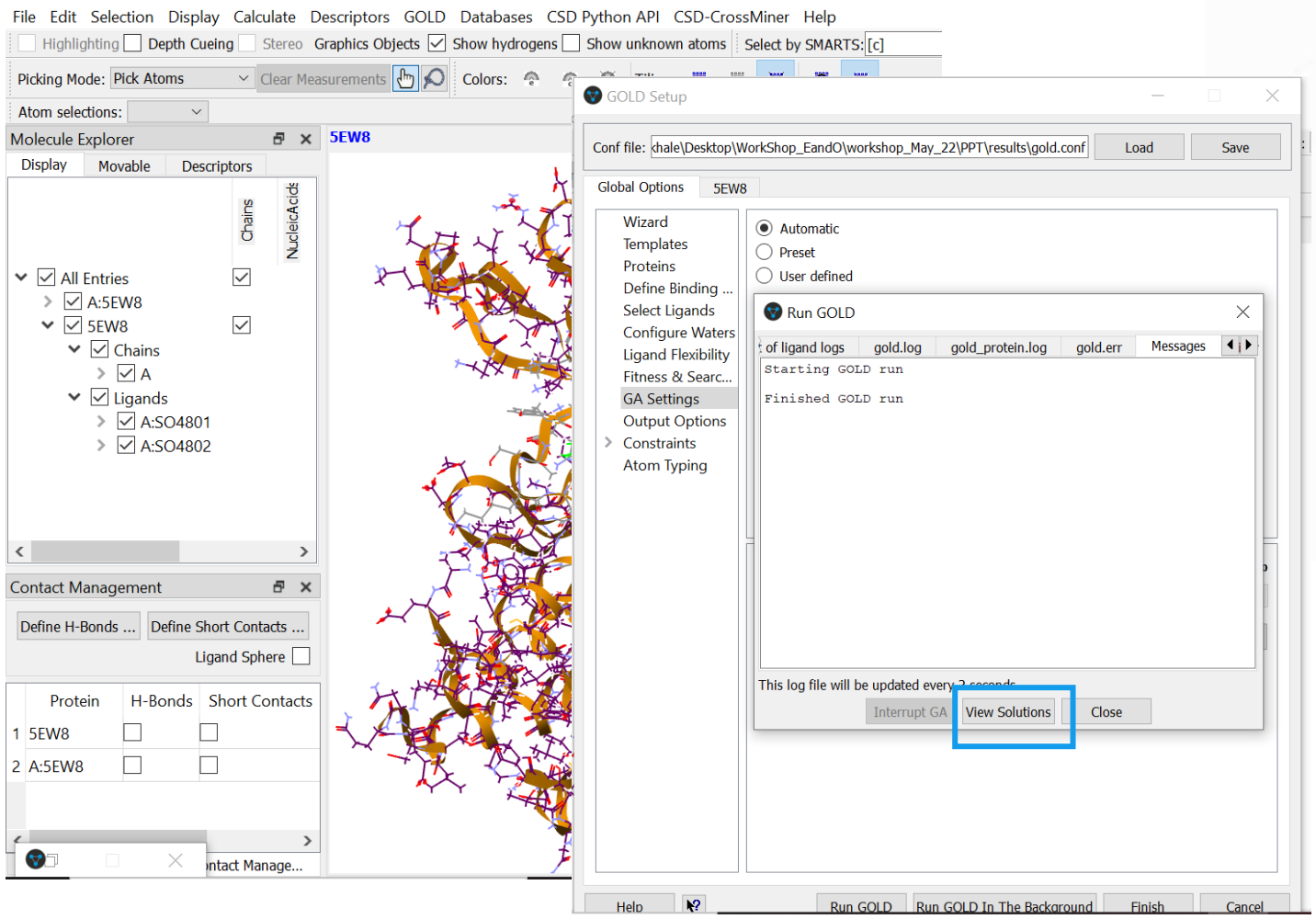
2. Run GOLD in Background: Results are seen in the output folder.

Docking with GOLD: Run the setup



- The run function asks for various options.
- The *gold.conf* file is the one with all the details. It is an editable file.

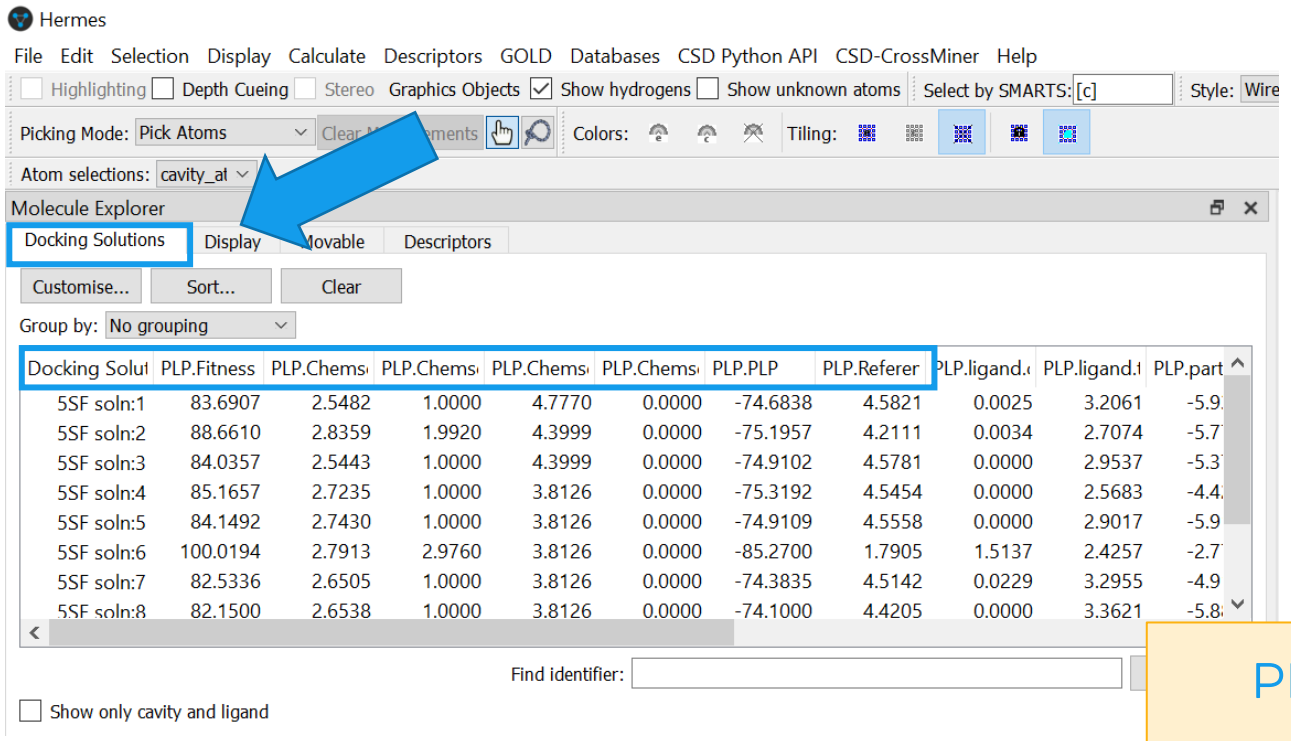
Docking with GOLD: Docking calculations



- You can check the progress of the calculation.
- The *Run GOLD* window has various tabs, which provide with on-the-fly status of your calculations.
- Once the run is complete, you can click *View Solutions*.

