The Cambridge Structural Database (CSD): An invaluable resource for drug discovery

Dr. Paul Sanschagrin (Dr. ポール・サンチャグレン)

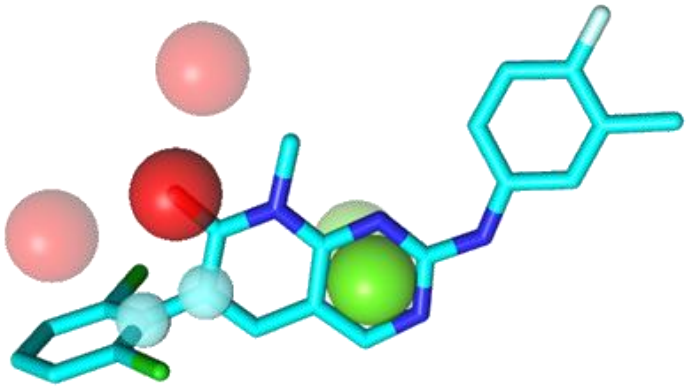
CCDC User Application Scientist, Princeton, NJ, USA

Dr. Francesca Stanzione (Dr.フランチェスカ・スタンツィオーネ)

CCDC User Application Scientist, Cambridge, UK

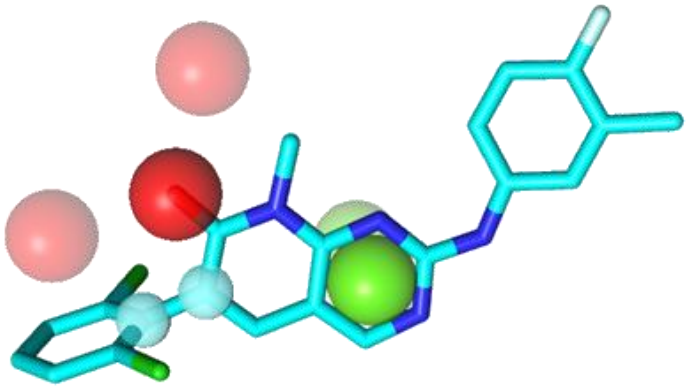
The 18th Annual Meeting of the Protein Science Society of Japan

28-Jun-2018, Niigata



Talk Outline

- **Introduction to the CCDC and the CSD**
- Introduction to CSD-CrossMiner software
- CSD-CrossMiner application example
- CSD-CrossMiner software demo



The Cambridge Crystallographic Data Centre (CCDC)

International Data Repository
Archive of crystal structure data
High quality scientific database

Scientific Software Provider
Search/analysis/visualisation tools
Scientific applications

Collaborative Research Organisation
New methodologies
Fundamental research

Education and Outreach
Workshops, Training
Teaching Materials

Employer of ~60 staff

Scientific editors
Software developers
Applications scientists

Cambridge UK

Princeton NJ USA

Originated in 1965

Financially self-supporting, not-for-profit

UK Registered Charity

University of Cambridge Partner Institute

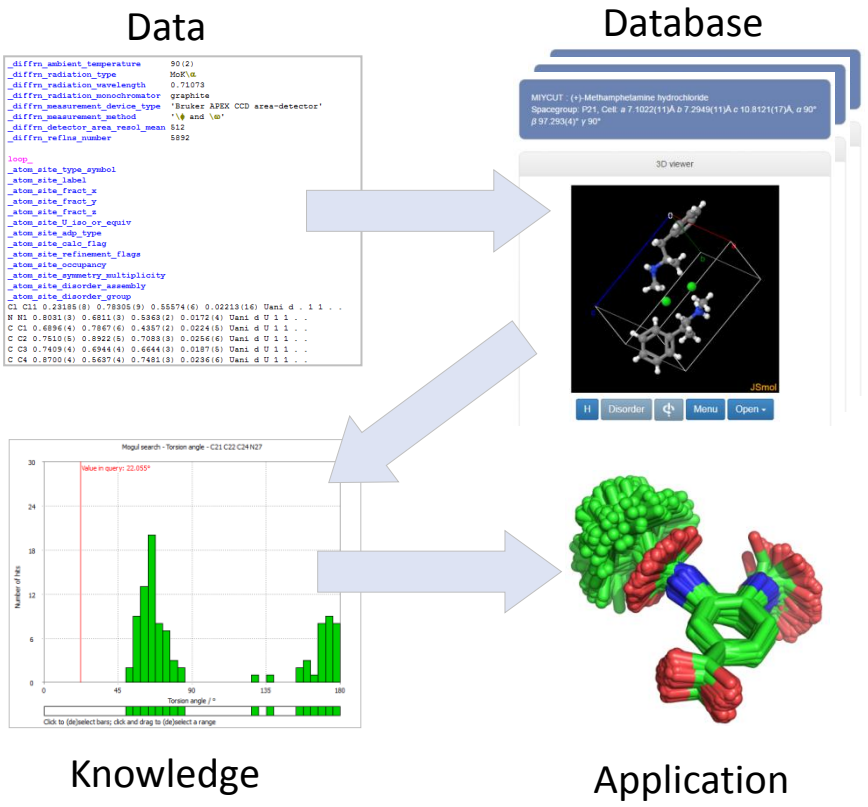
Unlocking Knowledge

International Data Repository
Archive of crystal structure data
High quality scientific database

Scientific Software Provider
Search/analysis/visualisation tools
Scientific applications

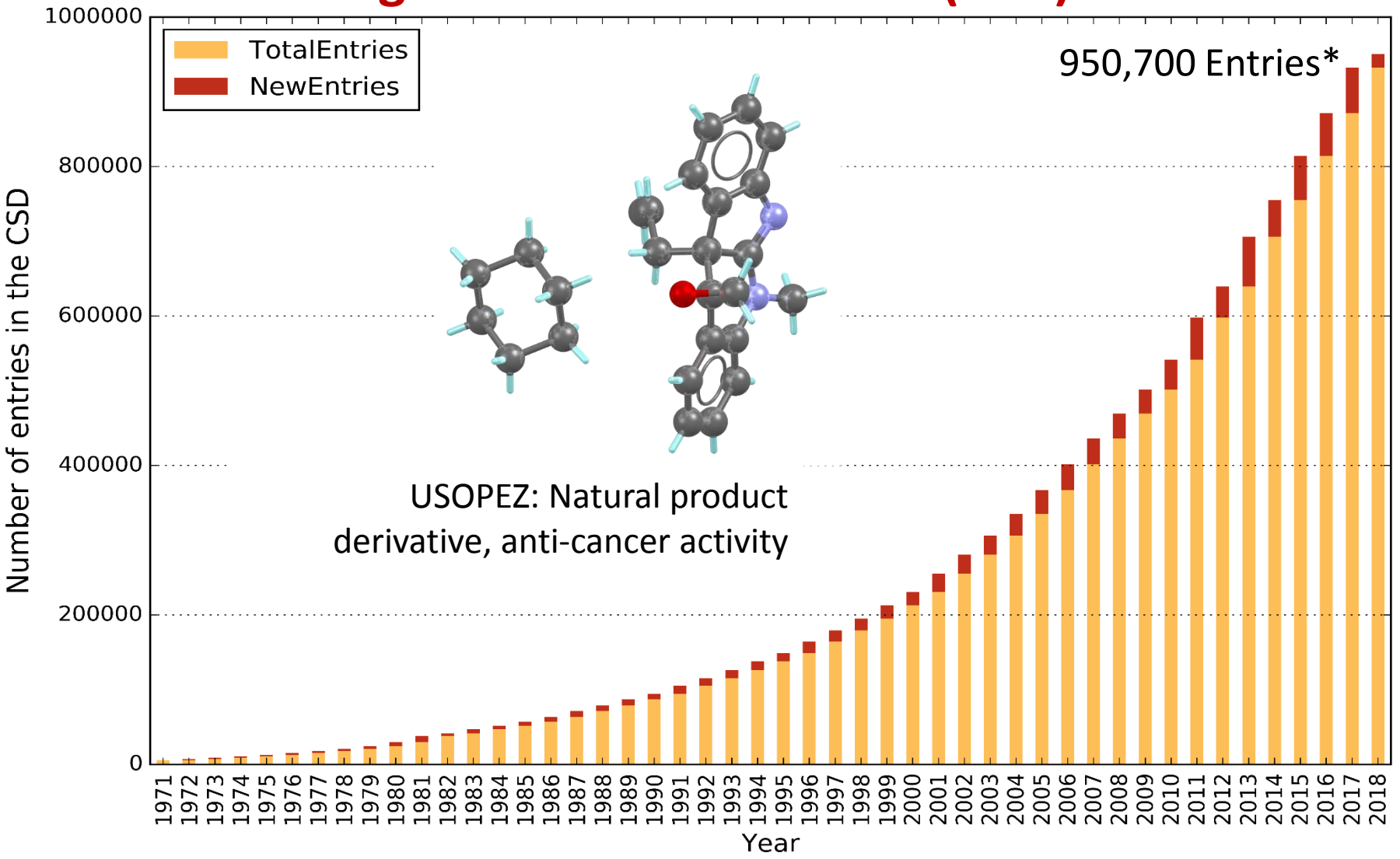
Collaborative Research Organisation
New methodologies
Fundamental research

Education and Outreach
Workshops, Training
Teaching Materials



Chemistry plus crystallography enables discovery and application of new knowledge

The Cambridge Structural Database (CSD)



*As of 17-May-2018, adding ~60,000/year

Groom, C. R.;. Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater. 2016, 72 (2), 171–179. DOI: 10.1107/S2052520616003954

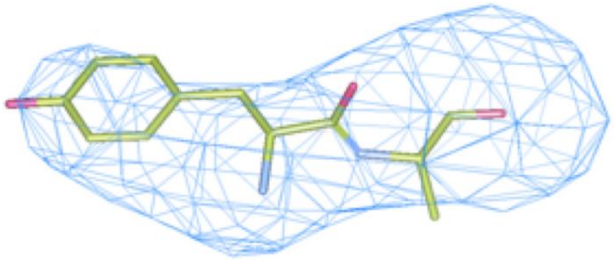
Top 200 small molecule drugs of 2015

Green = Exact Matches in the CSD (118 drugs)

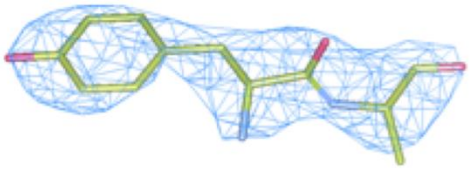
Grey = No exact match in CSD, it is likely something very similar is in the CSD

Crystallographic Resolution

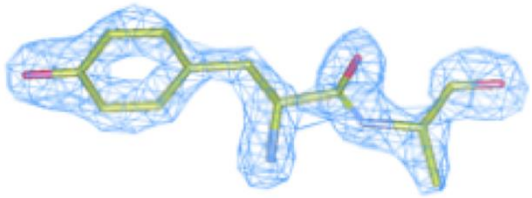
Crambin
PDB: 3nri
Experimental resolution: 0.48Å



