Systems level understanding and regulating of the disease with atomic resolution

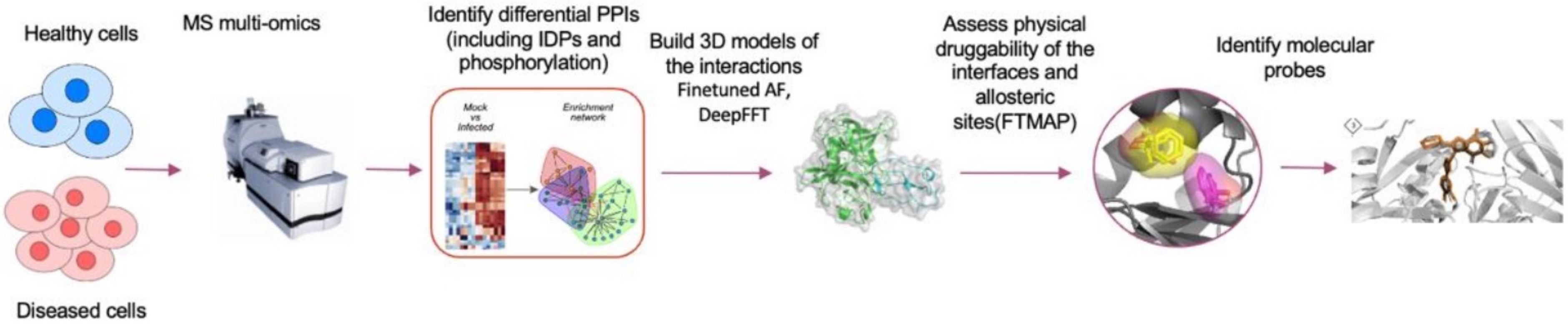
Dima Kozakov

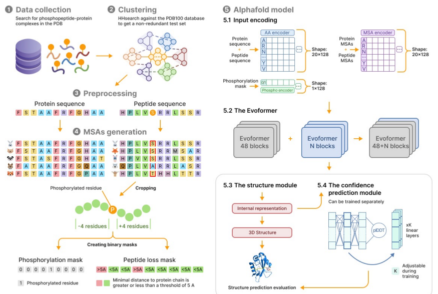
Laufer Center for Physical and Quantitative Biology

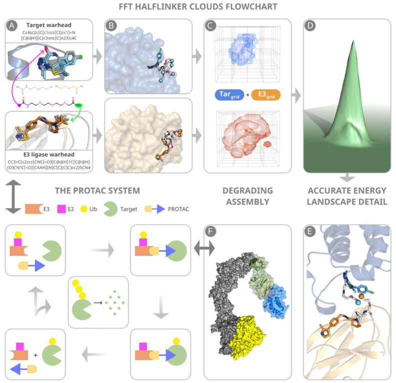
Stony Brook University

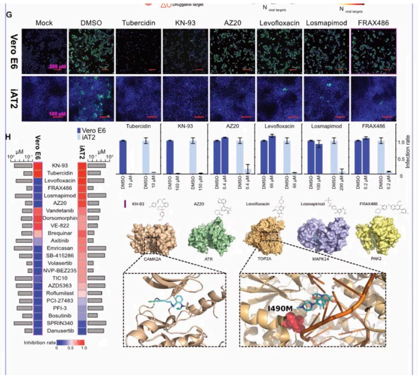
Stony Brook University

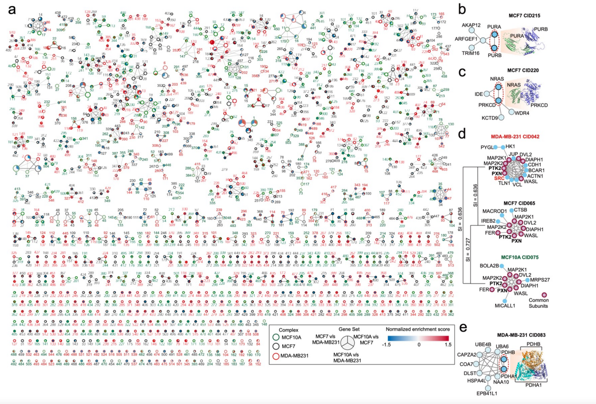
LAUFER CENTER FOR PHYSICAL AND QUANTITATIVE BIOLOGY

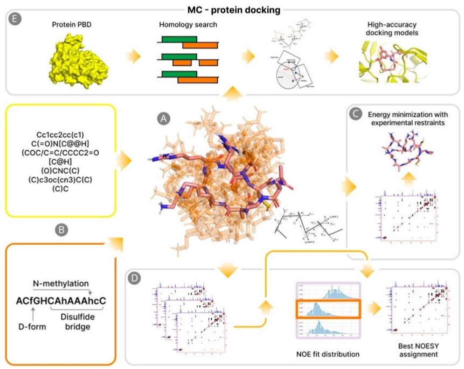

Healthy cells
Diseased cells
MS multi-omics
Identify differential PPIs (including IDPs and phosphorylation)
Mock vs Infected
Enrichment network
Build 3D models of the interactions Finetuned AF, DeepFFT
Assess physical druggability of the interfaces and allosteric sites(FTMAP)
Identify molecular probes

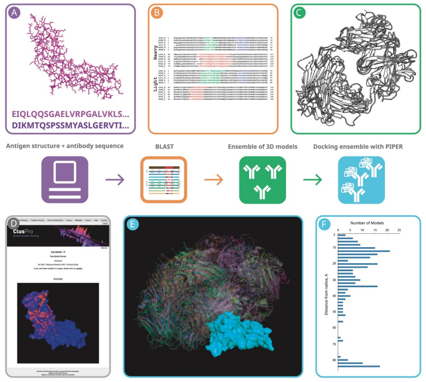

Data collection
Clustering
Alphafold model
Preprocessing
MSAs generation
5.2 The Evoformer
5.3 The structure module
5.4 The confidence prediction module

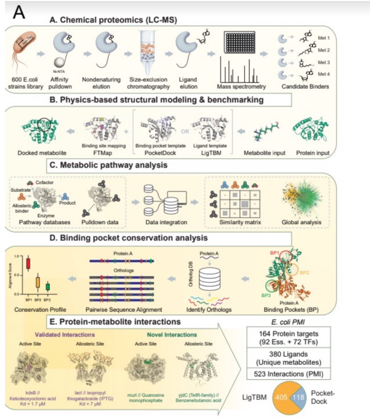

FFT HALFLINKER CLOUDS FLOWCHART
THE PROTAC SYSTEM
DEGRADING ASSEMBLY
ACCURATE ENERGY LANDSCAPE DETAIL

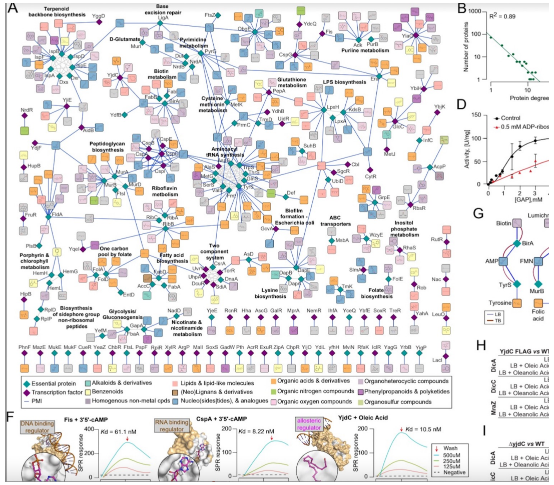





MC - protein docking
Protein PBD
Homology search
High-accuracy docking models
Energy minimization with experimental restraints
N-methylation
Disulfide bridge
D-form
NOE fit distribution
Best NOESY assignment

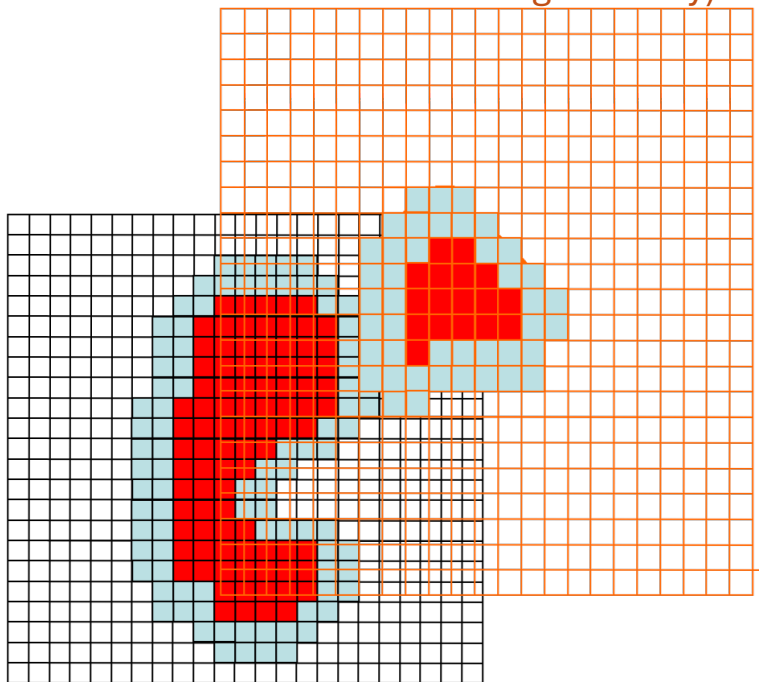

Antigen structure + antibody sequence
BLAST
Ensemble of 3D models
Docking ensemble with PIPER


A. Chemical proteomics (LC-MS)
B. Physics-based structural modeling & benchmarking
C. Metabolic pathway analysis
D. Binding pocket conservation analysis
E. Protein-metabolite interactions



FFT convolution approach enables global exhaustive macromolecular interaction sampling

Ligand representation (e.g. charge density)



Receptor representation (e.g. electrostatic potential)

