

A faint, light gray network diagram is visible in the background, consisting of interconnected nodes and lines forming a complex web-like structure.

ccDC

advancing structural science

# What's Up

## CCDC Monthly Update

24 October 2019



# Today's presenters



Suzanna Ward  
Head of Database



Ilenia Giangreco  
Discovery Science Team Leader



Stewart Adcock  
Technical Manager, Scientific Software

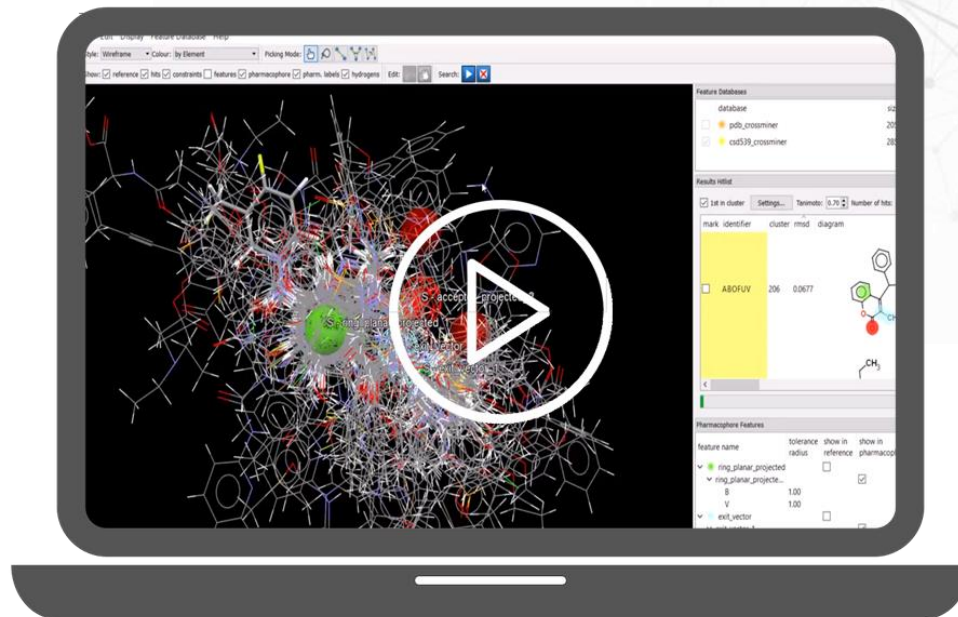


Ian Bruno  
Head of Strategic Partnerships

# Overview

A regular update on what's happening with CCDC and an opportunity for your news/feedback and suggestions:

- Welcome & introductions
- CCDC updates
- CrossMiner: Custom feature creation
- Coming soon: CSD Pipeline Pilot Components
- Open floor Q&A



# Warm up poll!

What sector are you in?



# What sector are you in?

Poll Results (single answer required):



# CCDC wide updates



Suzanna Ward  
Head of Database



# Welcome new team members

- Introducing
  - Nisshanth Dharmaraj – Deposition Coordinator



# Celebrating some of our CSD Heroes

## Leaderboard

Brian Skelton

Allan White

Arnie Rheingold

Peter Jones

Mike Hursthouse

Alex Slawin

Frank Fronczek

Judith Howard

Hoong-Kun Fun

Seik Weng Ng

## New blog series

### CSD Heroes: Alexandra Martha Zoya Slawin

Caroline Davies – October 16, 2019

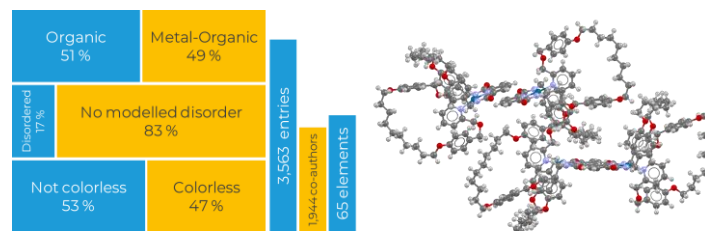
In the year that the CSD hit one million structures we are highlighting some of the most prolific contributors to the database. Our 5<sup>th</sup> CSD Hero in this series is Alexandra Martha Zoya Slawin who is currently 6<sup>th</sup> in our annual CSD author statistics. We wanted to thank Slawin for her contribution by doing what we do best – searching the CSD!

#### About Alexandra Martha Zoya Slawin

Alexandra Martha Zoya Slawin is Professor of Chemical Crystallography at the University of St. Andrews in the UK. Her research is primarily concerned with structure determination and its application to chemistry (1). In 2011 she was elected Fellow of the Royal Society of Edinburgh for being one of the world's leading chemical crystallographers (2). She completed her PhD at the University of Loughborough and her BSc in Chemistry at Imperial College London (3).

#### Alexandra M. Z. Slawin and the CSD

Alexandra is currently 6<sup>th</sup> in our annual list of most prolific authors in the CSD with over 3,500 entries dating back to 1984. She submitted a massive 276 entries in 2010, her busiest year yet! A graph showing the growth of Alexandra's entries in the CSD per year is shown below.



Alexandra has published structures in 107 different journals with 1,944 unique co-authors.

Her structures are diverse and include 65 different element types, a massive 297,142 atomic coordinates and cover 84 different space groups. One of her largest structures in terms of the number of atomic coordinates (616) is CSD Entry: QAVGIE which was published in *J. Am. Chem. Soc.* in 2005.

Get involved...

#CSDHeroes19  
#MyCSD1inaMillion



LinkedIn: [company/cambridge-crystallographic-data-centre](https://www.linkedin.com/company/cambridge-crystallographic-data-centre)

Twitter: @ccdc\_Cambridge

Facebook: @ccdc.cambridge

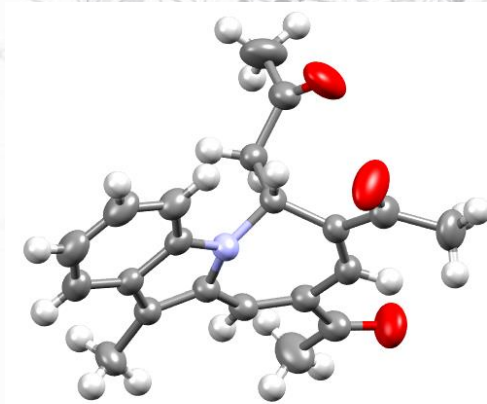
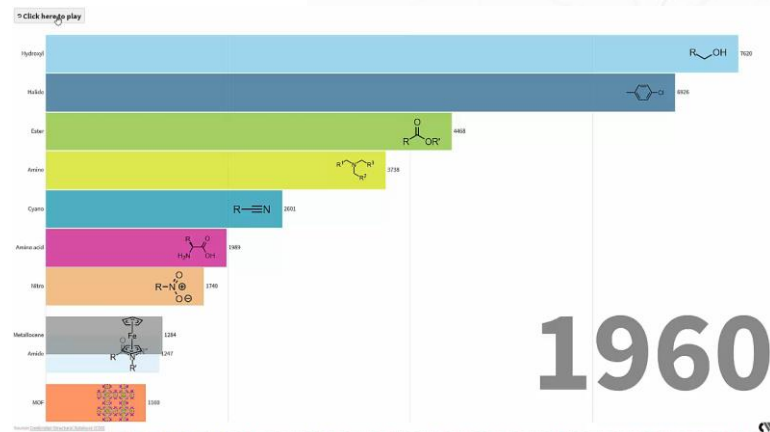
YouTube: CCDCCambridge

<https://www.ccdc.cam.ac.uk/Community/blog/>

CCDC

# Evolving the CSD

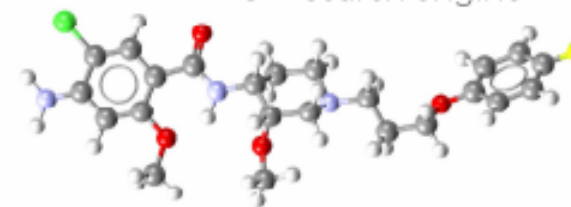
- Our database is evolving...
  - More data
  - New data integrity checks
  - Faster, more effective searches



## The rewriting of ConQuest's fundamental search engine means that:

- Many complex 2D substructure searches are now much faster
- Nearly all types of search are significantly faster than the previous version
- Element and formula searches are more effective
- Many text searches now give more accurate results, especially author searches which now properly handle international name conventions

2018: Complete C++ search engine



# Warm up poll!

What would you like us to add to the CSD next?

