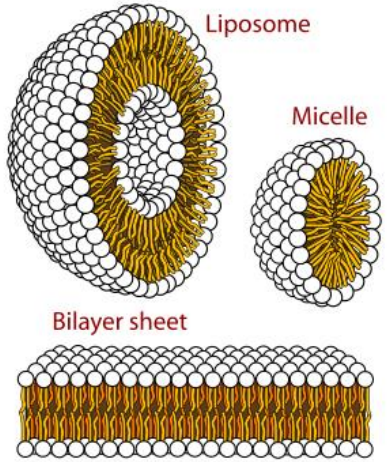
Machine Learning is Revolutionizing Structural Bioinformatics

Jian Wang

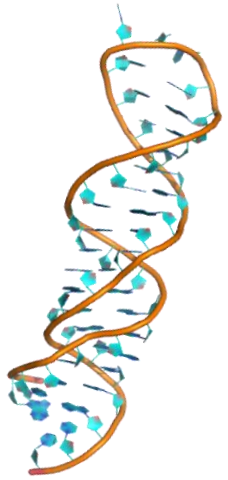
Nov 16, 2022

PennState

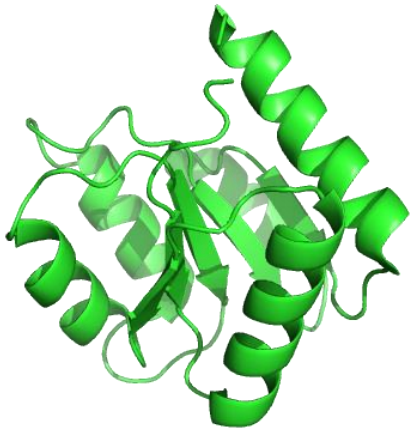
Biomolecules



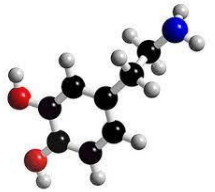
Lipids



Nucleic acids

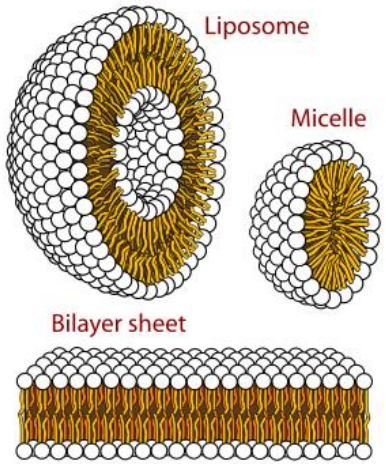


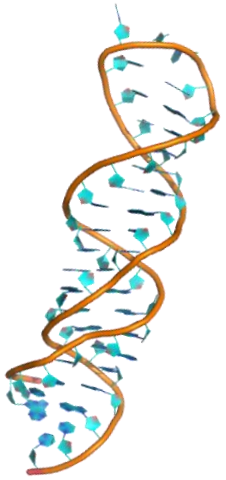
Proteins

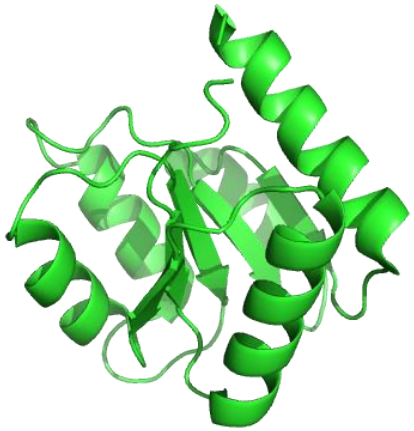


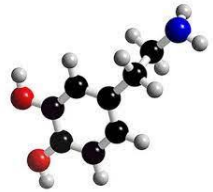
Chemical compounds

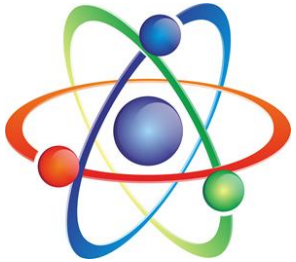
Quantum Computation











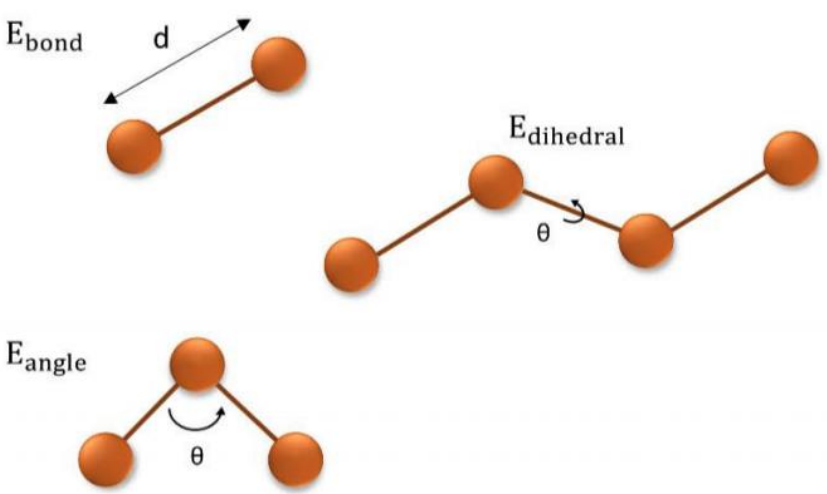
$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$

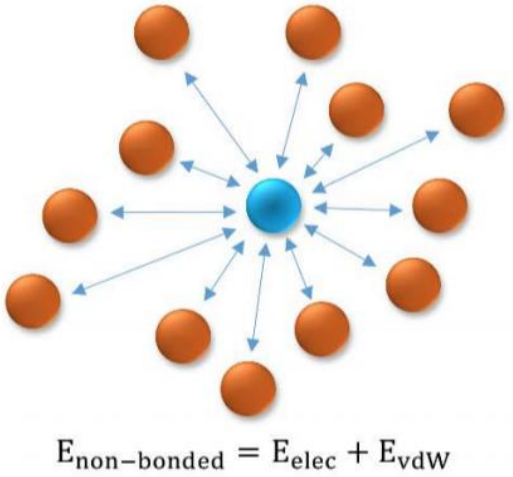
Schrödinger Equation

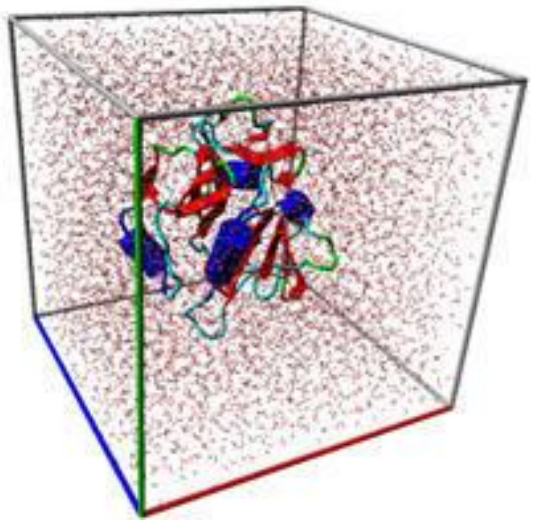
Quantum computation supports calculating the dynamics of a ***small molecule***

Molecular Dynamics Simulation

Force Field



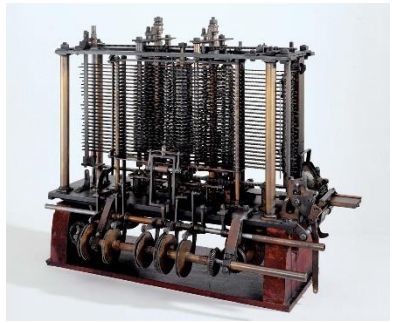




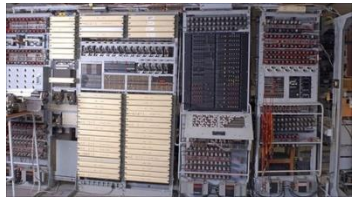
Newton's second law: $F = m \cdot a$

By using force field, molecular dynamics simulation supports the simulation of the dynamics of a ***protein***

Molecular Dynamics Simulation



Analytical Engine



ENIAC

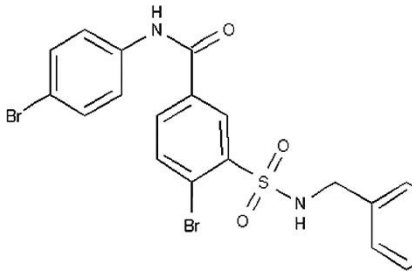
Modern Computer

HPC

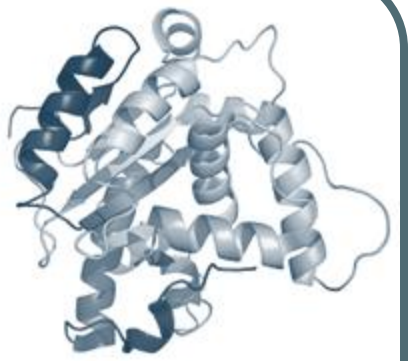
Quantum Computer

Moore's law

Atoms

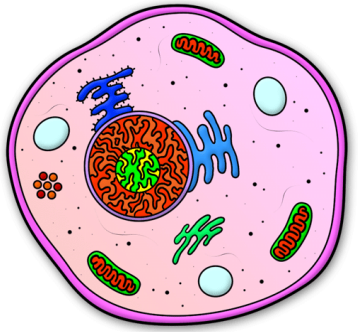


Small molecules



Proteins

Organelles

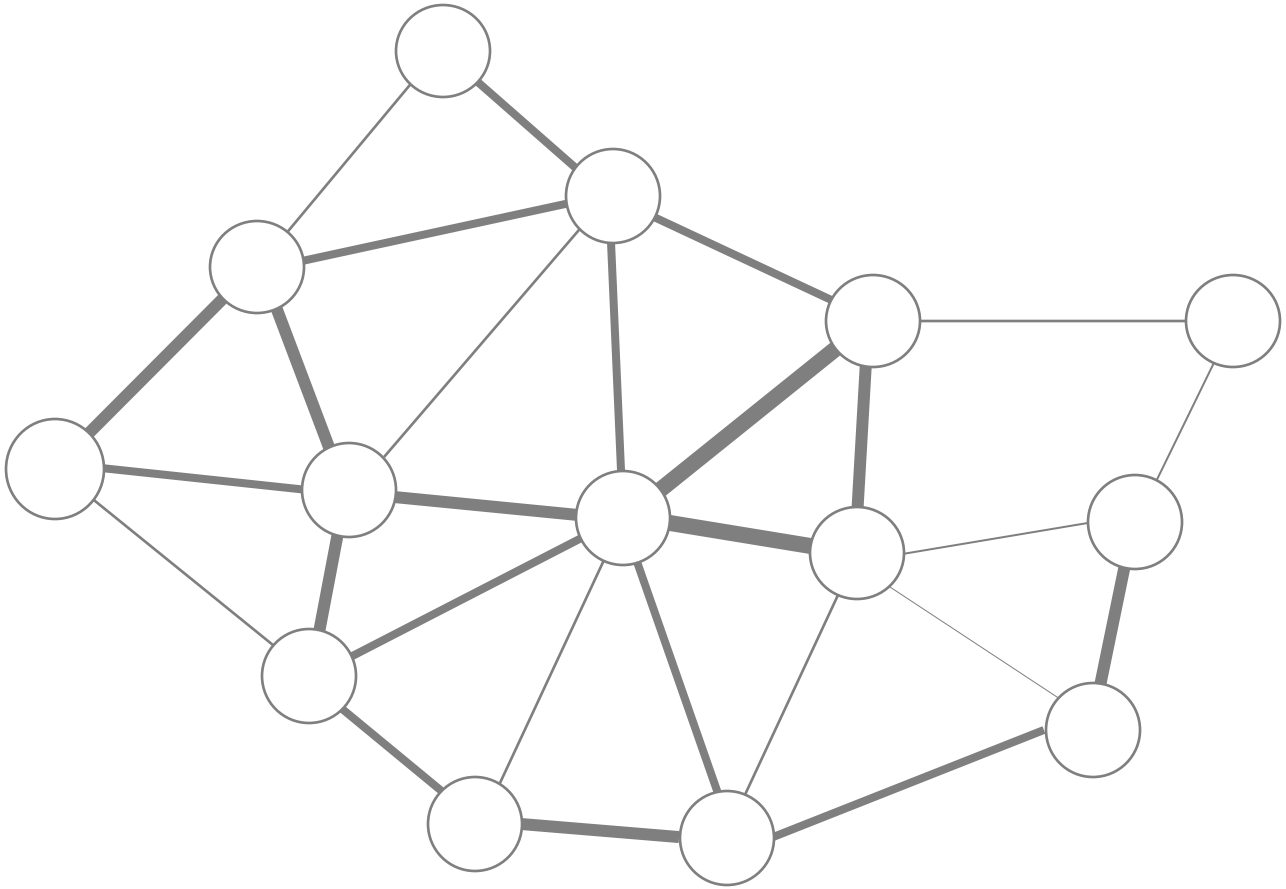


Cells

Graph

Network

Graph

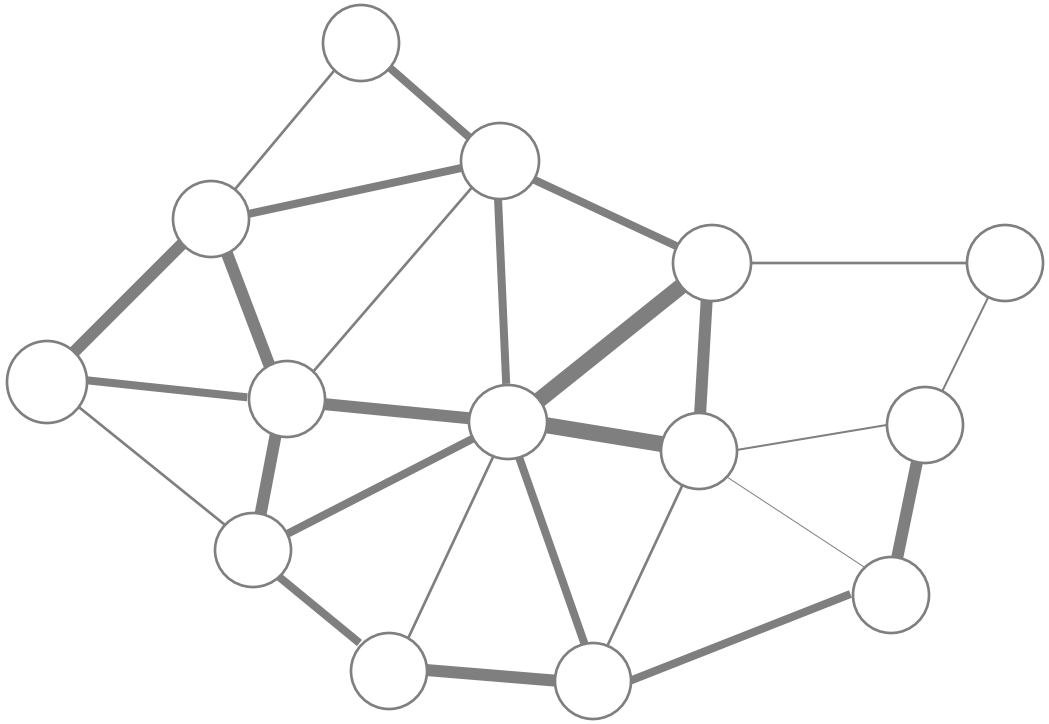


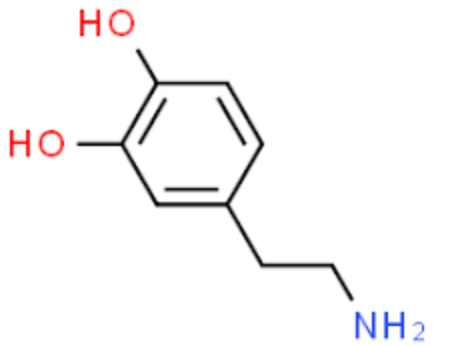
G(N, E)

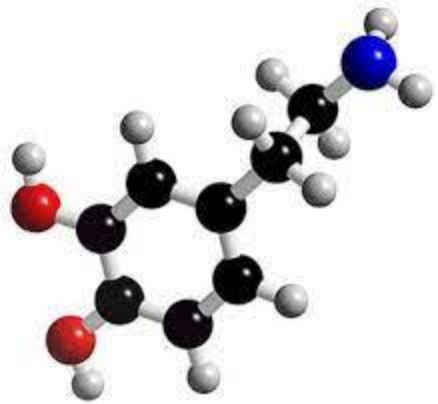
N: nodes

E: edges

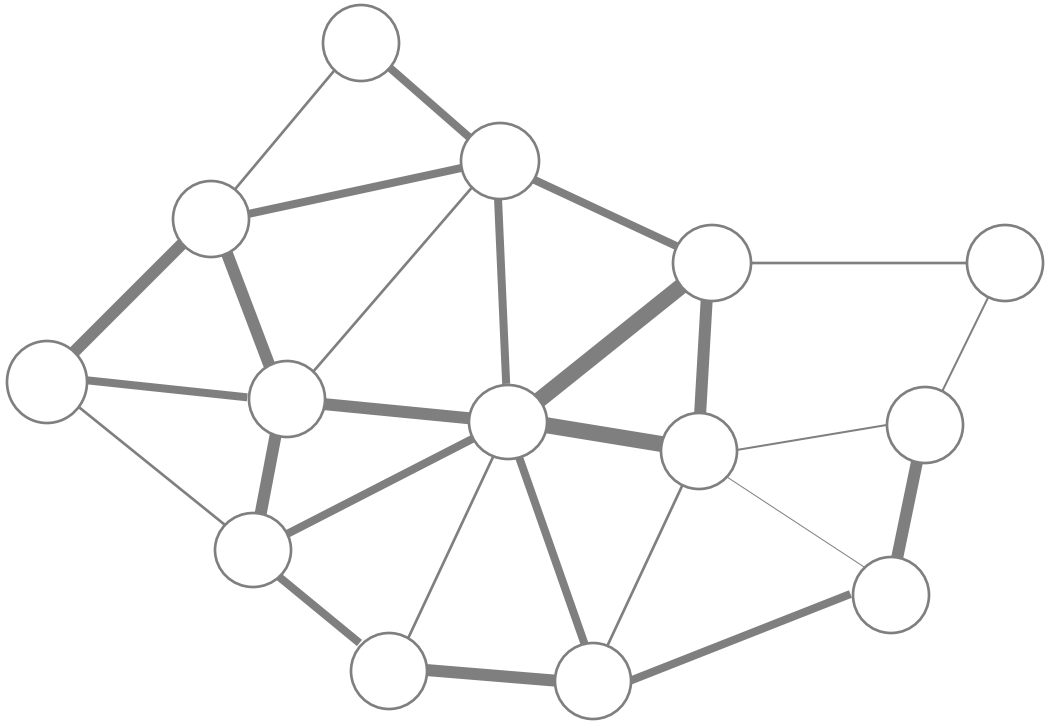
Chemical Compound

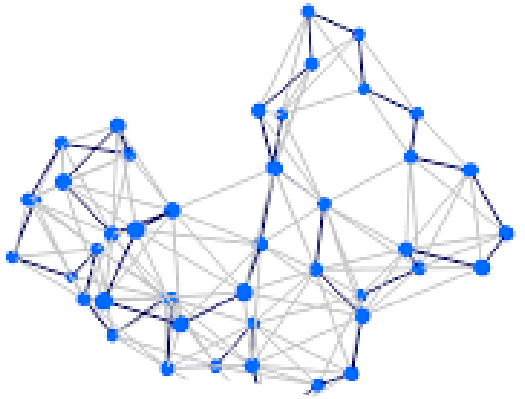




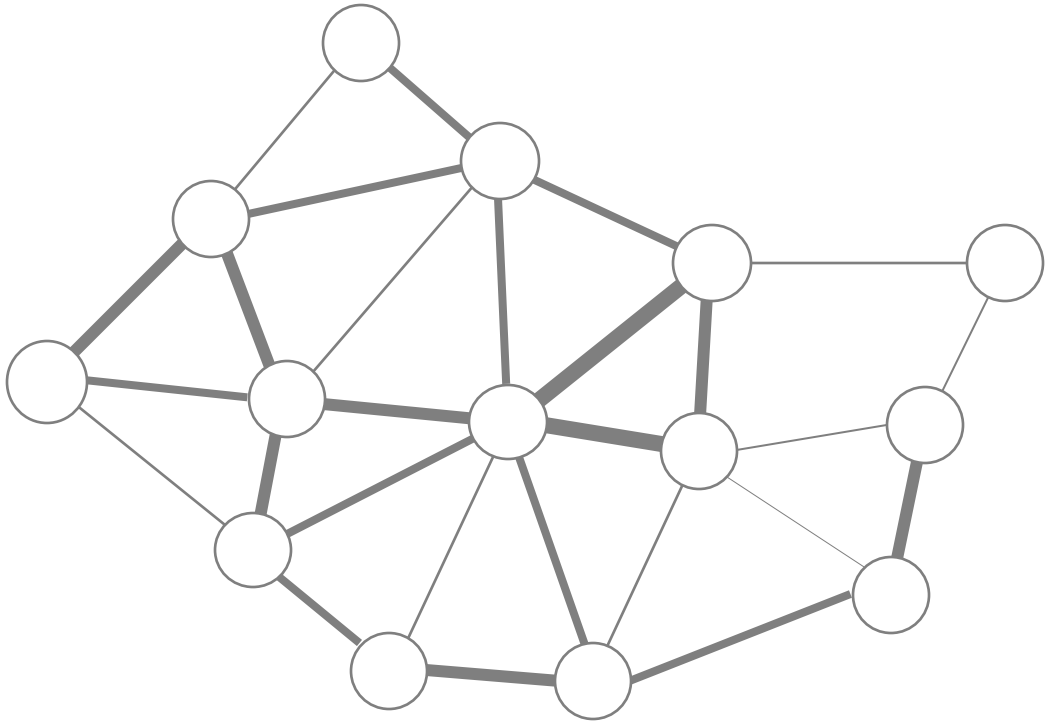


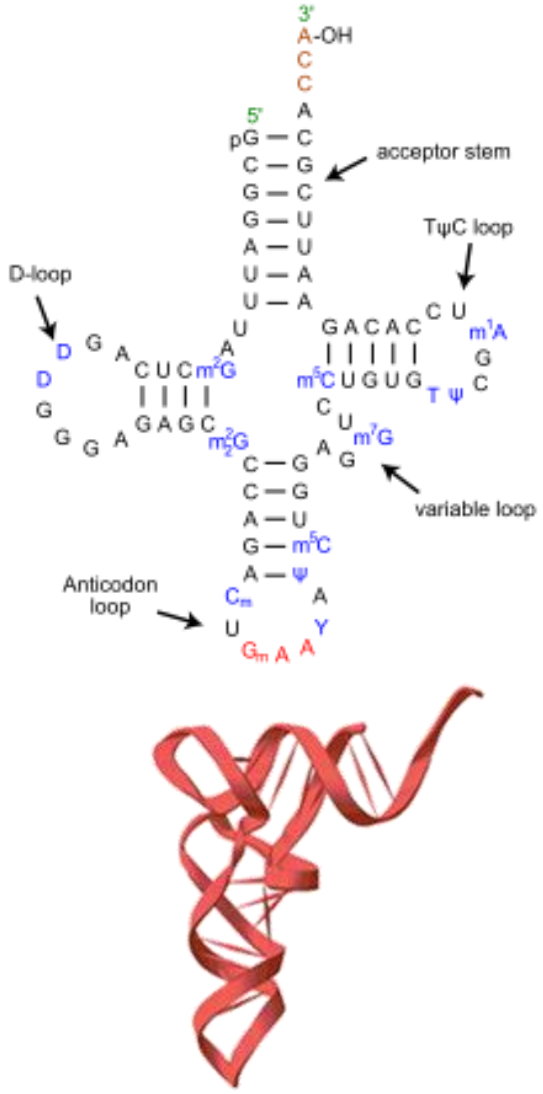
Protein Structure





RNA

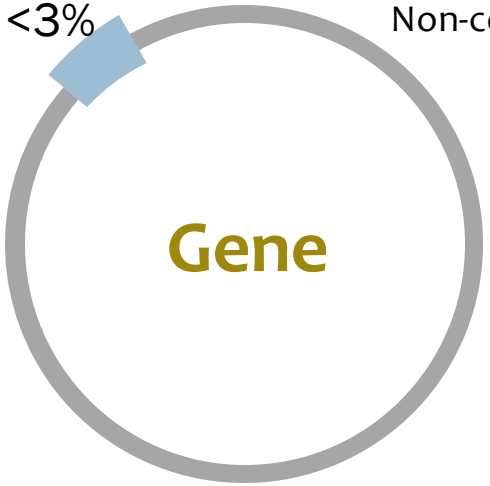



3'
A-OH
C
C
A
5'
pG — C
C — G
G — C
G — U
A — U
U — A
U — A
acceptor stem
TψC loop
D-loop
U
D G A
A
C U C m²G
D
G
G A G C
G G A
m²G
GACAC C U m¹A
m⁵C U G U G T ψ C
C
U
A m⁷G G
C — G
C — G
A — U
G — m⁵C
A — ψ
Cm A
U Y
Gm A A
variable loop
Anticodon
loop

Contents

Non-coding RNA 3D structure prediction

Protein encoding <3% Non-coding



1969

The first manually predicted tertiary structure of tRNA was regarded as a milestone in the emergence of bioinformatics

1989

The core of group I intron was solved based on extensive sequence comparisons, secondary structures, and published mutagenesis data.

