The 9th International Electronic Conference on Medicinal Chemistry (ECMC 2023)

01–30 November 2023 | Online

Sequence and pocket conservation across SARS-CoV-2 non-structural proteins - design of future therapeutics

Chaired by **Dr. Alfredo Berzal-Herranz** and **Prof. Dr. Maria Emília Sousa**

pharmaceuticals MDPI

Shubhangi Kandwal [1] and **Darren Fayne [2*]**

[1] Molecular Design Group, School of Biochemistry and Immunology, Trinity College Dublin, Ireland

[2] Senior Research Fellow and PI, Molecular Design Group, School of Biochemistry and Immunology, Trinity College Dublin, Ireland

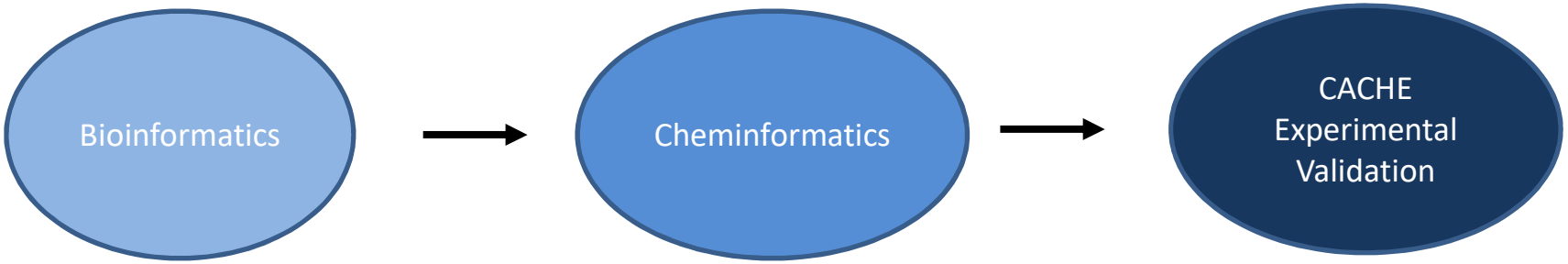
***** Corresponding author: fayned@tcd.ie

CACHE CHALLENGE ICHEC Irish Centre for High-End Computing TimoLassmann/ kalign Chemical Computing Group IRISH RESEARCH COUNCIL An Chomhairle um Thaighde in Éirinn Trinity College Dublin Coláiste na Tríonóide, Baile Átha Cliath The University of Dublin

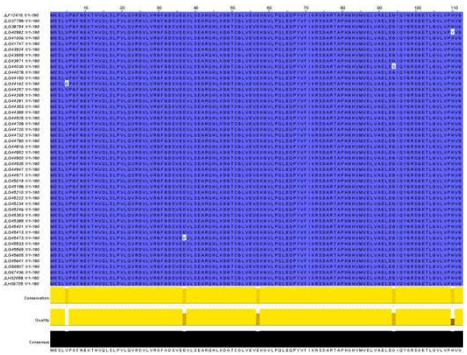
The 9th International Electronic Conference on Medicinal Chemistry
01–30 November 2023 | Online

Sequence and pocket conservation across SARS-CoV-2 non-structural proteins - design of future therapeutics

Graphical Abstract

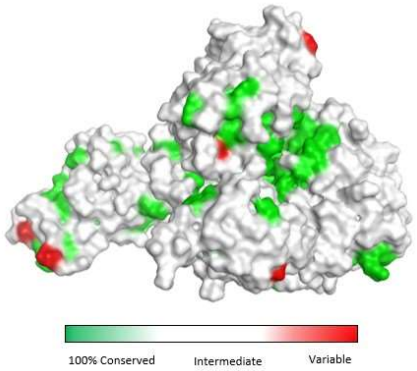


Phase 1



Nsp3 sequence alignment

Phase 2



Nsp3 tiered VS workflow to screen Enamine database

Phase 3

Abstract:

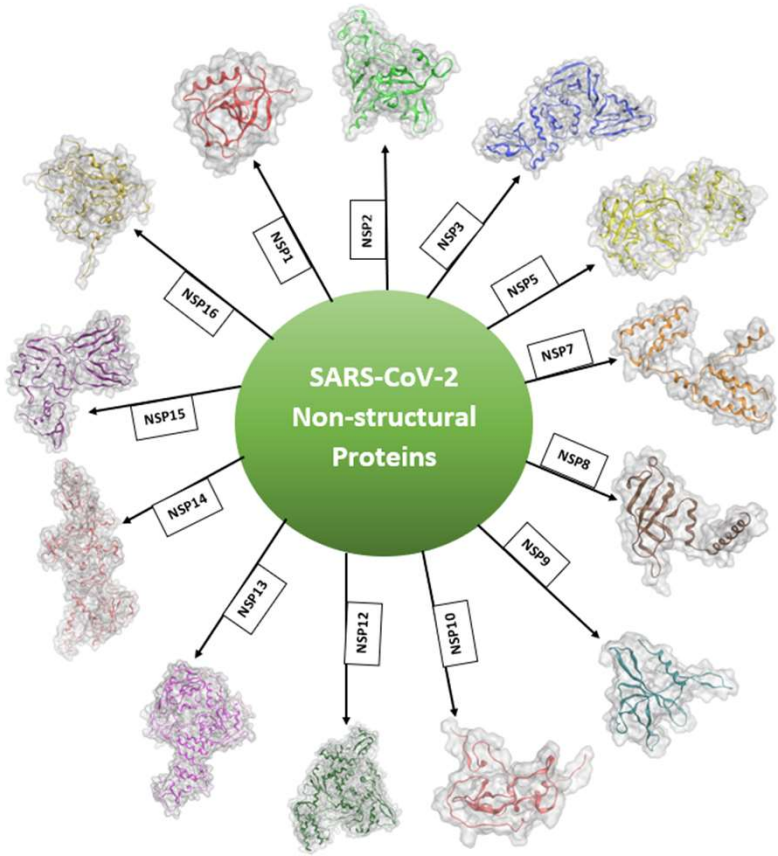
As new medications are used to treat COVID-19, many studies have reported that proteins such as spike, polymerase and proteases are prone to high levels of mutation that can create resistance to therapy over time. Thus, it becomes necessary to, not only target other viral proteins such as the non-structural proteins (nsp's), but to also target the most conserved residues of these proteins. A synergistic combination of bioinformatics, computer-aided drug-design and *in vitro* studies can feed into better understanding of SARS-CoV-2 (SC-2) and therefore help in the development of small molecule inhibitors against the nsp's.

