

# FRAGMENSTEIN



MATTEO FERLA

## FRAGMENSTEIN: STITCHING COMPOUNDS TOGETHER

30 MINUTE TALK



Centre for  
Medicines  
Discovery



p | g



wellcome  
centre  
human  
genetics

NIHR

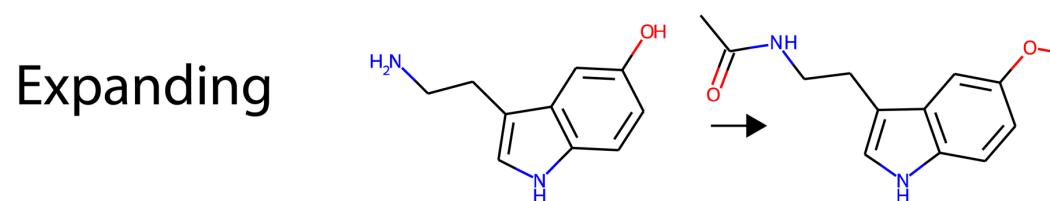
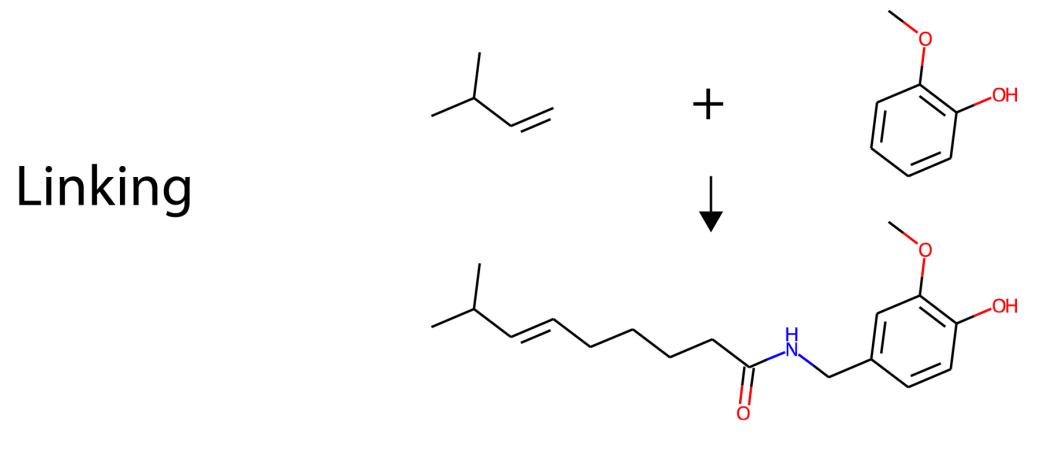
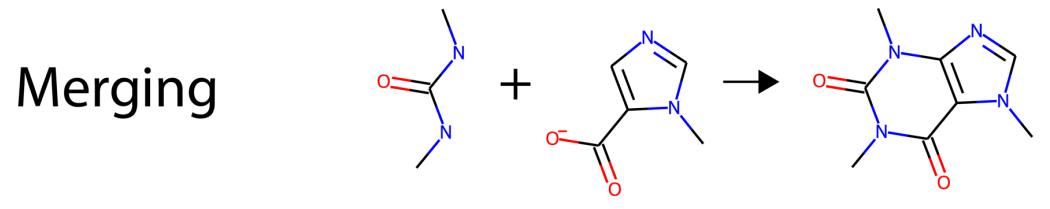
Oxford Biomedical  
Research Centre



UNIVERSITY OF  
OXFORD

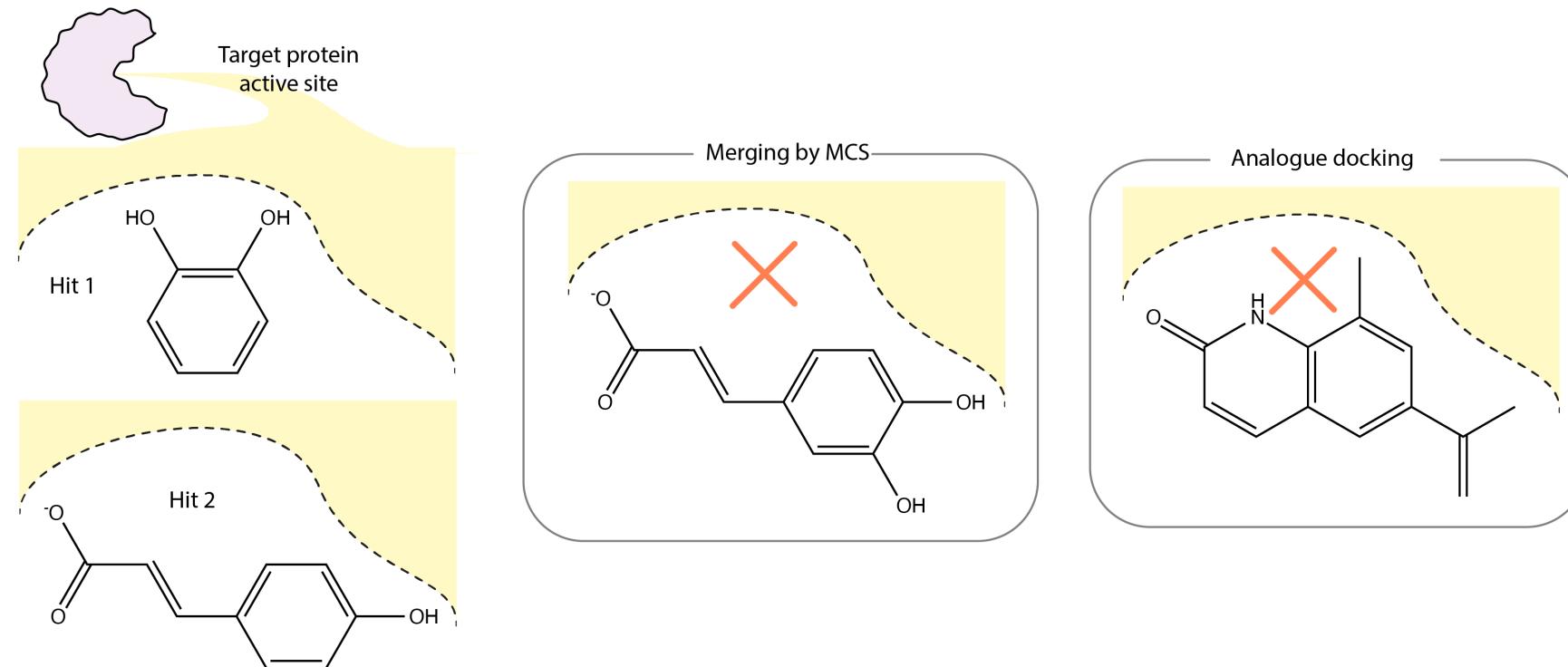
# FRAGMENT BASED DRUG DISCOVERY

- + Small libraries
  - + Ability to detect weak binding
  - + **Analogues bind similarly**



# LACK OF FIDELITY TO FRAGMENT HITS IN PREDICTIONS

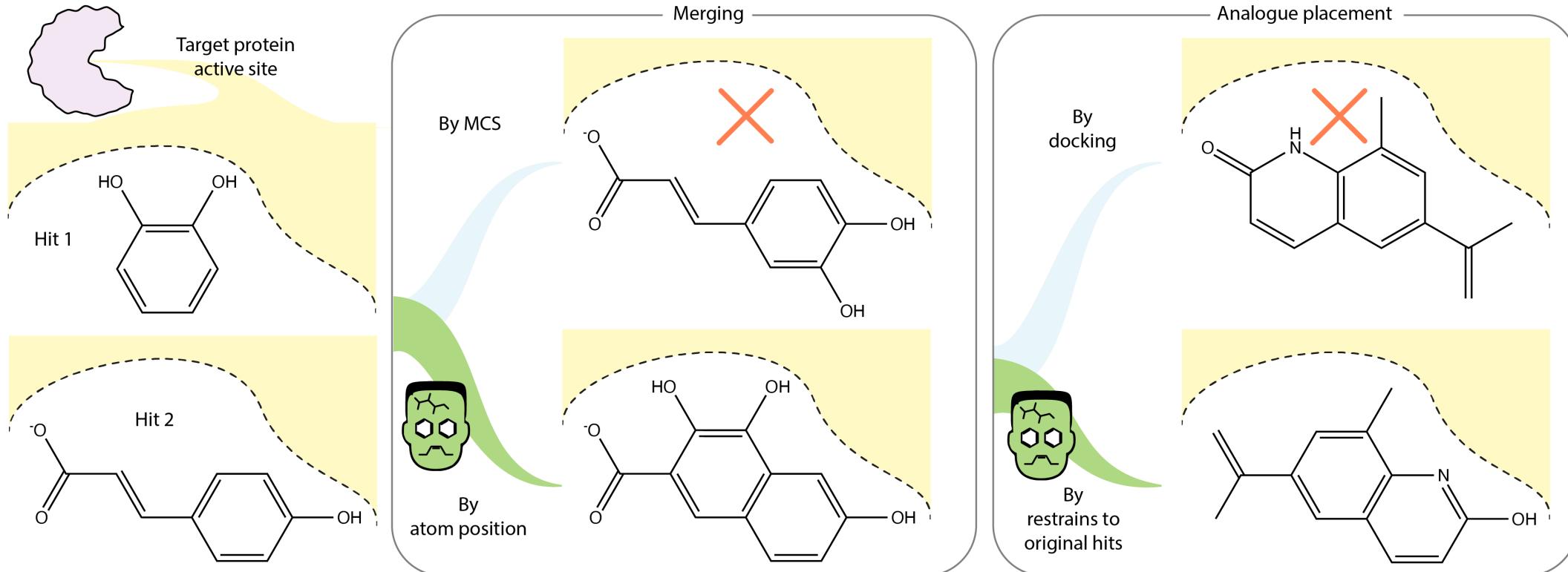
- Many docking and combinational methods not faithful to position of starting hits