1995

31 RNA 3D structures in PDB

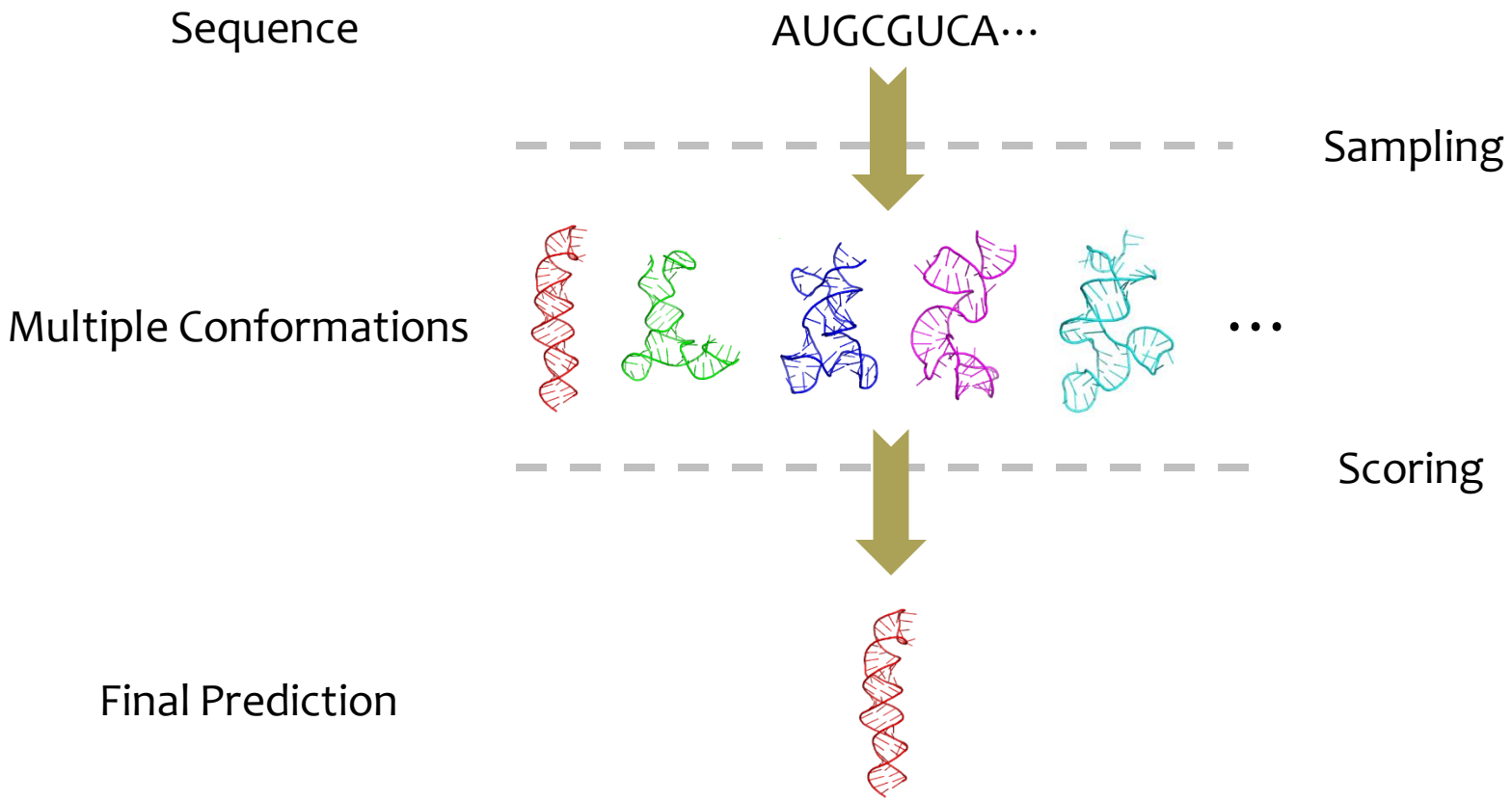
2005

462 RNA 3D structures in PDB

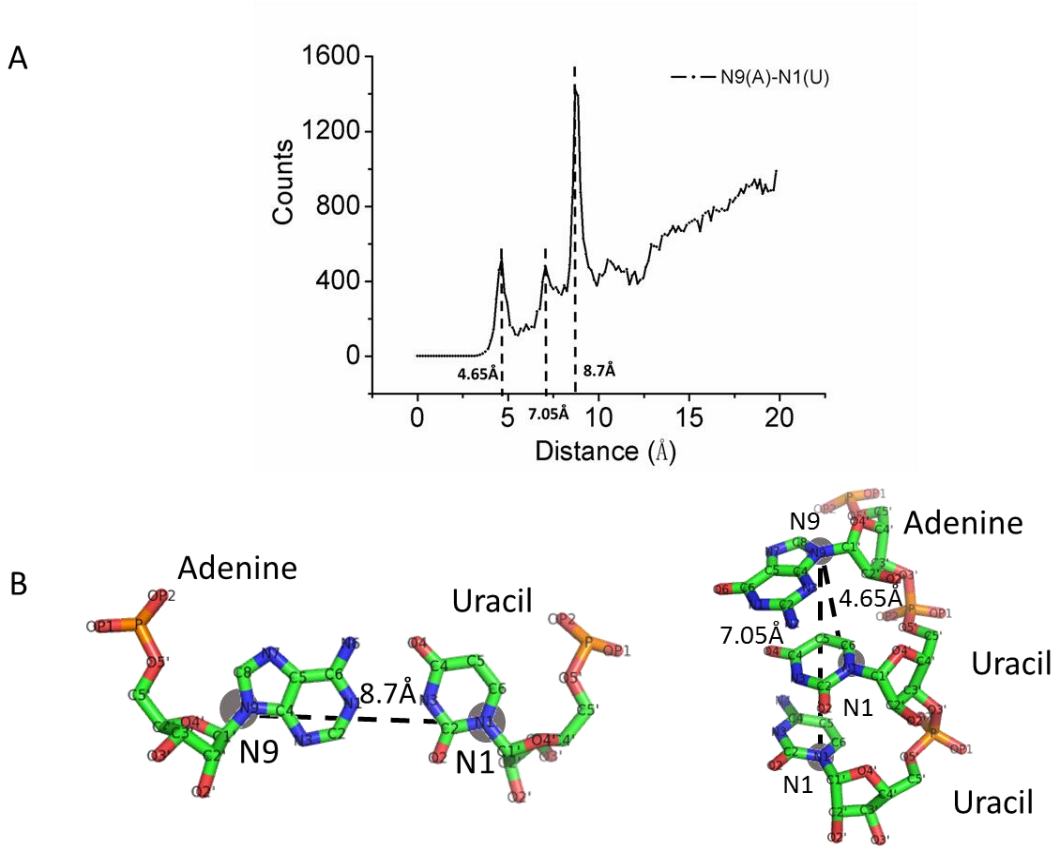
Now

Only **4924** RNA 3D structures in PDB (>**165,266** proteins in PDB)
22,776,905 non-coding RNA sequences in RNAcentral

Difficulties in RNA 3D structure prediction



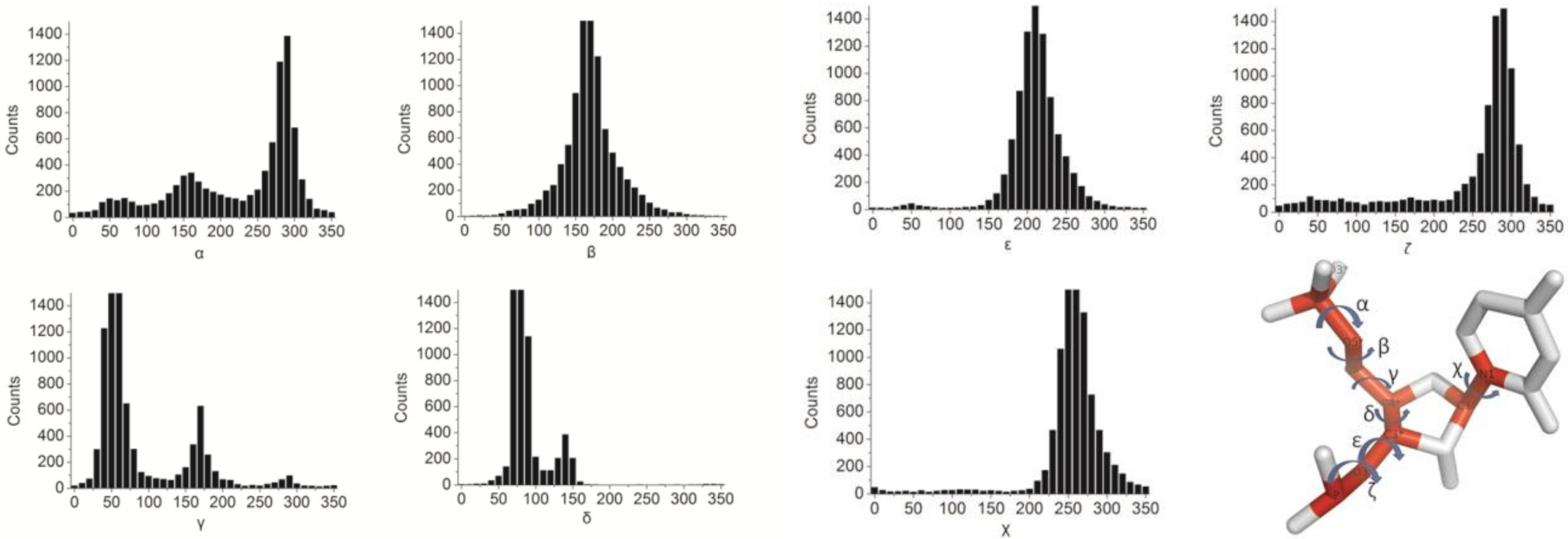
Distance distribution



Base pairing and base stacking information is implied in distance distribution

Wang, Jian, et al. Nucleic acids research 43.10 (2015): e63-e63.

Backbone torsion angle distribution



Backbone conformation is implied in torsion angle distribution

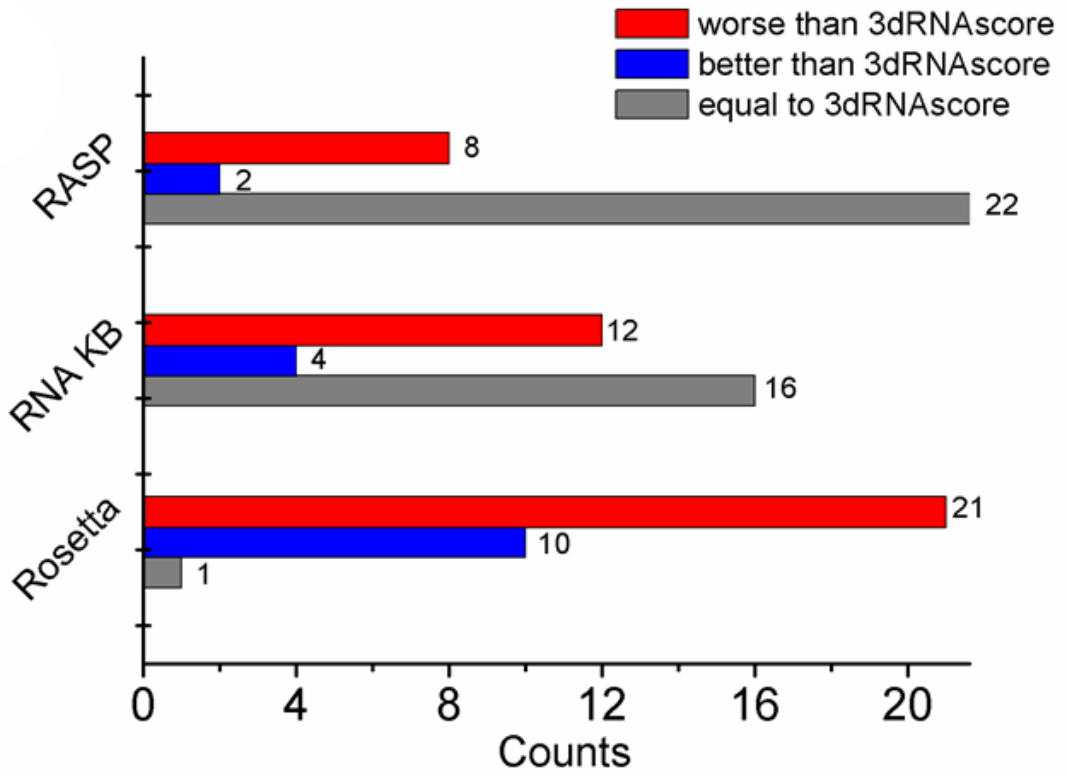
Wang, Jian, et al. Nucleic acids research 43.10 (2015): e63-e63.

Results in Test I, II, and III

$$E_{total} = E_{distance} + \omega E_{dihedral}$$

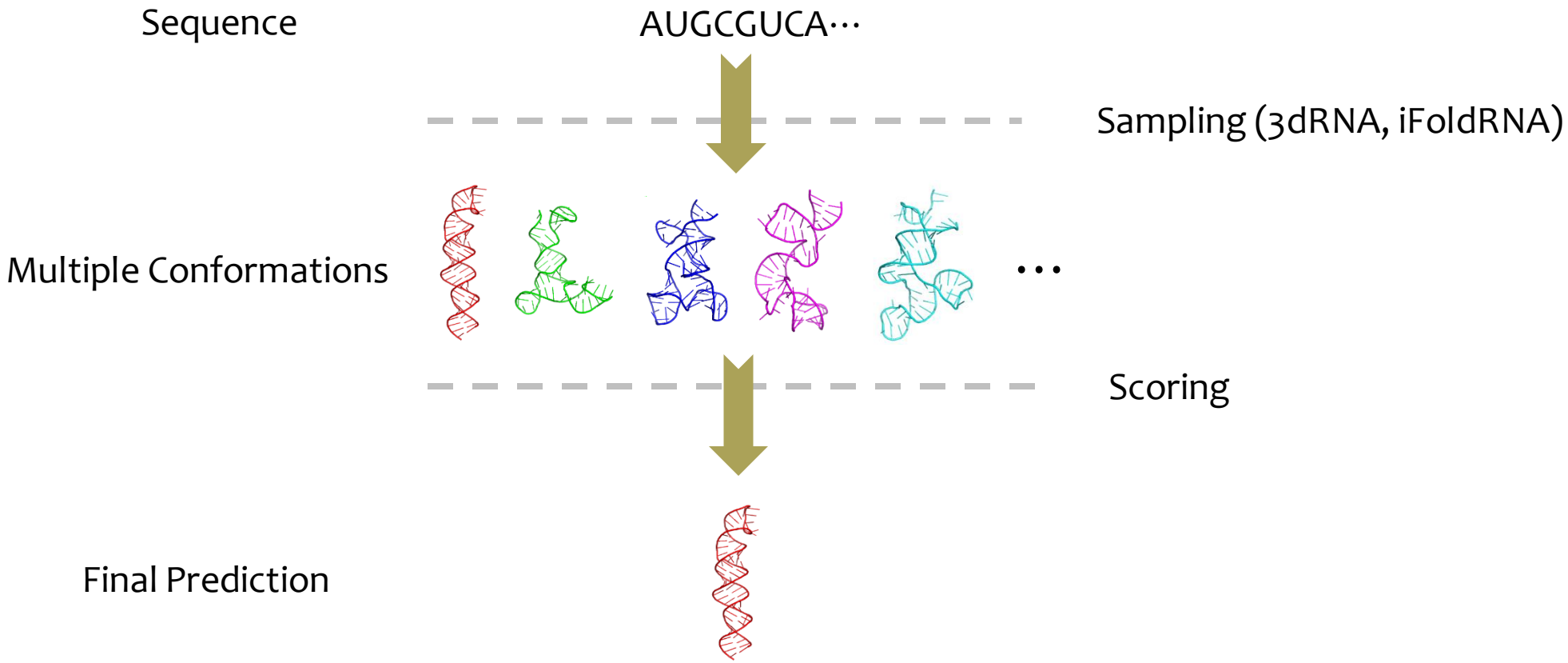
$$E_{distance} = -k_B T \sum_{ij} \ln \frac{f_{ab}^{OBS}(d_{ab}^{ij})}{f_{ab}^{REF}(d_{ab}^{ij})}$$

$$E_{dihedral} = -k_B T \sum_{i} \ln \frac{f_{a}^{OBS}(\theta_{a}^{i})}{f_{a}^{REF}(\theta_{a}^{i})}$$

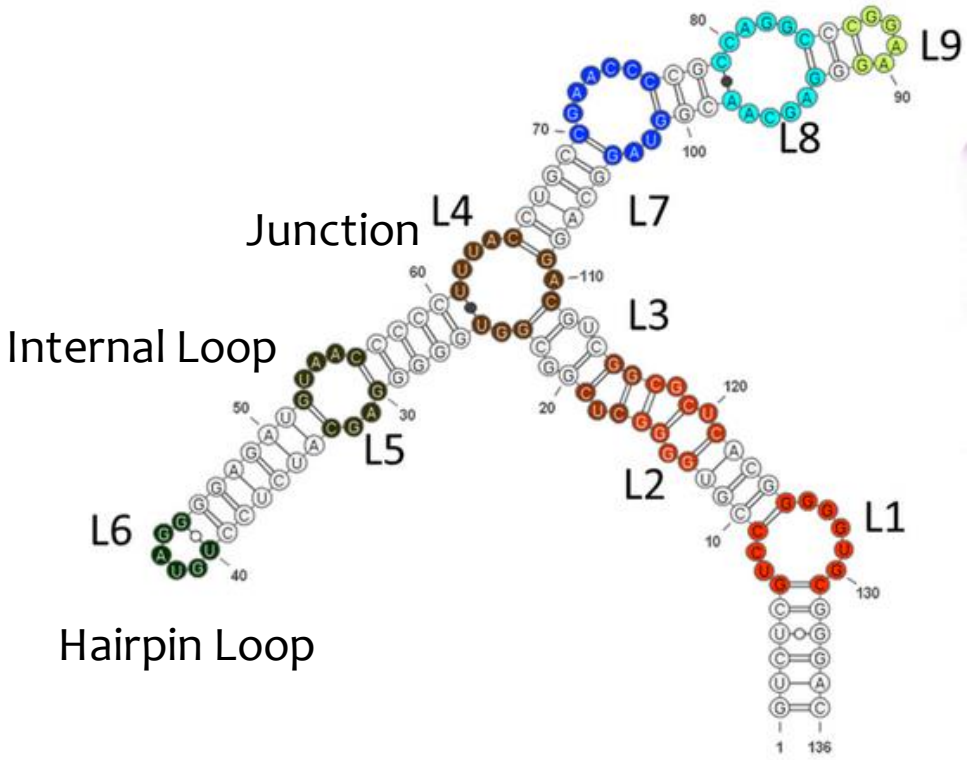


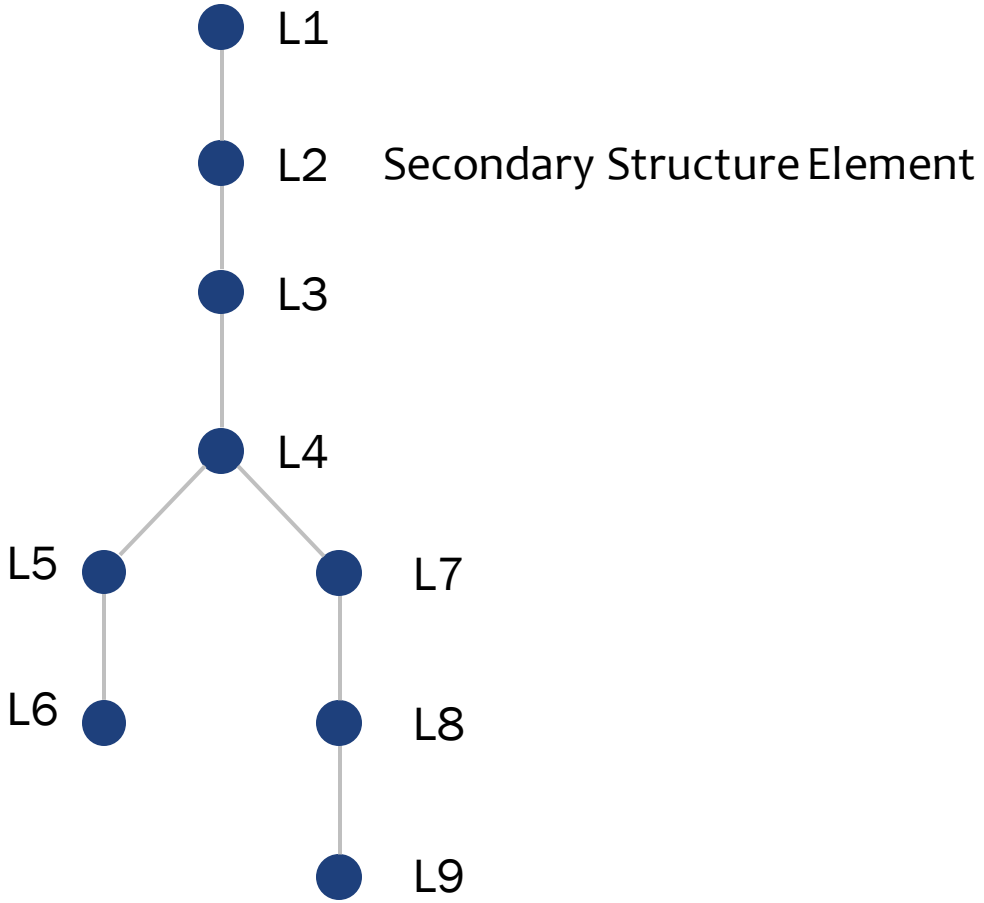
Wang, Jian, et al. Nucleic acids research 43.10 (2015): e63-e63.

Difficulties in RNA 3D structure prediction



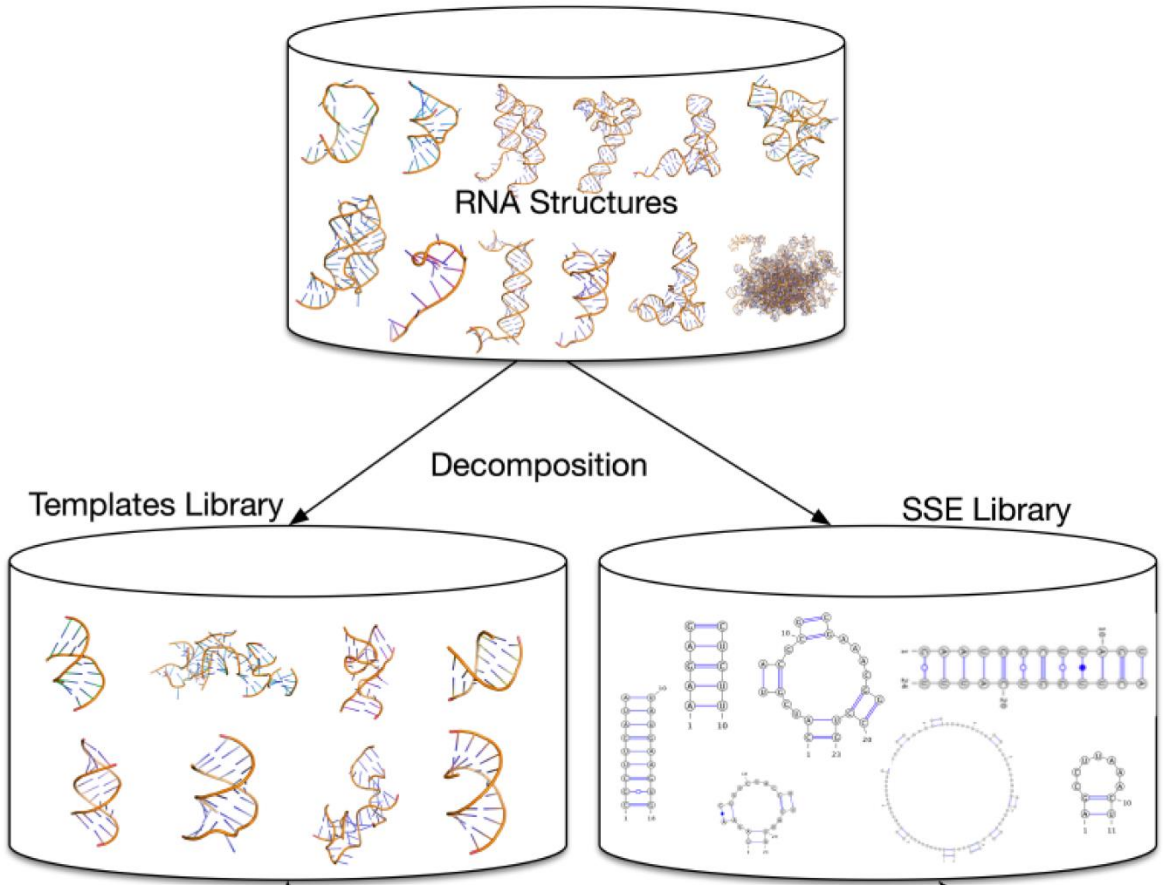
Secondary Structure Tree

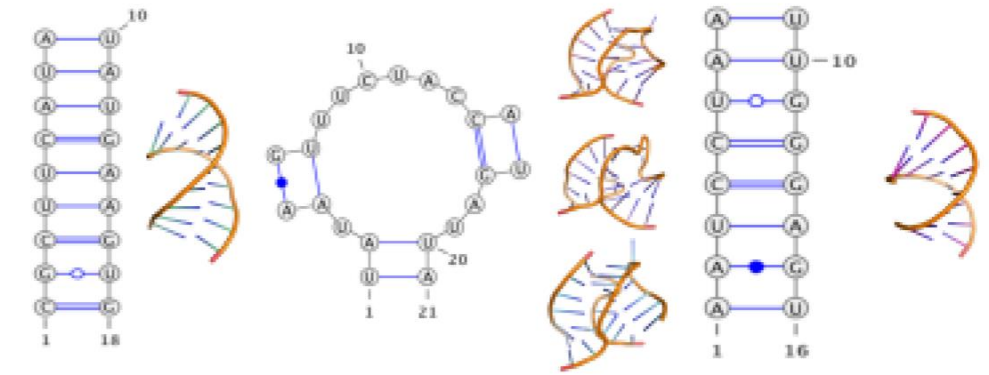


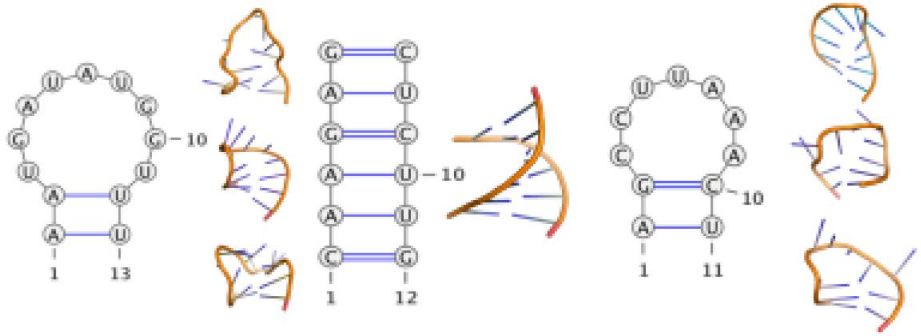


Wang, Jun, Wang, Jian, et al. IJMS. 20.17 (2019): 4116.