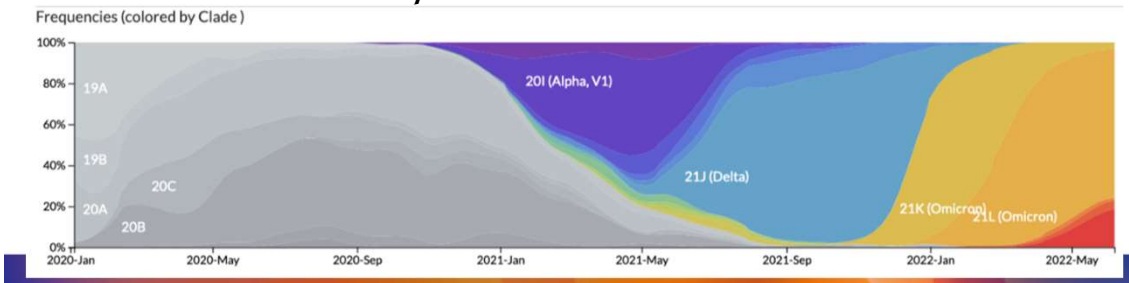
We have performed multiple sequence alignment studies on up to 11 million sequences of SC-2 Orf1ab to identify the most conserved residues. These residues were then visualized on 3D protein X-ray structures using MOE software. We found that there were known and novel binding pocket residues that were 100% conserved in our datasets. Our results indicate that these highly conserved pockets can be targeted for developing promising SC-2 inhibitors. Our group has recently been selected to enter two international challenges organized by the CACHE consortium to discover inhibitors of the RNA binding tunnel of SC-2 nsp13 and the Mac1 domain of SC-2 nsp3. We used a tiered screening workflow [volume/shape information of the binding pockets (fastROCS), in-house pharmacophore generation software (MoPBS/MOE) and docking in the binding pocket (FRED)] to rank hits for subsequent clustering and to identify compounds that bind to these conserved pockets. Our results on nsp3 will be presented here.

Keywords: CADD; drug design; SARS-CoV-2; Virtual Screening

Introduction

- The majority of SARS-CoV-2 (SC-2) therapeutic development work has focussed on targeting the spike protein, viral polymerase and proteases.
- As the pandemic progressed, many studies reported that these proteins are prone to high levels of mutation and can become drug resistant (Pacheti et al., 2020).
- Zhou et al., 2022 found that L50F and E166V mutations among others, showed up to 80-fold resistance to nirmatrelvir (Paxlovid- 3CLpro inhibitor) along with high reproductive fitness as compared to wild-type.
- Therefore, targeting other proteins involved in viral replication, in particular, targeting the most conserved residues present in the non-structural proteins (nsp's) becomes necessary.

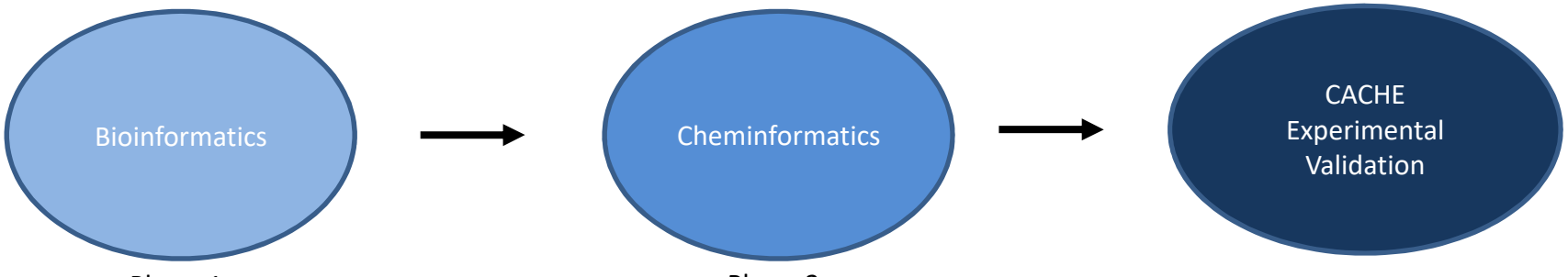


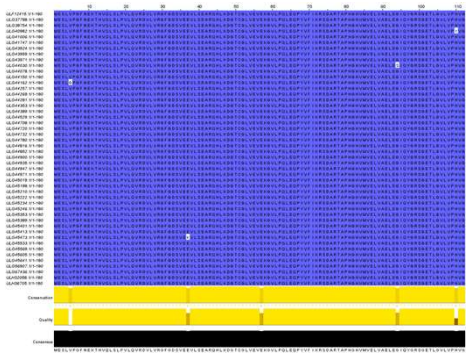


GISAID, 2022

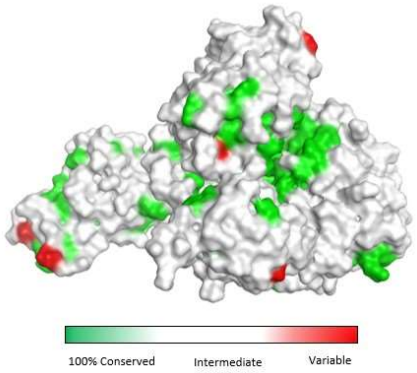
Approach

- Our group has recently been selected to enter two international challenges **(ranked in the top 4 globally)** organized by CACHE to computationally design inhibitors for the **RNA binding tunnel of SC-2 nsp13 (Challenge #2) and the Mac1 domain of SC-2 nsp3 (Challenge #3).**





NSP sequence conservation within SARS-CoV-2 (Multiple sequence alignments using Kalign). We have identified conserved hotspots across nsp's in our datasets ~93,000, ~200,000, ~1 million (NCBI-database) and ~11 million (GISAID-database) sequences.



Tiered virtual screening (VS) workflow to screen molecules from the Enamine database

Nsp3 – Challenge #3

- CACHE #3 challenge : Finding ligands targeting the macrodomain of SC-2 nsp3 that compete with the substrate, ADP-ribose (ADPr).
- Viral macrodomains are believed to counter or hijack host immunity by reversing the mono(ADP-ribosyl) modifications generated by host PARP enzymes, thereby interfering with interferon production (Grunewald et al., 2019).
- Mac1 domain is highly conserved across the SC-2 sequences.
- The mac1 conserved domains are also present in several positive-sense ssRNA (+ssRNA) viruses of the families *Hepeviridae*, *Togaviridae*, and *Coronaviridae*, such as hepatitis E virus (HEV), alphavirus, rubella virus, and all coronaviruses (Koonin et al., 1992, Snijder et al., 2003).

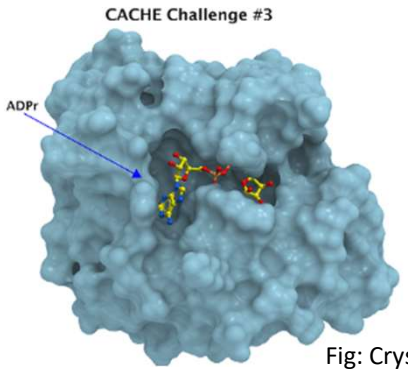


Fig: Crystal structure of Mac1 domain of Nsp3 (SC-2) co-cystalized with ADPr in the binding pocket

