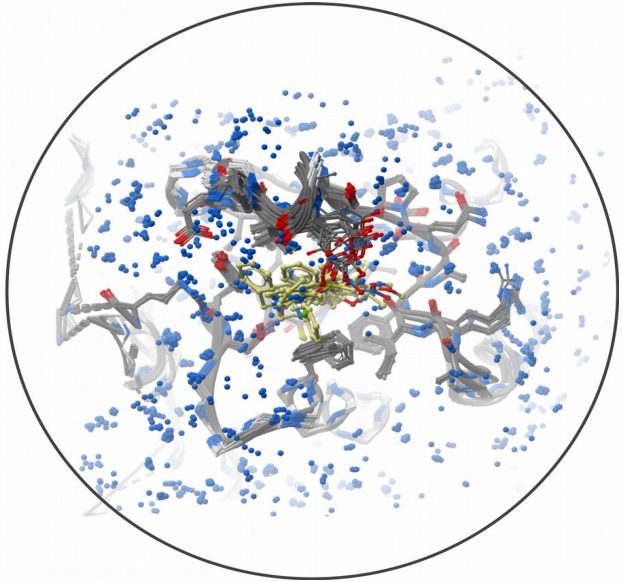
Fragment Network Implementation

Anthony Bradley, Frank van Delft
Diamond Light Source

Tim Dudgeon, Alan Christie
Informatics Matters Ltd.

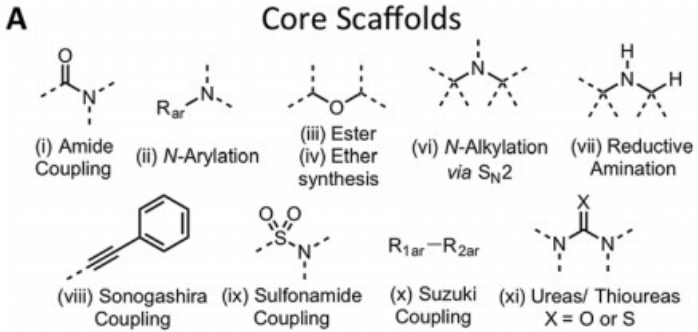
What do we do with this mess?

