

# Systems level understanding and regulating of the disease with atomic resolution

**Dima Kozakov**

**Laufer Center for Physical and Quantitative Biology**

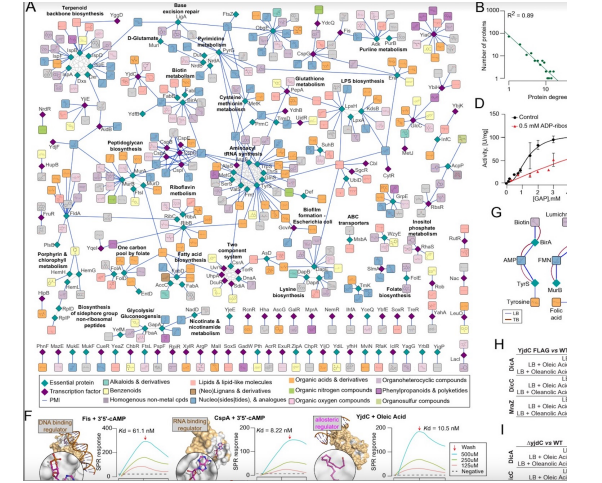
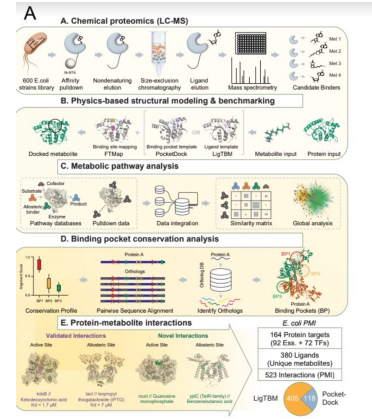
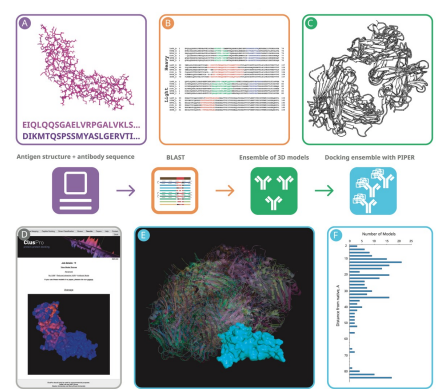
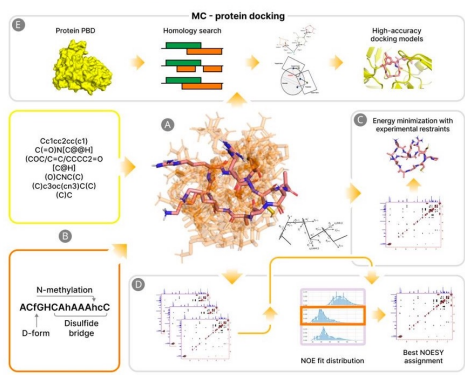
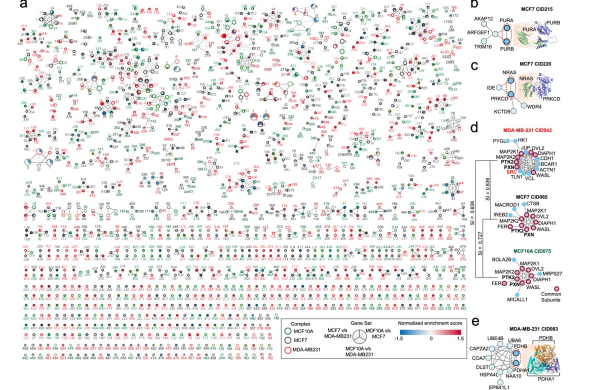
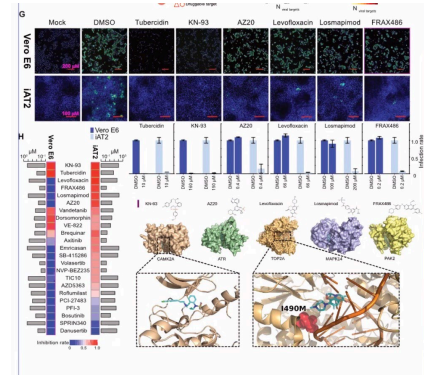
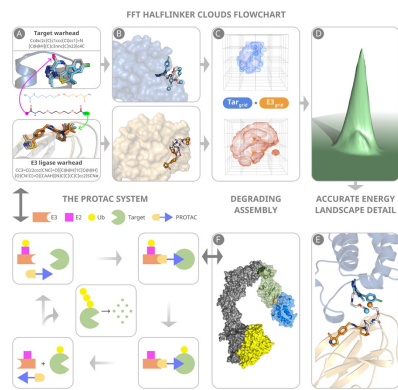
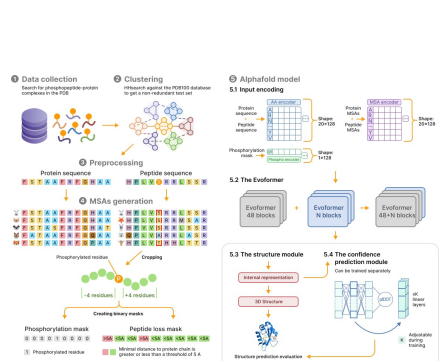
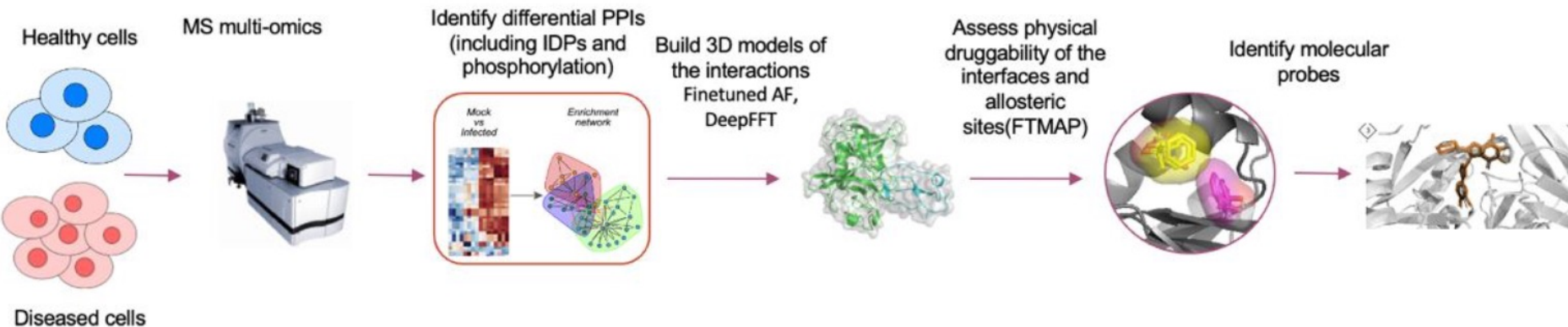
**Stony Brook University**



**Stony Brook  
University**

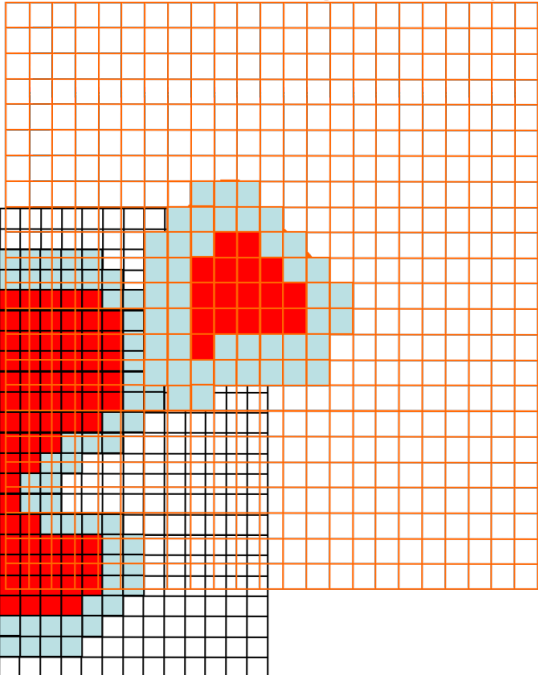


**LAUFER CENTER**  
FOR PHYSICAL AND QUANTITATIVE BIOLOGY



# FFT convolution approach enables global exhaustive macromolecular interaction sampling

Ligand representation (e.g. charge density)



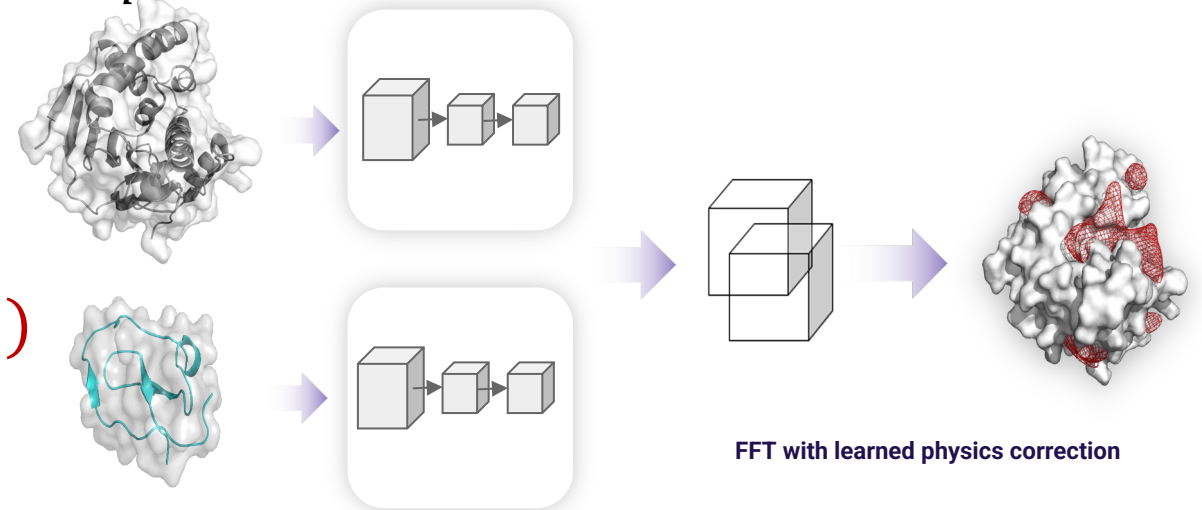
Interaction energy as a sum of FFT convolutions