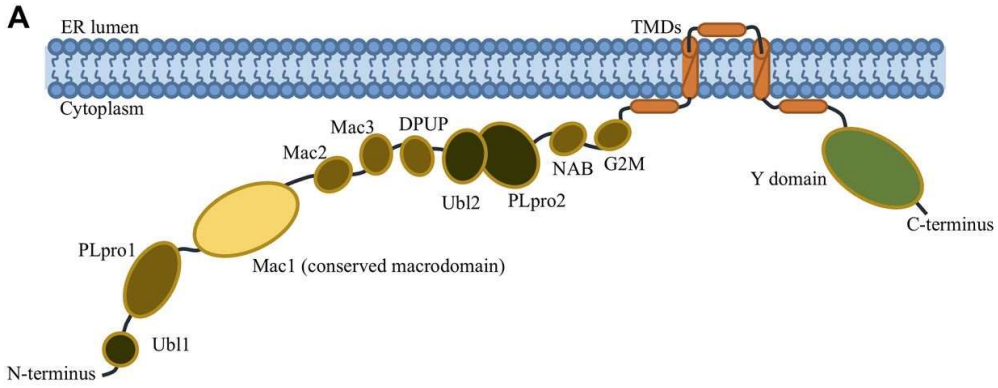


Fig: Nsp3 of SC-2 with all the domains Alhammad et al., 2021

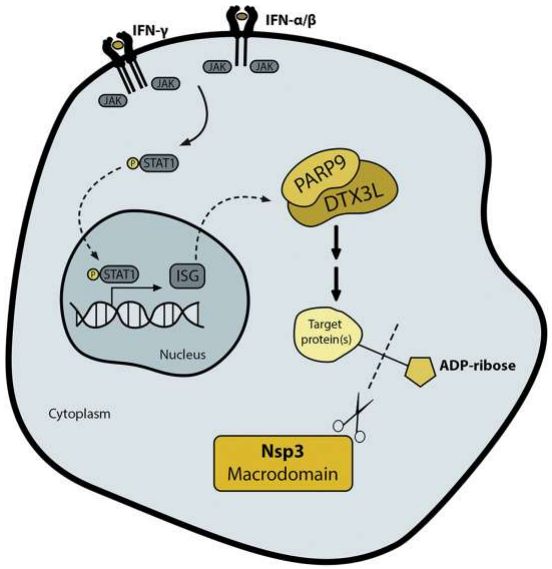


Fig: Mac1 domain of Nsp3 hijacking host immunity Russo et al., 2021

Phase 1 (Bioinformatics)

- The Mac1 domain of nsp3 is ~170 amino acid (AA) in length.
- The entire Mac1 domain of nsp3 is highly conserved across **~1.4 million** SC-2 sequences **(Conservation score: 98.99% - 99.99%).**
- The key residues responsible for binding of ADP-ribose and other related adenosine-derivatives with nsp3-macrodomain X are conserved across beta-CoVs **(Kandwal & Fayne, 2023).**

```
YP_009047231.1_MERS     EQTQNVTVKPKRLRKKRNVDPLSNFEHKVITECVTIVLGDAIQVAKCYGESVLVNAANTH   296
YP_009742610.1_SC2      ----------------TIEVNSFSGY--LKLTDNVYIKNADIVEEAKKVKPTVVVNAANVY   246
YP_009944368.1_SC1      ----------------EEPVNQFTGY--LKLTDNVAIKCVDIVKEAQSANPMVIVNAANIH   224
                                          *: ::.:    :*: * *   * :: *:     *:***** :

YP_009047231.1_MERS     LKHGGGIAGAINAASKGAVQKESDEYILAKGPLQVGDSVLLQGHSLAKNILHVVGPDARA   356
YP_009742610.1_SC2      LKHGGGVAGALNKATNNAMQVESDDYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNK   306
YP_009944368.1_SC1      LKHGGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGSCLLSGHNLAKKCLHVVGPNLNA   284
                        ******:***:* *:::.*:* ***:**  :*** **.* :*.**.***: ******: .

YP_009047231.1_MERS     KQDVSLLSKCYKAMNAYPLVVTPLVSAGIFGVKPAVSFDYLIREAKTRVLVVVNSQDVYK   416
YP_009742610.1_SC2      GEDIQLLKSAYENFNQHEVLLAPLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYD   366
YP_009944368.1_SC1      GEDIQLLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQTVRTQVYIAVNDKALYE   344
```

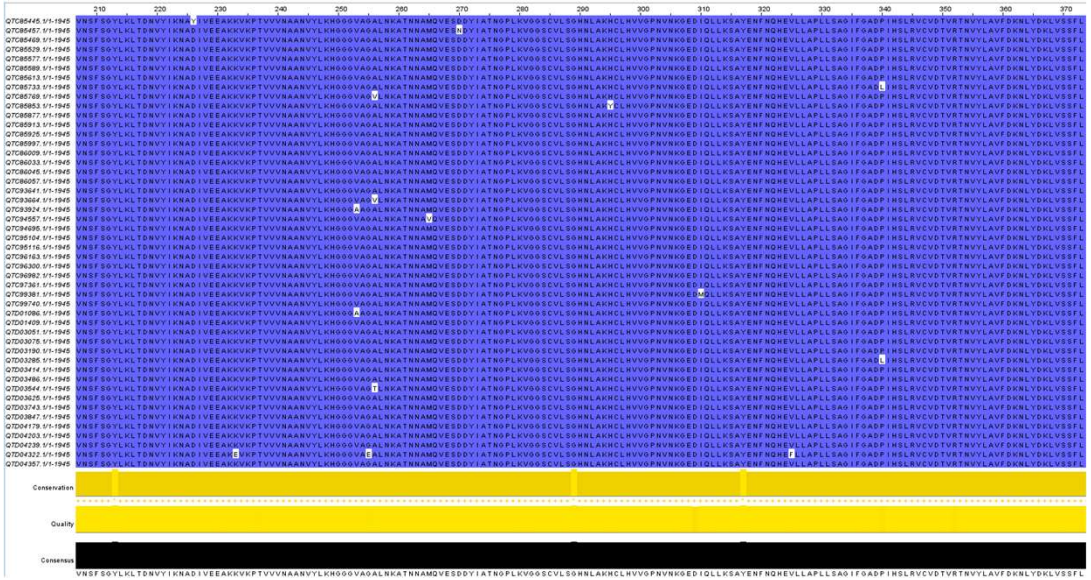


Fig: MSA of nsp3 (Mac1 domain) sequences of SC-2 (~1.4 million sequences - NCBI). Highlighted (indigo) residues are conserved across these sequences.

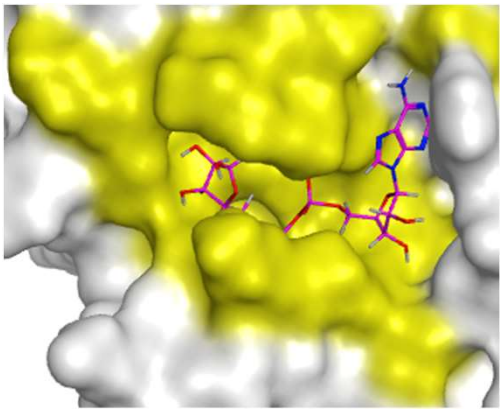


Fig: MSA of nsp3 sequences of MERS-CoV, SC-1 and SC-2. Highlighted (yellow) residues are a part of the ATPr binding pocket that are conserved across these sequences.

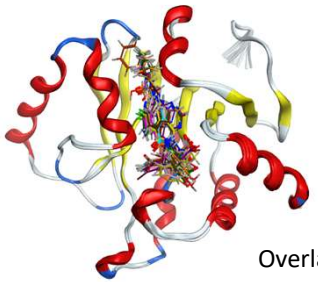
Kandwal, S.; Fayne, D. Genetic Conservation across SARS-CoV-2 Non-Structural Proteins – Insights into Possible Targets for Treatment of Future Viral Outbreaks. *Virology* **2023**, *581*, 97–115

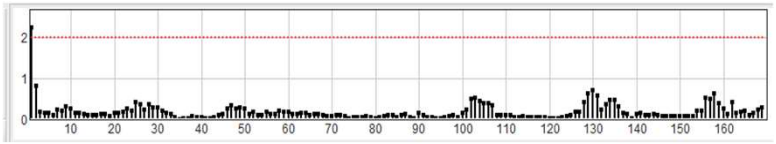
Phase 2 (Cheminformatics)

