RDKit UGM
Sep 20-21 2017, Berlin

XChem and SQUONK

Open source tools for fragment-based drug design using the RDKit

Tim Dudgeon
Informatics Matters
tdudgeon@informaticsmatters.com

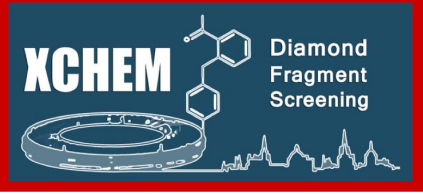
Anthony Bradley
Diamond Light Source and Oxford University
anthony.bradley@chem.ox.ac.uk

Topics

1. The challenge being addressed
2. The challenge at Diamond
3. CCP-CMC
4. Pipelines project
5. Docking Validation project
6. Conclusion

The challenge being addressed

Our Basic Premise

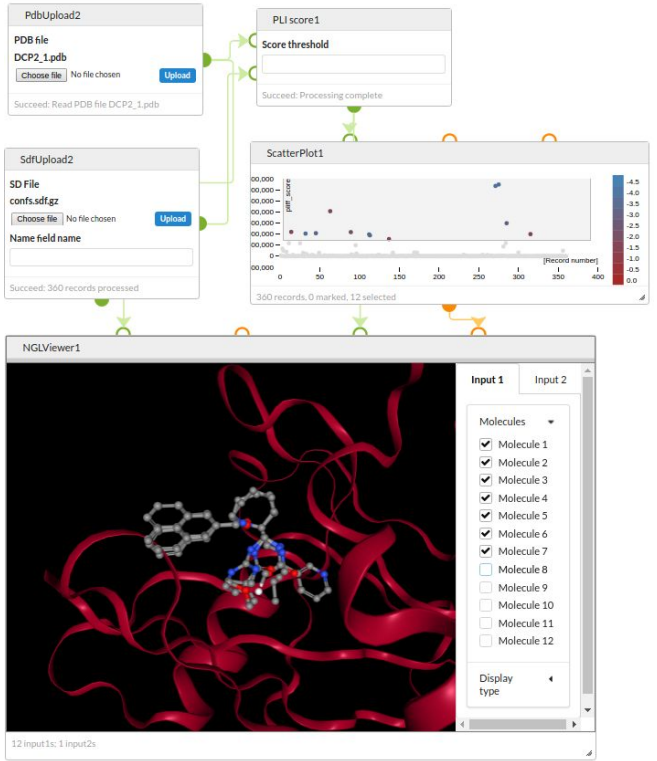


Lots of powerful computational tools exist, but they rarely get into mainstream use, often only being effectively used within the research groups that created them.

The primary reason for this is that most tools are difficult to access and do not integrate well into the overall workflow process.

Our Aim

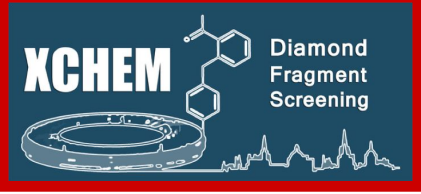
- Democratise cheminformatics & computational chemistry (and beyond)
- Make complex tools accessible to all
- Break down barriers to access
- Provide traceability and reproducibility
- Facilitate collaboration



