Systems level understanding and regulating of the disease with atomic resolution

Dima Kozakov

Laufer Center for Physical and Quantitative Biology

Stony Brook University
FFT convolution approach enables global exhaustive macromolecular interaction sampling

Interaction energy as a sum of FFT convolutions

\[ E(\tilde{t}, r) = \sum_n \sum_{\tilde{x}} R_p(\tilde{x}) L_{pr}(\tilde{t} - \tilde{x}) \]

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

Ligand representation (e.g. charge density)

Receptor representation (e.g. electrostatic potential)

\[ O(N^6) \rightarrow O(N^3 \ln N^3) \]

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

FFT with learned physics correction
Pytorch-AF – Customized Alphafold-style 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

Padhony 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;
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

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 Trapsha (UNC Chapel Hill)
Modeling Phosphorylated interactions

Glukhov et. al, Biorxiv 2024
Combination of FFT based architecture and AF for modeling Antigen Antibodies

Top performers in CAPRI – worldwide blind protein docking competition

Ghani et al Biorxiv 2022; Ashizawa et al in preparation
High Accuracy Epitope Detection

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

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)

- 599 E. coli strains library
- Affinity pulldown
- Nondenaturing elution
- Size-exclusion chromatography
- Ligand elution
- Mass spectrometry
- Candidate Binders

B. Physics-based structural modeling & benchmarking

- Docked metabolite
- Binding site mapping
- FTMMap
- Binding pocket template
- LigTBM
- Metabolite input
- Protein input

E. Protein-metabolite interactions

- Validated Interactions
  - Active Site
  - Allosteric Site
- Novel Interactions
  - Active Site
  - Allosteric Site

E. coli PMI Network

- 163 Protein targets (nr)
  - (92 Ess. + 75 TFs)
- 383 Ligands
  - (Unique metabolites)
- 476 Interactions (PMI)

LigTBM 406 118 PocketDock
Known and predicted protein-metabolite interactions & SPR validation

### Protein-Metabolite Interactions

<table>
<thead>
<tr>
<th>Protein</th>
<th>Metabolite</th>
<th>KD (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HemH</td>
<td>4-aminobenzamide</td>
<td>483</td>
</tr>
<tr>
<td>YjdC</td>
<td>Oleic acid</td>
<td>3.8</td>
</tr>
<tr>
<td>PurB</td>
<td>Phenazine</td>
<td>61.6</td>
</tr>
<tr>
<td>FldA</td>
<td>[2,2'-Bipyridine]-5-carboxylic acid</td>
<td>192</td>
</tr>
<tr>
<td>IspA</td>
<td>(5-Fluoro-2-oxo-2,3-dihydro-1H-indol-3-yl)-acetic acid</td>
<td>54.5</td>
</tr>
<tr>
<td>ObgE</td>
<td>1-isoquinolinyl(phenyl)methanol</td>
<td>119</td>
</tr>
<tr>
<td>FldA</td>
<td>Riboflavin</td>
<td>224</td>
</tr>
<tr>
<td>IspF</td>
<td>Phenazine</td>
<td>60.1</td>
</tr>
<tr>
<td>IspB</td>
<td>1-(3,4-Dimethoxy-phenyl)-ethylamine</td>
<td>118</td>
</tr>
<tr>
<td>TrmD</td>
<td>S-methyl-5'-thioadenosine</td>
<td>14.9</td>
</tr>
<tr>
<td>MetK</td>
<td>S-methyl-5'-thioadenosine</td>
<td>0.355</td>
</tr>
<tr>
<td>UvrY</td>
<td>4-chloro-2-hydroxybenzamide</td>
<td>170</td>
</tr>
<tr>
<td>PyrG</td>
<td>Guanosine</td>
<td>1250</td>
</tr>
<tr>
<td>IspE</td>
<td>Cytidine</td>
<td>67.7</td>
</tr>
<tr>
<td>MurB</td>
<td>Flavin mononucleotide</td>
<td>84500</td>
</tr>
<tr>
<td>HemL</td>
<td>Pyridoxine 5'-phosphate</td>
<td>0.941</td>
</tr>
</tbody>
</table>

### Surface Plasmon Resonance assay

The table above lists the known and predicted protein-metabolite interactions, along with their corresponding dissociation constants (KD) measured in nM. The interactions were validated using the Surface Plasmon Resonance (SPR) assay, a powerful technique for studying the kinetics of biomolecular interactions at the single molecule level.

Key interactions include:
- HemH with 4-aminobenzamide (KD = 483 nM)
- YjdC with Oleic acid (KD = 3.8 nM)
- PurB with Phenazine (KD = 61.6 nM)
- FldA with [2,2'-Bipyridine]-5-carboxylic acid (KD = 192 nM)
- IspA with (5-Fluoro-2-oxo-2,3-dihydro-1H-indol-3-yl)-acetic acid (KD = 54.5 nM)
- ObgE with 1-isoquinolinyl(phenyl)methanol (KD = 119 nM)
- FldA with Riboflavin (KD = 224 nM)
- IspF with Phenazine (KD = 60.1 nM)
- IspB with 1-(3,4-Dimethoxy-phenyl)-ethylamine (KD = 118 nM)
- TrmD with S-methyl-5'-thioadenosine (KD = 14.9 nM)
- MetK with S-methyl-5'-thioadenosine (KD = 0.355 nM)
- UvrY with 4-chloro-2-hydroxybenzamide (KD = 170 nM)
- PyrG with Guanosine (KD = 1250 nM)
- IspE with Cytidine (KD = 67.7 nM)
- MurB with Flavin mononucleotide (KD = 84500 nM)
- HemL with Pyridoxine 5'-phosphate (KD = 0.941 nM)
Phosphorylation effect of SARS-COV2 on infected lung cells –
Target identification

Hume et. al, Molecular Cell, 2021
Understanding EMT

Samples for PAMAF

<table>
<thead>
<tr>
<th>Cont</th>
<th>4 hours</th>
<th>1 day</th>
<th>2 days</th>
<th>3 days</th>
<th>4 days</th>
<th>5 days</th>
<th>6 days</th>
<th>8 days</th>
<th>12 days</th>
</tr>
</thead>
</table>

Epithelial → Intermediate hybrids → Mesenchymal

10 time points

Biological triplicates

Investigated omic layers

Cells → Conditioned media → Exosome Secretome

serial enrichment

Metabolome

Plasma membrane

Nuclear

mRNA

miRNA

Single-cell RNA sequencing

Example cancer-relevant mutations

CDH1, CTNNB1

ERRC2, GTF2H1

PPP2R1A, PPP2R2A

PPP2R1A, PPP2R5A

SET, NAP1L4

ELOC, ELOB

METTL3, METTL14

MAP2K2, MAP2K1

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

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

NIH, NSF
Computer time DOE INCITE leadership award