Fragment-assembly

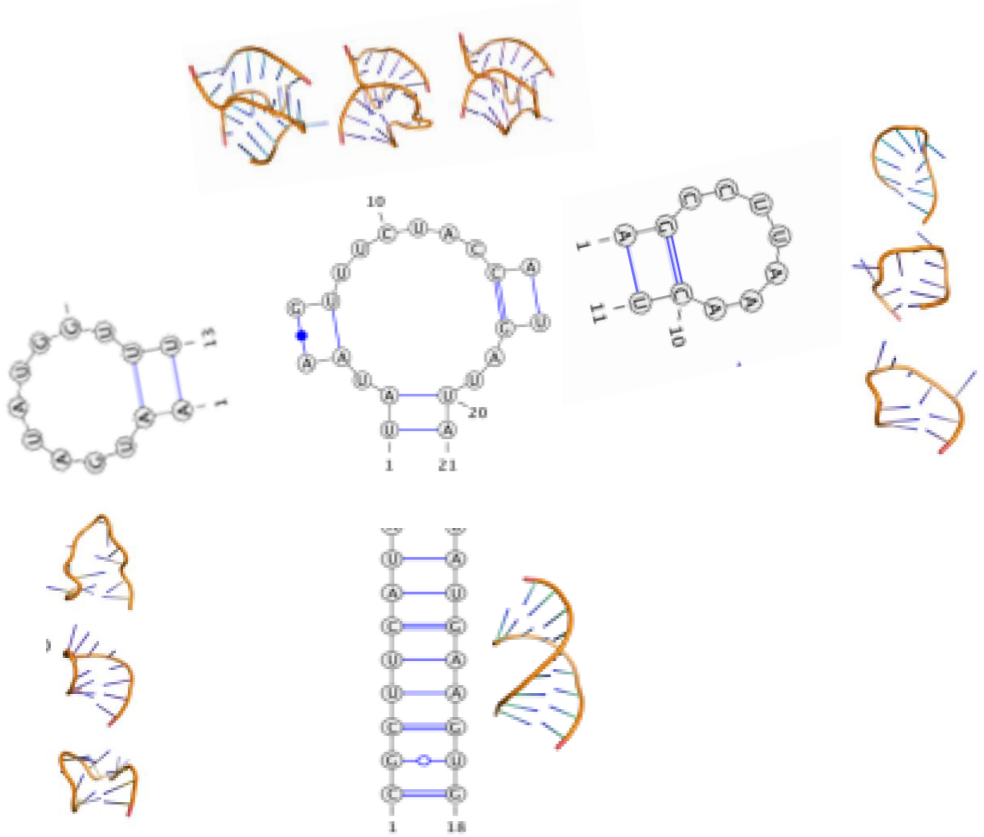






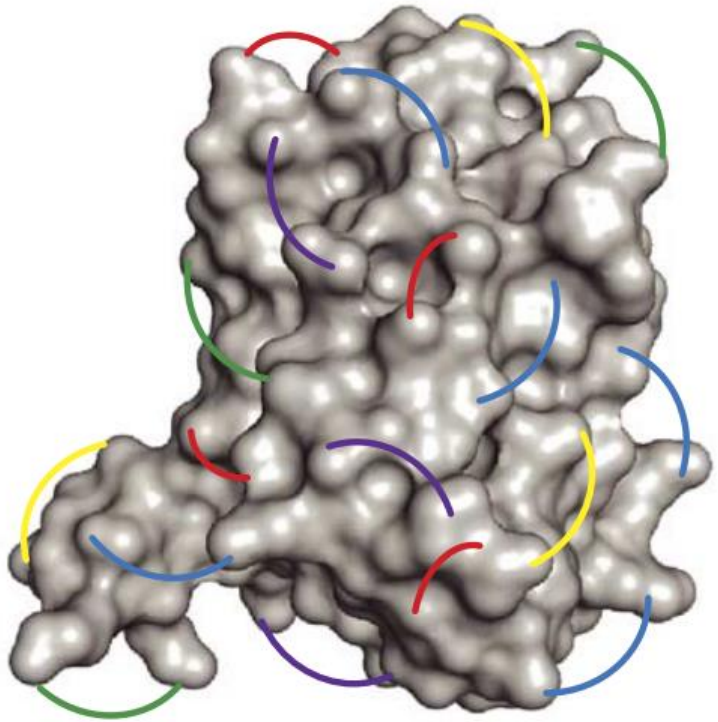
Wang, Jun, Wang, Jian, et al. IJMS. 20.17 (2019): 4116.

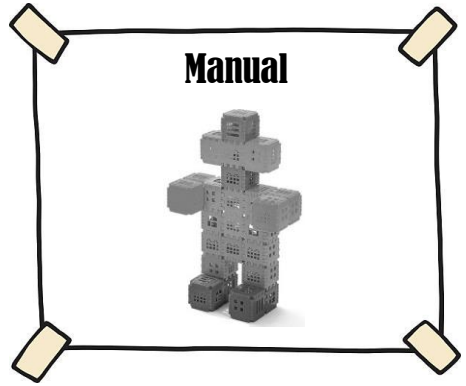
Fragment-assembly



Fragment assembly is like ***building blocks.***

Integrating Restraints



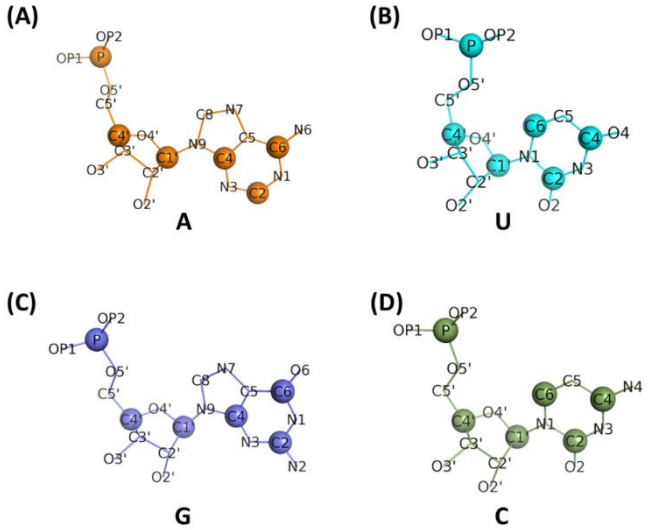


Restraints such as ***hydroxyl radical probing (HRP), cross-linking, and direct coupling analysis (DCA)*** *can be used as restraints to improve RNA modeling*

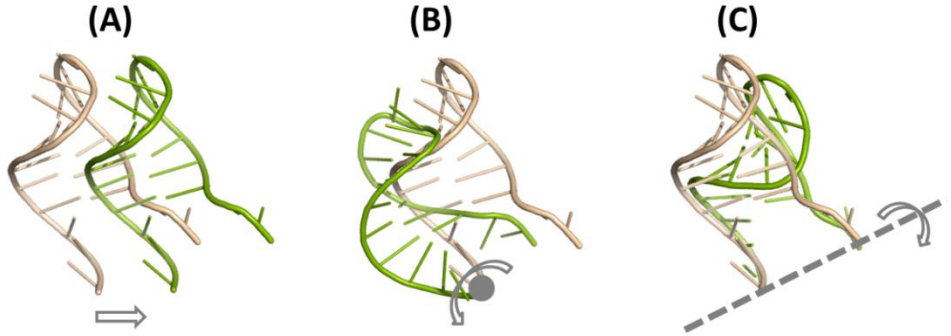
Wang, Jun, Wang, Jian, et al. IJMS. 20.17 (2019): 4116.

RNA 3D Structure Optimization

Coarse-grained model



Sampling



Energy Function

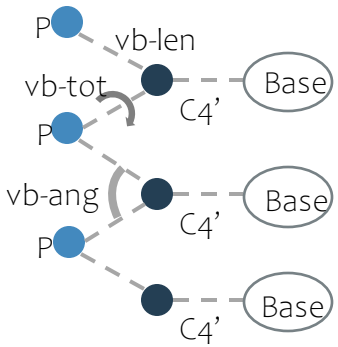
$$G = G_{vb-len} + G_{vb-ang} + G_{vb-tot} + G_{stacking} + G_{pairing} + G_{restr}$$

$$G_{vb-len} = k_l(l - l_0)^2$$

$$G_{vb-ang} = k_a(a - a_0)^2$$

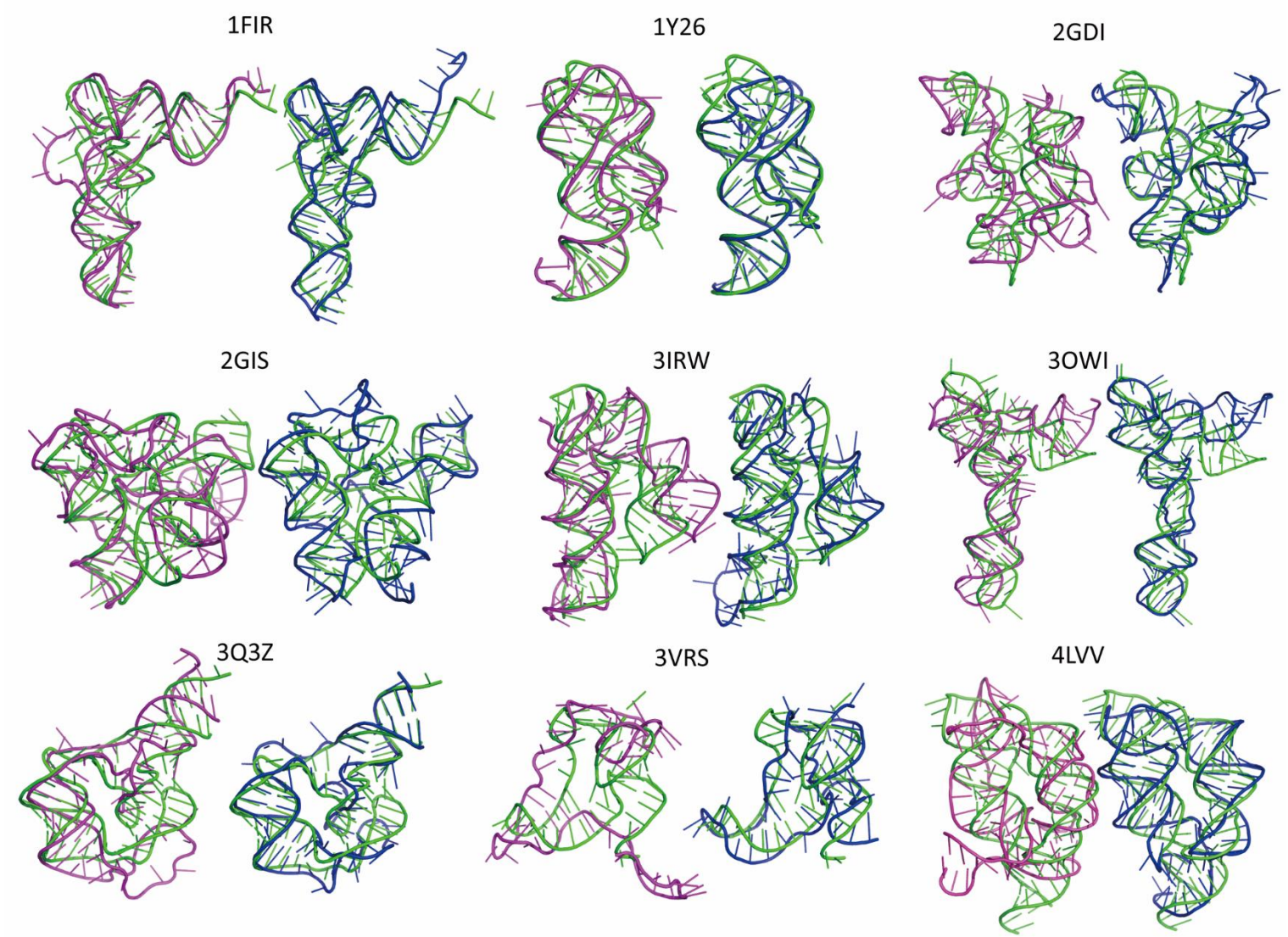
$$G_{vb-tor} = k_t \sin^2\left(\frac{t - t_0}{2}\right)$$

$$G_{restr} = k_r \sum_{\mathrm{n}} (r^n - r_0^n)^2$$



Wang, Jian, et al. Nucleic acids research 45.11 (2017): 6299-6309.

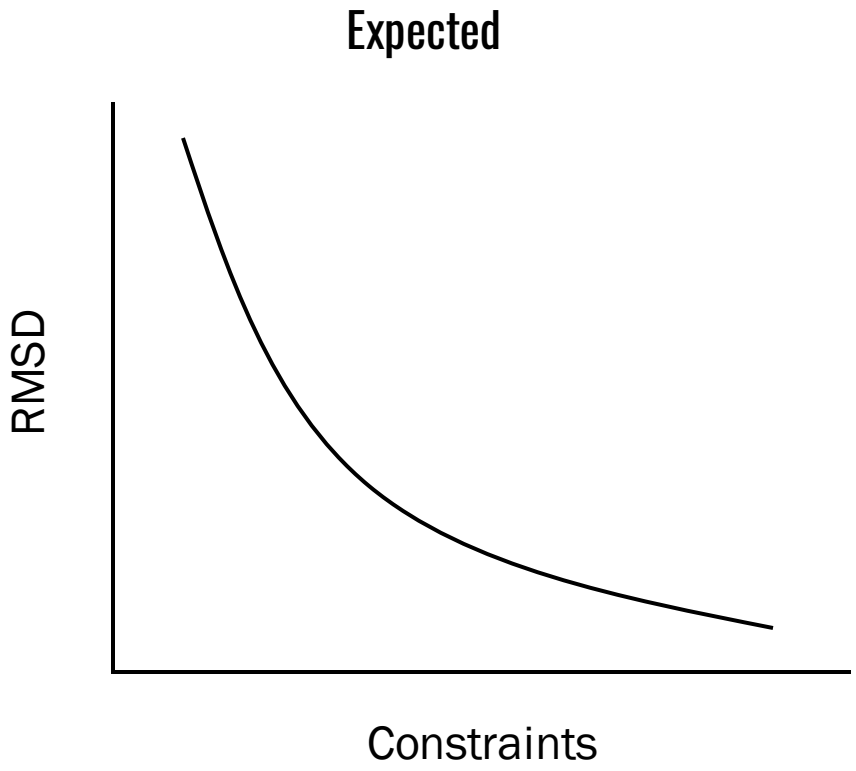
Test Results

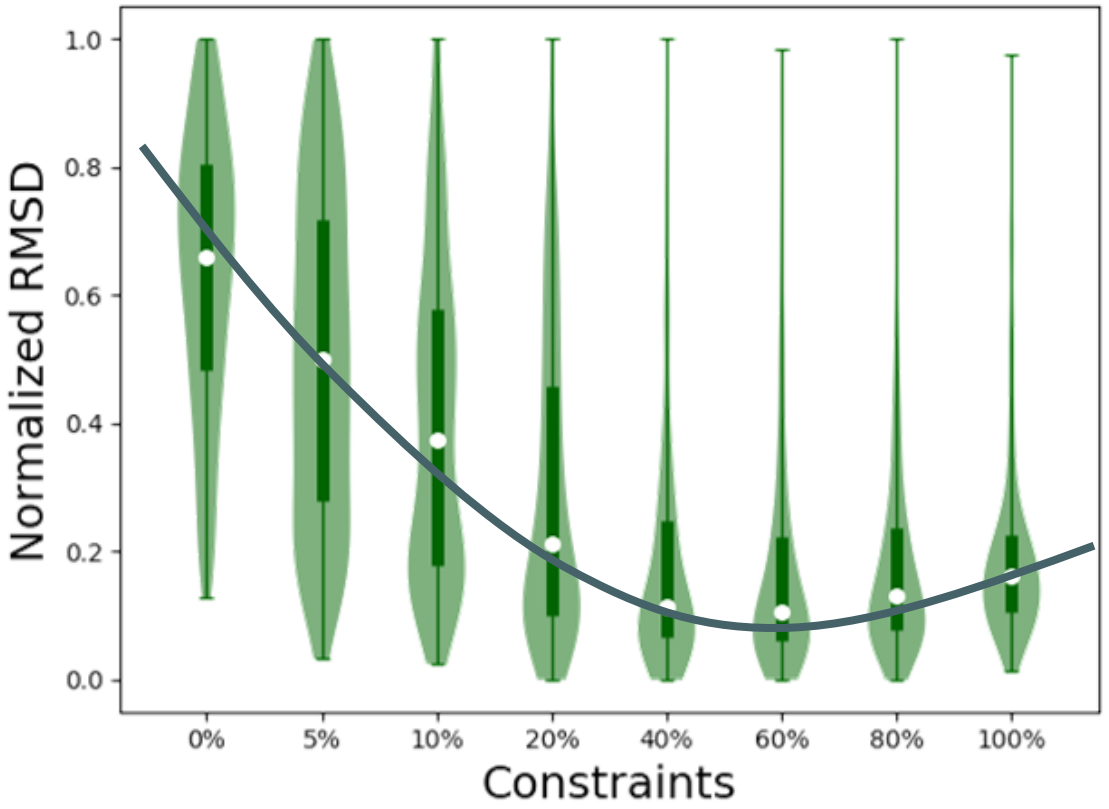


Native, Optimize w/o restraints, Optimize w/ restraints

Wang, Jian, et al. Nucleic acids research 45.11 (2017): 6299-6309.

How many restraints should we impose?





*When using more than **60**% of the contacts as constraints, the RMSD of the predicted model gets higher.*

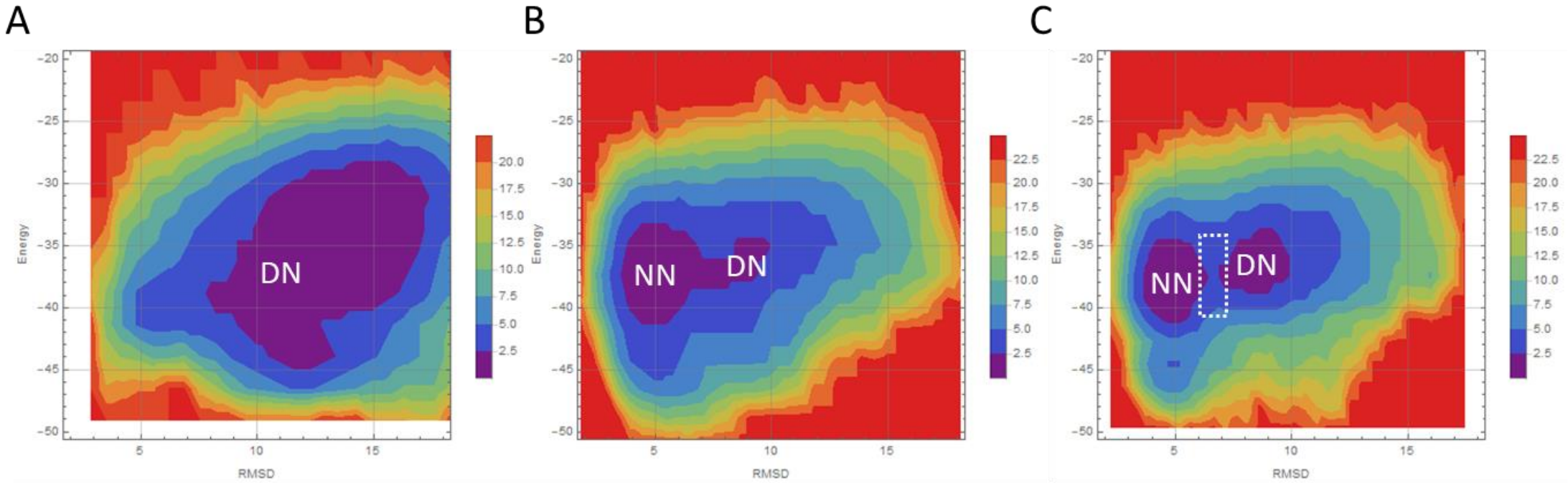
Wang, J, et al. Nucleic acids research. 47.11 (2019): 5563-5572.

Difficulties in RNA 3D structure prediction

No restraints

40% restraints

80% restraints

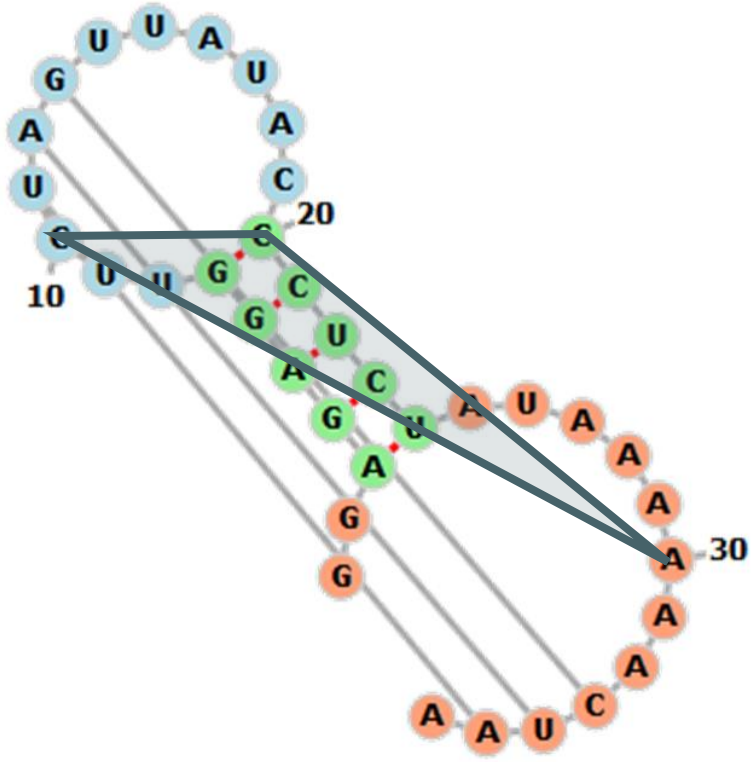


By using more restraints, there may be more energy barriers in the free energy landscape.

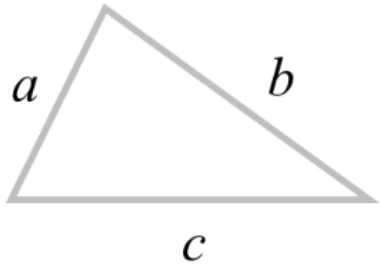
NN: local minimum state near the native state
DN: local minimum state distant from the native state

Wang, J, et al. Nucleic acids research. 47.11 (2019): 5563-5572.

Restraints Derivation



Triangle Inequality



$$c_{max} = a_{max} + b_{max}$$

$$c_{min} = \begin{cases} b_{min} - a_{max}, & b_{min} > a_{max} \\ a_{min} - b_{max}, & a_{min} > b_{max} \end{cases}$$

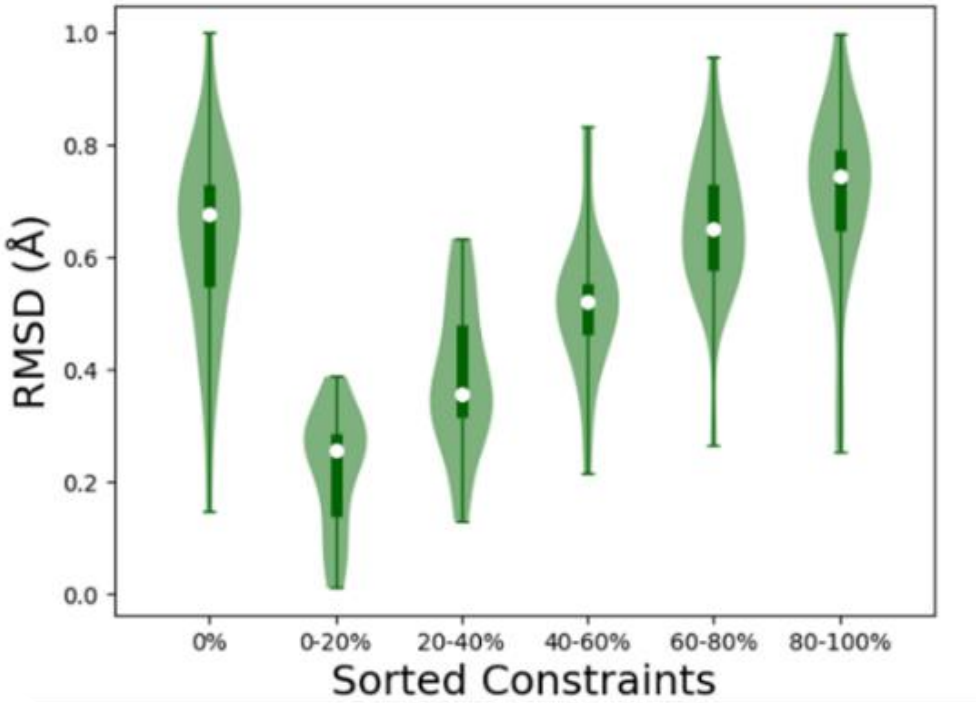
$$DV_c = c_{max} - c_{min}$$

Distance Variation (DV): the difference between the minimum distance and the maximum distance between two residues

*The larger the distance variation, the higher the **importance** of the corresponding restraint.*

Wang, J, et al. Nucleic acids research. 47.11 (2019): 5563-5572.