## QUICKPOLL

## What would you like us to add to the CSD next?

Poll Results (multiple answers allowed):

Additional CIF data fields	22%
Additional property data alongside the entries	44%
Better linking between CSD entries	28%
Data from different techniques such as NMR, ED and CSP	50%
More data quality metrics and filters	39%

# Recent CCDC events and activities

- Jeff Lengyel led a CCDC workshop at Yale
- US UGM in Philadelphia
  - Accelerating science through data and machine learning
  - Invited speakers
  - Product roadmap and future directions
- Various staff in Boston, US
  - Rational Solid Form Design and Development Summit 2019
  - Meeting with customers
  - Pistoia Alliance USA Conference 2019





# Current/upcoming CCDC events

- CCDC visit to Japan
  - Luncheon seminar
  - Japanese UGM
  - Customer visits
  - Andrew Maloney and Francesca Stanzione leading scientific discussions
- Crystal engineering conference in Israel
  - Contemporary Crystal Engineering and Solid-State Chemistry
  - Weizmann Institute of Science
  - Khaled Takieddin from our materials science team attending

**JAICI**  
化学情報協会

日本語 English

お問い合わせ 資料請求 Google カスタム検索

製品・サービス一覧 STN SciFinder<sup>n</sup> 調査依頼 (SHIPS) CAS登録番号サービス 結晶構造DB・化合物辞典

HOME > 最新情報一覧 > JAICI からのお知らせ

2019年9月26日 (木)

**CCDCユーザーミーティングのお知らせ**

弊協会は、以下の日程でCCDCユーザーミーティング 2019「Going for Growth」を開催いたします。CCDCに直接質問できるセッションも予定しておりますので、どうぞお気軽にお申込みください。

CSDは、The Cambridge Crystallographic Data Centre (CCDC) が運営する世界最大の有機低分子の結晶構造データベースです。1965年に結晶構造を集め始め、構造化学における金字塔として本年100万件目の構造を収録いたしました。CCDCでは、CSDに蓄積された構造情報を医薬分子の設計のみならず、材料開発に注目した製品開発にも取り組んでおります。今回のユーザー会では、CSD 1Mのデータをどのように生かすか、基本となるCSD-Systemおよびアップグレード版のCSD-Discoveryと CSD-Materialsの基本機能や活用方法をご紹介します。

プログラムの詳細は [こちら](#)。

日時	2019年10月28日(月) 10:00 - 17:00
会場	化学情報協会 中層ビル6階講習会室 (東京都文京区) <a href="https://www.jaici.or.jp/map.html">https://www.jaici.or.jp/map.html</a>
対象	企業、大学などの研究機関にて、創薬、製剤、分子設計に携わっている方、有機物の構造化学研究に携わっている方。
参加費	無料 (事前登録が必要)
申込締切	2019年10月18日 (金)

מכון ויצמן  
WEIZMANN INSTITUTE OF SCIENCE

Symposium commemorating  
G. M. J. Schmidt's 100th birthday anniversary

**Contemporary Crystal Engineering  
and Solid-State Chemistry**

23-24 October 2019  
The David Lopatie Conference Centre ► Weizmann Institute of Science

Accessibility

<https://www.ccdc.cam.ac.uk/News/Events/>

CCDC

# Research Data Alliance (RDA)



- Ian Bruno our Head of Strategic Partnerships is currently at the 14th Plenary RDA meeting in Helsinki
- The meeting brings together data experts, industry and policy-making from around the world under the theme "Data Makes the Difference".





# CrossMiner

Custom feature creation

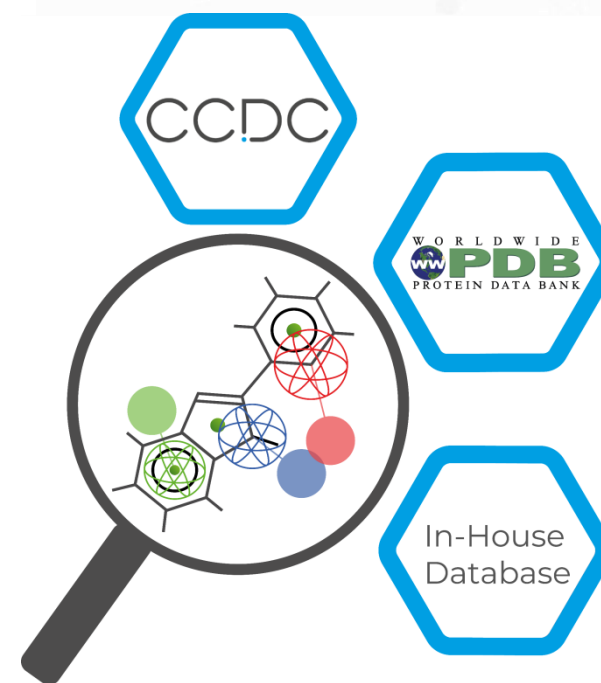


Ilenia Giangreco

Discovery Science Team Leader

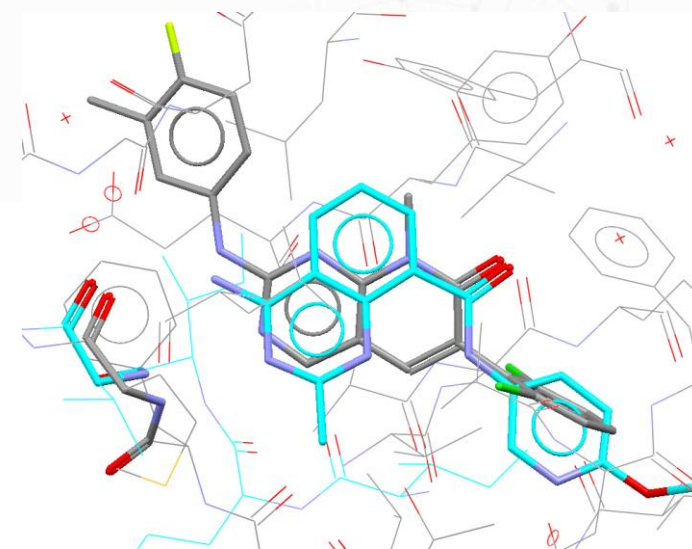
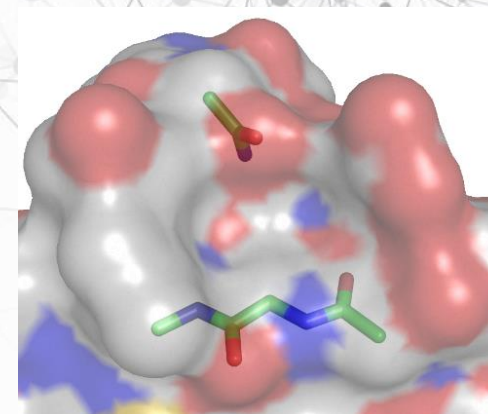
# CSD-CrossMiner overview

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



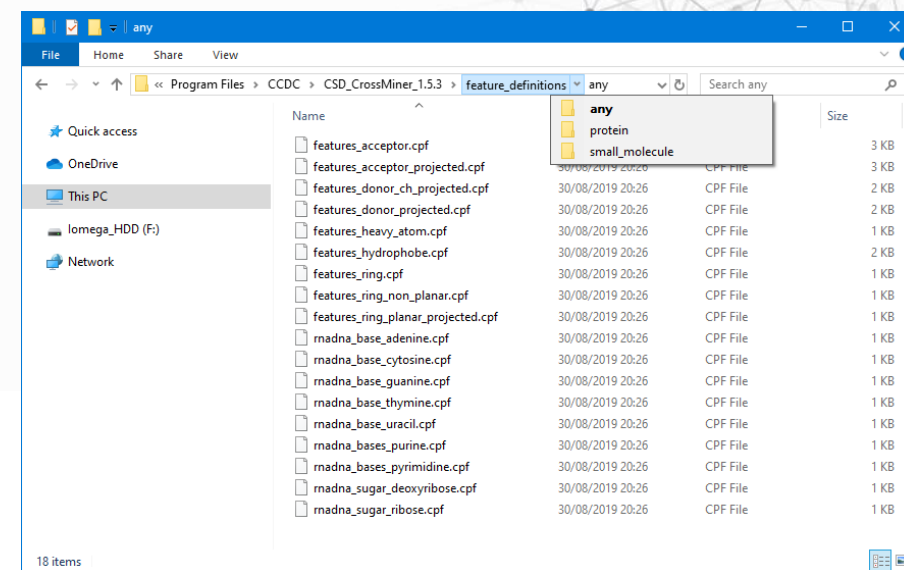
# CSD-CrossMiner applications

- 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 established ligands  
→ improve molecular properties; solve patent issue
  - 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