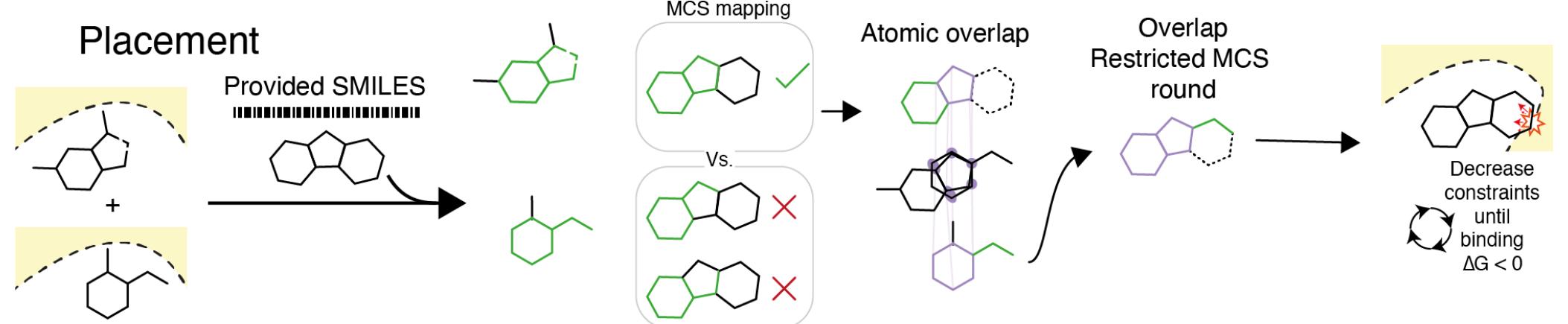
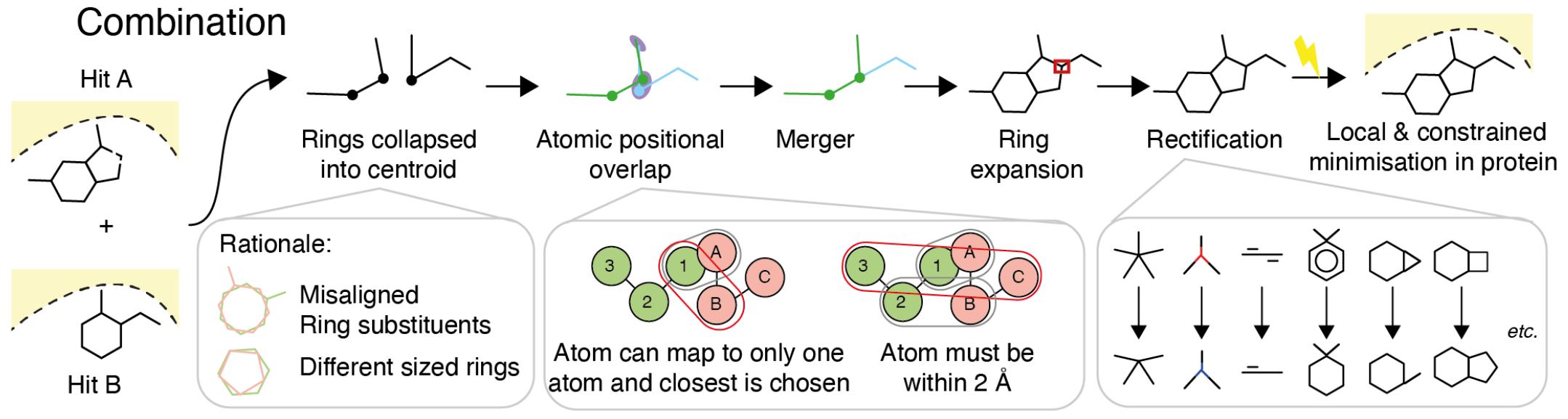


matteoferla/fragmenstein

# SOLUTION: FRAGMENSTEIN



# TWO ROUTES



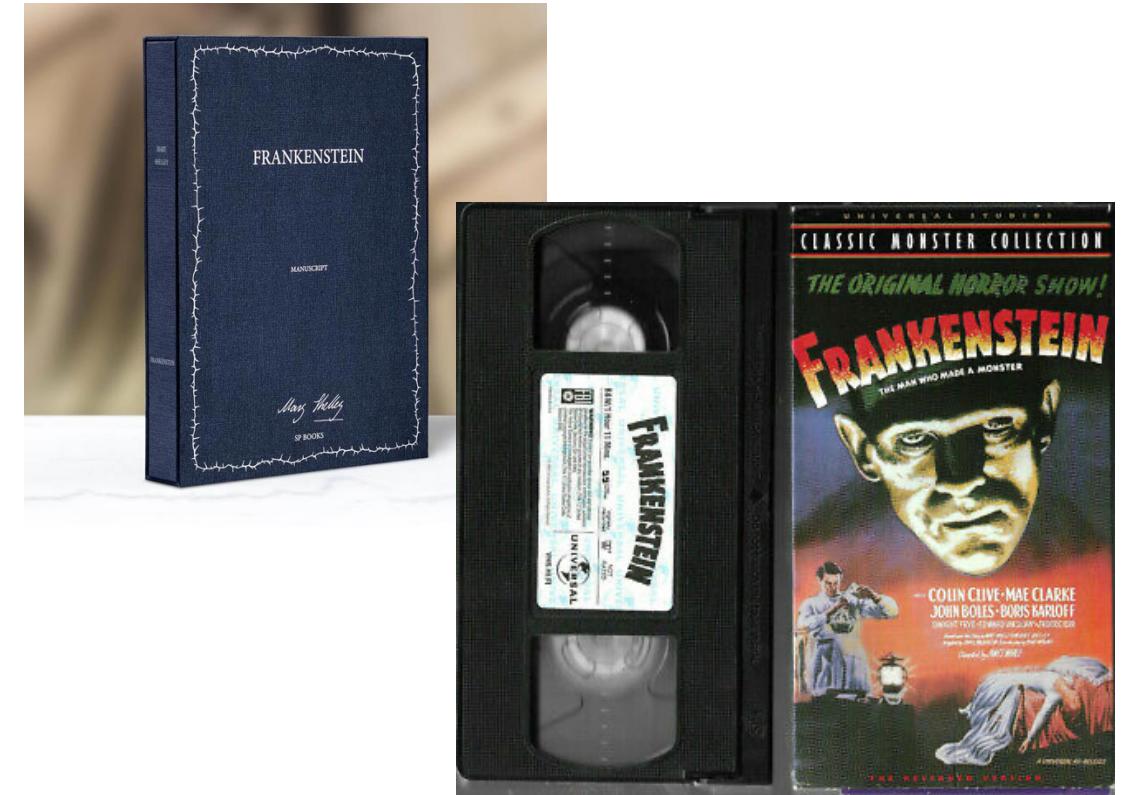


matteoferla/fragmenstein

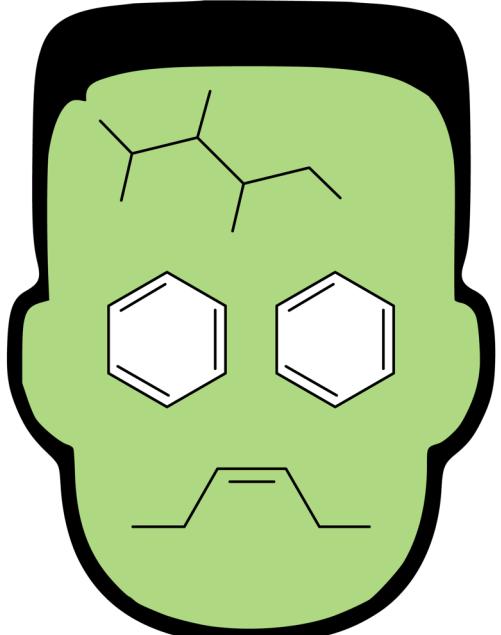
# FRAGMENSTEIN'S INNARDS

- Open source
- RDKit for molecular operations
- PyRosetta for protein/ligand minimisation
- Warning: jumble of both book and films