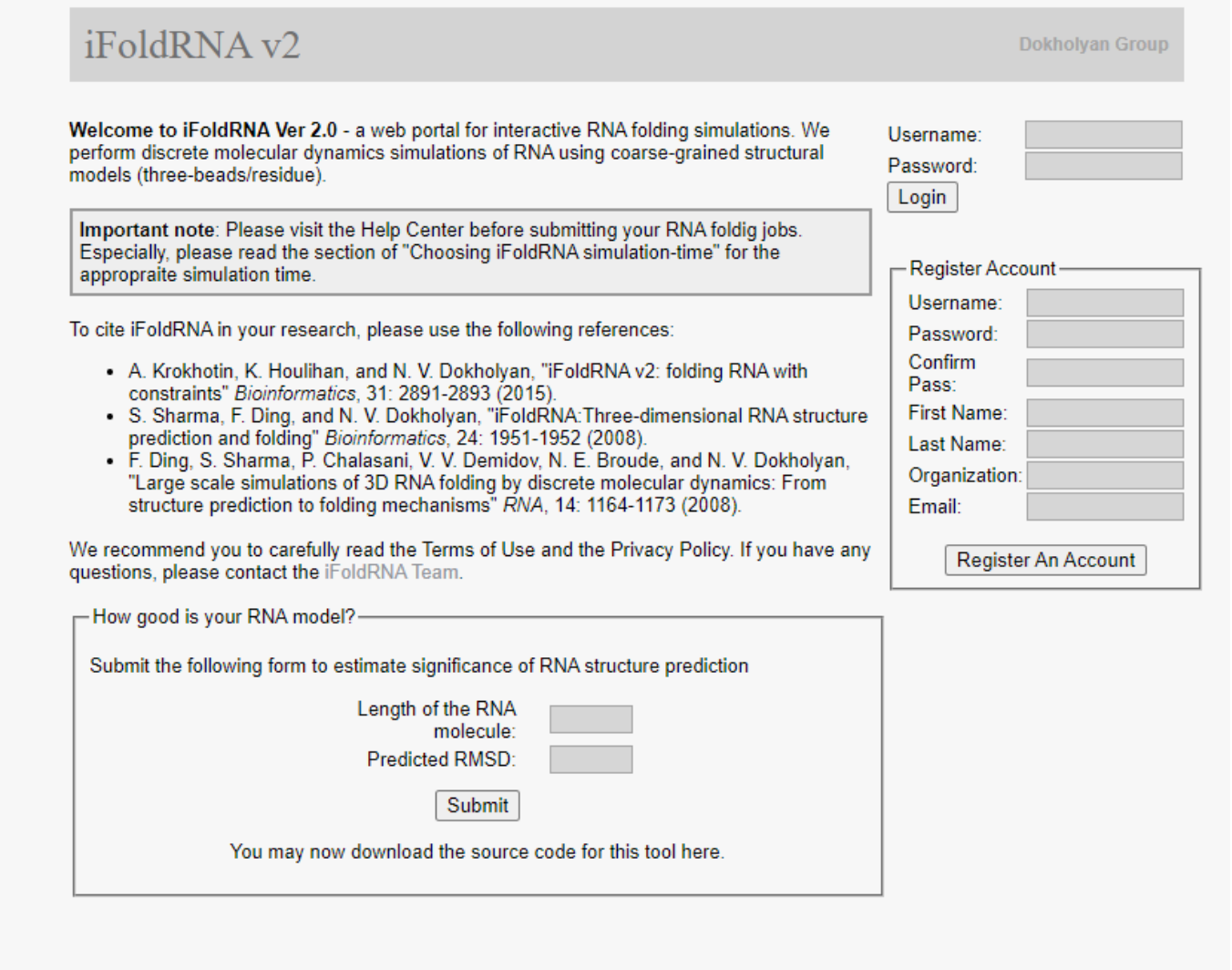
Difficulties in RNA 3D structure prediction



By sorting the constraints by the importance, we can use only the constraints that let us achieve the highest performance.

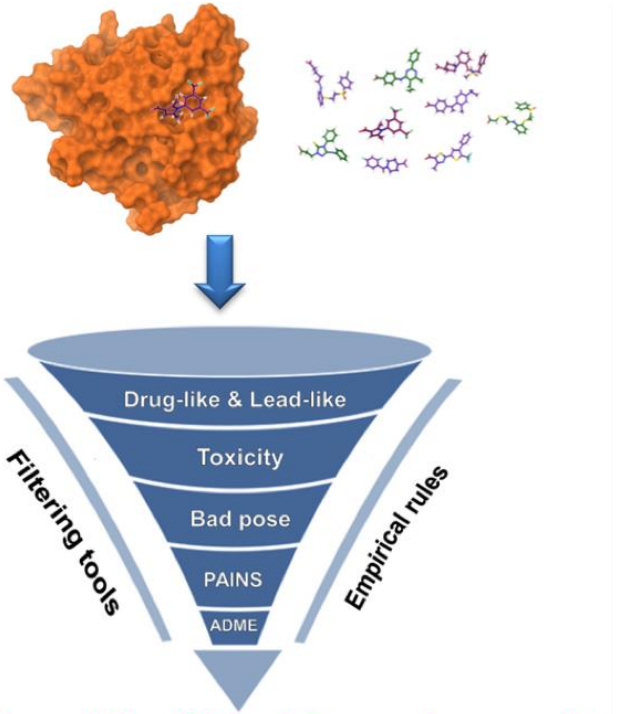
Wang, J, et al. Nucleic acids research. 47.11 (2019): 5563-5572.

iFoldRNA

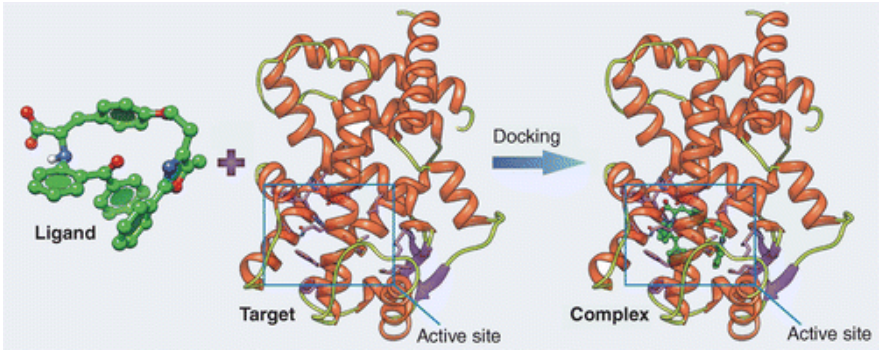

iFoldRNA v2
Dokholyan Group
Welcome to iFoldRNA Ver 2.0 - a web portal for interactive RNA folding simulations. We perform discrete molecular dynamics simulations of RNA using coarse-grained structural models (three-beads/residue).
Important note: Please visit the Help Center before submitting your RNA foldig jobs. Especially, please read the section of "Choosing iFoldRNA simulation-time" for the appropraite simulation time.
To cite iFoldRNA in your research, please use the following references:
A. Krokhotin, K. Houlihan, and N. V. Dokholyan, "iFoldRNA v2: folding RNA with constraints" Bioinformatics, 31: 2891-2893 (2015).
S. Sharma, F. Ding, and N. V. Dokholyan, "iFoldRNA:Three-dimensional RNA structure prediction and folding" Bioinformatics, 24: 1951-1952 (2008).
F. Ding, S. Sharma, P. Chalasani, V. V. Demidov, N. E. Broude, and N. V. Dokholyan, "Large scale simulations of 3D RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms" RNA, 14: 1164-1173 (2008).
We recommend you to carefully read the Terms of Use and the Privacy Policy. If you have any questions, please contact the iFoldRNA Team.
How good is your RNA model?
Submit the following form to estimate significance of RNA structure prediction
Length of the RNA molecule:
Predicted RMSD:
Submit
You may now download the source code for this tool here.
Username:
Password:
Login
Register Account
Username:
Password:
Confirm Pass:
First Name:
Last Name:
Organization:
Email:
Register An Account

Contents

Virtual Screening

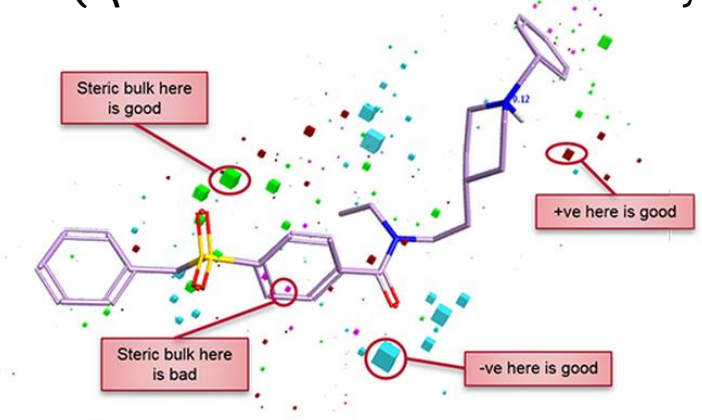


Molecular Docking



Time-consuming

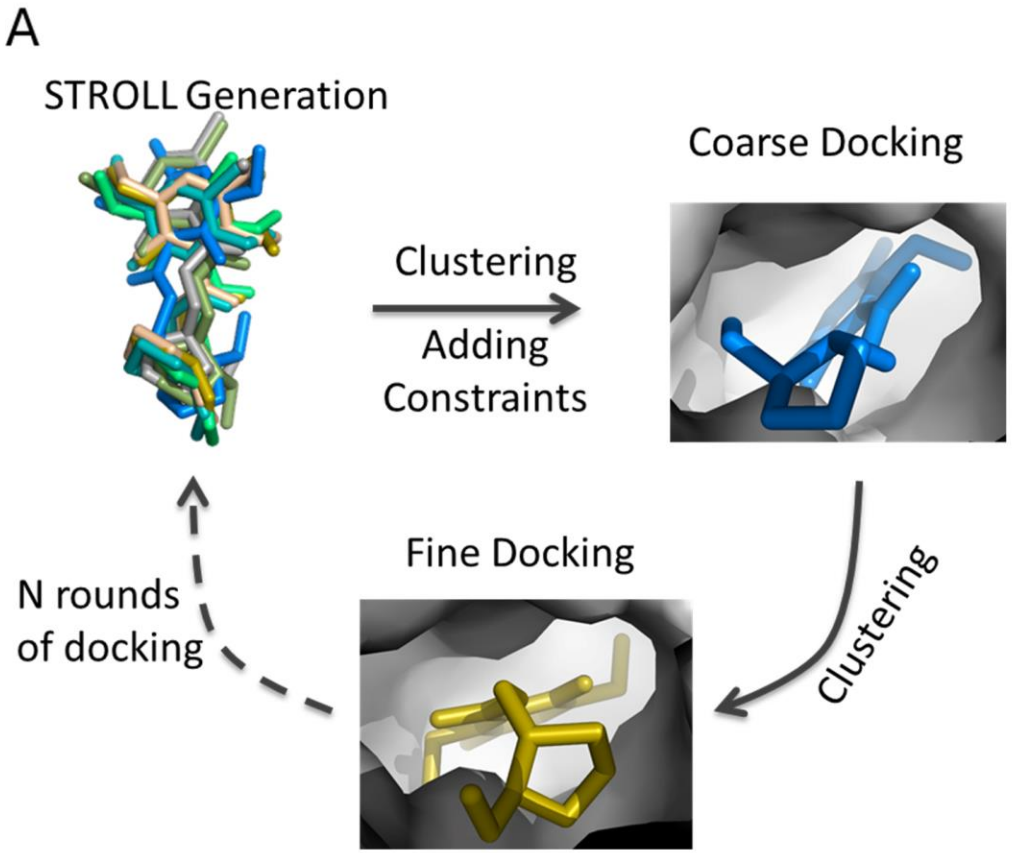
QSAR (Quantitative structure-activity relationship)

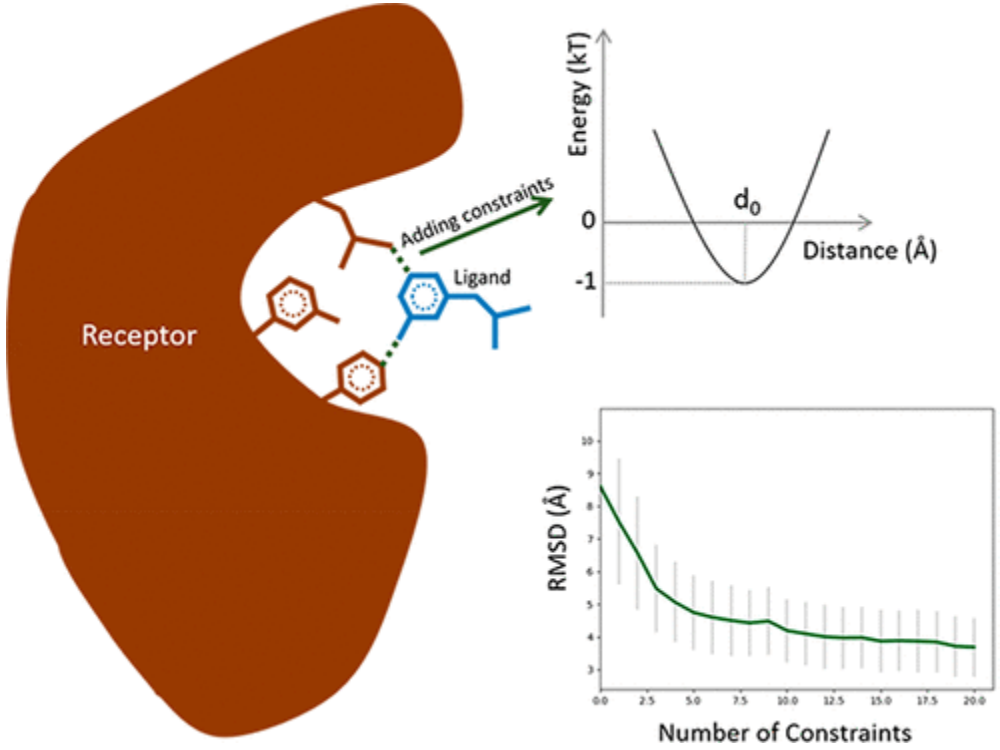


Risk of highly inaccurate predictions of pharmacological or biological activity

Wang, Jian, and Nikolay V. Dokholyan. JCIM 59.6 (2019): 2509-2515.
https://summerofhpc.prace-ri.eu/re-ranking-virtual-screening-results-in-computer-aided-drug-design
https://www.cresset-group.com/software/forge-qsar-models/

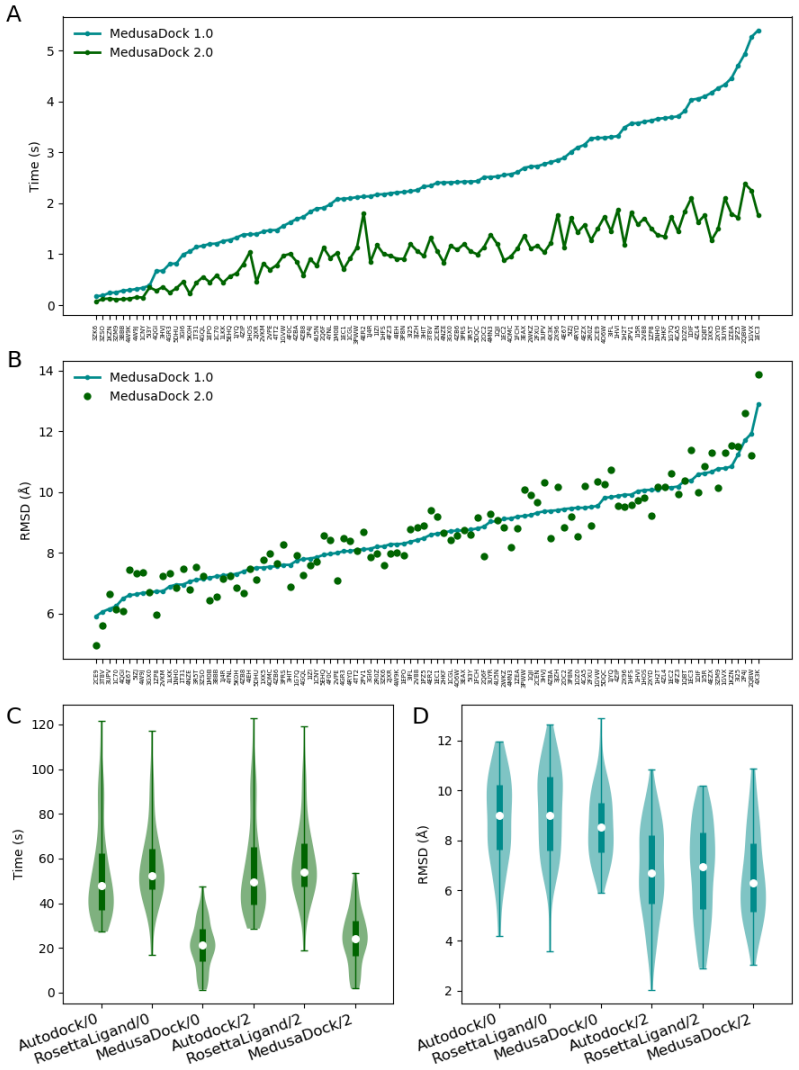
MedusaDock

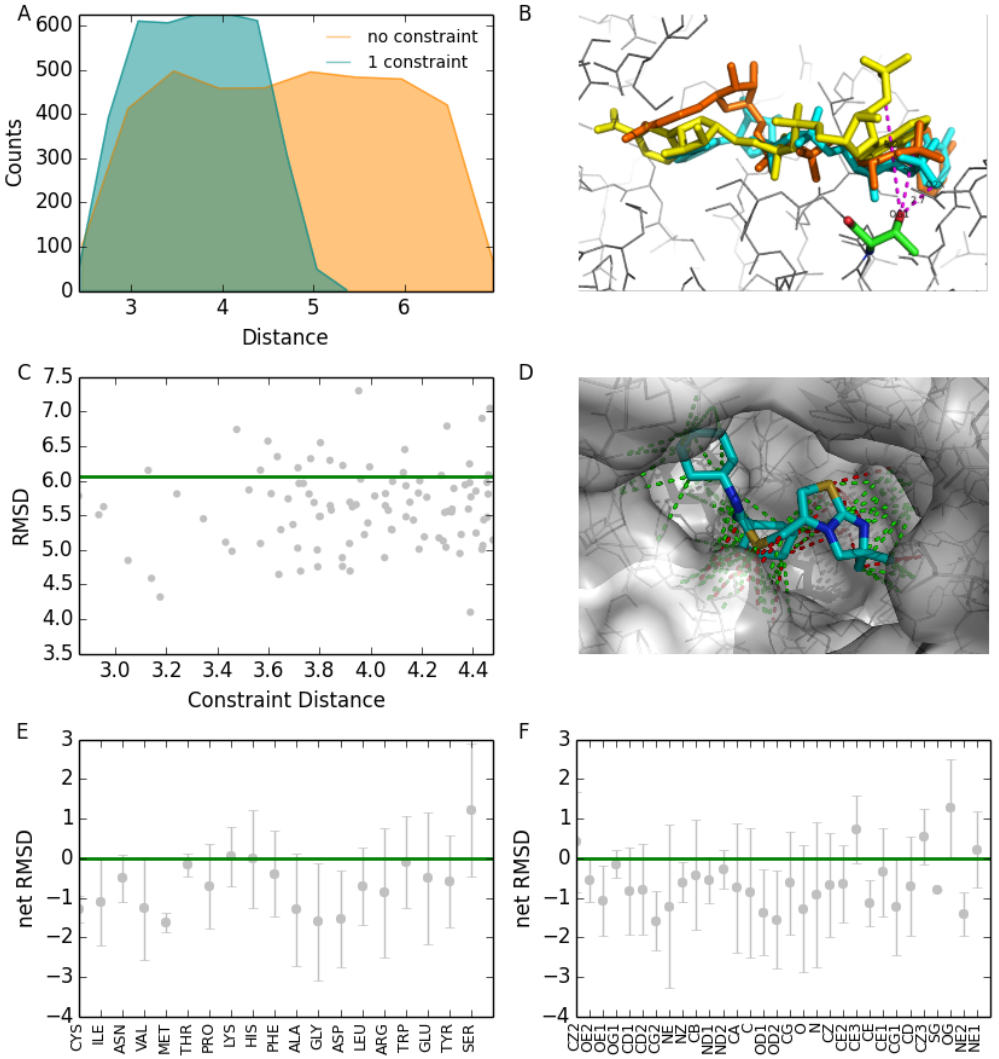




Wang, Jian, et al. JCIM. 59.6 (2019): 2509-2515.

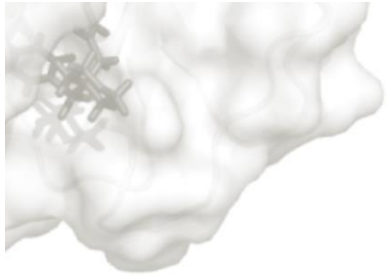
MedusaDock





Wang, Jian, et al. JCIM. 59.6 (2019): 2509-2515.

MedusaDock website

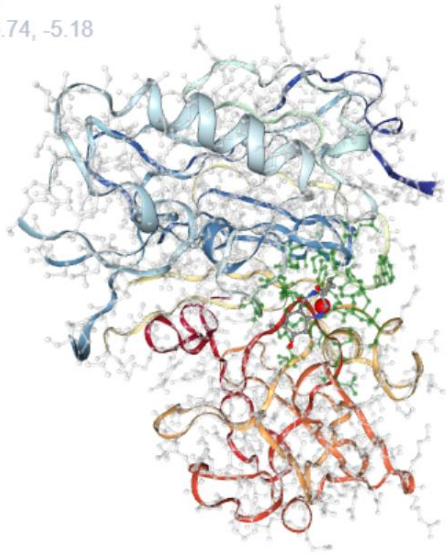


MedusaDock

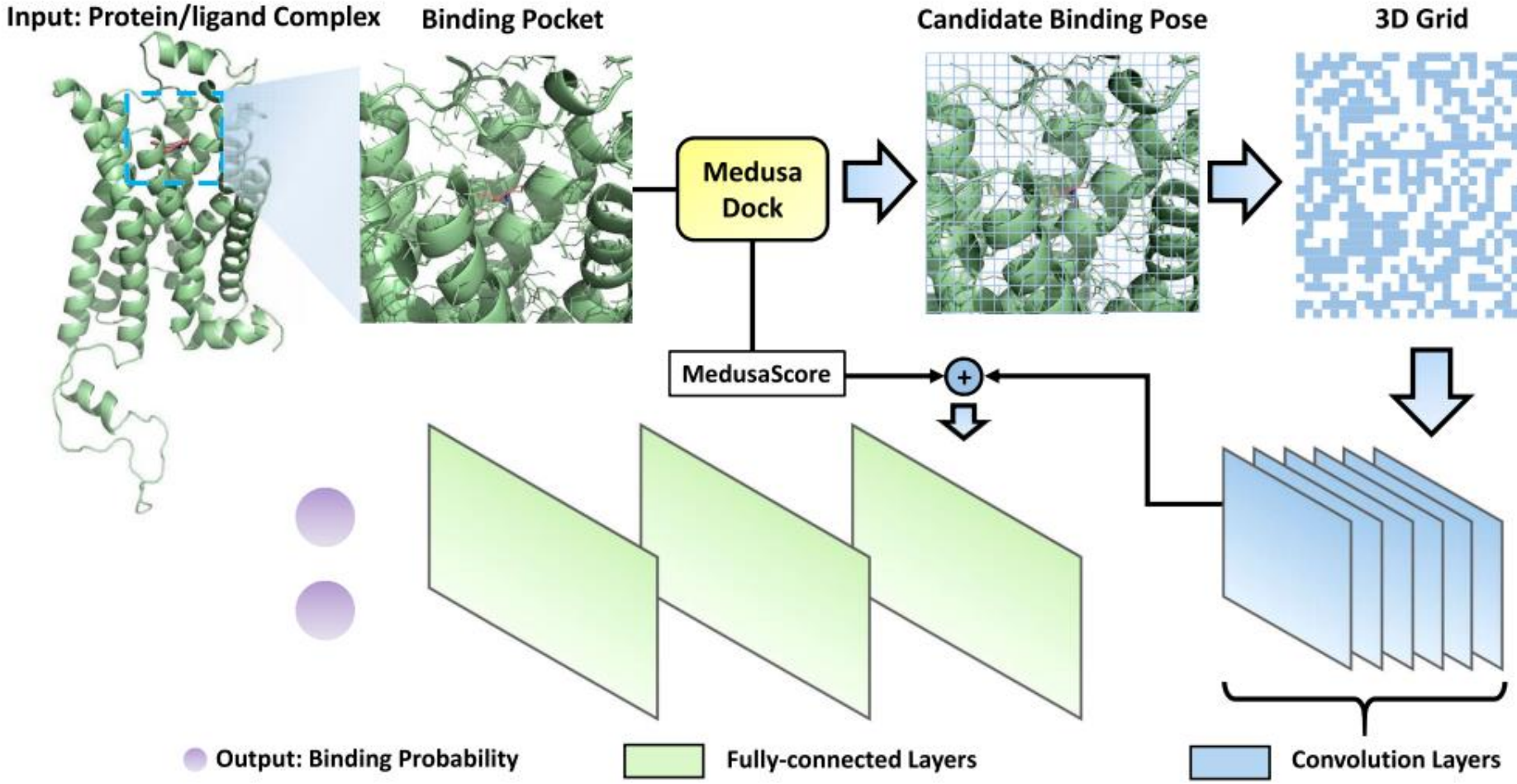
Select Receptor — Select Ligand — 3 Set Binding Site — 4 Set Constraints — 5 Submit

Please specify the position or upload the structure of the binding site.