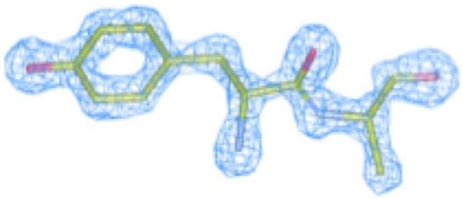
5.0 Å



3.0 Å

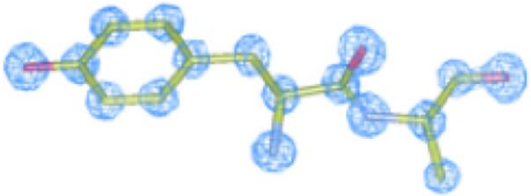


2.0 Å

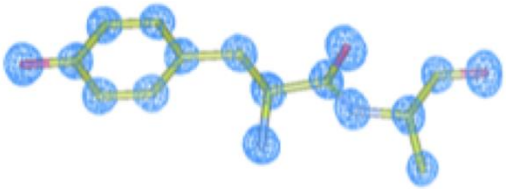


1.5 Å

Typical PDB Structures



1.0 Å



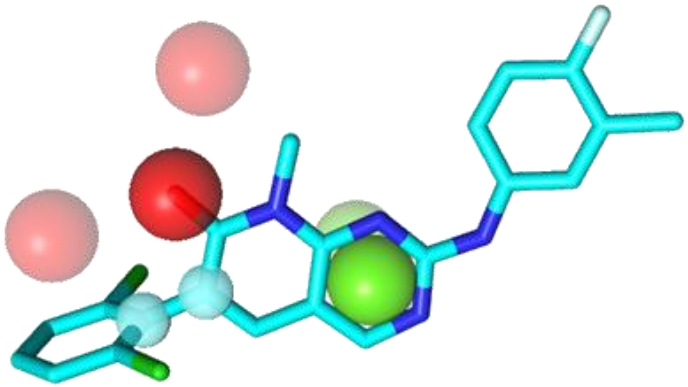
0.6 Å

Typical CSD Structures

Talk Outline

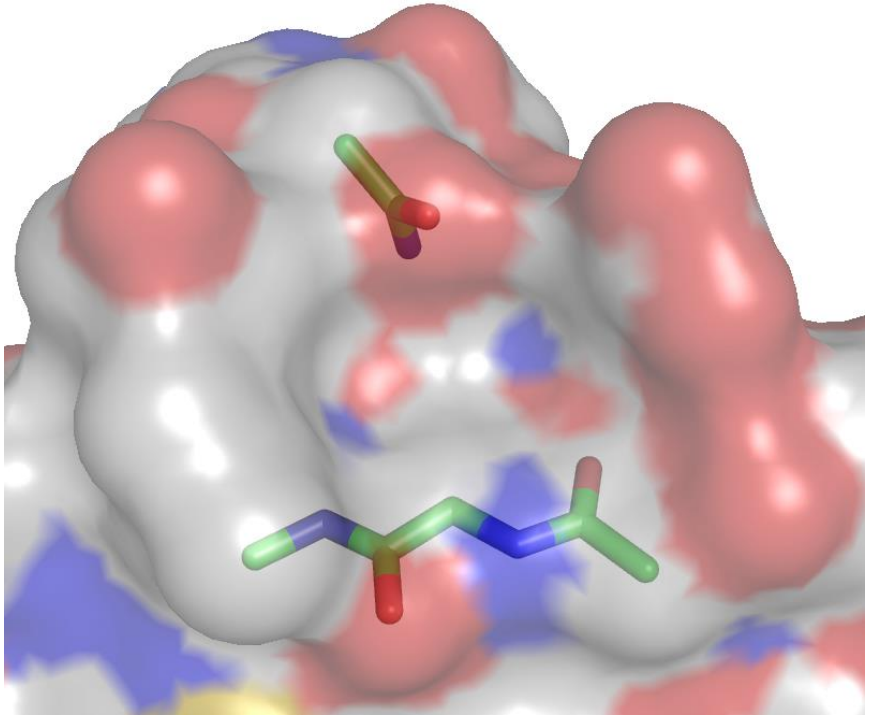
- Introduction to the CCDC and the CSD
- **Introduction to CSD-CrossMiner software**
- CSD-CrossMiner application example
- CSD-CrossMiner software demo

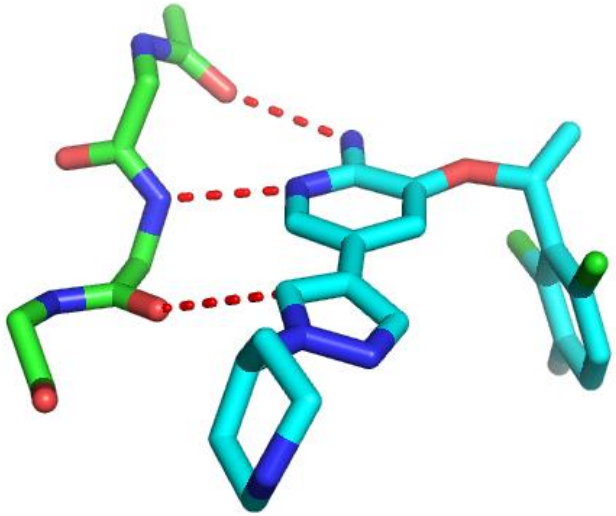
NEW!



Typical Questions Arising in Drug Design Projects

- Which structural motifs bind in a similar environment?





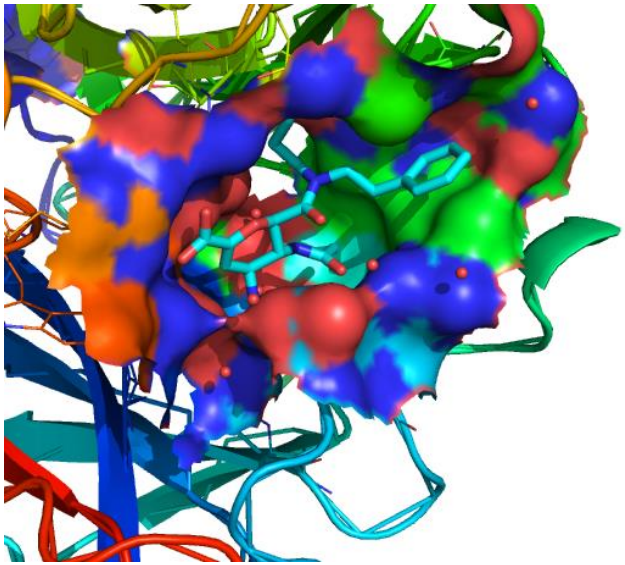
- Which ligand motifs have similar protein interaction pattern?
- Are there molecules with similar interacting groups in the database?

Grand Challenge: Enable a More Effective Usage of Structural Data in Drug Design Projects

protein-ligand binding sites

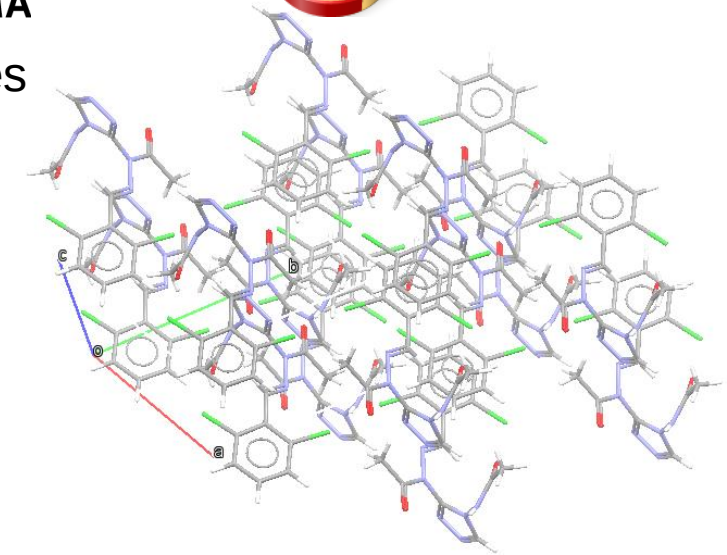
organic small molecules

PDBj
Protein Data Bank Japan



PS PHARMA

In-house databases



[1]Berman, *et al.*, *Nucl. Acids Res.*, **2000**, *28*, 235-242. DOI: 10.1093/nar/28.1.235
[2]Groom, *et al.*, *Acta Cryst. B*, **2016**, *72*, 171-179. DOI: 10.1107/S2052520616003954

CSD-CrossMiner: Pharmacophore Search Tool

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Interactive and Versatile Navigation of Structural Databases

Oliver Korb,[*,†,∥] Bernd Kuhn,[*,‡,∥] Jérôme Hert,[‡] Neil Taylor,[§] Jason Cole,[†] Colin Groom,[†] and Martin Stahl[‡]

[†]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, U.K.

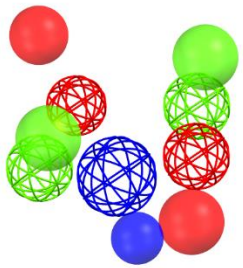
[‡]Roche Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, CH-4070 Basel, Switzerland

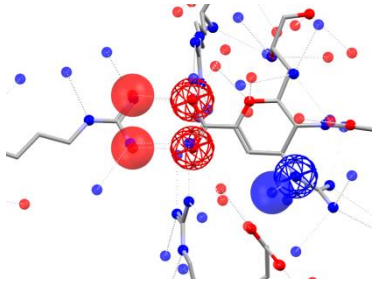
[§]Desert Scientific Software Pty Ltd., Level 5 Nexus Building, Norwest Business Park, 4 Columbia Court, Baulkham Hills, NSW 2153, Australia

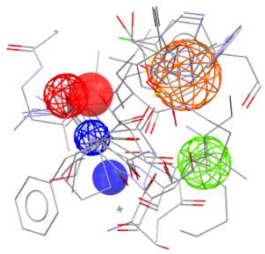
- Tool for pharmacophore-based searches in the PDB and CSD
  - From IUPAC: “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions”
  - Wikipedia: “abstract description of molecular features”
- Modify a hypothesis/results on the fly

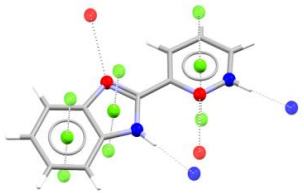
Why use Pharmacophores?