1. X-ray crystal structures of hit compounds in the binding pocket of Mac1 domain (Gahbauer et al., 2022).

2. These structures were overlaid with MOE with overall RMSD value of 0.23 Å.

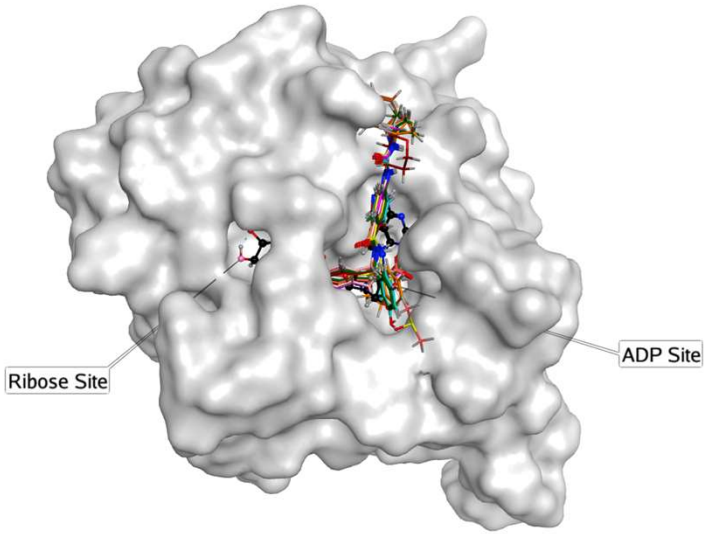
3. The graph below indicate that there was minimal amino acid (AA) movement within the X-ray structures.



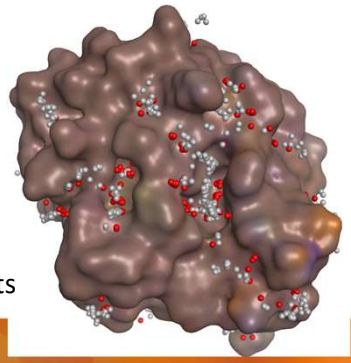


RMSD plot of overlaid protein structures

Overlaid protein structures (co-crystalised with lead-like molecules)

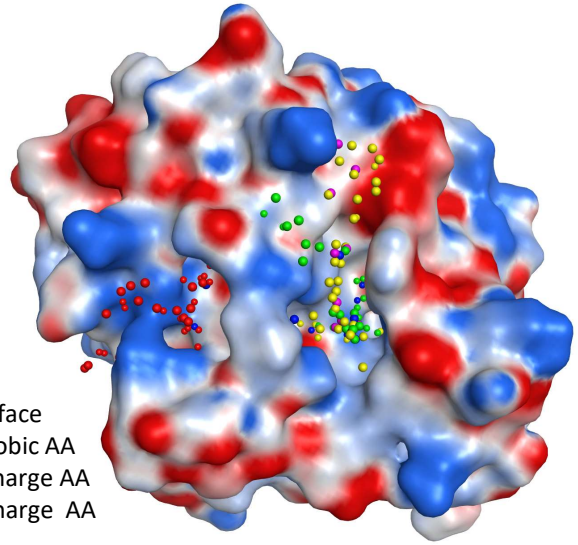


4. Site finder in MOE: Predicts putative binding pockets

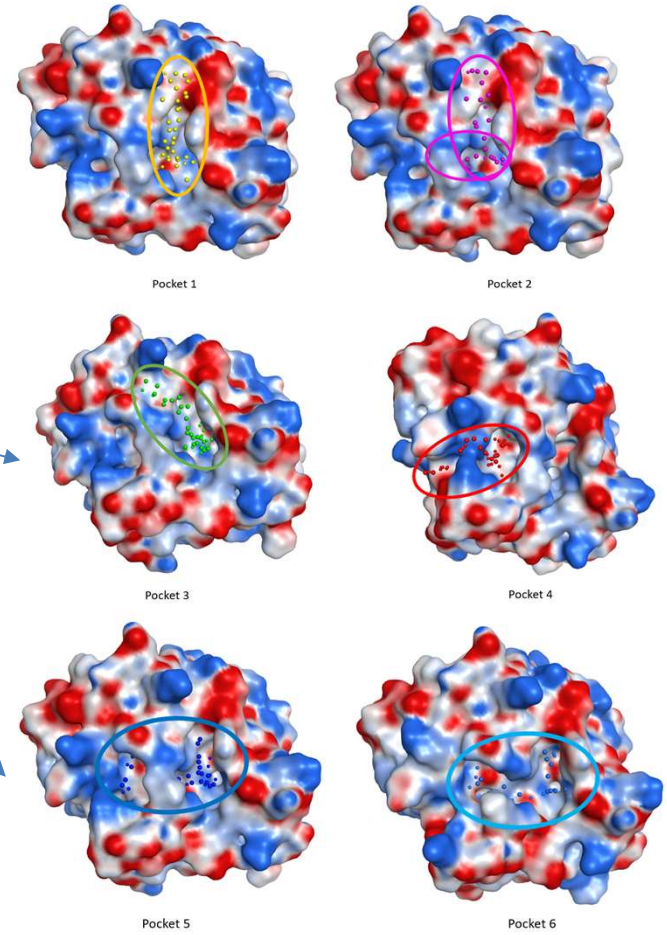


Red spheres: hydrophilic cavity points
Grey spheres: hydrophobic cavity points

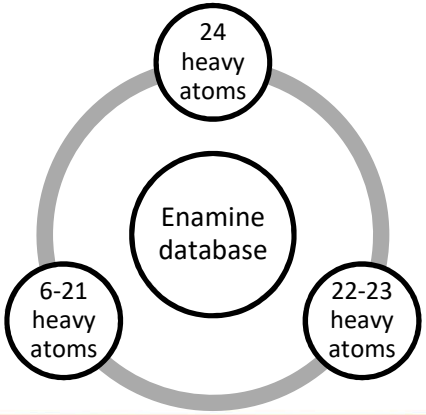
5. Site finder on MOE: Pocket selection



Electrostatic surface
White: hydrophobic AA
Red: Negative charge AA
Blue: Positive charge AA



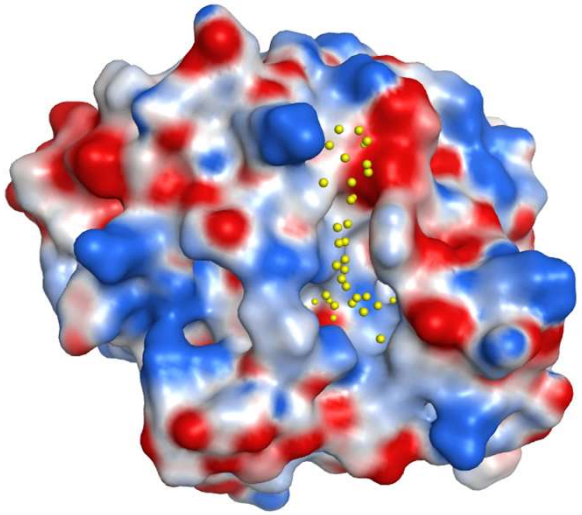
6. Database: Enamine



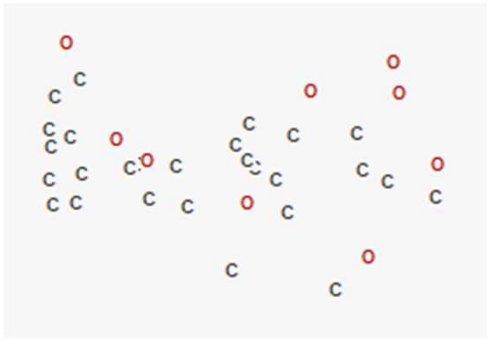
7. Workflow

Enamine Database (24HA, 23HA, 06-21HA) → Conformation generation (5) → Pocket selection after dummy atoms (MOE) → Dummy atoms volume queries (sdf format) → fastROCS (1st round screening) → Pharmacophore search (2nd round refine screening) → Filtration (logP, MW, PAINS, carboxylic acid) → Conformation generation (100) → FRED docking (3rd round ultra refine screening) → Clustering (Pipeline Pilot) Clusters for each pocket → Visual inspection (selection of 150 hits)