# CSD-CrossMiner feature definitions

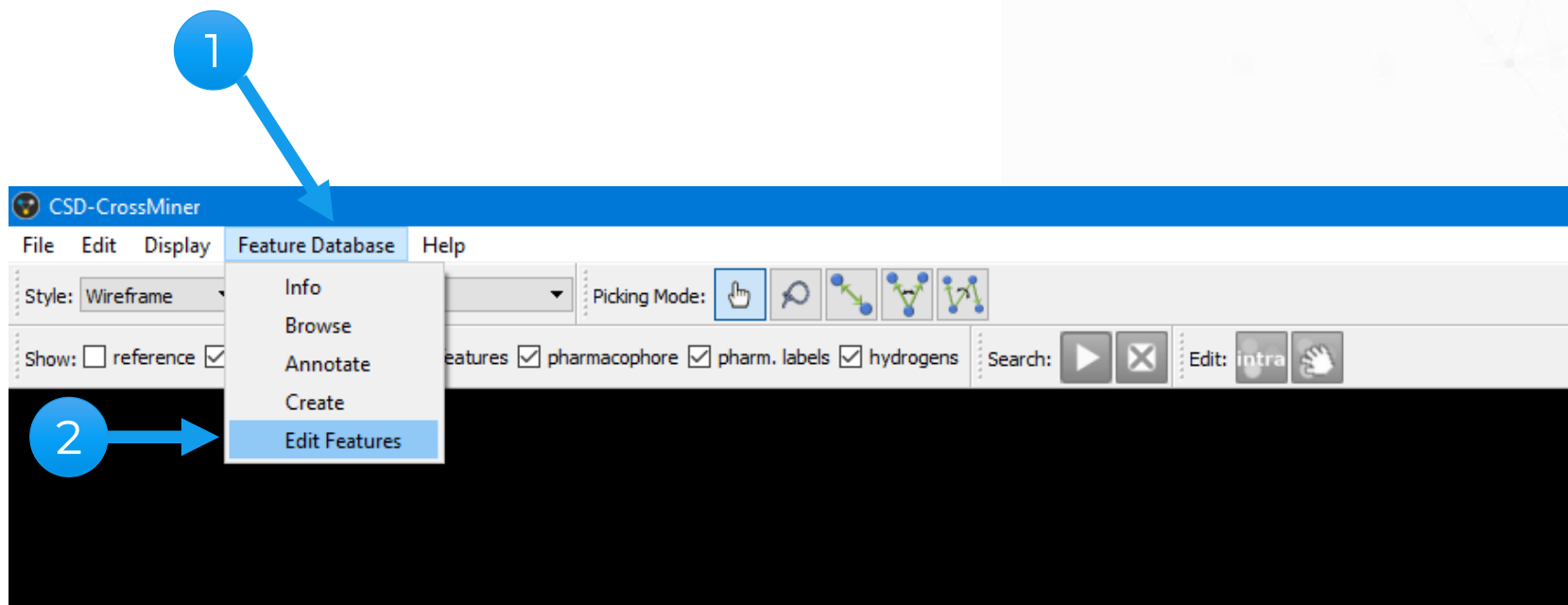
- Feature definitions are based on SMARTS substructure definitions available from <Installation folder>\feature\_definitions
- Features definitions can be created and/or edited



We will learn how to generate a custom feature definition for Zinc so that we can search for all Zinc-binding ligands in the database.

# Creating new feature definitions

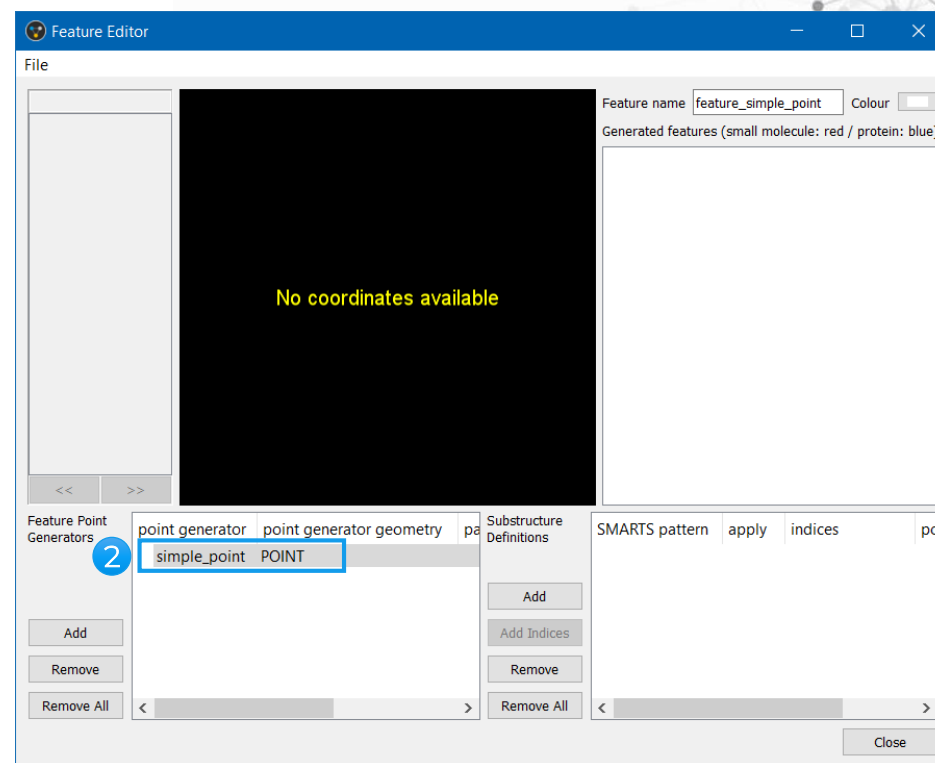
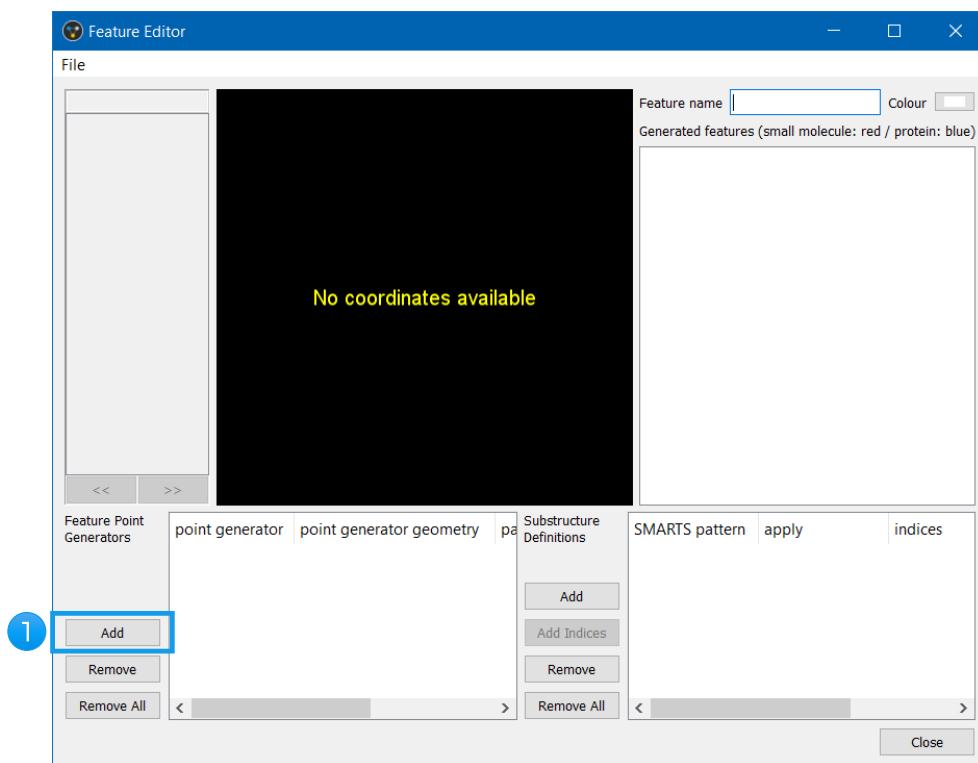
- Features can be created using the *Feature Editor* menu accessible from Feature Database → Edit Features