Receptor: dmyVrc.pdb
Ligand: Px2KI0.mol2
Binding Site: 8.96, 36.74, -5.18



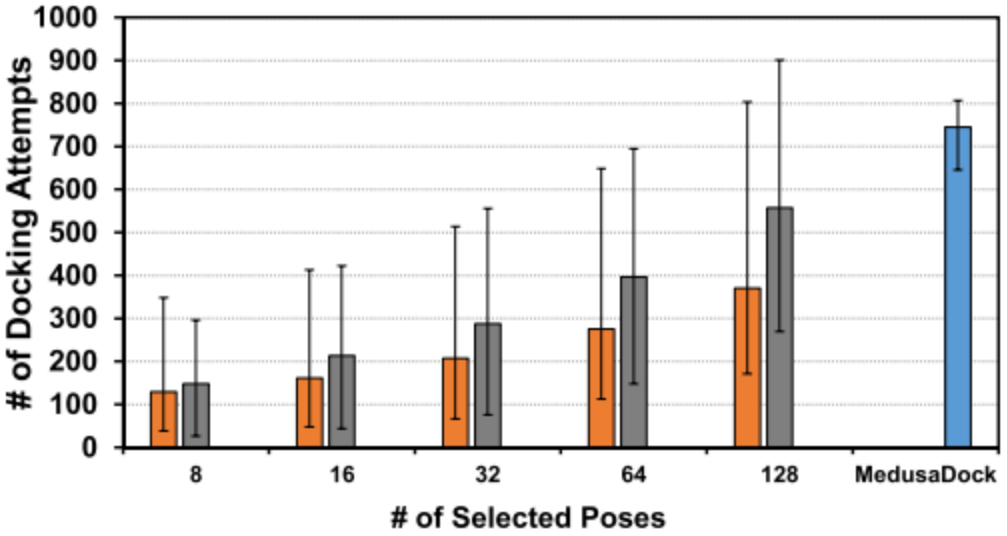
MedusaNet: Guiding Conventional Protein–Ligand Docking Software



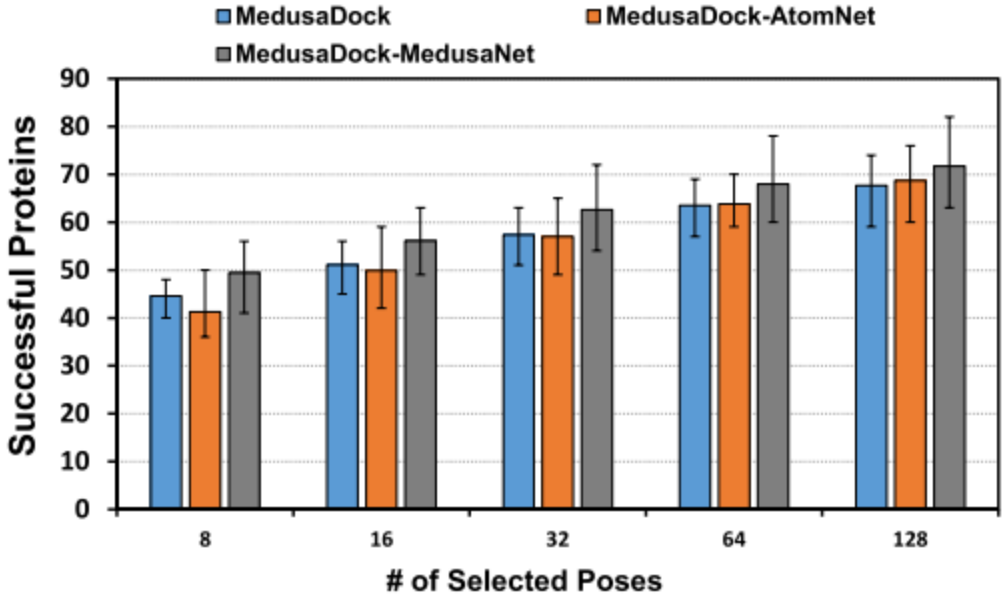
Huaipan Jiang

Dr. Mahmut Kandemir

MedusaNet: Guiding Conventional Protein–Ligand Docking Software



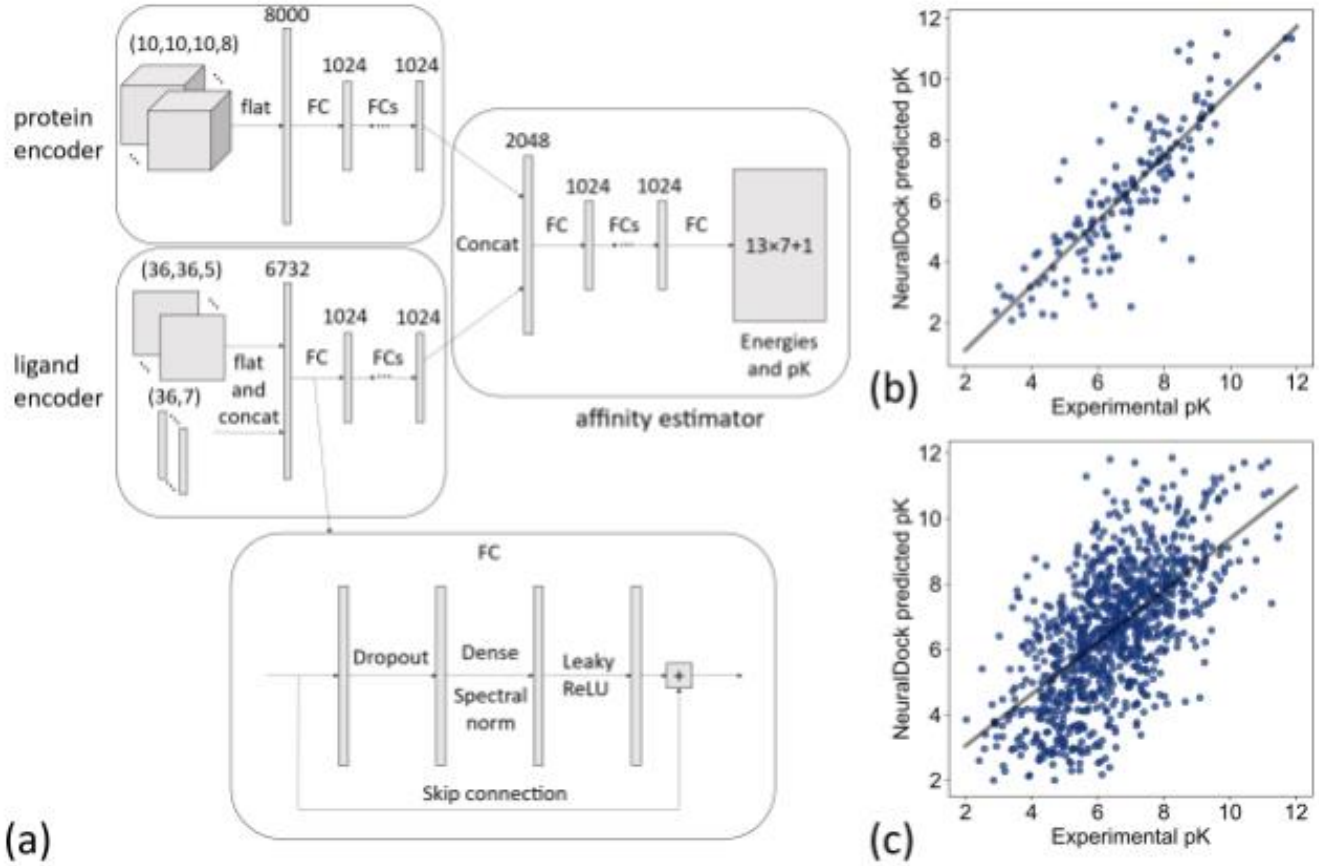
(a) MeudsaDock run less attempts with CNN guiding.



(b) Number of proteins found good pose by each approach.

MedusaNet improves both the efficiency and accuracy of MedusaDock

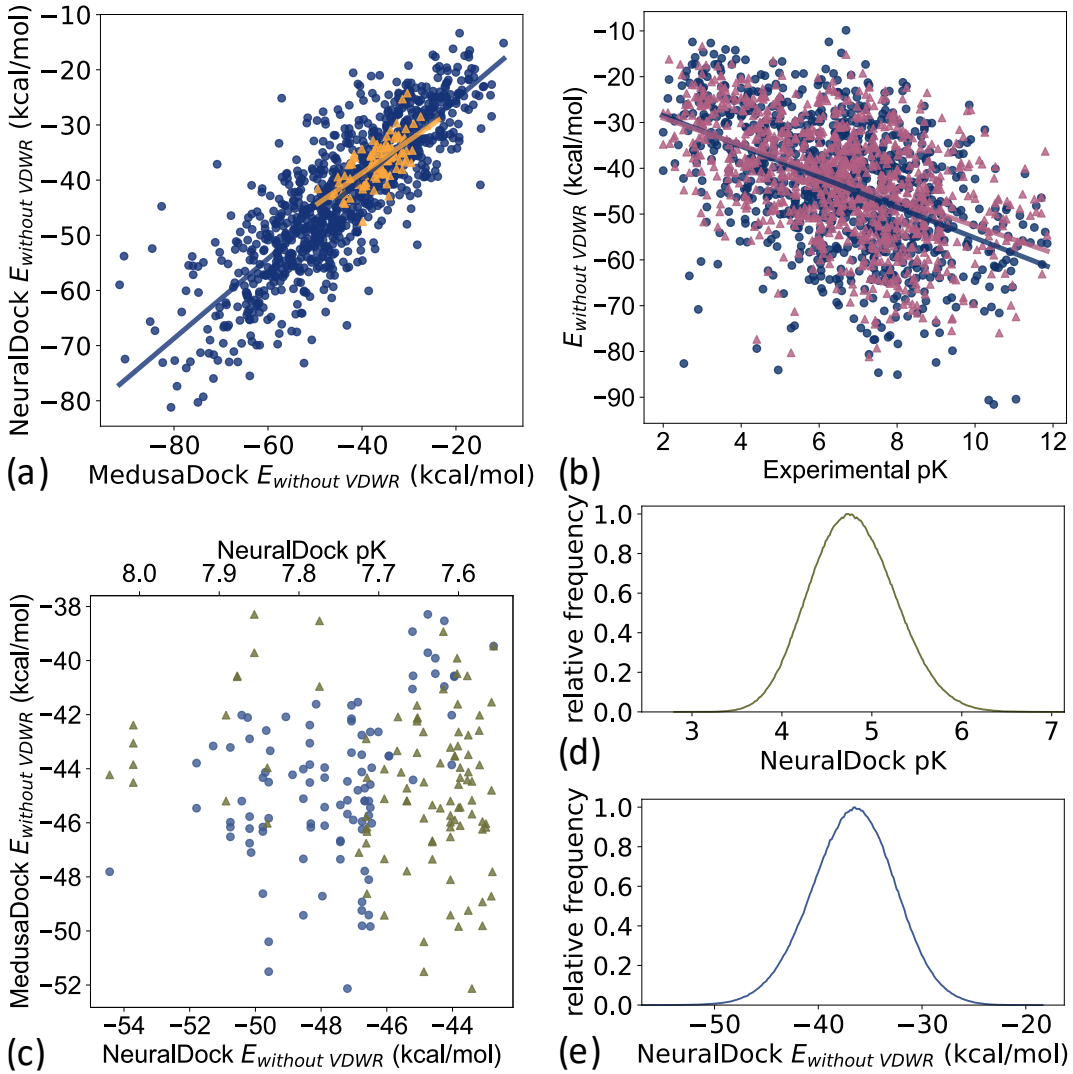
NeuralDock: Rapid and conformation-agnostic docking of small molecules



Congzhou Sha

Dr. Nikolay Dokholyan

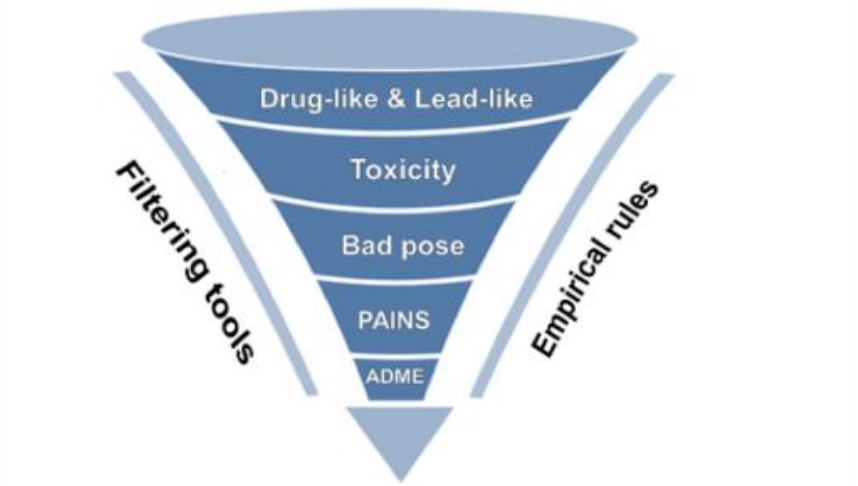
Comparison to traditional docking software



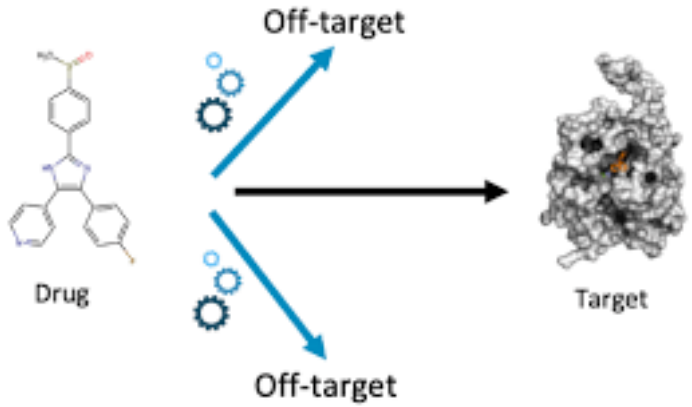
96000 protein-small molecule pairs
Tesla T4 GPU, training in a week
937 million ZINC compounds took 21 hours on 25 GPUs

Target identification

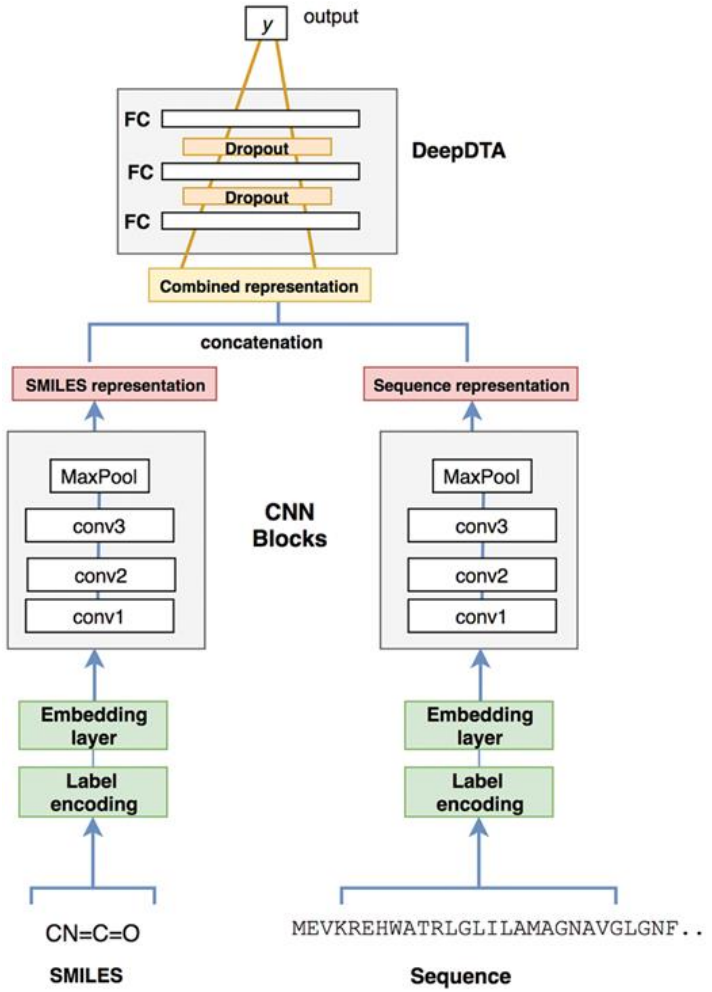
Virtual Screening



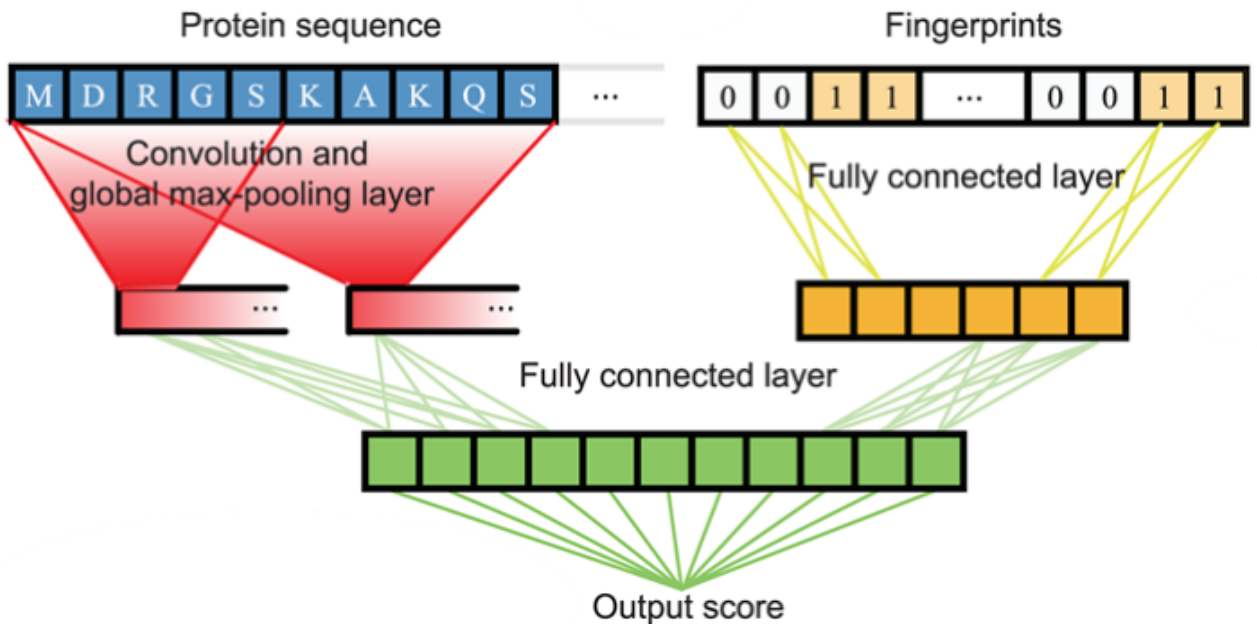
Target Identification



Current compound-protein interaction prediction models

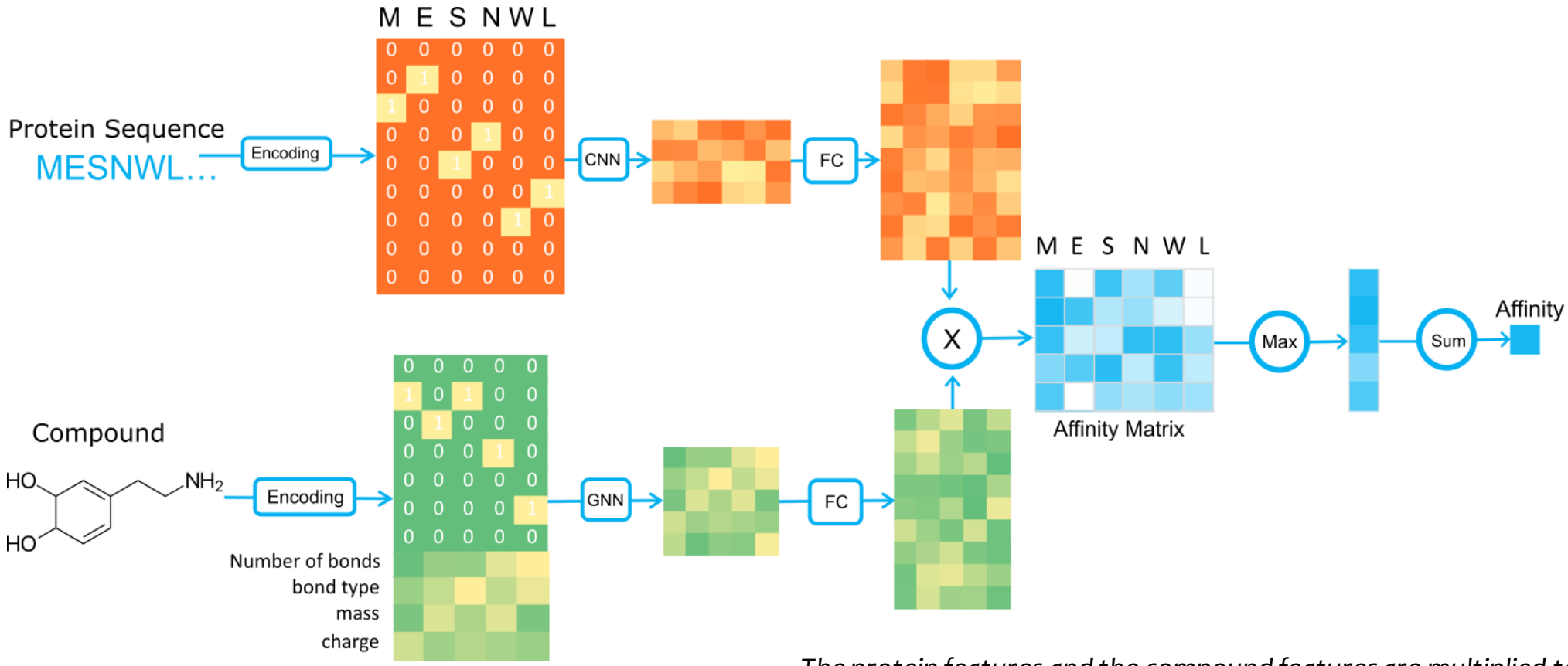


DeepDTA



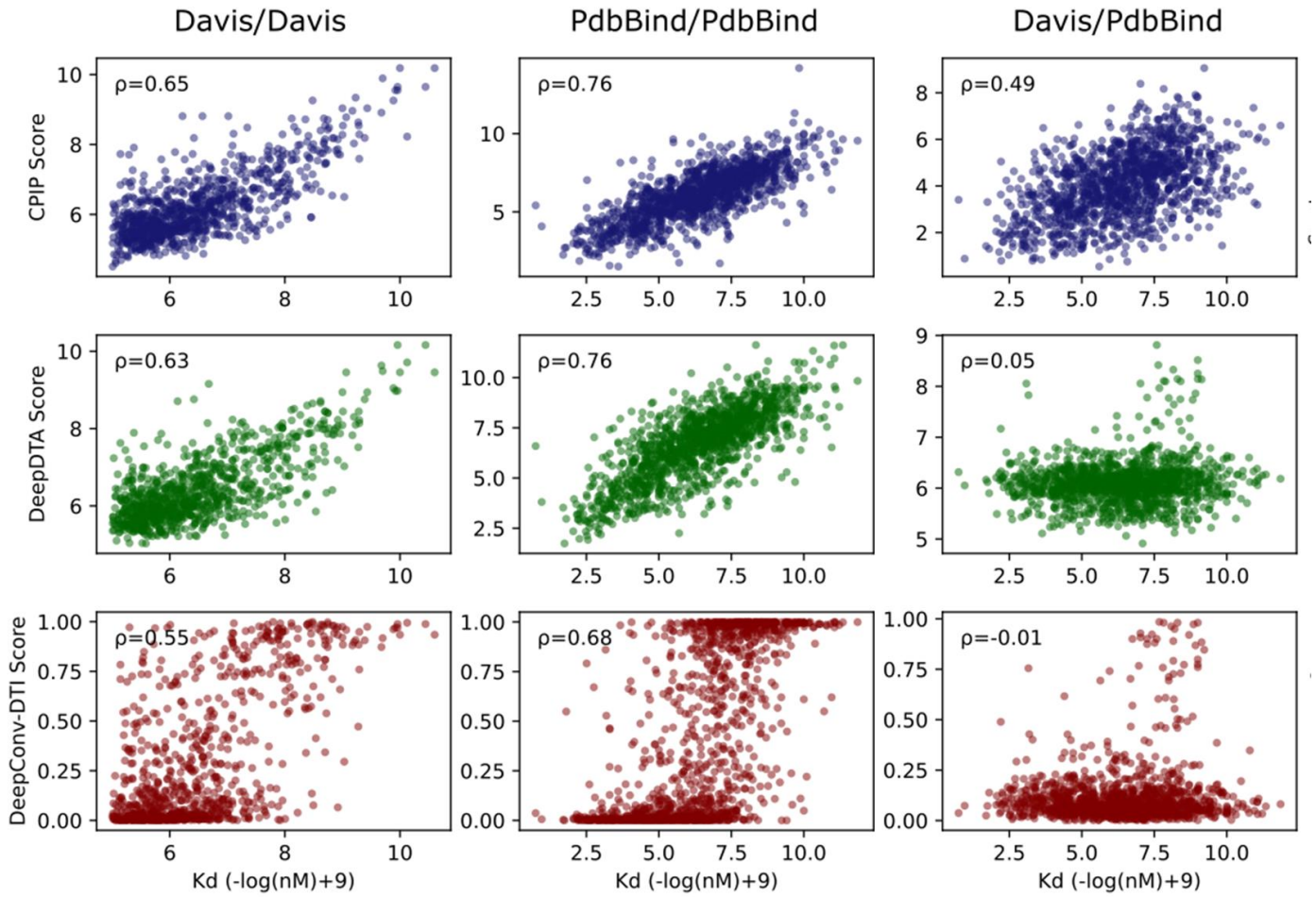
DeepConv-DTI

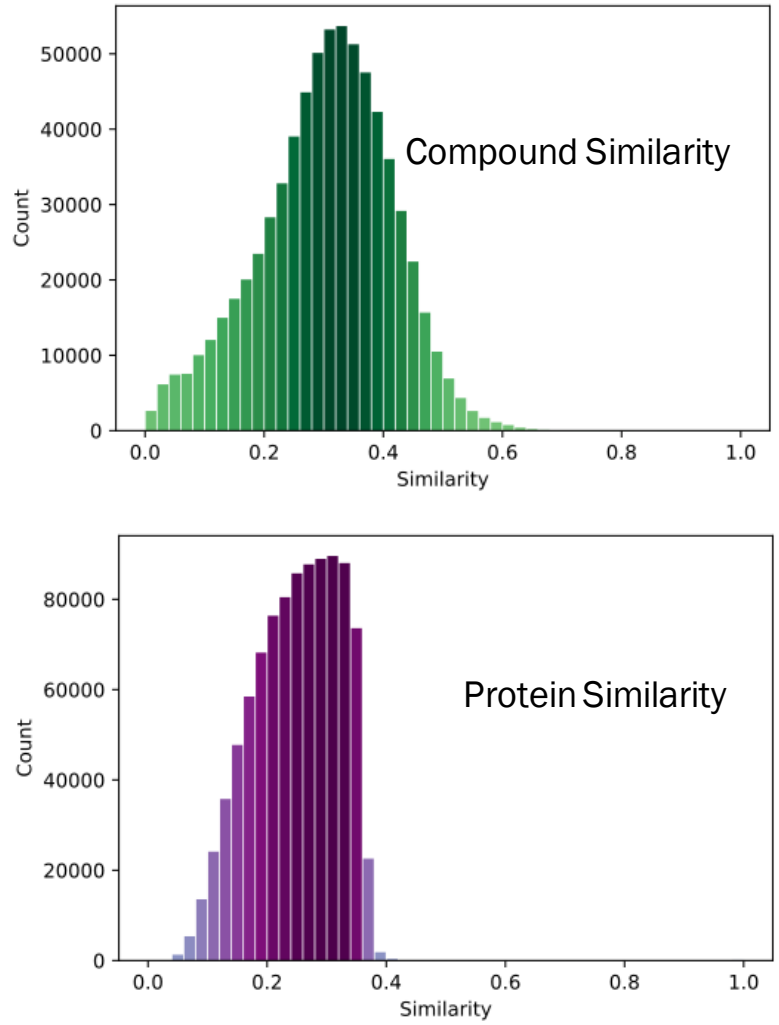
Compound-Protein Interaction Predictor (Yuel)



The protein features and the compound features are multiplied to evaluate the pairwise residue-atom interaction.

