Algorithms for progressing crystallographic fragment hits at XChem

Fragment merging using the Fragment Network

AViDD Open Science Forum

Stephanie Wills

DPhil student in the Oxford Protein Informatics Group (OPIG) and Centre for Medicines Discovery (CMD)

University of Oxford

Supervised by Dr Ruben Sanchez-Garcia, Prof Frank von Delft and Prof Charlotte M. Deane

Overview

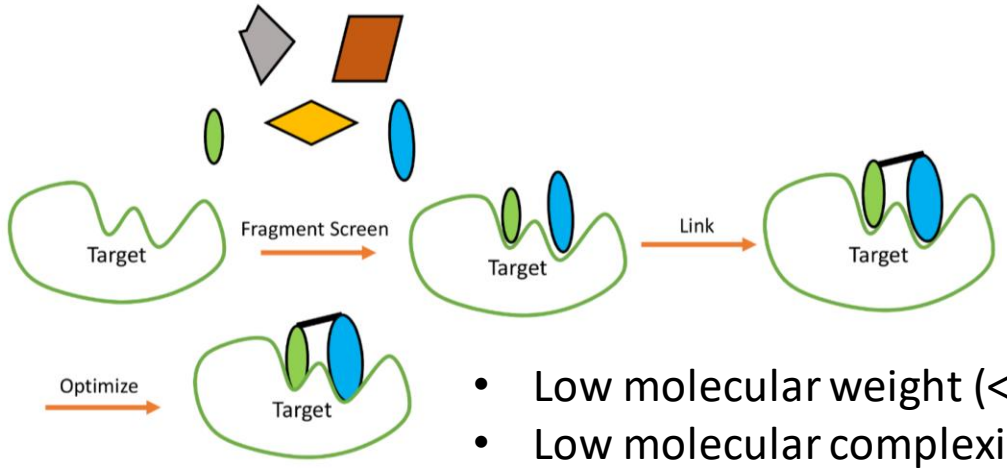
- **Background**
  - Fragment-based drug discovery
  - Fragment-to-lead progression
  - Catalogue search
- **Published work**
  - Fragment merging using the Fragment Network
- **Current work**
  - Expanding the chemical space used to find merges
- **Deployment of algorithms for active campaigns**

Background

Fragment-based drug discovery

Advantages

- Higher hit rates
- Smaller library sizes
- Cover a greater proportion of chemical space
- Greater **control over further synthesis and property optimization**
  - Step-by-step iterative process



- Low molecular weight (<300 Da)
- Low molecular complexity
- High µM to low nM affinity

Once we have fragment hits, what can we do with them?

Image from Emwas, A-H et al. *Molecules* **2020**, 25(20), 4597; https://doi.org/10.3390/molecules25204597.

Methods for elaborating fragment hits

- How can we use data from fragment screens to identify compounds that **recapitulate atoms seen in the fragments**?
- Structure-guided optimization
  - Increase **potency**
  - Maximize number of **interactions**
  - Elucidate **structure–activity relationship** (SAR)
- Three main approaches to fragment elaboration

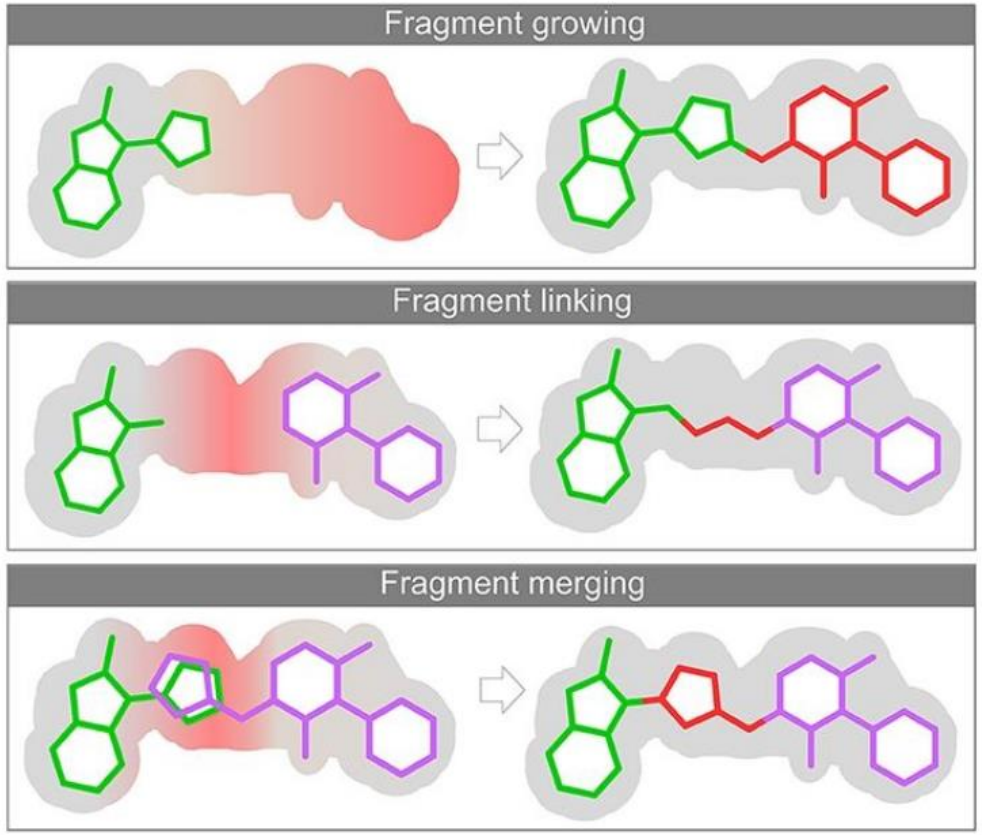
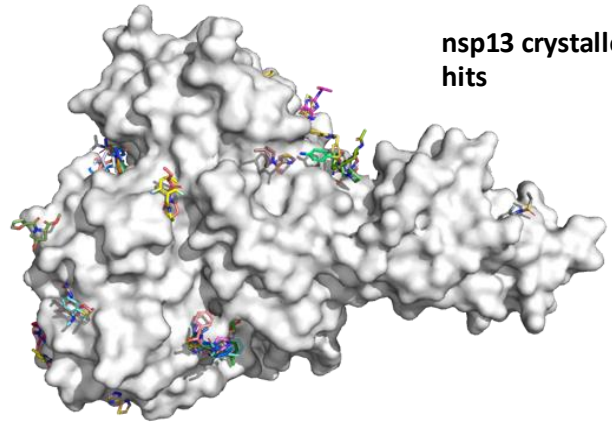


Image from De Souza Neta, L.R. et al. *Front. Chem.* **2020**, 8:93; https://doi.org/10.3389/fchem.2020.00093.

Fragment screening using X-ray crystallography

- **Biophysical techniques** typically used to detect weakly binding fragments
- Fragment screening with **X-ray crystallography** allows:
  - Confirmation of **binding pose**
  - Maps **possible interactions**
  - **Structure-guided** optimization



nsp13 crystallographic fragment hits

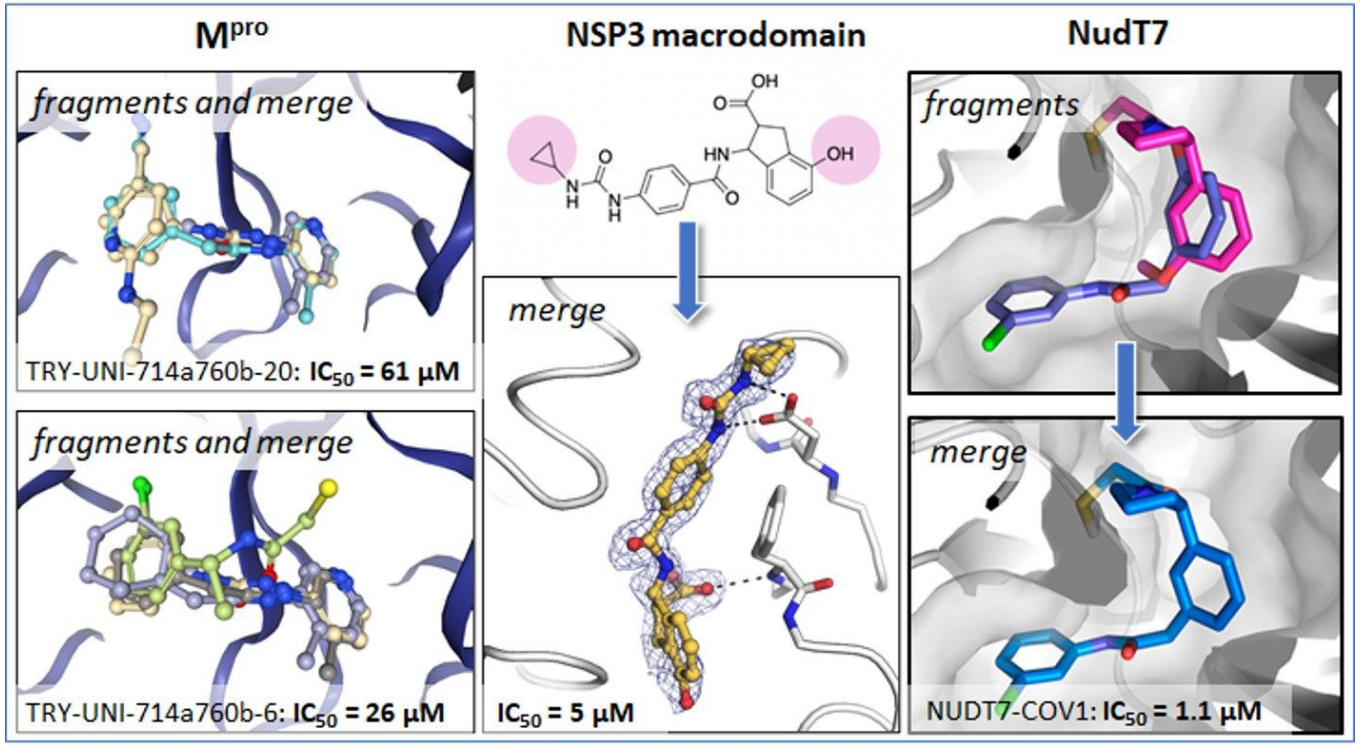
XCHEM Diamond Fragment Screening

diamond

Newman, J.A. et al. *Nat. Commun.* **2020**, 12:4848; https://doi.org/10.1038/s41467-021-25166-6.
https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html

Fragment merging is highly efficient for increasing potency

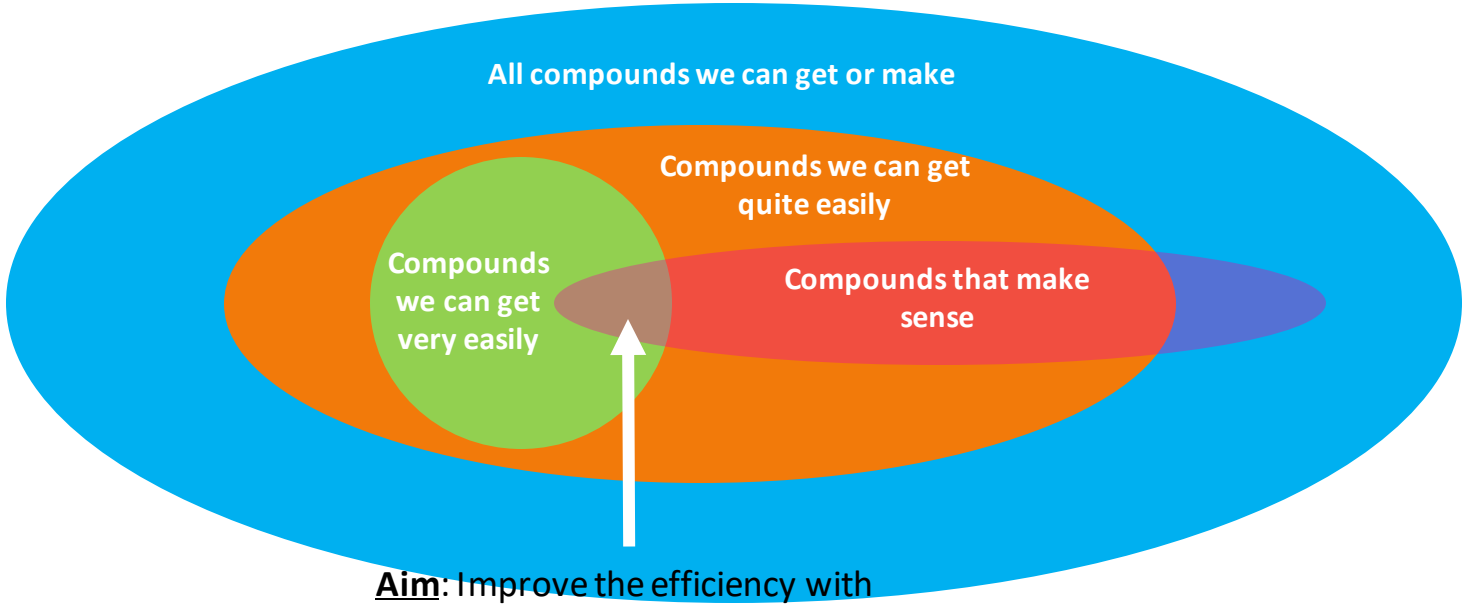
- Fragment merging and linking are **efficient approaches for increasing potency**
- Merging can lead directly to **on-scale affinity**



The COVID Moonshot Consortium. ChemRxiv **2020**; https://doi.org/10.26434/chemrxiv-2021-585ks-v2. Resnick, E. et al. *J. Am. Chem. Soc.* **2019**, 141, 22, 8951-8968; https://doi.org/10.1021/jacs.9b02822. Gahbauer, S. et al. *PNAS* **2023**, 120, 2, e2212931120; https://doi.org/10.1073/pnas.2212931120.