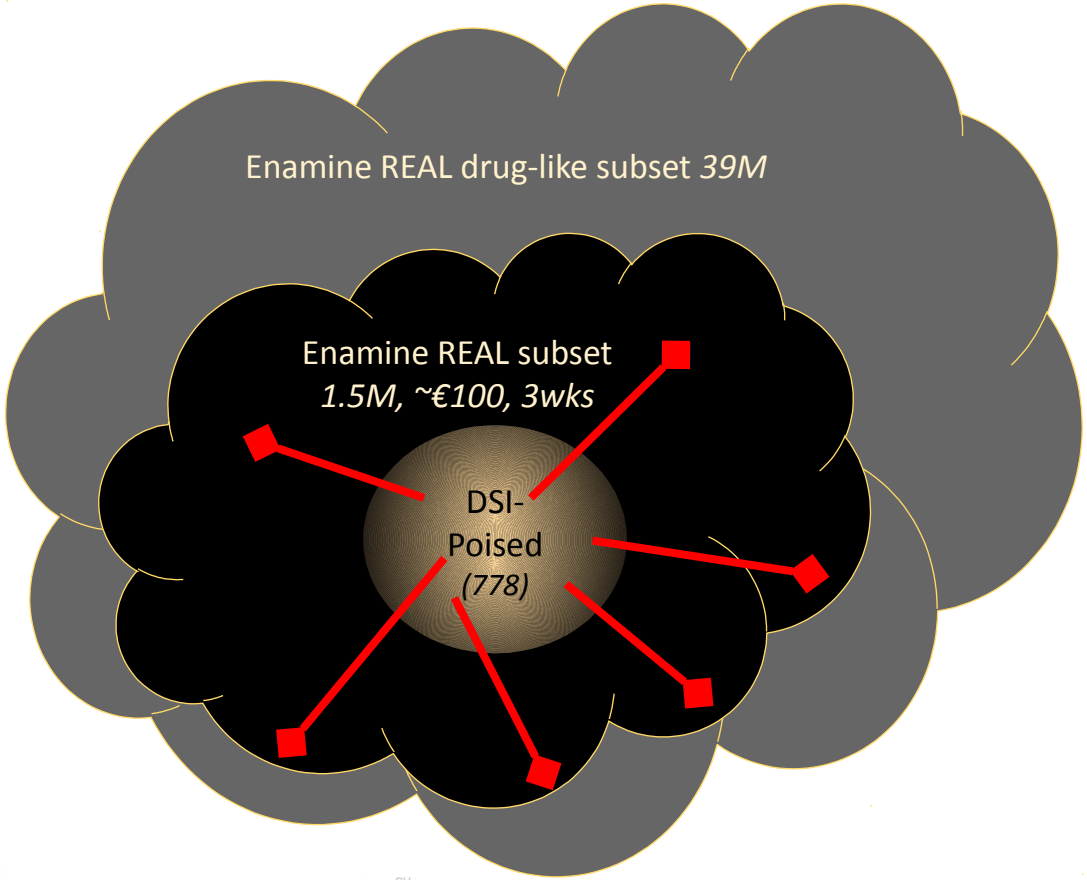


- What *can* we make – FAST AND CHEAPLY?
- What *should* we make?
- Will it look like the drug we want? (Penetration, soluble, …)
- How can we use ALL the information? OBSERVED *AND* PRIORS.

Compound logistics: quick, cheap(ish) follow-up

- DSI poised: now delivered by Enamine
- Reselected original poised library to align with Enamines's REAL dataset