Classes include: Monster, Igor, Victor, Laboratory, Walton



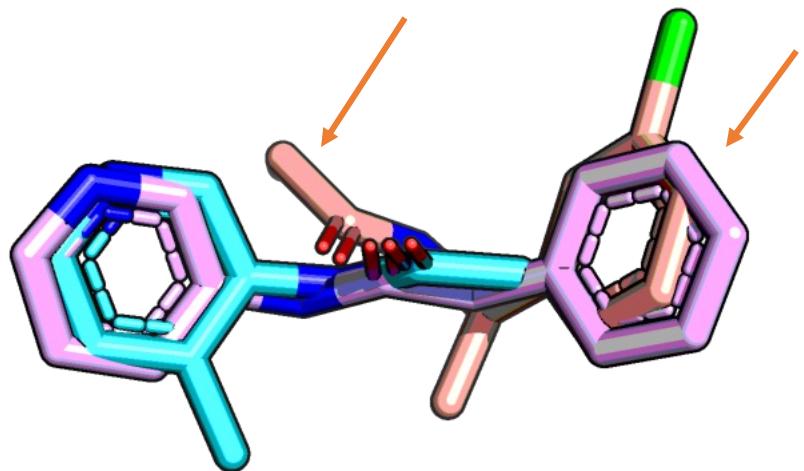
**FRAGMENSTEIN — COMBINE**



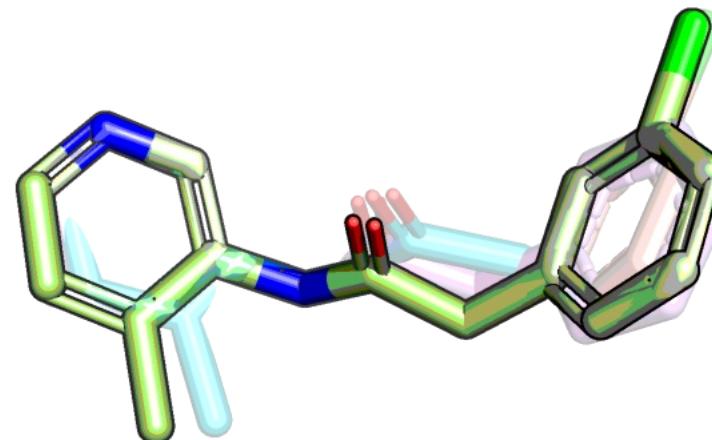
UNIVERSITY OF  
**OXFORD**

# IMPERFECT OVERLAPS

**Inspiration hits**  
(Mpro mooshot project)

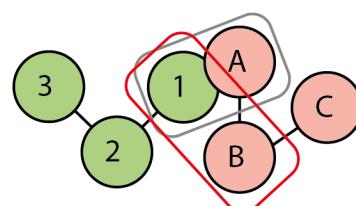


**Crystallised merger/expansion**  
(followup)

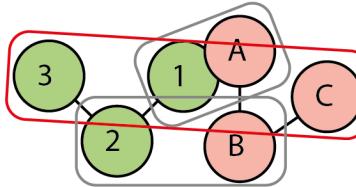


# RULES FOR ATOMIC OVERLAPS

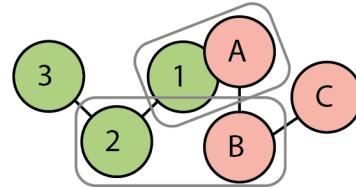
- Fragmenstein merges by atoms, not by bonds like Schrödinger BREED



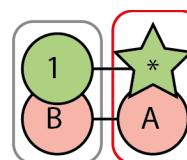
Atom can map to only one atom and closest is chosen



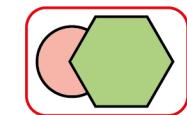
Atom must be within 2 Å



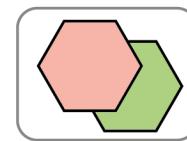
Mapping generated



Regular atoms cannot map with R-groups



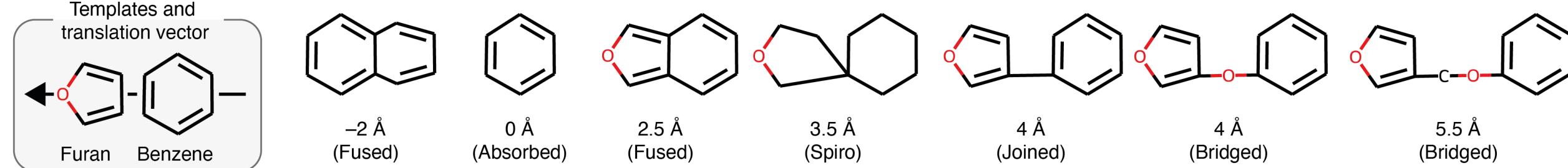
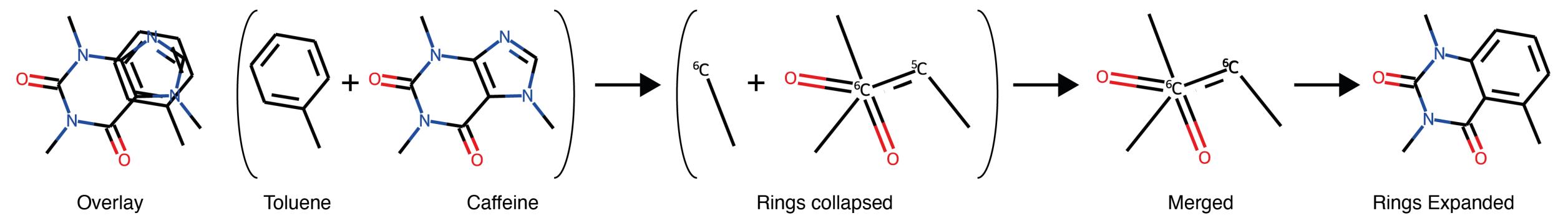
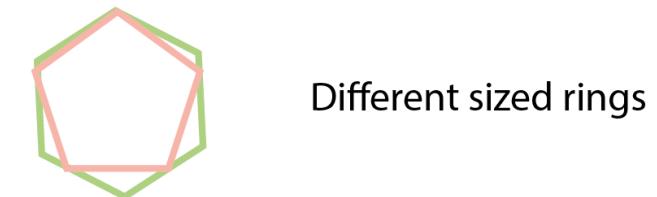
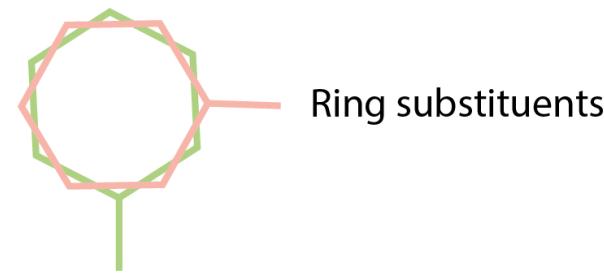
Regular atoms cannot map with ring placeholders



Ring placeholders can map with ring placeholders

# COLLAPSING RINGS

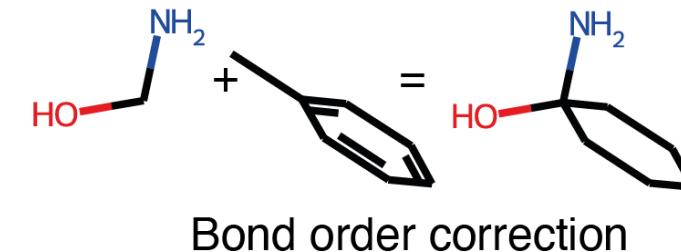
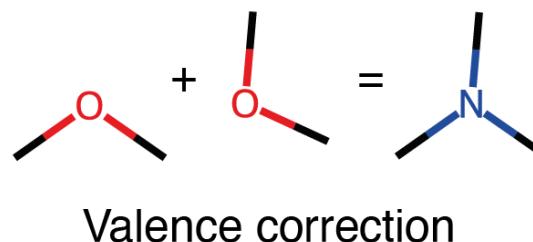
- Rings are problematic:  
misaligned substituents & different sizes
- Merging a ring placeholder resolves this





# RECTIFICATION

- A rules based algorithm corrects chemistry errors allowing it to be sanitised
- Analogue search circumvents intractable chemistry





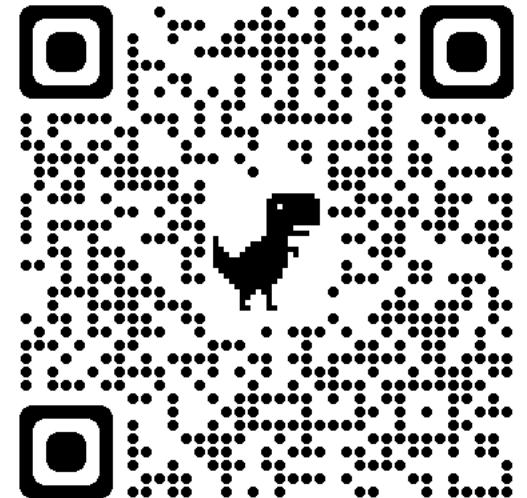
# IGOR'S ENERGY MINIMISATION

---

- The merger is stitched together and distorted:  
unfeasible bonds and dihedrals
- Constrained energy minimisation with PyRosetta in the target protein
- Decreasing weights until  $\Delta G$  of binding is negative

# BENCHMARK

- Mac1 (nucleosyl-peptide hydrolase, SARS-COV-2)
  - Mpro (cysteine protease, SARS-COV-2)
  - DSi poised library
- 
- Combinations (mergers and linkers) generated



[michelangelo.sgc.ox.ac.uk/  
r/fragmenstein-mpro-DSiP](https://michelangelo.sgc.ox.ac.uk/r/fragmenstein-mpro-DSiP)

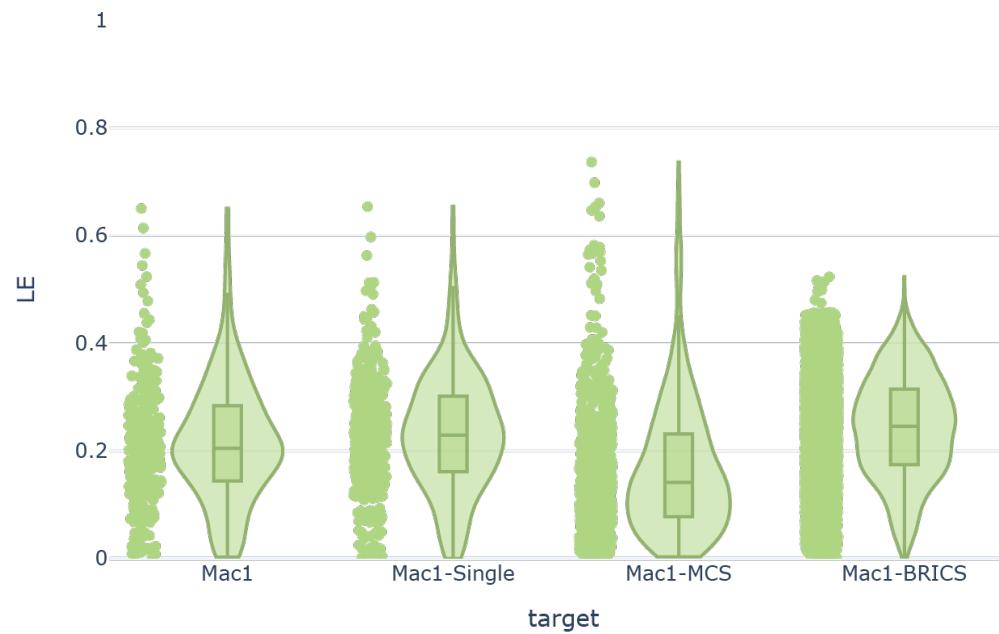
# BENCHMARK RESULTS

	MPro	Mac1
Number of hits used	34	44
Number of acceptable mergers	157	263
Number of failed mergers due to equal size to one hit	13	34
Number of failed mergers due to > 5 Å minimum distance between hits	918	1438
Number of failed mergers due to strain ( $\Delta\Delta G > 0$ kcal/mol or >1 Å RMSD)	33	149
Number of failed mergers due to technical issues	1	8
median mol. wt of acceptable subset	356.1	305
median QED of acceptable subset	0.79	0.66
Number of acceptable compounds with SA < 0.2	54	27
Number of acceptable compounds with SA≤0.4	71	40
Number of acceptable compounds that are purchasable	5	2
Number of acceptable compounds with purchasable analogues in Enamine Real differing by 2 or fewer atoms	26	22
Number of acceptable compounds accessible via a one-step synthesis	28	10
Number of acceptable compounds accessible via a two-step synthesis	16	10