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

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

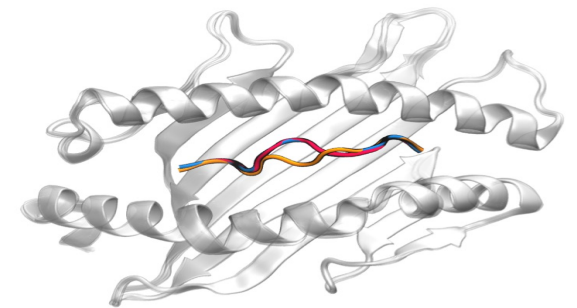
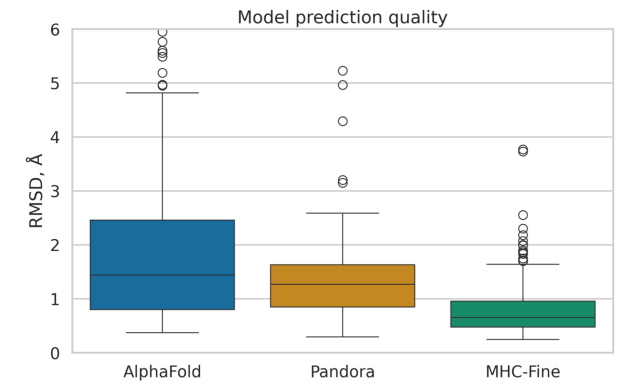
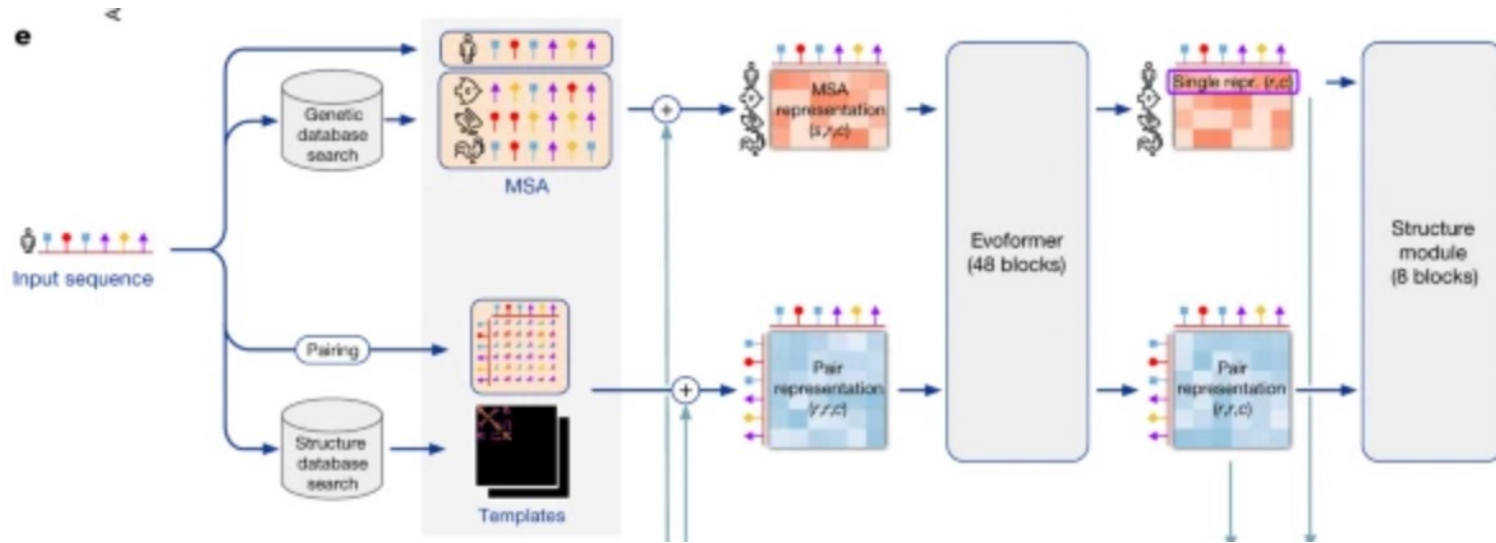
$$\begin{aligned} E_{full} &= E_{vdw} + E_{elec} + E_{pair} \\ E_{vdw} &= E_{rep} + E_{attr} \\ E_{elec} &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{q_i q_j}{\left( r_{ij}^2 + D^2 \exp\left(-\frac{r_{ij}^2}{4D^2}\right) \right)^{1/2}} \\ E_{pair} &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \varepsilon(i, j)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

Receptor representation  $O(N^6) \rightarrow O(N^3 \ln N^3)$   
(e.g. electrostatic potential)



FFT with learned physics correction

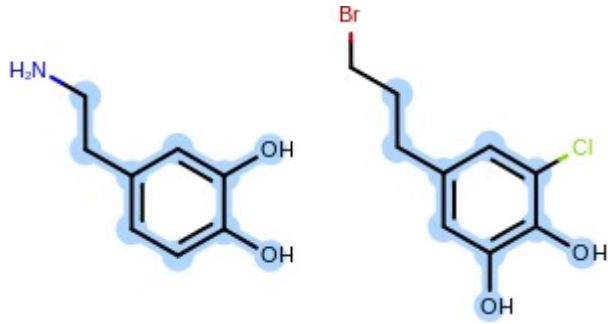
# Pytorch-AF – Customized Alphafold-stlye architecture



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

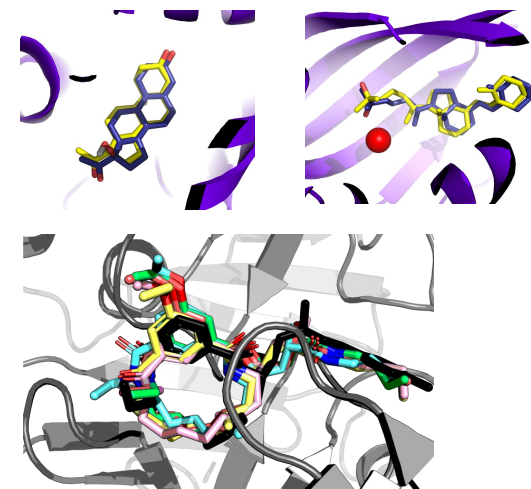
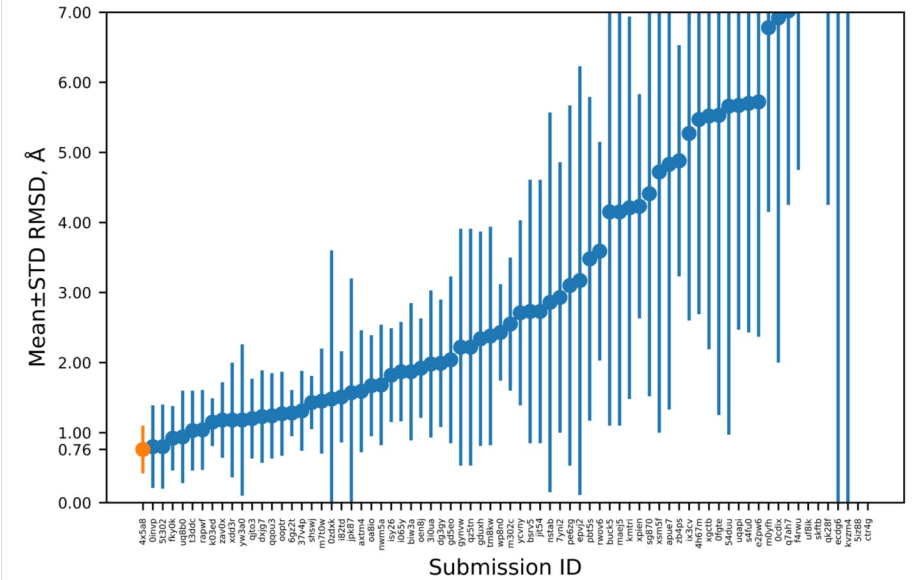
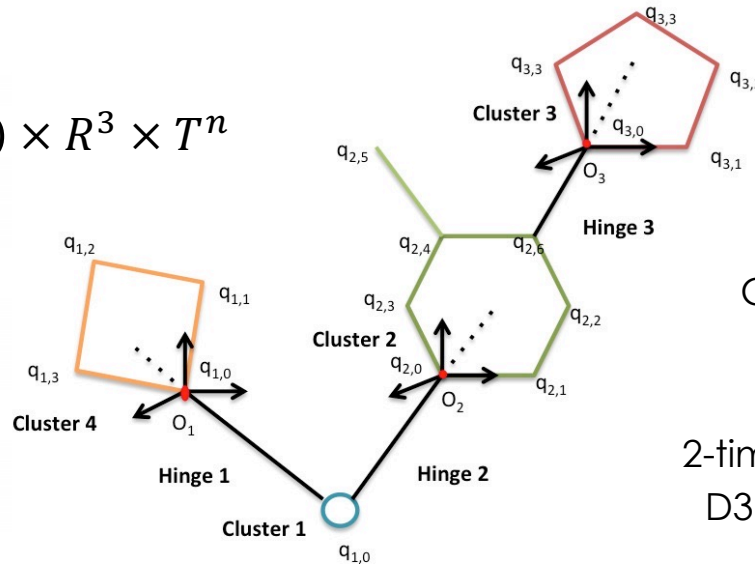
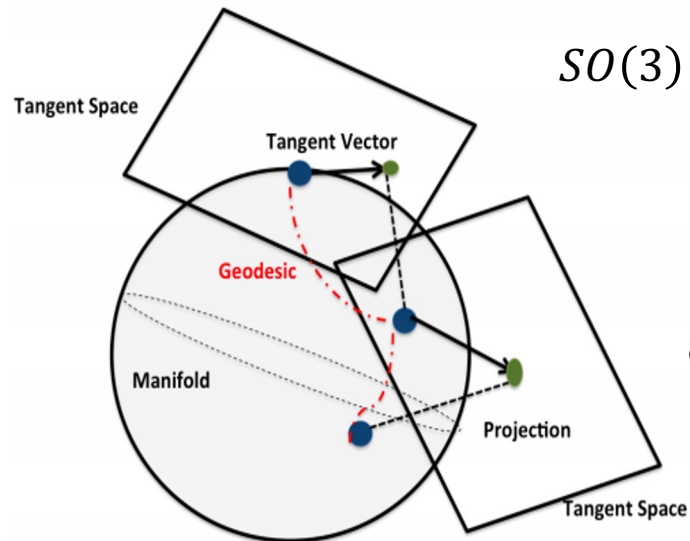
# LigTBM protein-ligand docking