# Creating new feature definitions

## Adding a feature point generator

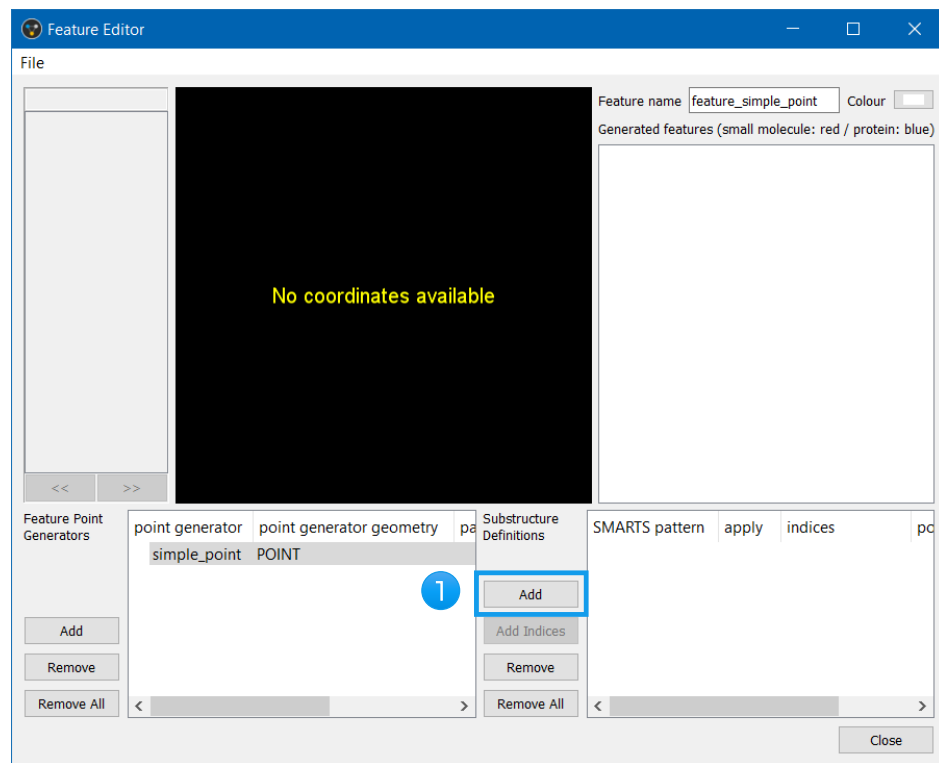
- A *simple\_point* feature is generated at the selected atom position



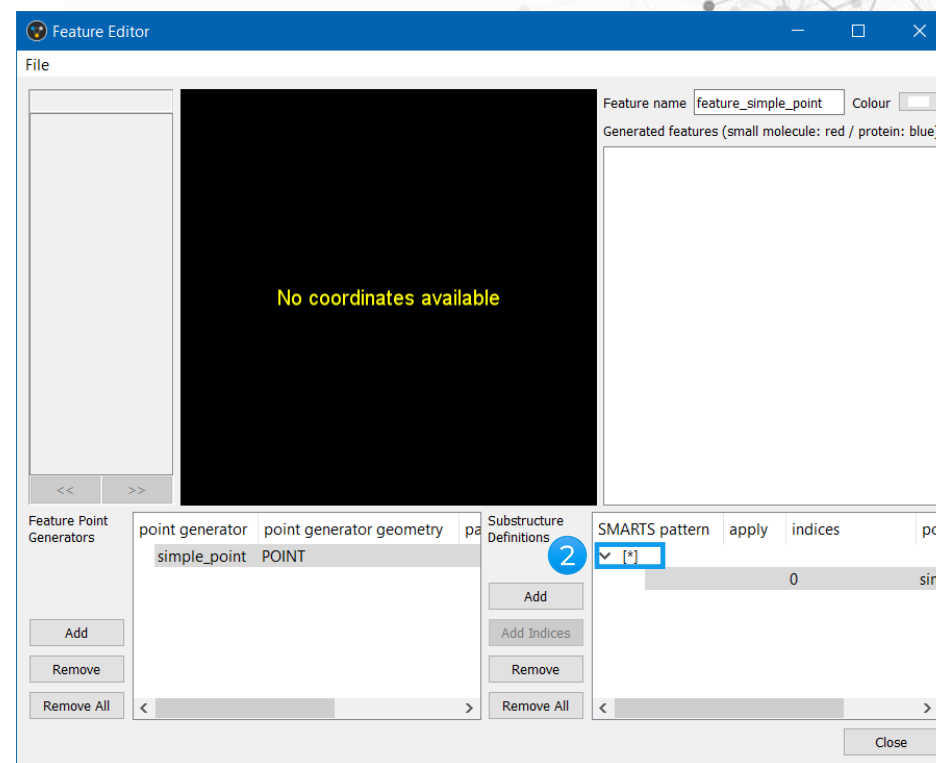


# Creating new feature definitions

## Adding a substructure definition



1. By default a SMARTS pattern of [\*] is generated - this corresponds to every atom



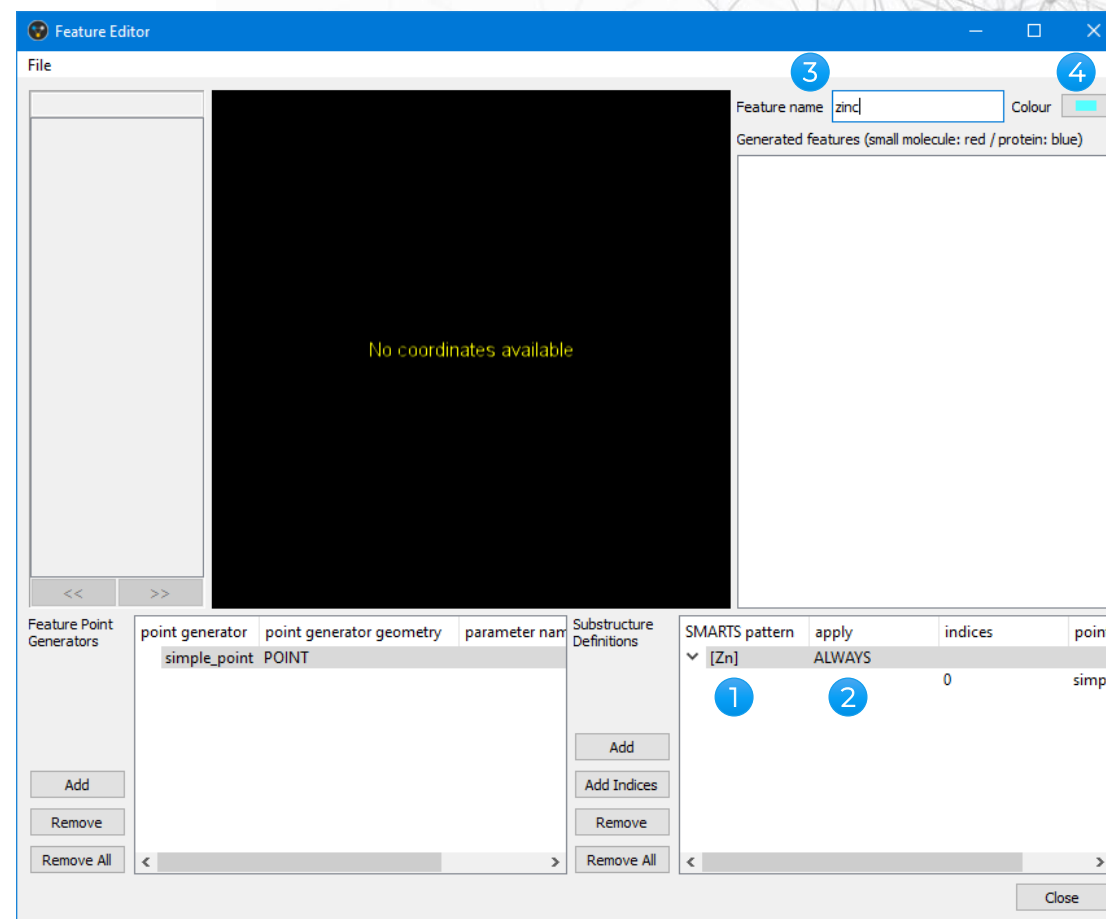
2. Double clicking on [\*] allows to edit the SMARTS pattern