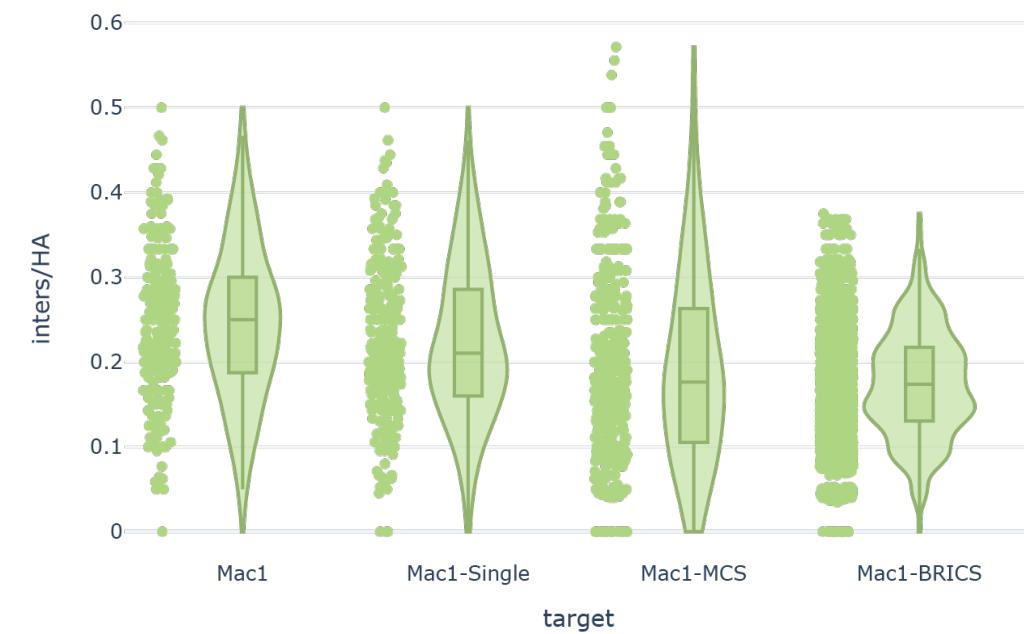
# MISLEADING POTENTIAL ENERGY

Poses of mergers generated by Fragmenstein may have worse energies per atom than structure-naïve approaches, but form more interactions per atom

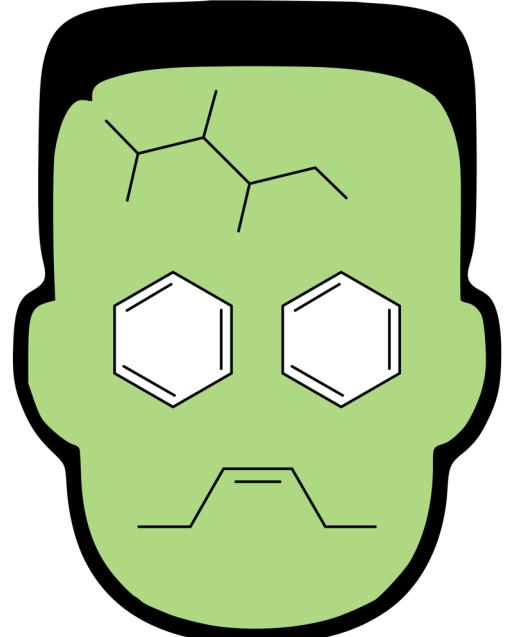
Distribution of number of ligand Efficiency of acceptable mergers



Distribution of number of interactions per heavy atom of acceptable

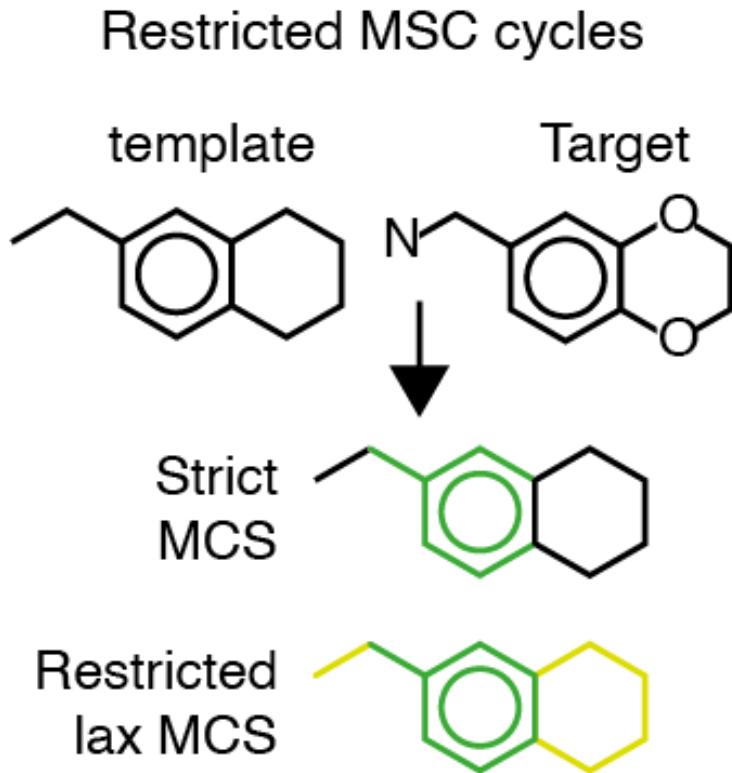


# FRAGMENSTEIN — PLACEMENT



UNIVERSITY OF  
**OXFORD**

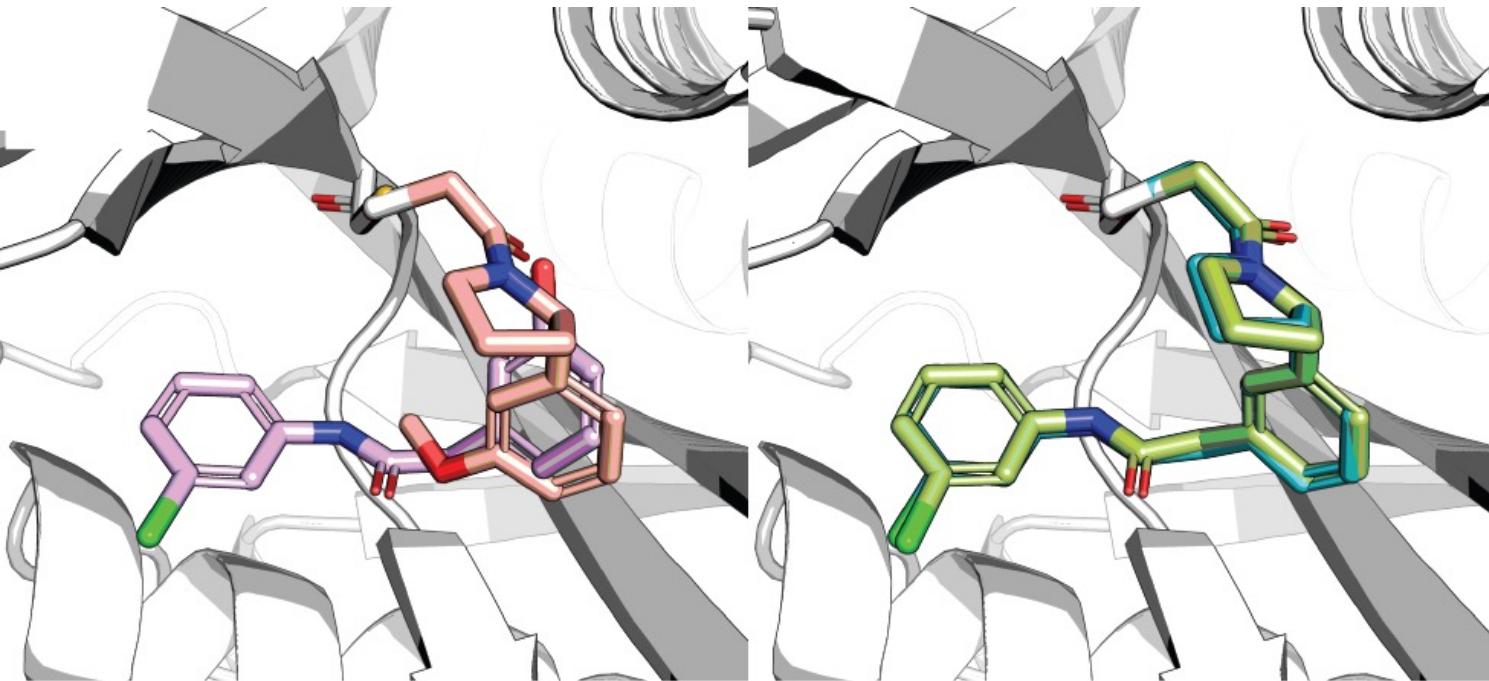
# MAPPING CHALLENGE



- Travelling salesman problem: which atom in the follow-up maps to which atom in the inspiration?
- Strict MCS search followed by laxer MCS searches constrained by the former
- Up to 3 hampering atoms discarded, e.g. misleading substituents