Maximum Common Substructure



Diffusion on the manifolds

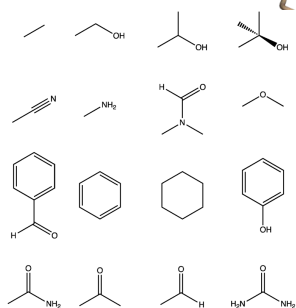
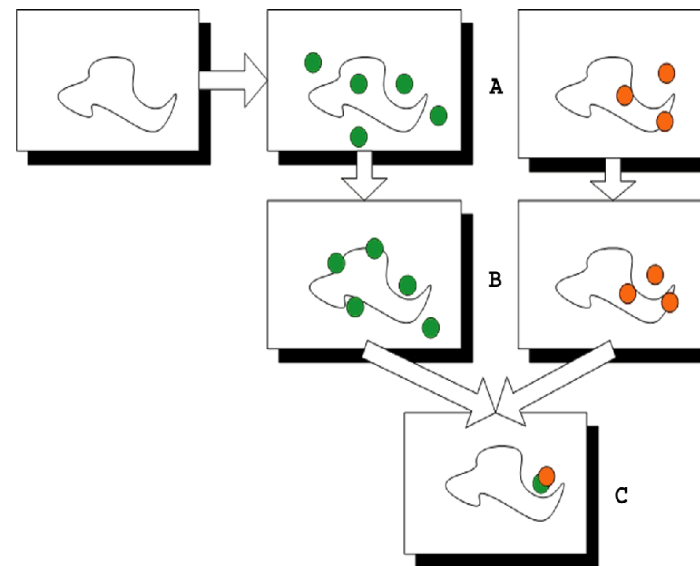
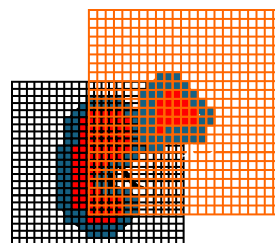
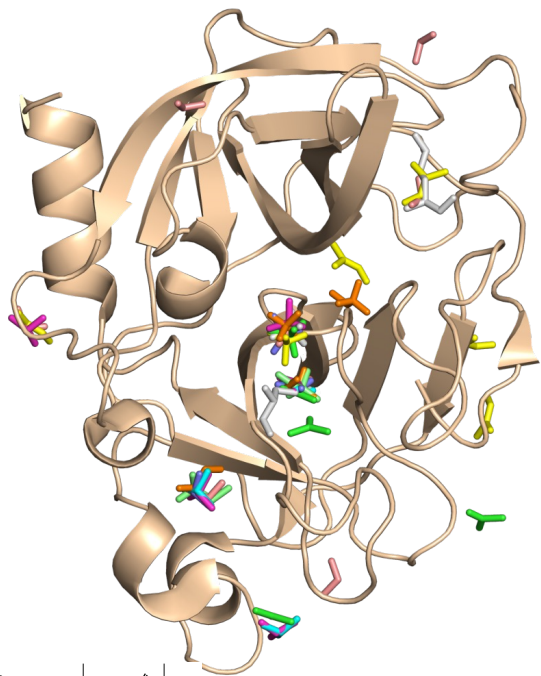
$$SO(3) \times R^3 \times T^n$$



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

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

# FTMap – computational solvent mapping

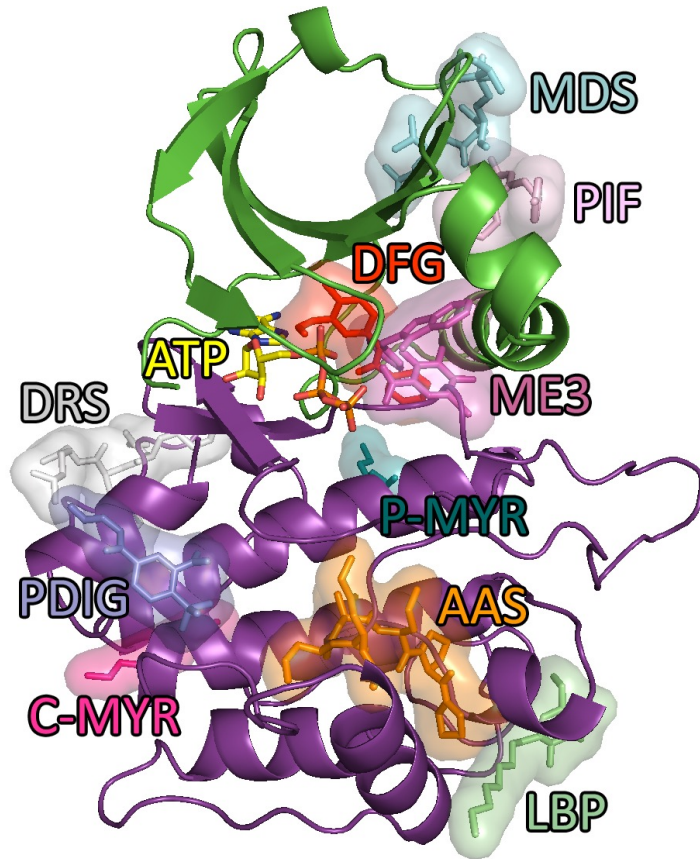


Druggable sites bind a variety of small molecules

“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

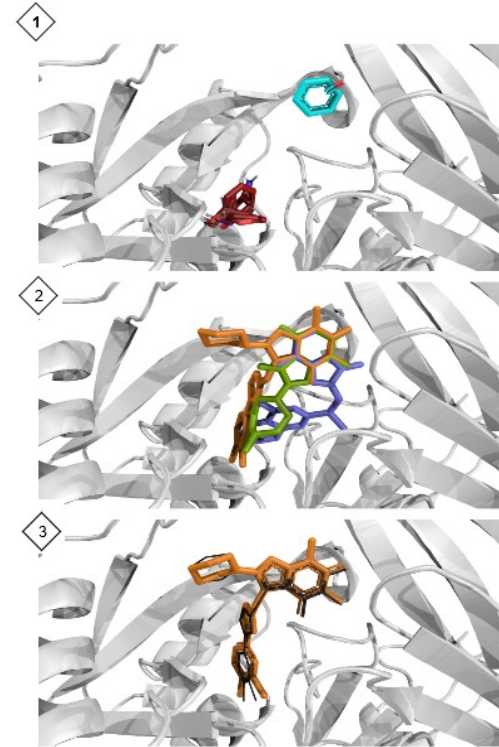
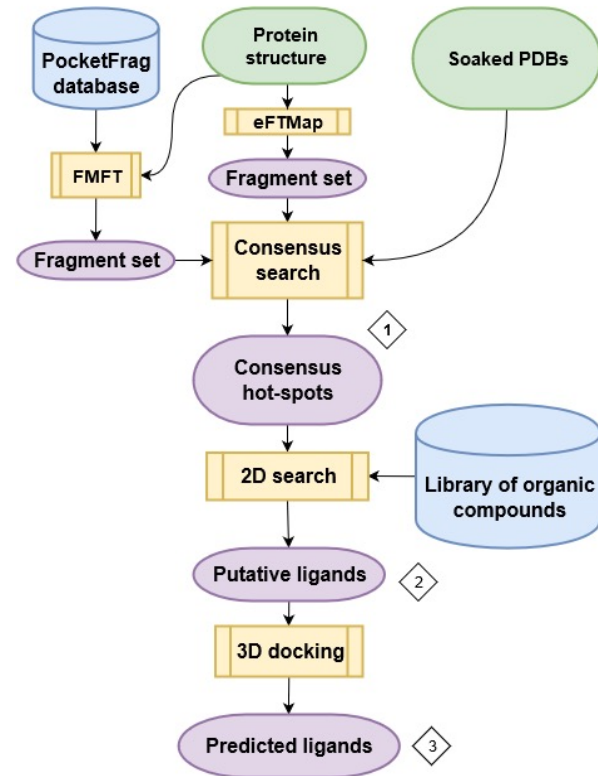
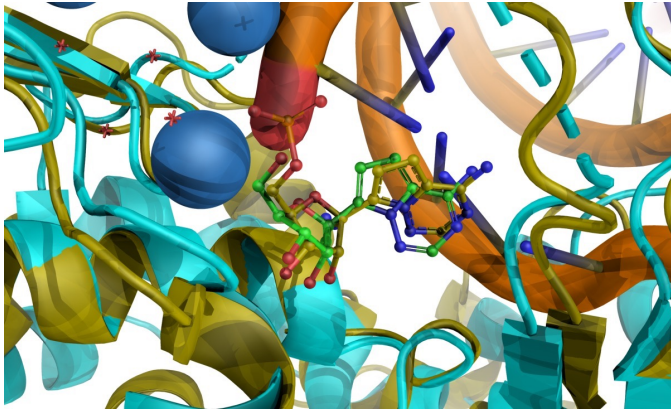
325 Different kinase families