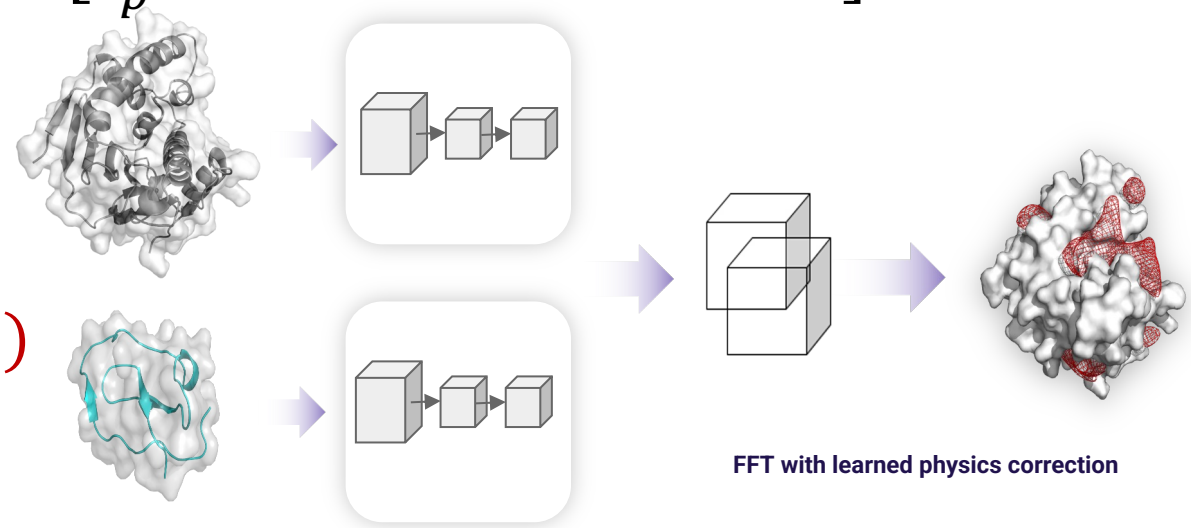
Interaction energy as a sum of FFT convolutions

$$E(\bar{t},r) = \sum_{p}\sum_{\bar{x}} R_p(\bar{x})L_{pr}(\bar{t}-\bar{x})$$

$$E(\bar{t},r) = IFT\left[\sum_{p} FT^{*}\{R_p(\bar{x})\}FT\{L_{pr}(\bar{x})\}\right]$$

$$O(N^6) \longrightarrow O(N^3 lnN^3)$$

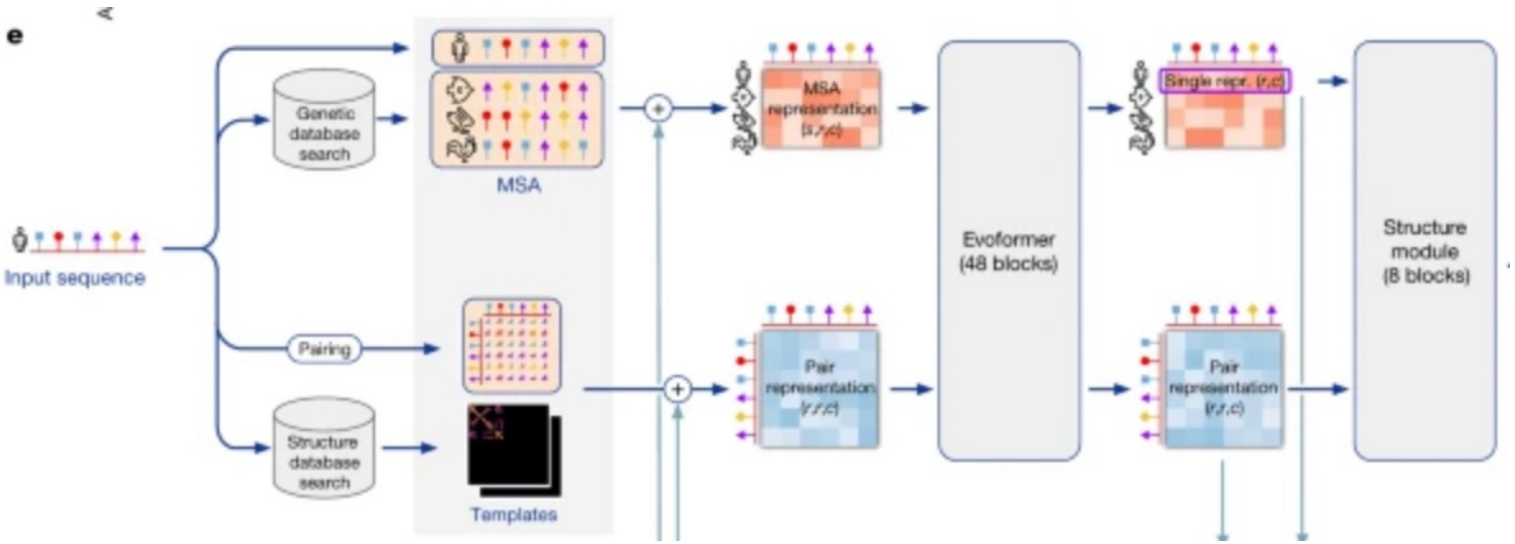
$$\begin{aligned}
E_{full} &= E_{vdw} + E_{elex} + E_{pair} \\
E_{vdw} &= E_{rep} + E_{attr} \\
E_{elec} &= \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \frac{q_i q_j}{\left(r_{ij}^2 + D^2 exp\left(-\frac{r_{ij}^2}{4D^2}\right)\right)^{1/2}} \\
E_{pair} &= \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \varepsilon(i,j) = \begin{cases} 0, r_{ij} > D \\ \varepsilon_{ij}, d < r_{ij} < D \end{cases} \\
& \varepsilon_{ij} = \sum_t \varepsilon_{ti}\lambda_t\varepsilon_{tj}
\end{aligned}$$

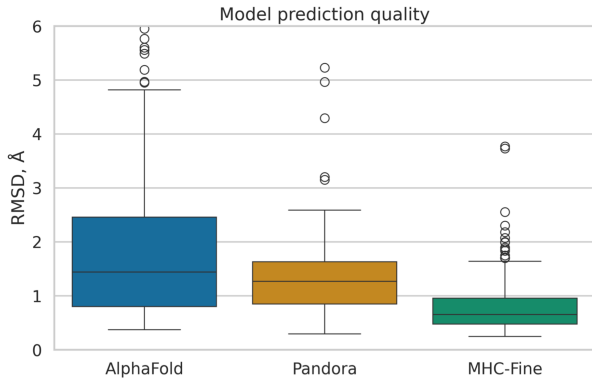


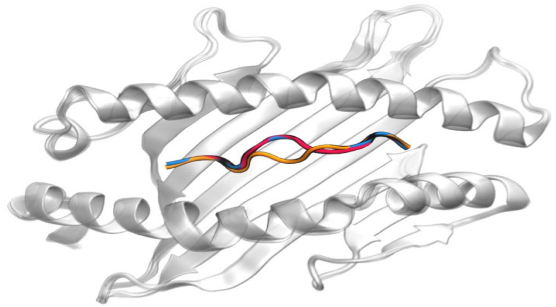
FFT with learned physics correction

Padhorny et. al PNAS 2016; Kozakov et. al Nature Protocols 2017;Desta et.al Nature Protocols 2023; Ignatov et. al JACS 2023