donor
acceptor
planar ring





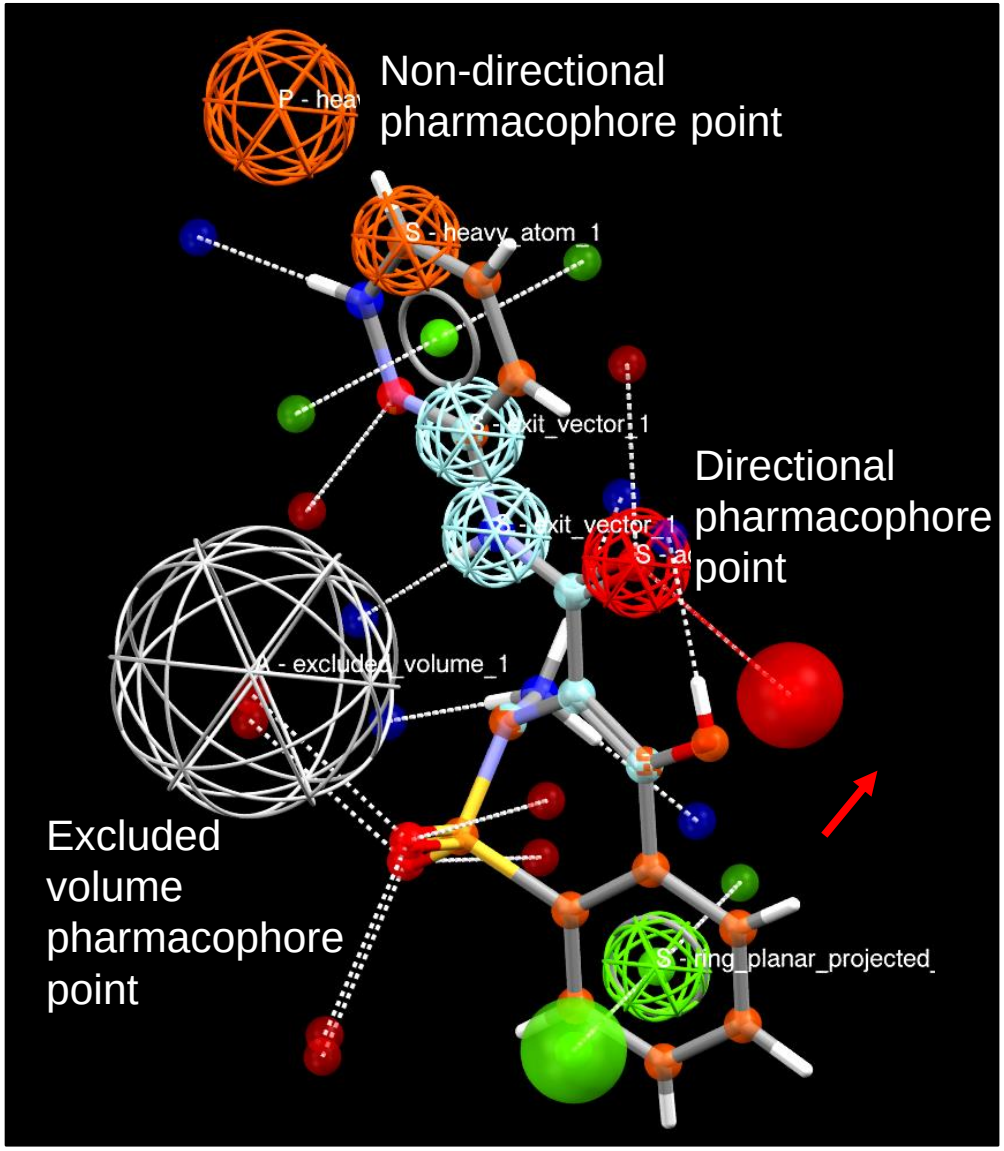




- Intuitive as the query can be specified relative to a 3D structure
  a) Can easily be refined
- *"what you see is what you get"*
  a) Hits will be overlaid onto the query and can be analysed in the same context
- Concept well-understood by medicinal chemists
- Flexible feature definitions
  a) User can control the level of chemical abstraction

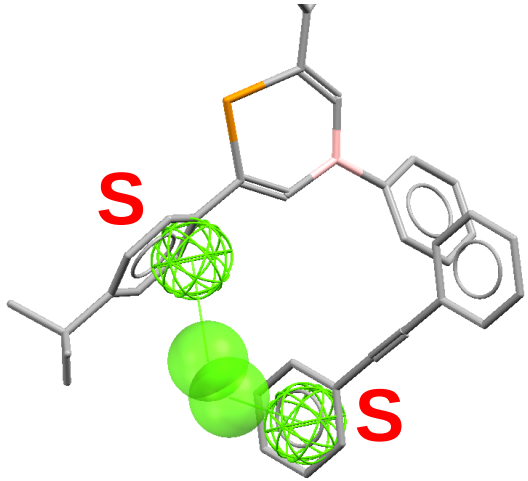
From Features to Pharmacophores

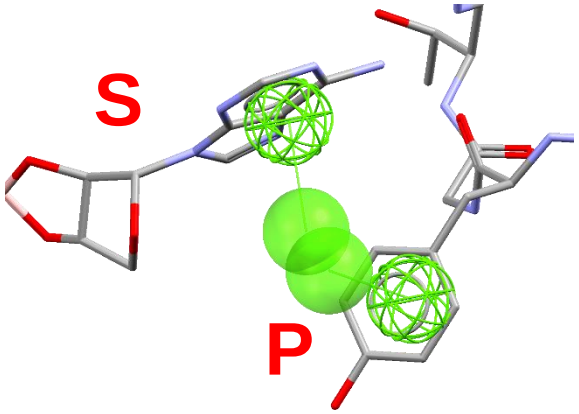
- Molecular structures are annotated with *features*
  - Based on SMARTS patterns
  - Stored in feature database
- Pharmacophore query is based on *tolerance spheres*
  - Sphere radii reflect uncertainty in the position of the features
- Pharmacophore points can be
  - Single point
  - Directional (two points)

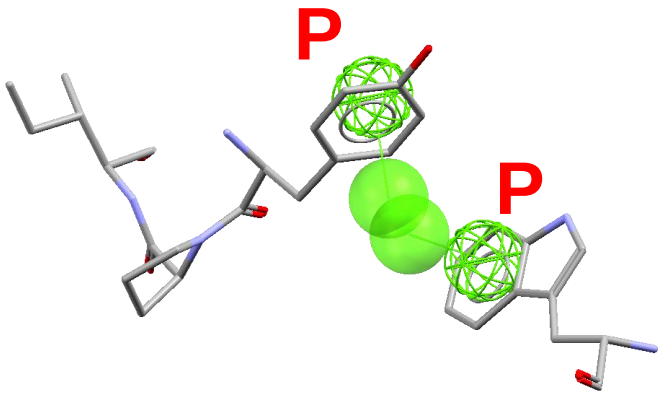


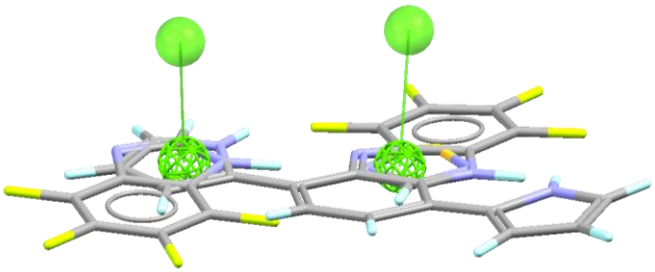
Constraints

Protein (P)
Small molecule (S)