https://github.com/InformaticsMatters/squonk
Apache 2.0 license

https://squonk.it

Squonk Architecture



Computational Notebook

workflows analysis

collaboration provenance

App 2

App 3

Services

calculate predict transform data

Runtime

docker kubernetes OPENSHIFT keycloak

Hardware

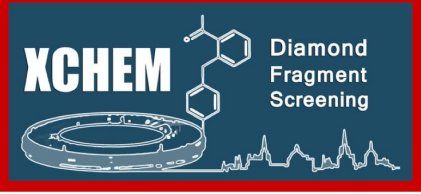
Cloud/VPC
Bare metal
Laptop

Today's talk is about this

How to plug in interoperable RDKit based services

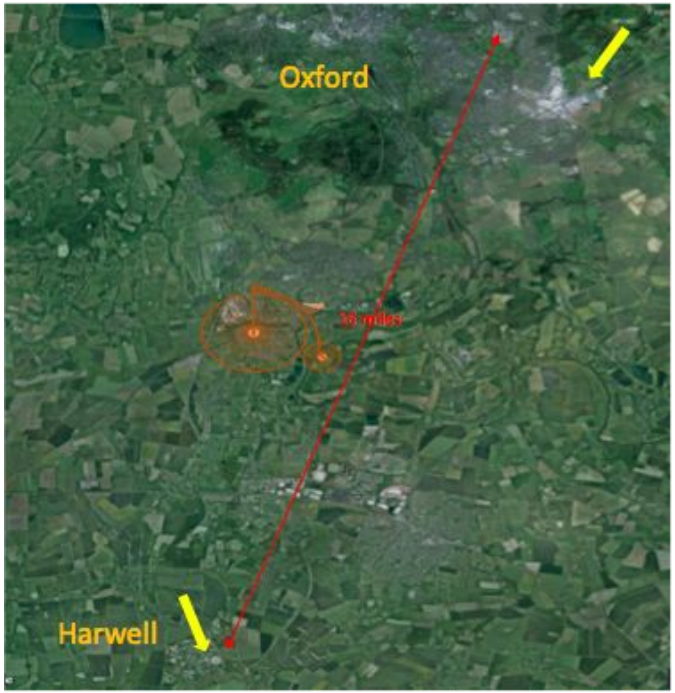
The Challenge at Diamond

Me

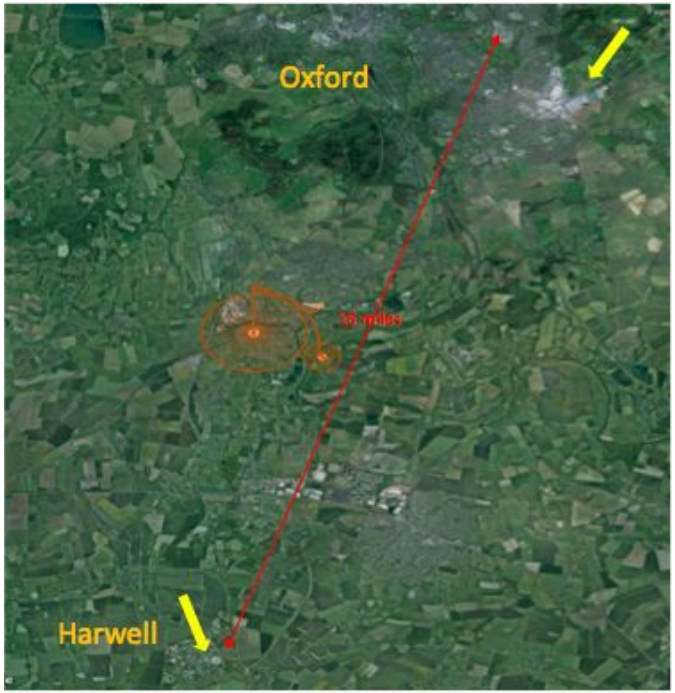


Anthony Bradley

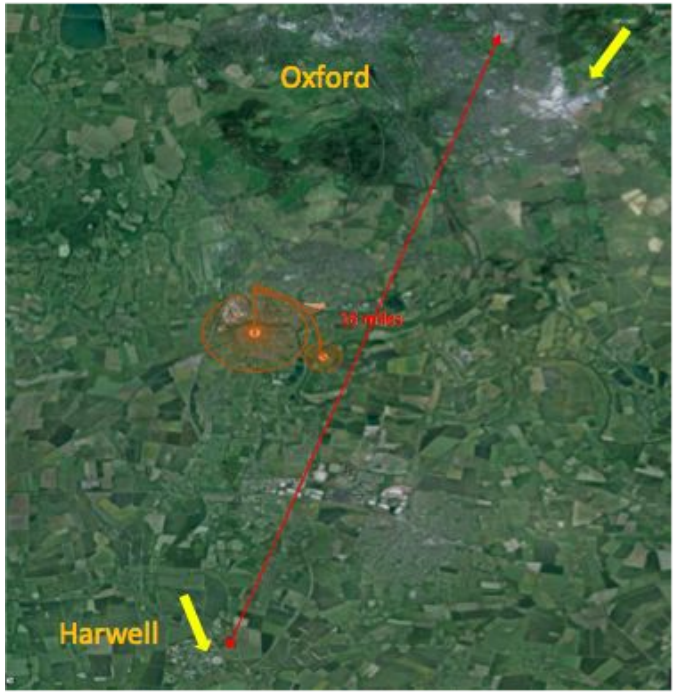
Diamond



Diamond

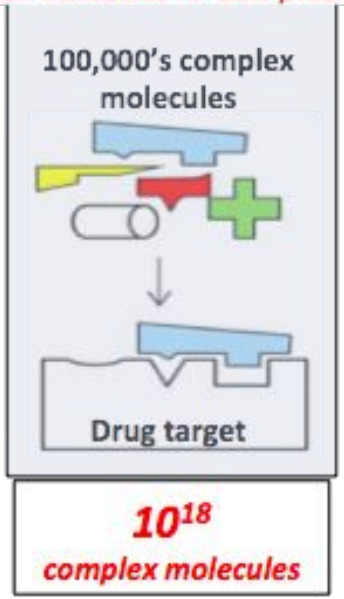


Diamond



Conventional screening

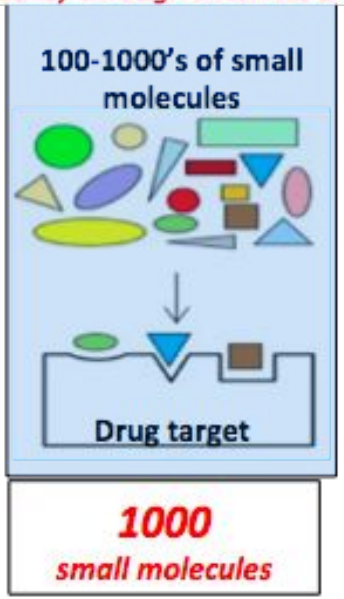
- *Searching for potent molecules*
- *Complex molecules → Low probability*



≈

Fragment screening

- *Guaranteed binding – but weak*
- *Potency through chemical elaboration*

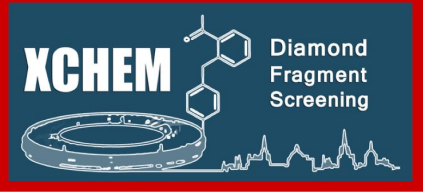


Diamond Background

Diamond Background

Spurlino, Meth Enz, 2011

Diamond current state



@Diamond: true user programme – regular experiments

(>1/wk)

793 hits across 20 targets

(Oct 2016)

(8 hits)