~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

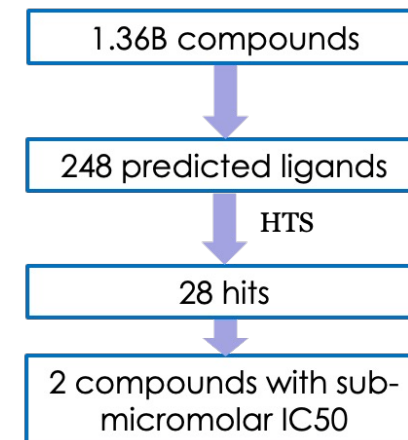
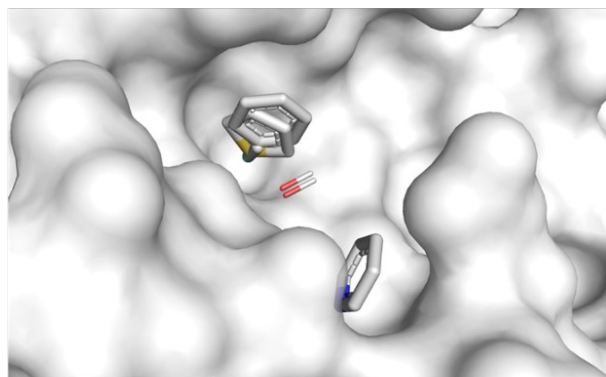
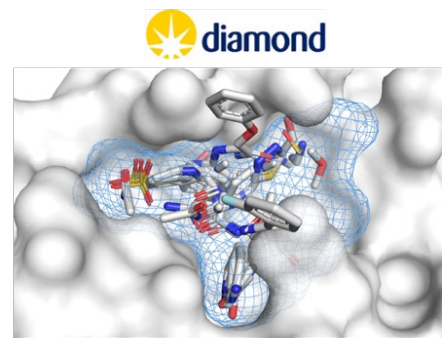
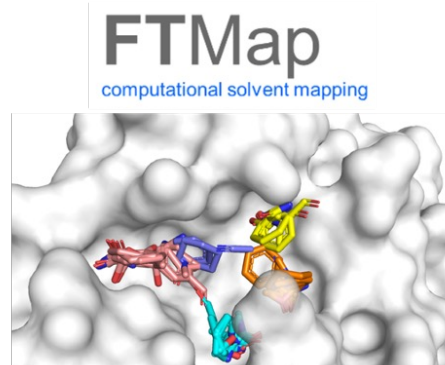
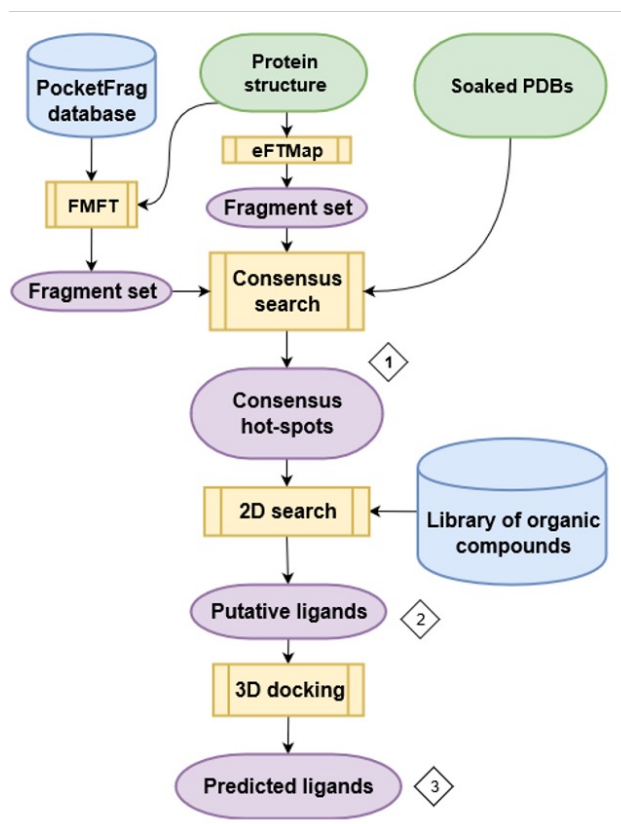
## Screening of Giga-size libraries



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

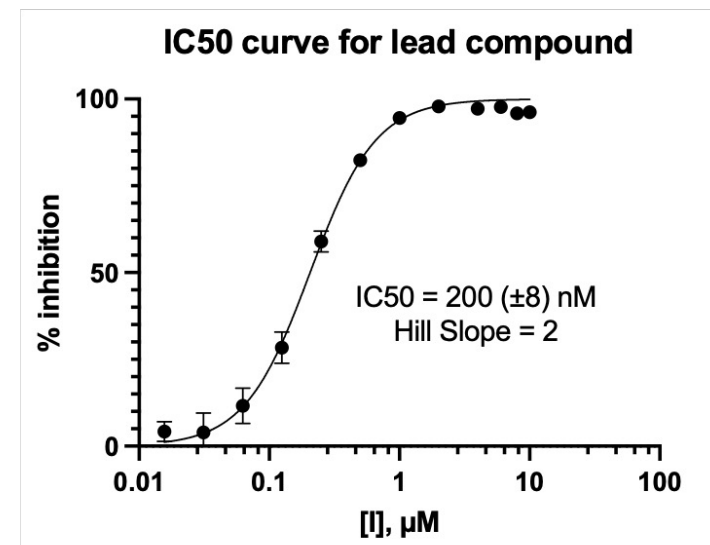


# Virtual screening of giga-size libraries – SARS-CoV-2 Mpro



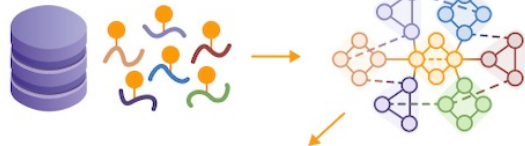
Currently in process of determining the X-ray structure of protein-ligand complex;

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



# Modeling Phosphorylated interactions

- 1 Data collection**  
Search for phosphopeptide-protein complexes in the PDB
- 2 Clustering**  
HHsearch against the PDB100 database to get a non-redundant test set



### 3 Preprocessing

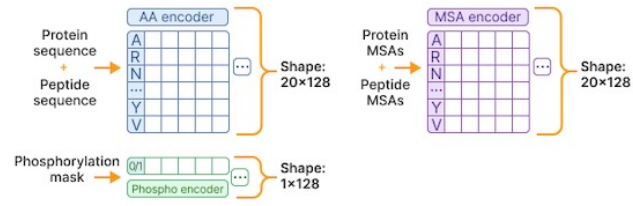
Protein sequence: F S T A A F R F G H A A  
 Peptide sequence: H P L V S R R L S S R

### 4 MSAs generation



### 5 Alphafold model

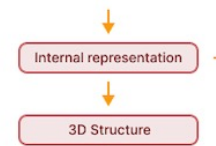
#### 5.1 Input encoding



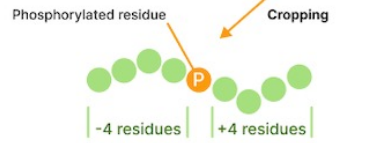
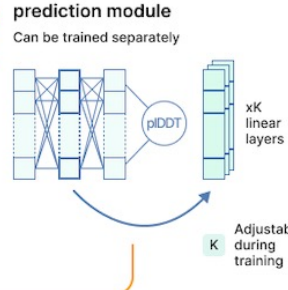
#### 5.2 The Evoformer



#### 5.3 The structure module



#### 5.4 The confidence prediction module



Creating binary masks

Phosphorylation mask

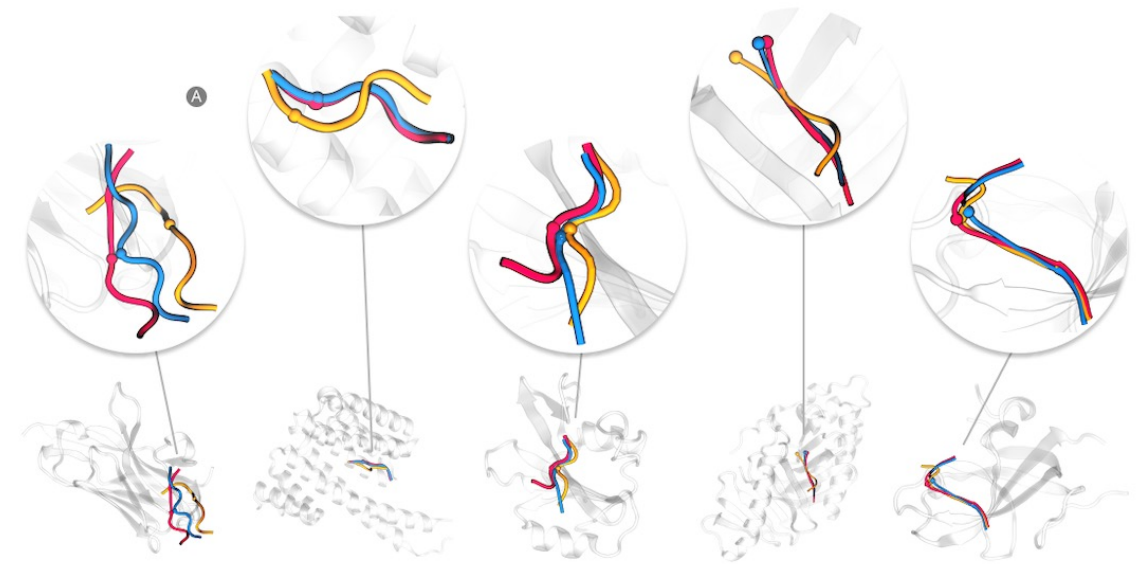
0 0 0 0 1 0 0 0 0

1 Phosphorylated residue

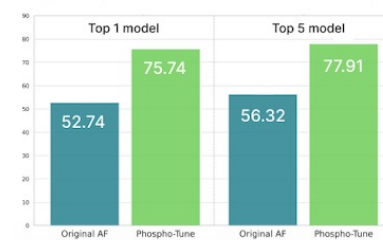
Peptide loss mask

>5A <5A <5A >5A <5A <5A <5A <5A <5A

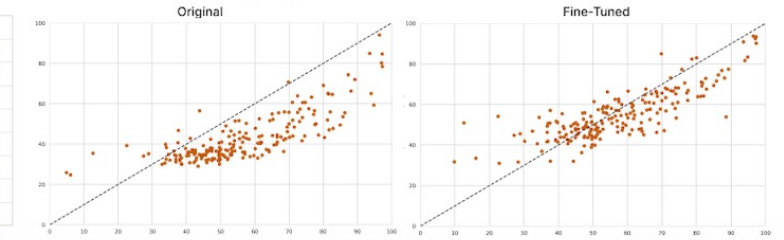
Minimal distance to protein chain is greater or less than a threshold of 5 Å