Docking with GOLD: Analysis of results

- Results are displayed in the *Molecular Explorer*.

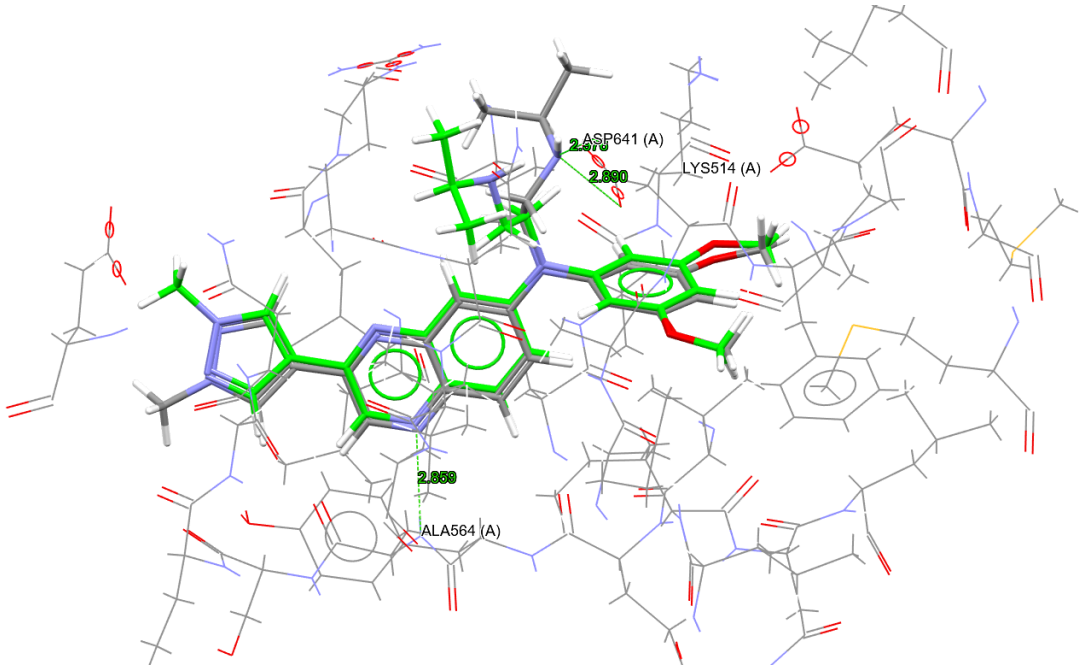


- In most cases we are interested in the PLP fitness score and the RMSD.
- For PLP scoring function, higher scores and lower RMSD means better results.

PLP: Piecewise Linear Potential, it is an empirical fitness functions optimised for pose prediction.

CCDC

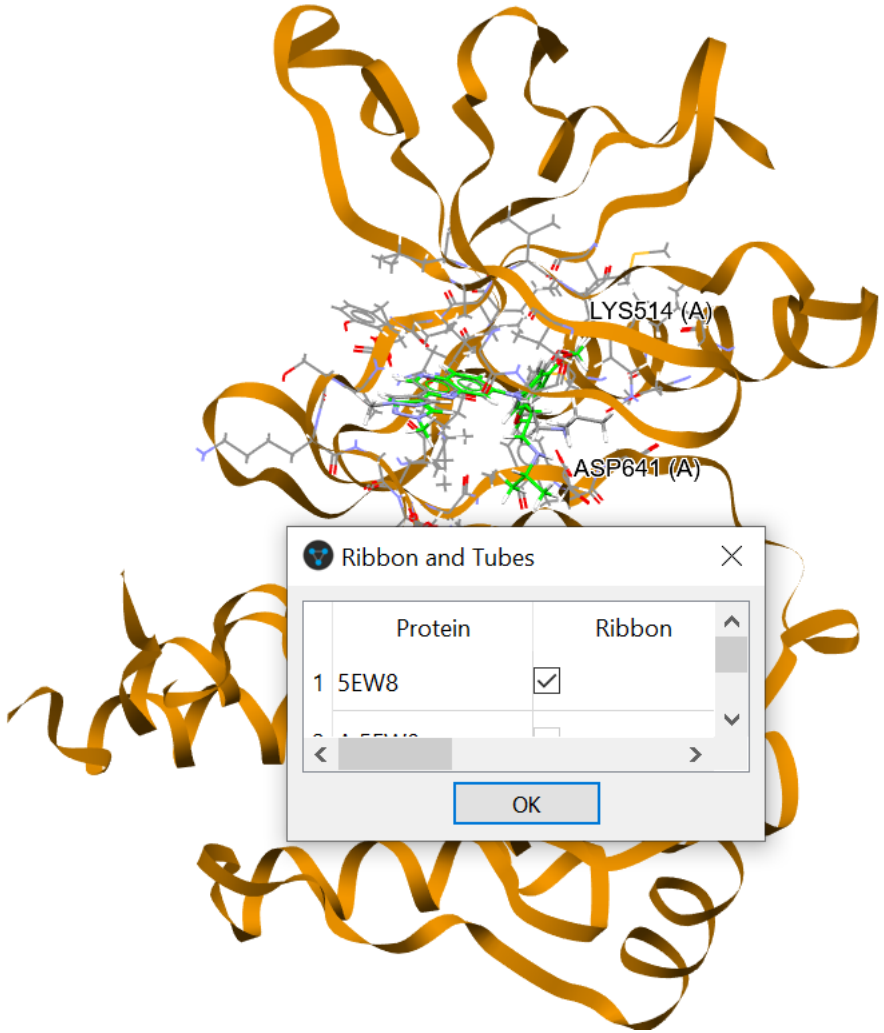
Docking with GOLD: Analysis of results



- Use the *Molecular Explorer* to display docking solutions, component of the system and the molecules.
- Manage the views and study the interactions.