Pytorch-AF – Customized Alphafold-stlye architecture



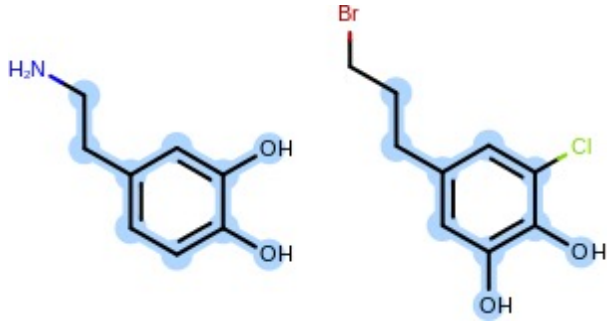


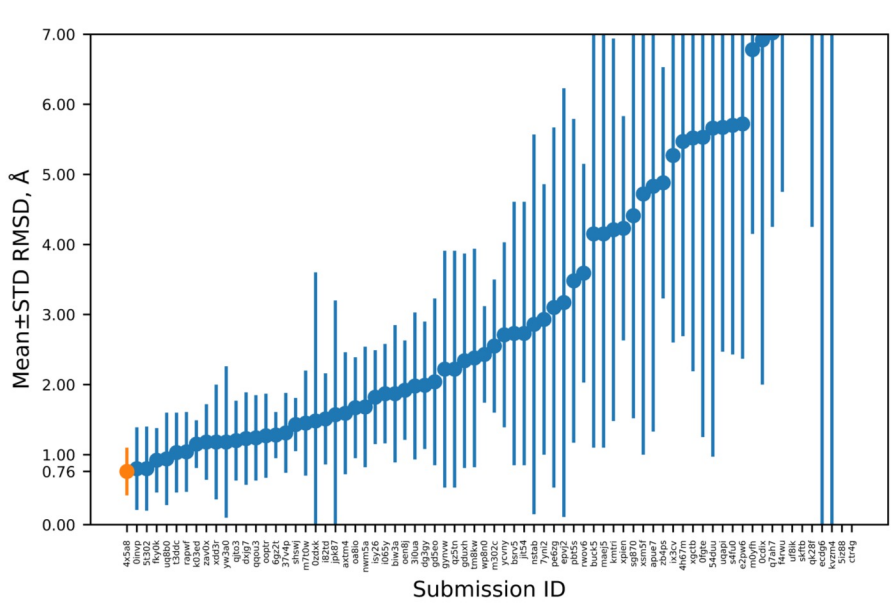


Jumper et. al. 2021; Glukhov et. al 2023

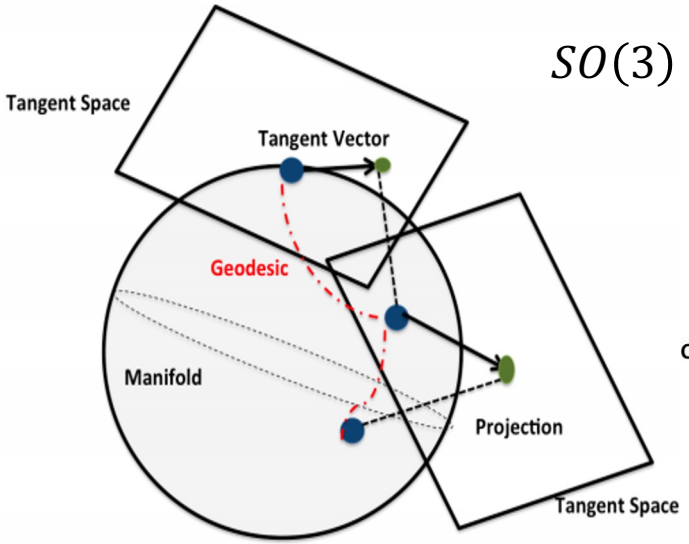
LigTBM protein-ligand docking

Maximum Common Substructure

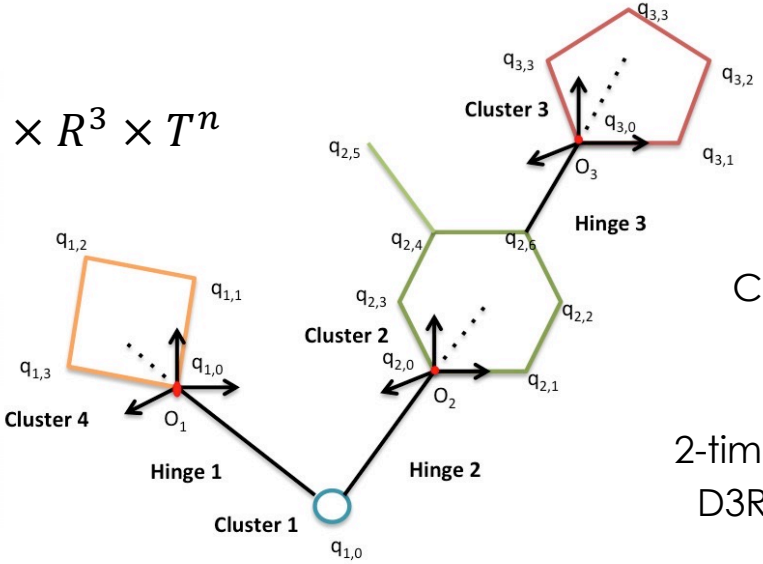


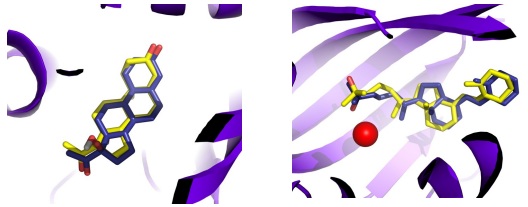


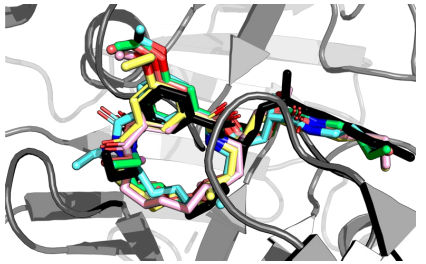
Diffusion on the manifolds



$$SO(3) \times R^3 \times T^n$$





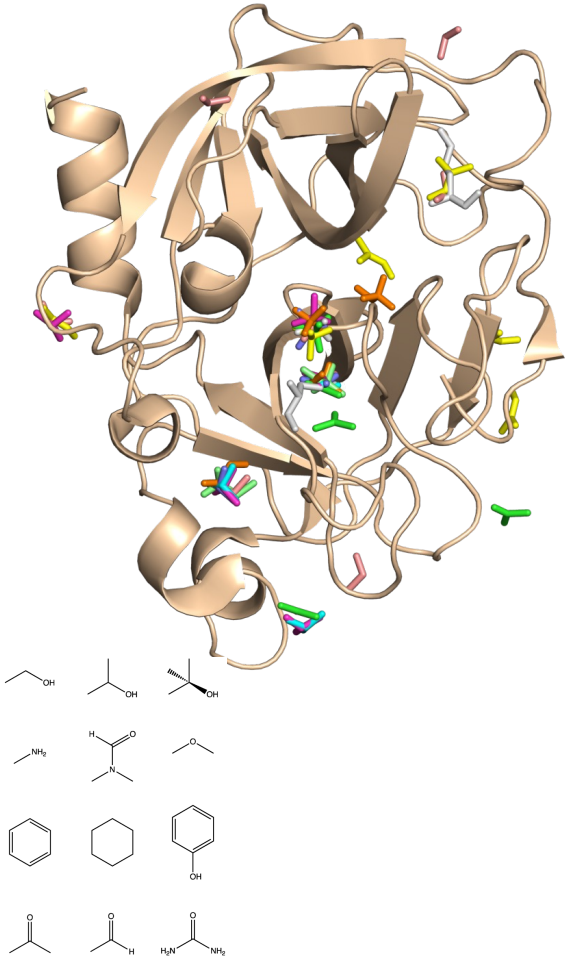


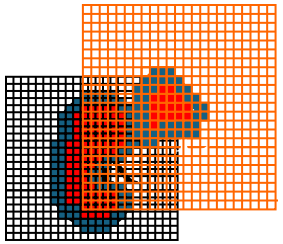
Top performer in the latest CASP (ligand prediction) and GPCR Dock competitions

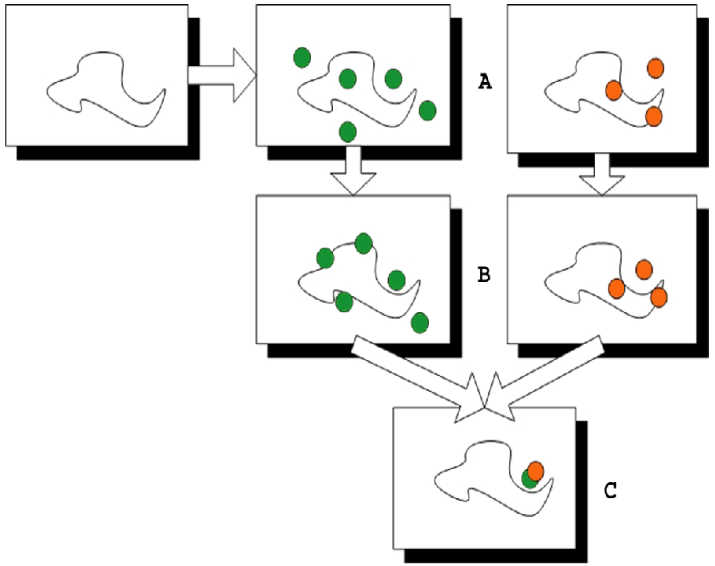
2-times top performer in the latest D3R ligand docking competition

Padhorny et al., JCAMD 2018; Ignatov et al., JCAMD 2018; Kotelnikov et al., JCAMD 2019; Alekseenko et al., JMB 2020; Kotelnikov et al., Proteins 2023;

FTMap – computational solvent mapping





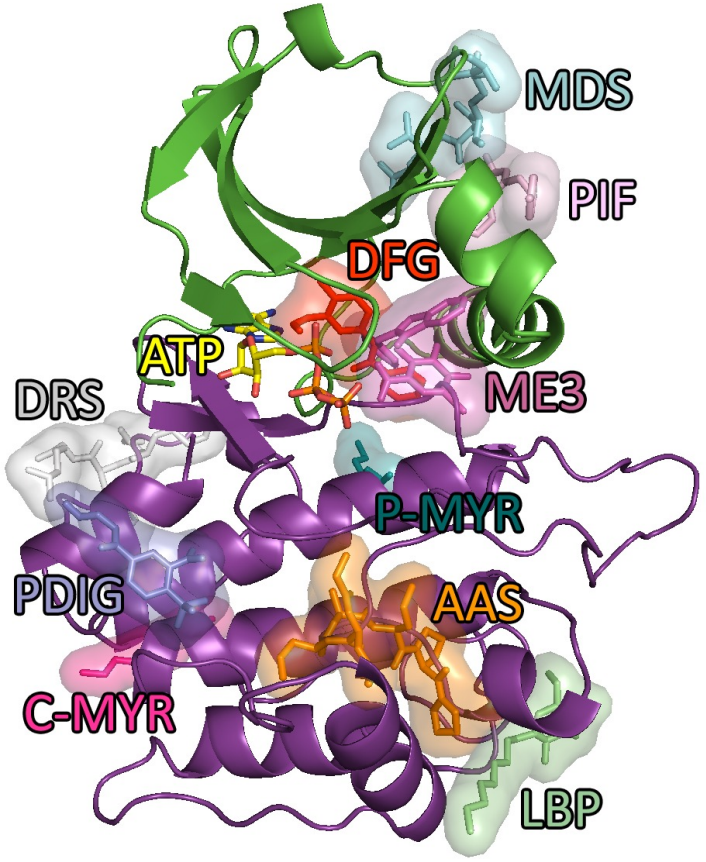


Druggable sites bind a variety of small molecules

“Hit rate” is a predictor of druggability

Brenke et. al 2009 Bioinformatics; Kozakov et al PNAS 2011;
Villar et al Nature Chem Bio 2014; Kozakov et al. Nature Protocols 2015;
Kozakov et. al PNAS 2015; Kozakov et. al J Med Chem 2015; Beglov et. al PNAS 2018; Yueh. J Med. Chem 2019; Egbert et. al 2021; Khan et. al 2023

Kinase Allostery Atlas



PDB Contains more than 3000 structures

325 Different kinase families

~250 human kinases + 250 more AF structures

Known regulatory sites:

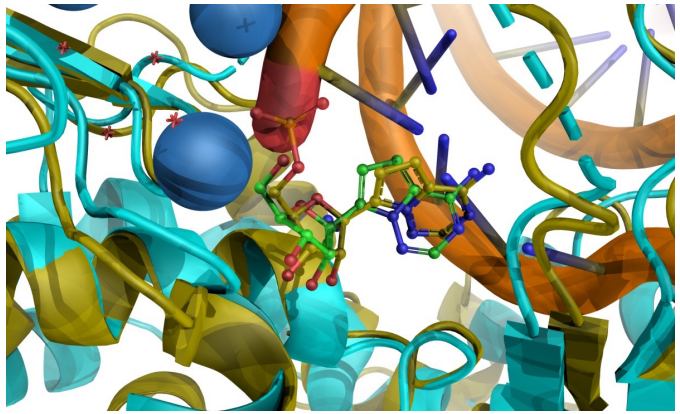
1) DFG loop pocket

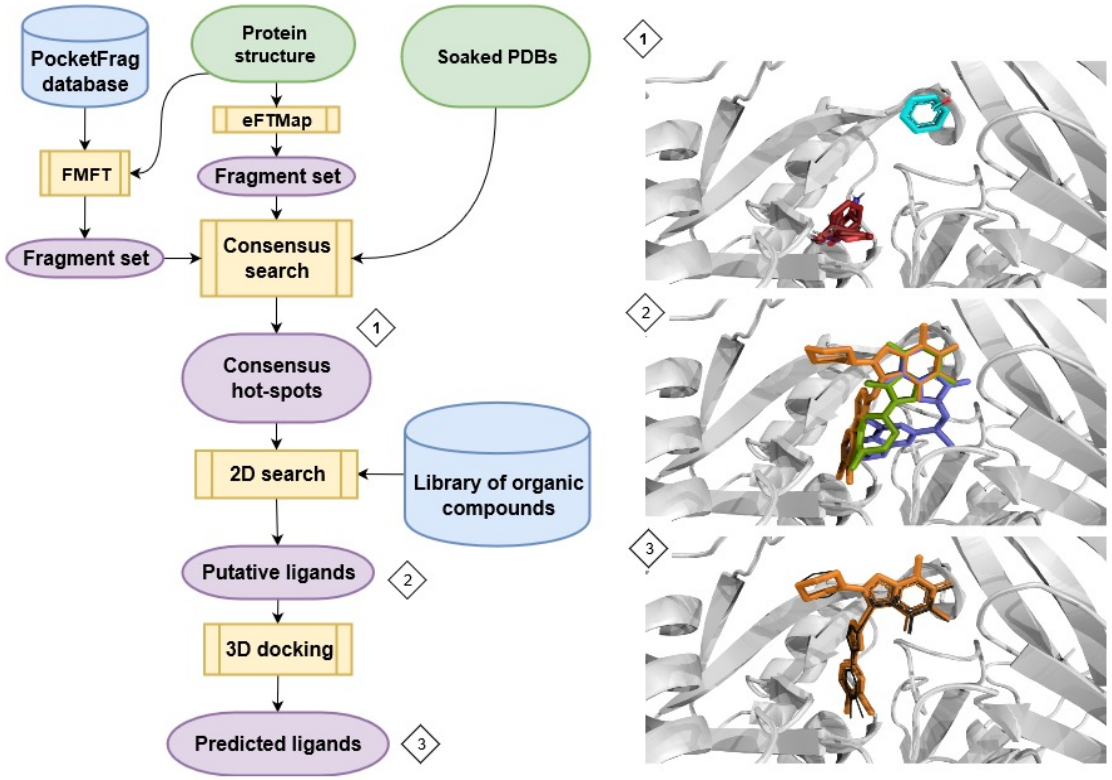
2) PIF pocket few compounds reported for PDK1