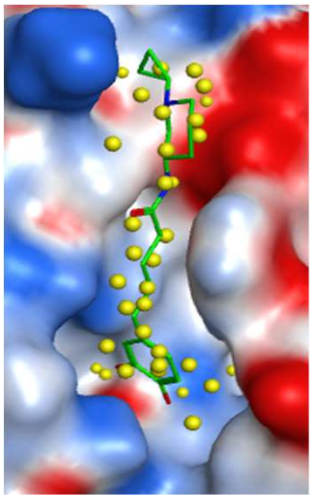
A. fastROCS volume query search using dummy atoms (1st round of VS)



Pocket selection using Dummy atoms (yellow)



Dummy atoms volume query (carbon-hydrophobic and oxygen-hydrophilic)

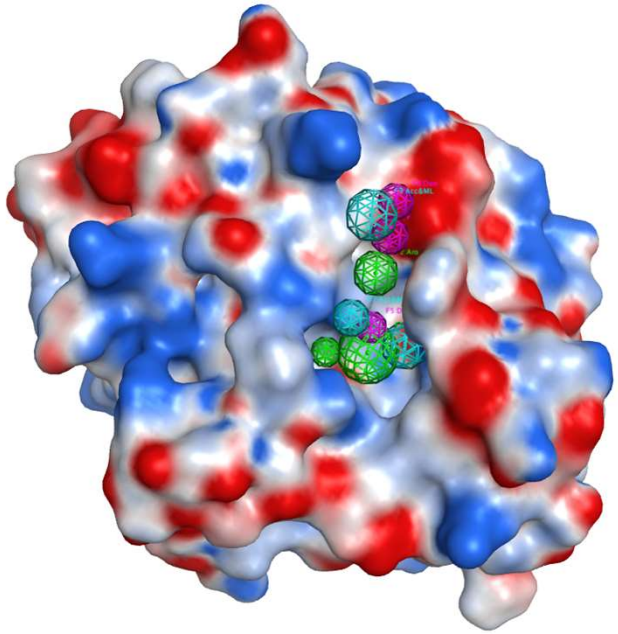


Pocket dummy atoms with a predicted fastROCS hit

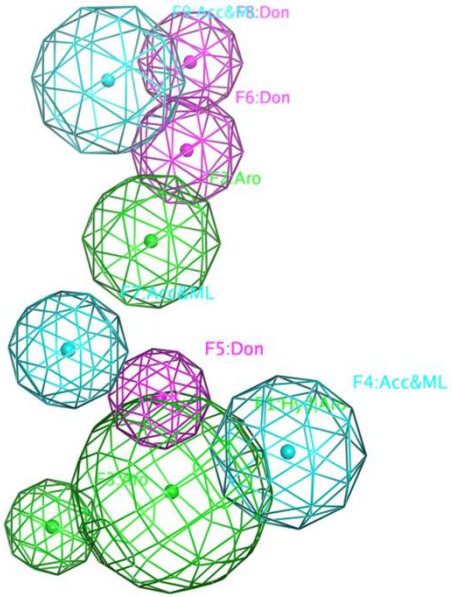
The 9th International Electronic Conference on Medicinal Chemistry
01–30 November 2023 | Online
MDPI

Enamine Database (24HA, 23HA, 06-21HA) → Conformation generation (5) → Pocket selection after dummy atoms (MOE) → Dummy atoms volume queries (sdf format) → fastROCS (1st round screening) → Pharmacophore search (2nd round refine screening) → Filtration (logP, MW, PAINS, carboxylic acid) → Conformation generation (100) → FRED docking (3rd round ultra refine screening) → Clustering (Pipeline Pilot) Clusters for each pocket → Visual inspection (selection of 150 hits)

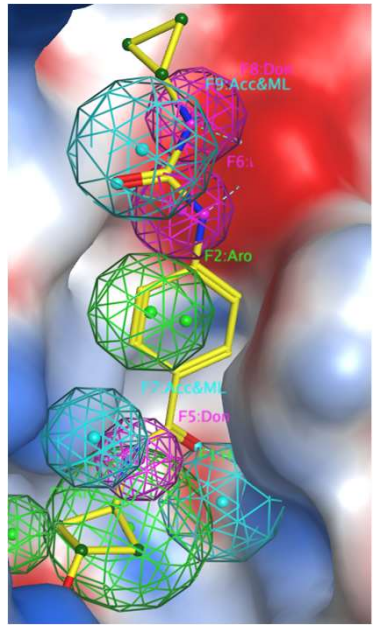
B. MOE pharmacophore search on fastROCS hits (2nd round)



Pocket 1 with Pharmacophore features



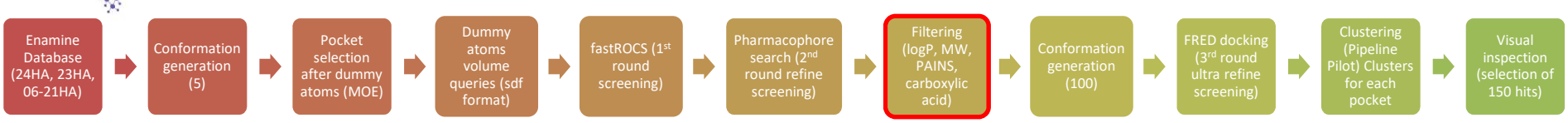
Pharmacophore features generated using Consensus Ph4 (overlay of 22 PDB structures) on MOE



Pharmacophore features mapping with a hit

Note: Pharmacophore features were predicted using Consensus Ph4 on MOE for pocket 1.
Whereas, for rest of the pocket an in-house software called MoPBS **(Braun & Fayne, 2022)** was used to predict the Ph4 features.

Braun, J.; Fayne, D. Mapping of Protein Binding Sites Using Clustering Algorithms - Development of a Pharmacophore Based Drug Discovery Tool. *Journal of Molecular Graphics and Modelling* **2022**, *115*,108228



Filtering

- MOE database viewer was used to filter the hit lists (prior to docking).
- Molecules having MW > 400 Da, logP > 3.5, Pan-assay interference compounds (PAINS) and carboxylic acids.