CCDC

Docking with GOLD: Analysis of results



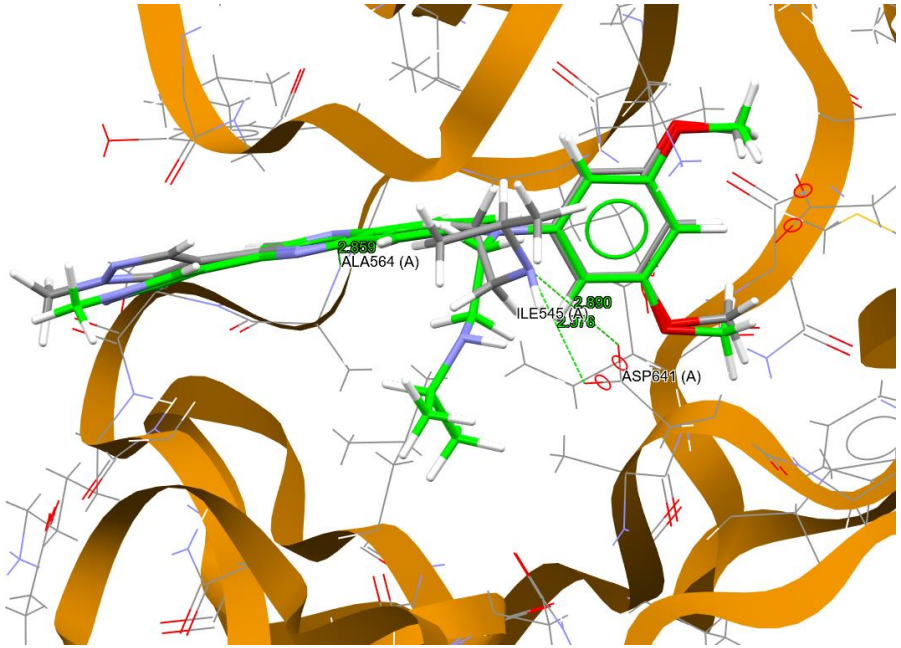
- Select Display

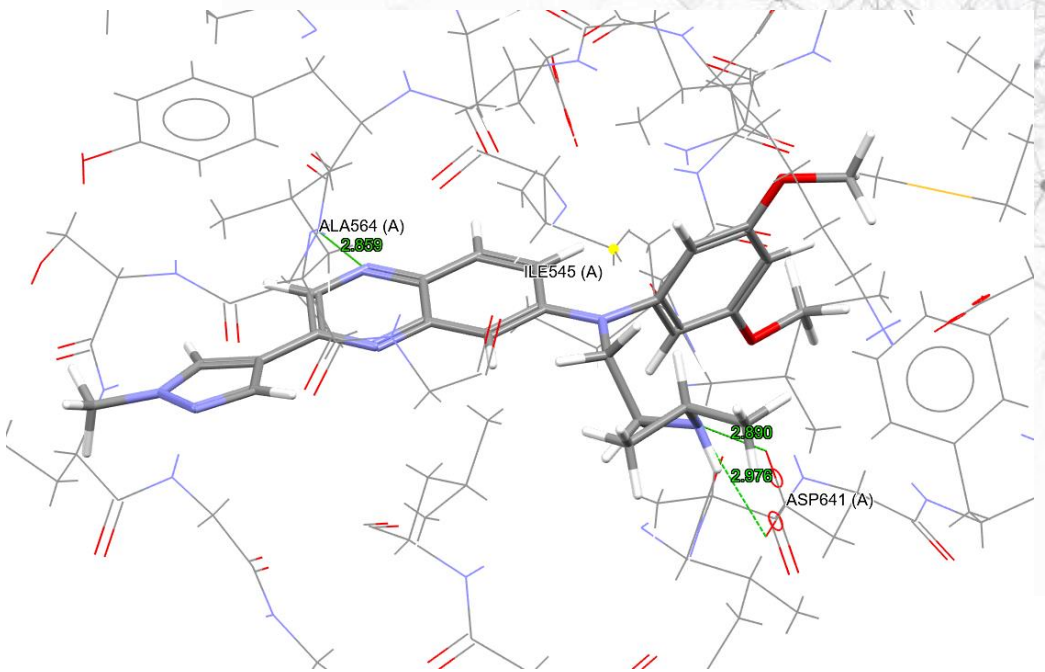
Ribbons and Tubes

Explore various options to create various colour combinations and displays

CCDC

Docking with GOLD: Analysis of results

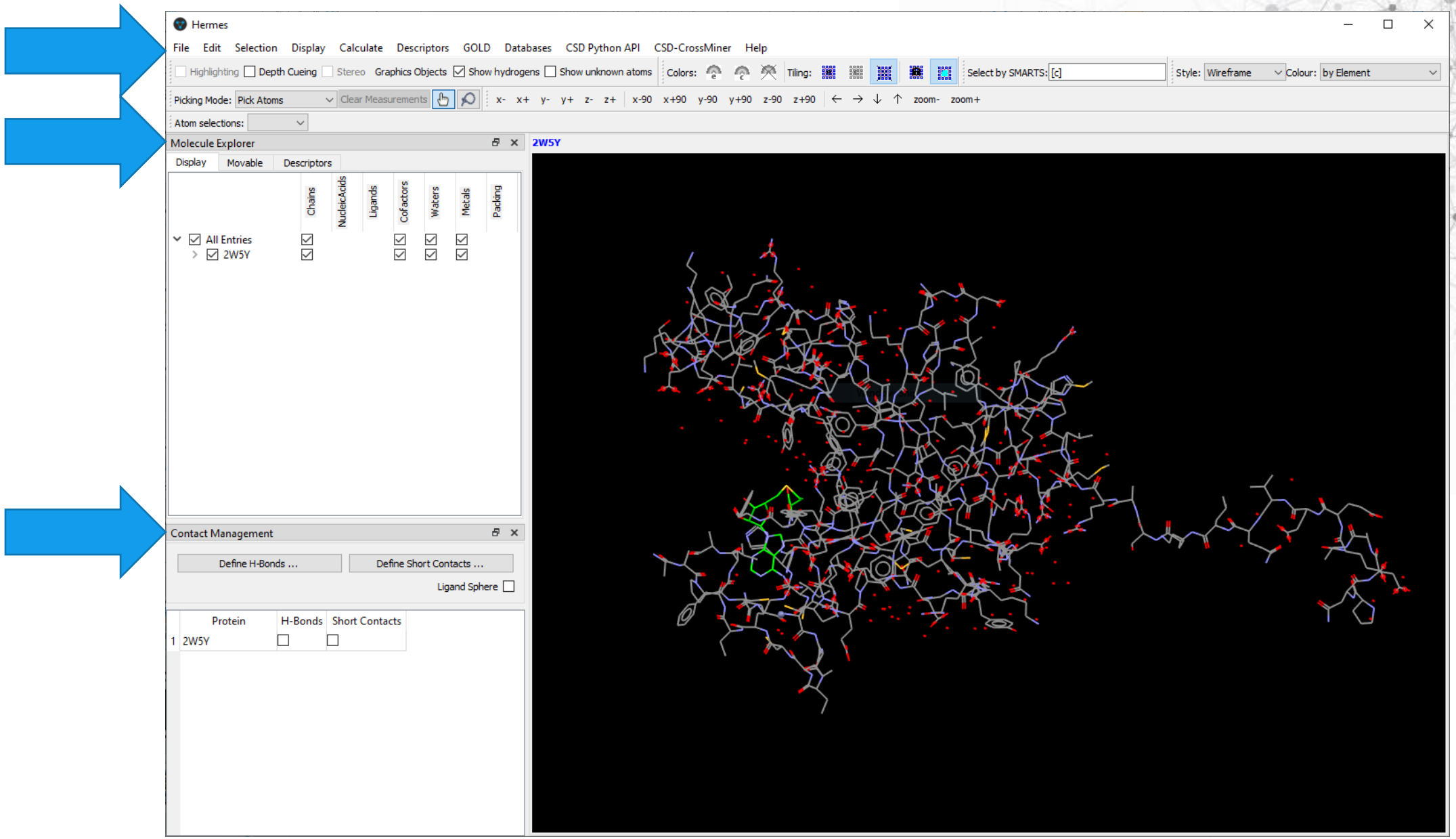




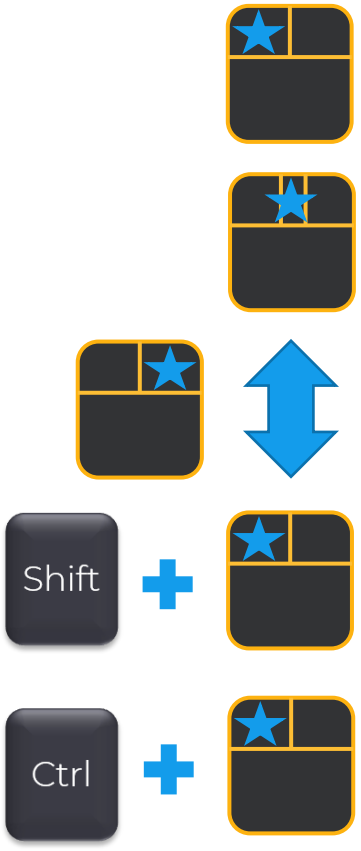
- Results can be visualised in various ways and representations.