# Creating new feature definitions

## Editing the SMARTS definition

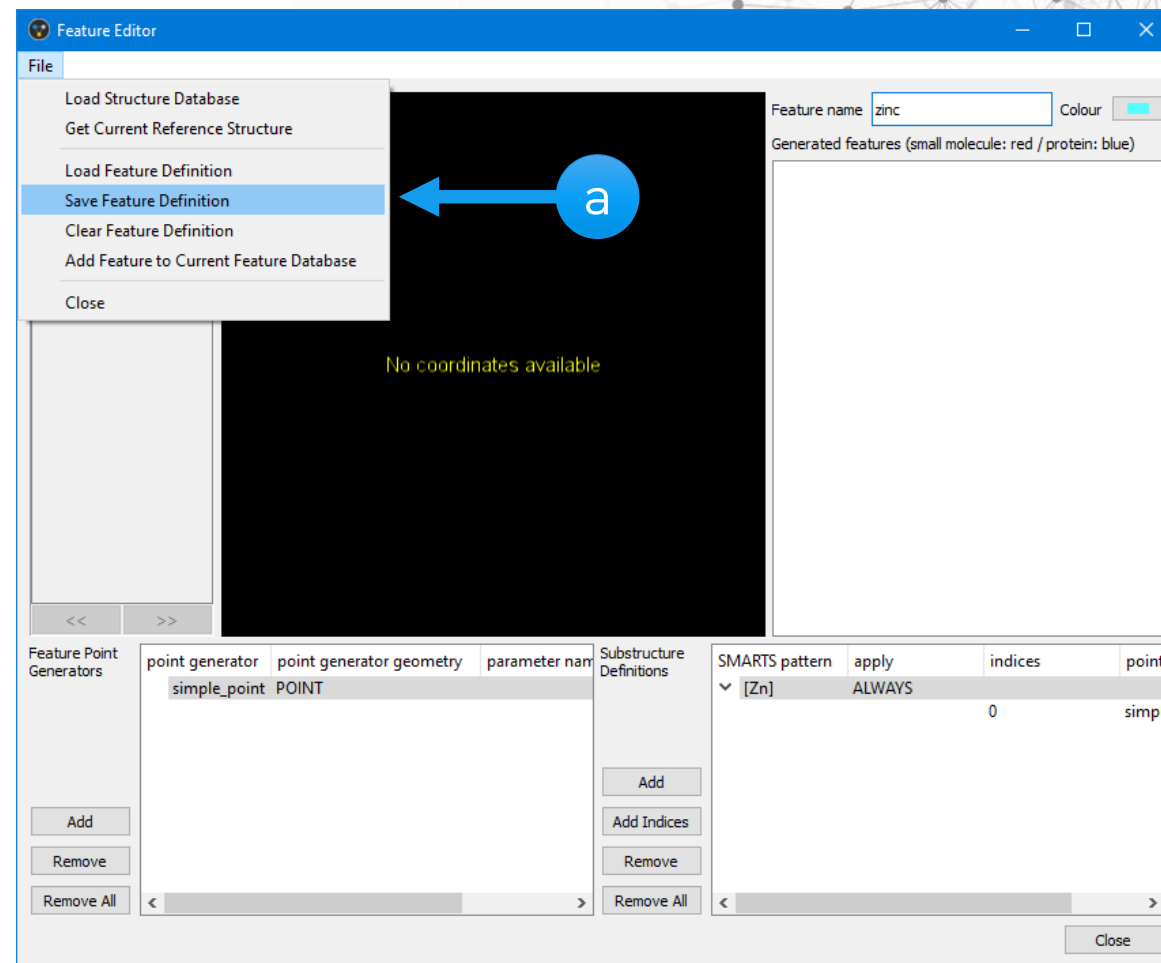
1. In the *SMARTS* pattern replace [\*] with [Zn]
2. In the *apply* section select **ALWAYS**
3. Change the *Feature name* to *zinc*
4. Pick a new *Colour* for the newly created zinc feature



# Creating new feature definitions

## Exporting the new feature

- a) The new feature can be saved and used for example to regenerate the feature database



# Creating new feature definitions

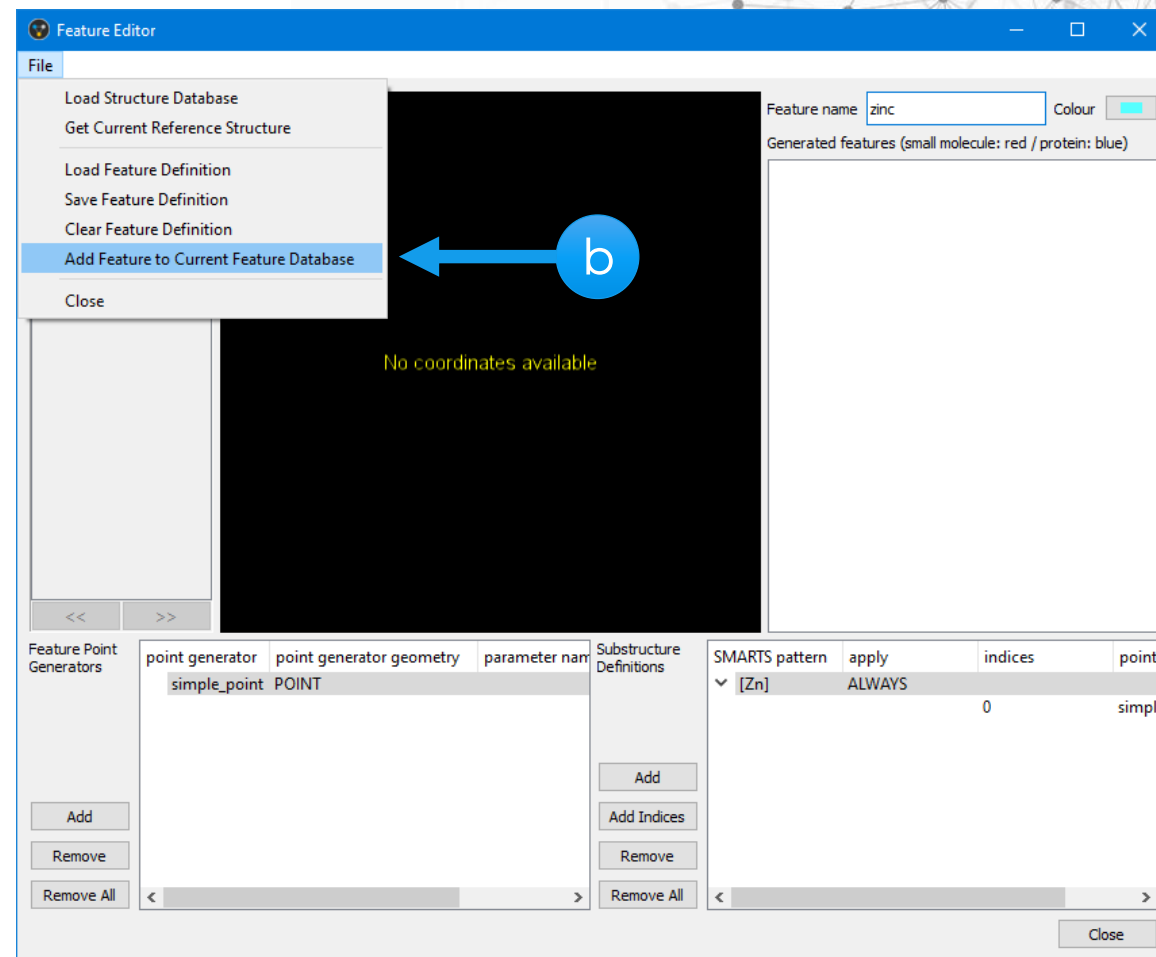
## Exporting the new feature

a) The new feature can be saved and used for example to regenerate the feature database

OR

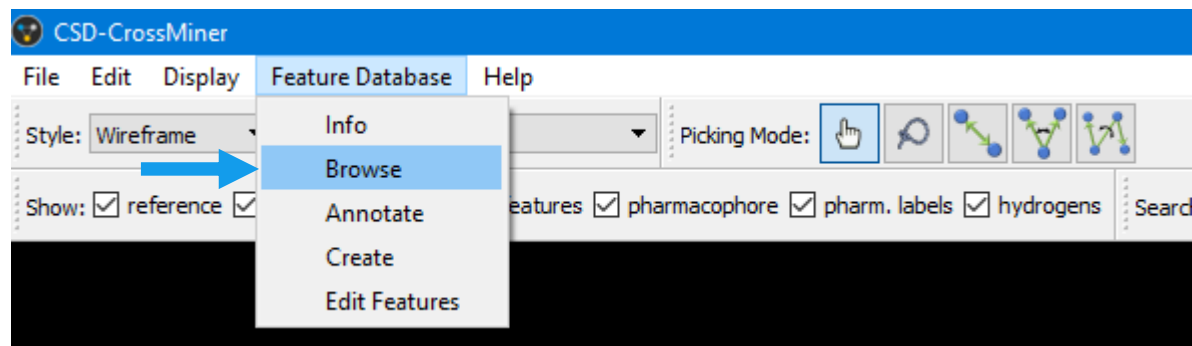
b) The new feature can be added to the loaded feature database and appended to the *Pharmacophore features* window

Pharmacophore Features			
feature name	tolerance radius	show in reference	show in pharmacophore
TRP		<input type="checkbox"/>	
TYR		<input type="checkbox"/>	
VAL		<input type="checkbox"/>	
excluded_volume		<input type="checkbox"/>	
annotation_filter		<input type="checkbox"/>	
substructure_filter		<input type="checkbox"/>	
zinc		<input checked="" type="checkbox"/>	