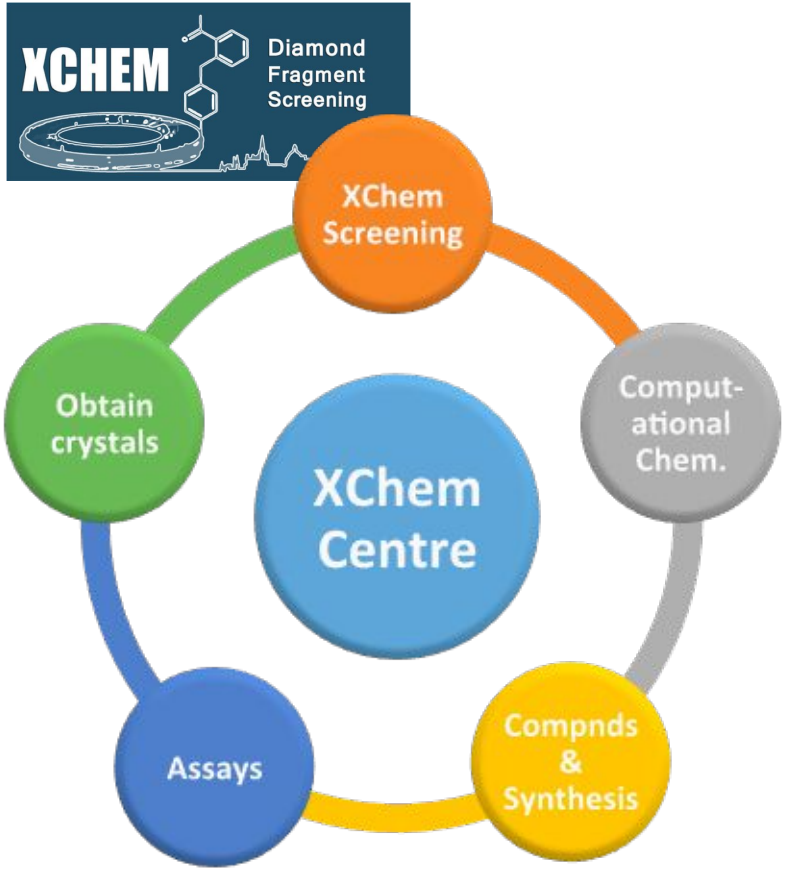
peptide

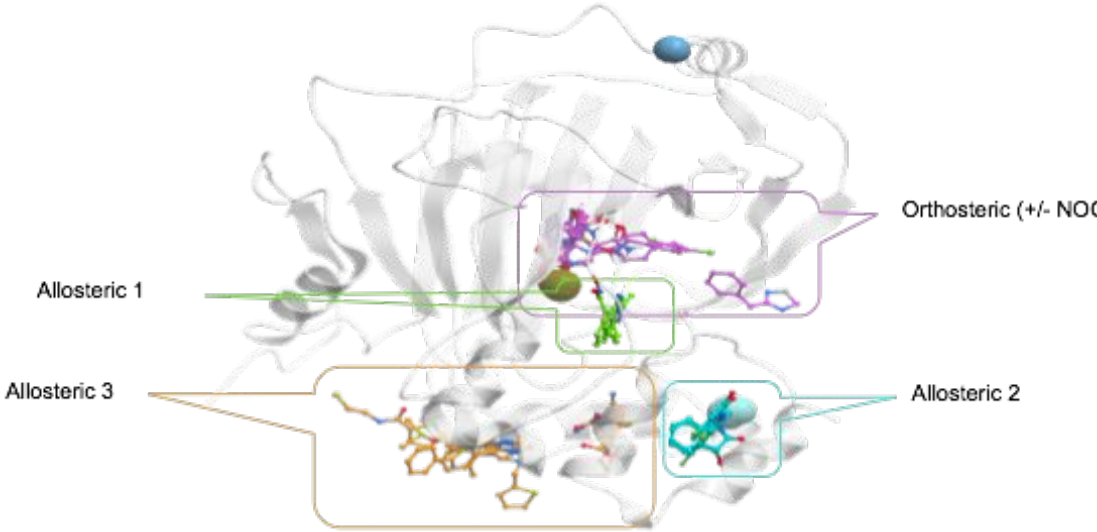
site (3 h

hotspot

(9 hits)