What are the existing approaches for identifying follow-up compounds?

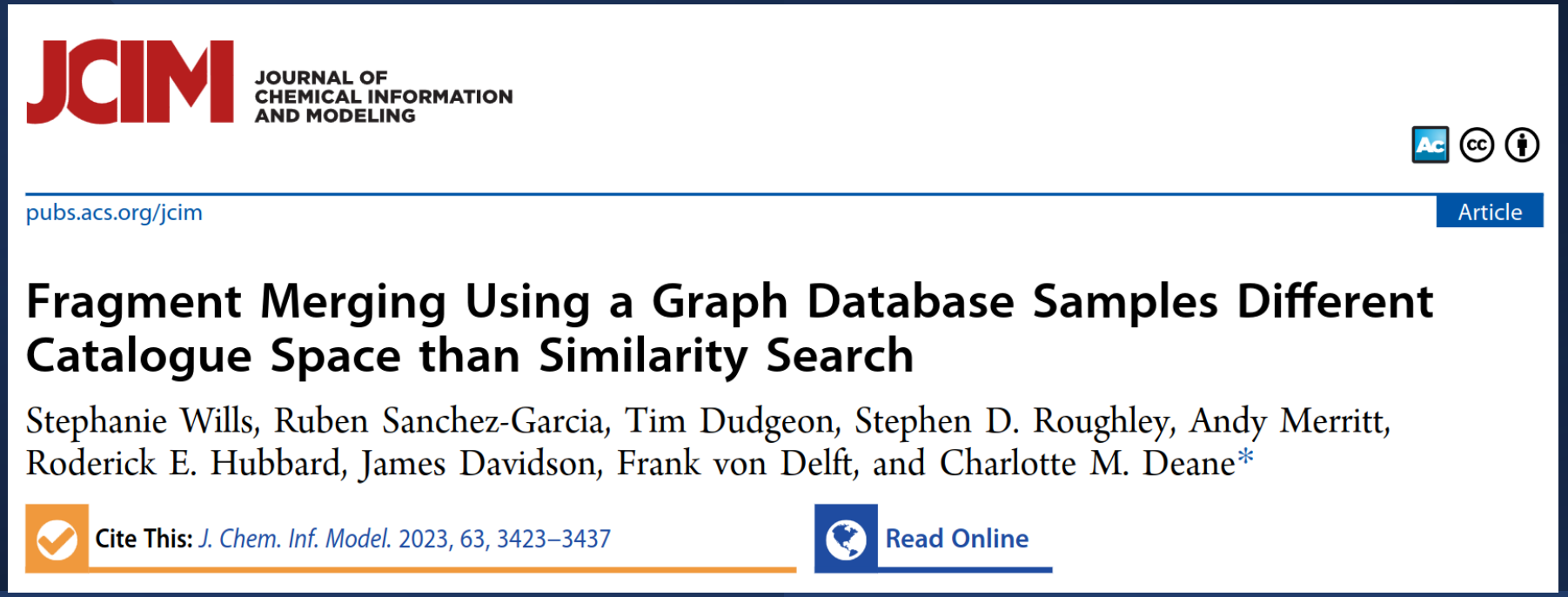
- Existing approaches
  - Manual design
    - Difficult to **scale-up**
  - *De novo* design
    - Limited by **synthetic accessibility**
  - Catalogue search
    - Similarity and substructure search
    - **Rapid and cheap i**dentification of follow-up compounds
    - Lacks formalized workflows



Aim: Improve the efficiency with which we sample this area of accessible chemical space

Fragment merging using the Fragment Network

Pipeline v1



JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

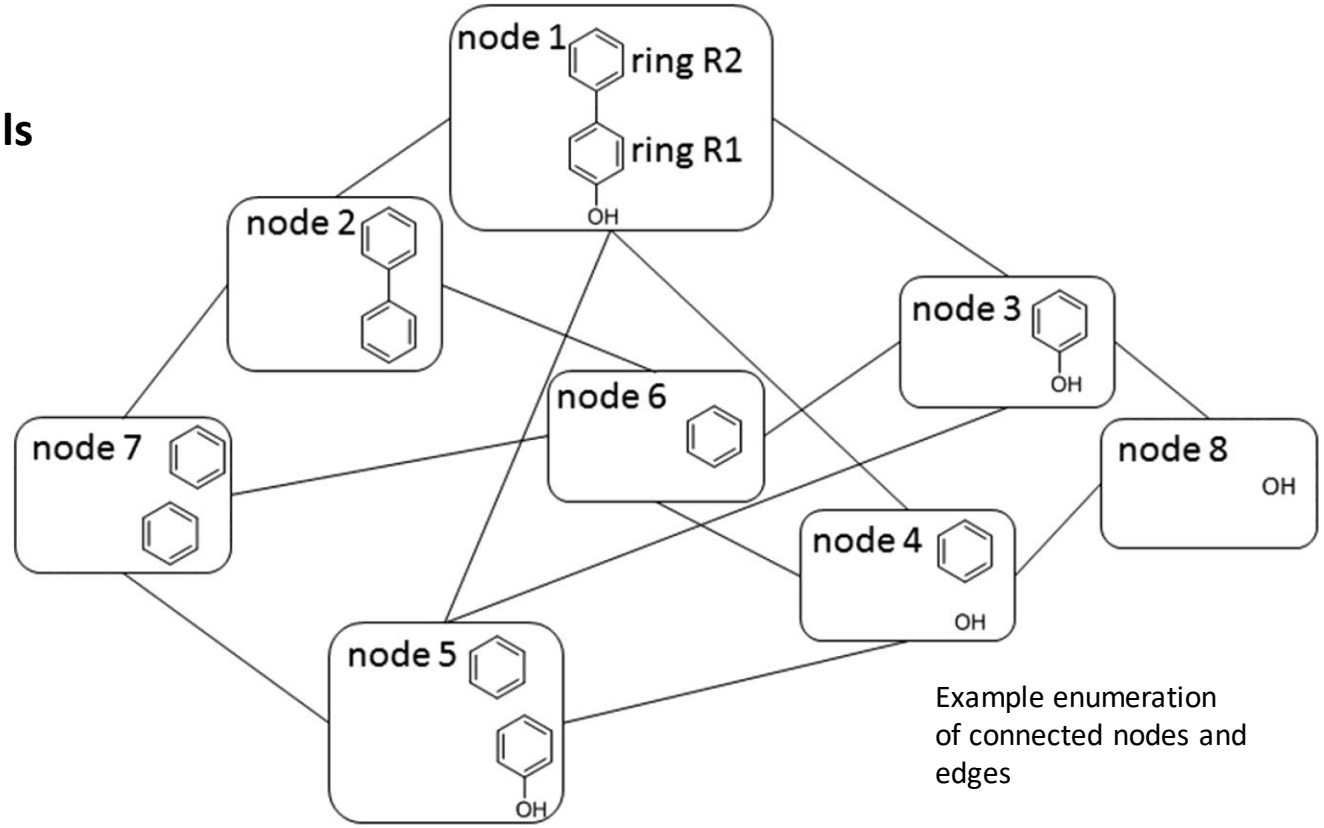
Fragment Merging Using a Graph Database Samples Different Catalogue Space than Similarity Search

Stephanie Wills, Ruben Sanchez-Garcia, Tim Dudgeon, Stephen D. Roughley, Andy Merritt, Roderick E. Hubbard, James Davidson, Frank von Delft, and Charlotte M. Deane*

Cite This: *J. Chem. Inf. Model.* 2023, 63, 3423–3437 Read Online

The Fragment Network

- First described in 2017 by **Astex Pharmaceuticals**
- Architecture
  - Nodes represent molecules
  - Edges represent transformations
  - More **chemically intuitive**
- Populated with **commercial catalogues**
- **>120M compounds** at time of publishing
  - Current version contains >200M compounds

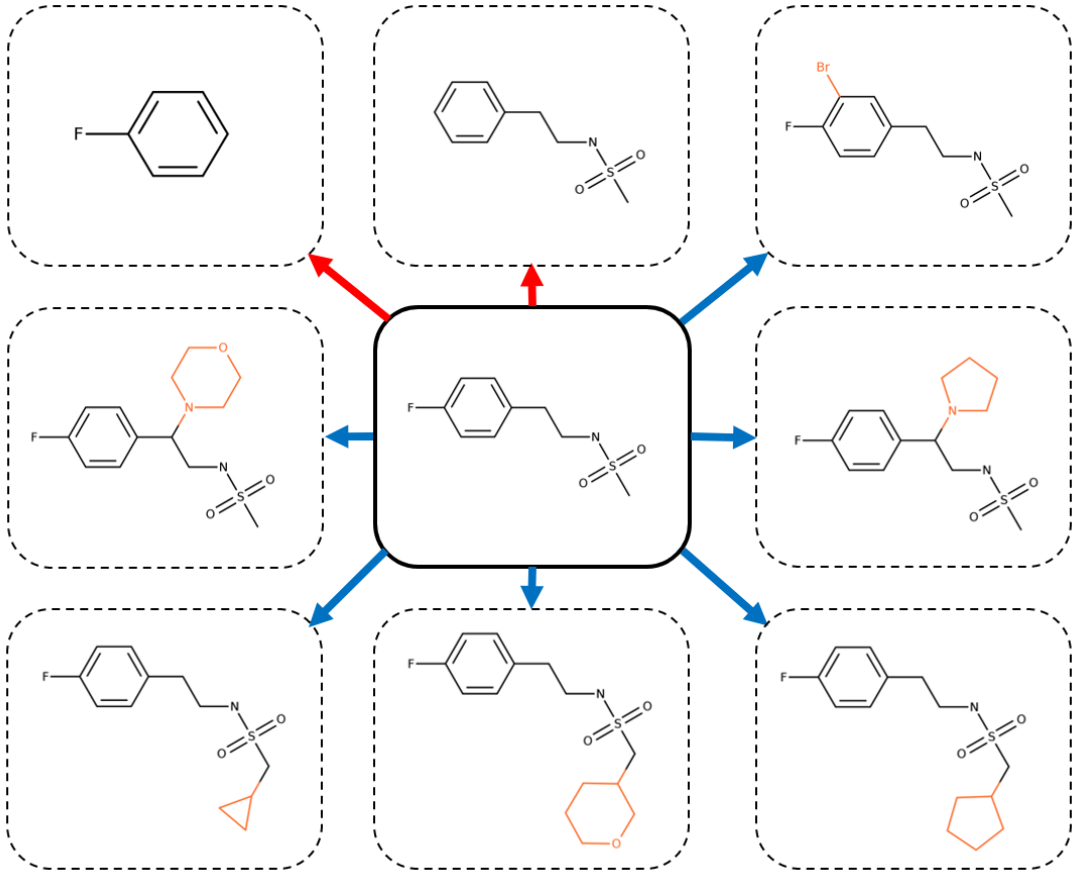


Example enumeration of connected nodes and edges

Image and Fragment Network from Hall, R.J. et al. *J. Med. Chem.* **2017**, 60, 14, 6440-6450; https://doi.org/10.1021/acs.jmedchem.7b00809.

The Fragment Network

- First described in 2017 by **Astex Pharmaceuticals**
- Architecture
  - Nodes represent molecules
  - Edges represent transformations
  - More **chemically intuitive**
- Populated with **commercial catalogues**
- **>120M compounds** at time of publishing
  - Current version contains >200M compounds



Example transformations for nsp13 hit fragments

Image and Fragment Network from Hall, R.J. et al. *J. Med. Chem.* **2017**, 60, 14, 6440-6450; https://doi.org/10.1021/acs.jmedchem.7b00809.