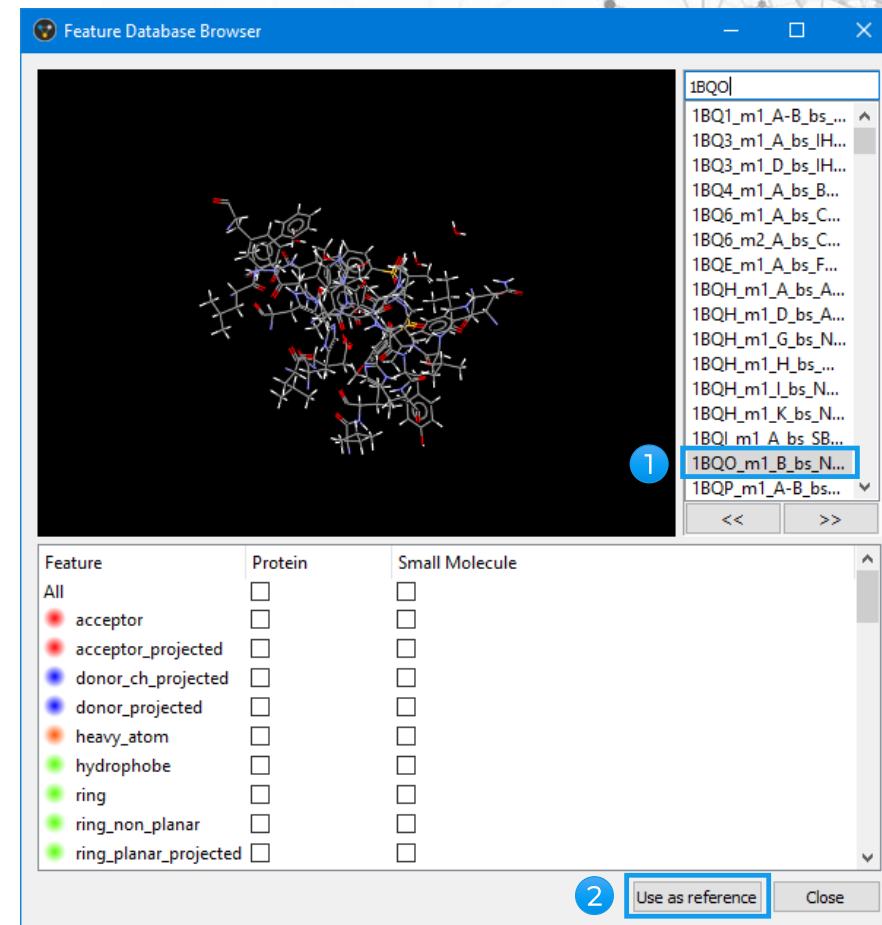


# Using new feature definition on the fly

## Loading a structure from the feature database

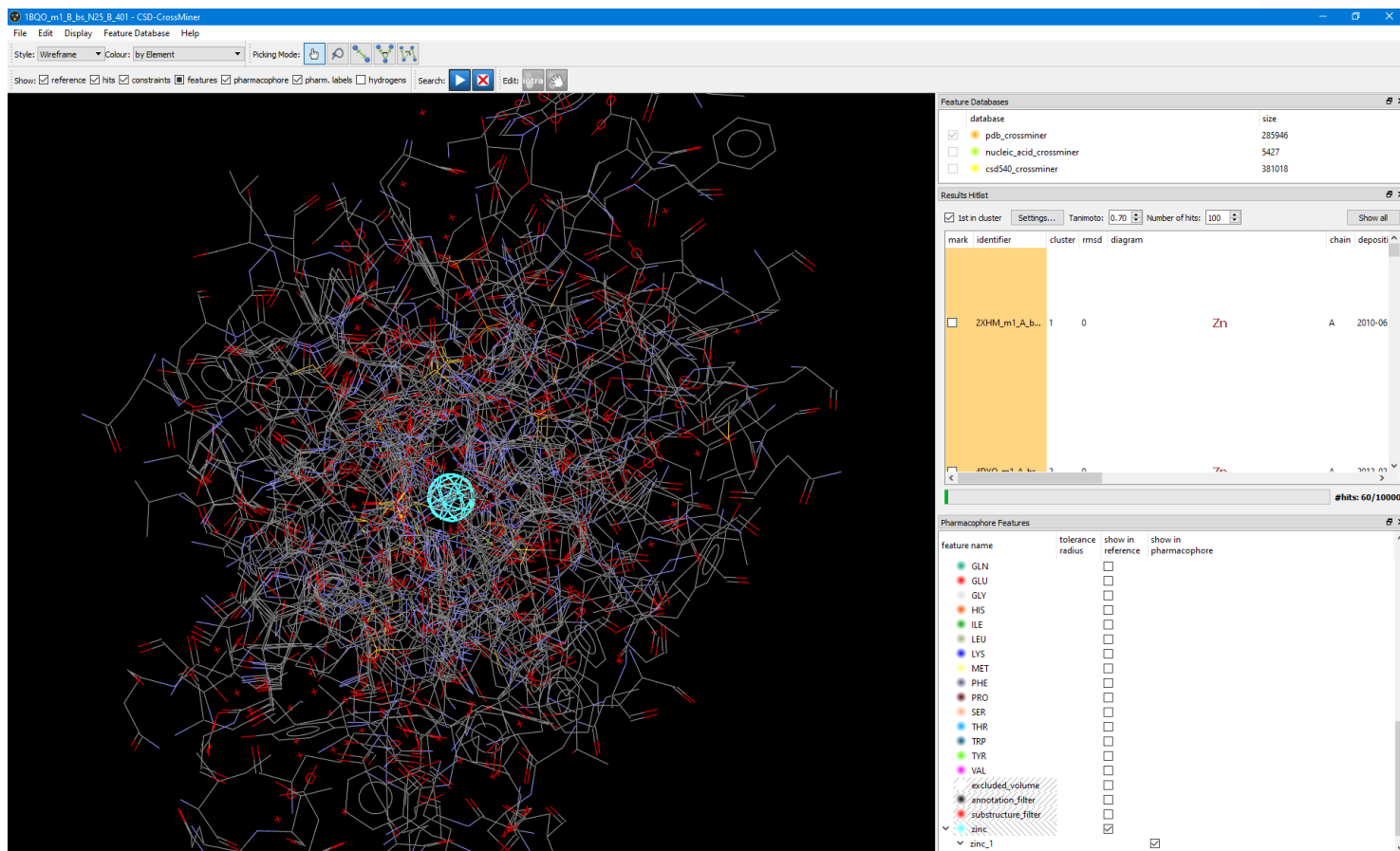


1. Select *1BQO\_m1\_B\_bs\_N25\_B\_401* from the list of structures
2. Use it as reference to define a pharmacophore query



# Using new feature definition on the fly

## Searching for Zinc-binding ligands



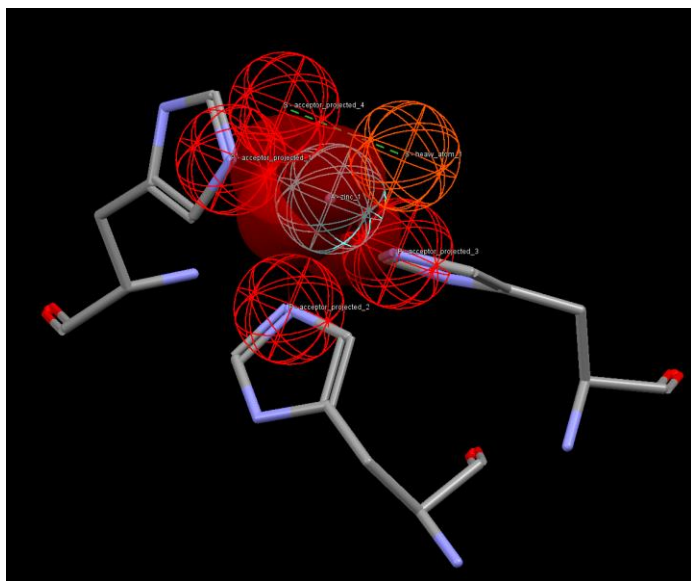
- Searching for all Zinc-containing structures in the database



# Using new feature definition on the fly

## Searching for Zinc-binding ligands

- Searching for bidentate Zinc-binding ligands



**Feature Databases**

database	size
pdbs_crossminer	285946
nucleic_acid_crossminer	5427
csd540_crossminer	381018

**Results Hitlist**

mark	identifier	cluster	rmsd	diagram	chain	deposit
<input type="checkbox"/>	1A85_m1_A_hs_3	3	0.704		A	1998-04

**Pharmacophore Features**

feature name	tolerance radius	show in reference	show in pharmacophore
All		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acceptor		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acceptor_projected		<input type="checkbox"/>	<input type="checkbox"/>
donor_ch_projected		<input type="checkbox"/>	<input type="checkbox"/>
donor_projected		<input type="checkbox"/>	<input type="checkbox"/>
heavy_atom		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
heavy_atom_1		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
B	1.00	<input type="checkbox"/>	<input type="checkbox"/>
hydrophobe		<input type="checkbox"/>	<input type="checkbox"/>
ring		<input type="checkbox"/>	<input type="checkbox"/>
ring_non_planar		<input type="checkbox"/>	<input type="checkbox"/>
ring_planar_projected		<input type="checkbox"/>	<input type="checkbox"/>
purine		<input type="checkbox"/>	<input type="checkbox"/>
pyrimidine		<input type="checkbox"/>	<input type="checkbox"/>
adenine		<input type="checkbox"/>	<input type="checkbox"/>
cytosine		<input type="checkbox"/>	<input type="checkbox"/>
guanine		<input type="checkbox"/>	<input type="checkbox"/>
thymine		<input type="checkbox"/>	<input type="checkbox"/>
uracil		<input type="checkbox"/>	<input type="checkbox"/>
deoxyribose		<input type="checkbox"/>	<input type="checkbox"/>