3)Several others, which are not validated

Yueh et al. J. Med Chem 2019; Jones et. al n preparation;

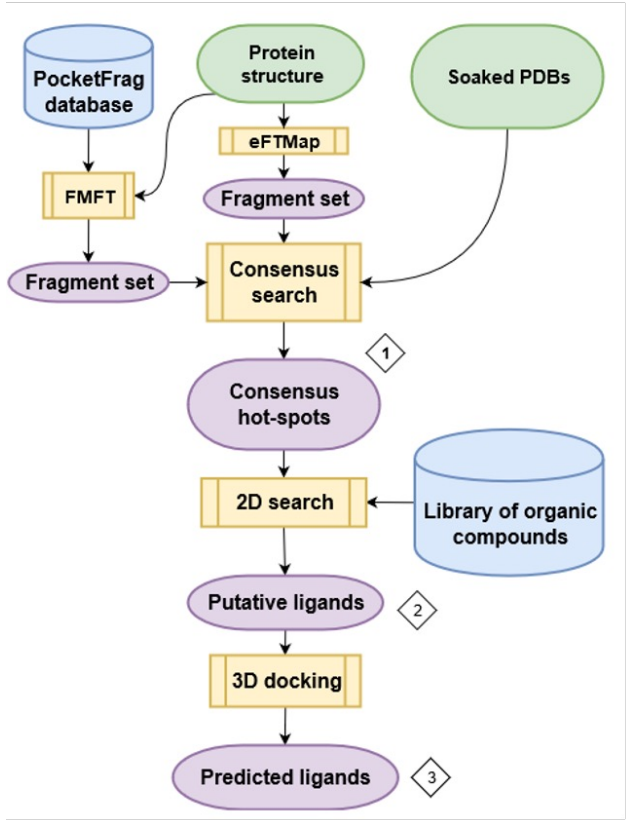
Screening of Giga-size libraries



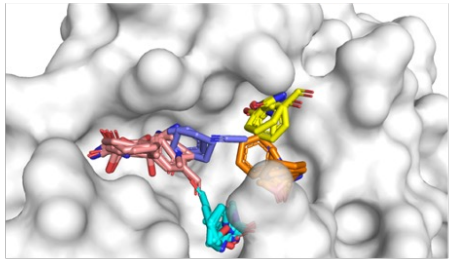


Identified nanomolar and low macromolar
hits to a number of COVID targets NSP3, NSP13, Mpro

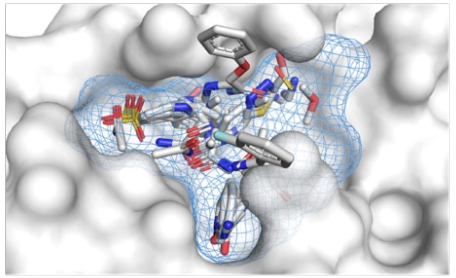
Virtual screening of giga-size libraries – SARS-CoV-2 Mpro

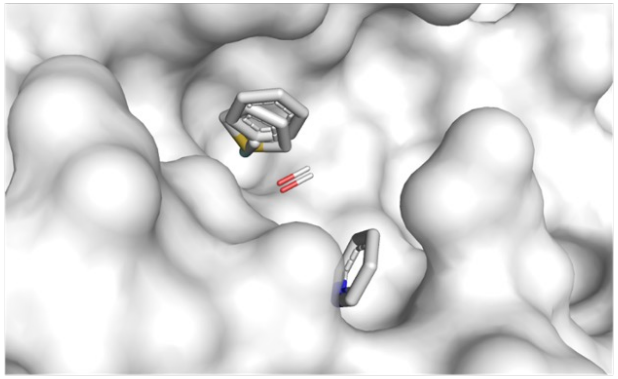


FTMap
computational solvent mapping



diamond





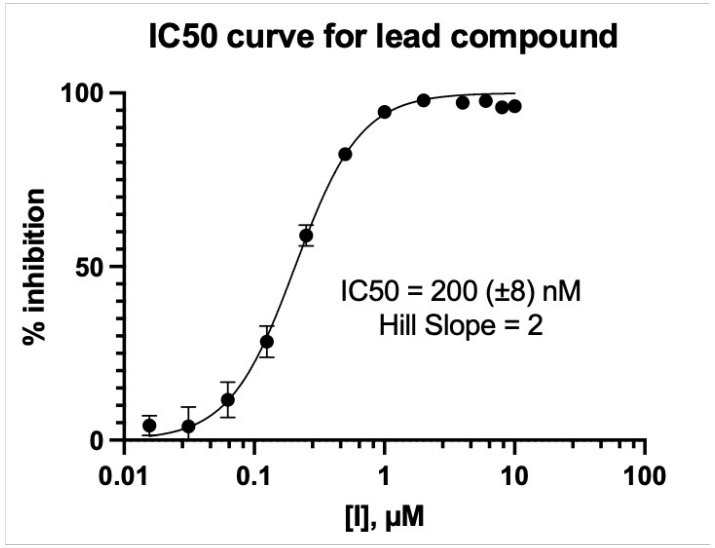
1.36B compounds

248 predicted ligands

HTS

28 hits

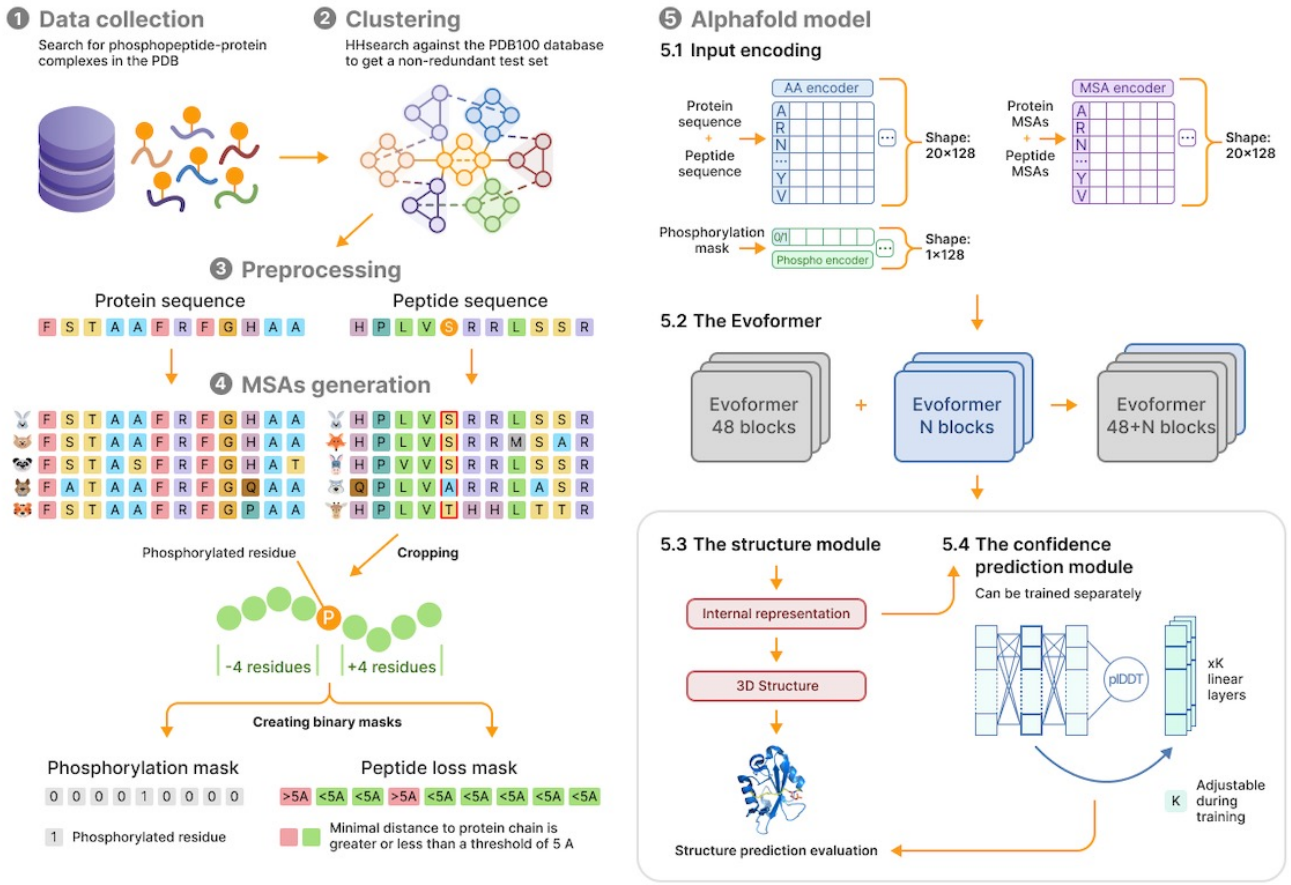
2 compounds with sub-micromolar IC50

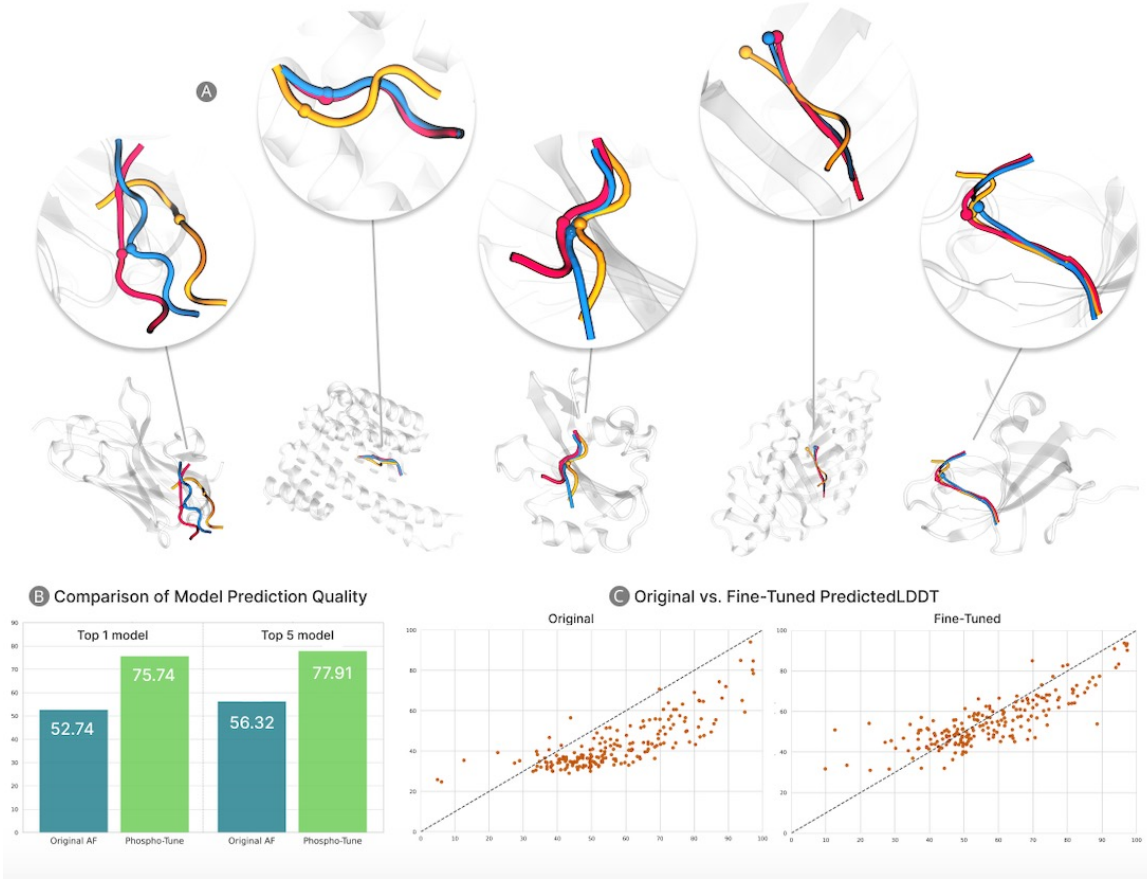


Currently in process of determining the X-ray structure of protein-ligand complex;