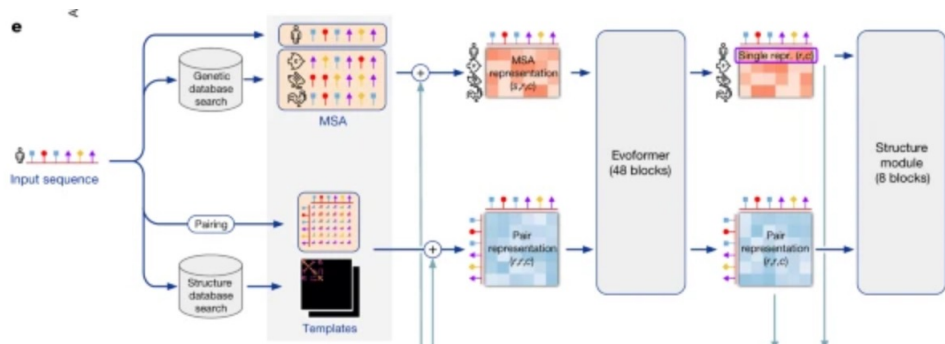
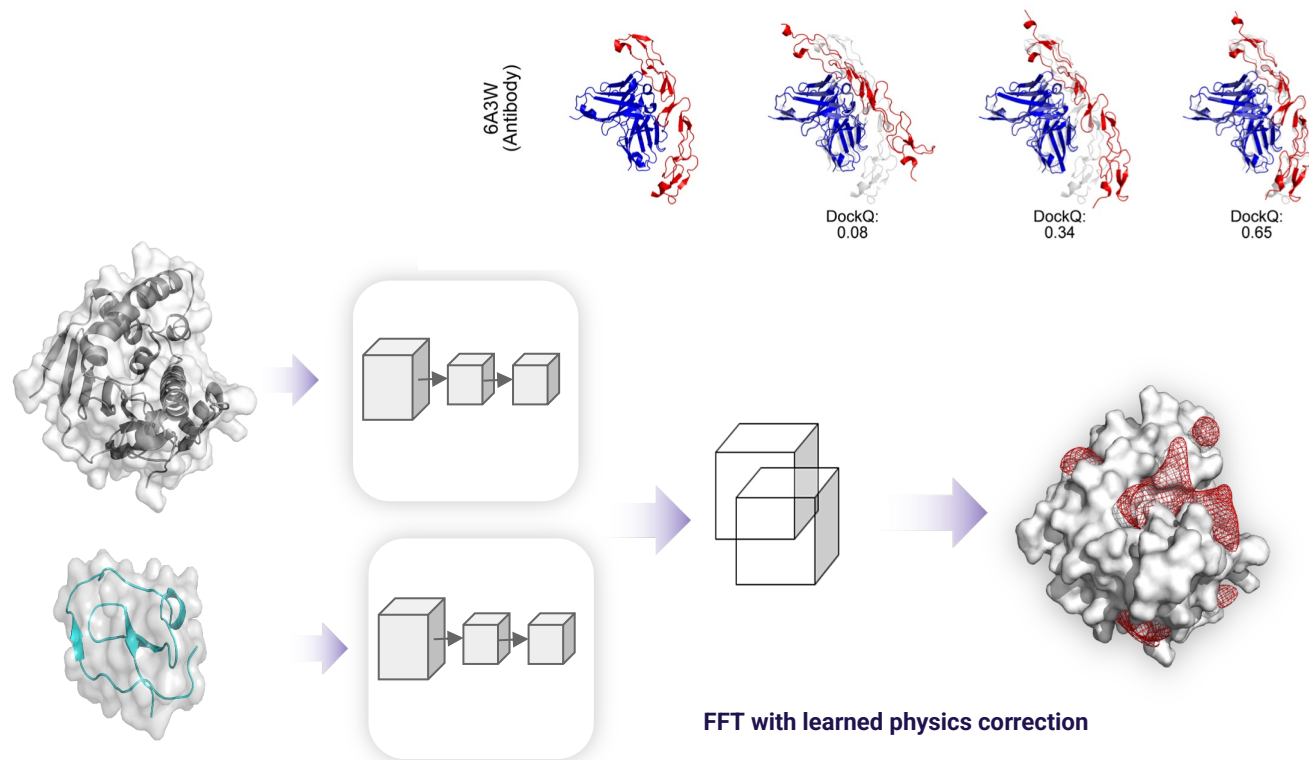
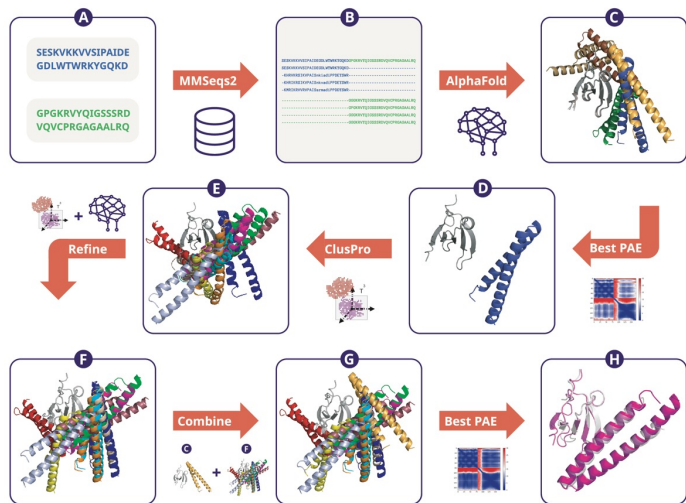
### B Comparison of Model Prediction Quality



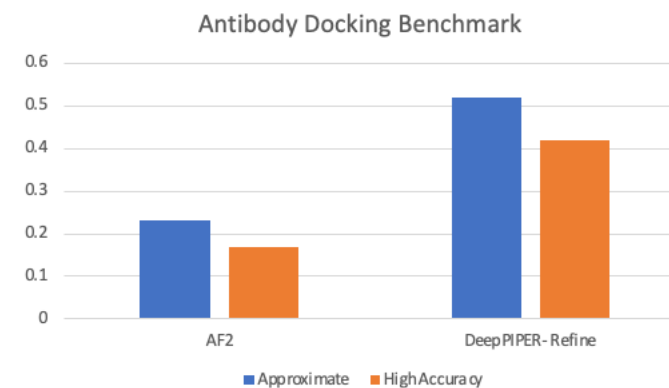
### C Original vs. Fine-Tuned PredictedLDDT



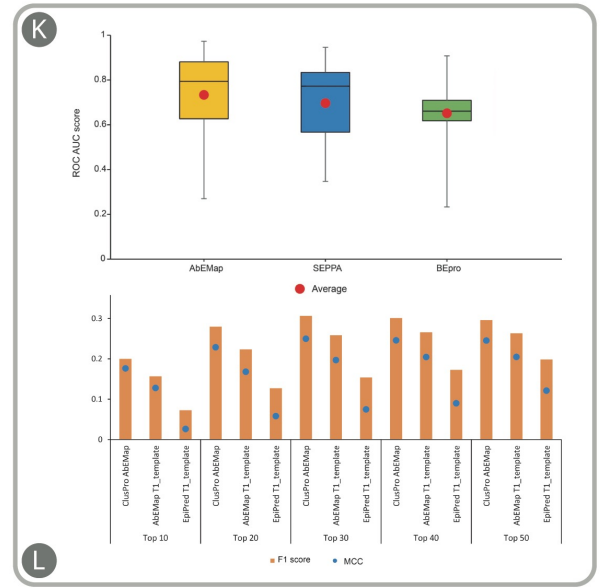
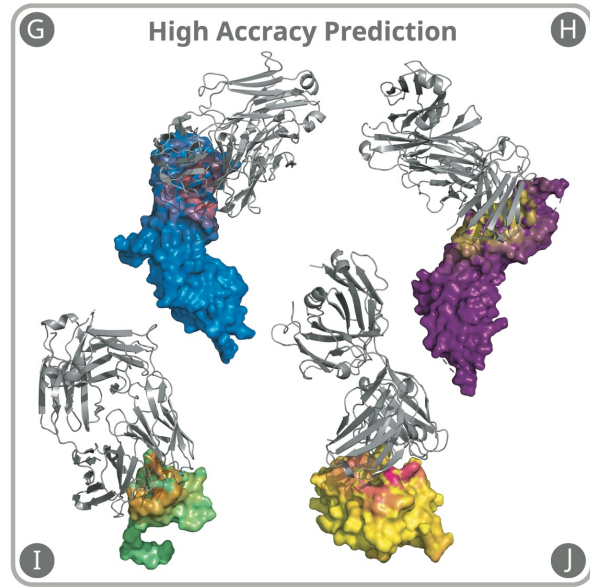
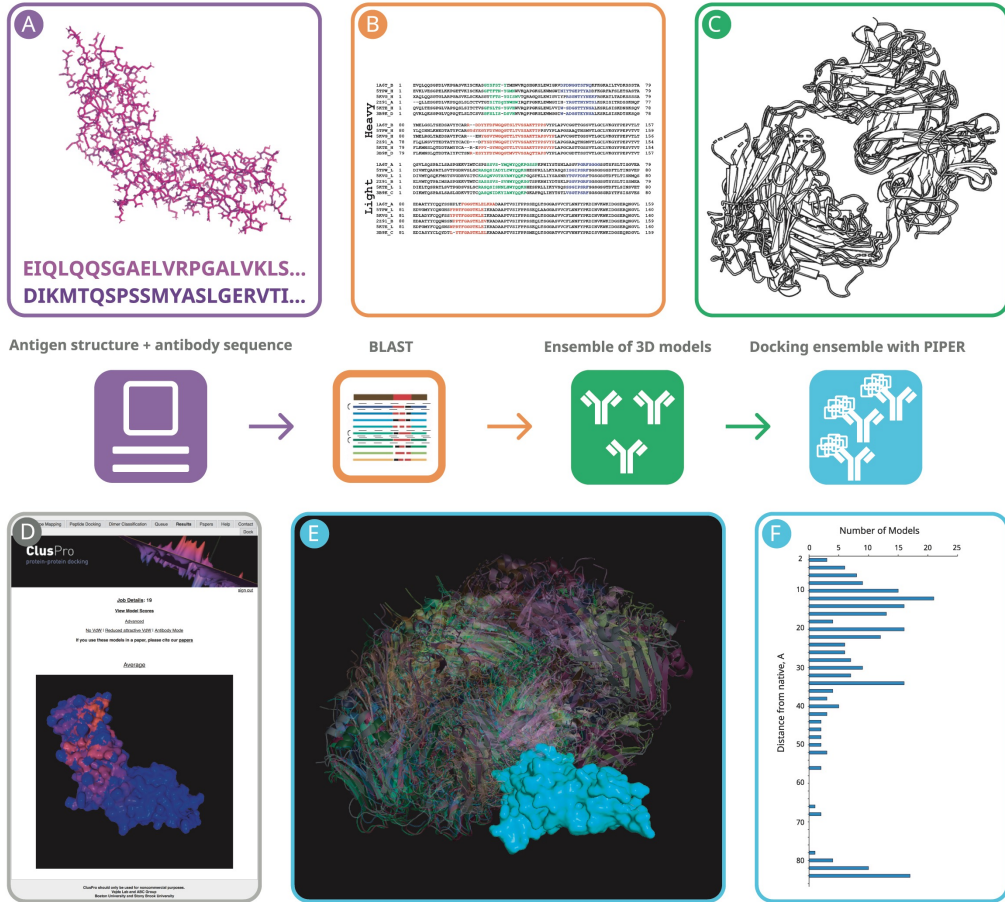
# Combination of FFT based architecture and AF for modeling Antigen Antibodies