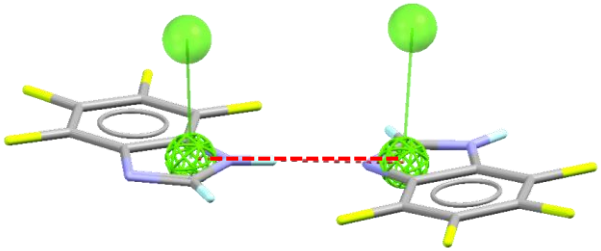




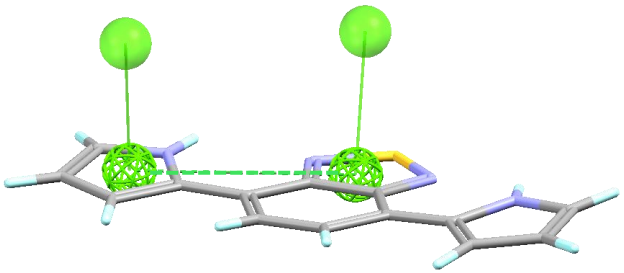




No constraint

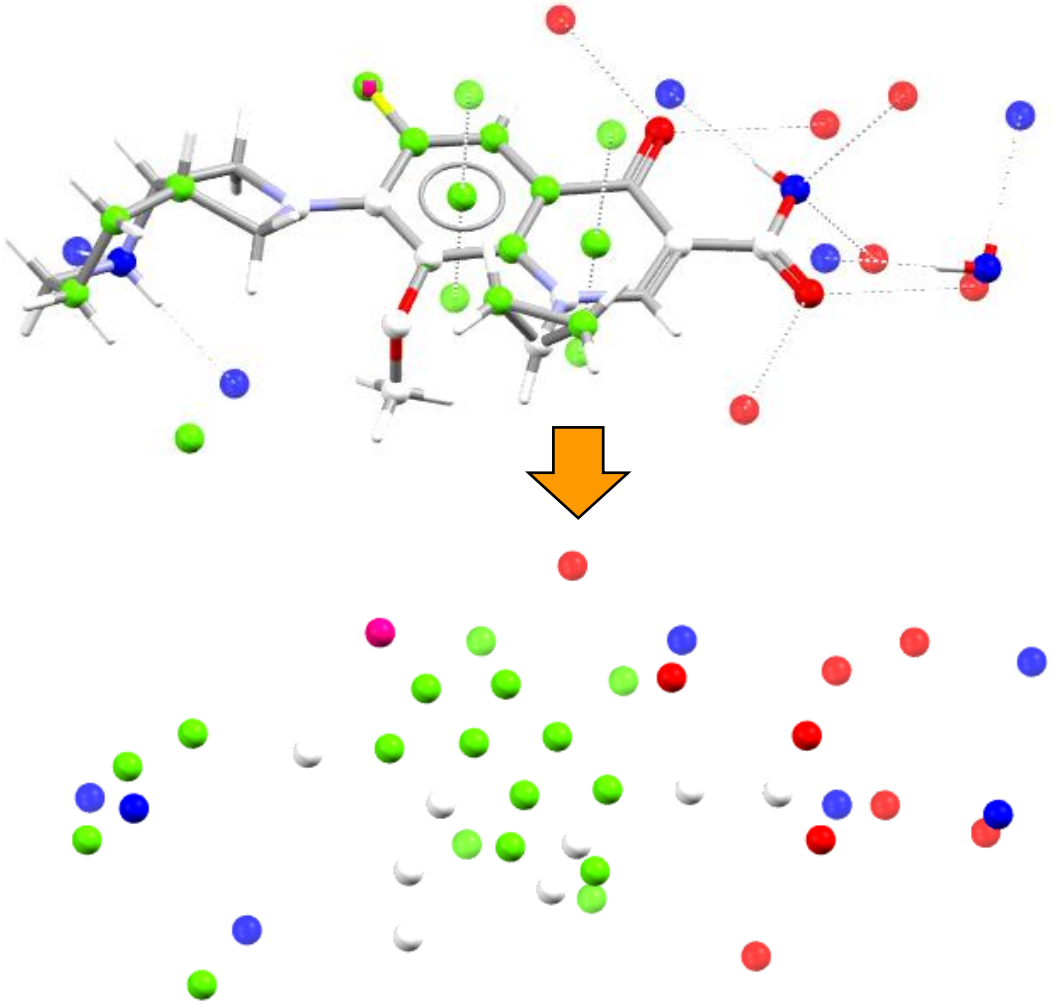


Inter-molecular constraint



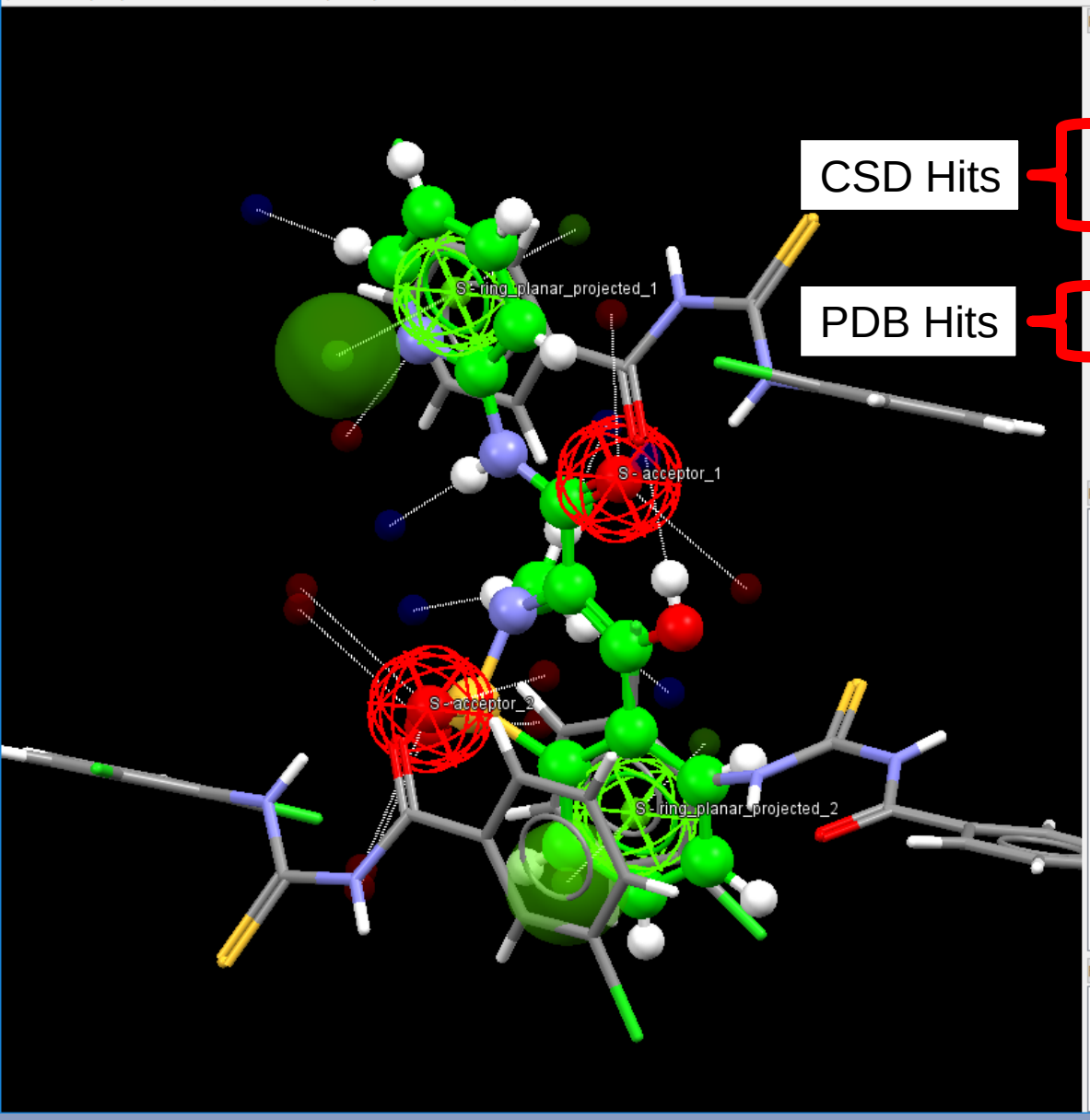
Intra-molecular constraint

Feature Abstraction



the actual search algorithm doesn’t know about molecules

Software Interface



BIYSEH - CSD-CrossMiner

File Edit Display Feature Database Help

Style: Ball and Stick Colour: by Element Picking Mode:

Show: reference hits constraints features pharmacophore pharm. labels hydrogens Search: Edit: intra

x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90 zoom- zoom+

CSD Hits

PDB Hits

Results Hitlist

1st in cluster Settings... Tanimoto: 0.70 Number of hits: 370 Show all

mark	identifier	cluster	rmsd	chain	deposition_date	ec_number
	MEHLUH_1	36	0.802			
	MEHMER	37	0.929			
	MEHMIW_1	38	0.713			
	MEHMOC_1	39	0.766			
	MEHNAP	40	0.981			
	1EO3_m1_A-B_...	41	0.812	B	2000-03-21	3.1.21.4
	MEHNIY_1	42	0.759			
	5OE3_m1_D_bs_...	43	0.971	D	2017-07-07	6.2.1.32
	5P91_m1_A_bs_...	44	0.975	A	2016-08-29	2.1.1.6
	5P9H_m1_A_bs...	45	0.865	A	2016-09-20	2.7.10.2
	5PAC_m1_B_bs...	46	0.837	B	2016-11-10	3.4.21.21
	MEHPAR_1	47	0.79			

#hits: 795/10000

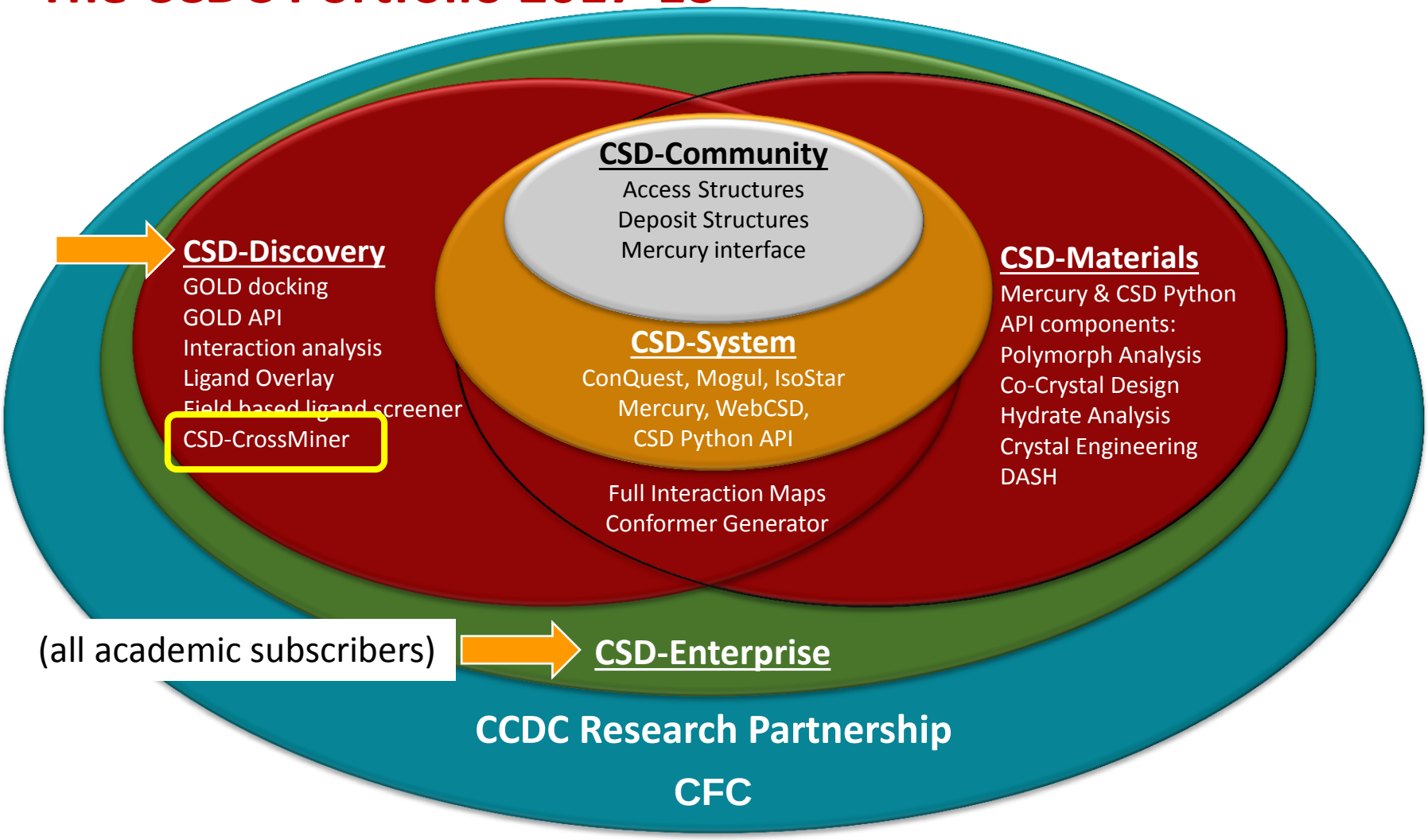
Pharmacophore Features

feature name	tolerance radius	show in reference	show in pharmacophore
acceptor_1			☑
B	1.00		
acceptor_2			☑
B	1.00		
acceptor_projected		☑	
donor_ch_projected		☑	
donor_projected		☑	
heavy_atom		☑	
hydrophobe		☐	
ring		☐	
ring_non_planar		☐	
ring_planar_projected		☑	
ring_planar_project...			☑
B	1.00		
V	1.00		

Feature Databases

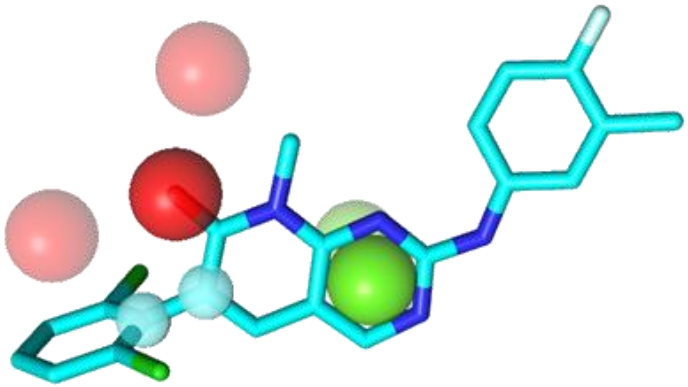
database	size
pdb_crossminer	205194
csd539_crossminer	285437