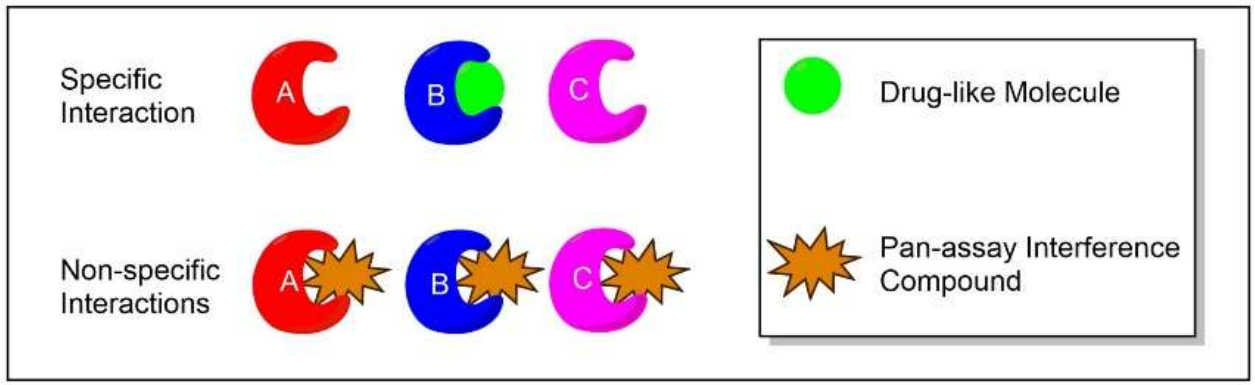
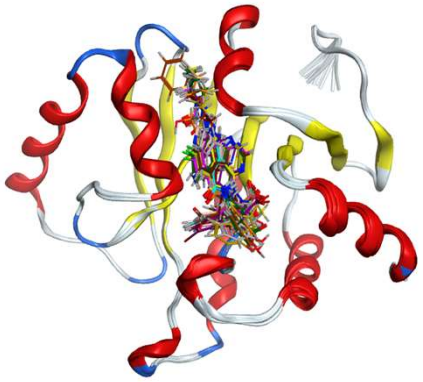


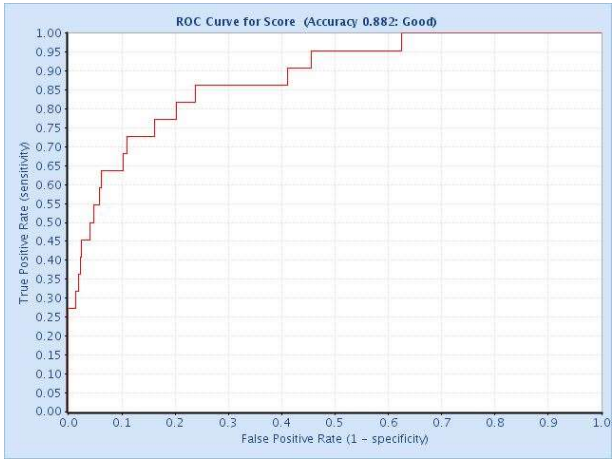
Diagram depicting a representative pan-assay interference compound. The drug-like molecule specifically interacts with target B, but the PAINS-like compound non-specifically interacts with multiple targets. Figure taken from Wikipedia.

8. Positive control docking to check the efficiency of the docking software

- The efficiency of the docking software was checked using DataPype (in-house software)
- Decoys (inactives) were generated using the DUD-E protocol
- **Haystack (1122): Actives (22) + Decoys (1100)**
- FRED docking software was able to rank the actives (needle) at the top
- ROC-AUC was 0.88

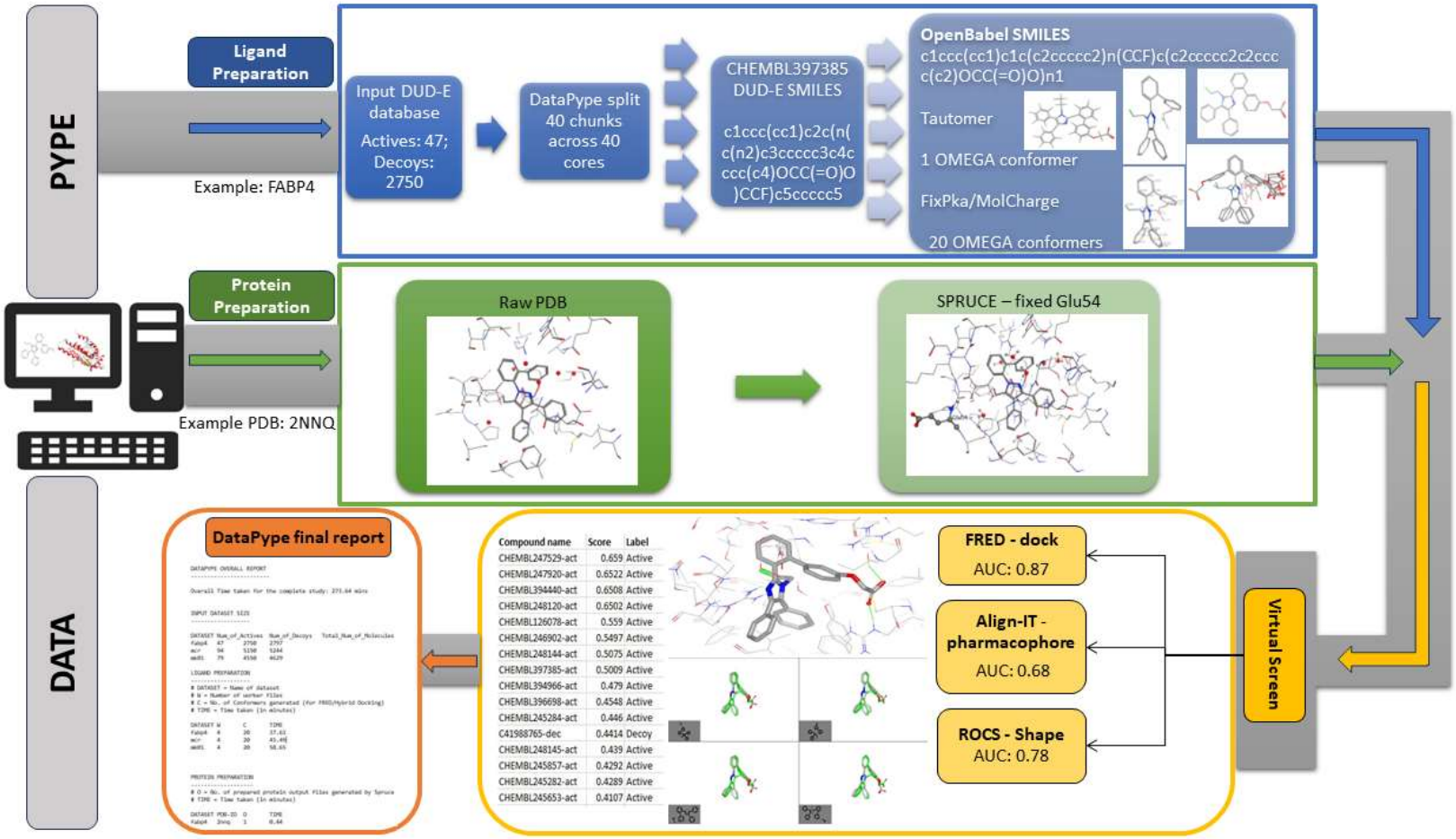
Title	FRED Chemgauss4 score	Label
5SRZ	-11.880444	Active
5SRL	-11.570402	Active
5SQ6	-11.516068	Active
5SRY	-11.296618	Active
5SS9	-11.185315	Active
C96451909	-10.907855	Decoy
5SQO	-10.483145	Active
C08758435	-10.143887	Decoy
C50672345	-10.125989	Decoy
C59487988	-10.073594	Decoy
C06454265	-9.577257	Decoy
C05386036	-9.550199	Decoy
C01794214	-9.467897	Decoy
C36378815	-9.355917	Decoy
C10217688	-9.325662	Decoy
C04869773	-9.322695	Decoy
C21334060	-9.312538	Decoy
C10217685	-9.311275	Decoy





The 9th International Electronic Conference on Medicinal Chemistry
01–30 November 2023 | Online