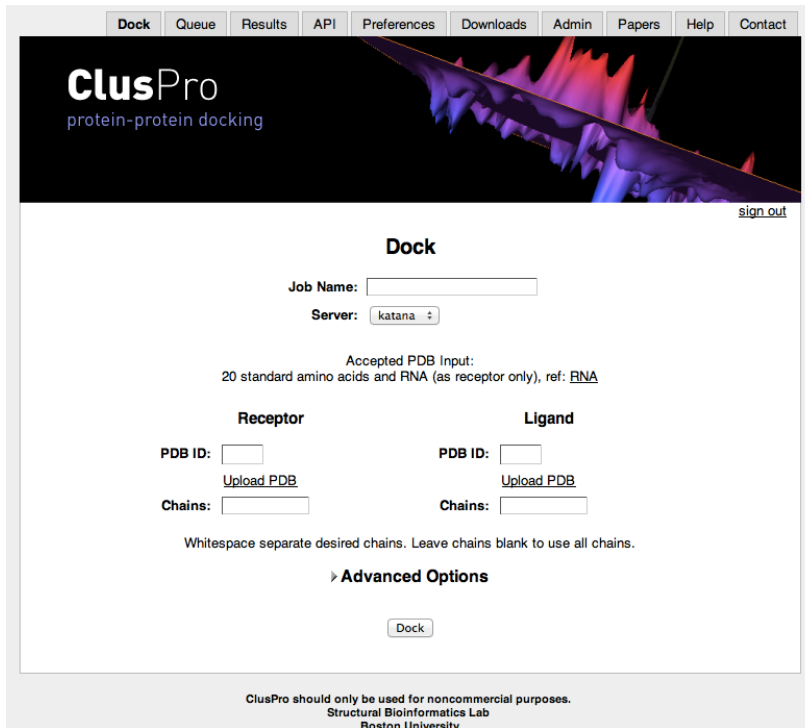
**Top performers in CAPRI – worldwide blind protein docking competition**



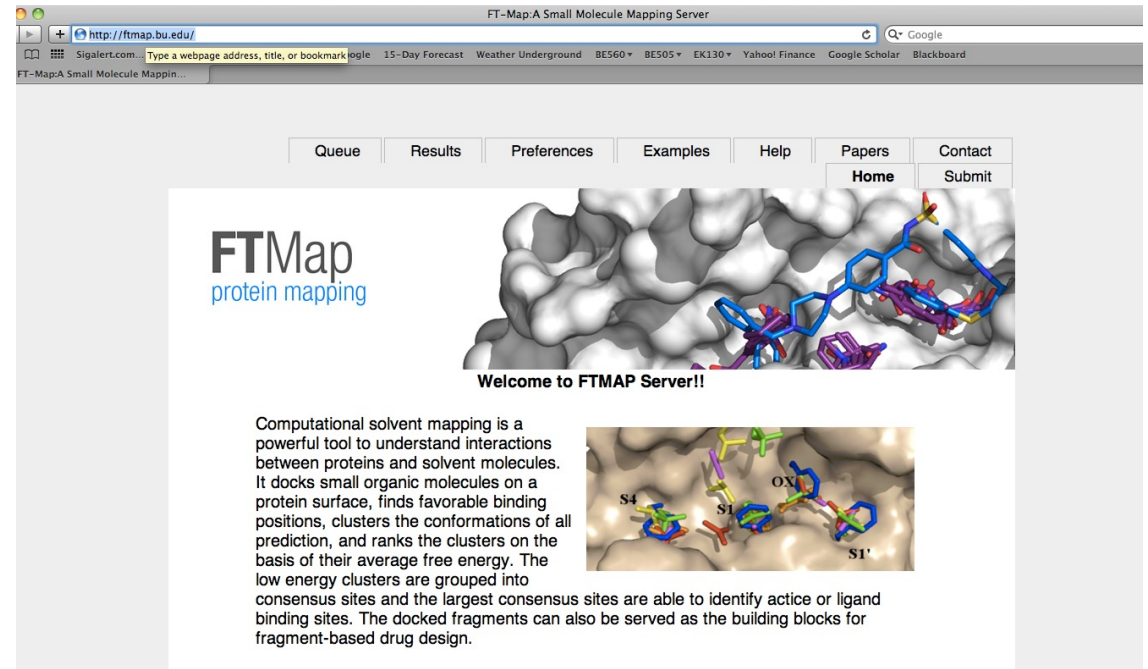
# High Accuracy Epitope Detection



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



The screenshot shows the ClusPro web interface. At the top, there is a navigation menu with links for Dock, Queue, Results, API, Preferences, Downloads, Admin, Papers, Help, and Contact. Below the menu is a header with the ClusPro logo and the text "protein-protein docking". A "sign\_out" link is visible in the top right corner. The main content area is titled "Dock" and contains a form for submitting a docking job. The form includes a "Job Name" input field, a "Server" dropdown menu currently set to "katana", and a note about "Accepted PDB Input: 20 standard amino acids and RNA (as receptor only), ref: RNA". There are two columns for input: "Receptor" and "Ligand". Each column has a "PDB ID" input field with an "Upload PDB" link below it, and a "Chains" input field. A note states "Whitespace separate desired chains. Leave chains blank to use all chains." Below the input fields is a section for "Advanced Options" with a "Dock" button. At the bottom of the page, there is a disclaimer: "ClusPro should only be used for noncommercial purposes. Structural Bioinformatics Lab Boston University".



The screenshot shows the FTMap web interface. At the top, there is a navigation menu with links for Queue, Results, Preferences, Examples, Help, Papers, Contact, Home, and Submit. Below the menu is a header with the FTMap logo and the text "protein mapping". A large 3D visualization of a protein surface with a docked ligand is shown. Below the visualization is the text "Welcome to FTMAP Server!!". To the right of the text is a smaller 3D visualization of a protein surface with docked fragments labeled S4, S1, OX, and S1'. Below the smaller visualization is a paragraph of text: "Computational solvent mapping is a powerful tool to understand interactions between proteins and solvent molecules. It docks small organic molecules on a protein surface, finds favorable binding positions, clusters the conformations of all prediction, and ranks the clusters on the basis of their average free energy. The low energy clusters are grouped into consensus sites and the largest consensus sites are able to identify active or ligand binding sites. The docked fragments can also be served as the building blocks for fragment-based drug design."

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

# Modeling PROteolysis Targeting Chimeras (PROTACs)

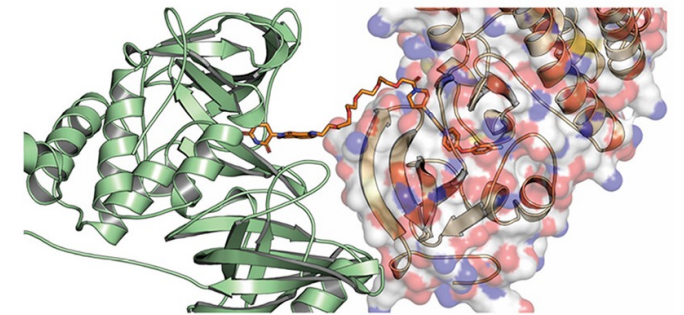
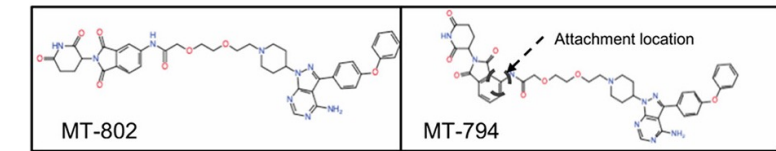
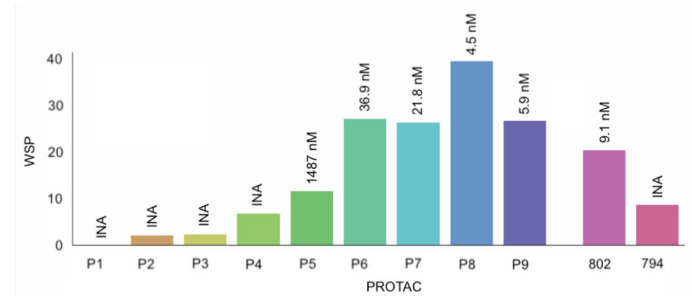
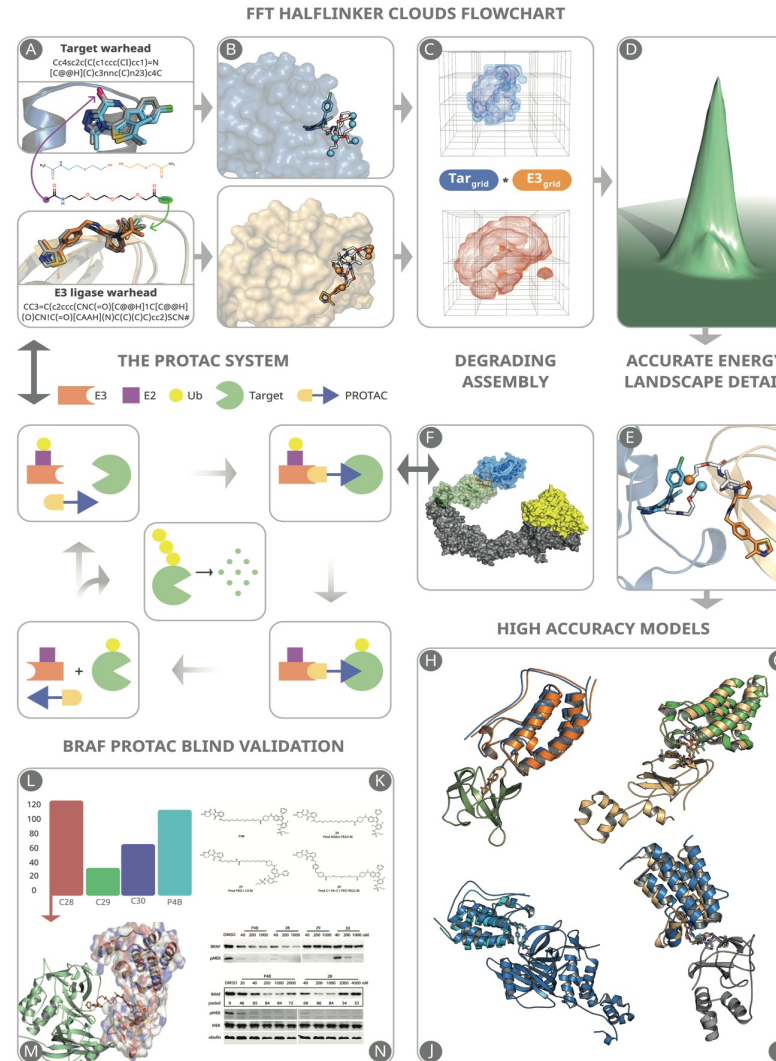
Goal: hijack ubiquitin-proteasome system to degrade target protein