The CCDC Portfolio 2017-18

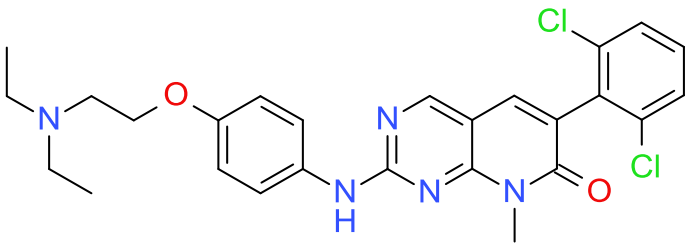


Talk Outline

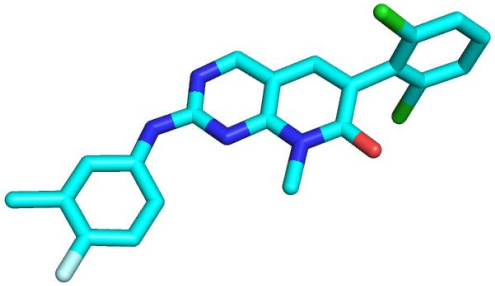
- Introduction to the CCDC and the CSD
- Introduction to CSD-CrossMiner software
- **CSD-CrossMiner application example**
- CSD-CrossMiner software demo



Kinase Inhibitor Scaffold Hopping based on Ligand Features

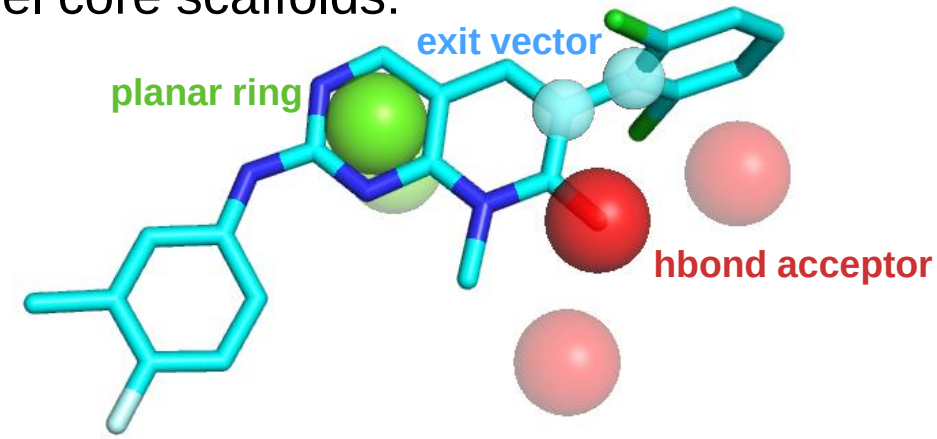


PD166285 is known tyrosine kinase inhibitor



PDB:2hzi
(Abl kinase)

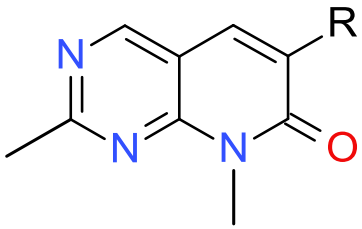
Search for novel core scaffolds:

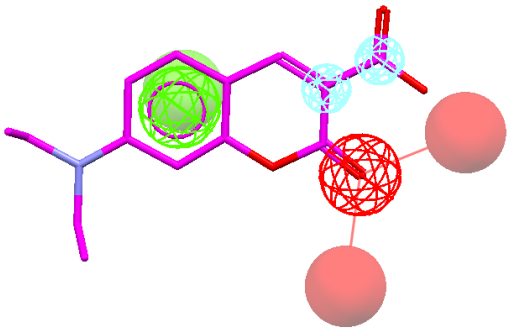


Furet et al., *Bioorg. Med. Chem. Lett.* 2008, *18*, 897 DOI: 10.1016/j.bmcl.2007.12.041 (Novartis)

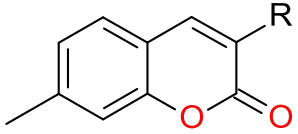
Top-ranking Hit Structures in CSD

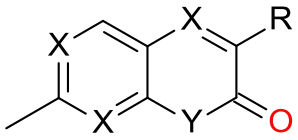
Template:



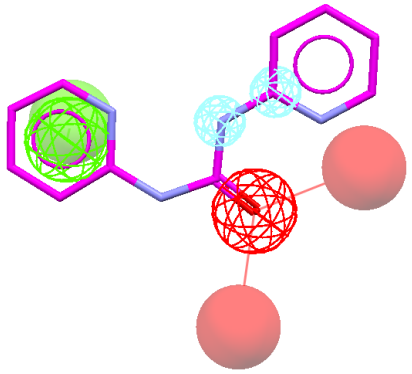


IDUQOO_1

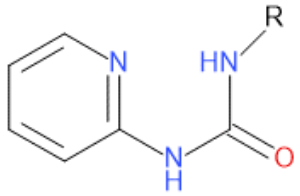


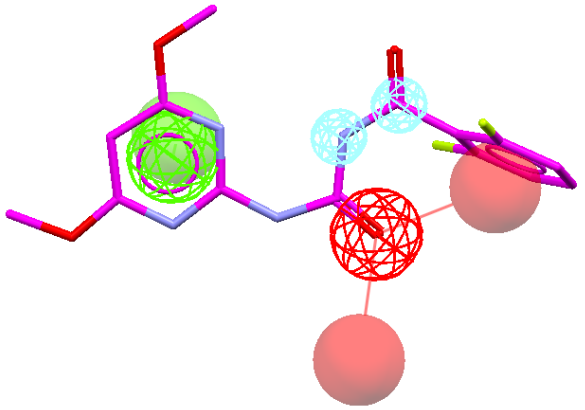


X: C,N
Y: O,N

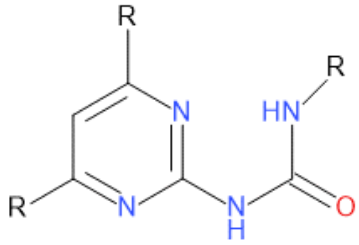


ALOJIW_1



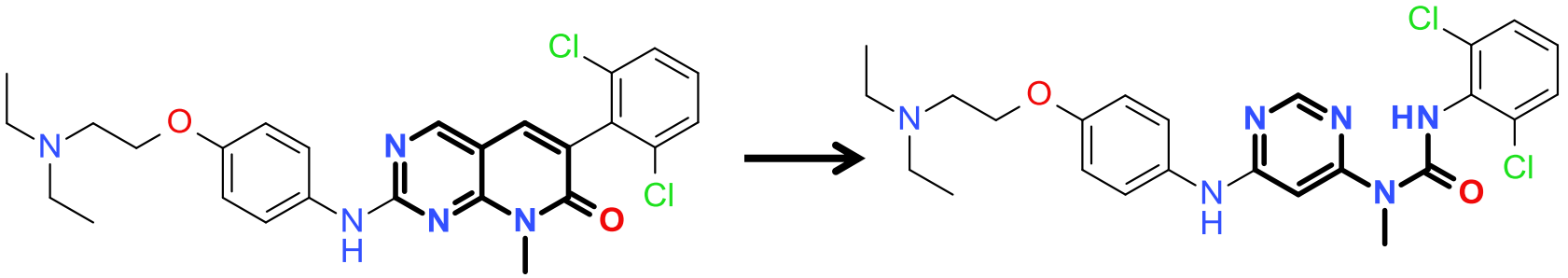


IFAFEB_1



Furet et al., *Bioorg. Med. Chem. Lett.* 2008, *18*, 897 DOI: 10.1016/j.bmcl.2007.12.041 (Novartis)

Scaffold Hopping Example from Novartis



These pseudo rings are considered to be particularly stable and represent a new and non-conventional strategy to mimic established lead compounds.

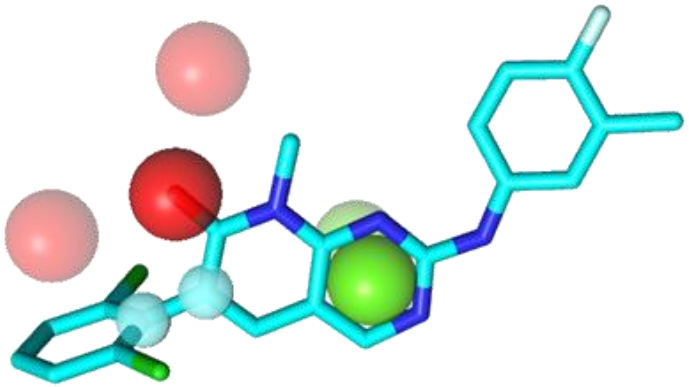
IC_{50} < 1 uM against different kinases