# Pipeline Pilot

CSD component collection



Stewart Adcock

Technical Manager, Scientific Software



# CSD Pipeline Pilot Component Collection

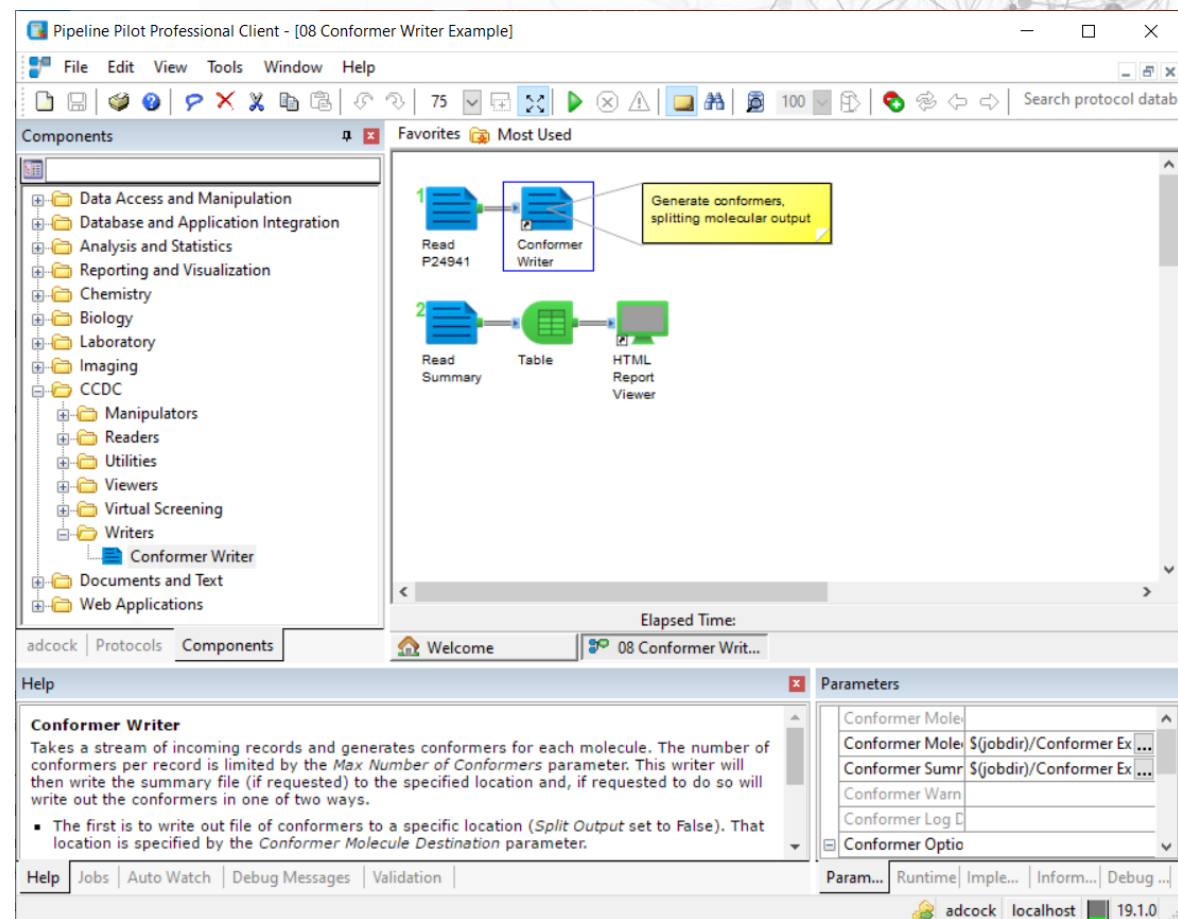
- How can we make CSD data more accessible and useful?
- We want to unlock valuable CSD data and features from our various scientific methods outside of our pre-packaged desktop and web-based applications.
- The CSD Python API largely does this already.
  - The CSD Python API is our recommended interface for integrating with 3<sup>rd</sup> party tools.
- We can now use the API to relatively easily build Pipeline Pilot components that expose the CSD.



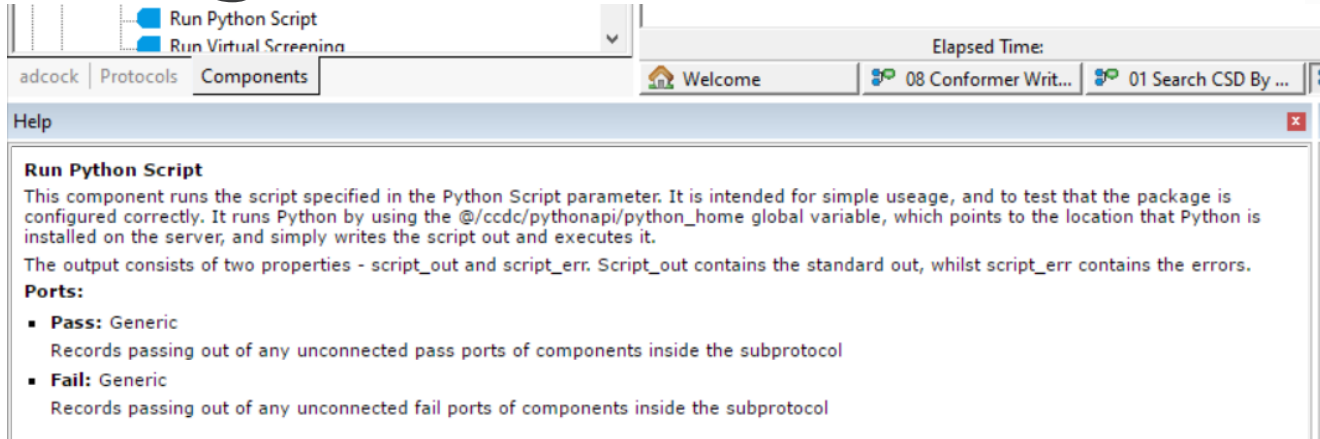
We will learn about the upcoming CSD Pipeline Pilot component collection.

# CSD PP Component Collection v.1

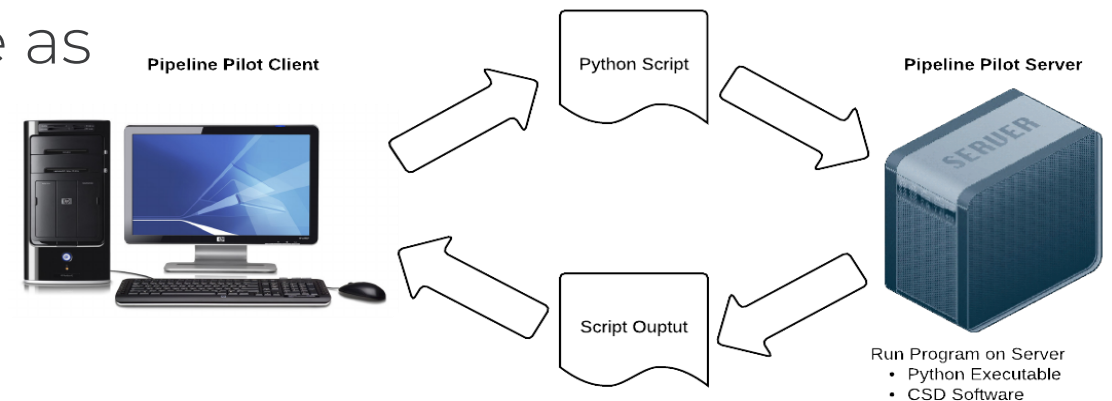
- Components
  - CSD data access and search nodes
  - CSDS Viewer nodes
  - Conformer generation
  - Very basic virtual screening
- Protocols
  - CSD Searching
  - Python Examples
  - Virtual Screening and Conformer Generation



# Integration Scheme

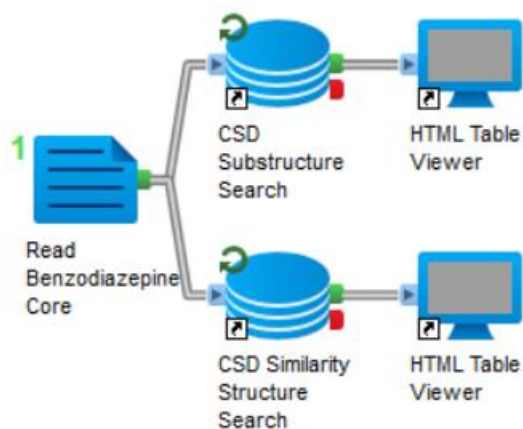


- Our solution is to make arbitrary Python Scripts, including those that use the CSD Python API, available as PP components.
  - Alongside tools for parsing and manipulating CSD data types.
  - CSDS visualisation tools available as viewer components.