- Can display hydrogen bonds for the docked and reference ligands.

CCDC

Show One: Hermes Interface



The 3D window basics

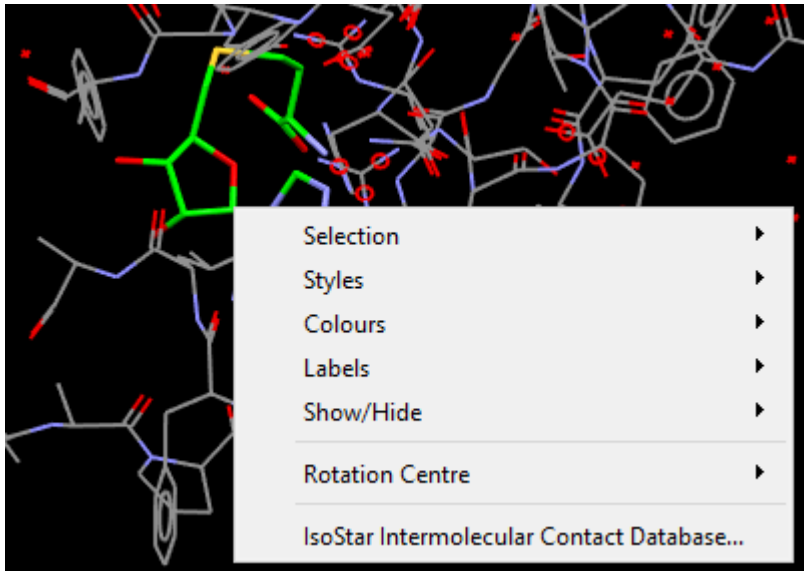


- Left mouse button and move – rotate molecules
- Middle Mouse wheel – move molecules up and down
- Right mouse button and move up and down – zoom in and out of molecules
- Shift + Left mouse button and move - rotate in the plane molecules
- Ctrl + Left mouse button and move - translate molecules