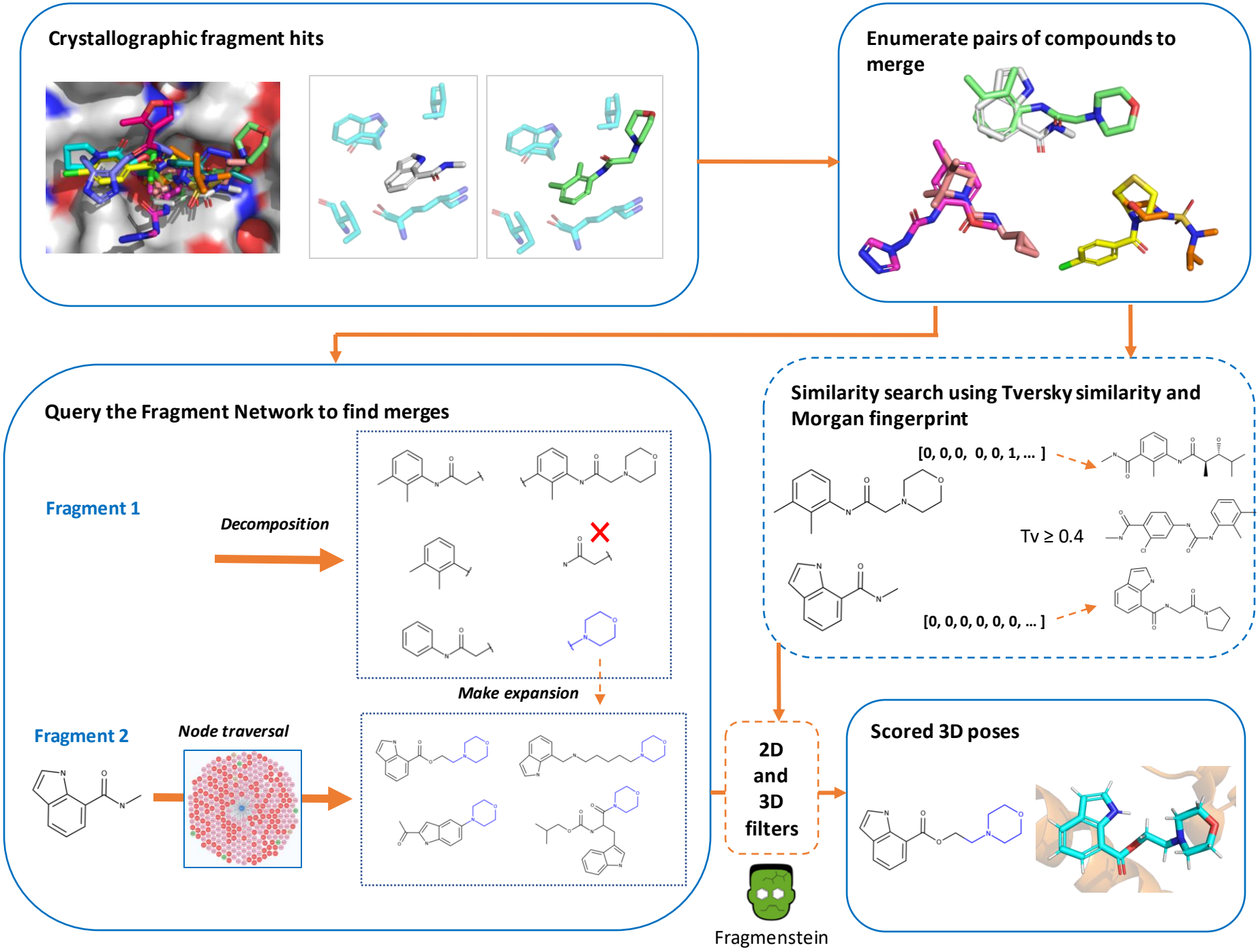
Aims

- Can the Fragment Network be used to find **purchasable fragment merges**?
- Can we find merges that **maintain substructures** of the crystallographic fragment hits?
- Can we find merges that **recapitulate the binding pose and interactions** of the parent fragments?
- Do merges identified by the Fragment Network have any use beyond those identified using a more **traditional similarity search**?

Methodology

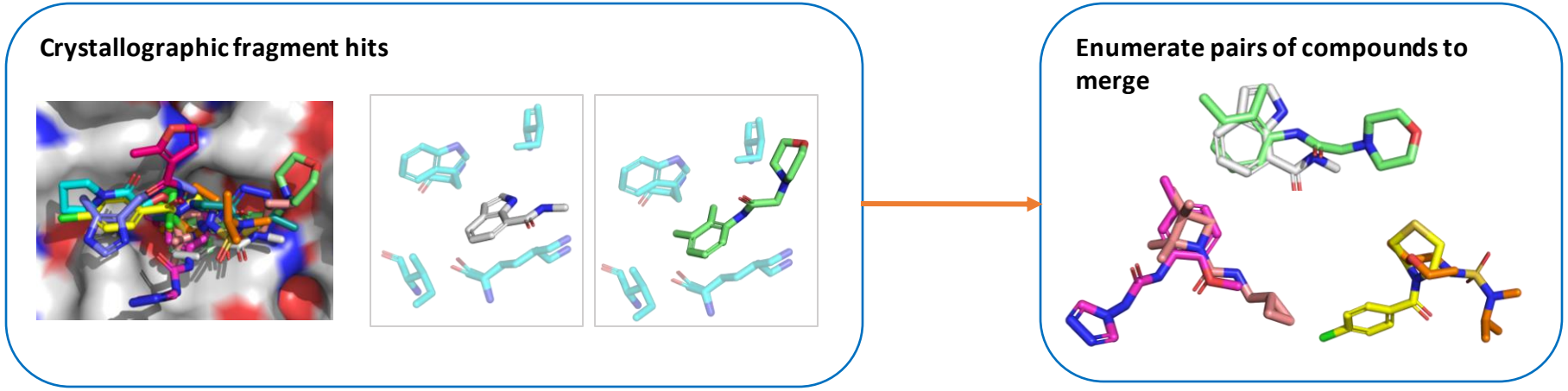
- We compared the ability of the **Fragment Network** and a more traditional, **fingerprint-based similarity search** to find fragment merges
- Four test systems (XChem targets) were used
  - SARS-CoV-2 main protease (**Mpro**)
  - SARS-CoV-2 helicase non-structural protein 13 (**nsp13**)
  - *P. gingivalis* dipeptidyl peptidase 11 (**DPP11**)
  - Human poly(ADP-ribose) polymerase 14 (**PARP14**)
- Retrospective analyses with existing experimental data was performed
  - Inhibitors of **Mpro** proposed by the COVID Moonshot project (COVID Moonshot Consortium, *ChemRxiv*, 2020)
  - Inhibitors of *M. tuberculosis* transcriptional repressor protein **EthR** (Nikiforov et al., *Org. Biomol. Chem.,* 2016)

The COVID Moonshot Consortium. ChemRxiv **2020**; https://doi.org/10.26434/chemrxiv-2021-585ks-v2.
Nikiforov, P.O. et al. *Org. Biomol. Chem.* **2016**, 14(7): 2318-2326; https://doi.org/10.1039/c5ob02630j.


Crystallographic fragment hits
Enumerate pairs of compounds to merge
Query the Fragment Network to find merges
Fragment 1
Decomposition
Make expansion
Fragment 2
Node traversal
Similarity search using Tversky similarity and Morgan fingerprint
[0, 0, 0, 0, 0, 1, ...]
Tv ≥ 0.4
[0, 0, 0, 0, 0, 0, ...]
2D and 3D filters
Fragmenstein
Scored 3D poses

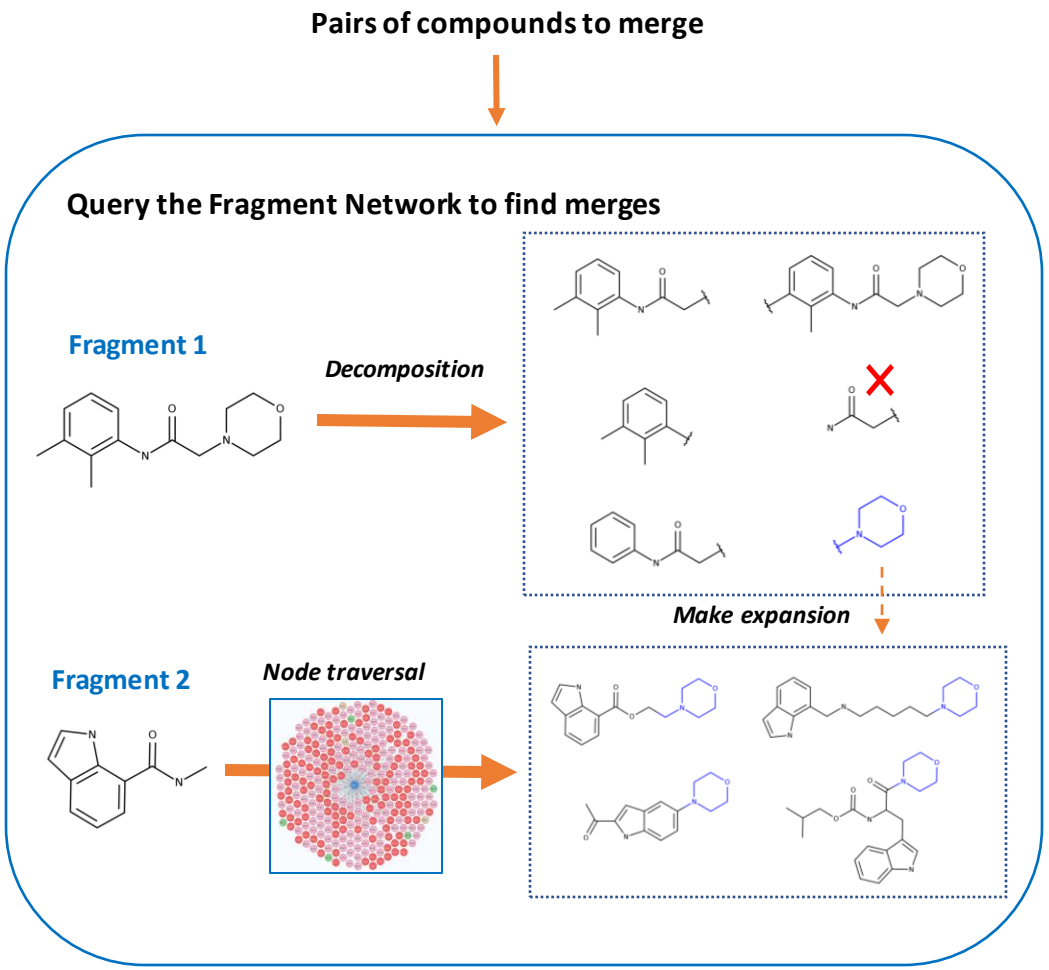
Pipeline: enumerating fragment pairs for merging

- Fragment screening data available to download from **Fragalysis**
- All **pairs of compounds enumerated** for querying the database
  - Remove pairs that are too **far apar**t or **similar**