In collaboration with Professor Peter J. Tonge (Chemistry Dept.)
And Qun Liu (Brookhaven National Laboratory)
Alexander Tropsha (UNC Chapel Hill)

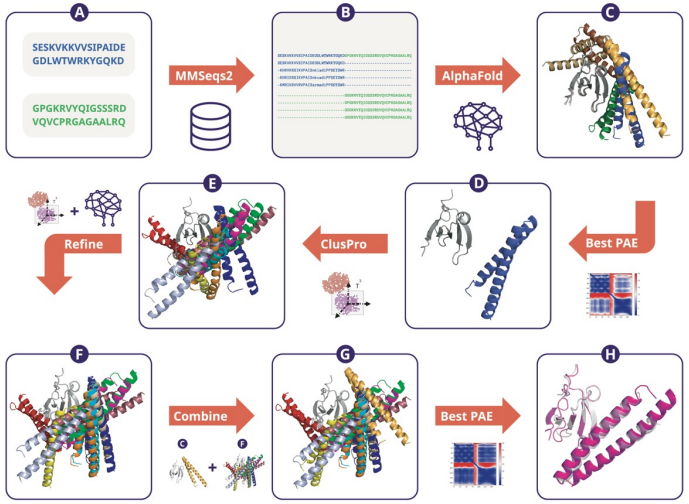
Modeling Phosphorylated interactions

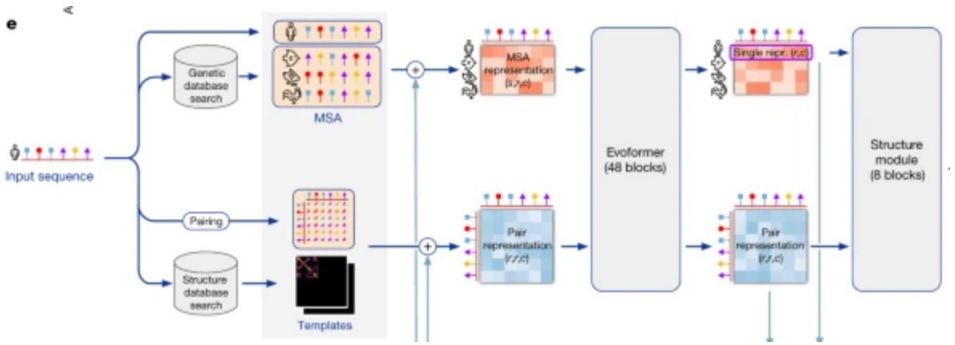


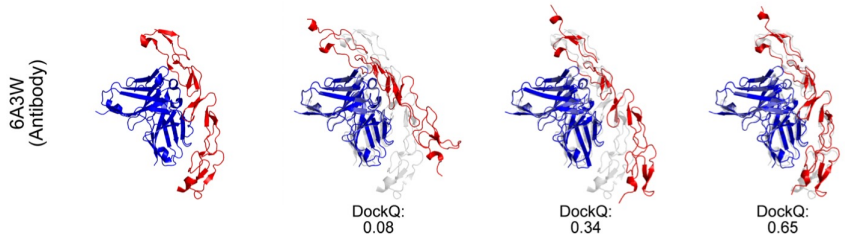


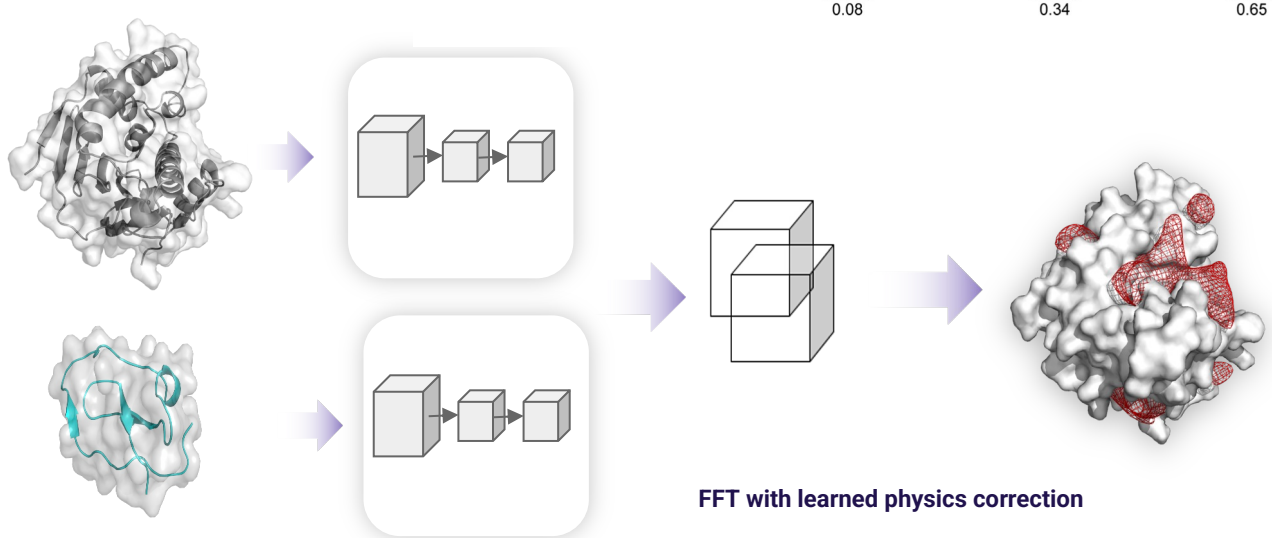
Glukhov et. al, Biorxiv 2024

Combination of FFT based architecture and AF for modeling Antigen Antibodies



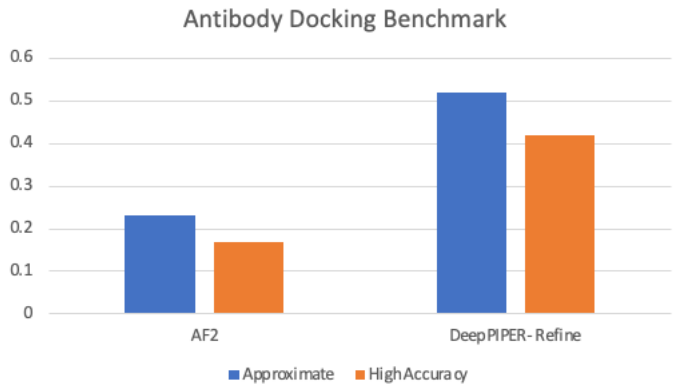






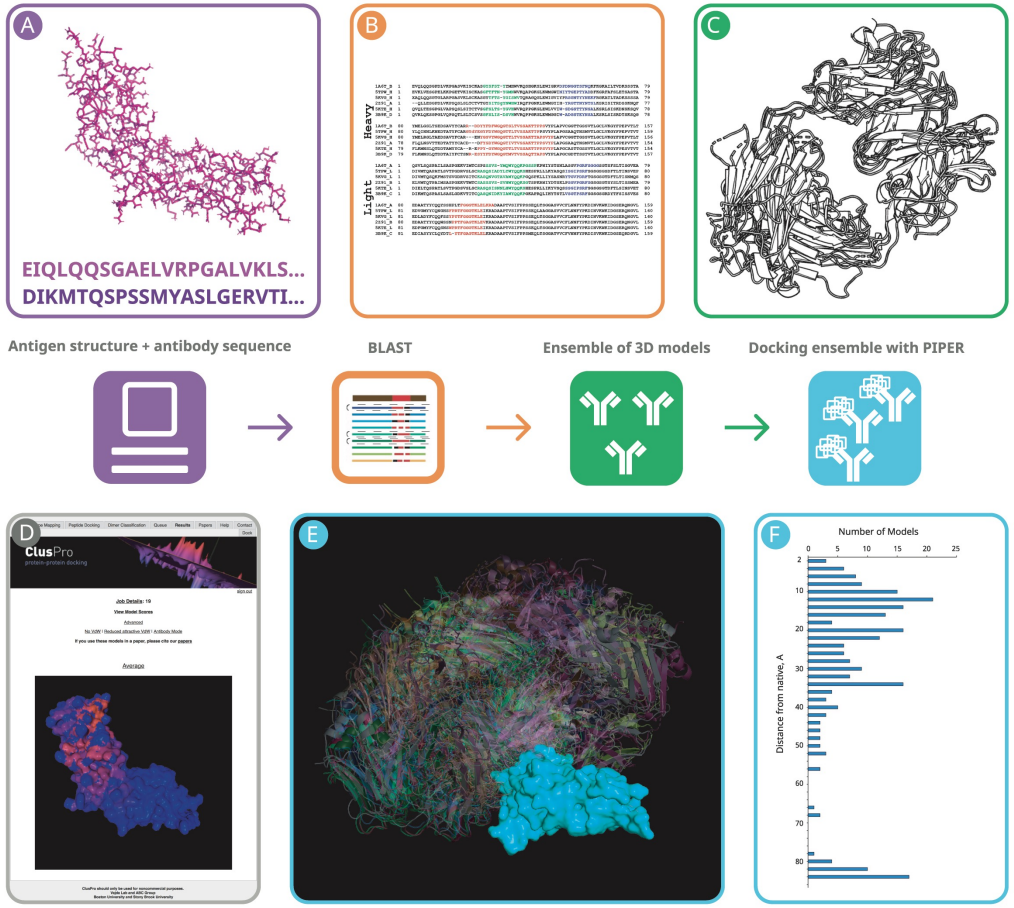
FFT with learned physics correction

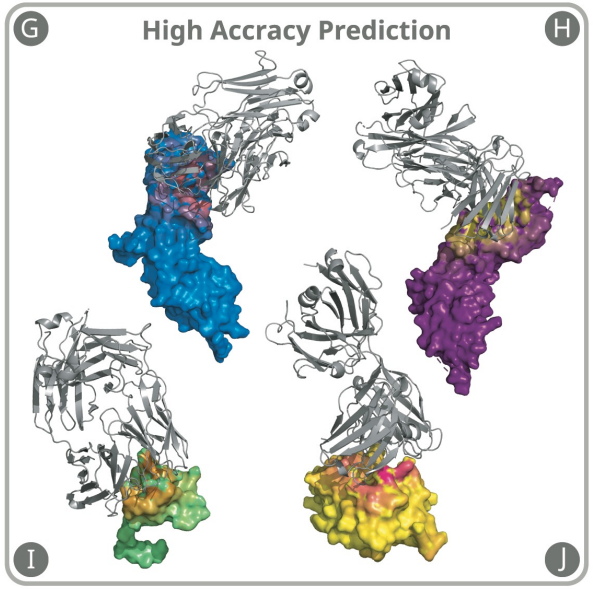
Top performers in CAPRI – worldwide blind protein docking competition