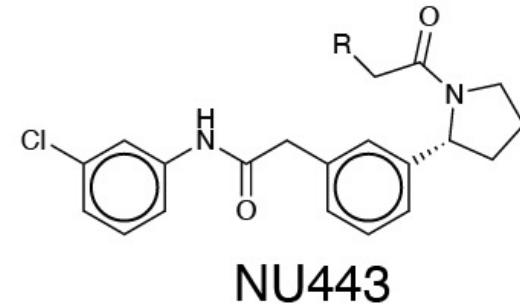
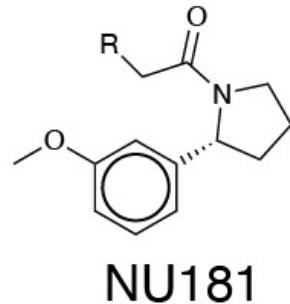
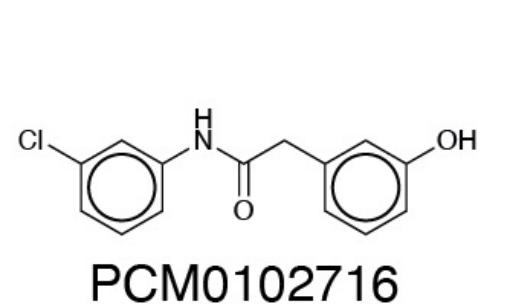
# TOLERANCE TO MISLEADING ATOMS

Two inspiration hits

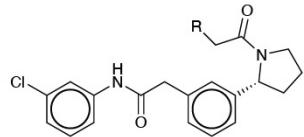
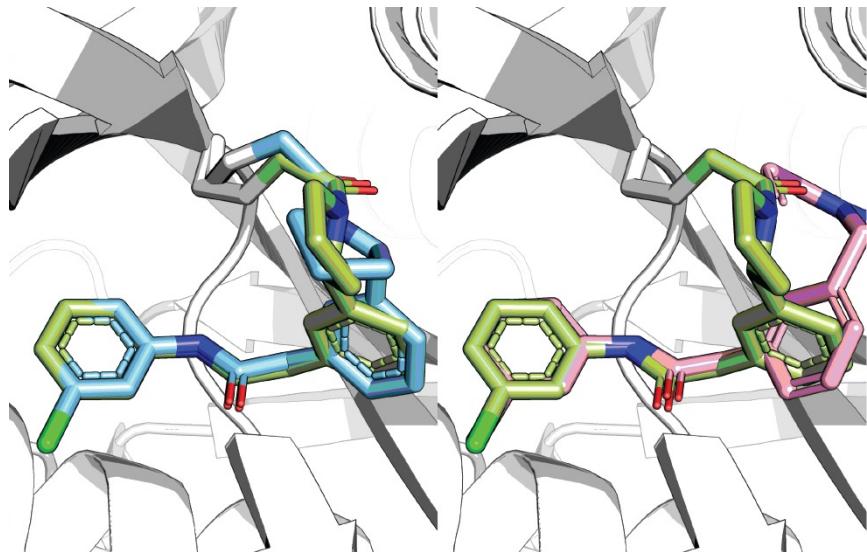


Crystal (blue), Fragmenstein (green)

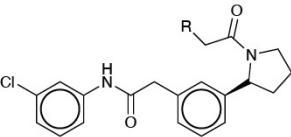
- Test on historical data
- NUDT7 from Resnick et al. 2020
- Several atoms do not overlap cleanly, while others are misleading
- Combined RMSD against the inspiration hits
  - of model = 0.65 Å
  - of structure = 0.61 Å
- RMSD Fragmenstein vs. crystal  
0.28 Å



# STRAINED FAILED BINDER



NU443



NU442

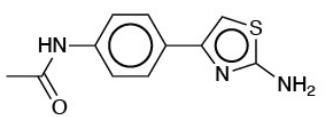
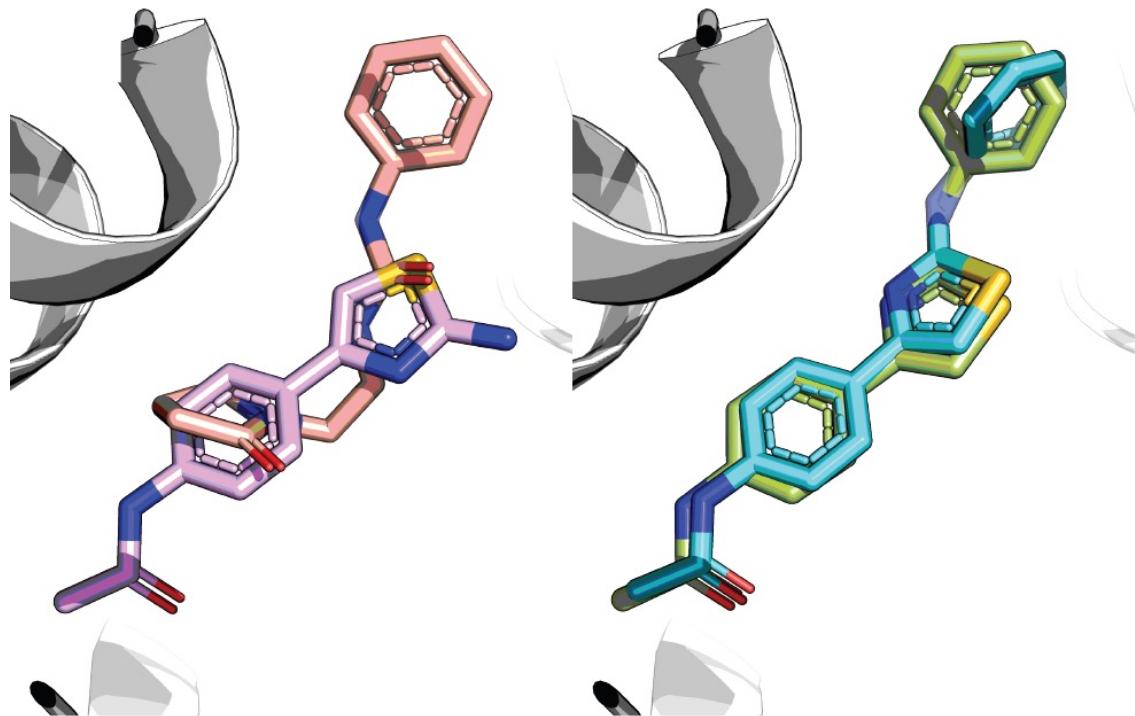
Blue: Previous structure

Puce: Isomer, non reaction

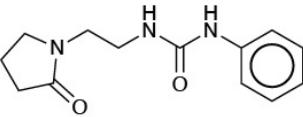
Green: Isomer, predicted

- Isomer did not react
- Covalent model:
  - (higher RMSD)
  - strained cysteine sulfur
  - 10% worse potential
- Fragmenstein prefers the correct isomer