Diamond current state

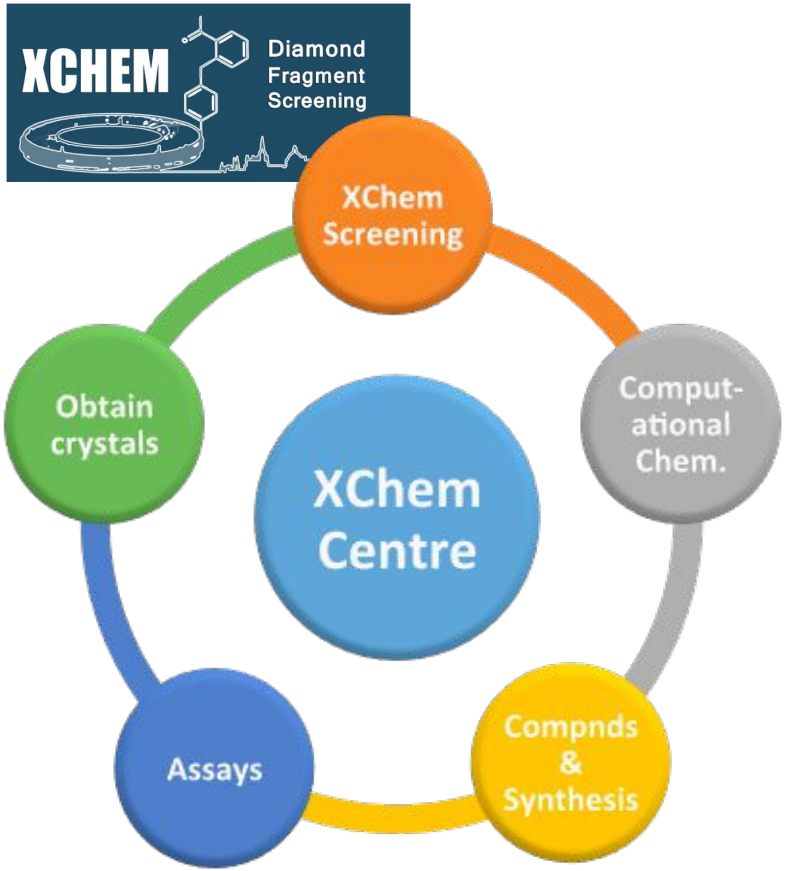


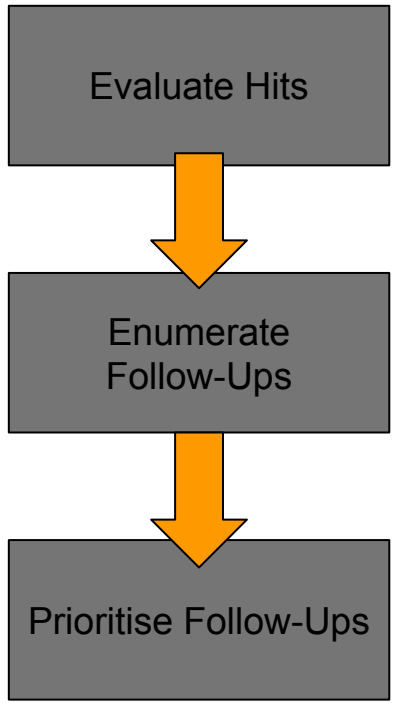


ARE THESE INTERESTING???