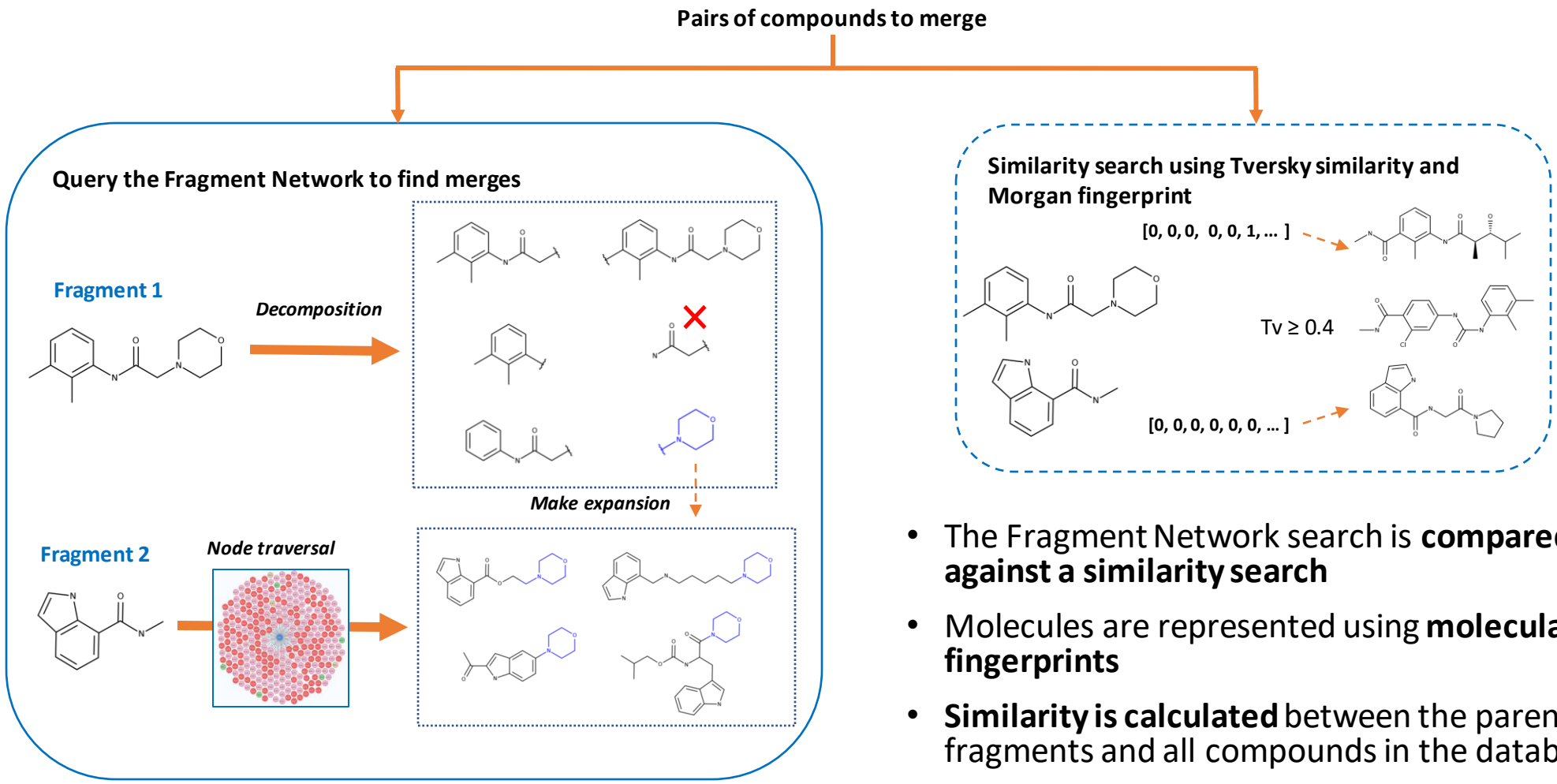


Pipeline: querying the database to find merges

- A number of **hops are made from the seed fragment** to generate diversity
- The second fragment is decomposed into **substructures**
- An expansion is made from the seed fragment in which a **substructure from the other fragment is incorporated**
- The query retrieves compounds that contain **substructures from both parent fragments**

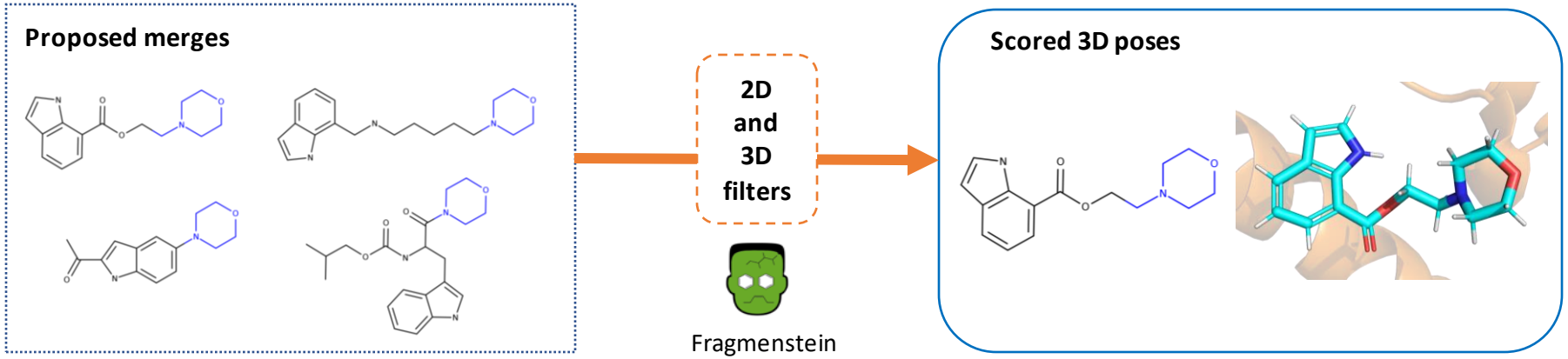


Pipeline: finding merges using a similarity search

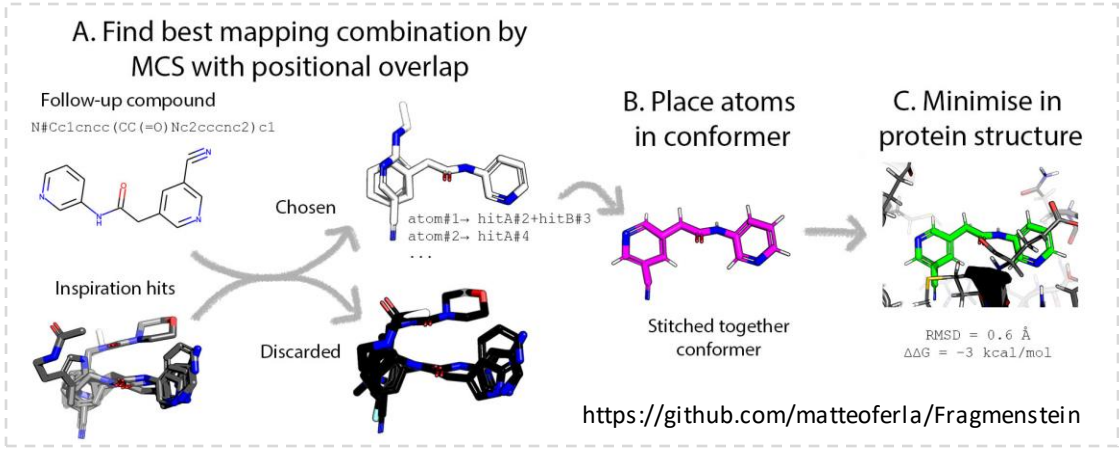


- The Fragment Network search is **compared against a similarity search**
- Molecules are represented using **molecular fingerprints**
- **Similarity is calculated** between the parent fragments and all compounds in the database

Pipeline: filtering proposed fragment merges



- **2D filters**
  - Cheaper filters (molecular descriptors, rule out of expansions)
- **3D filters**
  - More computationally intensive
  - Generate compounds based on the conformation of the parent fragments using **Fragmenstein**
    - Filter out those that don't **maintain binding pose**
    - Filter out those that don't **fit the binding pocket**
    - Filter out those with **energetically unfavourable** conformations

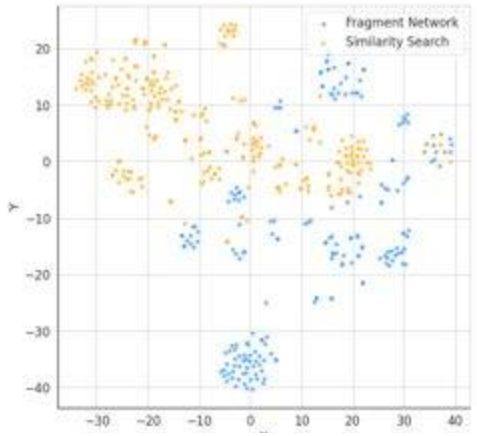


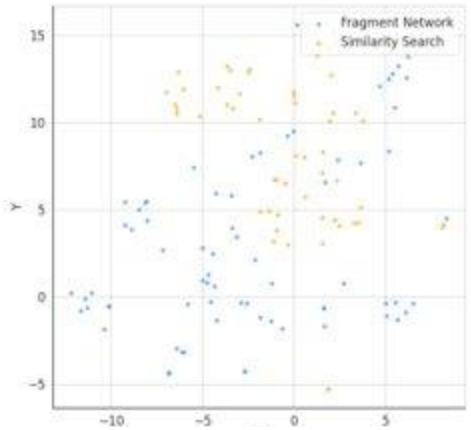
Results

The two search techniques show complementary results

- The search methods operate in **different areas of chemical space**

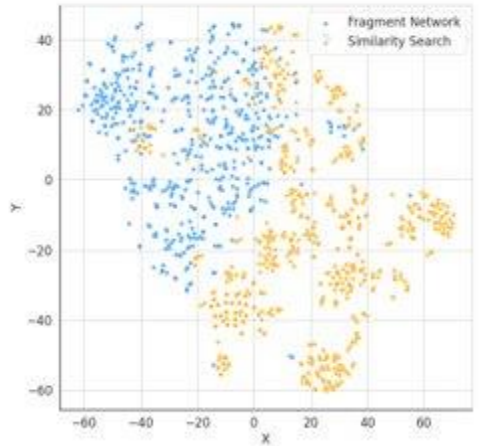
DPP11

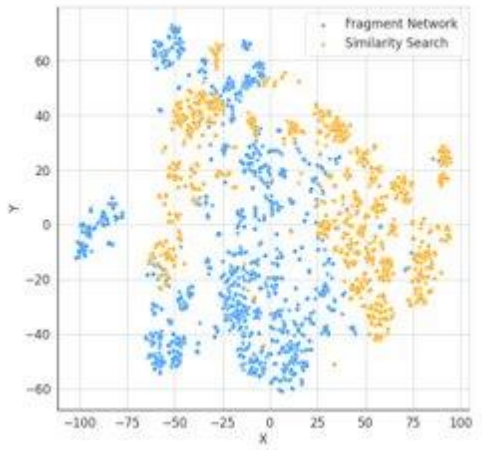




PARP14

nsp13





Mpro

	Number of hits	Search type[a]	Before filtering	After filtering	% filtered	Number of overlap
DPP11	11	FN	22,903	198	0.9	8
		SS	85,919	271	0.3	
PARP14	13	FN	48,320	70	0.1	0
		SS	78,116	56	0.1	
nsp13	9	FN	36,239	503	1.4	1
		SS	40,102	530	1.3	
Mpro	19	FN	109,012	952	0.9	4
		SS	169,424	918	0.5	

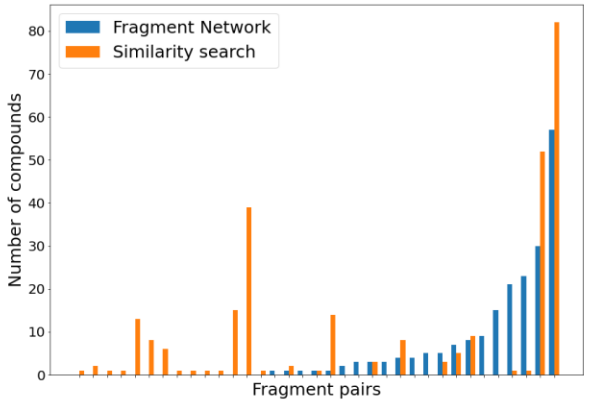
Low-dimensional representations of chemical space occupied by filtered compounds (using T-SNE plots)

Results

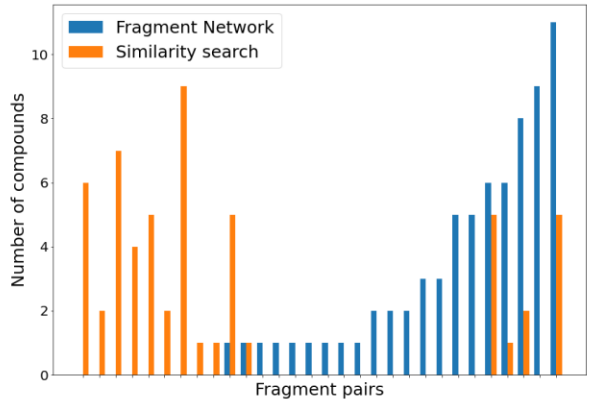
The two search techniques show complementary results

- The search methods identify **merges for fragment pairs not represented by the other**

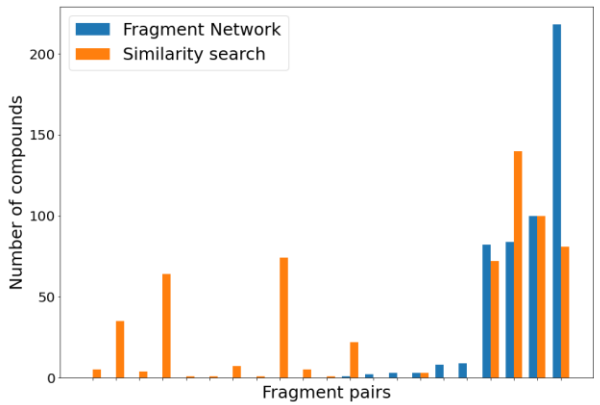
DPP11



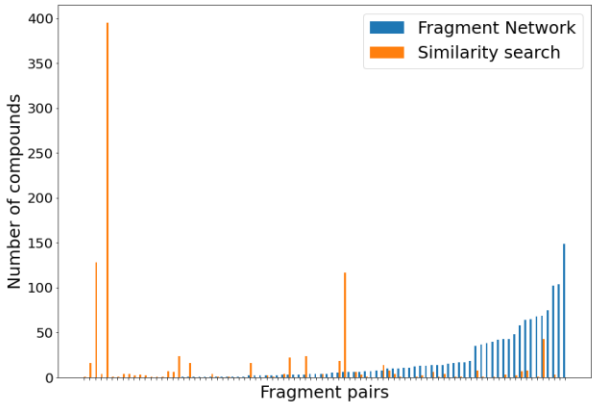
PARP14



nsp13



Mpro



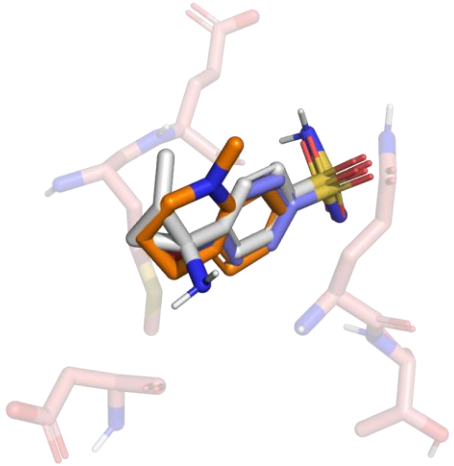
Numbers of filtered compounds identified for each pair of fragments