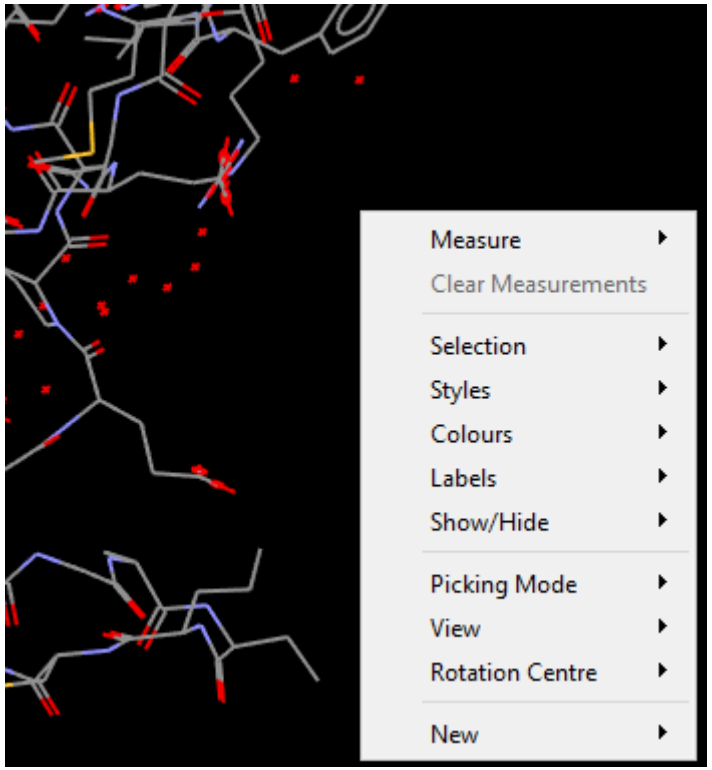
CCDC

The 3D window basics – Right click

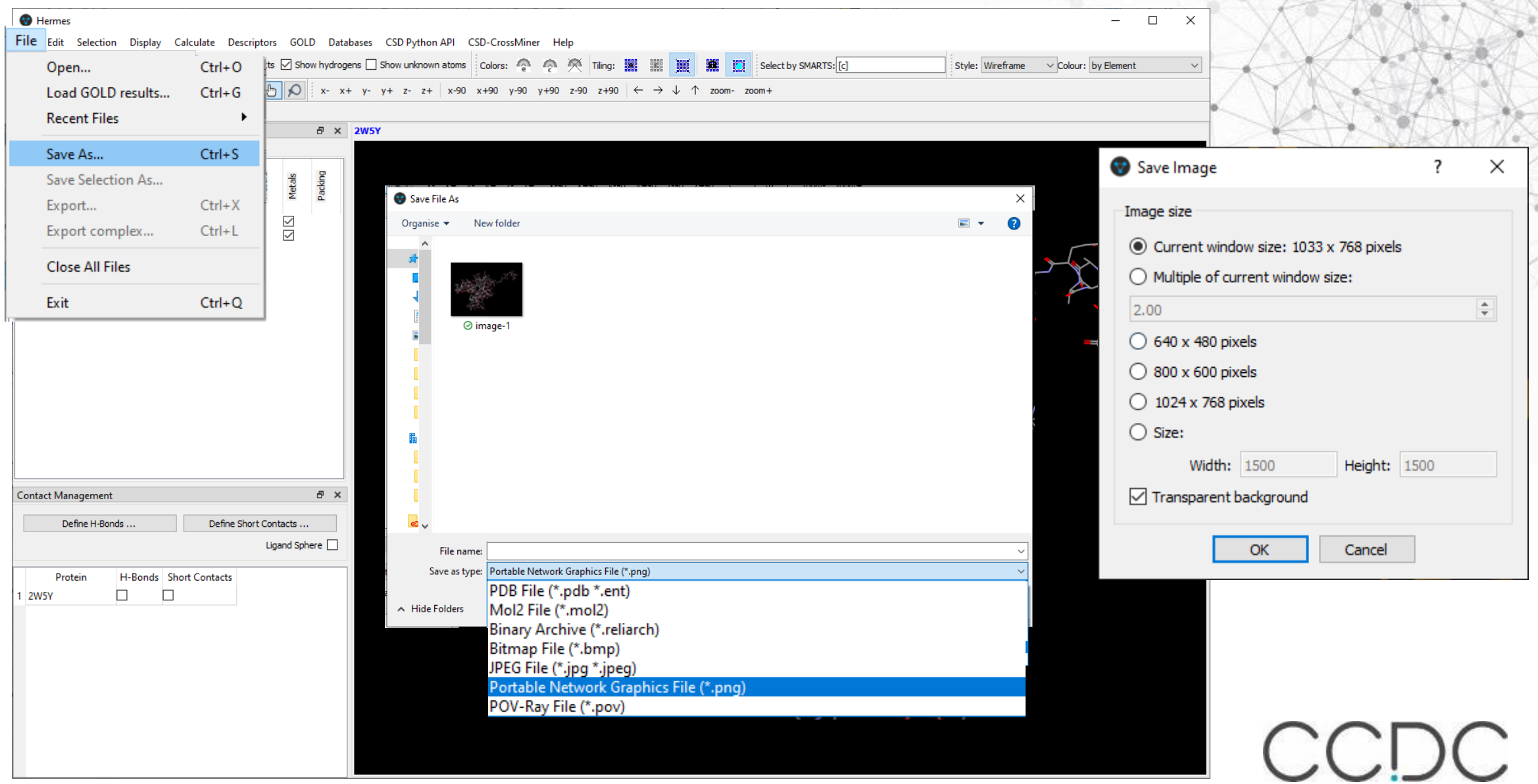
On a feature



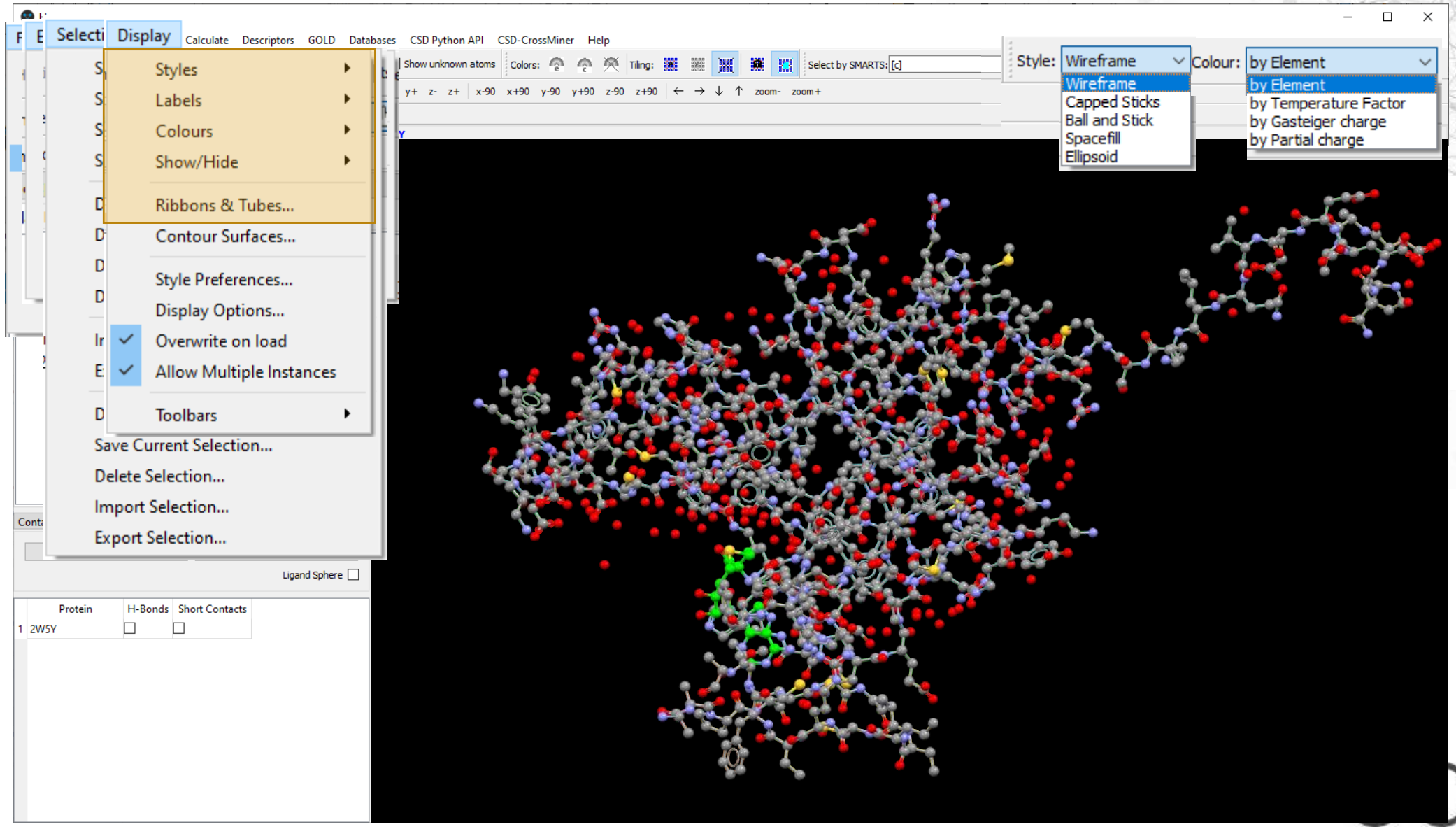
Away from a feature



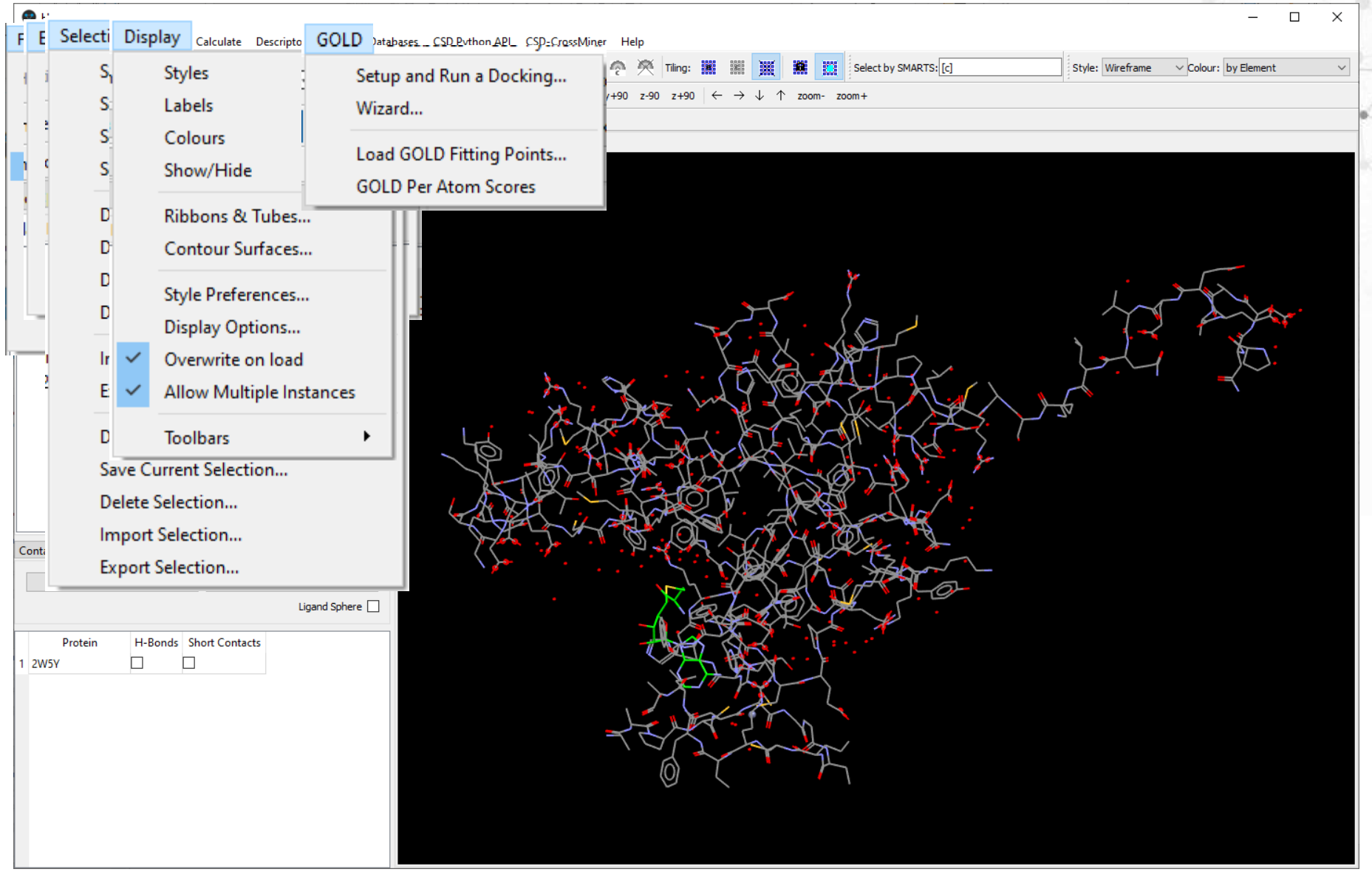
Show One: Hermes interface



Show One: Hermes interface



Show One: Hermes interface



Explore More: what can you do with GOLD?

- There is not enough time to explore more advanced functionality today, but we will briefly introduce some extra tips and examples.

CCDC

GOLD: Complete molecular docking suite

Pose prediction

Validate your ligand docking results and optimise hits to leads.

Highly configurable constraints

Use your existing knowledge of the system to bias results and focus on known features and behaviours.