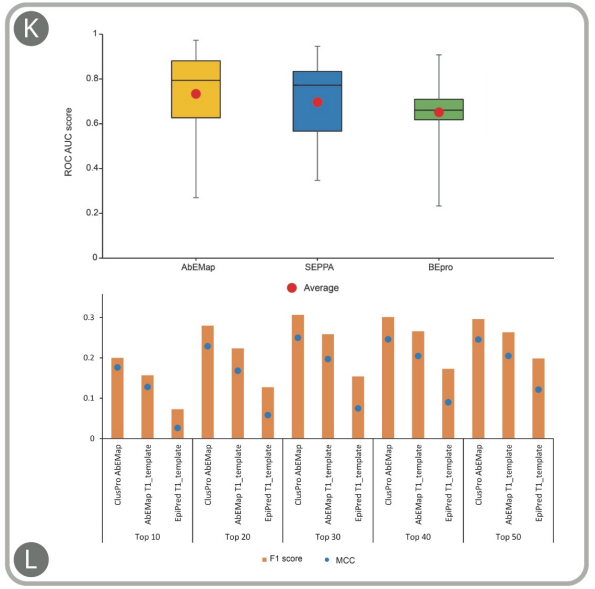


Ghani et al Biorxiv 2022; Ashizawa et. al in preparation

High Accuracy Epitope Detection

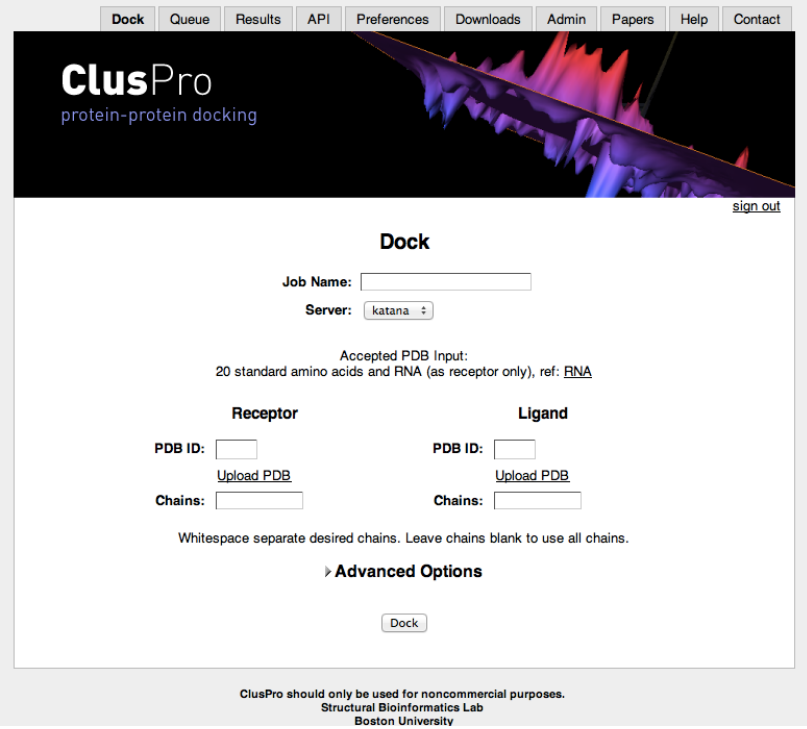


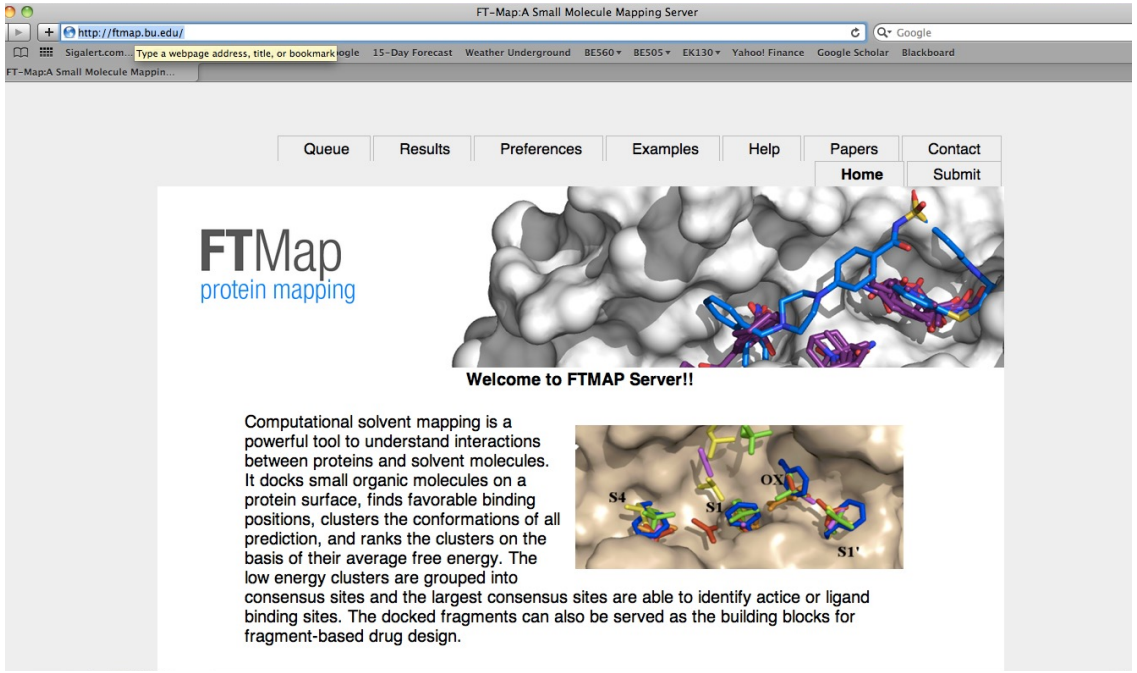




Desta et. al. Proteins 2023; Desta, Kotelnikov et. al Nature Protocols 2023.

ClusPro server 20000 users; FTMap server: over 5000 registered users





Kozakov et. al, Nature Protocols, 2015; Kozakov et. al Nature Protocols 2017; Desta et. al Nature Protocols 2023

Modeling PROteolysis TArgeting Chimeras (PROTACs)

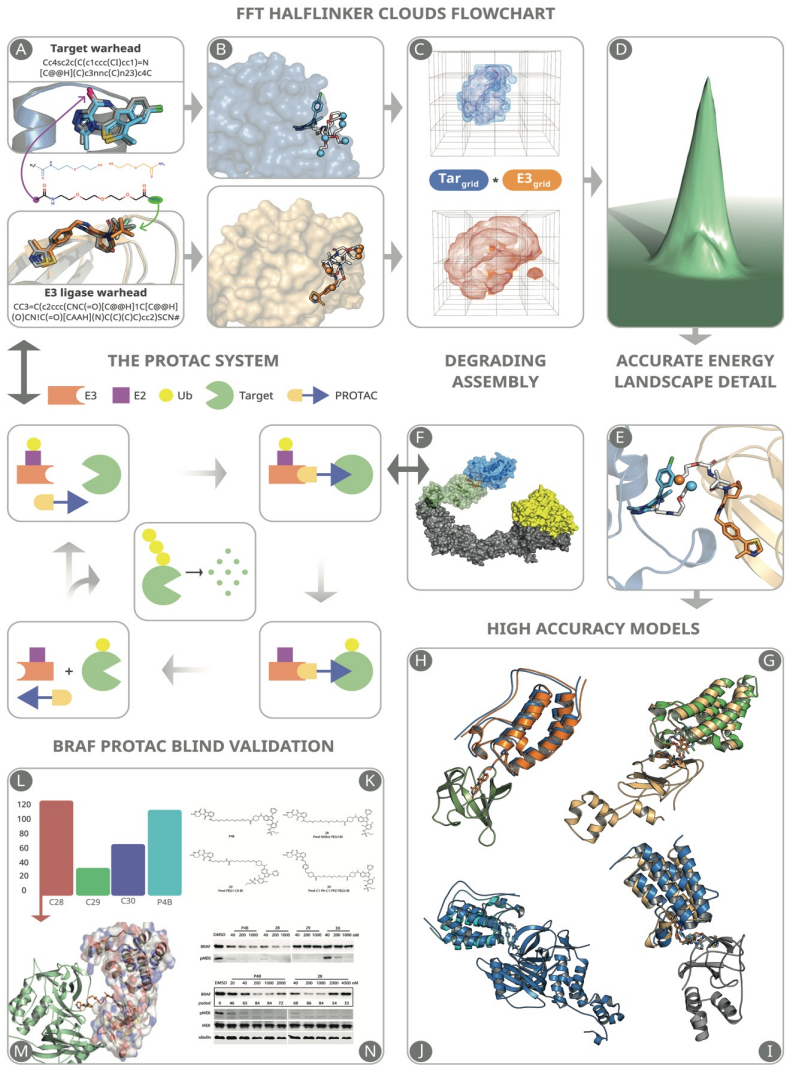
Goal: hijack ubiquitin-proteasome system to degrade target protein

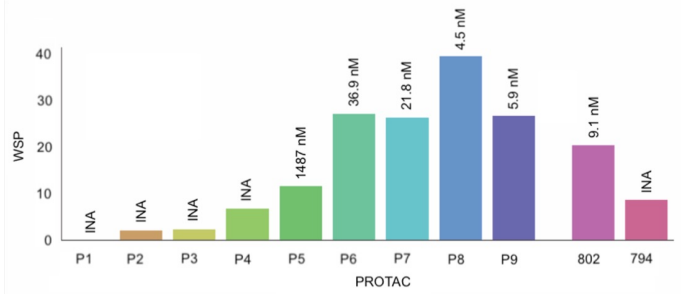
We want to aid PROTAC design:

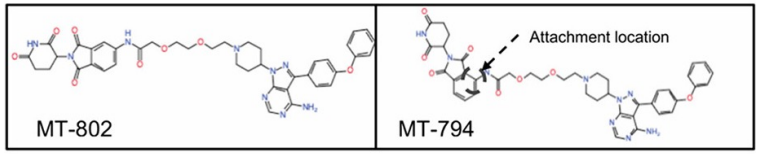
- PROTAC ternary complex structure
- PROTAC efficiency

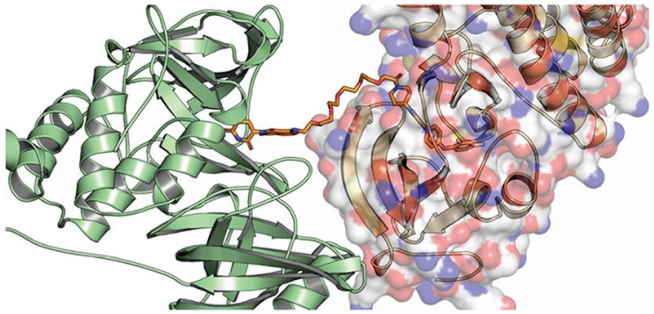
Challenging sampling problem:

- PROTAC linker might have non-trivial chemistry and conformational space
- Multiscale modeling
- Non-native protein-protein interaction
- Suboptimal interface

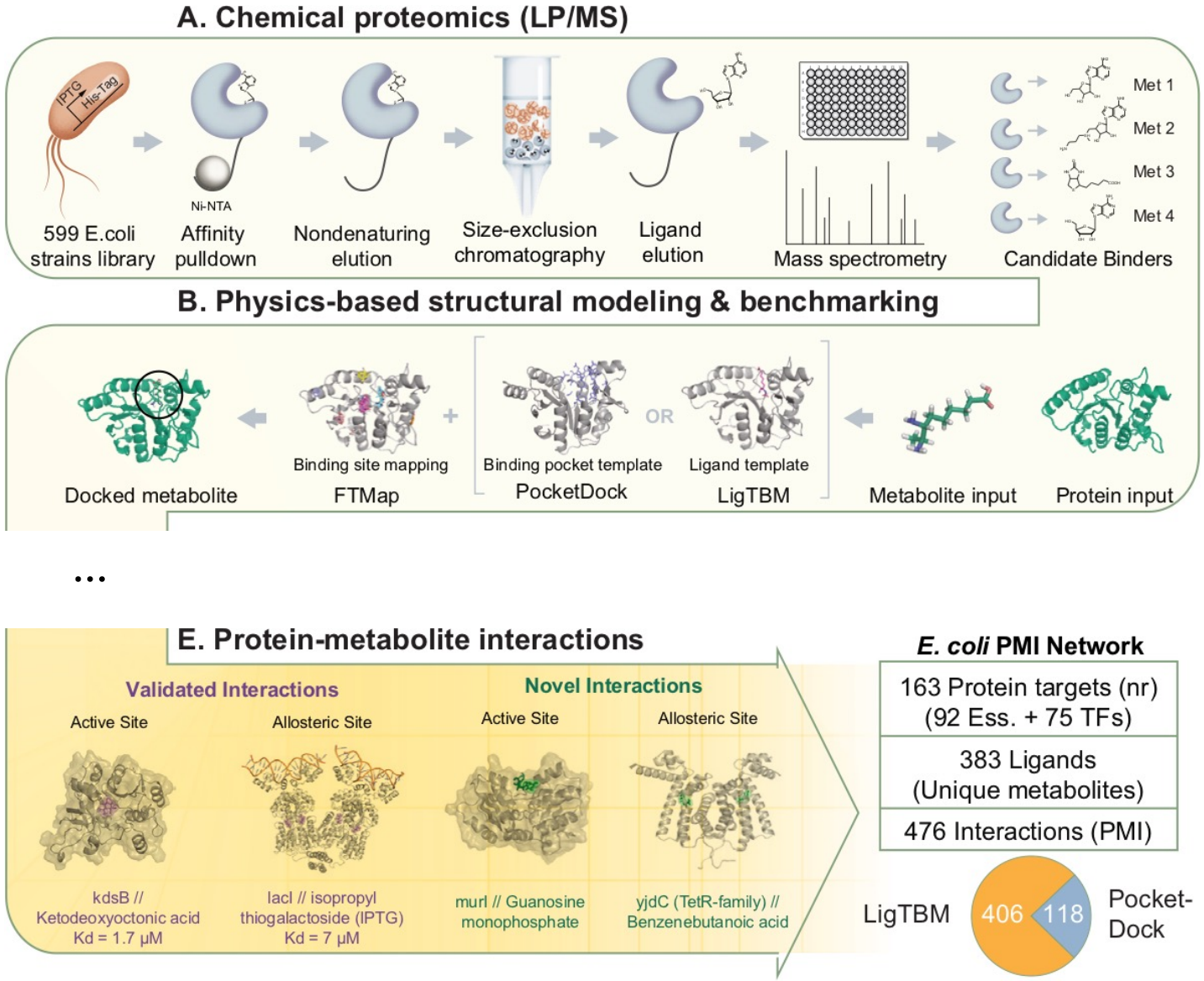




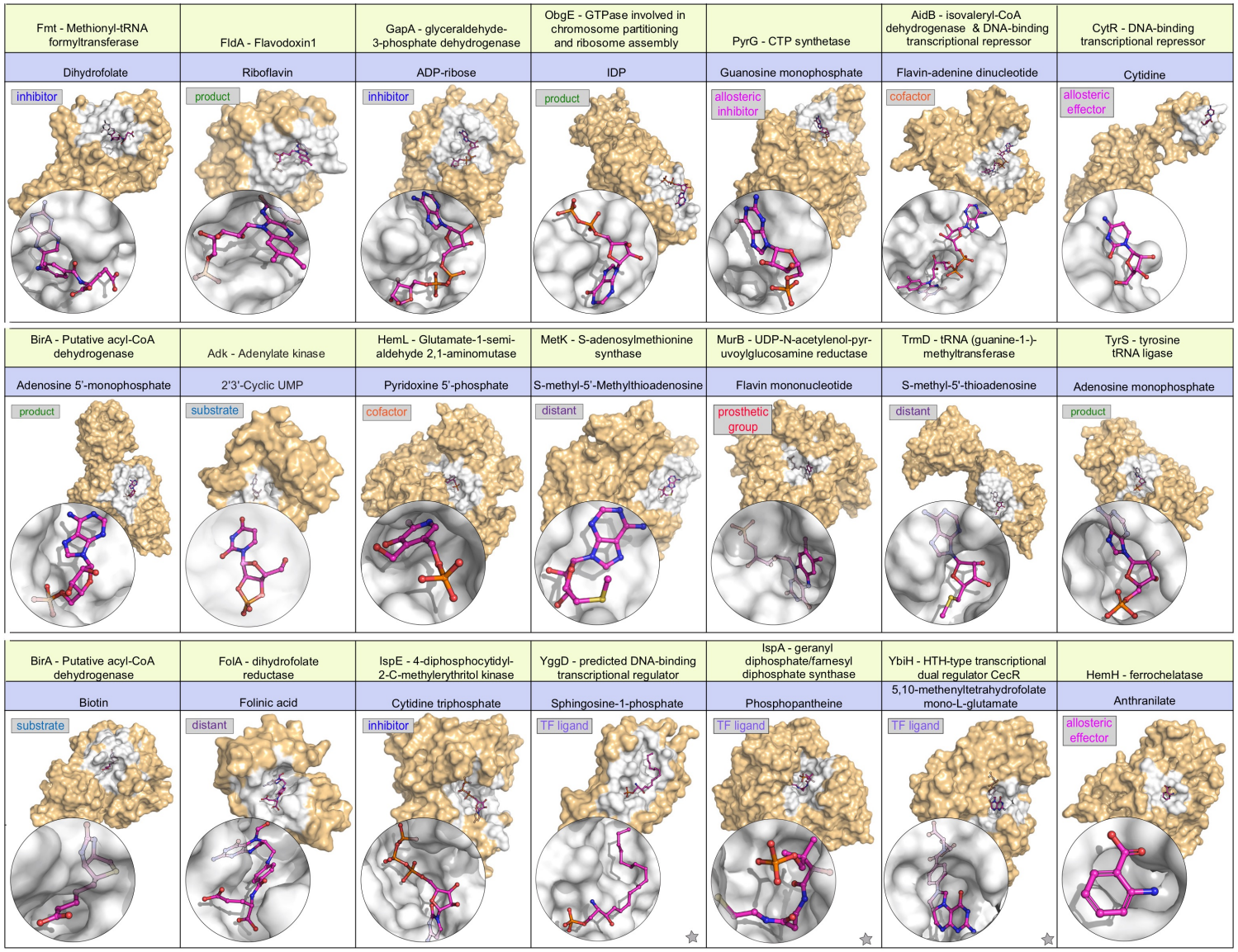




Large-scale mapping of native protein-metabolite interactions in E. coli using Mass Spec & LigTBM



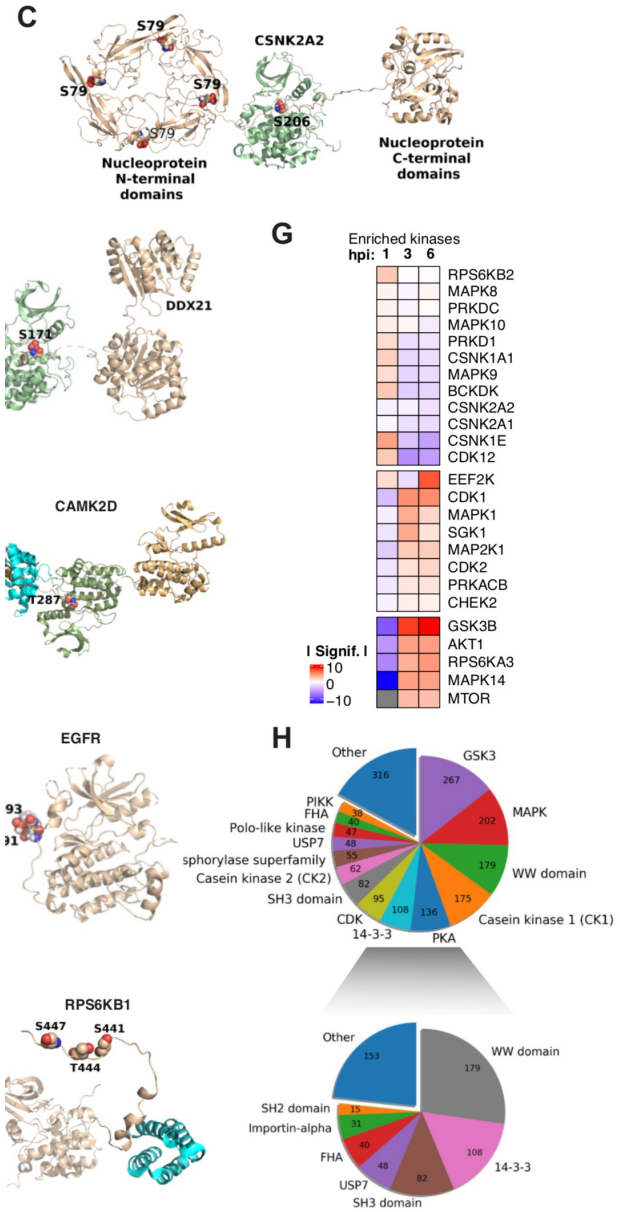
Known and predicted protein-metabolite interactions & SPR validation