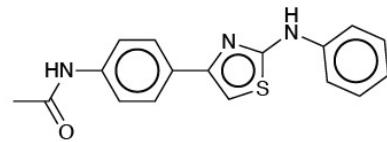
# USER DEFINED ATOM MAPPING



F36



F04



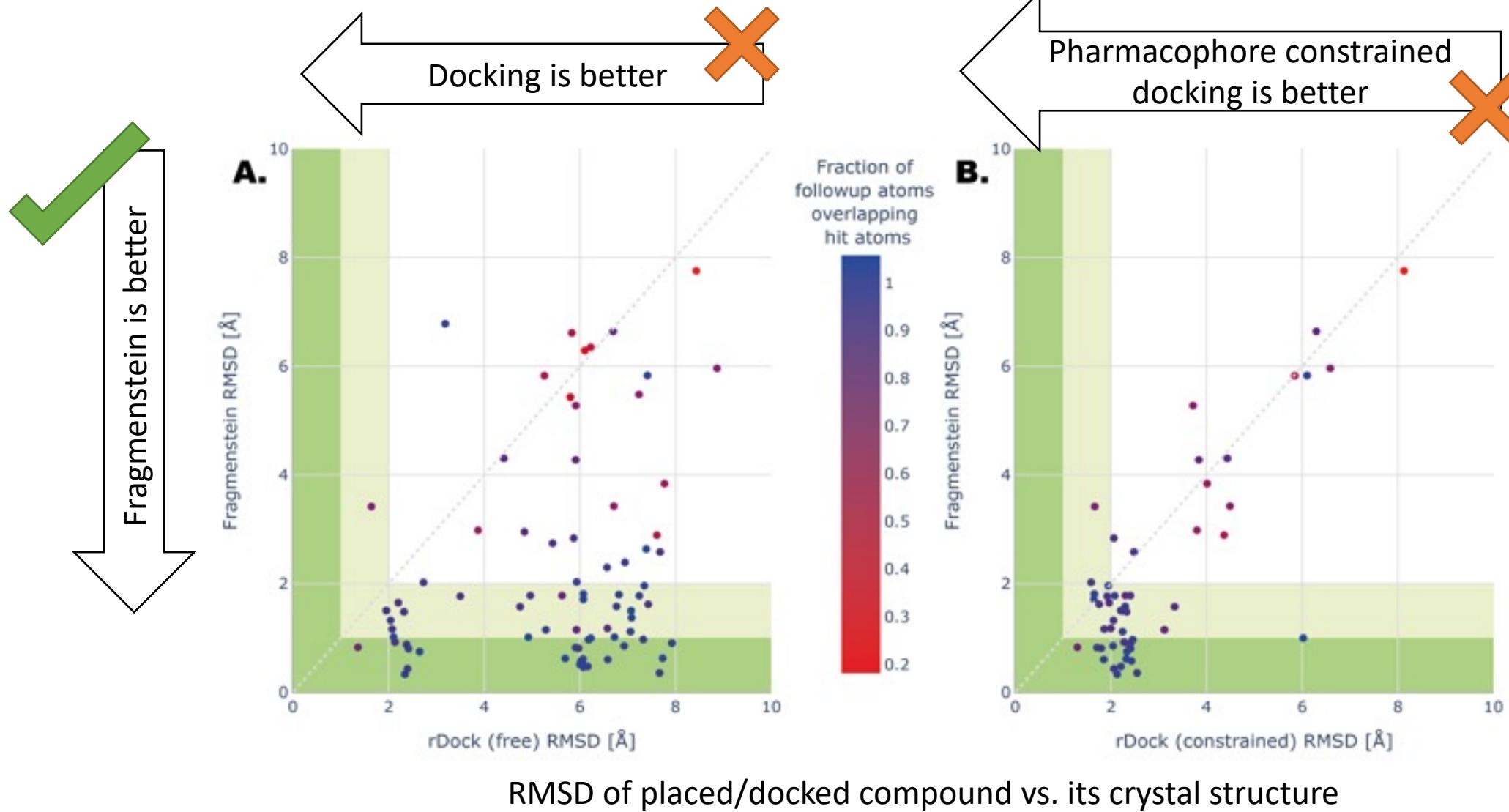
Todalam-4

- Test on historical data
- Tubulin from Mühlethaler *et al.* 2022
- A flipped ring in the follow-up relative to an inspiration (density for both orientations)
- User can explicitly specify an atom mapping

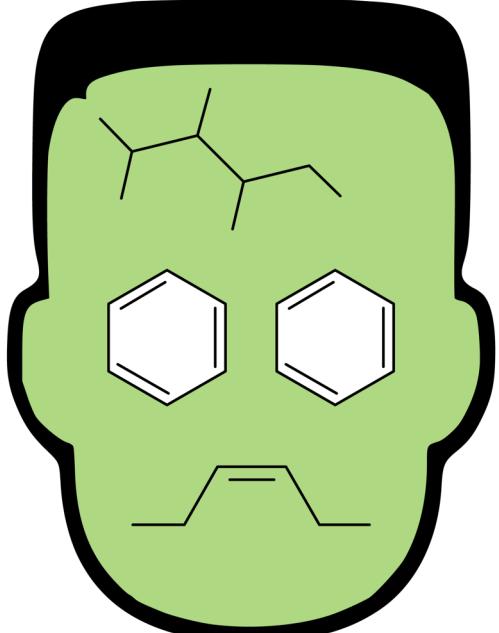
# MPRO MOONSHOT SUBMISSIONS



Rubén Sanchez-García,  
OPIG,  
U of Ox

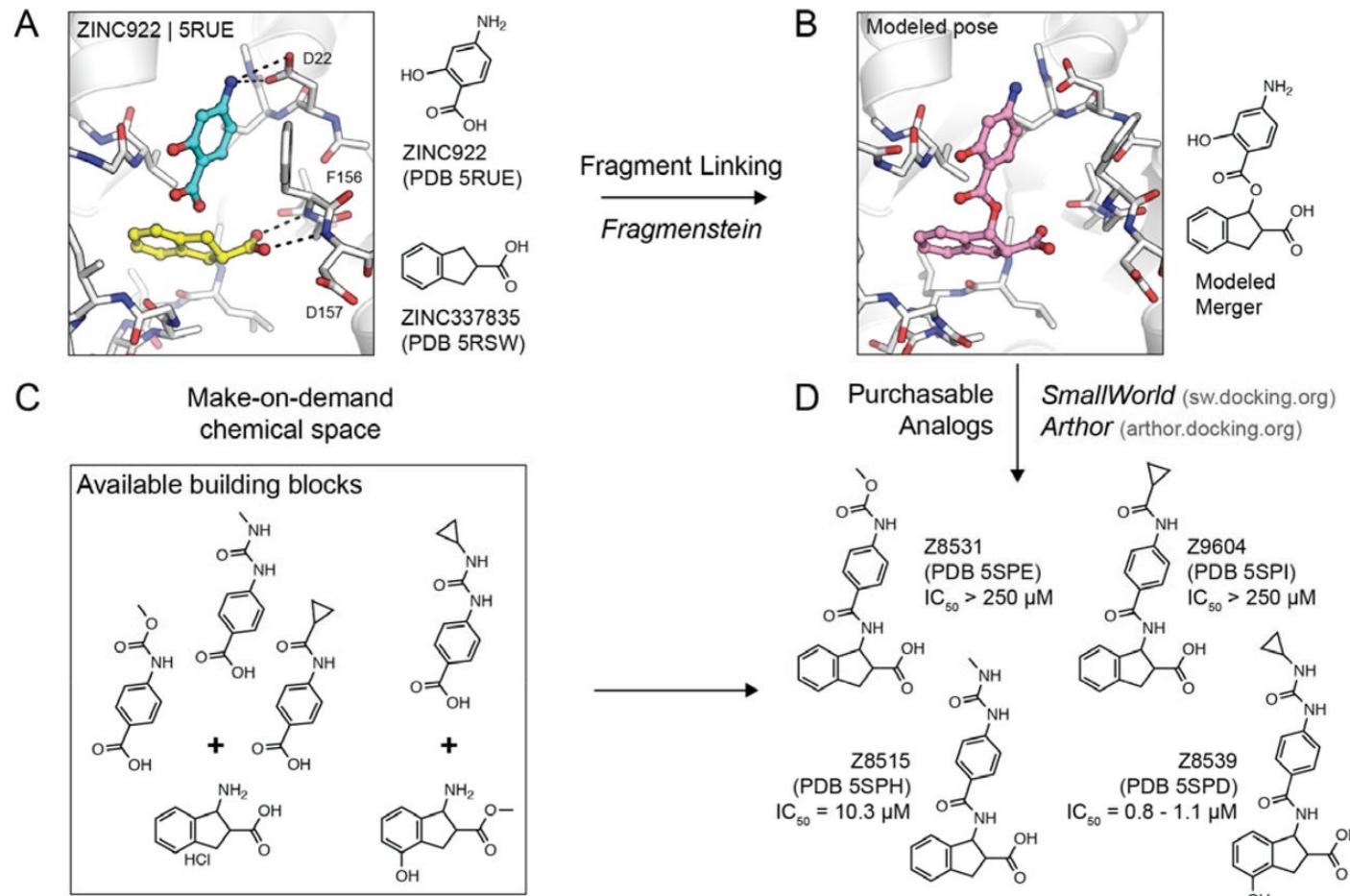


# FRAGMENSTEIN — USAGE



UNIVERSITY OF  
**OXFORD**

# USAGE: GAHBAUER ET AL. 2022

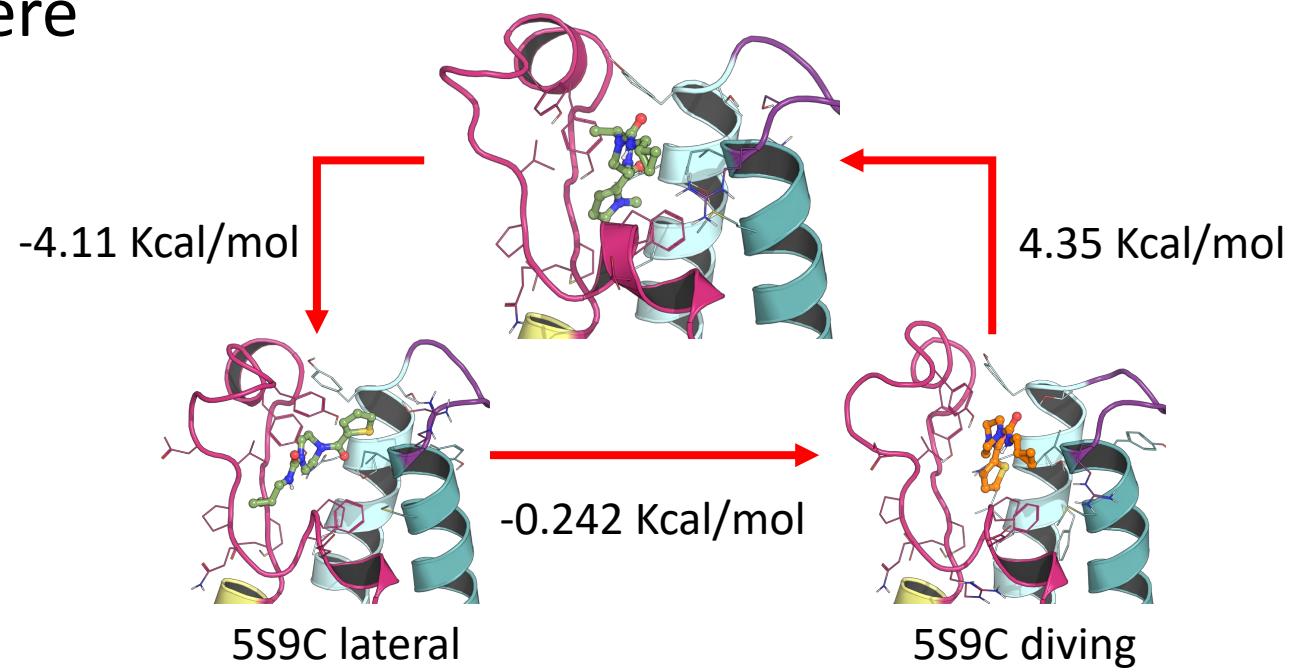


- Pairwise combinations with analogue search via SmallWorld server
- On of top 10 hits gave submicromolar  $IC_{50}$  (fortuitous novel substituents)