Journal of Medicinal Chemistry

Article

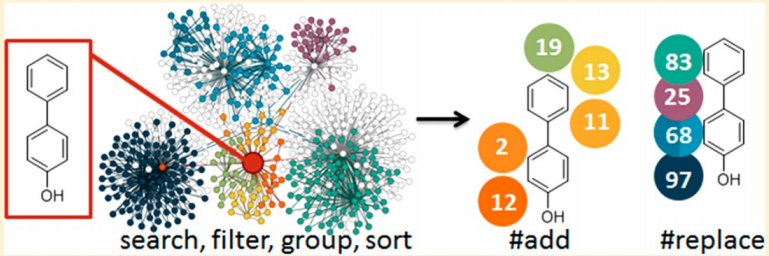
pubs.acs.org/jmc

The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database

Richard J. Hall,* Christopher W. Murray, and Marcel L. Verdonk

Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom

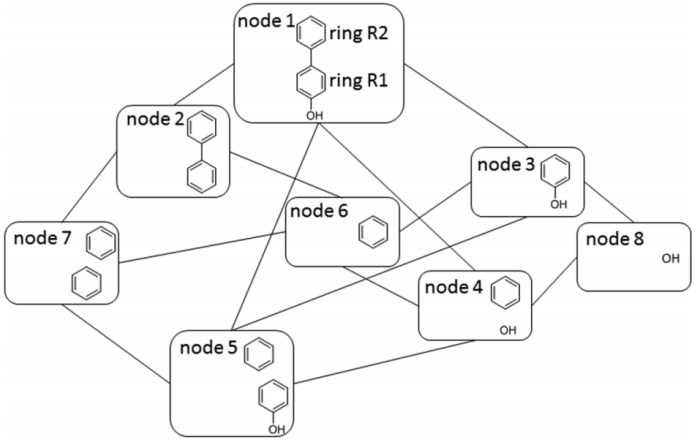
Supporting Information



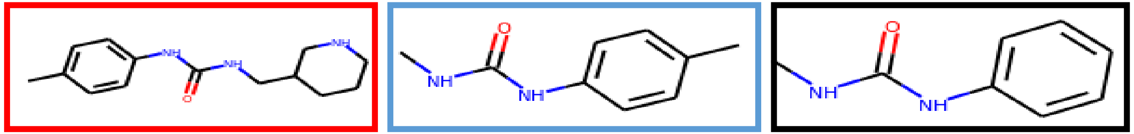
ABSTRACT: The hit validation stage of a fragment-based drug discovery campaign involves probing the SAR around one or more fragment hits. This often requires a search for similar compounds in a corporate collection or from commercial suppliers. The Fragment Network is a graph database that allows a user to efficiently search chemical space around a compound of interest. The result set is chemically intuitive, naturally grouped by substitution pattern and meaningfully sorted according to the number of observations of each transformation in medicinal chemistry databases. This paper describes the algorithms used to construct and search the Fragment Network and provides examples of how it may be used in a drug discovery context.

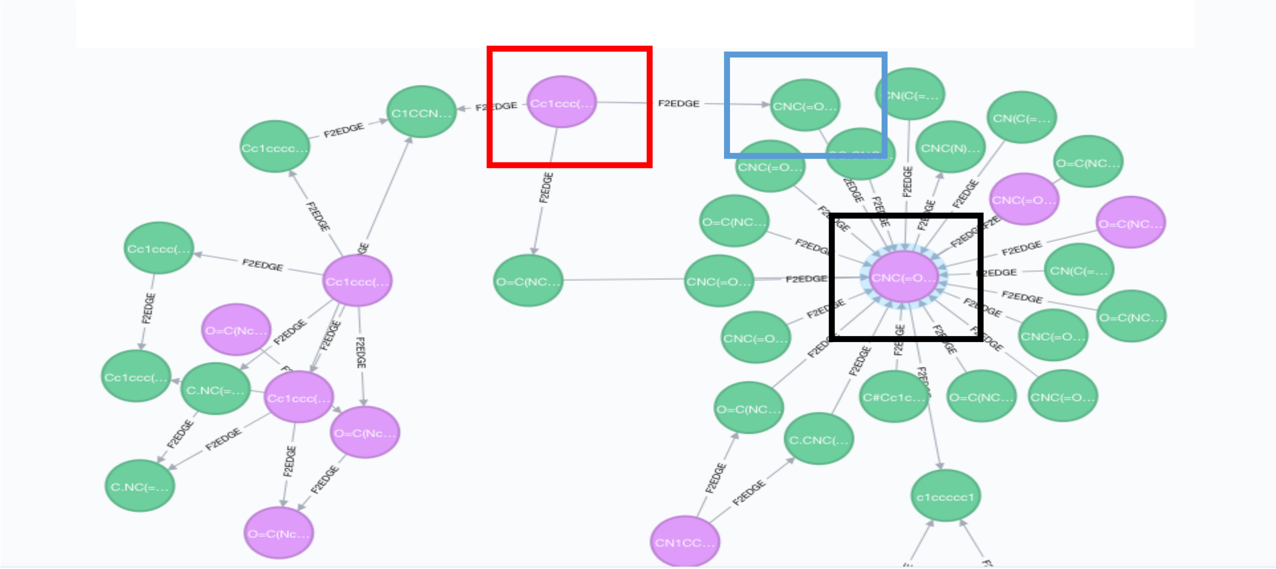
DOI: 10.1021/acs.jmedchem.7b00809

- Offers Intuitive method for navigating chemical space
- Very scalable system (ms queries)
- Allows easy analysis
  - Follow-ups available (Enamine space of Millions compounds)
  - Vectors not explored (rapid analysis)
  - Path between molecules



How does it work?