Evaluation of the predictability of Yuel





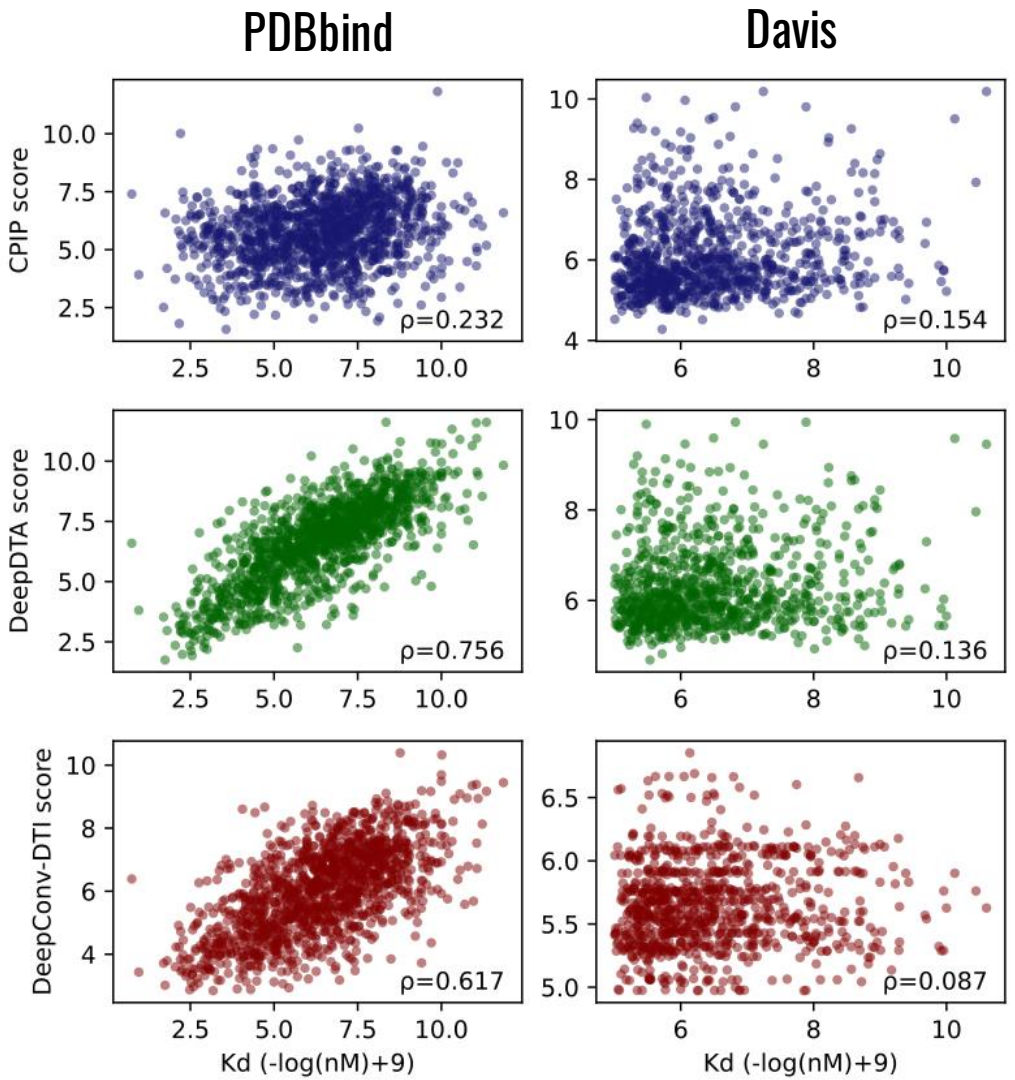
Davis/Davis: Models are trained on Davis and tested on Davis
PDBbind/PDBbind: Models are trained on PDBbind and tested on PDBbind
Davis/PDBbind: Models are trained on Davis and tested on PDBbind

*Davis and PDBbind are **dissimilar***

Testing CPIP in the datasets with the protein sequence shuffled

Shuffle the protein sequence in the test sets

*When **shuffling** the protein sequences, DeepDTA and Deep-Conv-DTI still predict high affinity between the shuffled protein and the compounds.*



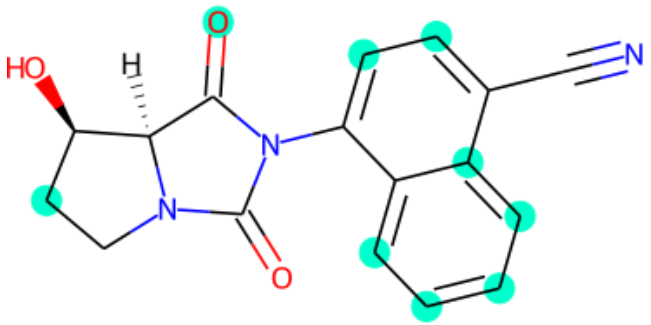
Yuel

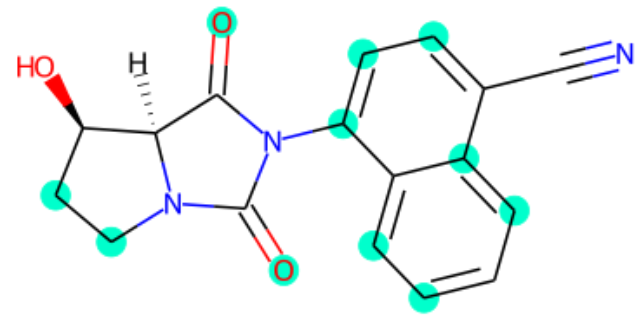
DeepDTA

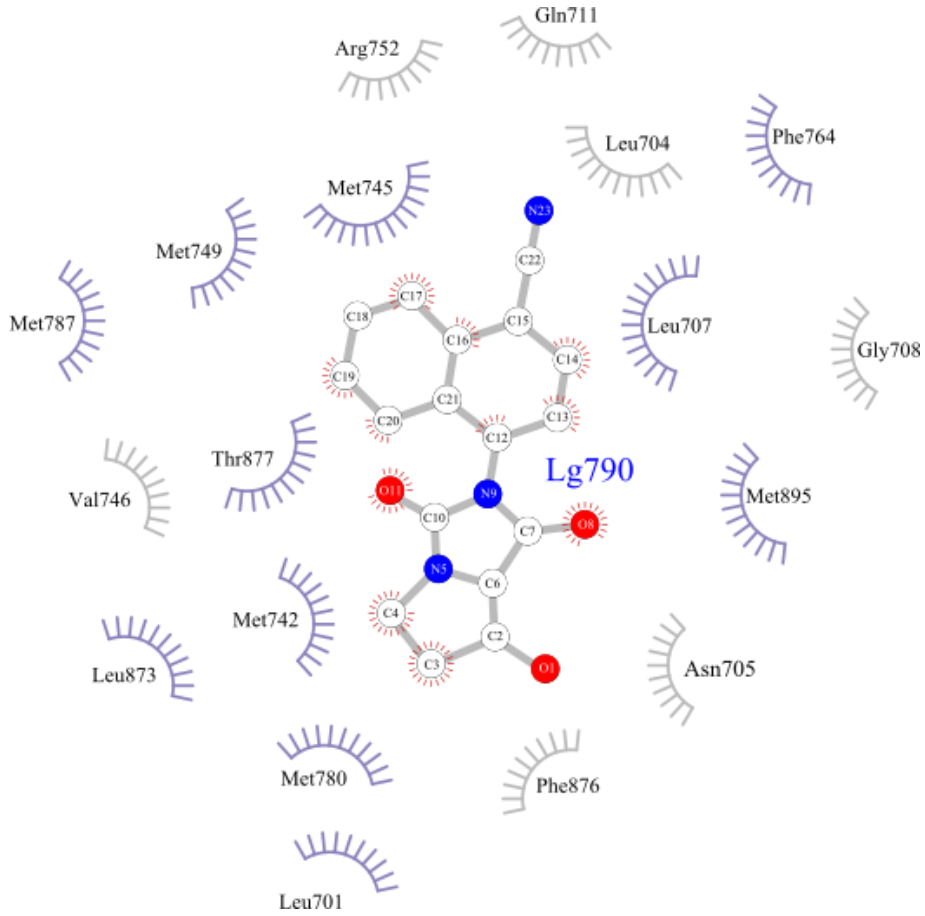
DeepConv-DTI

Yuel can predict hotspot atoms and residues

PDBID: 2IHQ





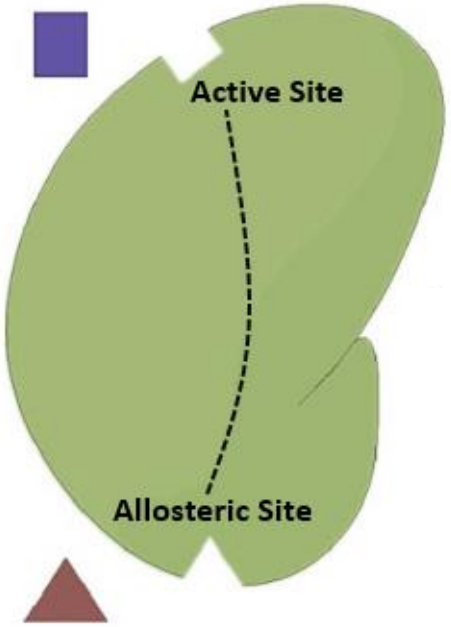


*Yuel can predict compound atoms that interact with the protein (**hotspot atoms**) as well as protein residues that interact with the compound **(hotspot residues).***

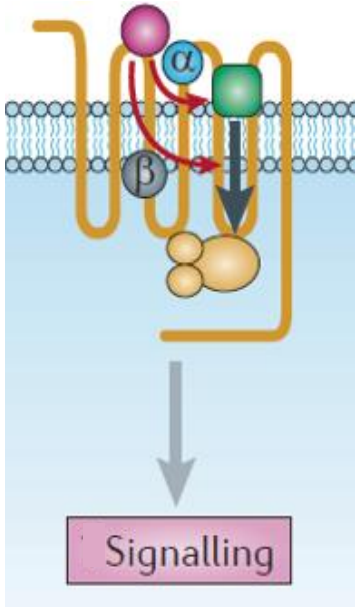
Contents

Allostery in various heterogeneous materials

Allostery

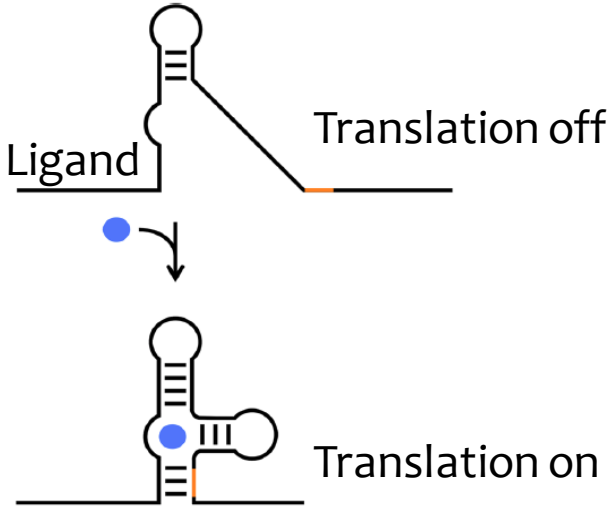


Allostery in GPCR



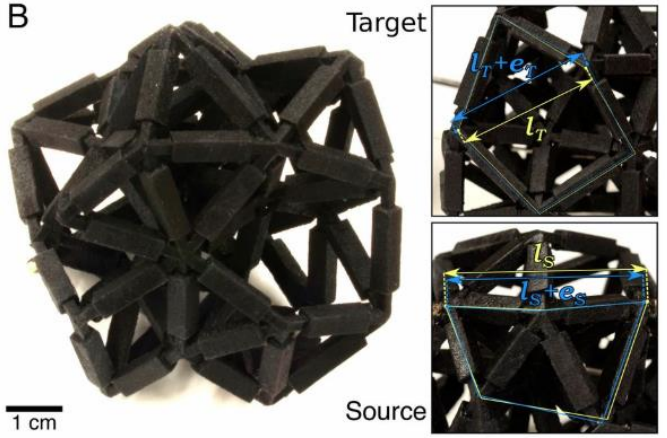
Wootten, D., et al. *Nat. Rev. Drug Discov* 2013; **12**: 630.

Allostery in RNA



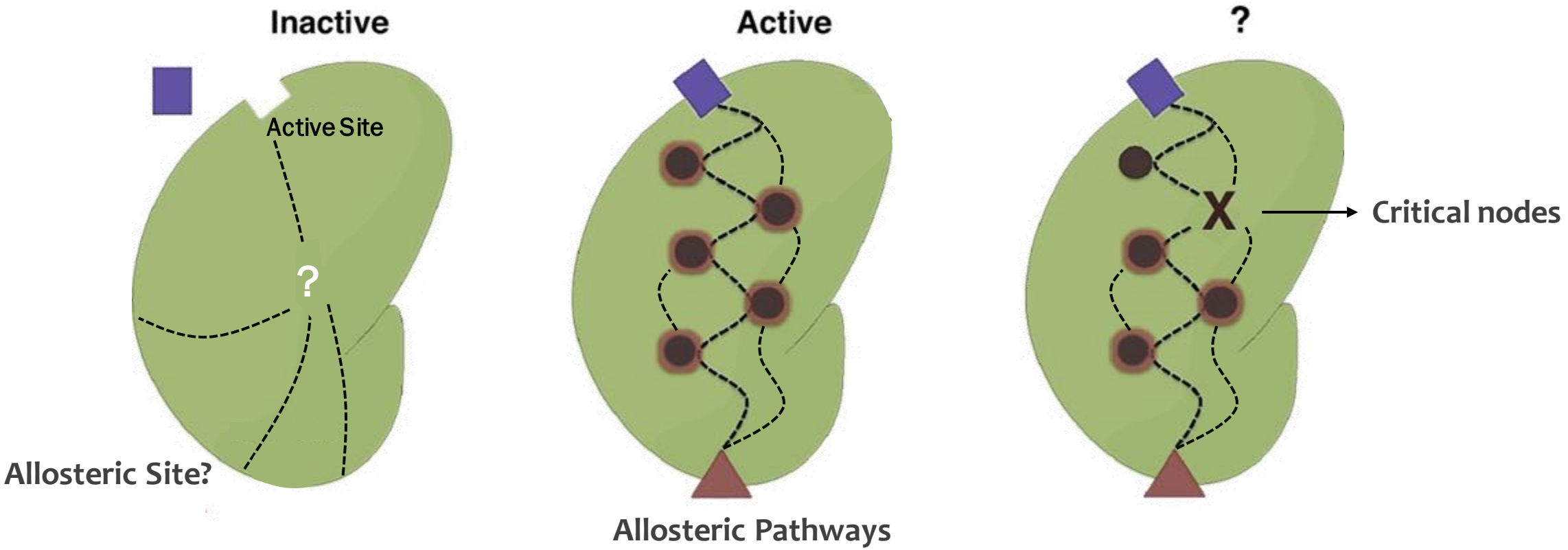
Ren, A., et al. *Nature* 2012; **486**: 85.

Allostery in mechanical networks



Rocks, J. W., et al. *PNAS* 2017; **114**: 2520-2525.

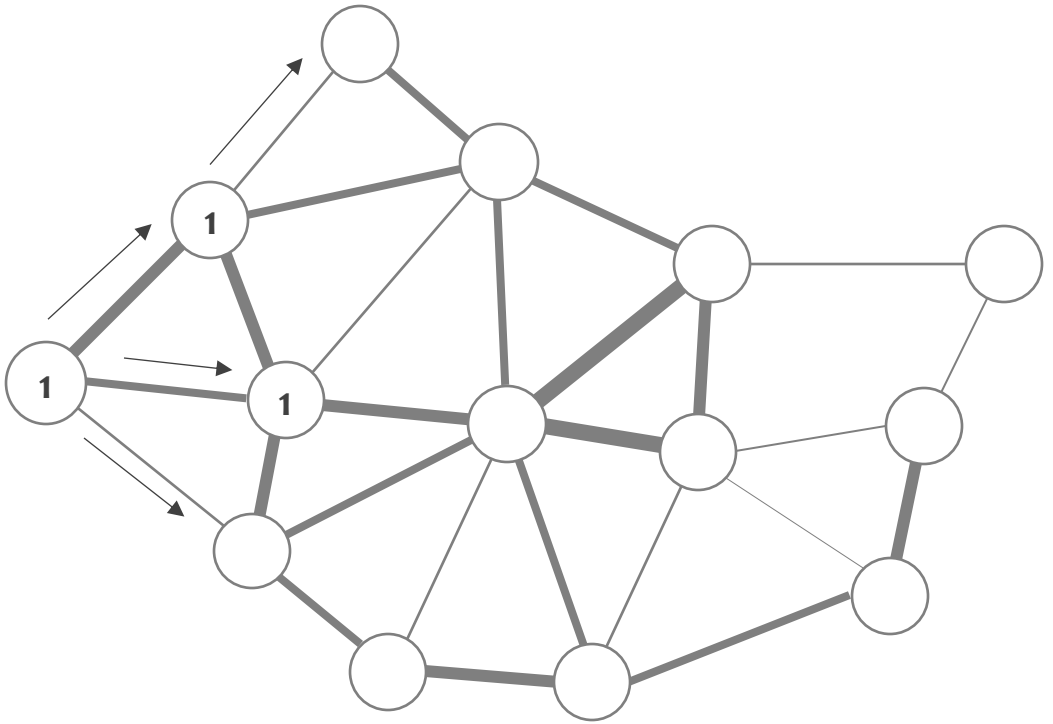
Allosteric Pathways and Critical Nodes



Residue Interaction Network

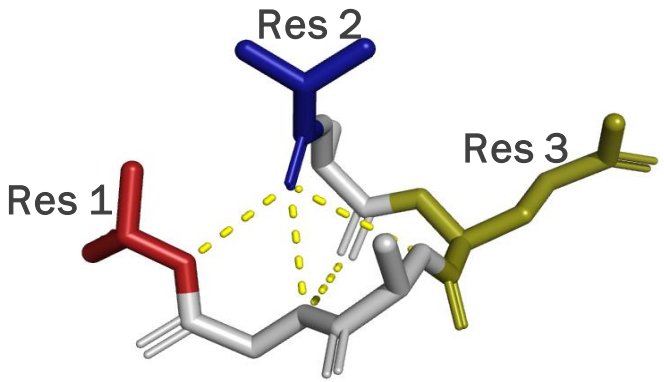
Social Network

Residue Interaction Network



Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Perturbation Propagation Algorithm



Contact Matrix (C)

$$C_{ij} = \sum_{a,b} H\left(r_0 - \left|\overrightarrow{r_a^i} - \overrightarrow{r_b^j}\right|\right)$$

$$N_{ij} = \frac{C_{ij}}{C_i}, N_{ij} = \frac{C_{ij}}{C_i}$$

Matrix of number of contacts per atom (N)

1.67	0.83	0.17	1.03	1.67
0.83	1.57	0.67	1.45	0.33
0.17	0.67	1.70	1.64	0.36
1.03	1.45	1.64	2.00	1.25
1.67	0.33	0.36	1.25	1.97

$$P_{ij} = 1 - e^{-\alpha \cdot N_{ij}}$$

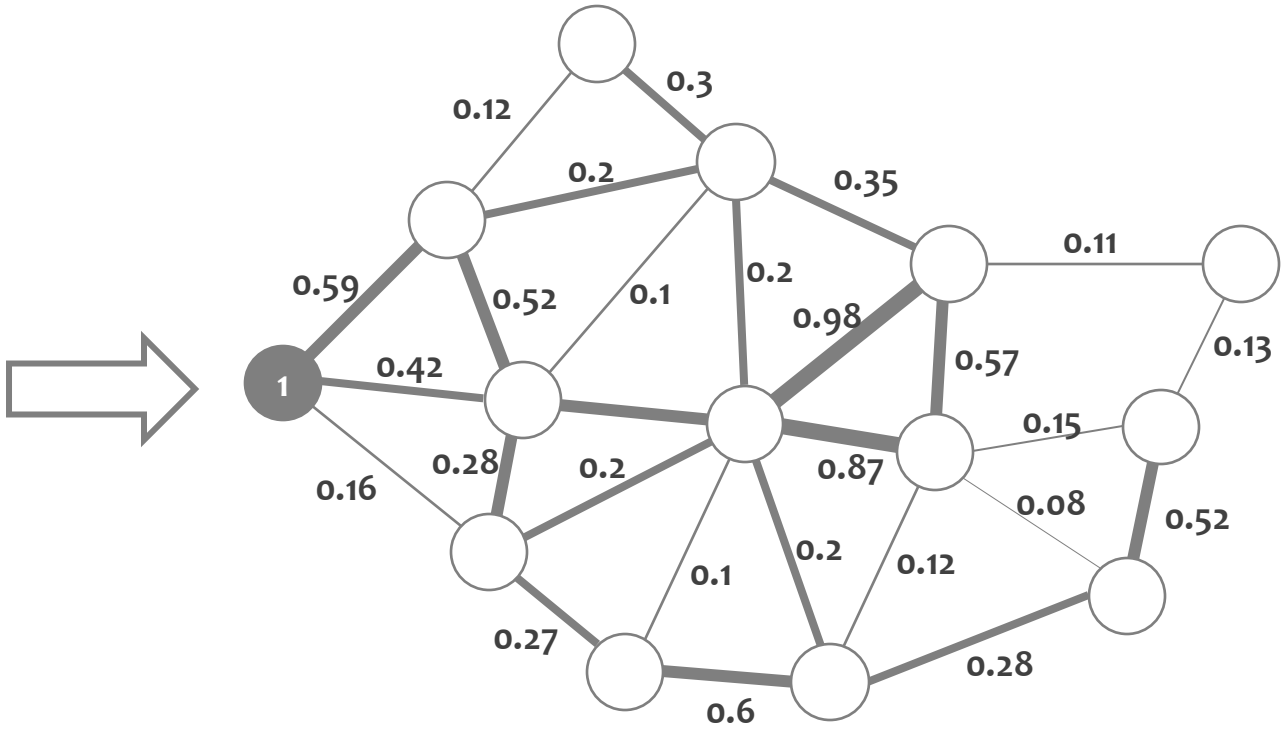
Perturbation Transfer Probability Matrix (P)

0.00	0.50	0.10	0.62	1.00
0.50	0.00	0.40	0.87	0.20
0.10	0.40	0.00	0.98	0.22
0.62	0.87	0.98	0.00	0.75
1.00	0.20	0.22	0.75	0.00

Perturbation Propagation Algorithm

Perturbation Transfer Probability Matrix (P)

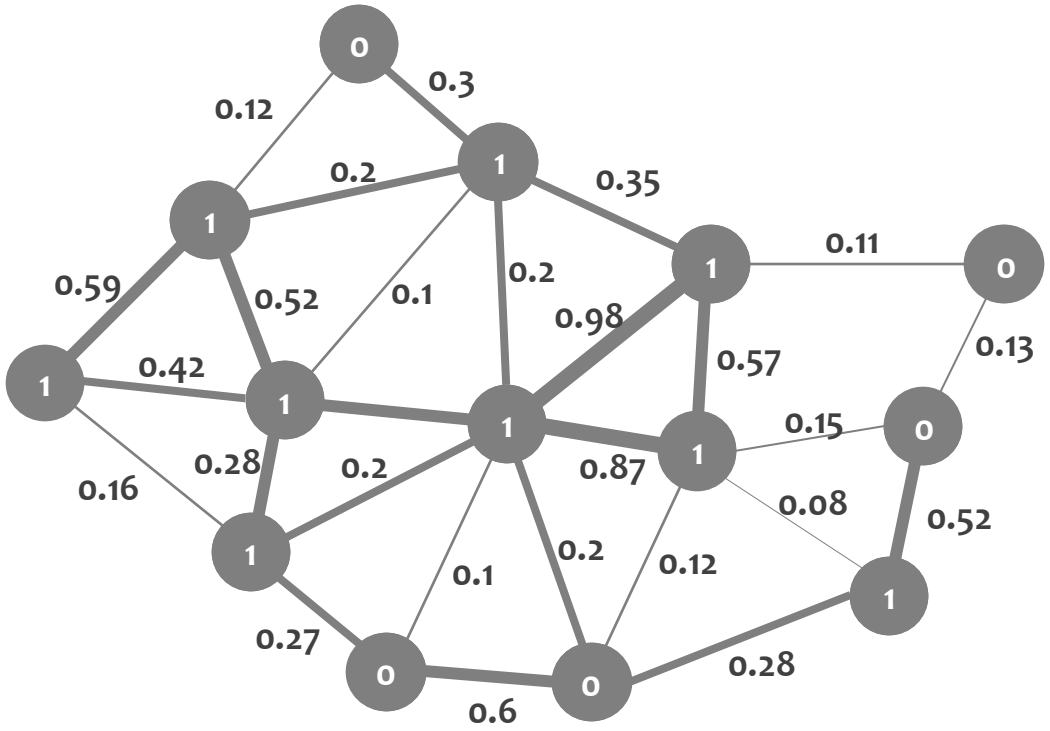
0.00	0.50	0.10	0.62	1.00
0.50	0.00	0.40	0.87	0.20
0.10	0.40	0.00	0.98	0.22
0.62	0.87	0.98	0.00	0.75
1.00	0.20	0.22	0.75	0.00



Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Perturbation Propagation Algorithm

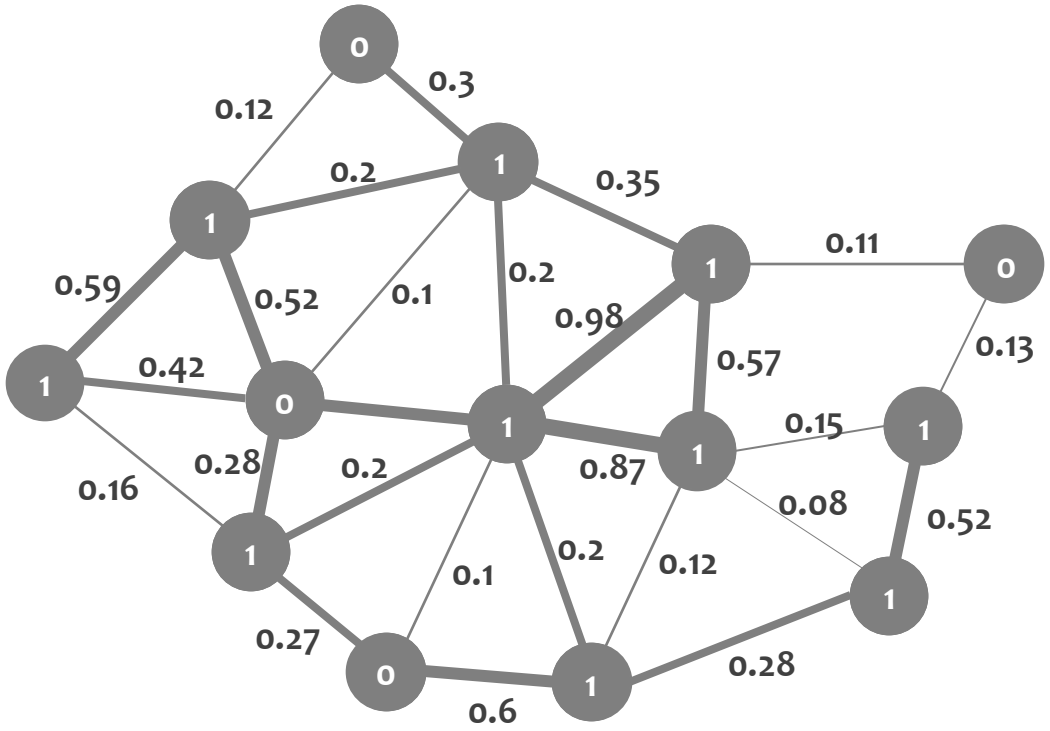
Round 1



Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Perturbation Propagation Algorithm