Need for rapid, cheap and accesible chemistry - and comp-chem

Pipeline Needed





WONKA - pharmacophores

Detect Vectors (using SMARTS)

Reaction SMARTS

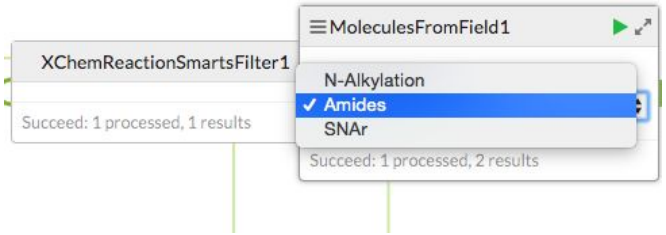
Reaction Vectors (Garrett Morris)

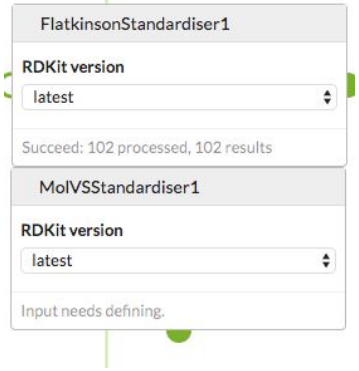
Sanitisation (MolVS and Flatkinson)

Lipinski / Rule of 3 filters

Conformer generation

Pipeline Needed

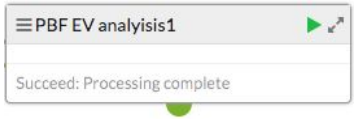




Evaluate Hits

WONKA - pharmacophores

Detect Vectors (using SMARTS)



Enumerate Follow-Ups

Reaction SMARTS

Reaction Vectors (Garrett Morris)

Prioritise Follow-Ups

Sanitisation (MolVS and Flatkinson)

Lipinski / Rule of 3 filters

Conformer generation

A solution

XCHEM Diamond Fragment Screening

Computer Nerd Land

No need to write installer

Can share best practice

Don’t worry about running

Biology Nerd Land

No need to install

Default values can be inserted

Instant HPC support

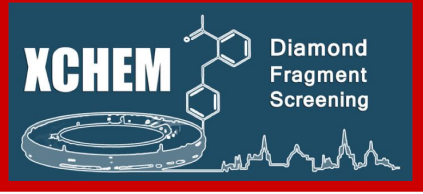
RDKit

Open-Source Cheminformatics and Machine Learning

STRUBI

CCP-CMC

CCP CMC



Purpose:

1. Naive user access to comp-chem best-practice and tools
2. Academic route-to-market for tools and technologies
3. Novel method development that speaks to the needs of pharma

Headline Outcomes:

1. Two workshops held at Diamond and Cambridge and attended by over 20 people
2. Contributor to open-source community surrounding SQUONK platform. Including introduction of third party tools and commitment to continue to do so
3. Third meeting being organised for October and training workshop to be held this year