Molecules processed and fragmented using RDKit/Python

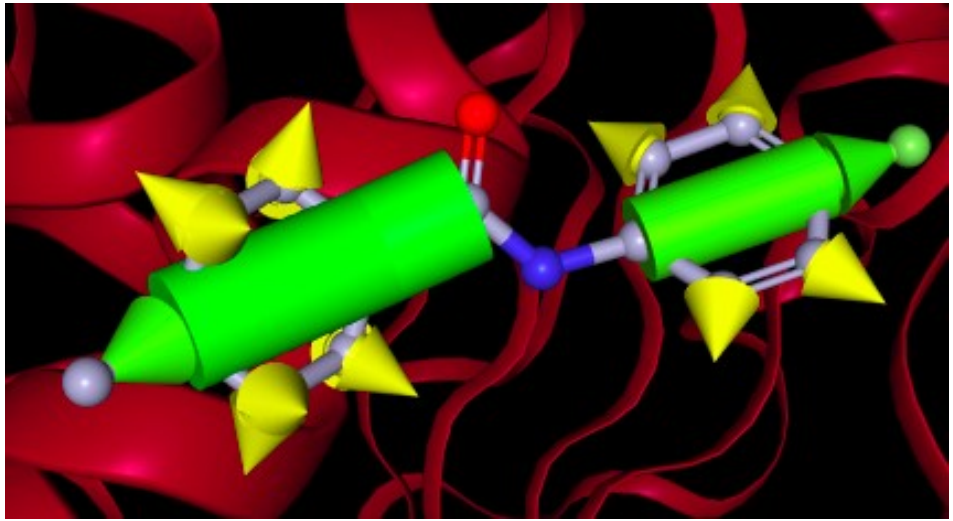
Nodes and edges loaded into Neo4J graph database

Fragment networks built for Enamine REAL datasets:

1.5M (5.4M nodes, 24M edges)

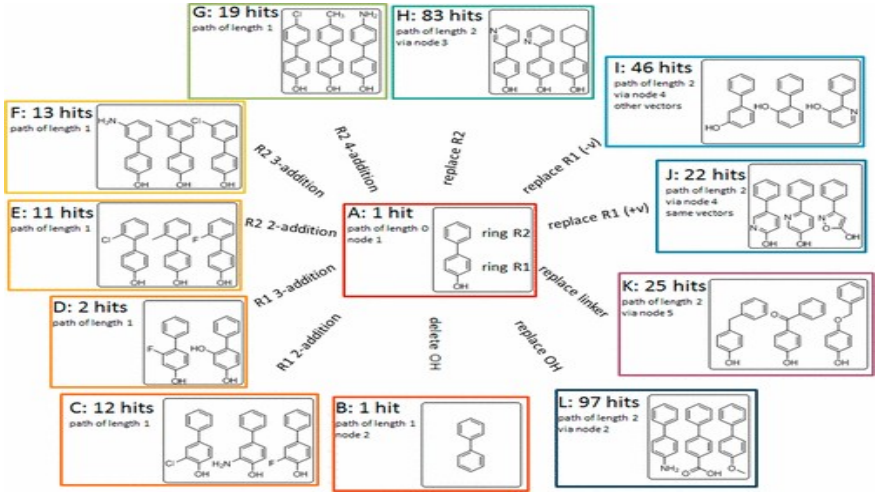
39M (200M nodes, 1,175M edges)

Exploration in action



Web based UI provided using React and NGL viewer

Allows to explore vectors where fragments can be expanded



Current Status

Molecules processed and fragmented using RDKit/Python.
Workflow parallelised on compute cluster.

Fragment networks built for 1.5M and 39M Enamine REAL datasets.
Methodology being improved and extended to other datasets.

Nodes and edges loaded into Neo4J graph database.

Web based UI provided using React and NGL viewer, Django middle tier

Going alpha any day now!
(test) https://fragalysis.apps.xchem.diamond.ac.uk/viewer/react/landing
(alpha) https://fragalysis.diamond.ac.uk/viewer/react/landing

What interest is there in this outside of Diamond?

The Code

https://github.com/xchem/fragalysis

https://github.com/xchem/fragalysis-backend
https://github.com/xchem/fragalysis-frontend
https://github.com/xchem/fragalysis-stack
https://github.com/InformaticsMatters/dls-fragalysis-stack-openshift

All under Apache 2.0 or MIT license