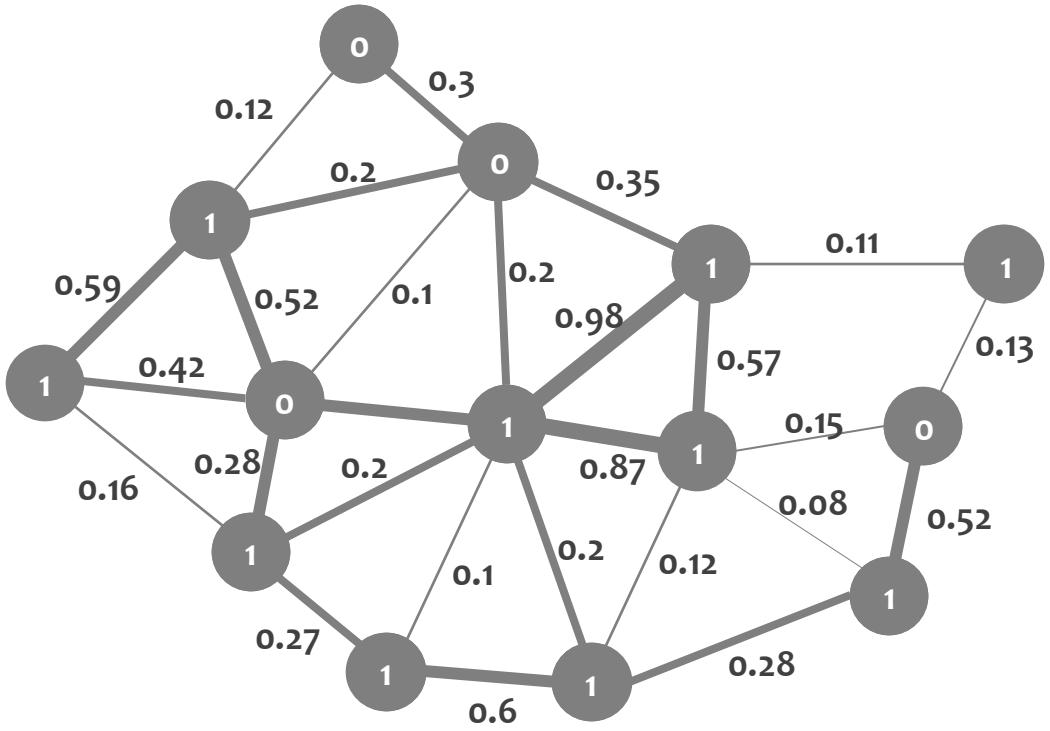
Round 2



Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Perturbation Propagation Algorithm

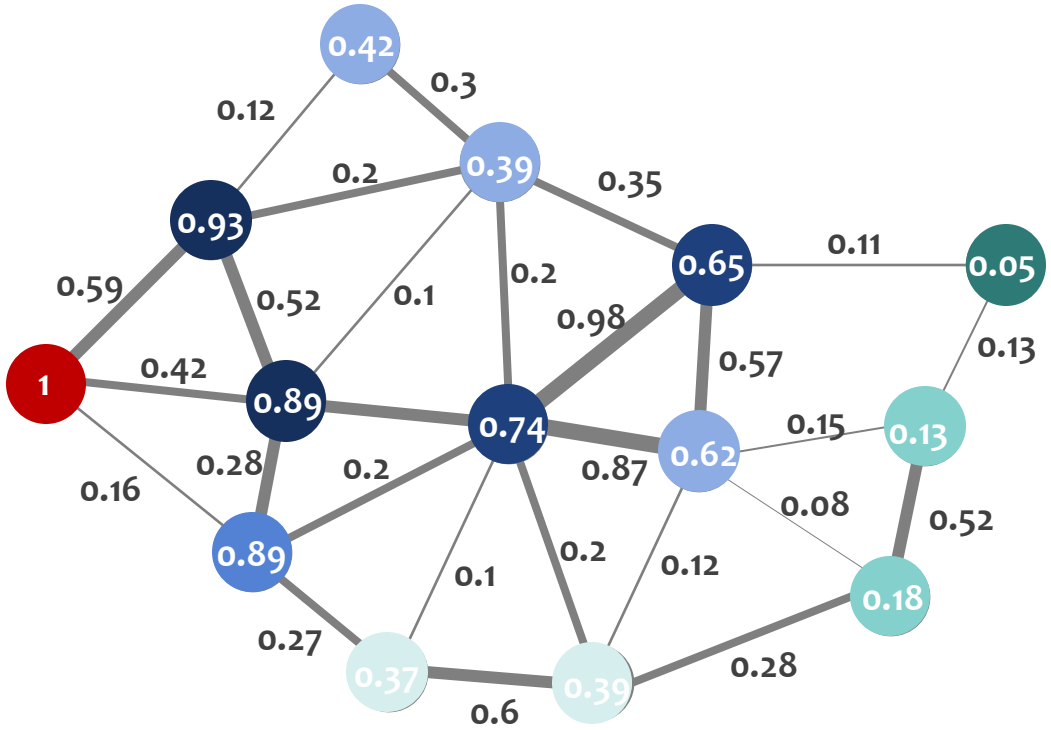
Round 3



Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Perturbation Propagation Algorithm

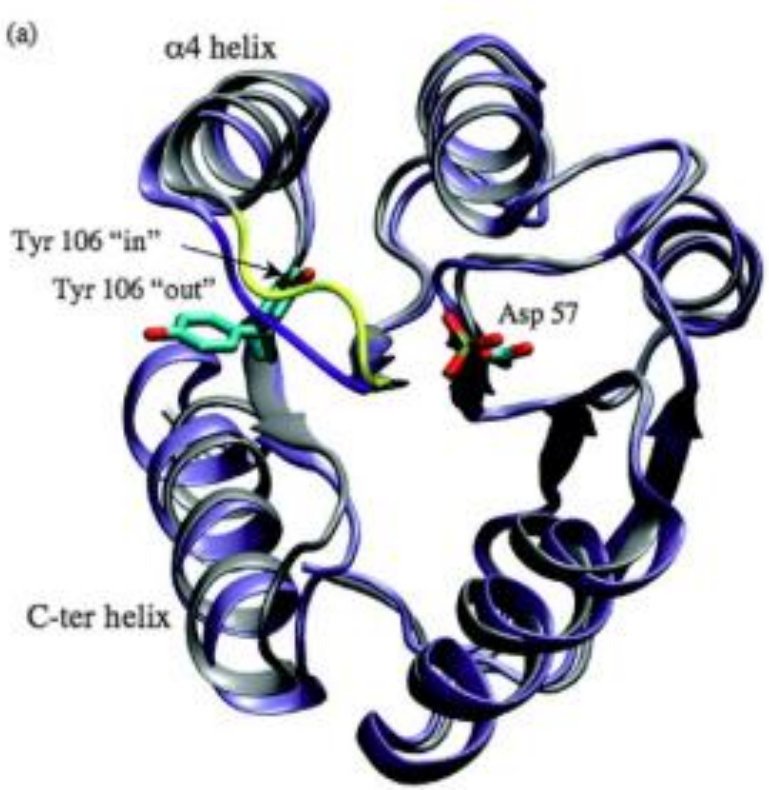
Repeat 10000 times

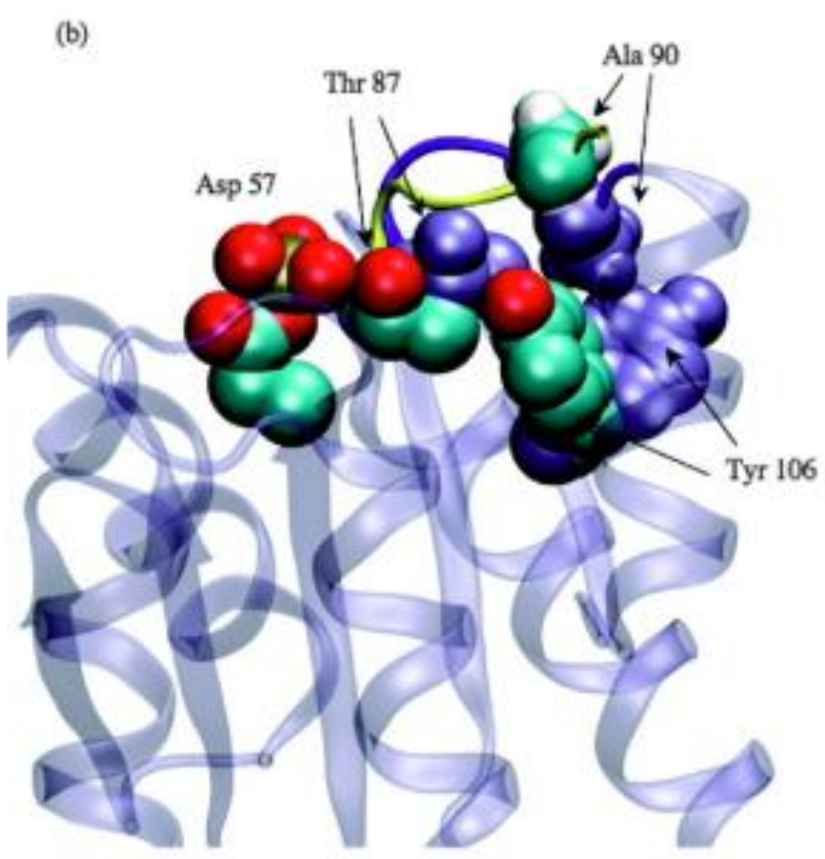


Allosteric Correlation Intensity

Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Chemotaxis protein Y (CheY)



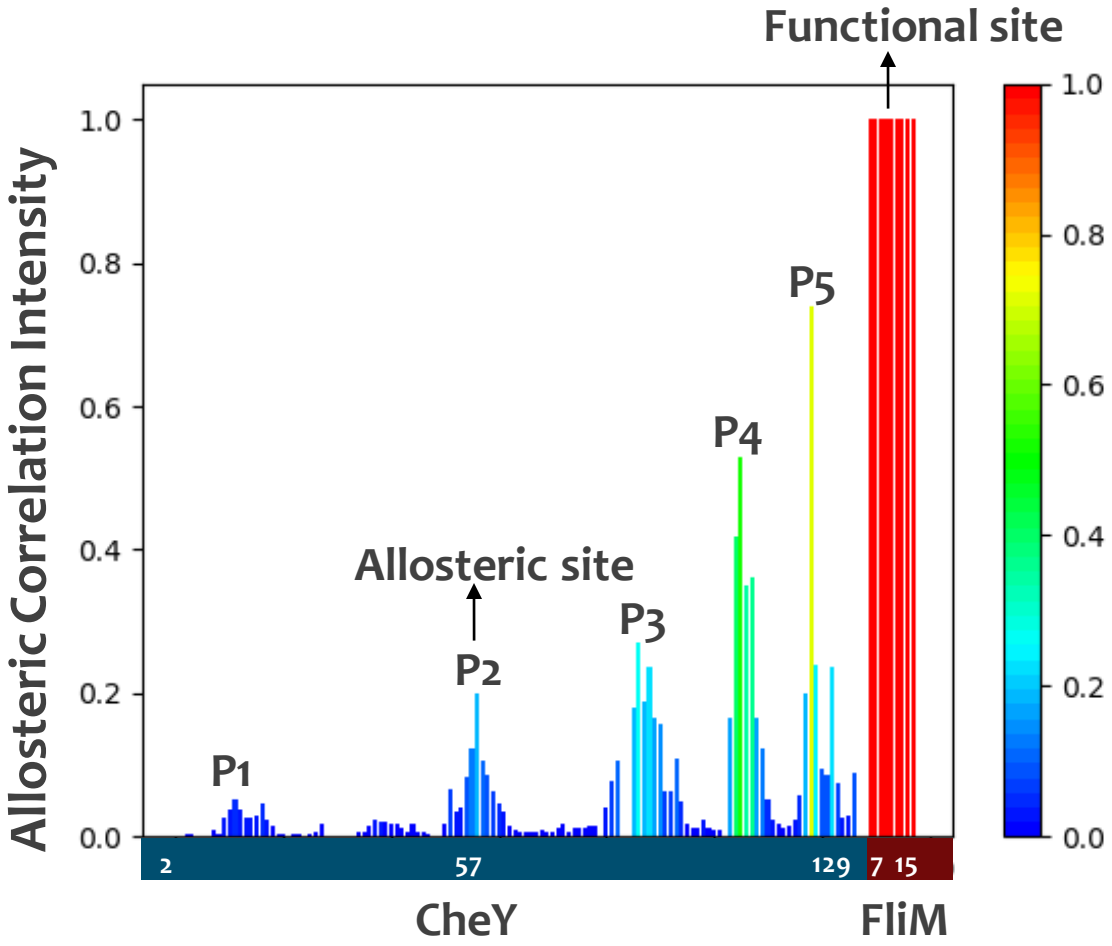


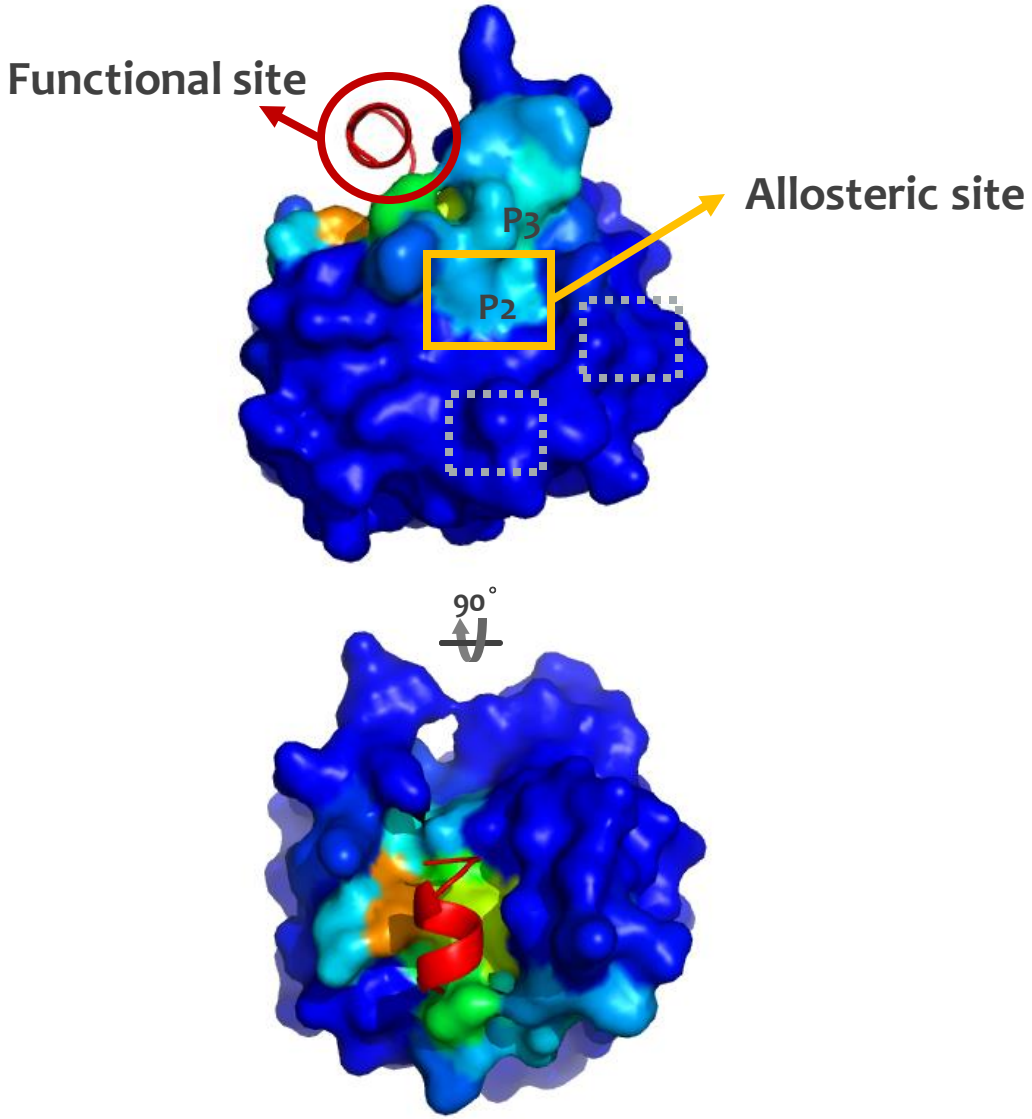
"Y–T" coupling scheme

The phosphorylation of D57 residue of CheY can activate the binding of FliM and other flagellar motors at the distal binding surface.

Lee S-Y, et al. Nat Struct Biol 2001; 1: 52–56.
Formaneck, et al. Proteins 2006; 63: 846-867.

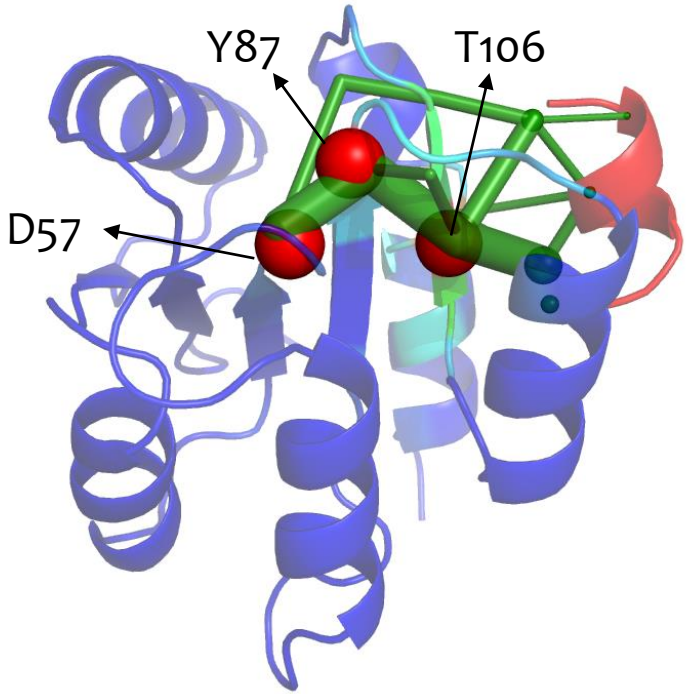
Identification of the allosteric site in CheY

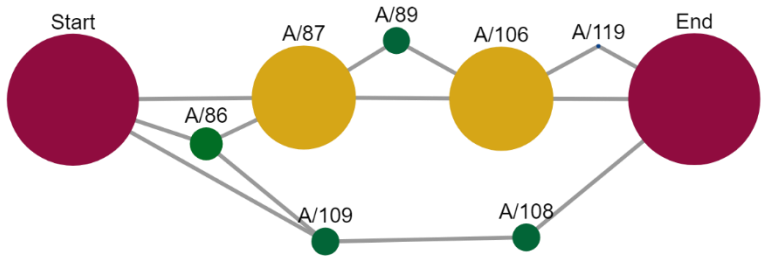


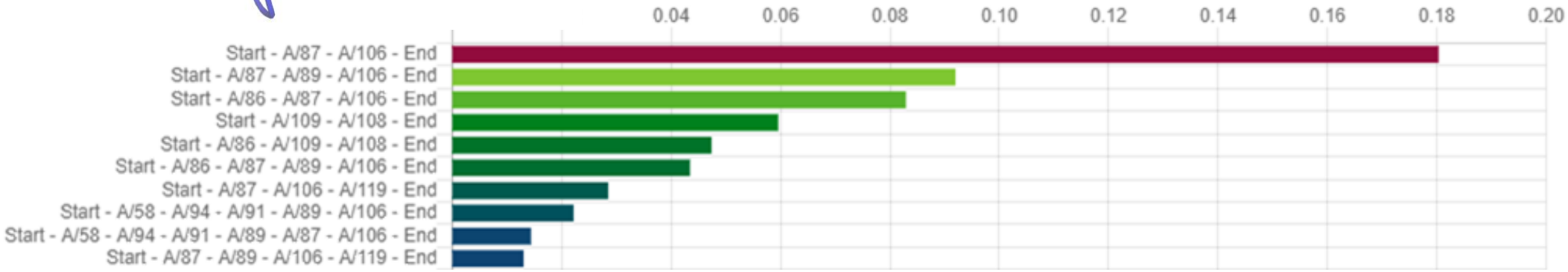


Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Critical residues in allosteric pathways of CheY

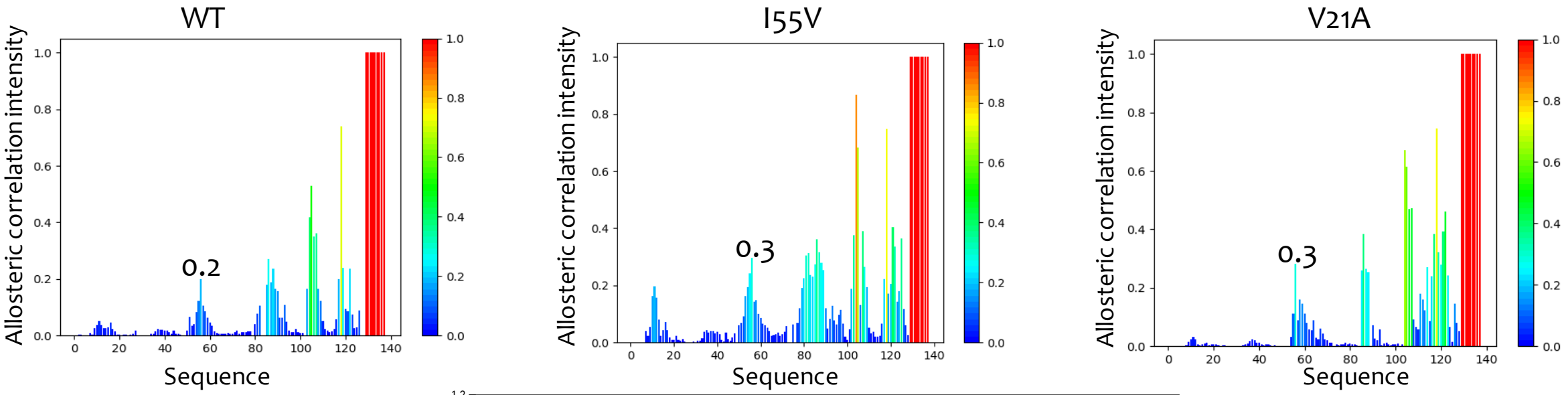


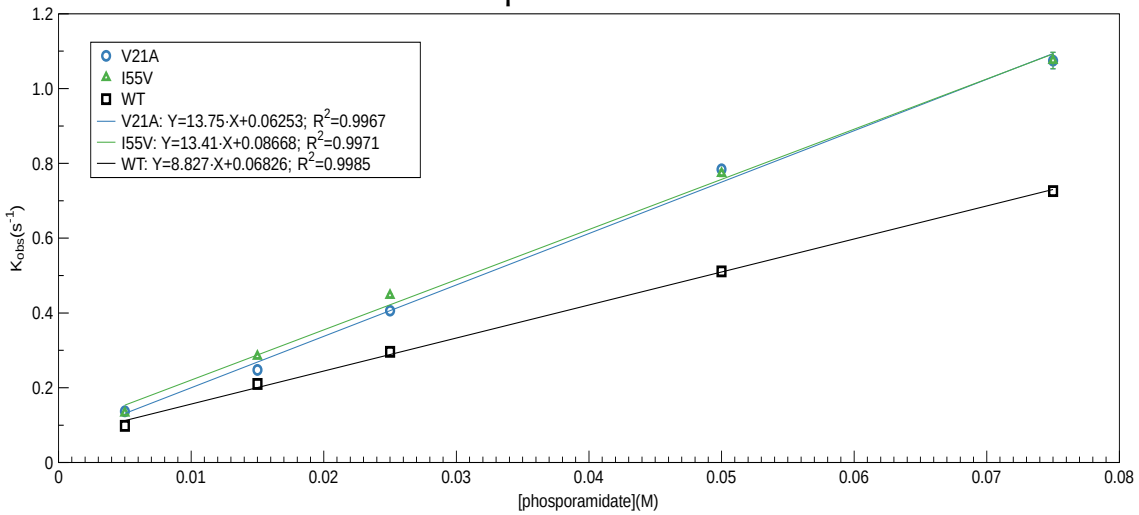




Wang, J., et al. Nature communications 11.1 (2020): 1-13.

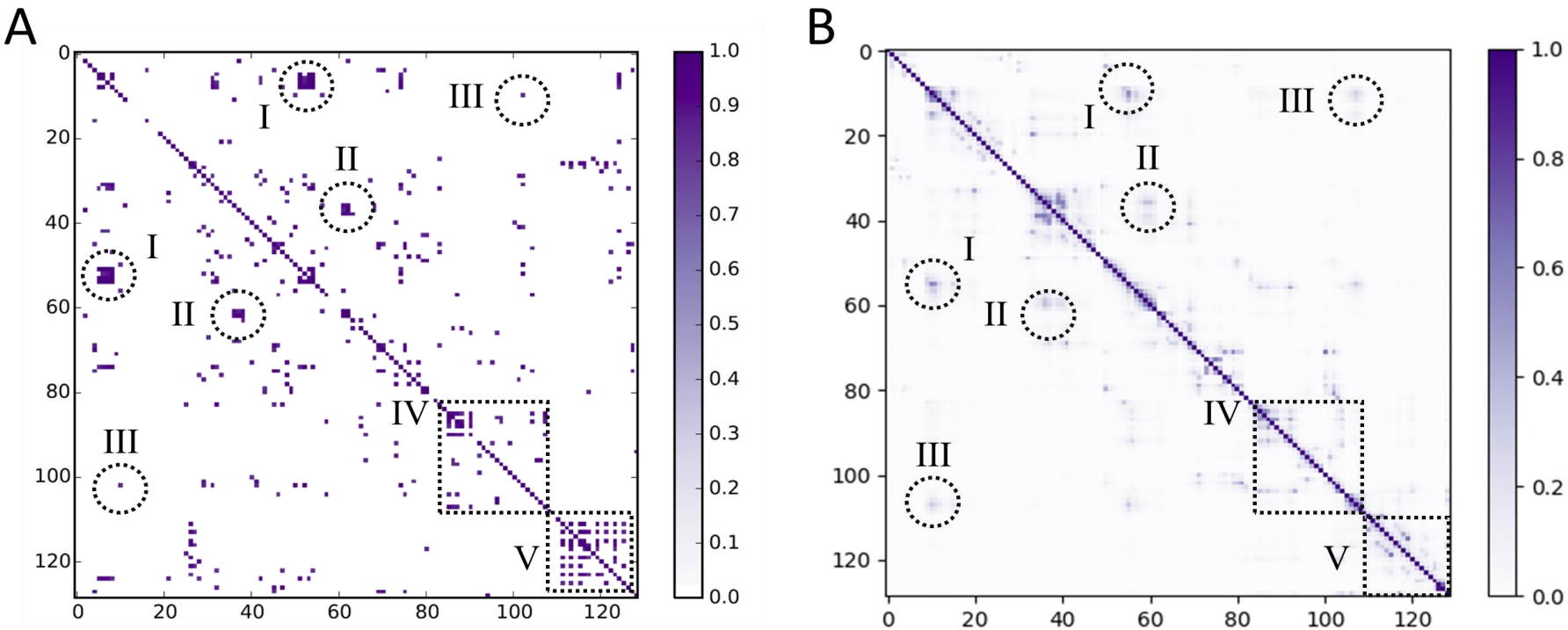
Allosteric correlation intensities in mutations of CheY





Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Prediction of All Residue-wise Allosteric Correlations



NMR Chemical Shift Covariance Analysis (CHESCA)

OHM

Wang, J., et al. Nature communications 11.1 (2020): 1-13.