# USAGE: HYPOTHESIS TESTING



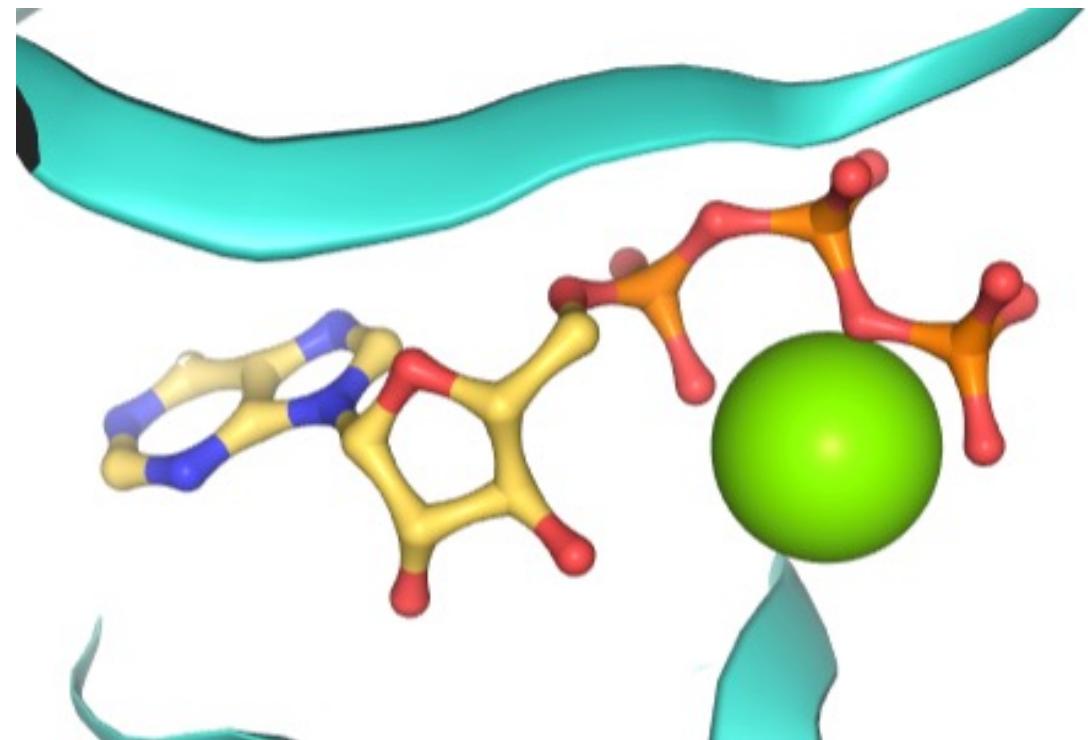
- A set of follow-up compounds in FBDD crystallised differently everting a loop (mimicking substrate bound form)
- To test if crystal-artefacts, they were placed with Fragmenstein in the expected conformation and MD simulated



Harold Grosjean  
Phil Biggin group,  
U of Ox  
@GrosjeanHarold

# USAGE: LIGAND NATIVISATION

- Many crystal structures contain cofactor or substrate analogues
- Fragmenstein has been used to convert GDP+NO<sub>3</sub> to GTP in Pagnamenta et al. 2022



# FRAGMENSTEIN BALANCES

## Pros

- Is faithful to inspirations
- Can work with more than one inspiration molecule
- can deal with covalent molecules
- Is open-source
- Is user-interaction driven

## Cons

- For full operations needs PyRosetta which requires a licence
- Takes 30+ seconds per combination/placement
- (combine) Distance of analogues in make-on-demand space
- (placement) Performs poorly with red herring inspirations/groups

# DEMO

Two colab notebooks are present in  
the GitHub Repo  
[github.com/matteoferla/Fragmenstein](https://github.com/matteoferla/Fragmenstein)



# THANK YOU FOR LISTENING



matteo.ferla@stats.ox.ac.uk  
@matteoferla  
/matteoferla

## UNIVERSITY OF OXFORD

- Charlotte Deane (OPIG)
- Rubén Sanchez-Garcia, Rachael Skyner, Frank von Delft, Brian Marsden (CMD)
- Jenny Taylor (BRC genomic medicine subtheme)
- Harold Grosjean (Biggin group)

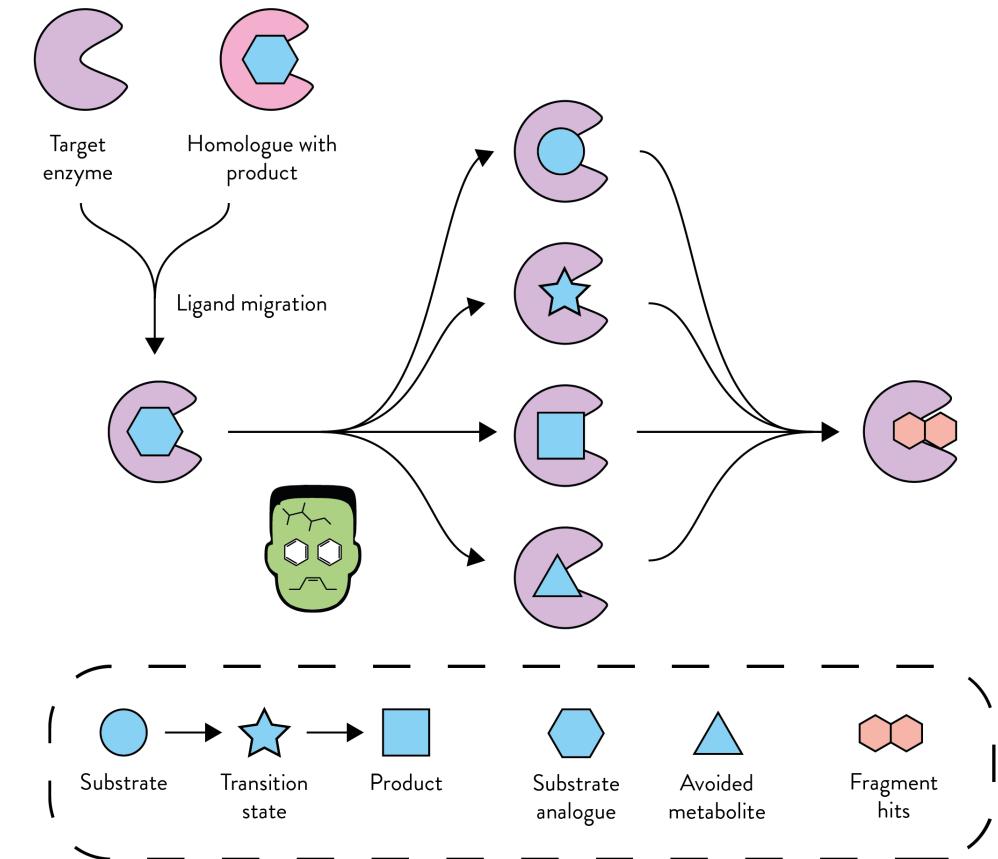
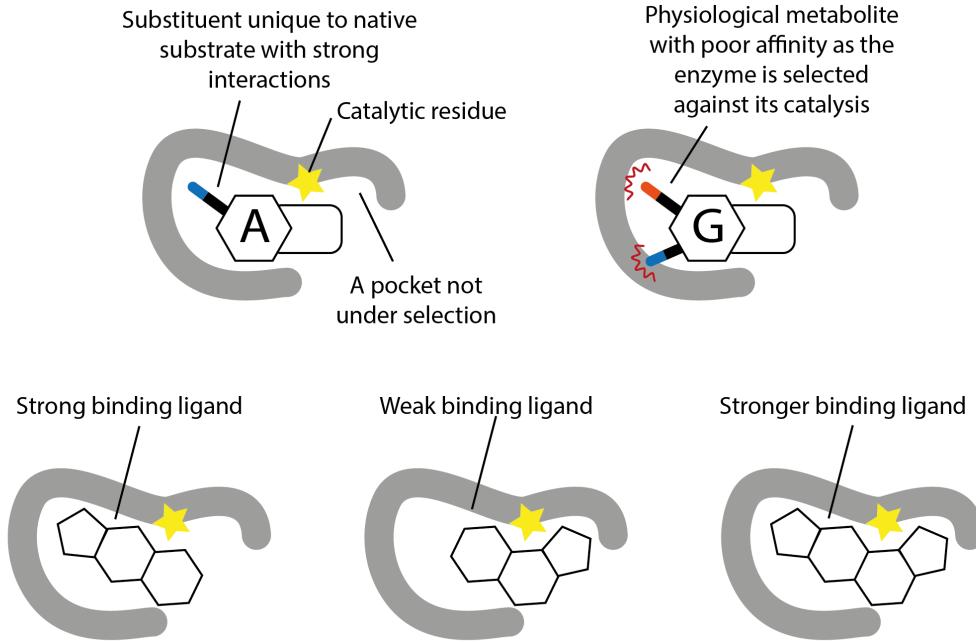
## UNIVERSITY OF CALIFORNIA — SAN FRANCISCO

- Stefan Gahbauer (Schochet group)