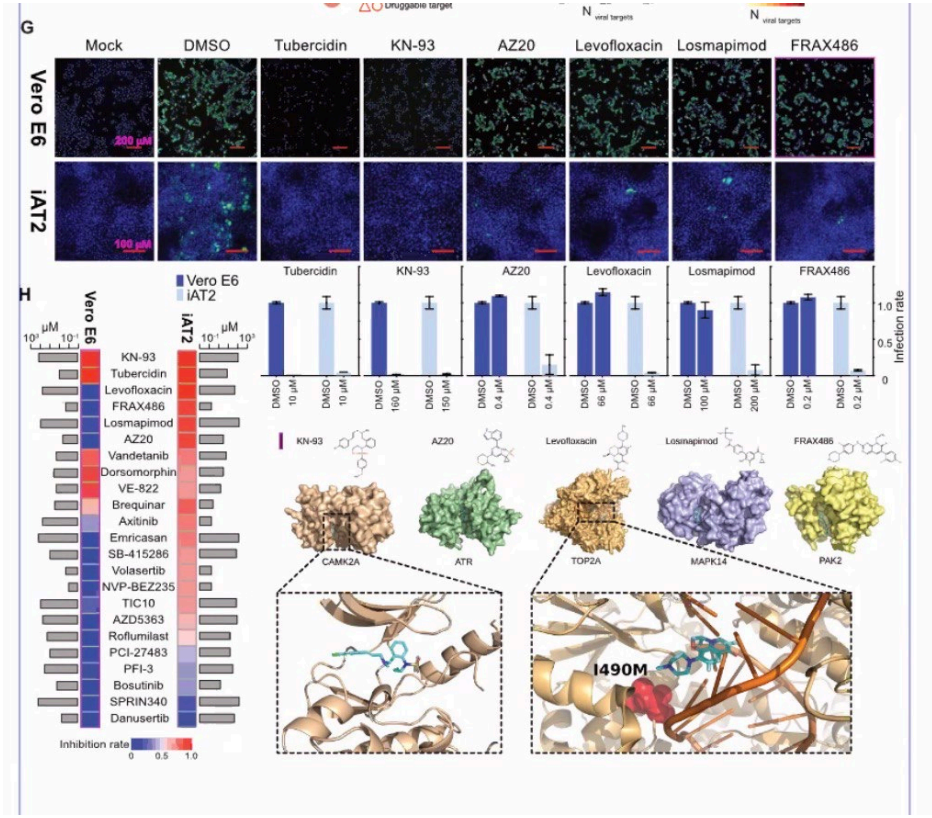


Surface Plasmon Resonance assay

	KD (nM)
HemH - 4-aminobenzamide	483
YjdC - Oleic acid	3.8
PurB - Phenazine	61.6
FldA - [2,2'-Bipyridine]-5-carboxylic acid	192
IspA - (5-Fluoro-2-oxo-2,3-dihydro-1H-indol-3-yl)-acetic acid	54.5
ObgE - 1-isoquinolinyl(phenyl) methanol	119
FldA - Riboflavin	224
IspF - Phenazine	60.1
IspB - 1-(3,4-Dimethoxy-phenyl)-ethylamine	118
TrmD - S-methyl-5'-thioadenosine	14.9
MetK - S-methyl-5'-thioadenosine	0.355
UvrY - 4-chloro-2-hydroxybenzamide	170
PyrG - Guanosine	1250
IspE - Cytidine	67.7
MurB - Flavin mononucleotide	84500

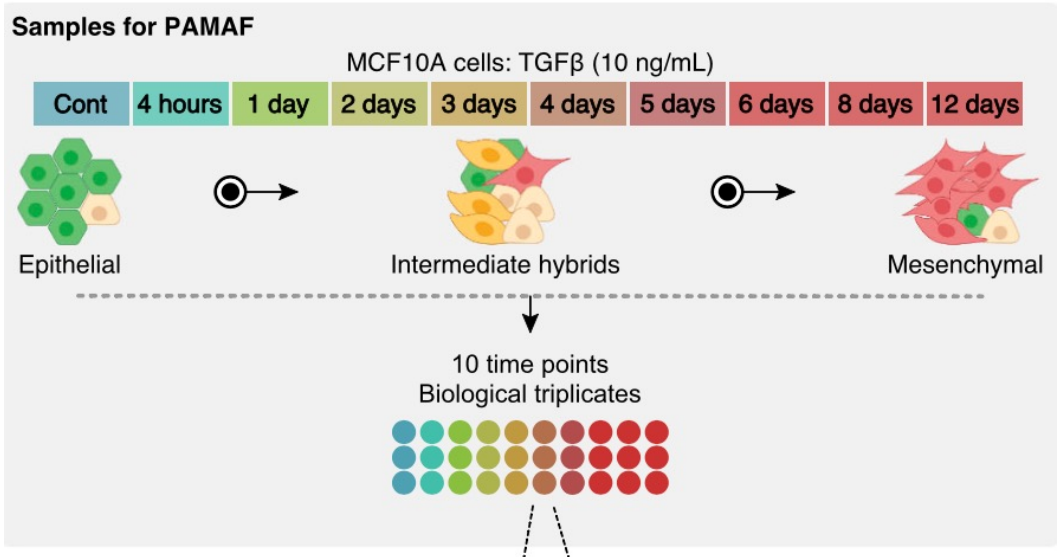
Phosphorylation effect of SARS-COV2 on infected lung cells – Target identification

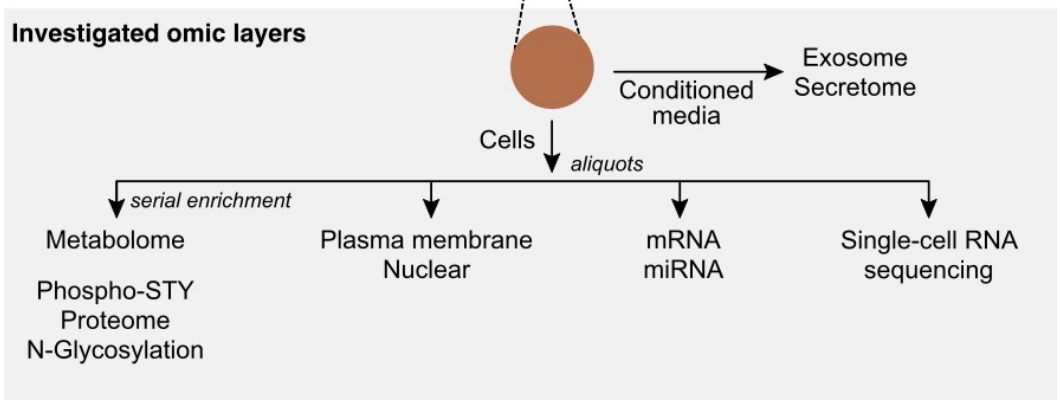


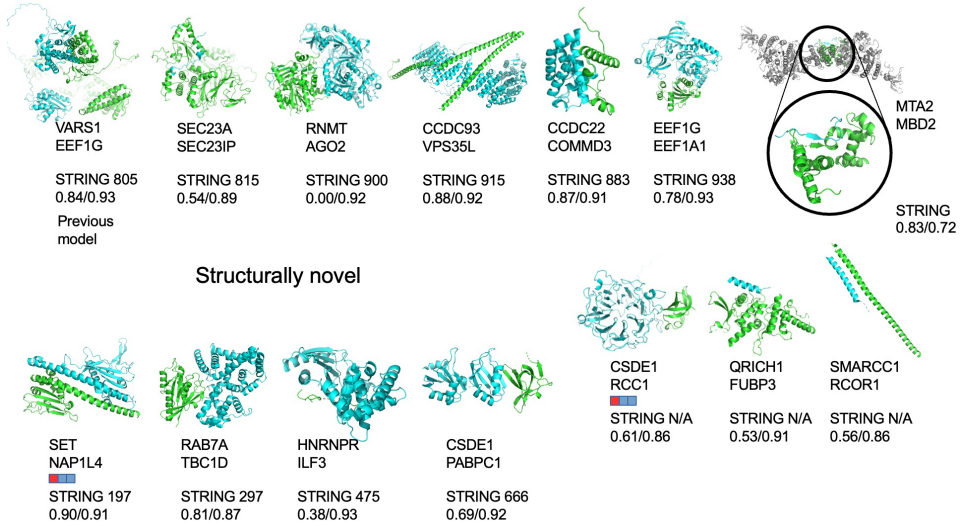


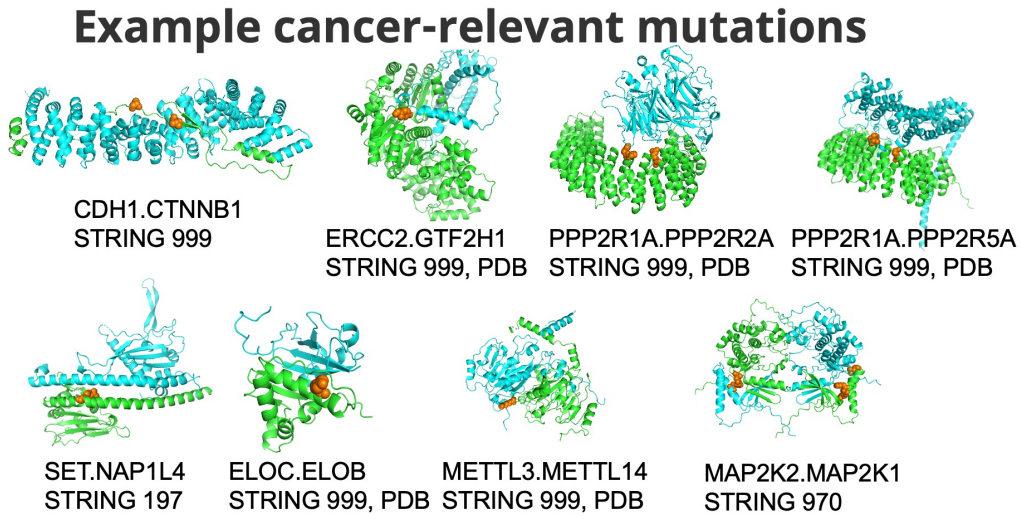
Hume et. al, Molecular Cell, 2021

Understanding EMT





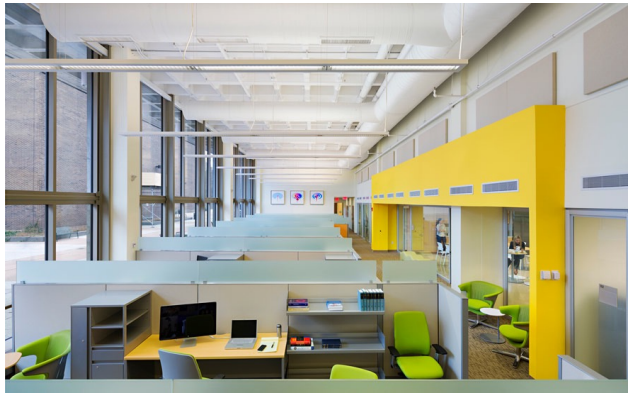




Havugimana, 2022 Nature Communication; Indranil et. al, Nature communications, 2023; Indranil , Padhorny in preparation

ABC Lab members

Dr. Dzmitry Padhorny
Sergei Kotelnikov
Dr. Mikhail Ignatov
Ernest Glukhov
Ryota Ashizawa
Xiaogang Li
Dr. Mark Lukin
George Jones
Thu Nguyen
Yimin Zhu
Derara Haligeorgious
Veranika Averkova
Dmytro Kalitin



NIH , NSF
Computer time DOE INCITE leadership award

SBU

Ken Dill
Carlos Simmerling
Ivet Bahar
Peter Tonge
David Thanassi
Vageli Coutsias

U of Toronto, OHSU

Andrew Emili

BU

Sandor Vajda
Adrian Whitty

UNC

Alex Tropsha
Tim Wilson