Results

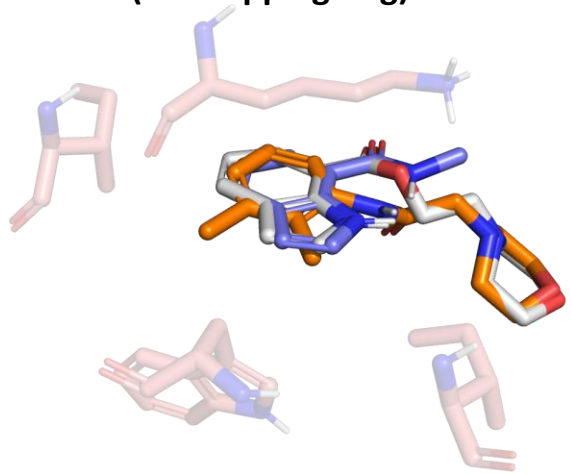
The two search techniques are suited to identifying different types of merge

- Merging opportunities **classified according to degree of overlap** between two fragments
- The two search techniques **differ in the types of merge they identify**
- The Fragment Network is more efficient at identifying types 3 & 4
  - **Hypothesis** for merging is **non-obvious**
  - Unclear which substructures should be used in the final merge

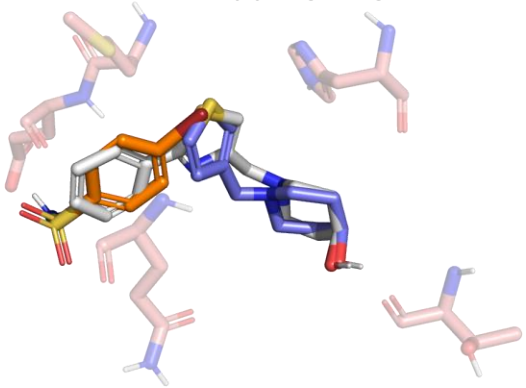
1. Completely overlapping merges



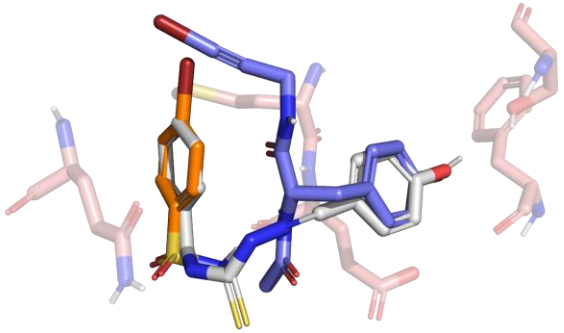
2 . Partially overlapping merges (overlapping ring)



3 . Partially overlapping merges (no overlapping ring)



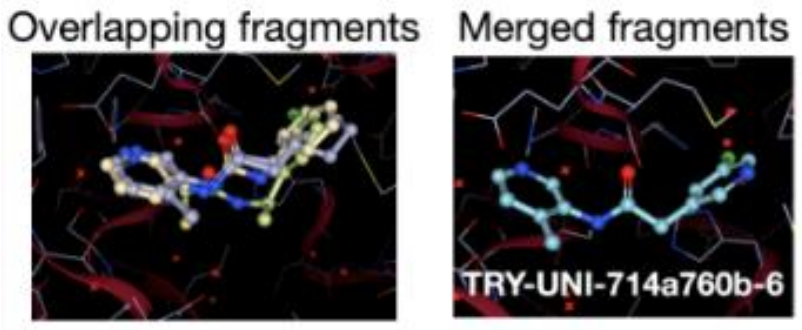
4 . Non-overlapping fragments

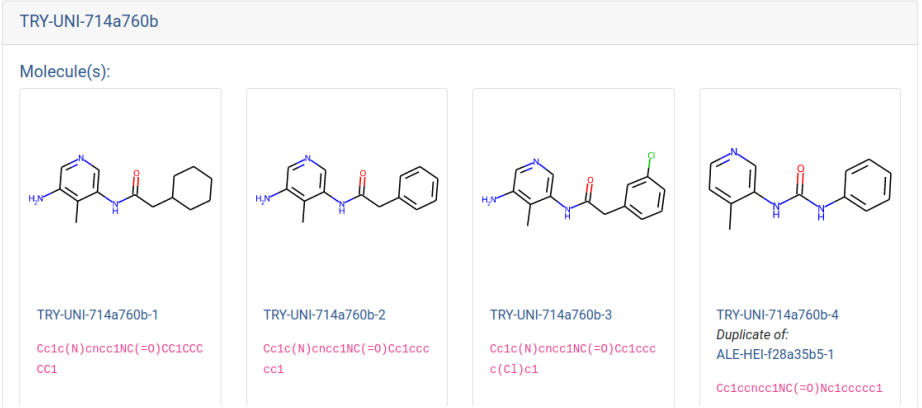


Results

The Fragment Network identifies known and potential binders against Mpro

- Literature examples of manually designed merges used as test cases
  - Crowd-sourced designs for **SARS-CoV-2 Mpro** from **COVID Moonshot project**
  - 5 fragments used as inspiration for manual design of 24 compounds (collective label: **TRY-UNI-714a760b**)
  - 8 compounds with IC_{50} values
- Fragment merging pipeline run on the five fragments and the results compared against compounds with experimental data



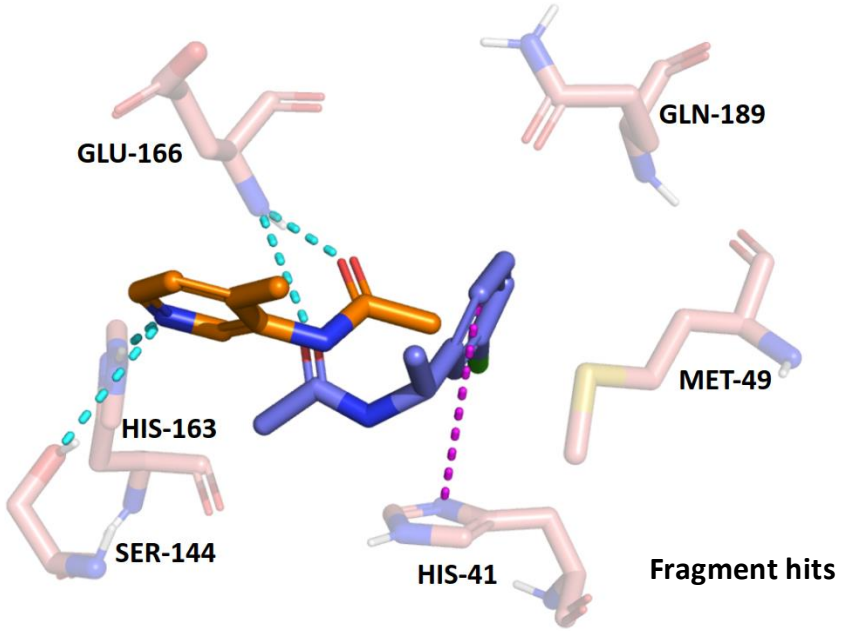


The COVID Moonshot Consortium. ChemRxiv **2020**; https://doi.org/10.26434/chemrxiv-2021-585ks-v2.
PostEra. Mpro activity data https://covid.postera.ai/covid/activity_data.

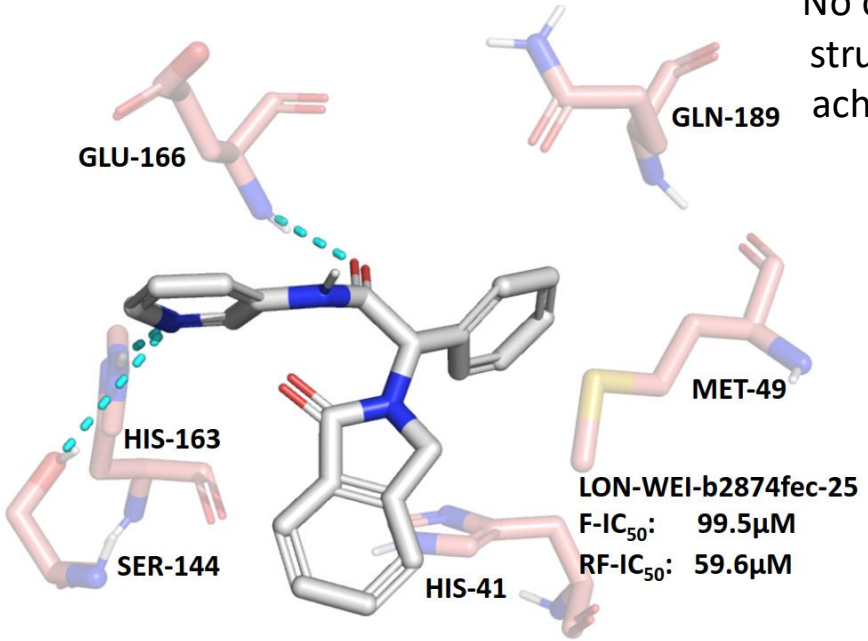
Results

The Fragment Network identifies known and potential binders against Mpro

- The Fragment Network identifies a **known binder** against Mpro



Fragment hits



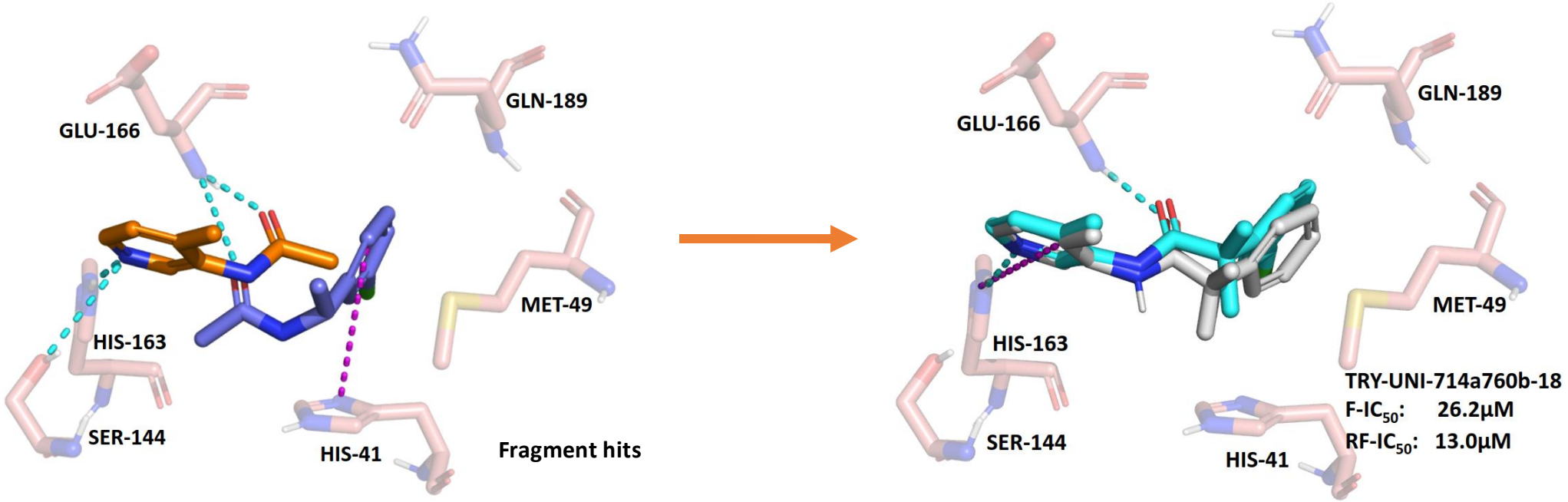
No crystal structure achieved

Fragmenstein-predicted structure of known binder shown in white

Results

The Fragment Network identifies known and potential binders against Mpro

- The Fragment Network identifies several **close analogues** of known binders against Mpro
- Fragmenstein-predicted **poses match crystal structures**