Multiple scoring functions

Score and rescore to build a full picture of your system or perform consensus scoring.

Flexible docking

Perform ensemble docking or handle flexible side-chains with soft potentials.

Water handling

Assess how structural waters affect binding, see if the ligand displaces waters or mediates the interaction during docking.

CCDC

GOLD: Complete molecular docking suite

Virtual screening

Unlimited potential with virtual screening powered by cloud or cluster (HPC).

Python API access

Run dockings programmatically - for parameter optimisation and workflow incorporation.

KNIME component

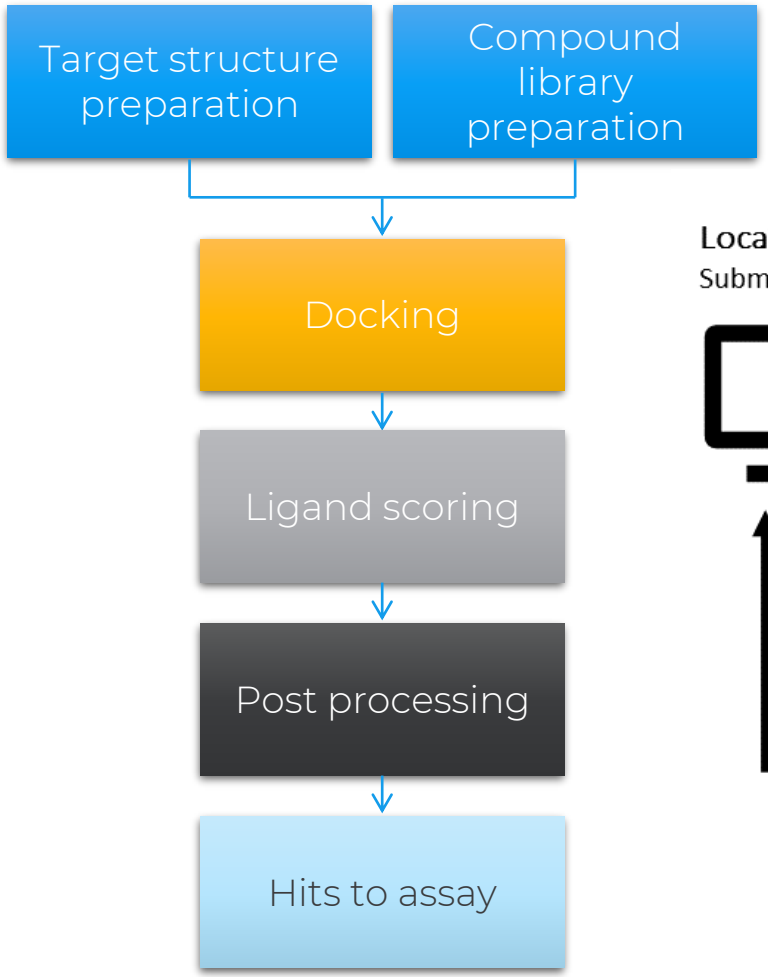
Perform protein-ligand docking in the KNIME interface to easily build into pipelines of work.