We want to aid PROTAC design:

- PROTAC ternary complex structure
- PROTAC efficiency

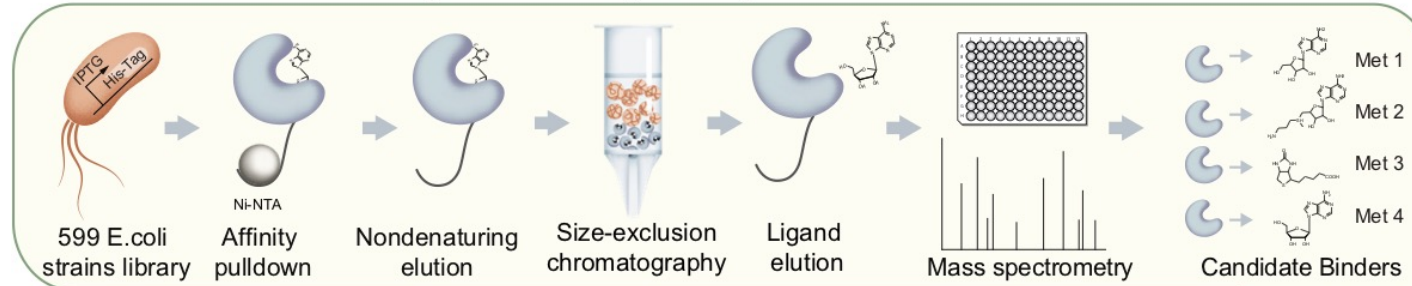
Challenging sampling problem:

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

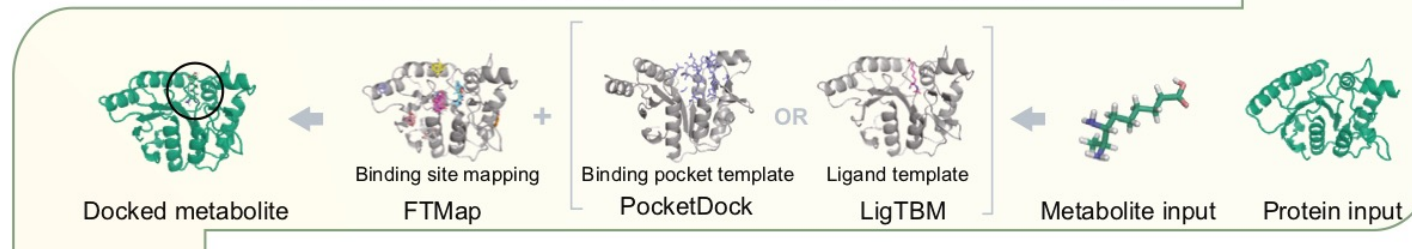


# Large-scale mapping of native protein-metabolite interactions in *E. coli* using Mass Spec & LigTBM

## A. Chemical proteomics (LP/MS)

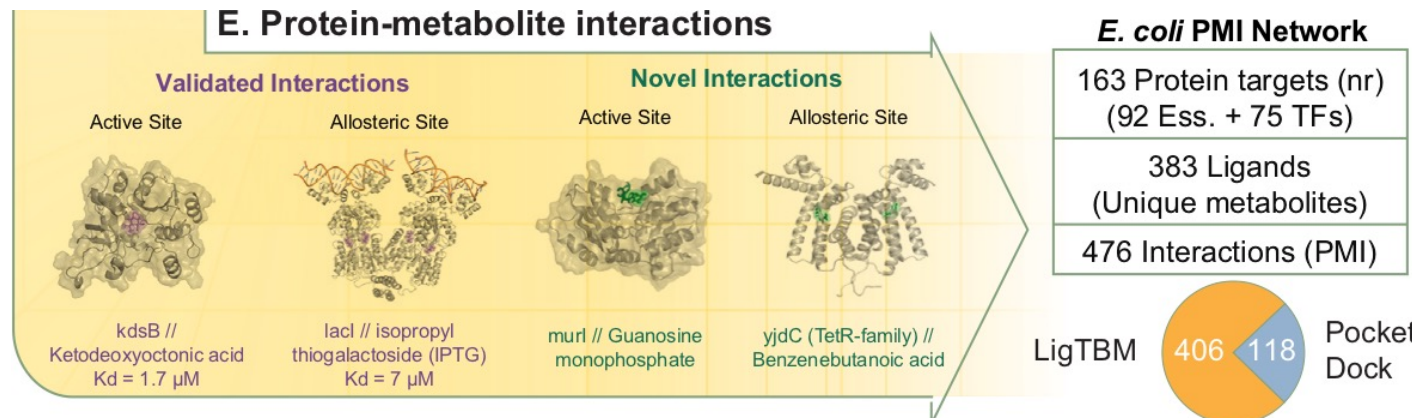


## B. Physics-based structural modeling & benchmarking



...

## E. Protein-metabolite interactions



# Known and predicted protein-metabolite interactions & SPR validation

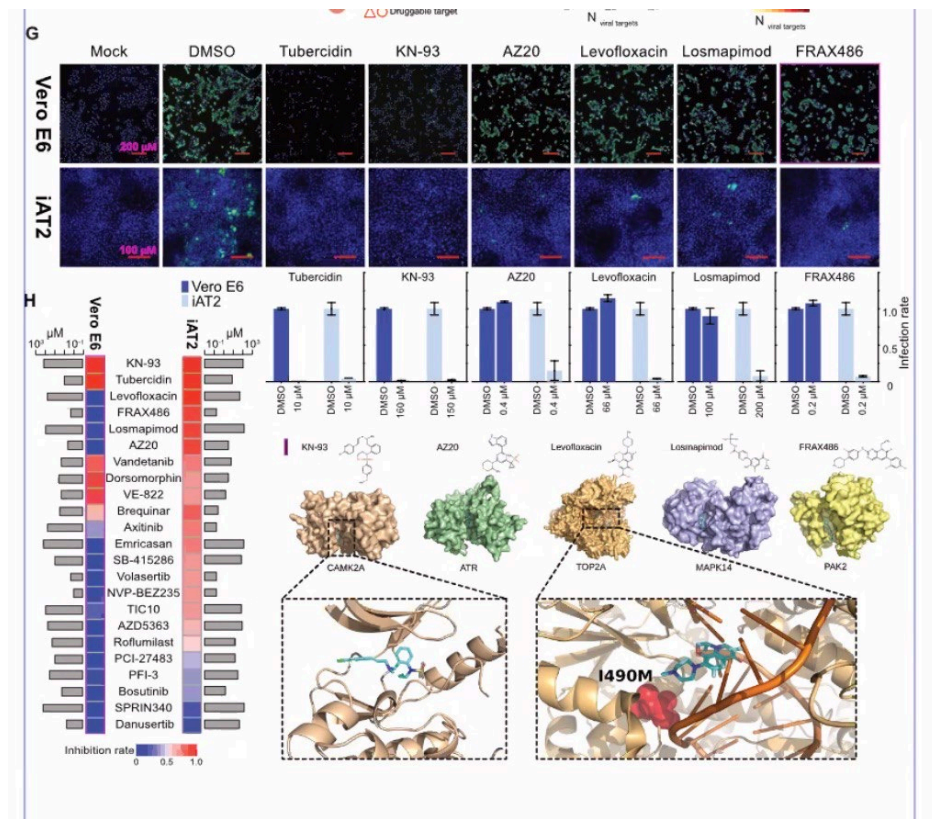
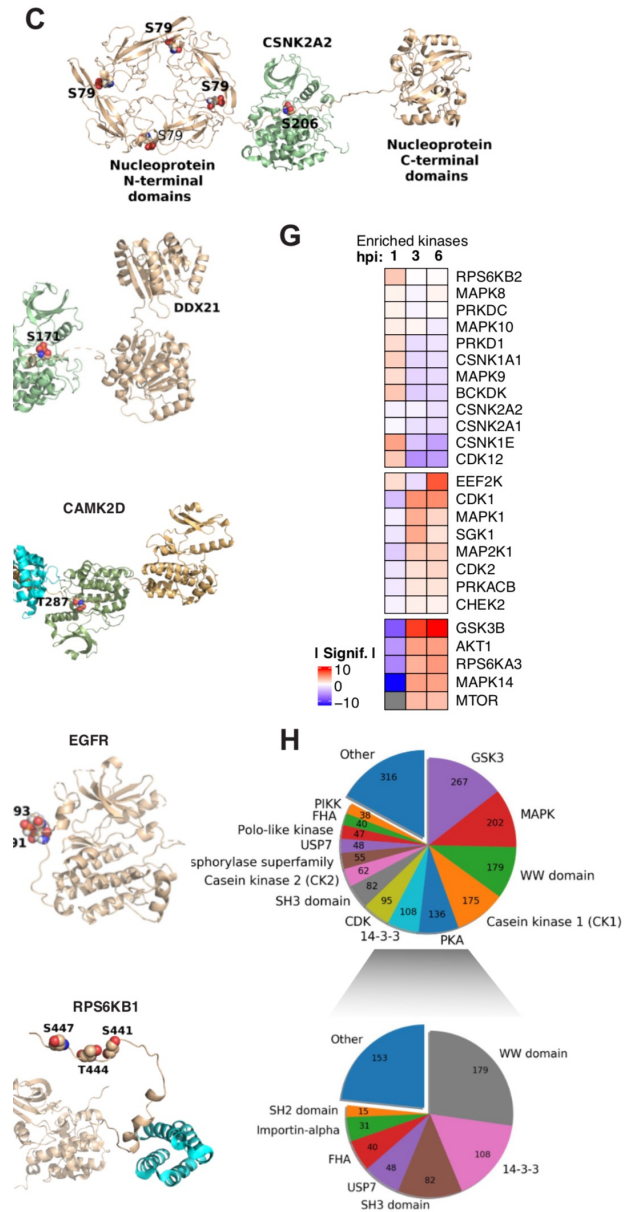
Fmt - Methionyl-tRNA formyltransferase	FldA - Flavodoxin1	GapA - glyceraldehyde-3-phosphate dehydrogenase	ObgE - GTPase involved in chromosome partitioning and ribosome assembly	PyrG - CTP synthetase	AidB - isovaleryl-CoA dehydrogenase & DNA-binding transcriptional repressor	CytR - DNA-binding transcriptional repressor
Dihydrofolate	Riboflavin	ADP-ribose	IDP	Guanosine monophosphate	Flavin-adenine dinucleotide	Cytidine
inhibitor	product	inhibitor	product	allosteric inhibitor	cofactor	allosteric effector
BirA - Putative acyl-CoA dehydrogenase	Adk - Adenylate kinase	HemL - Glutamate-1-semialdehyde 2,1-aminomutase	MetK - S-adenosylmethionine synthase	MurB - UDP-N-acetylenol-pyruvoylglucosamine reductase	TrmD - tRNA (guanine-1)-methyltransferase	TyrS - tyrosine tRNA ligase
Adenosine 5'-monophosphate	2'3'-Cyclic UMP	Pyridoxine 5'-phosphate	S-methyl-5'-Methylthioadenosine	Flavin mononucleotide	S-methyl-5'-thioadenosine	Adenosine monophosphate
product	substrate	cofactor	distant	prosthetic group	distant	product
BirA - Putative acyl-CoA dehydrogenase	FoIA - dihydrofolate reductase	IspE - 4-diphosphocytidyl-2-C-methylerythritol kinase	YggD - predicted DNA-binding transcriptional regulator	IspA - geranyl diphosphate/farnesyl diphosphate synthase	YbiH - HTH-type transcriptional dual regulator CccR	HemH - ferrochelatase
Biotin	Folic acid	Cytidine triphosphate	Sphingosine-1-phosphate	Phosphopantetheine	5,10-methylenetetrahydrofolate mono-L-glutamate	Anthranilate
substrate	distant	inhibitor	TF ligand	TF ligand	TF ligand	allosteric effector