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

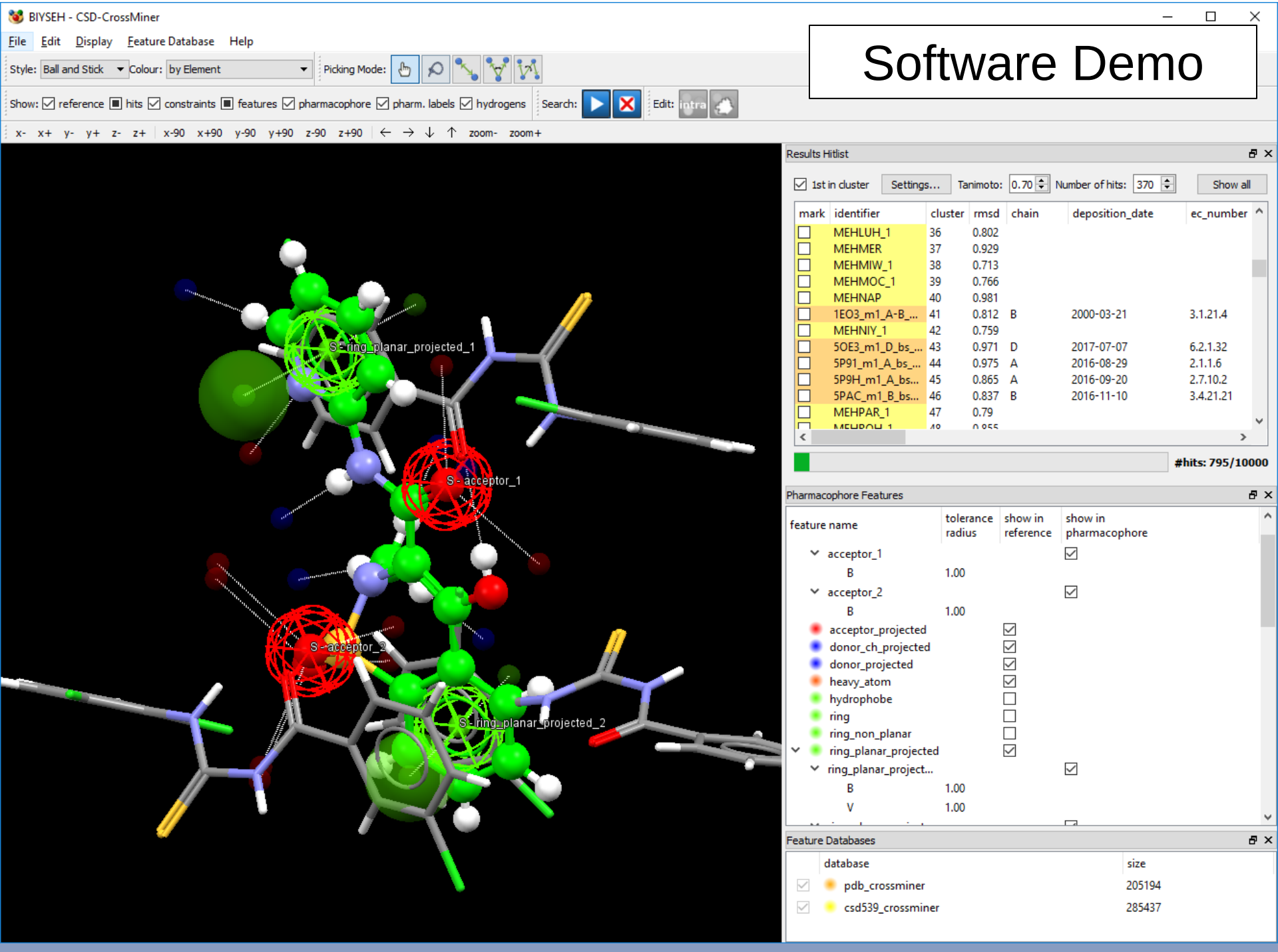
Furet et al., *Bioorg. Med. Chem. Lett.* 2008, *18*, 897 DOI: 10.1016/j.bmcl.2007.12.041 (Novartis)

Talk Outline

- Introduction to the CCDC and the CSD
- Introduction to CSD-CrossMiner software
- CSD-CrossMiner application example
- **CSD-CrossMiner software demo**

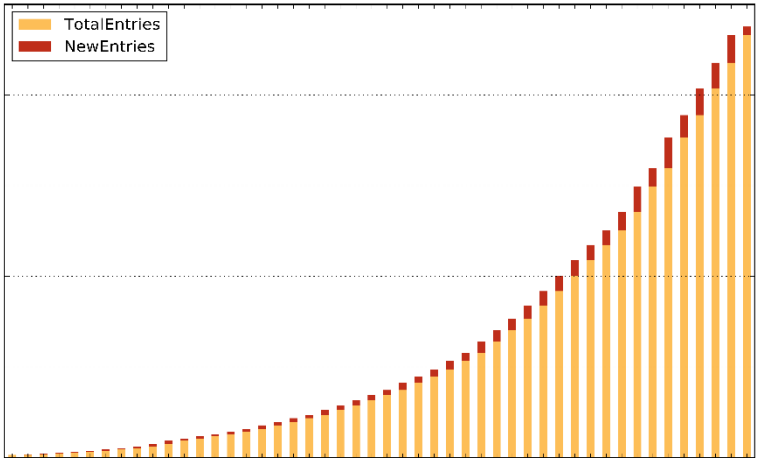


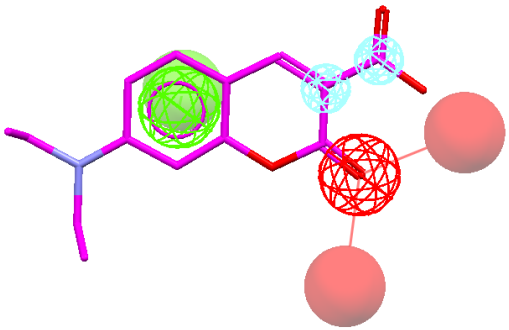
Software Demo

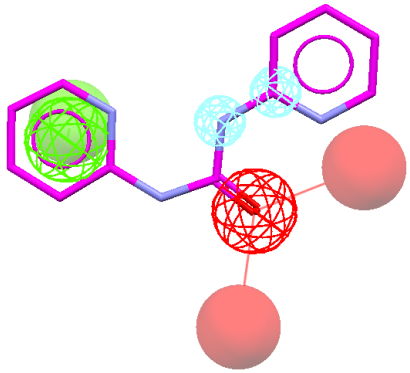


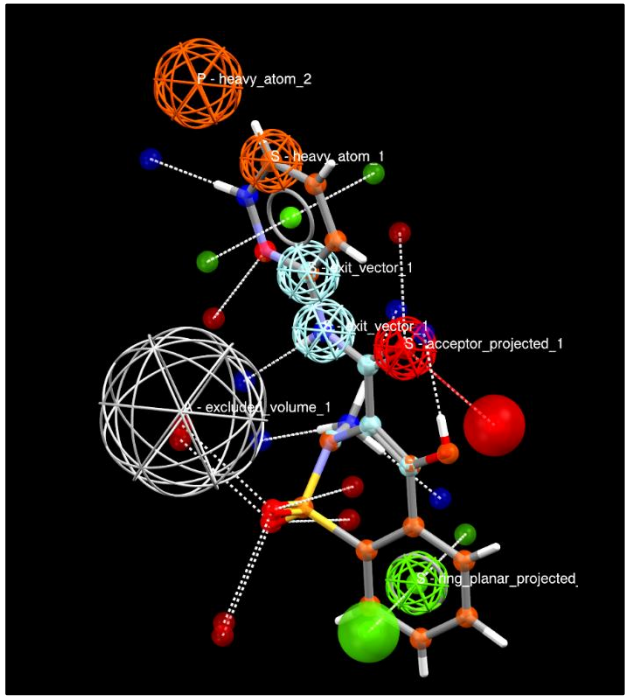
Summary

- Introduction to the CCDC and the CSD
- Introduction to CSD-CrossMiner software
- CSD-CrossMiner application example
- CSD-CrossMiner software demo









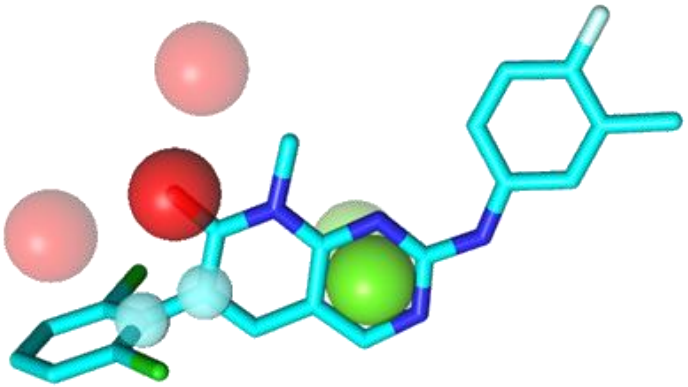
Thank you
どうもありがとうございました

support@ccdc.cam.ac.uk

企業向け：化学情報協会 <crystal@jaici.or.jp>

アカデミック向け：大阪大学蛋白質研究所

下家 <kshimoke@protein.osaka-u.ac.jp>



【CSDパッケージソフトの構成】

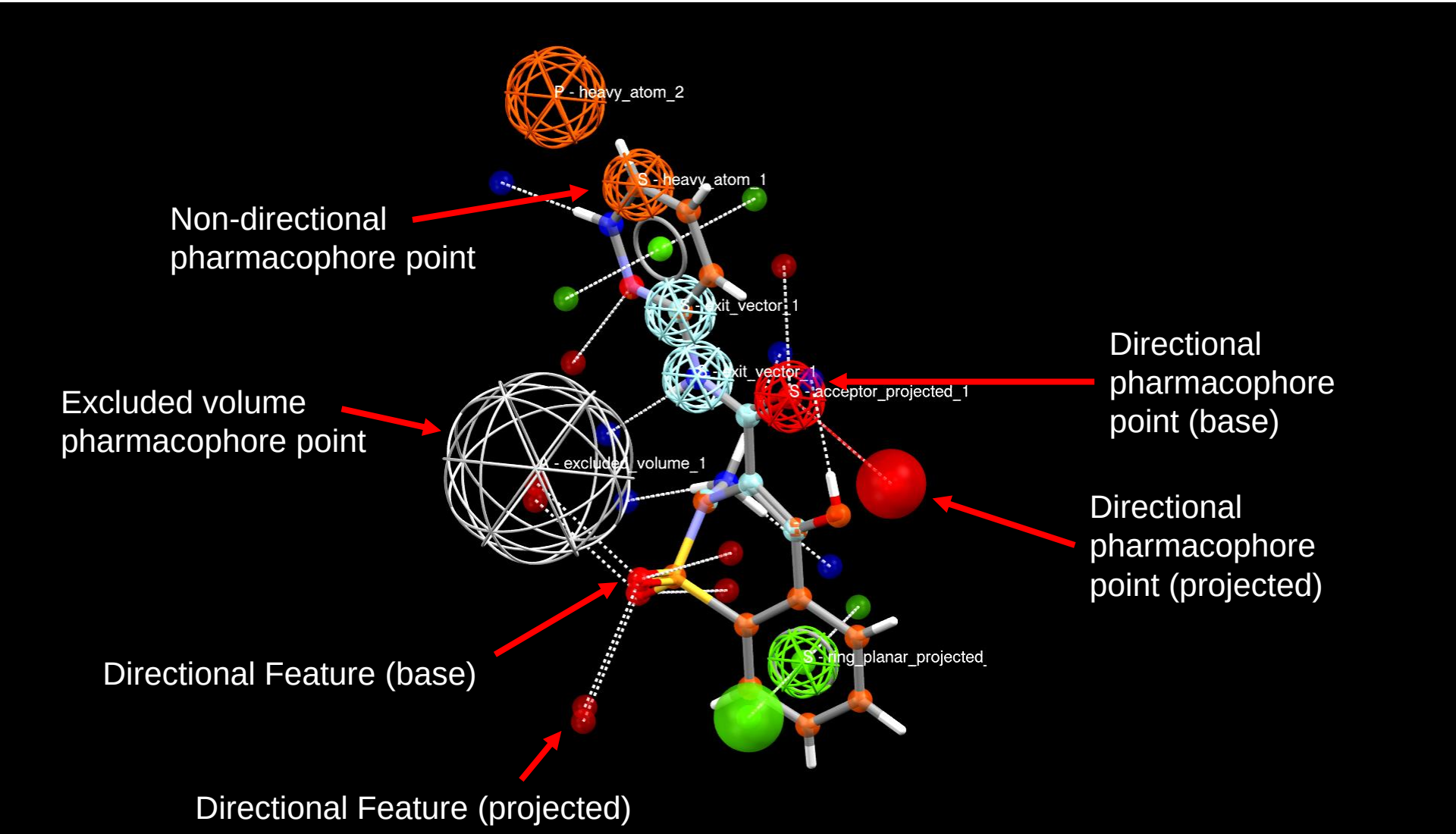
◆The Cambridge Structural Database (CSD): 90万件を超える有機分子の結晶構造を集めたデータベース.
◆CSD-System: CSDの基本セット．データ，検索，解析，表示ソフト付．分子間相互作用のライブラリー
分子ジオメトリーのsubライブラリーMogul付．(企業向け)
◆CSD-Discovery: CSD-System + 創薬向けツールが付属．(企業向け， CSDのアップグレード版)
◆CSD-Materials: CSD-System + 製剤系ツールが付属．(企業向け， CSDのアップグレード版)
◆CSD-Enterprise: CSD-System + すべてのツールが付属.
(アカデミック向けは，これ1本に統合．企業はCSDのアップグレード版として利用可)