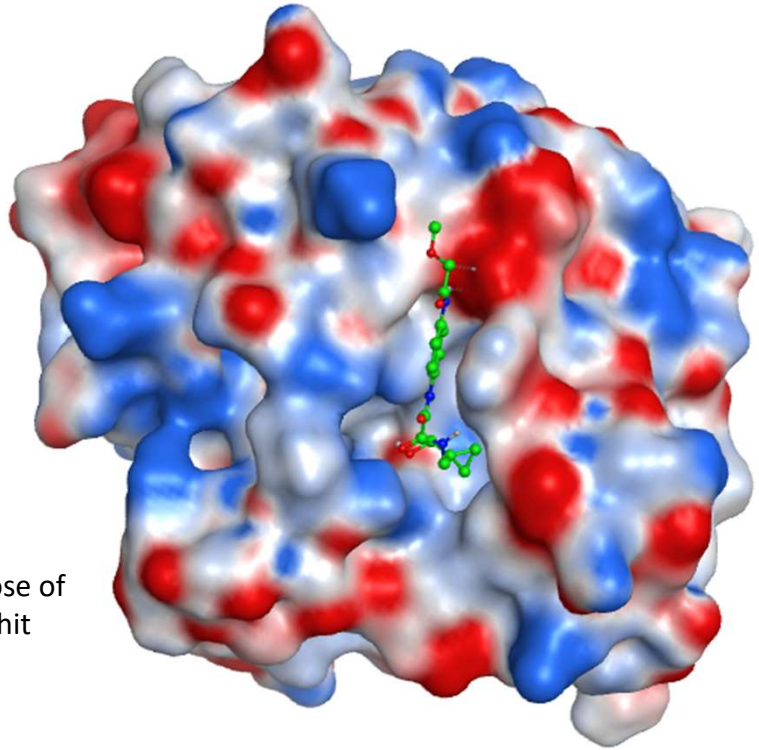
DataPype overview



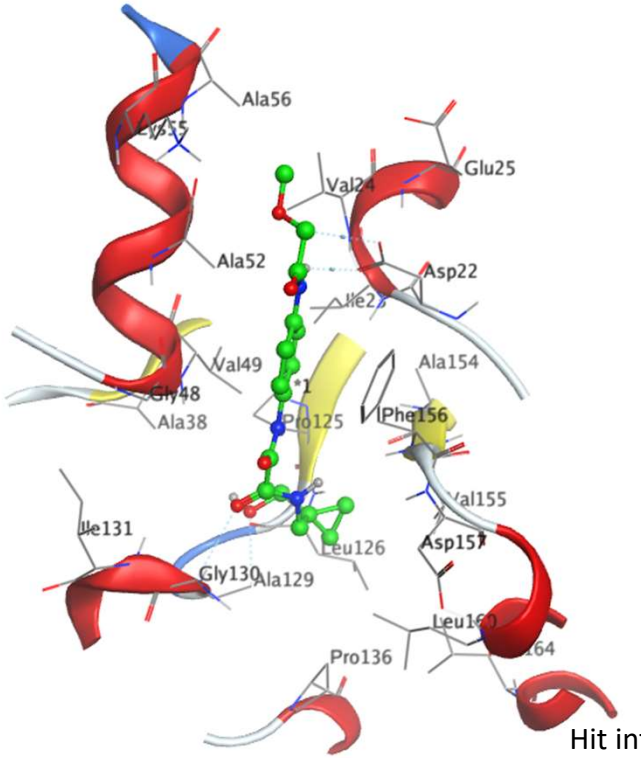
Khan, M. F.; Kandwal, S.; Fayne, D. DataPype: A Fully Automated Unified Software Platform for Computer-Aided Drug Design. *ACS Omega* **2023**, acsomega.3c05207. https://doi.org/10.1021/acsomega.3c05207.

Enamine Database (24HA, 23HA, 06-21HA) → Conformation generation (5) → Pocket selection after dummy atoms (MOE) → Dummy atoms volume queries (sdf format) → fastROCS (1st round screening) → Pharmacophore search (2nd round refine screening) → Filtering (logP, MW, PAINS, carboxylic acid) → Conformation generation (100) → FRED docking (3rd round ultra refine screening) → Clustering (Pipeline Pilot) Clusters for each pocket → Visual inspection (selection of 150 hits)

C. FRED docking (3rd round of VS)



Docked pose of predicted hit



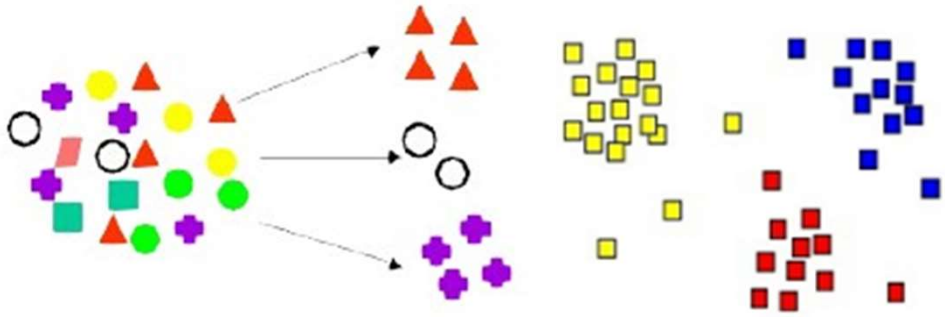
Hit interaction with AA residues

- The residues resulting in the interaction i.e., Asp22, Ala129 and Gly130 are conserved across Beta-CoVs.
- Asp22 (99.97%), Ala129 (99.94%) and Gly130 (99.96%) residues are also highly conserved across SC-2 sequences (~1.4 million sequences).

Enamine Database (24HA, 23HA, 06-21HA) → Conformation generation (5) → Pocket selection after dummy atoms (MOE) → Dummy atoms volume queries (sdf format) → fastROCS (1st round screening) → Pharmacophore search (2nd round refine screening) → Filtering (logP, MW, PAINS, carboxylic acid) → Conformation generation (100) → FRED docking (3rd round ultra refine screening) → Clustering (Pipeline Pilot) Clusters for each pocket → Visual inspection (selection of 150 hits)