## 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
HemH - Ferrochelatase	2.241

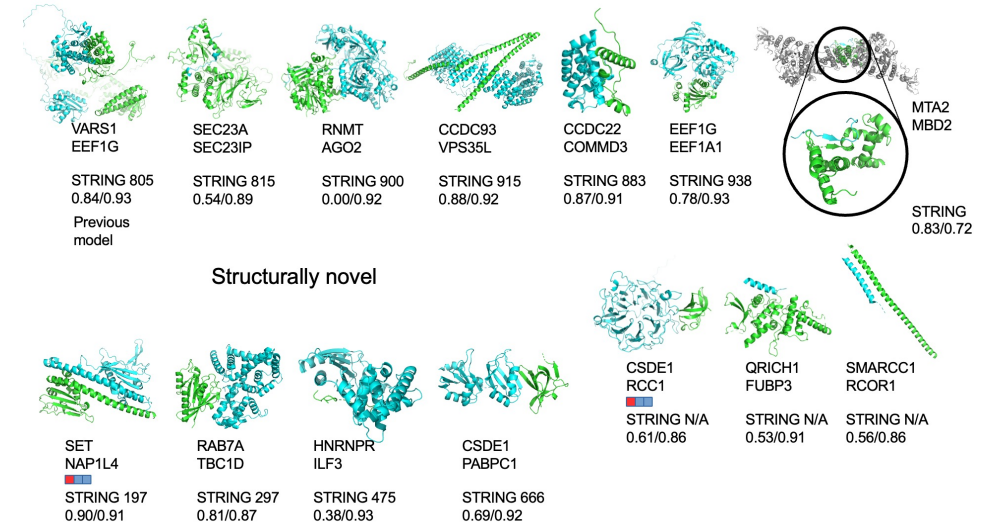
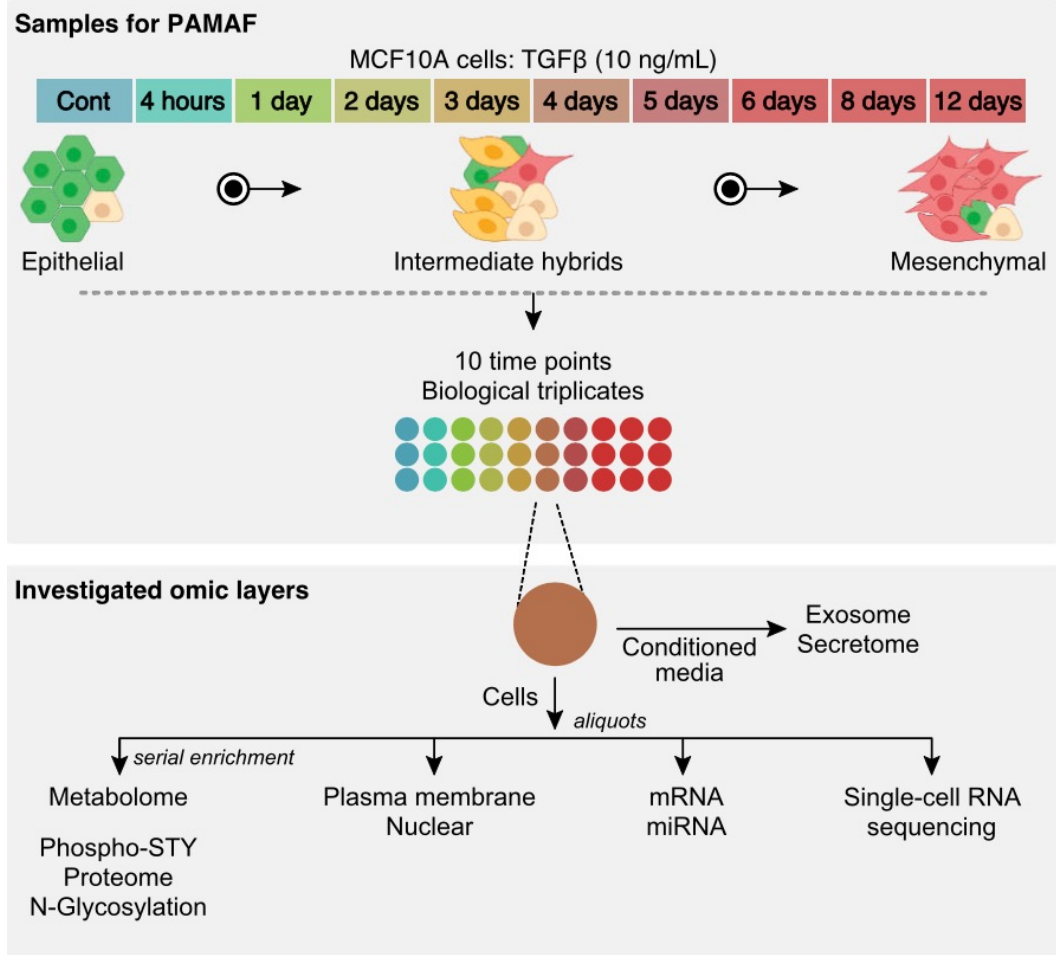


# Phosphorylation effect of SARS-COV2 on infected lung cells – Target identification

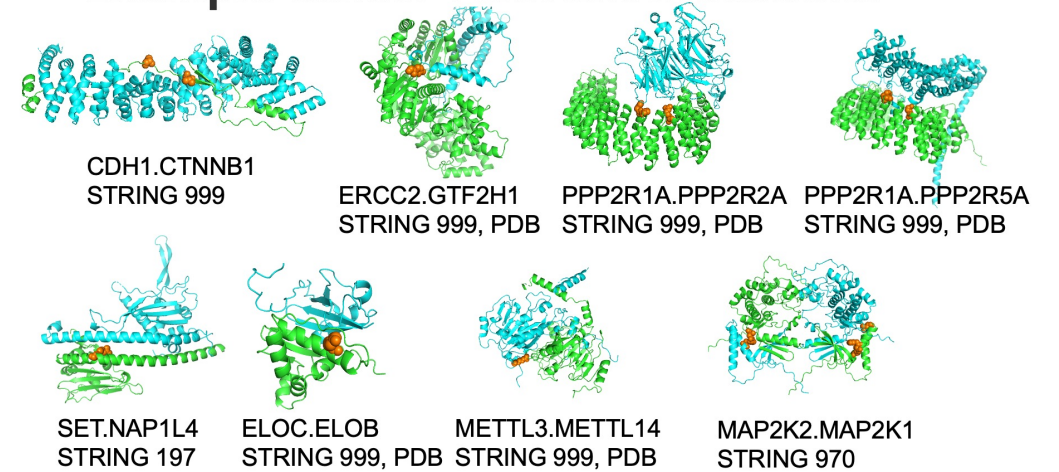


Hume et. al, Molecular Cell, 2021

# Understanding EMT

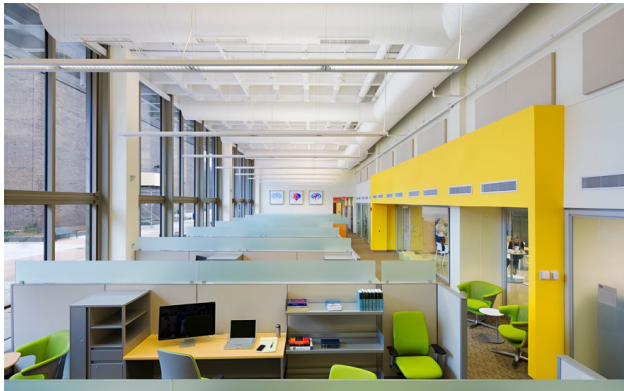


## Example cancer-relevant mutations



## 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 Coutsiass

## U of Toronto, OHSU

Andrew Emili

## BU

Sandor Vajda  
Adrian Whitty

## UNC

Alex Tropsha  
Tim Wilson