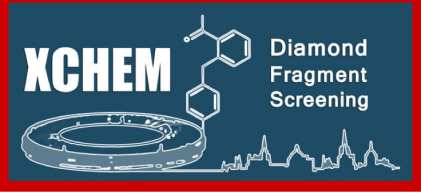
CCP-CMC

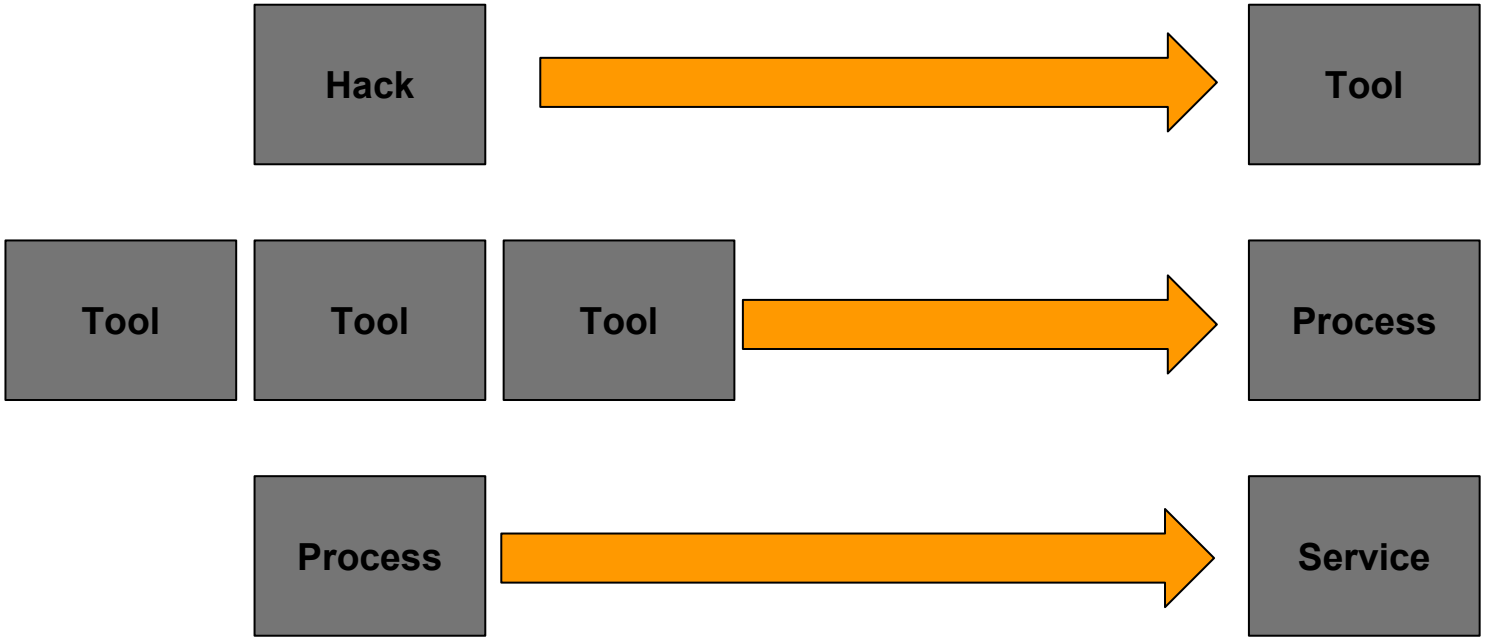
	Providers	Users
Academic	Simple route to broader application	Access to easy-to-use tools and workflows
Industry	Make tools easier to get into the hands of consumers	Access to pre-competitive tools and infrastructure

CCP-CMC

	Providers	Users
Academic	Simple route to broader application	Access to easy-to-use tools and workflows
Industry	Make tools easier to get into the hands of consumers	Access to pre-competitive tools and infrastructure

CCP-CMC





E.g. PBF EV analysis **(pipelines project)**

E.g. Docking workflows

E.g. Make a REST service

Pipelines project

Simple Pipeable Tools

https://github.com/InformaticsMatters/pipelines

Apache 2.0 license

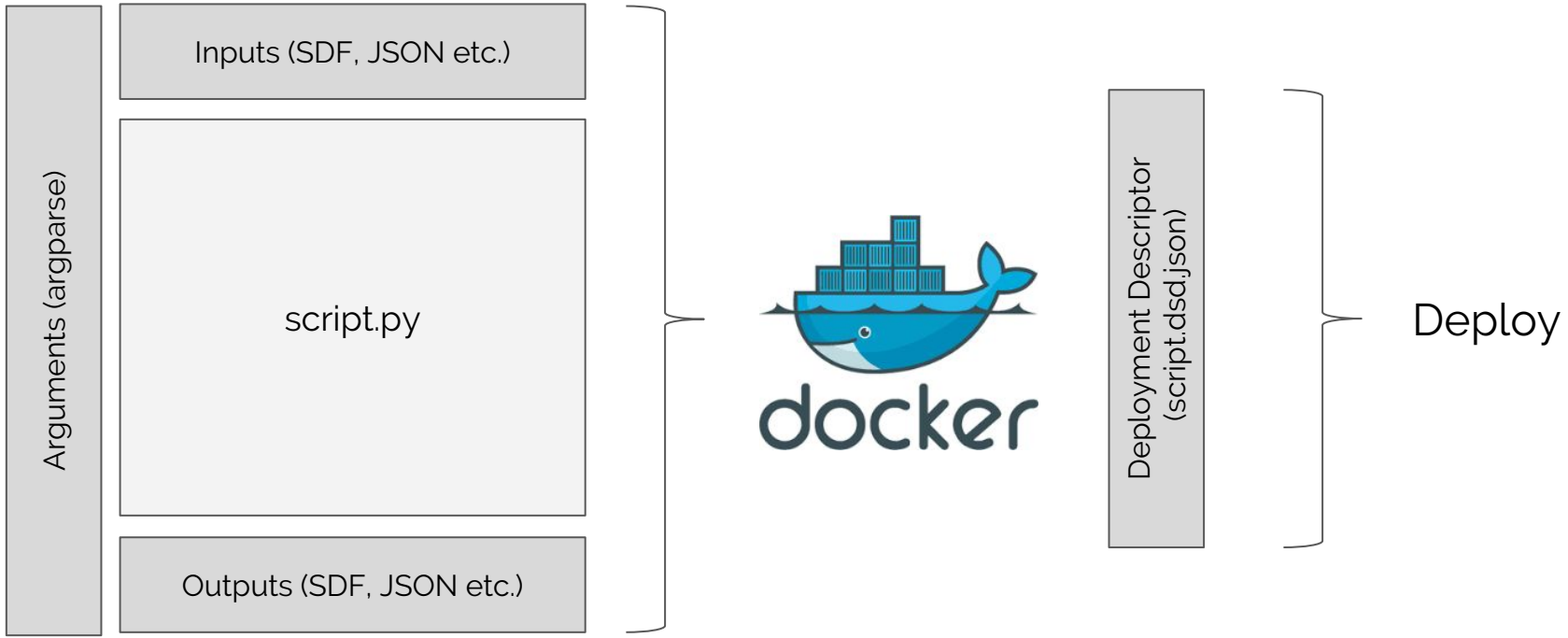
Follow the Unix pipes principle:
- do one thing and do it well
- output of one process becomes input of next process