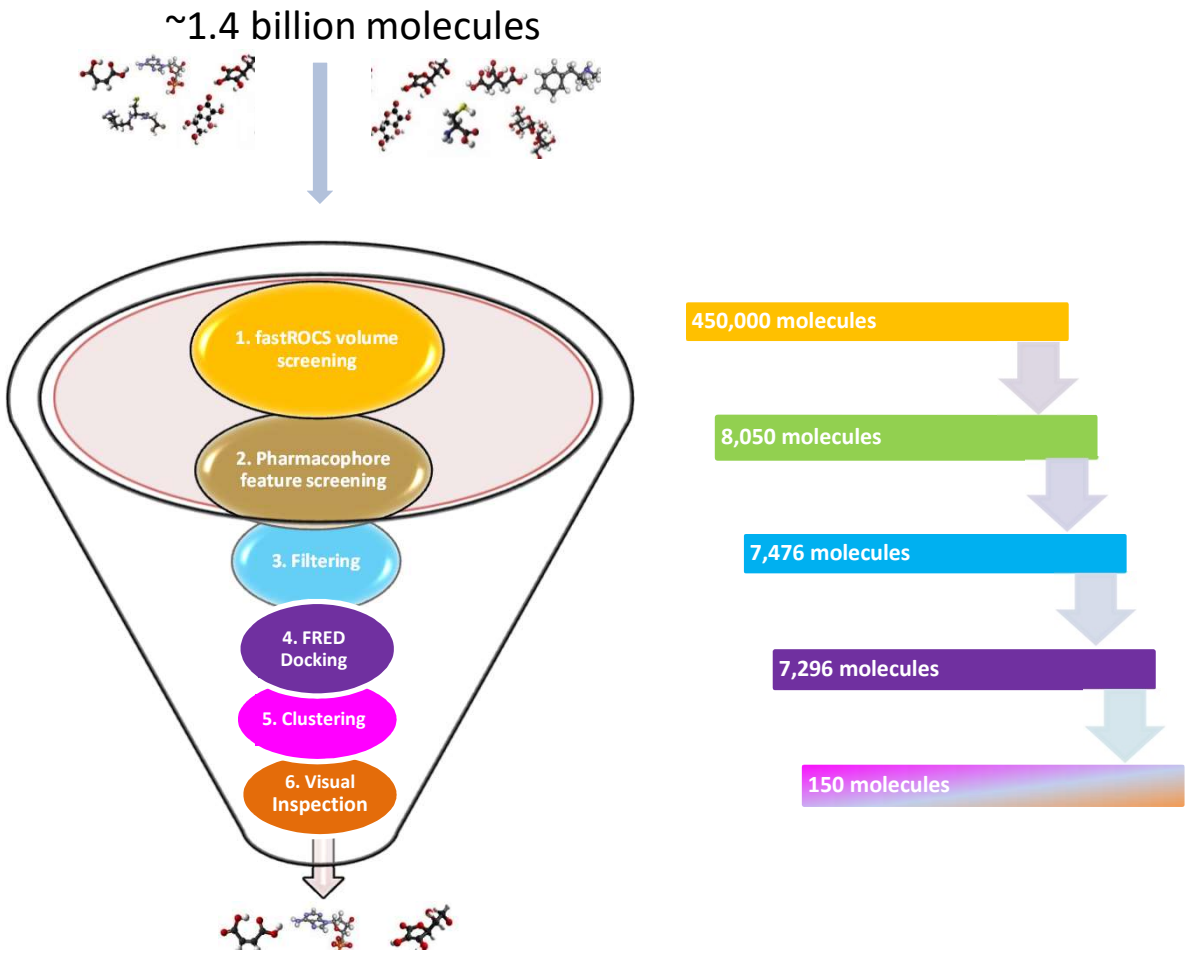
Clustering

- Clustering was performed using BIOVIA Pipeline Pilot (after docking) with ECFP4 fingerprint.
- Similar groups of compounds can be identified using clustering and from these clusters diverse set of representative compounds can be picked.



Examples of clustering
Figure taken from Data Mining - Clustering (Function|Model) | Data Mining | Datacadamia - Data and Co

MDG tiered virtual screening protocol



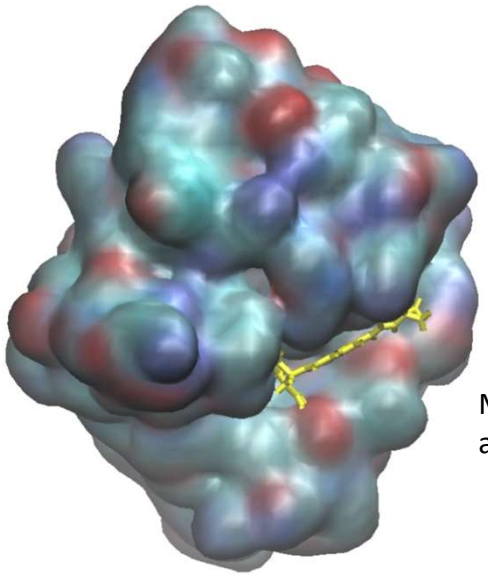
Conclusions

1. The bioinformatics analysis on SC-2 sequences suggested high level of conservation in the binding pocket of nsp3 (ADPr binding pocket).

2. Our tiered VS workflow has predicted hits for the binding pockets of nsp3 making interactions with conserved AA residues.

3. The experimental validation results of the predicted hits are expected in mid-November for nsp3.

4. If any of our hits show activity against nsp3, we will then work on model refinement i.e., create analogues followed by relative FEP calculations.



MD simulation of one of the actives (nsp3) using GROMACS

Future Directions

❖ My group also aims to target the conserved binding pockets of other SC-2 nsp's.

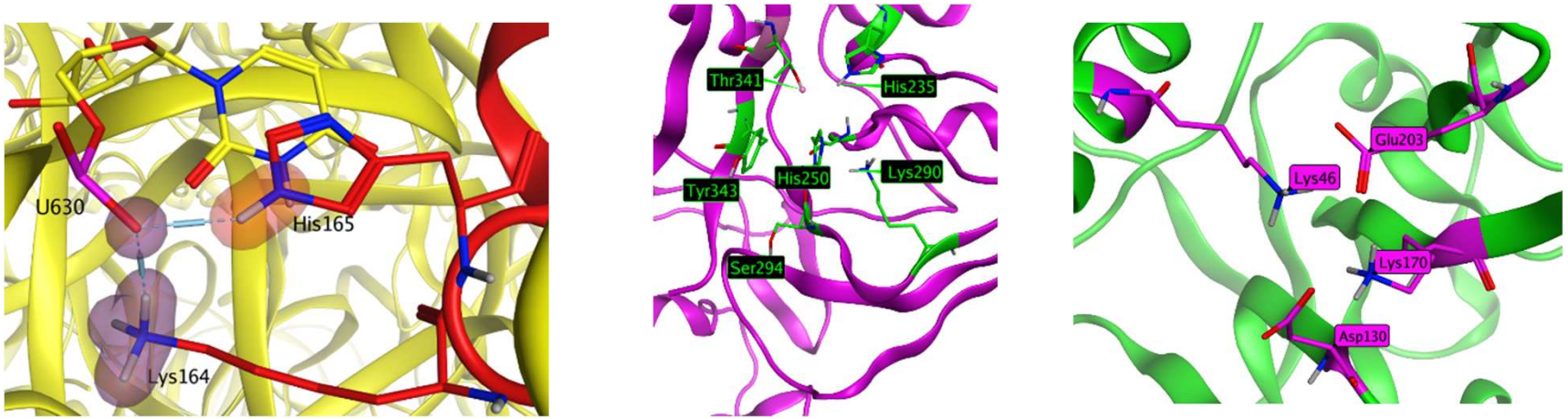


Fig: Conserved residues in nsp1, nsp15 and nsp16 respectively (**Kandwal & Fayne, 2023***)

A synergistic combination of bioinformatics, computer-aided drug-design and *in vitro* studies can feed into better understanding of SARS-CoV-2 (SC-2) and therefore help in the development of small molecule inhibitors against the nsp's.

As the inhibitors target the conserved AA residues, this can lead to the development of pan-coronavirus inhibitors.

* Kandwal, S.; Fayne, D. Genetic Conservation across SARS-CoV-2 Non-Structural Proteins – Insights into Possible Targets for Treatment of Future Viral Outbreaks. *Virology* **2023**, *581*, 97–115. https://doi.org/10.1016/j.virol.2023.02.011.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Acknowledgements

PhD candidate
Shubhangi Kandwal (IRC)

Molecular Design Group (MDG)
Dr Ivan Cmelo (IRC)
Cian Moore
Claire Ott

Smurfit Institute of Genetics
Dr Karsten Hokamp
Dr Fiona Mary Roche

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Thank You…

Contact me if you are interested in joining the molecular design adventure (and if you like pizza…)