**CONSULTANT**  
Atlas

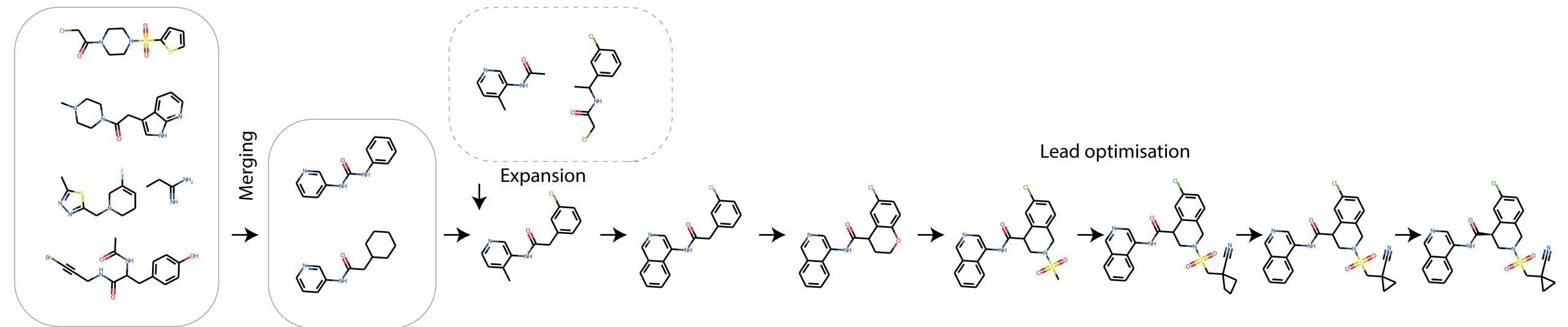


# FULL CATALYTIC REACTION?



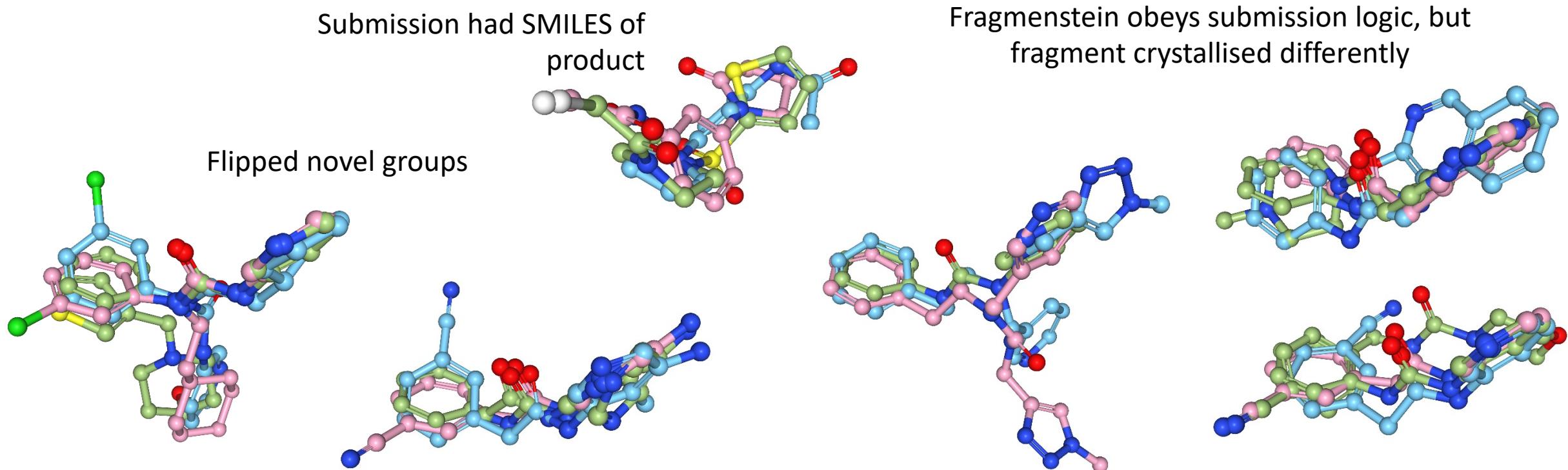
# THE COVID MOONSHOT

- SAR-CoV-2 protease drug campaign
- Pan-Dataset Density Analysis (PanDDA) method
- Collaborative: public call for follow-up compounds
- Diverse strategies



# DATASET SCORES IN PERSPECTIVE

- Fragmenstein: 28% less than 1 Å & 56% less than 2 Å
- 64% of follow-up compounds preserved the binding of the inspiration (<2 Å). Below green=inspirations, blue=crystal, puce=Fragmenstein

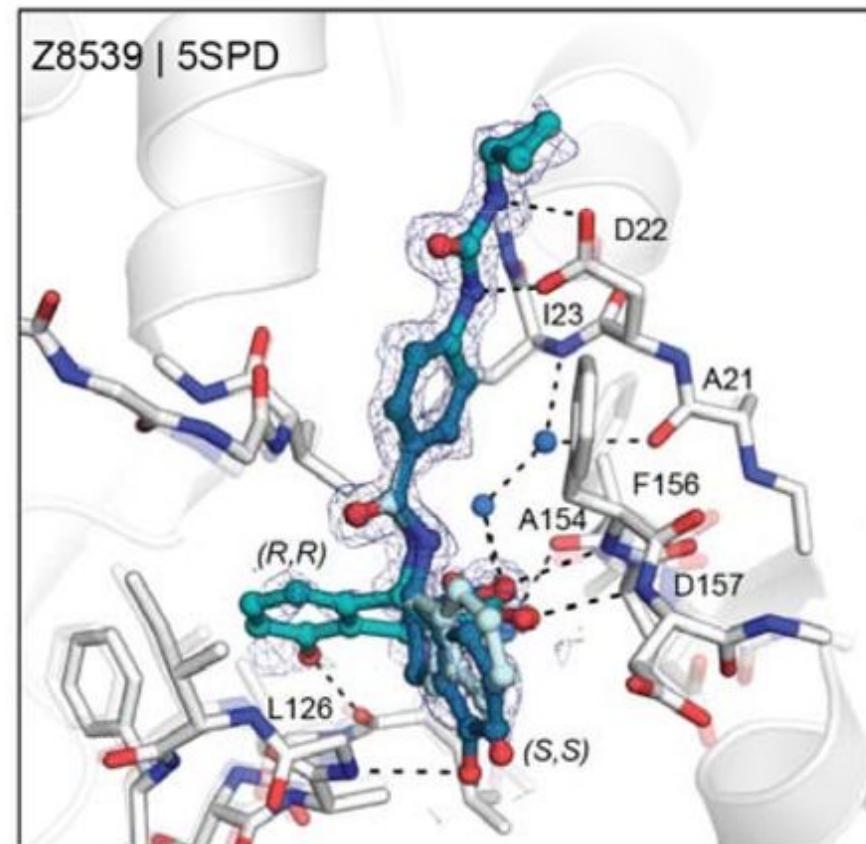


# GAHBAUER ET AL. CONT'D



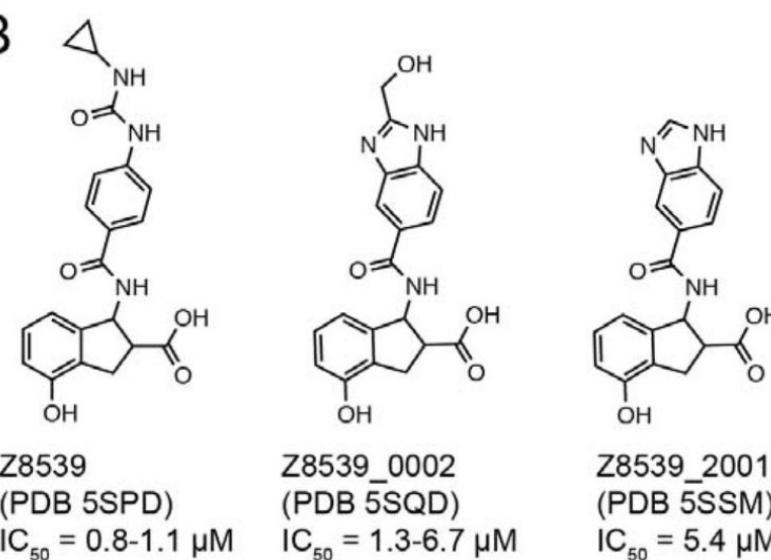
Diverted  
traffic

E



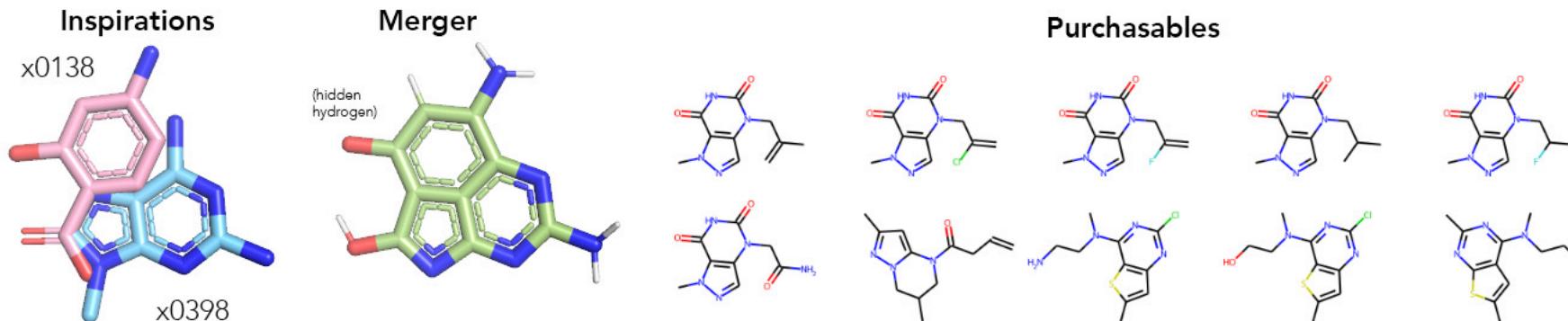
- enantiomers
- Carboxylic group isostere issue
- Cornered in make-on-demand space  
(acyl chloride chemistry barred)

B



# DISTANCE IN MAKE-ON-DEMAND SPACE

- Fragmenstein mergers often have distant analogues in Enamine REAL
- Substructure searches will be more focused on the core rather than the substituents
- In Gahbauer et al. preprint, submicromolar compound was found but was isolated in make-on-demand space, i.e. poor ‘elaborability’



See [michelangelo.sgc.ox.ac.uk/r/fragmenstein\\_nsp3](https://michelangelo.sgc.ox.ac.uk/r/fragmenstein_nsp3) for more