	CSD-System	GOLD	Relibase+	CSD-CrossMiner	SuperStar	Conformer Generator/Full Interaction Maps	Materials module	DASH
CSD-System	●							
CSD-Discovery	●	●	●	●	●	●		
CSD-Materials	●					●	●	●
CSD-Enterprise	●	●	●	●	●	●	●	●

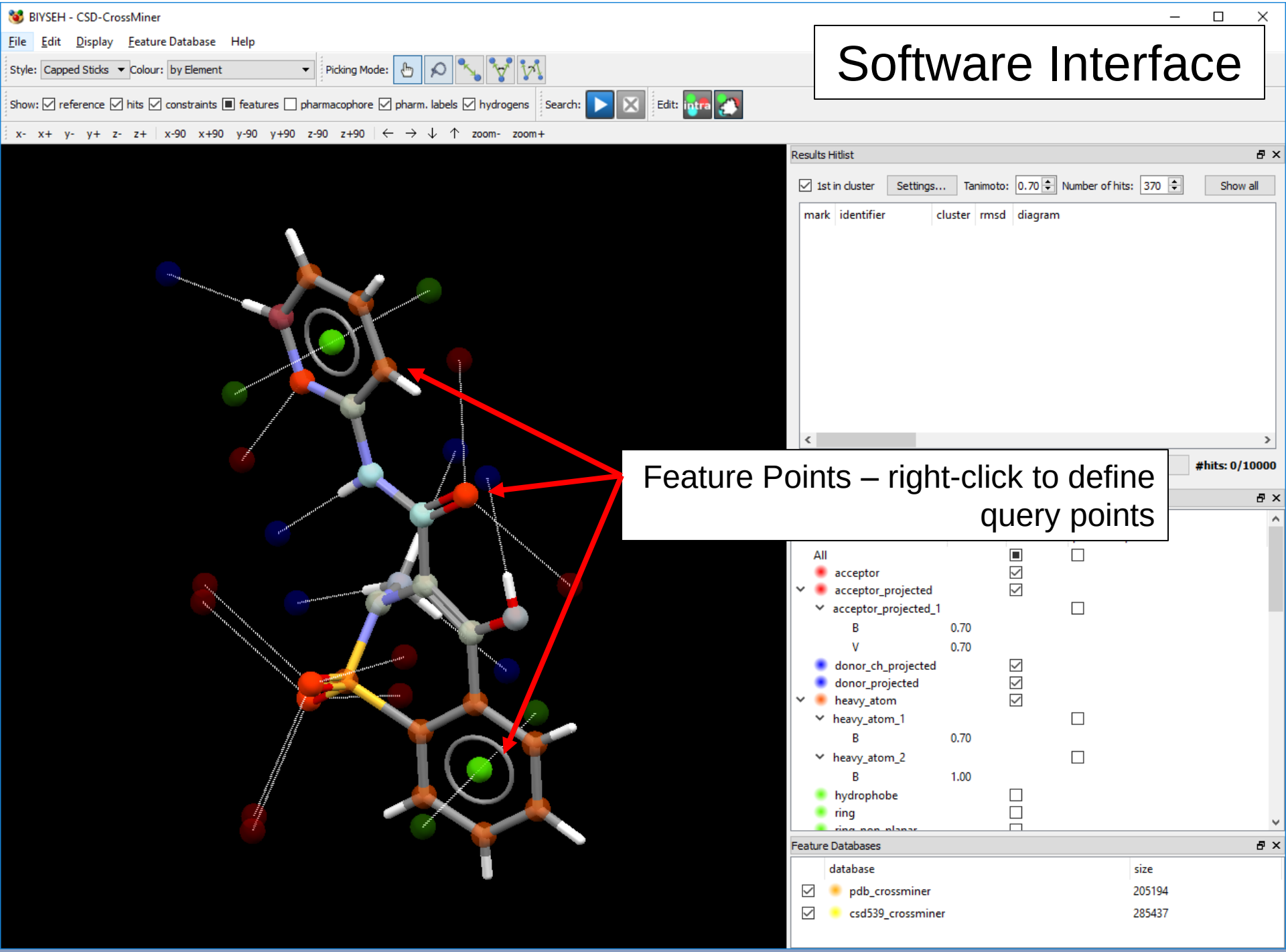
◇GOLD: 遺伝的アルゴリズムを使ったドッキングプログラム．多様なconstrainが容易に設定可能.
◇Relibase+: PDBの3Dを自在に検索可能．PDBの構造も定期的にupdate．(2018年11月頃をもって提供終了を予定)
◇CSD-CrossMiner: 今回リリースされた，CSDおよびPDBのデータをPharmacophoreを元に検索.
◇SuperStar: 注目している蛋白質の結合部位に対し，相互作用のhotspotをCSDのデータをもとに視覚化.
◇Conformer Generator: CSDの実験データをもとにconformerを自動発生．経験値からの配座発生のため，瞬時に発生可能.
◇Full Interaction Map:注目している分子に対して，周りの相互作用をCSDのデータをもとに視覚化.
◇Materials Module: 類似パッキングの検索，水素結合様式に基づく，多形結晶の可能性プロット，水和物の解析など，
固体結晶の解析に役立つツール.
◇DASH: 粉末結晶構造解析ソフト．粉末パターンから結晶構造を導きます.

詳細，お問い合わせは，https://www.ccdc.cam.ac.uk/solutions/
アカデミック：http://www.protein.osaka-u.ac.jp/csd/csd.html
企業：https://www.jaici.or.jp/wcas/csdpackage.htm

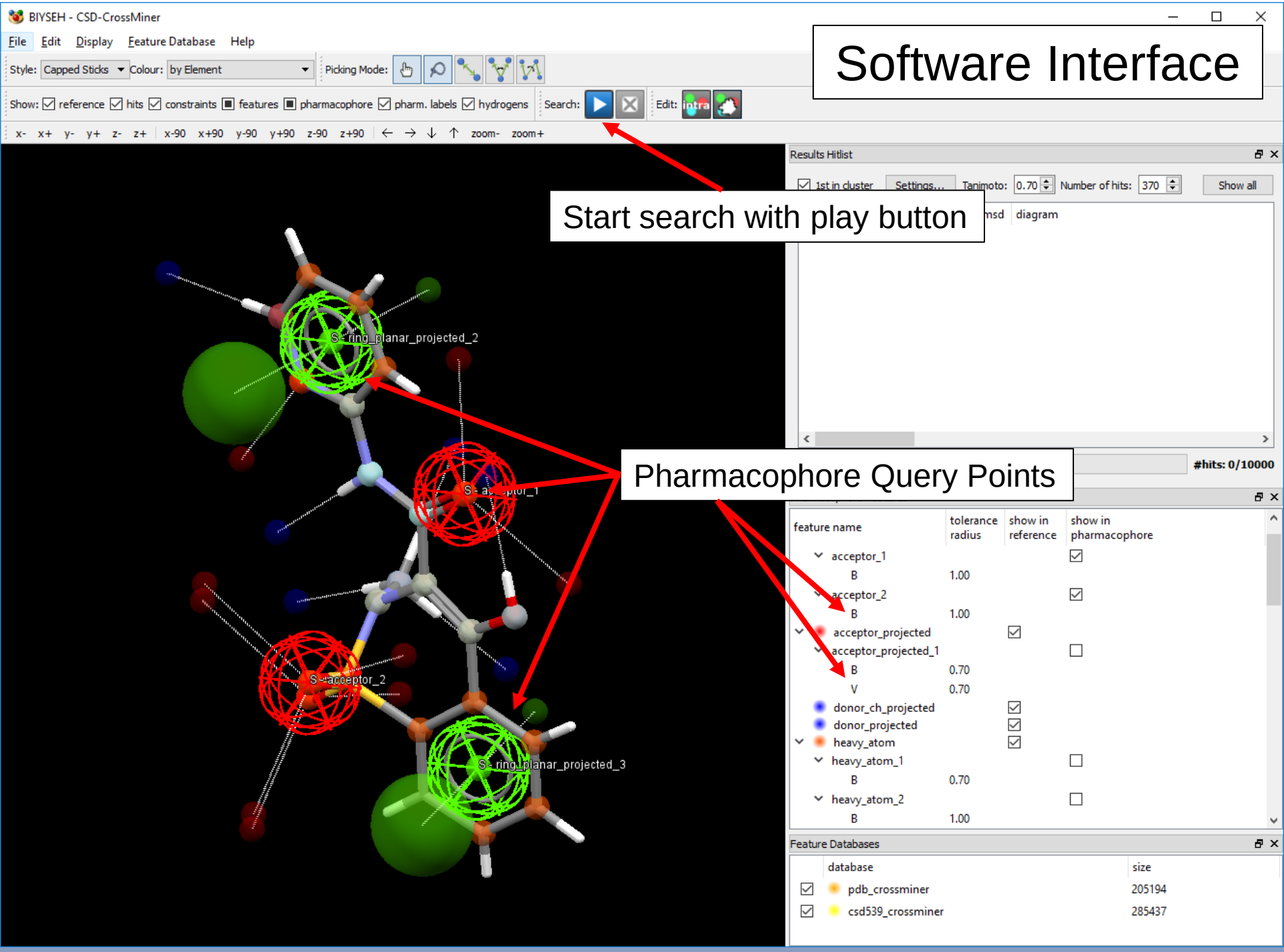
Feature and Pharmacophore Representation