Covalent docking

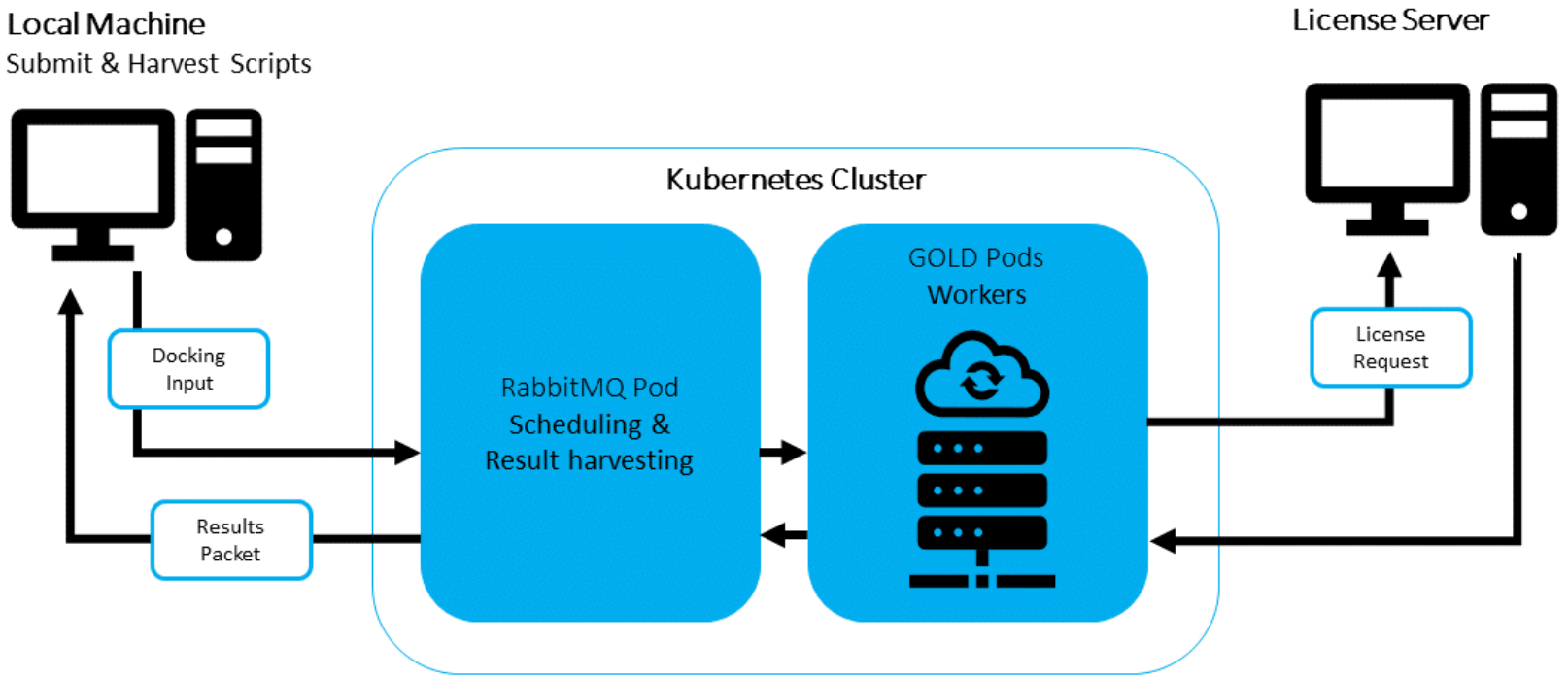
Understand irreversible binding with covalent docking to explore cancer, immunology and infectious disease targets.

CCDC

Docking virtual screening (VS) with GOLD



Typical workflow of a docking-based VS

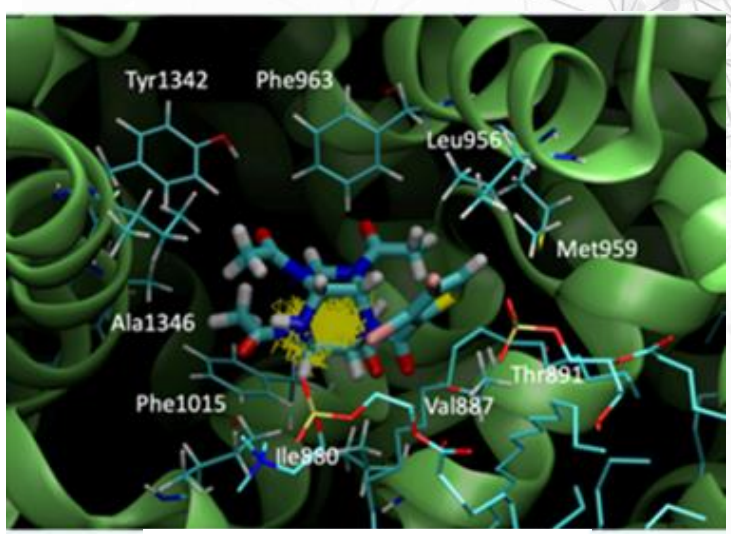


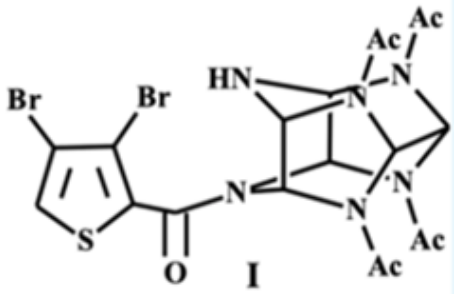
Gold High Performance Computing

CCDC

Discovery of a novel non-narcotic analgesic

- Thiowurtzine is identified as a new potent analgesic molecule for chronic pain treatment
- It is derived from CL-20 explosive
- Mechanism of action is evaluated with docking, molecular modeling, and molecular dynamics
- Thiowurtzine may target the mu opioid receptor, the TRPA1 ion channel, and the Ca_V voltage-gated calcium channel.