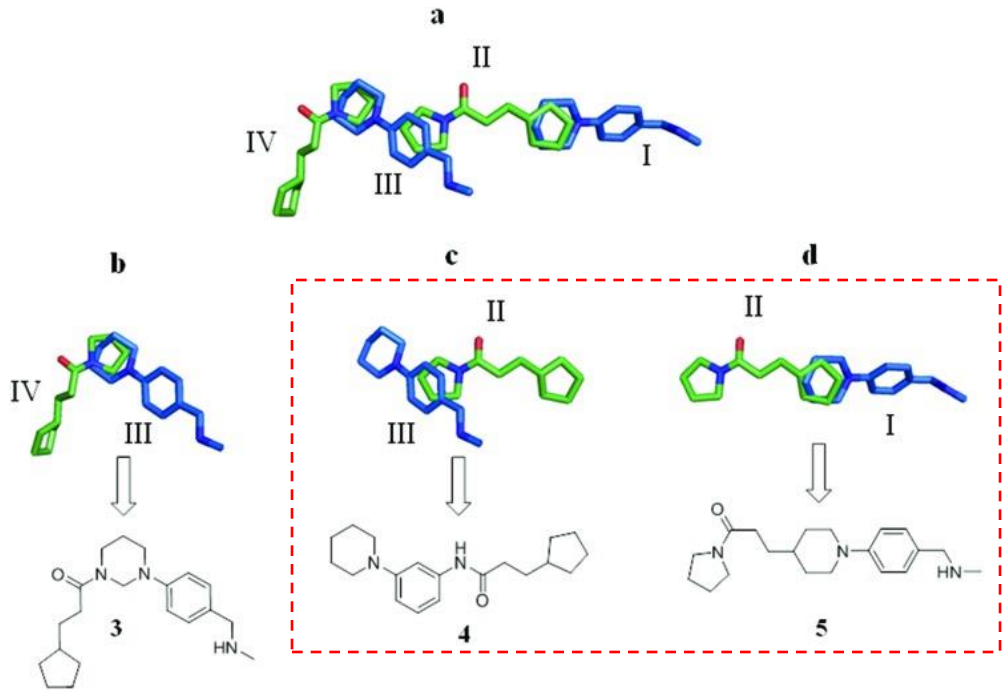


Fragmenstein-predicted structure shown in white
Crystal structure of known binder in cyan

Results

The Fragment Network identifies known and potential binders against EthR

- Manually designed merges against *M. tuberculosis* transcriptional repressor protein **EthR**
- **Two fragments hits** each bind in two conformations (Nikifirov et al, *Org. Biomol. Chem.*, 2016)
- Manually designed merges 4 and 5 overlap with parent fragments
- **Fragment merging pipeline run for the two fragments** and poses generated using all combinations of fragment conformations

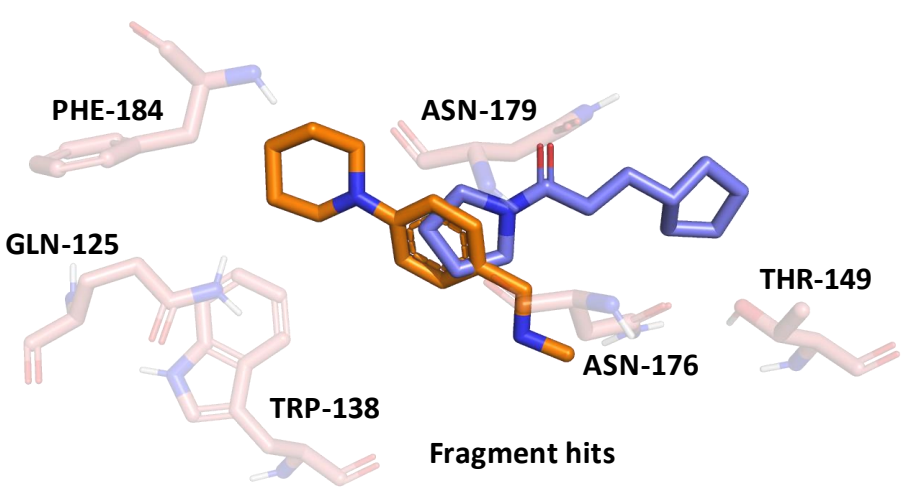


Nikiforov, P.O. et al. *Org. Biomol. Chem.* **2016**, 14(7): 2318-2326; https://doi.org/10.1039/c5ob02630j.

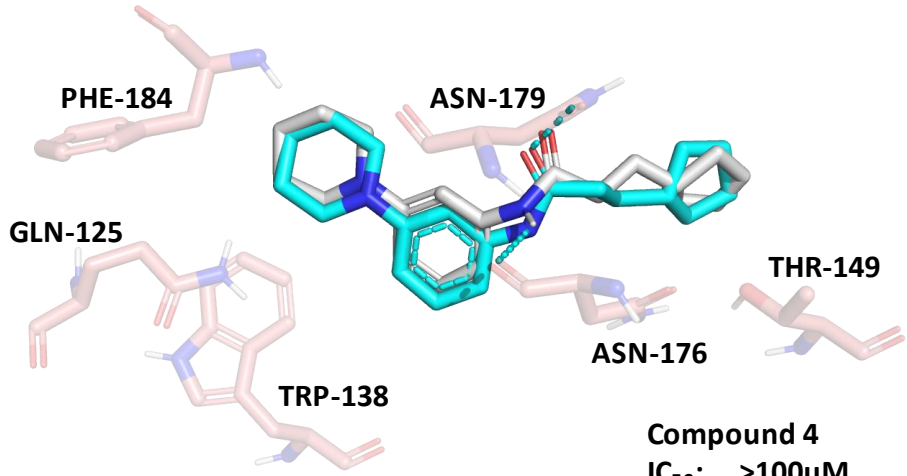
Results

The Fragment Network identifies known and potential binders against EthR

- The Fragment Network **identifies known binder** after filtering (compound 4)



Fragment hits



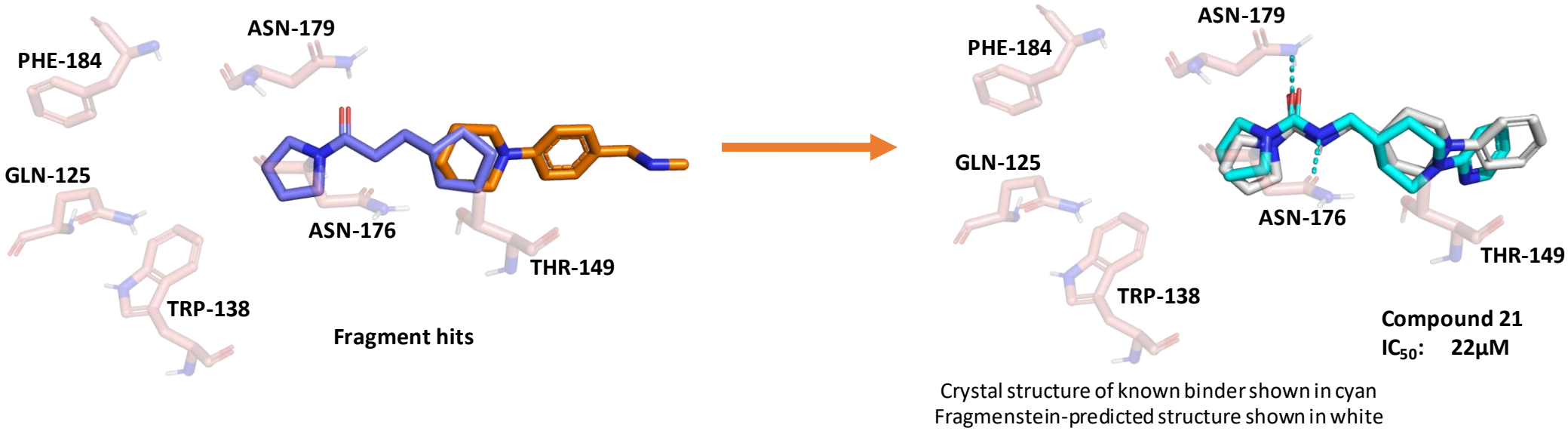
Compound 4
IC_{50}: >100µM

Crystal structure of known binder shown in cyan
Fragmenstein-predicted structure shown in white

Results

The Fragment Network identifies known and potential binders against EthR

- The Fragment Network identifies **close analogues** to several other known binders



Crystal structure of known binder shown in cyan
Fragmenstein-predicted structure shown in white

Summary

- We provide a **flexible pipeline for finding purchasable follow-up compounds** for crystallographic fragment hits using merging and linking-like strategies
- The Fragment Network identifies merges that **maintain exact substructures** of the parent fragments
- The Fragment Network search is **complementary to a more traditional fingerprint-based similarity search** and can be used to **increase the productivity** of a catalogue search
- The Fragment Network is **able to identify potential binders** against two targets
- The Fragment Network search provides a **more computationally efficient** approach to searching large chemical libraries

Expanding the chemical space explored by the Fragment Network search

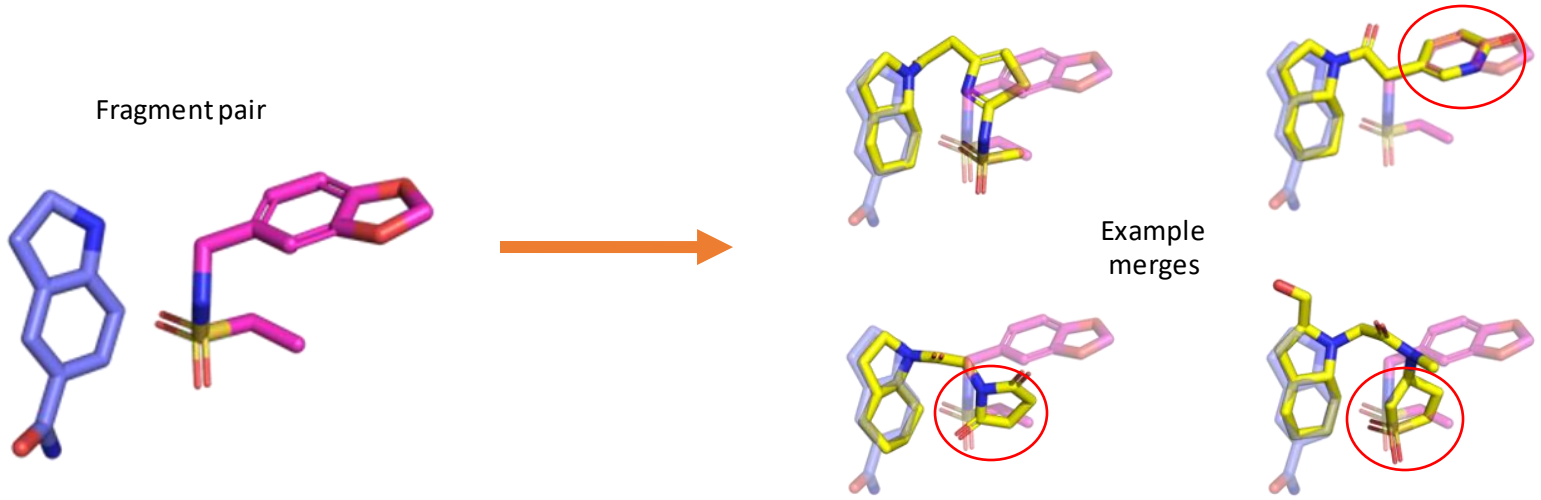
Pipeline v2

Motivations

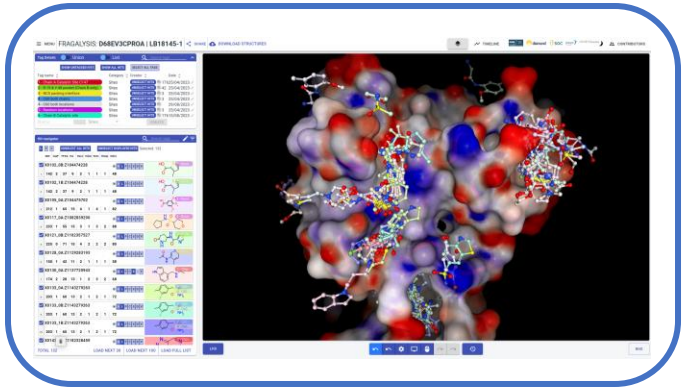
- How can we **maximize the chemical space** explored by fragment merging techniques while still remaining **faithful to the parent fragments**?
- The first iteration of the Fragment Network merging pipeline **fails to find catalogue-based merges for certain pairs of fragments**
- Can we **improve the hit rate** by looking for merges that **recapitulate interactions** of the parent fragments without incorporating the exact substructures?
- Can we do this in a **computationally efficient** way without screening an entire library of compounds?