Typically 100 - 200 line programs. Many are based on Python + RDKit, many come straight from the RDKit Cookbook.

Examples:

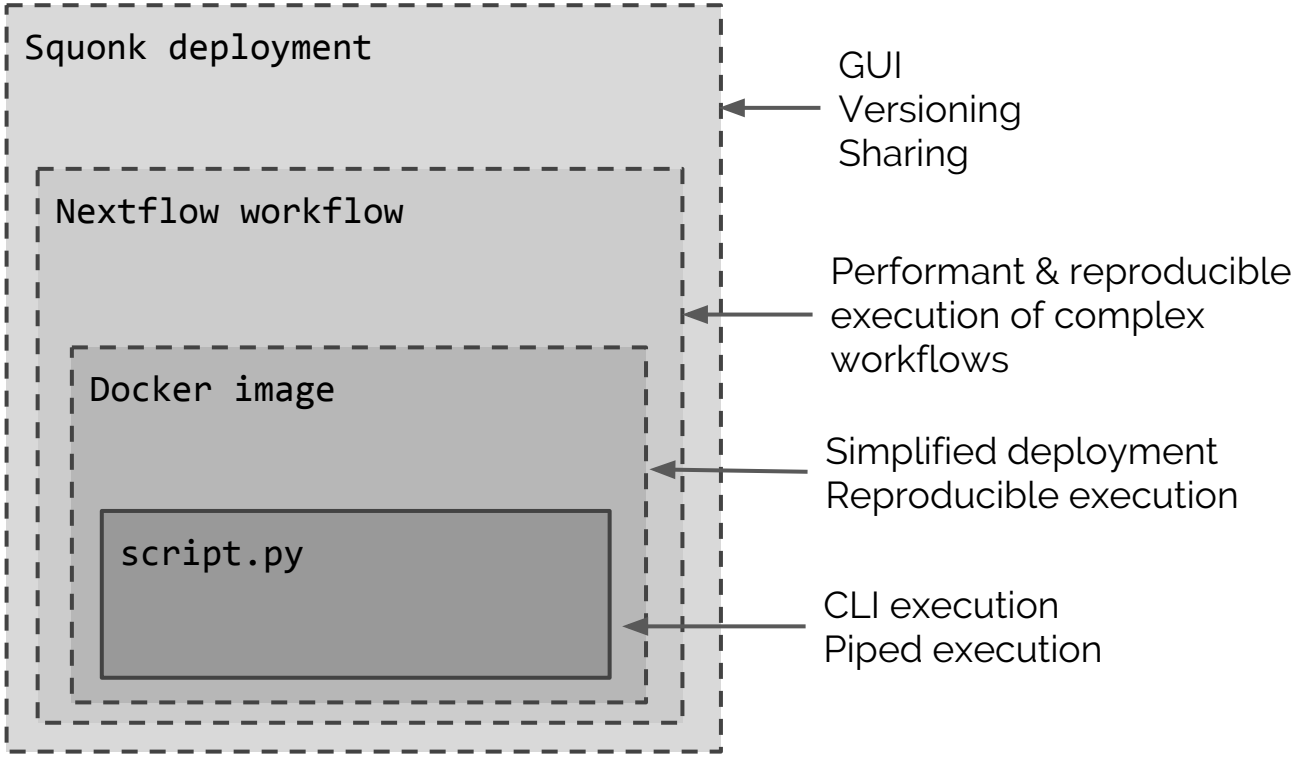
cluster_butina.py	splitter.py	sanifier.py
screen.py	filter.py	constrained_conf_gen.py
o3DAlign.py	conformers.py	rxn_maker.py