Software Interface



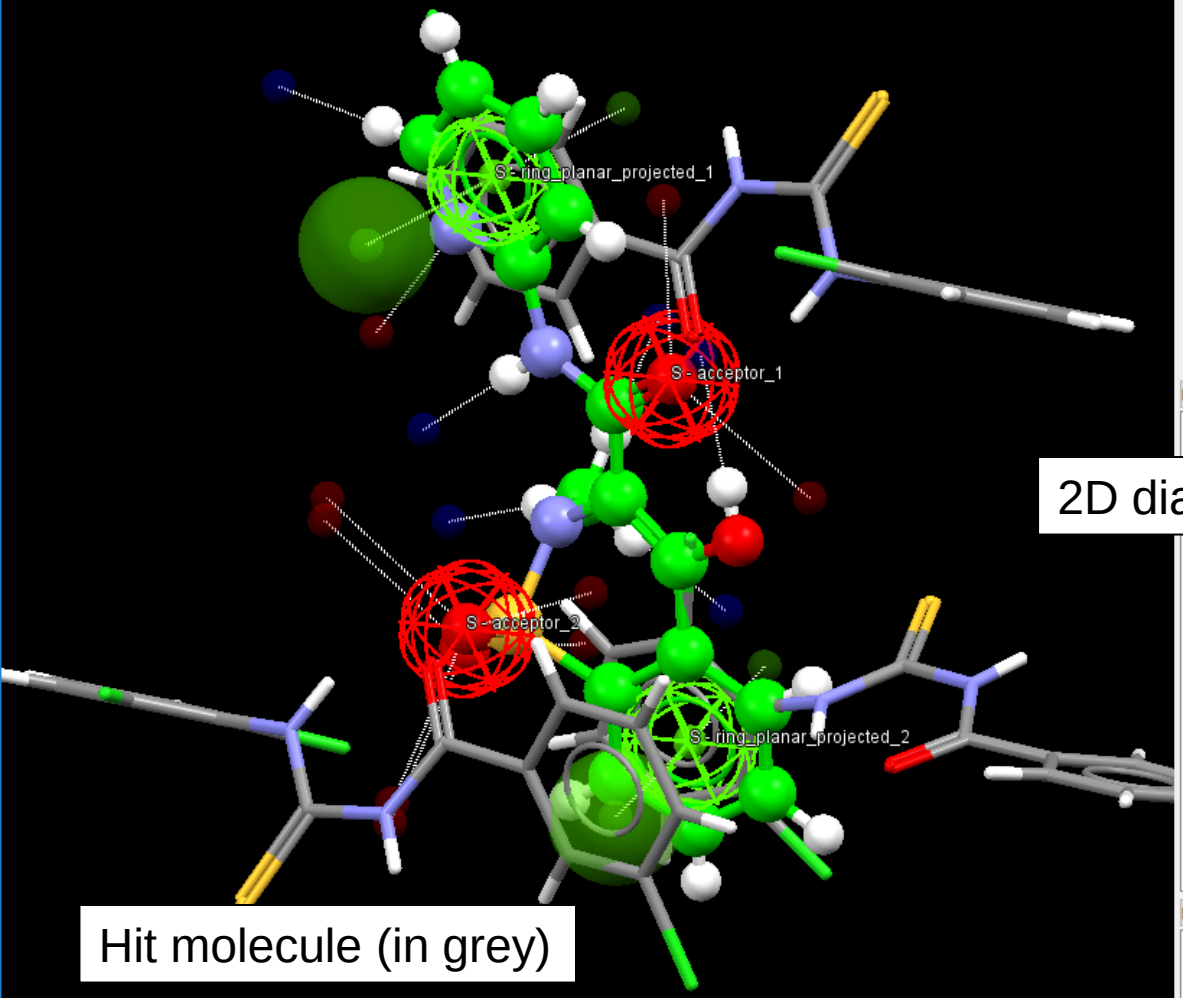
Feature Points – right-click to define query points


BIYSEH - CSD-CrossMiner
File Edit Display Feature Database Help
Style: Capped Sticks Colour: by Element Picking Mode:
Show: reference hits constraints features pharmacophore pharm. labels hydrogens Search: Edit:
x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90 zoom- zoom+
Software Interface
Start search with play button
Pharmacophore Query Points
S - ring_planar_projected_2
S - acceptor_2
S - ring_planar_projected_3
Results Hitlist
1st in cluster Settings... Tanimoto: 0.70 Number of hits: 370 Show all
msd diagram
#hits: 0/10000
feature name tolerance radius show in reference show in pharmacophore
acceptor_1
B 1.00
acceptor_2
B 1.00
acceptor_projected
acceptor_projected_1
B 0.70
V 0.70
donor_ch_projected
donor_projected
heavy_atom
heavy_atom_1
B 0.70
heavy_atom_2
B 1.00
Feature Databases
database size
pdb_crossminer 205194
csd539_crossminer 285437

Software Interface

As search is running, hits appear as they are found

Reference molecule in green



Hit molecule (in grey)

2D diagram with query points aligned

BIYSEH - CSD-CrossMiner

File Edit Display Feature Database Help

Style: Ball and Stick Colour: by Element Picking Mode:

Show: reference hits constraints features pharmacophore pharm. labels hydrogens Search: Edit: intra

x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90 zoom-

1st in cluster Settings... Tanimoto: 0.70 Number of hits: 370 Show all

mark	identifier	cluster	rmsd	diagram
☐	MEHJUF_1	33	0.808	

#hits: 795/10000

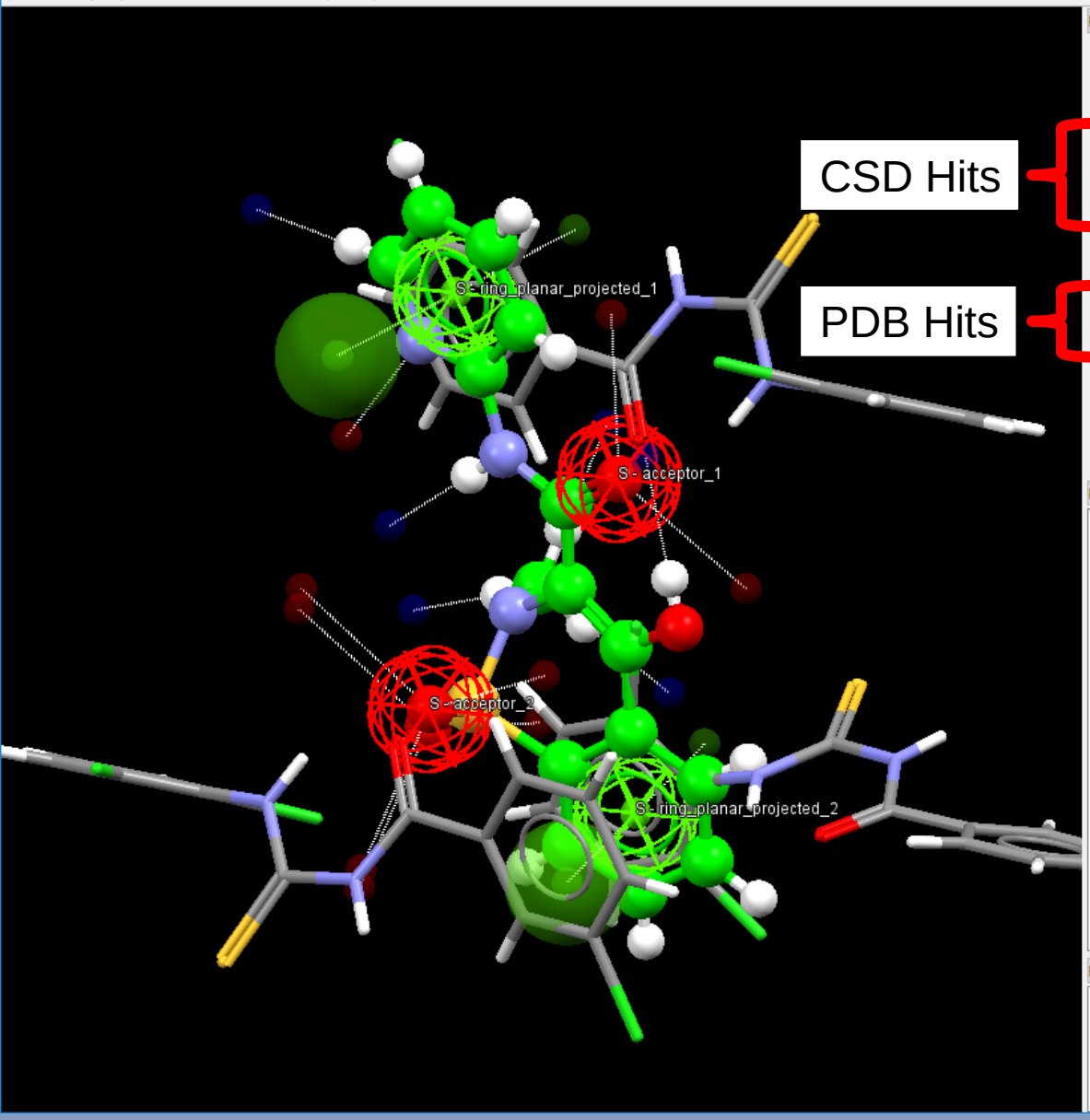
Pharmacophore Features

feature name	tolerance radius	show in reference	show in pharmacophore
acceptor_2			☑
B	1.00		
acceptor_projected		☑	
donor_ch_projected		☑	
donor_projected		☑	
heavy_atom		☑	
hydrophobe		☐	
ring		☐	
ring_non_planar		☐	
ring_planar_projected		☑	
ring_planar_project...			☑
B	1.00		
V	1.00		

Feature Databases

database	size
pdb_crossminer	205194
csd539_crossminer	285437

Software Interface



BIYSEH - CSD-CrossMiner

File Edit Display Feature Database Help

Style: Ball and Stick Colour: by Element Picking Mode:

Show: reference hits constraints features pharmacophore pharm. labels hydrogens Search: Edit: intra

x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90 zoom- zoom+

CSD Hits

PDB Hits

Results Hitlist

1st in cluster Settings... Tanimoto: 0.70 Number of hits: 370 Show all

mark	identifier	cluster	rmsd	chain	deposition_date	ec_number
	MEHLUH_1	36	0.802			
	MEHMER	37	0.929			
	MEHMIW_1	38	0.713			
	MEHMOC_1	39	0.766			
	MEHNAP	40	0.981			
	1EO3_m1_A-B_...	41	0.812	B	2000-03-21	3.1.21.4
	MEHNIY_1	42	0.759			
	5OE3_m1_D_bs_...	43	0.971	D	2017-07-07	6.2.1.32
	5P91_m1_A_bs_...	44	0.975	A	2016-08-29	2.1.1.6
	5P9H_m1_A_bs...	45	0.865	A	2016-09-20	2.7.10.2
	5PAC_m1_B_bs...	46	0.837	B	2016-11-10	3.4.21.21
	MEHPAR_1	47	0.79			

#hits: 795/10000

Pharmacophore Features

feature name	tolerance radius	show in reference	show in pharmacophore
acceptor_1			☑
B	1.00		
acceptor_2			☑
B	1.00		
acceptor_projected		☑	
donor_ch_projected		☑	
donor_projected		☑	
heavy_atom		☑	
hydrophobe		☐	
ring		☐	
ring_non_planar		☐	
ring_planar_projected		☑	
ring_planar_project...			☑
B	1.00		
V	1.00		

Feature Databases

database	size
pdb_crossminer	205194
csd539_crossminer	285437