Expanding the search to 'imperfect' merges

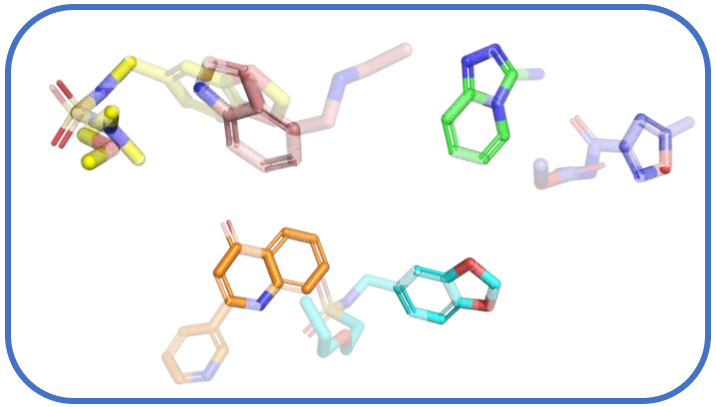
- 'Imperfect' merges
  - How can we identify merges that **replicate pharmacophoric properties** of the parent fragments without incorporating exact substructures?
  - Can we still remain **close enough to the parent fragments** to ensure that the **binding pose is maintained**?



Pipeline for identifying imperfect merges



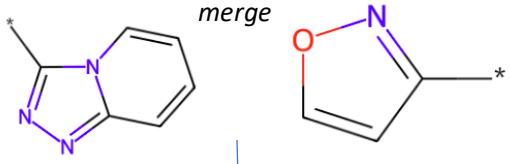
Fragment hits



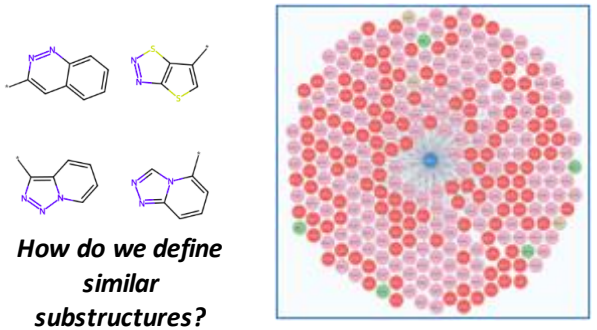
Enumerate compatible substructure for merging

Query the Network for 'imperfect' merges

Substructures to merge

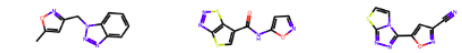


Fragment Network

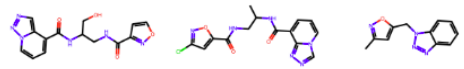


How do we define similar substructures?

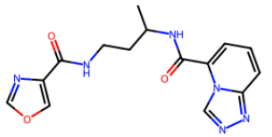
Reverse query

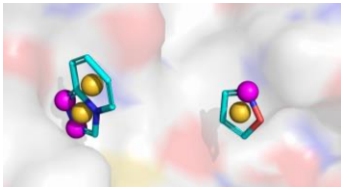


Proposed 2D merges



Generate conformers for the proposed imperfect merges

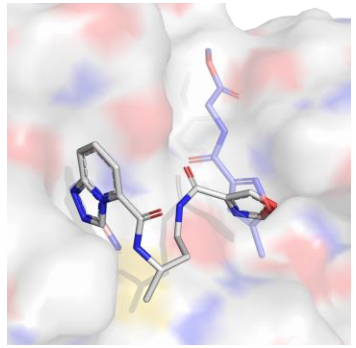




How do we place merges using pharmacophores?

Conformers generated using pharmacophore embedding and Fragmenstein

Minimization

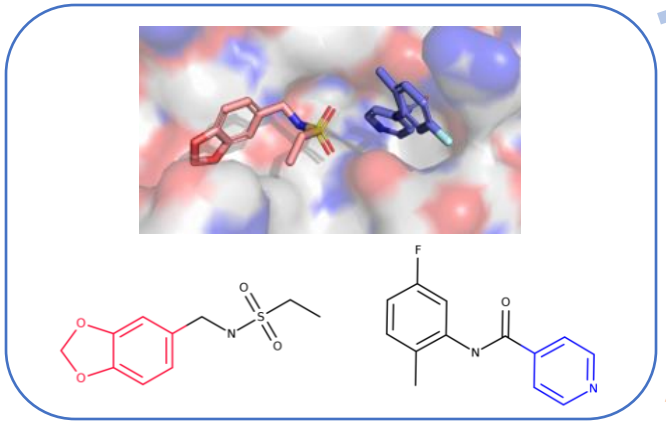


Initial results for imperfect merging pipeline

Enterovirus D68 3C protease

- Enumeration of substructure pairs
  - 24 fragments binding to catalytic site
  - 404 fragment pairs after applying distance and overlap filters
  - 368 substructure pairs for querying
    - Representing 155 fragment pairs
- Querying
  - Fragment Network queried for merges replacing one substructure using similarity calculation
  - Pharmacophore fingerprint used to find similar substructures
- Filtering
  - Conformers generated for a maximum of 500 merges per fragment pair (20,914 compounds placed)
  - Top 30 merges for each pair subject to reverse query and alignment (9,419 additional compounds placed)
  - 417 compounds with SuCOS (shape and colour score) > 0.55

Fragment hits

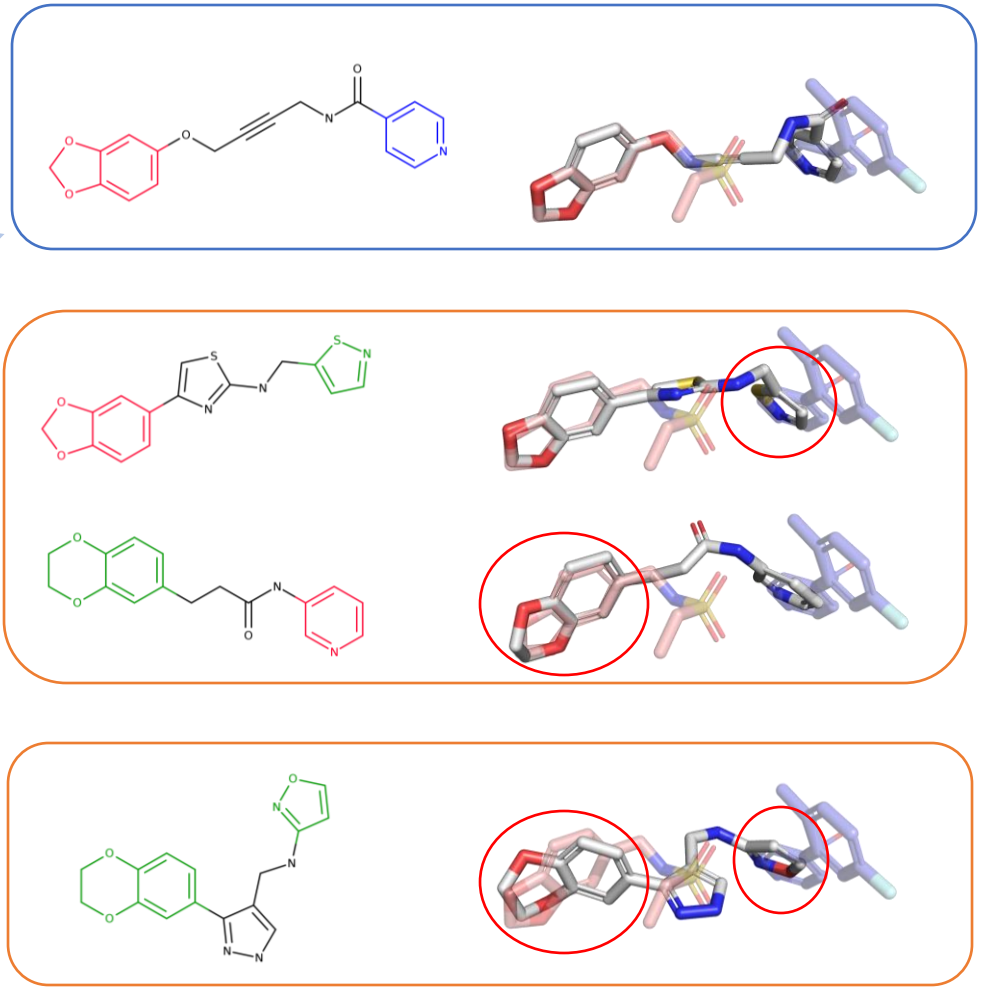


Perfect merge

Replace one substructure

Replace both substructures

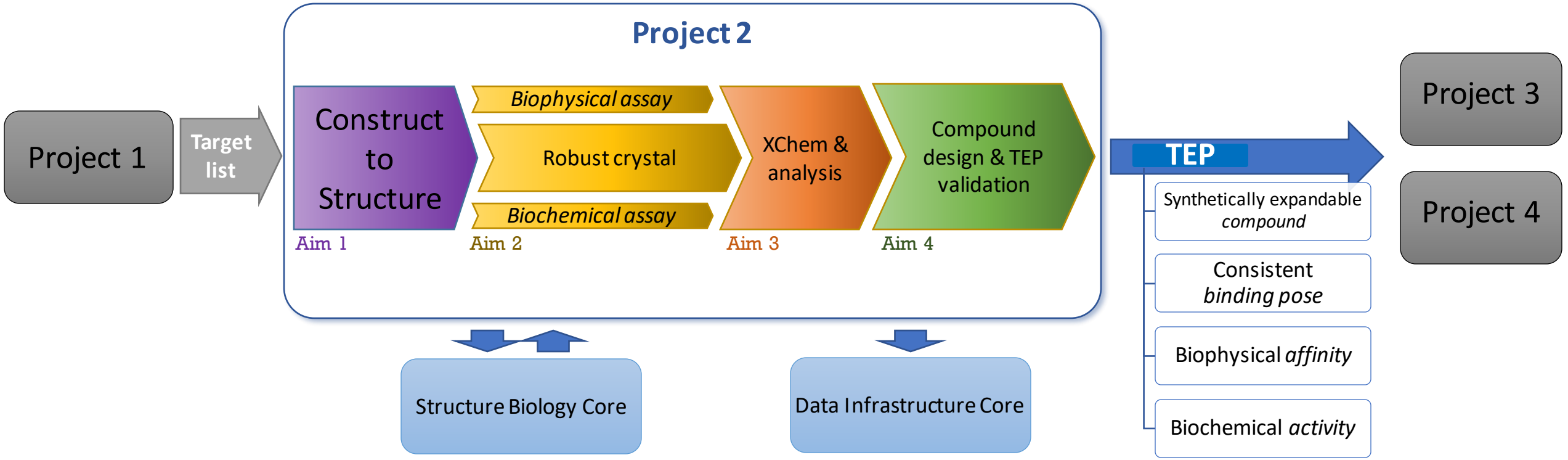
Example merges



SuCOS score from https://github.com/MarcMoesser/SuCOS

Deployment of fragment elaboration algorithms at XChem

Using ASAP to validate the approach

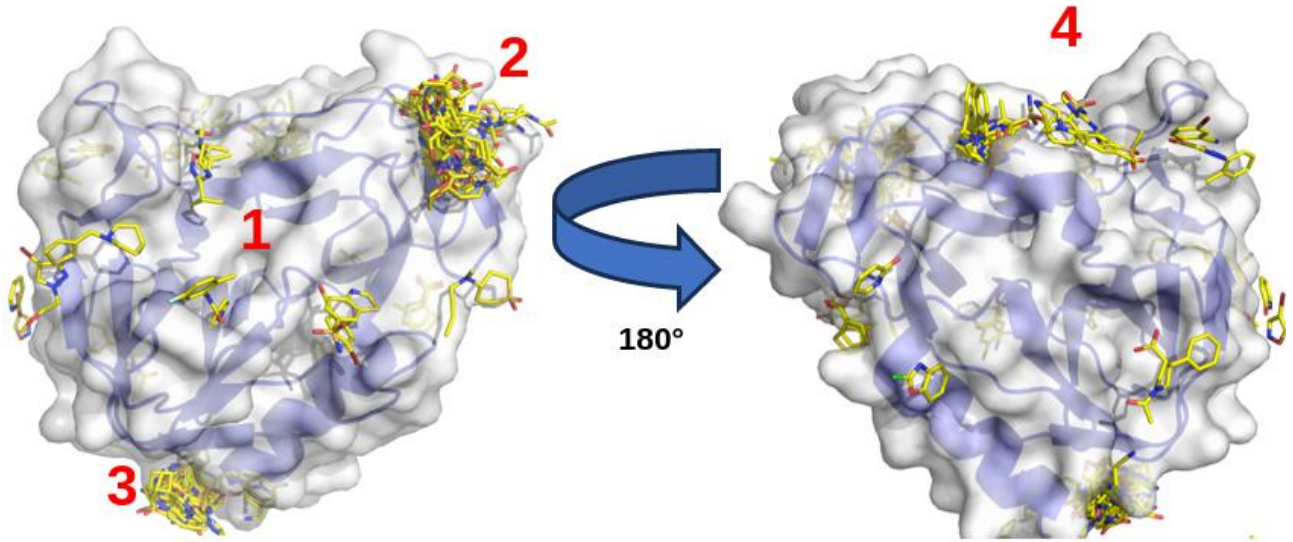


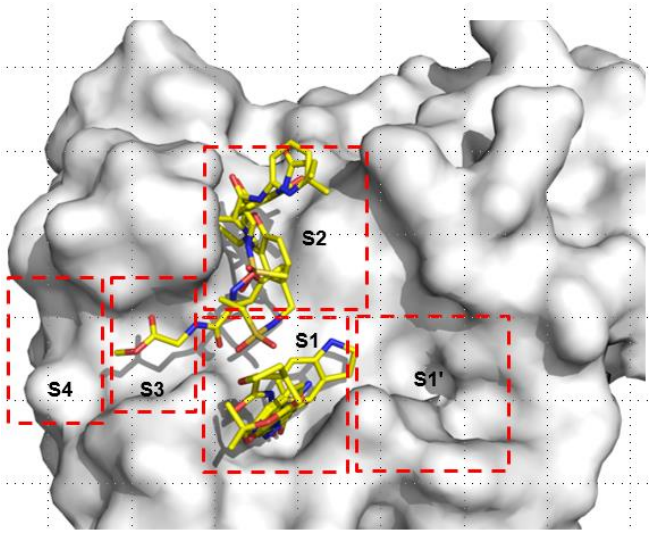
- Work being done to establish process for **Aim 4** for ASAP targets
- Enables prospective validation for design algorithms