Ohm website

PennState College of Medicine

Ohm Submit Queue Help Contact

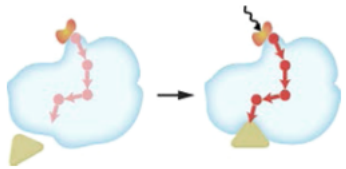
Explore Allostery

Identify allosteric sites and pathways

Predict Allosteric Sites | Identify Allosteric Pathways | Perturb Allosteric Pathways | Predict Allosteric Correlations

What is Ohm?

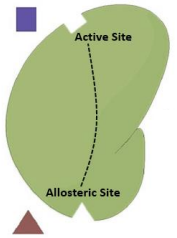
Allostery is a natural phenomenon in proteins whereby distal structural elements are dynamically coupled. The origins of the allosteric phenomenon are rooted in physical properties of inter-atomic interactions in heterogenous media. Protein sequences are heterogeneous and their corresponding structures represent a diverse range of forces between amino acids that shape their structures. In proteins perturbations of amino acids will propagate non-uniformly throughout the structure: a residue that is stronger coupled to a neighbor, would be more affected by the perturbation of the neighbor.



ohm.dokhlab.org

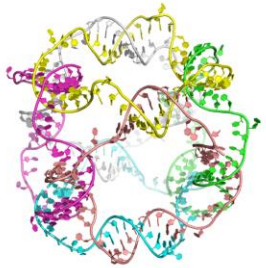
Applications of Structural Bioinformatics

Protein Allostery



Identifying new druggable site

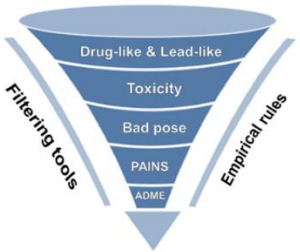
RNA Modeling



Drug delivery

Antisense drugs

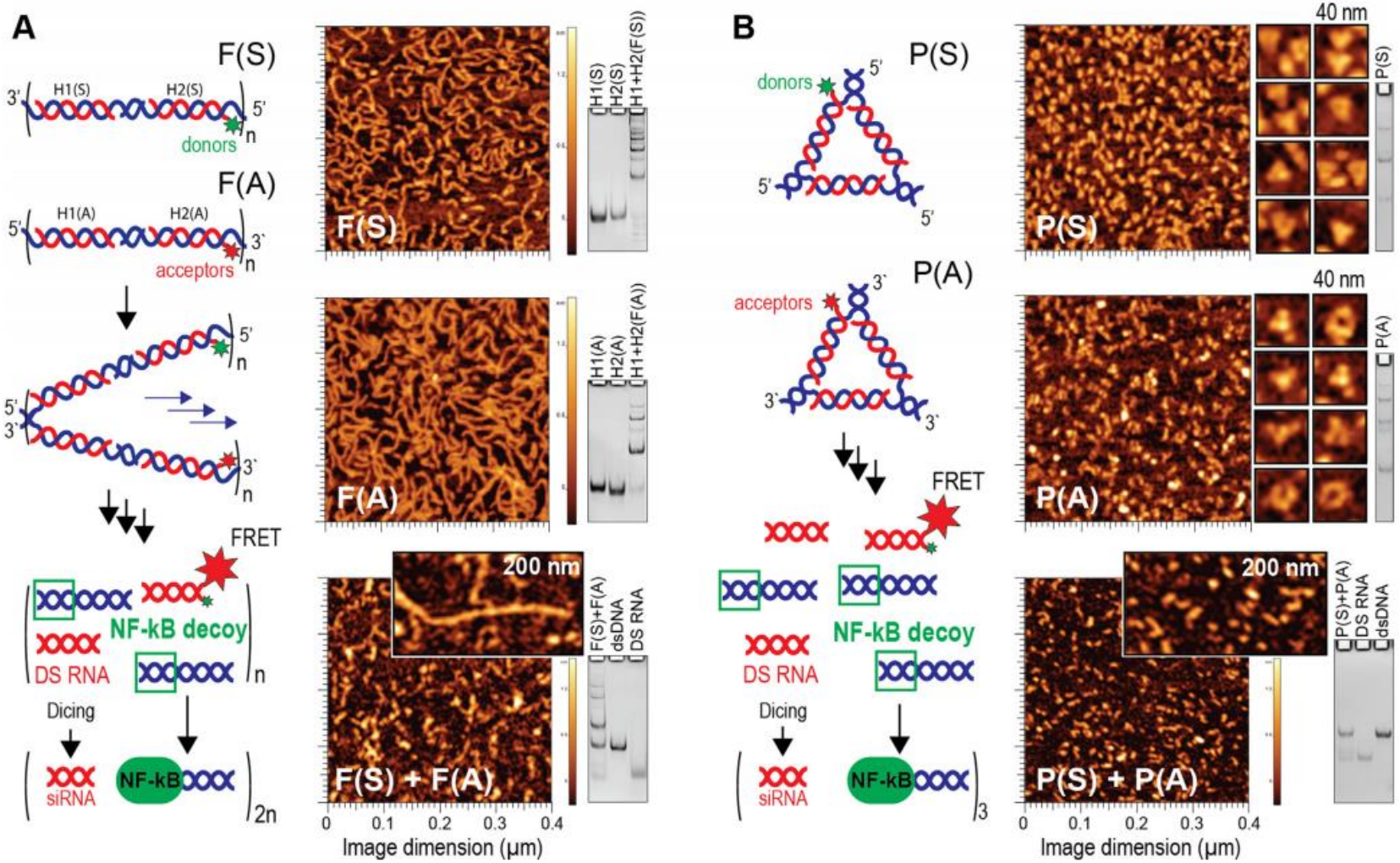
Compound-protein interaction



Drug screening, Target identification

Contents

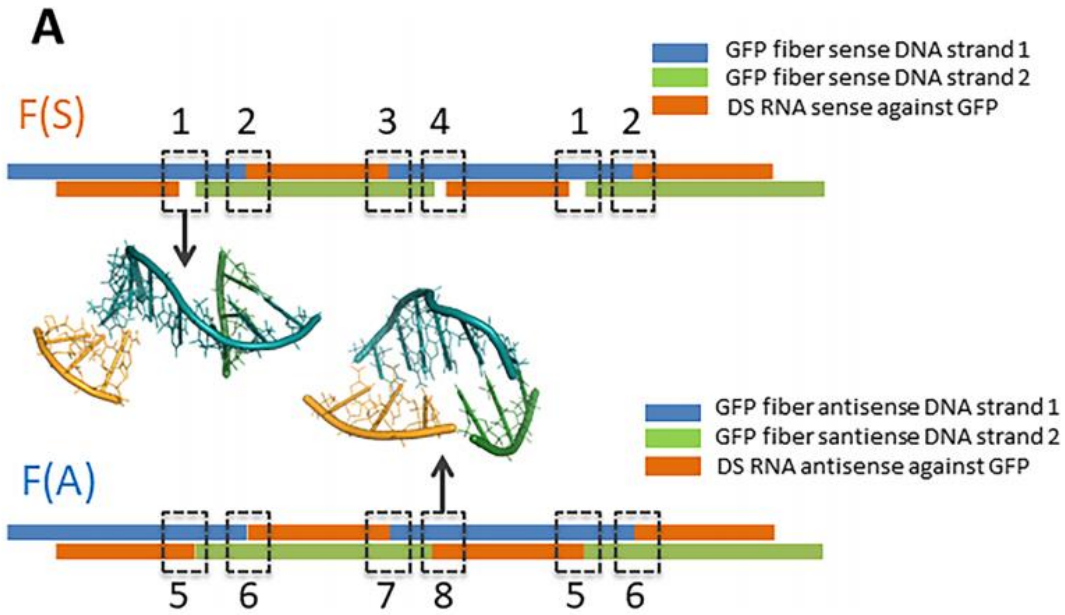
RNA Modeling in Nucleic Acid Nanoparticle Design

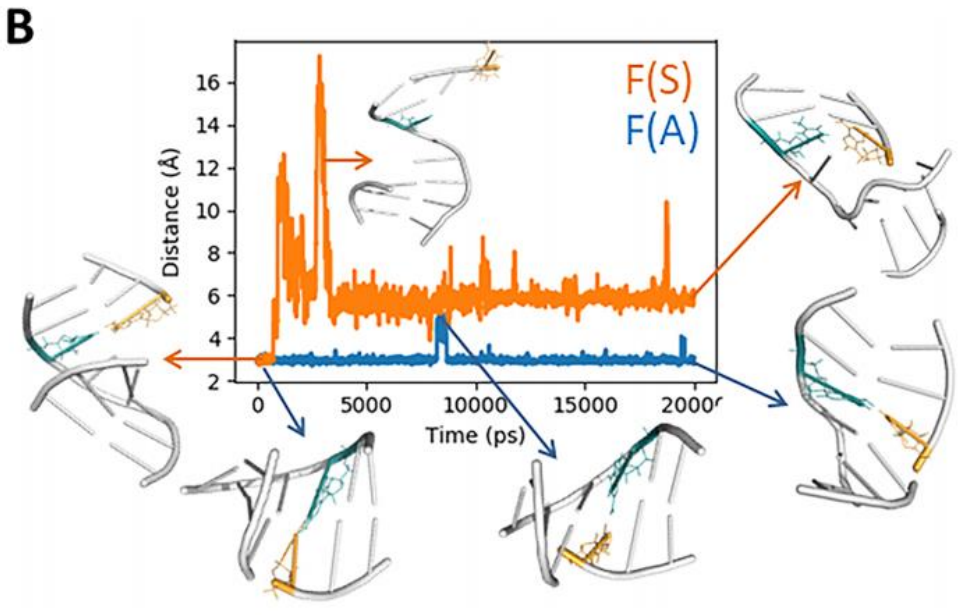


UNC CHARLOTTE

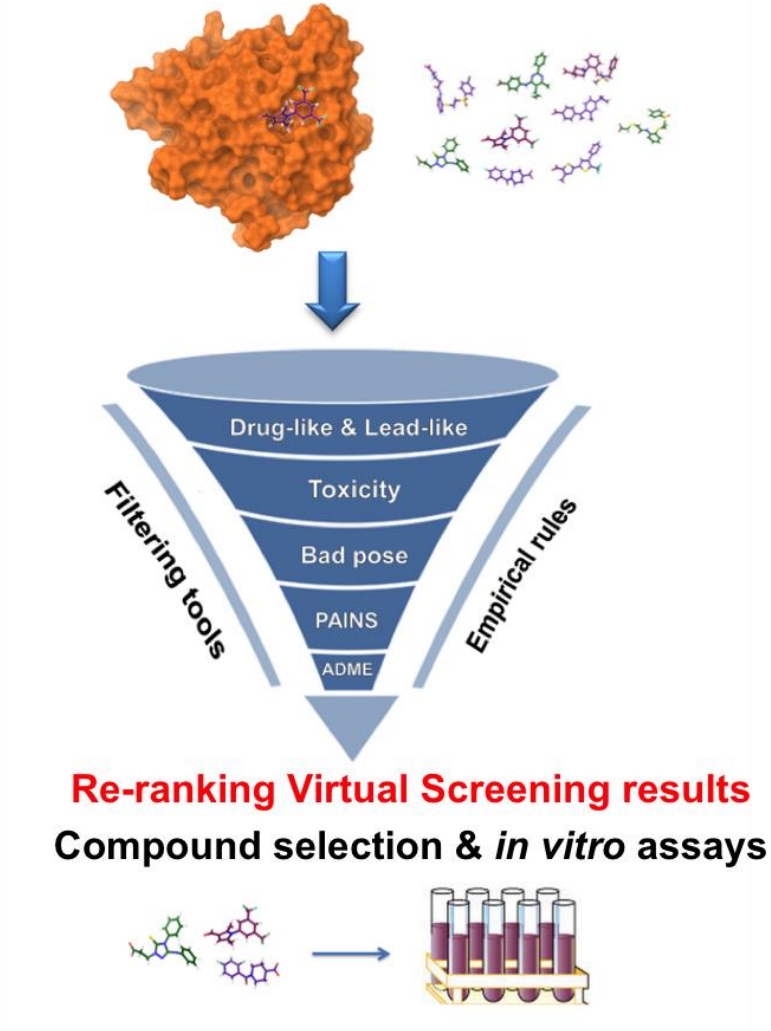
Dr. Kirill Afonin

RNA Modeling in Nucleic Acid Nanoparticle Design

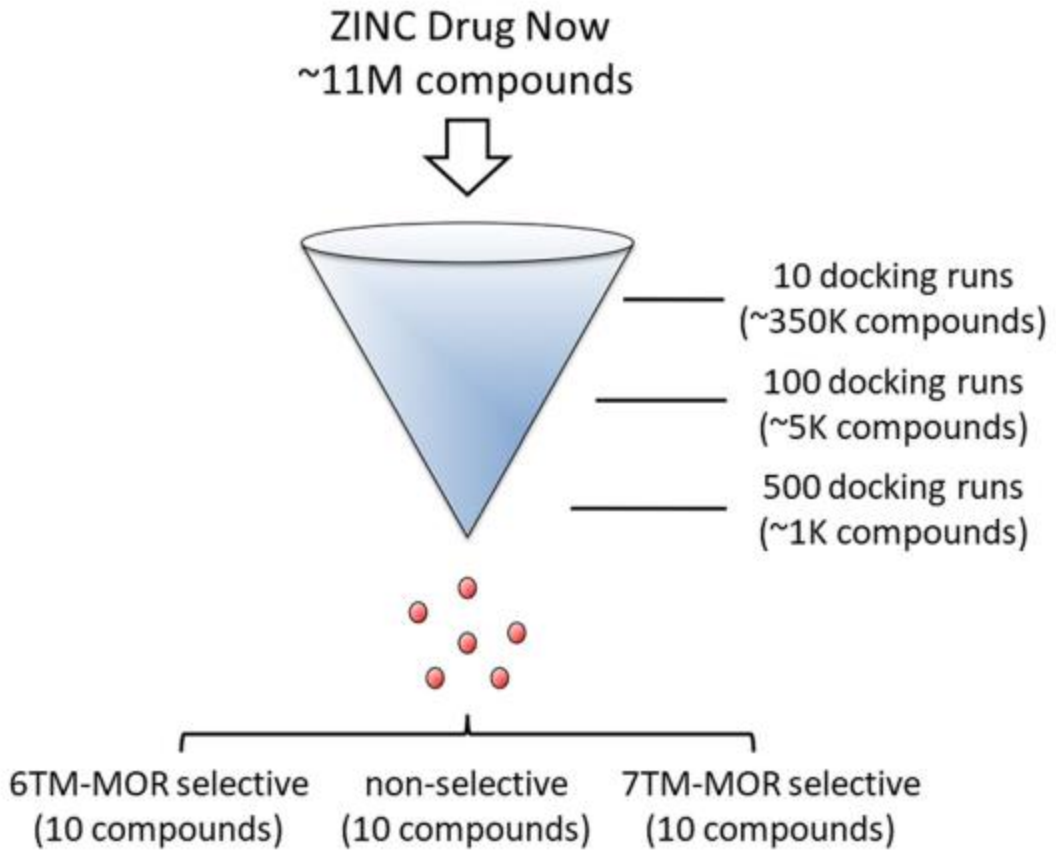




Virtual Screening



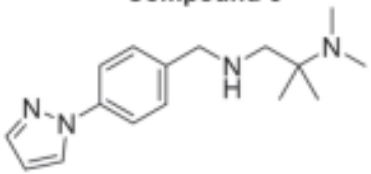
Targeting 6- and 7-transmembrane μ-opioid receptor isoforms



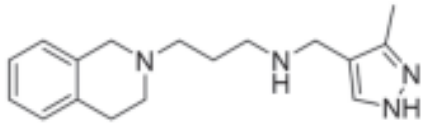
Dr. Luda Diatchenko

UNC

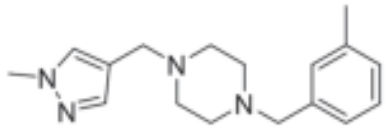
Compound 5



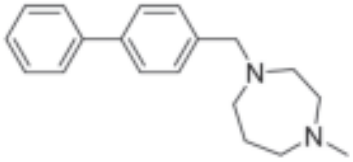
Compound 10



Compound 11

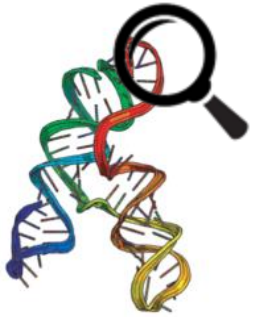


Compound 25

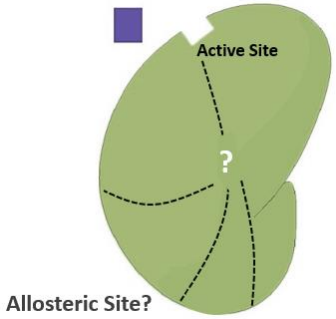


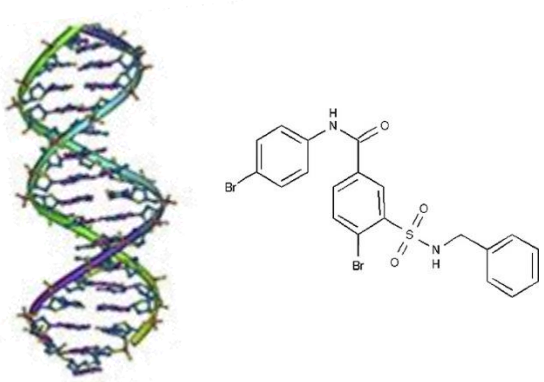
Conclusion

RNA Scoring

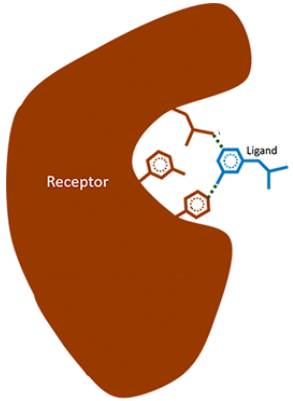


Protein Allostery Mapping

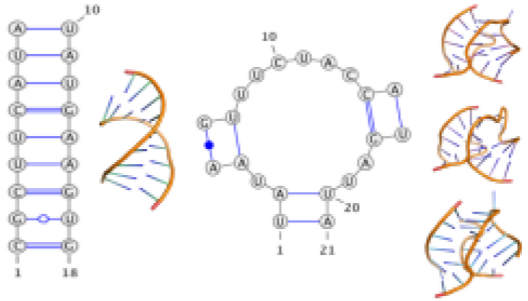


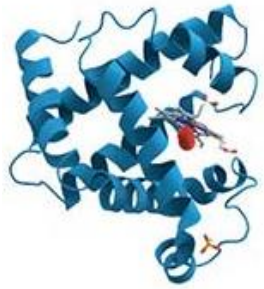


CPI Evaluation



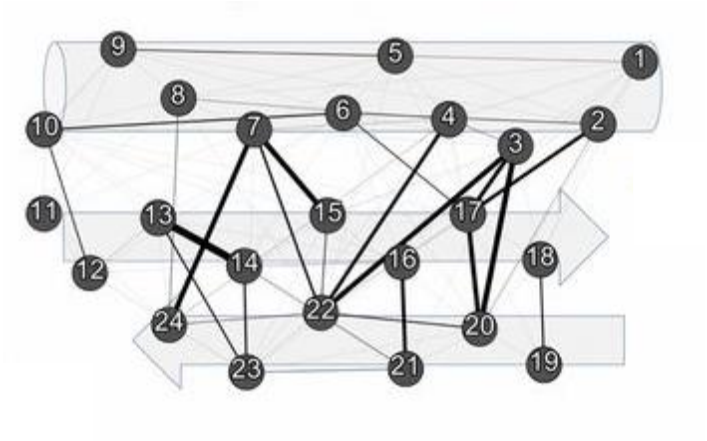
RNA Modeling





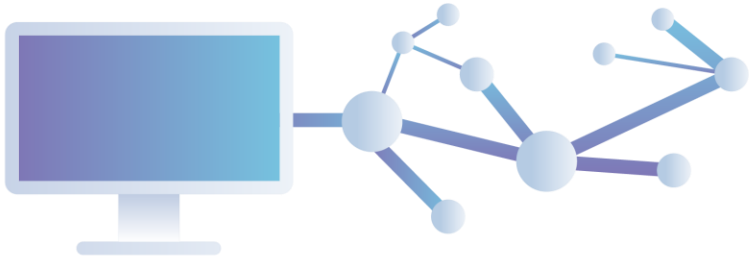
Understanding → Prediction → Design

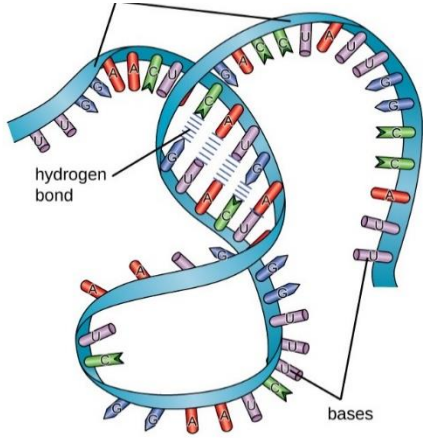
Designing new biomolecules



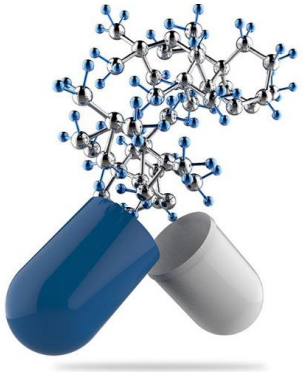
Protein Design

Graph Design



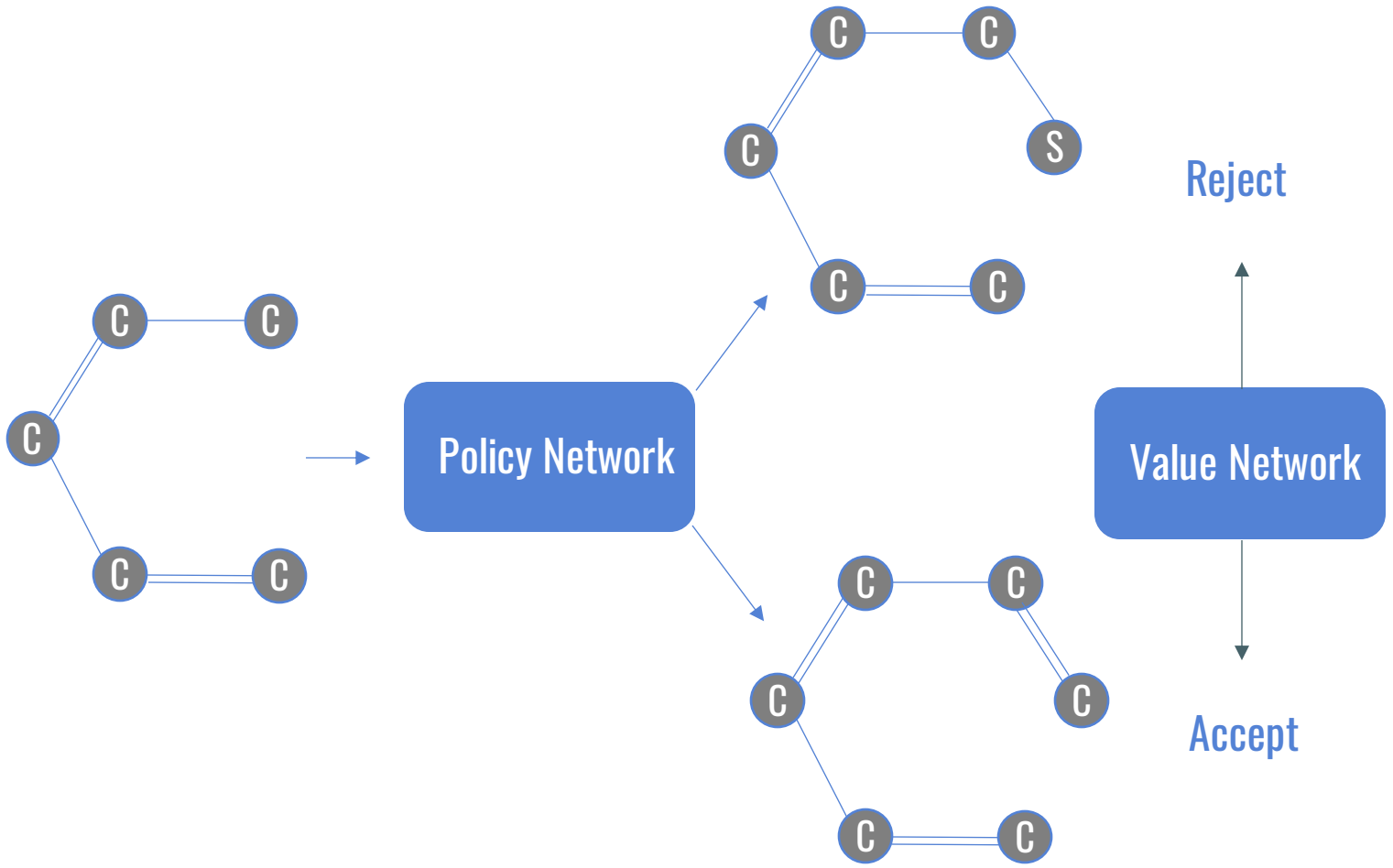


RNA Design



Drug Design

Graph Design



Policy network ***makes decisions*** *and Value network* ***evaluates the situation***

Questions?

PennState