# Exemplar Protocols

- Purpose at this stage is to demonstrate use of the components.

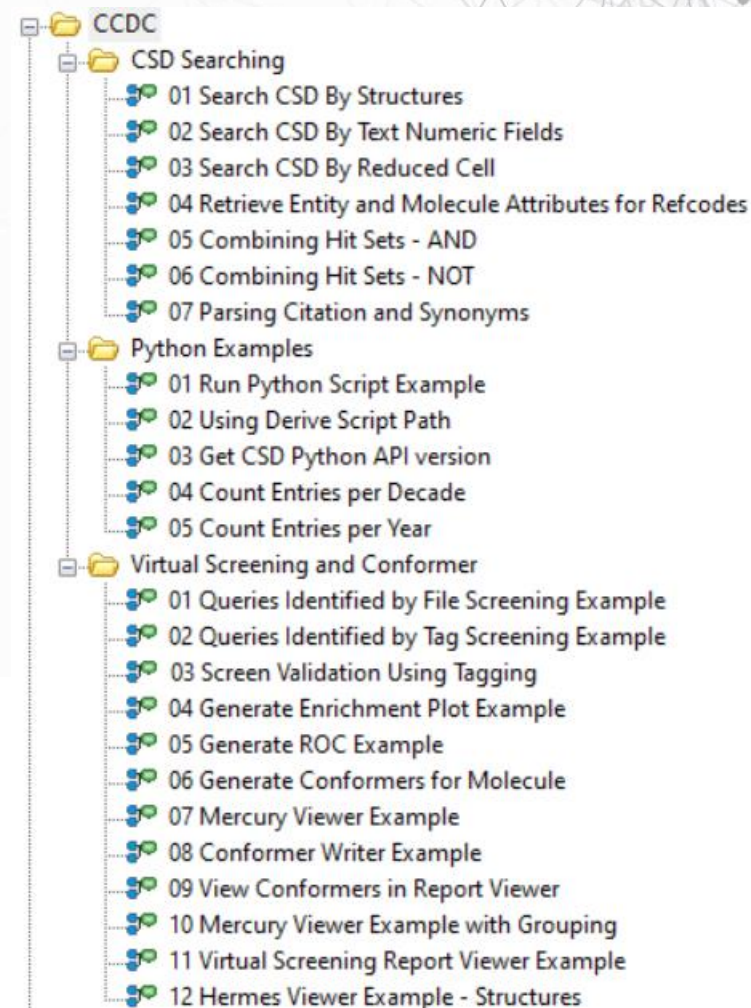


## 01 Search CSD By Structures

This protocol uses the benzodiazepine core structure, shipped as part of the Chemistry collection, and searches CSD for structures which contain this as a substructure (top) or as a similar compound (bottom).

The search component outputs as a minimum the refcode(s) for hit structures.

Similarity searches take into account the similarity threshold - which all hit structures must exceed. To aid in the interpretation of the results, the component outputs an additional property which is the similarity value for this compound to the query for which it was found.





# Running any arbitrary python script

- This is the basis of most CSD components.
- It need not rely on the CSD Python API



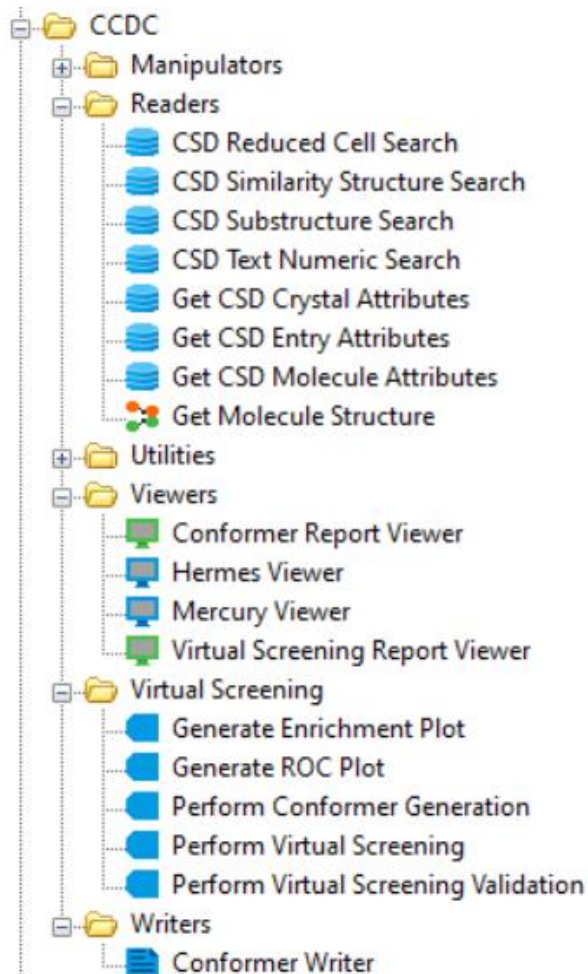
## Run Python Script Example

This protocol takes a simple example script and displays it to the user before executing it. The script is simple, taking a set of numeric inputs and summing them together.

Switching the final argument to `--max` will return the max value instead.

It demonstrates how the Run Python component can be configured to point to any Python script and execute it.

# Components



- Search:
  - Reduced Cells
  - Similarity
  - Substructure
  - Textual/Numerical Data
- Extract data with *Get CSD Crystal/Entry/Molecule Attributes*
- Access 3D structures with *Get Molecule Structure*

# Release Plan

- At the time of our next CSD-System release, we will make this component collection available to research partners.
- Over the following months, we will roll these out to other user categories.
- Depending on the uptake and feedback, we can extend to cover other features of the API.
- This is our call for discussion.
  - What do you need to most effectively use the CSD in your protocols?



# Benefits to users

- Existing users of PP can access the CSD.
- Existing users of PP can solve their problems using supported CCDC software.
- Search tasks can be implemented/scripted within automated workflows.
- No additional platforms or GUIs to install or learn how to navigate.

# Upcoming Events

- CSP Database Working Group Kick-Off: Nov 6, Cambridge
- AsCA: Dec 18-20, Singapore

Email [hello@ccdc.cam.ac.uk](mailto:hello@ccdc.cam.ac.uk) for more info

# Q&A

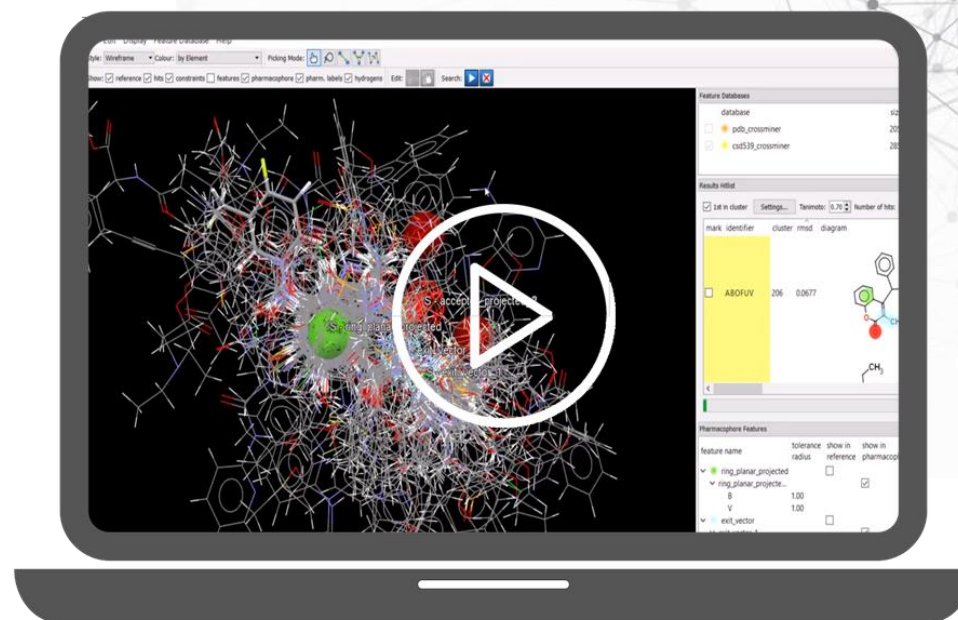
The floor is yours



# Next What's Up Webinar

- Next webinar: November 21
- Send us your ideas and news

[hello@ccdc.cam.ac.uk](mailto:hello@ccdc.cam.ac.uk)



# Thank you