Enterovirus D68 3C protease

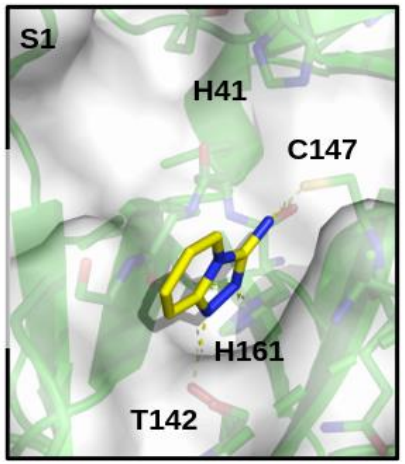
Summary

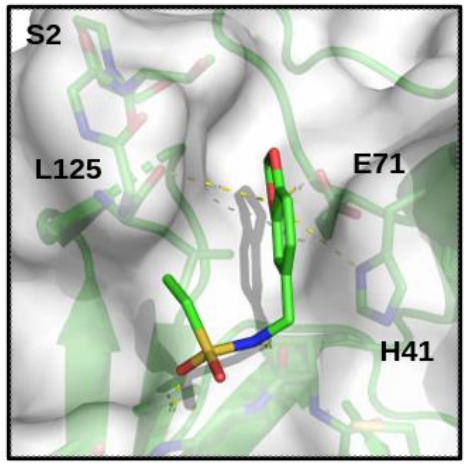
- 1231 compounds screened
- 126 hits identified in 4 sites
  1. Catalytic site – Monomer A
  2. R19 & Y48 Pocket – Monomer B
  3. ASU interface
  4. C60 – Both monomers





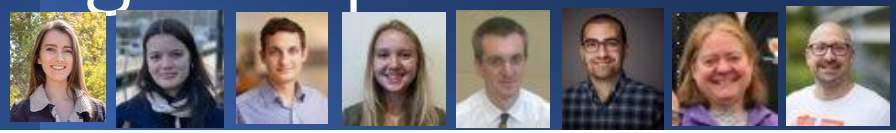
18 hits
Predominantly in S1 and S2





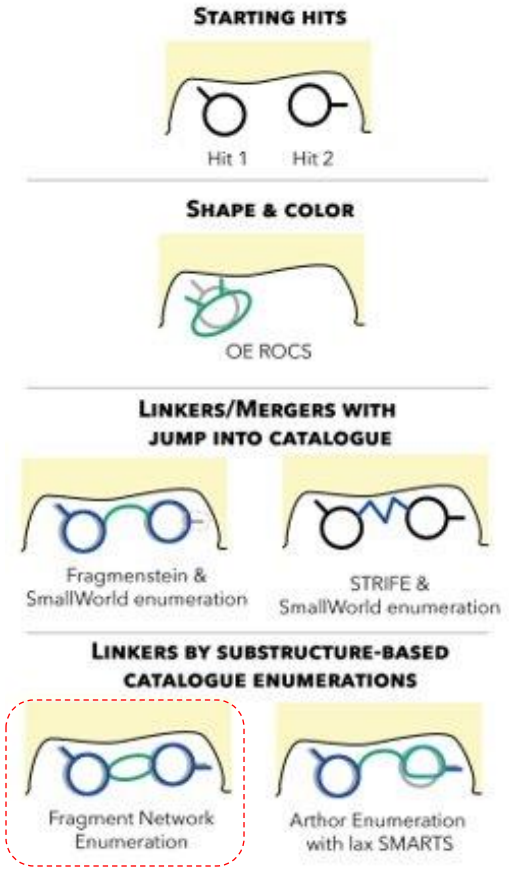
ASAP
Accessible Antivirals to Prevent Pandemics

Pipeline for designing and ordering compounds



Compound design

Compounds are proposed using direct enumeration from catalogues or by generative models plus analogue search



Multifactor ranking

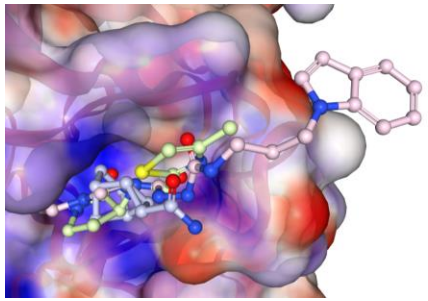
Compounds from design algorithms are scored using shape and colour, predicted interactions (PLIP), etc.

Medicinal chemists

Compounds are sanity checked and curated by medicinal chemists

Compound ordering

Process established for ordering compounds from Enamine

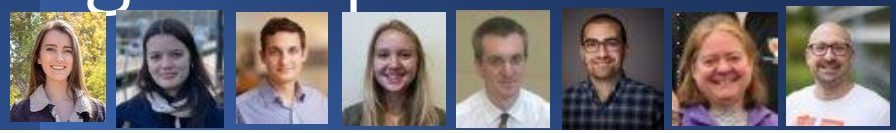


Compound designs accessible on Fragalysis

- Enterovirus D68 3C protease used as an initial target
- Pipeline established for designing, ordering and assaying compounds from Enamine
- First iteration took 7 weeks from design to new structures
- More data incoming from further iterations and targets
- Compounds will feed into next-generation of work that explores moving away from the catalogue for SAR exploration

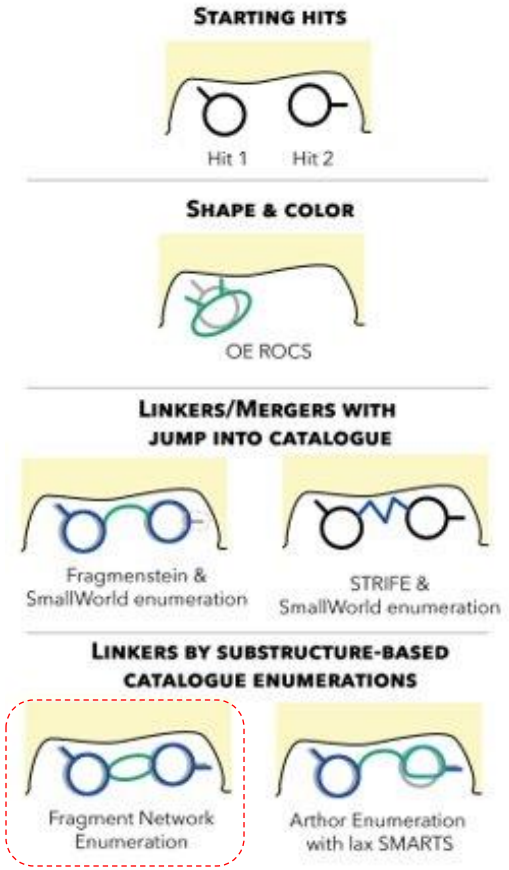
https://fragalysis.diamond.ac.uk/

Pipeline for designing and ordering compounds



Compound design

Compounds are proposed using direct enumeration from catalogues or by generative models plus analogue search



Multifactor ranking

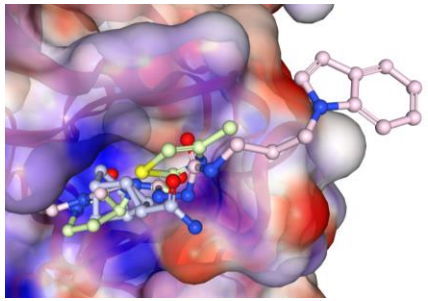
Compounds from design algorithms are scored using shape and colour, predicted interactions (PLIP), etc.

Medicinal chemists

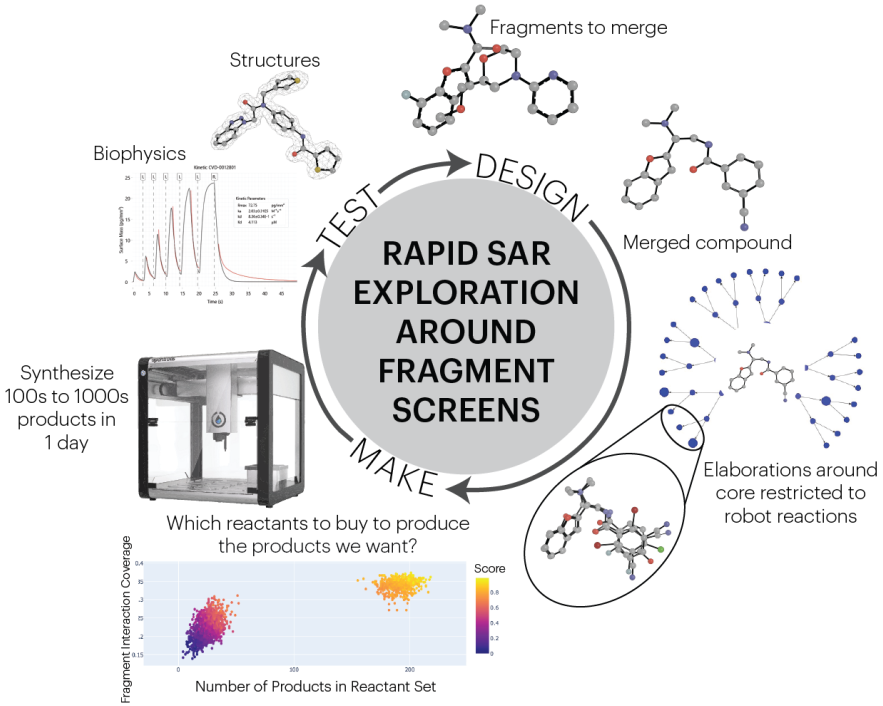
Compounds are sanity checked and curated by medicinal chemists

Compound ordering

Process established for ordering compounds from Enamine



Compound designs accessible on Fragalysis



https://fragalysis.diamond.ac.uk/

Summary

- The Fragment Network merging tool provides a **pipeline for finding merges** from the **catalogue** that **recapitulate substructures** from **crystallographic fragment hits**
- We show initial development on an updated pipeline that **expands the chemical space** explored for finding these merges using **pharmacophores** without the requirement for performing a virtual screen on an entire library of compounds
- The ASAP projects have helped to develop a **workflow for ordering follow-up compounds on a rapid timescale**, moving from compound design to structures in a matter of weeks, and providing **validation for the design algorithms** used at XChem

Acknowledgements

- Ruben Sanchez-Garcia (OPIG)
- Charlotte Deane (OPIG)
- Frank von Delft (DLS & CMD)
- Tim Dudgeon (Informatics Matters)
- Rod Hubbard (Vernalis)
- Steve Roughley (Vernalis)
- James Davidson (Vernalis)
- Andy Merritt (LifeArc)
- Matteo Ferla (OPIG & XChem)
- Warren Thompson (XChem)
- Kate Fieseler (OPIG & XChem)
- Lucy Vost (OPIG)
- Beth MacLean (XChem)
- Ryan Lithgo (XChem)
- Ed Griffen (Medchemica)
- Tryfon Zarganis-Tzitzikas (CMD)

AND MANY MORE…

Thank you for listening