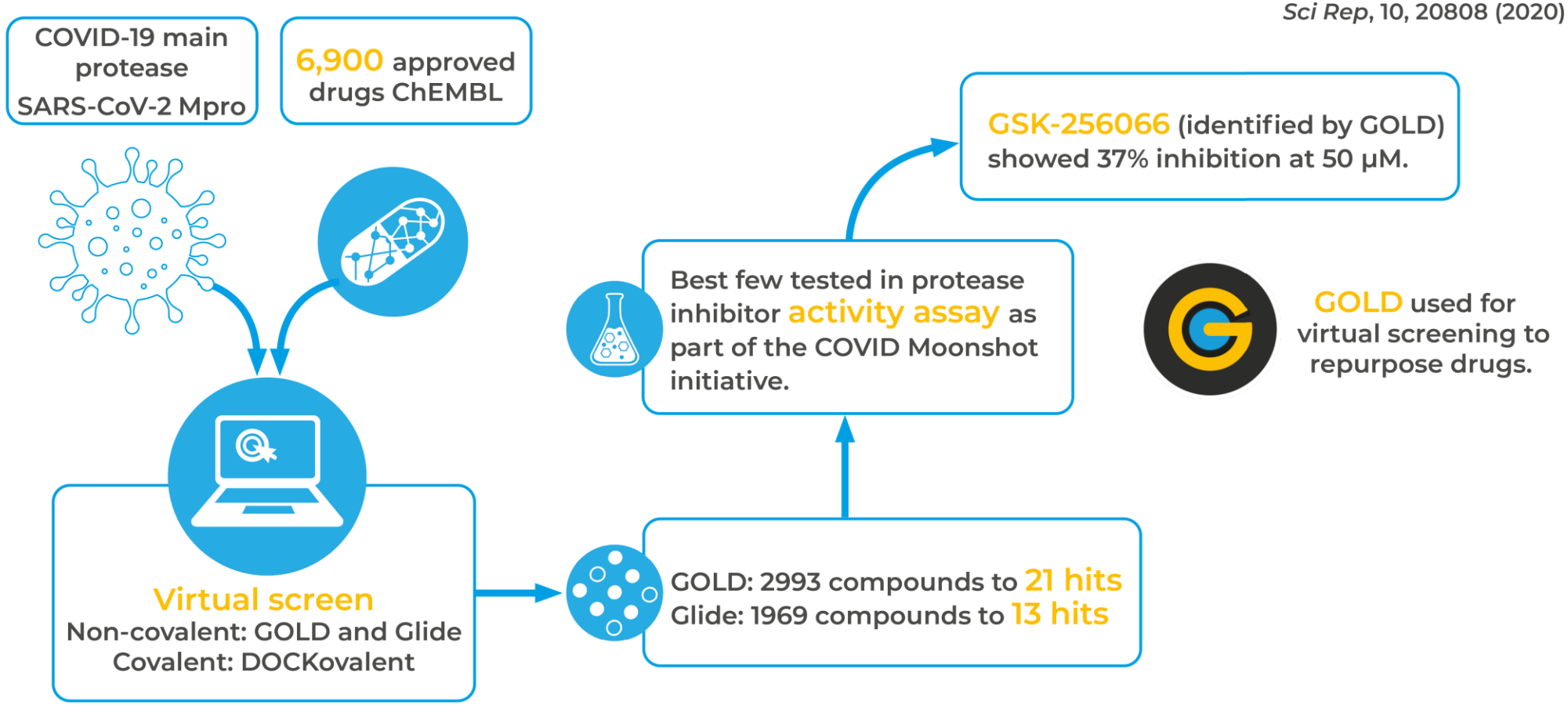
Thiowurtzine

Aguero S. *et al.*, 2021, 6, 23, 15400-15411 DOI:10.1021/acsomega.1c01786

CCDC

Virtual screening to repurpose drugs for COVID-19

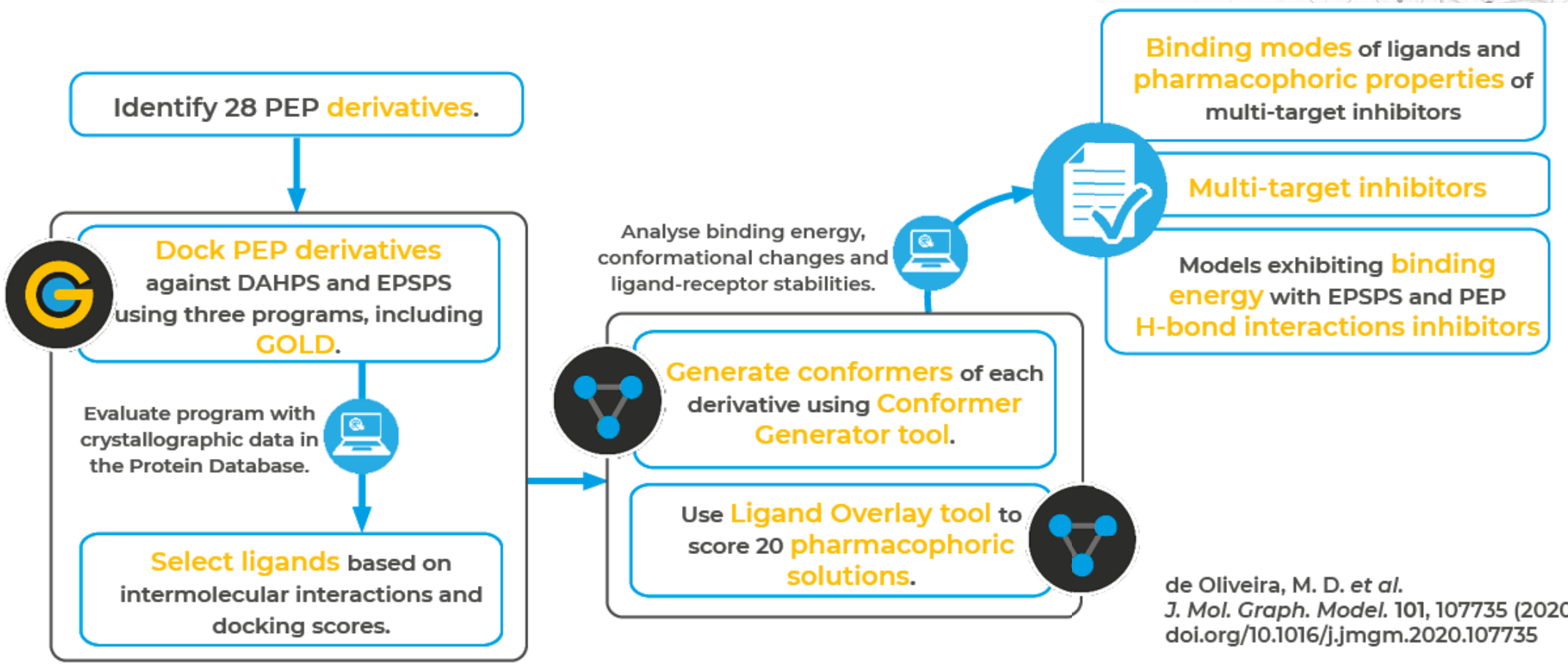
Alina Shitrit *et al.*
Sci Rep, 10, 20808 (2020)



https://www.ccdc.cam.ac.uk/Community/blog/gold-virtual-screen-covid-drug-repurposing/

CCDC

Ligand-based drug design of competitive inhibitors against DAHPS and EPSPS



de Oliveira, M. D. *et al.*
J. Mol. Graph. Model. 101, 107735 (2020).
doi.org/10.1016/j.jmgm.2020.107735

https://www.ccdc.cam.ac.uk/Community/blog/ligand-based-drug-design-of-competitive-inhibitors/

CCDC

CSD-Discovery overview

GOLD: Protein-ligand docking and virtual screening

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.**

SuperStar: Analyse, predict and understand protein and ligand interactions

CCDC

https://www.ccdc.cam.ac.uk/Community/educationalresources/

Want to explore more?

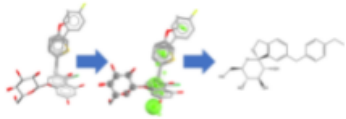
Training and Educational Resources

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

The CCDC and our colleagues continually produce educational materials for use in classroom and computer lab settings, or as independent study modules. Many of these materials make use of the Teaching Subset - a freely available set of over 750 structures that can be investigated with the free version of our Mercury visualisation and analysis program. Of database of over one million entries are available for free through our Access Structures portal.

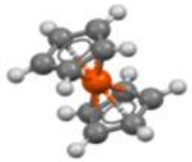
n educator looking for supplementary teaching materials, find out more about the Teaching Database **here**. If you have developed your own modules using the CSD and o share them with the broader community, please contact us at education@ccdc.cam.ac.uk.

to date with the latest news from education and outreach at the CCDC, sign up for the Education and Outreach Newsletter **here**.

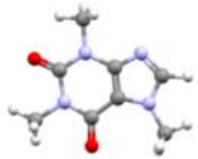


CSD-Discovery

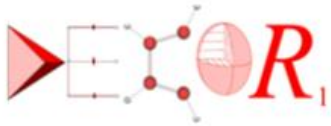
For pharmaceutical and agrochemical researchers, tools for discovering new molecules and performing protein docking studies.



formation on the Teaching Subset



Access a series of teaching modules for use in the classroom



DECOR: Educational Resources for Teaching Crystallography

Register for E&O newsletter

Self-guided workshops

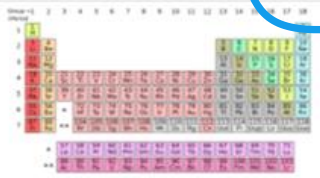
Download a series of self-guided workshop materials for CCDC tools and features

Watch software training and support videos

CSDU modules

On-demand modules with completion certificate

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



Explore the Periodic Table through Crystal Structures

YouTube and LabTube channels

CCDC