Built Upon a Common Pattern

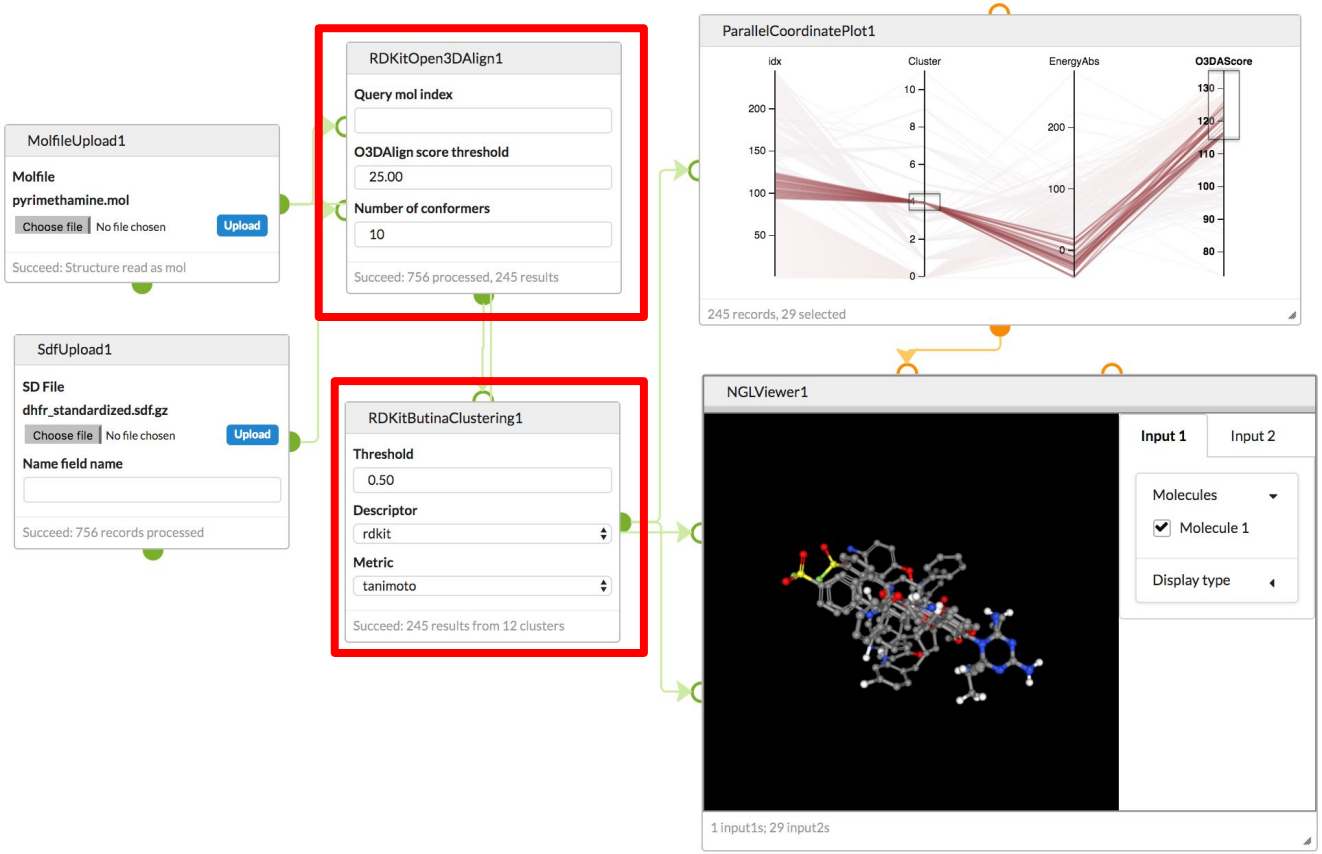


https://github.com/InformaticsMatters/pipelines/tree/master/src/python/pipelines/rdkit

Layered Approach Facilitates Reuse



Execution in Squonk



Docking Validation project

Docking Validation

https://github.com/InformaticsMatters/docking-validation

Apache 2.0 license

Similar principles to Pipelines but focussed on target based virtual screening

- Aim to incorporate a range of docking and scoring functions
- “Simple” Docking (rDock, VINA/SMINA, PLANTS, ...)
- MD techniques (MM-{P,G}BSA, DuCK, FEP, ...)
- Focus on validation, benchmarking and establishing best practice

Interest in processes for standardizing molecules such as preparation for docking

Conclusions

Conclusions

- XChem and SQUONK working together to build and incorporate open-source tools and frameworks for FBDD
- Open-source infrastructure for incorporating your own tools
- Truly trivial to do - and work not just applicable to SQUONK
- Hackathon Ideas ->
    - 1) Incorporate your own tool in SQUONK (talk to me or Tim)
    - 2) Generic frameworks to incorporate tools (into any framework perhaps)

https://squonk.it/