AI-Aided Structure-Based Drug discovery

www.sgo.or.kr

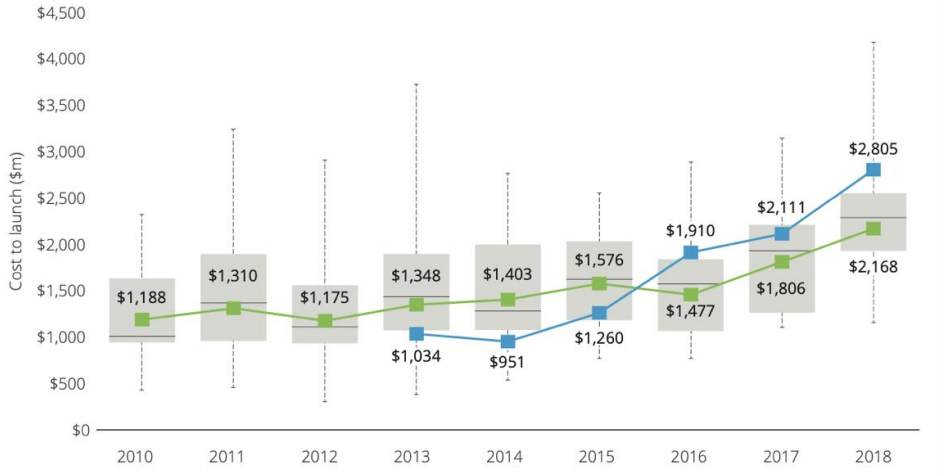
Woong-Hee Shin
Department of Biomedical Informatics
Korea University College of Medicine

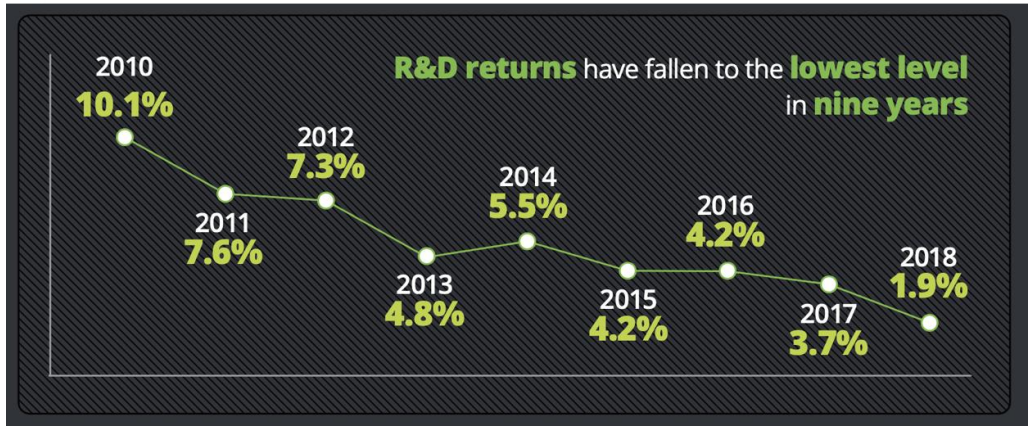
DECLARATION OF INTERESTS

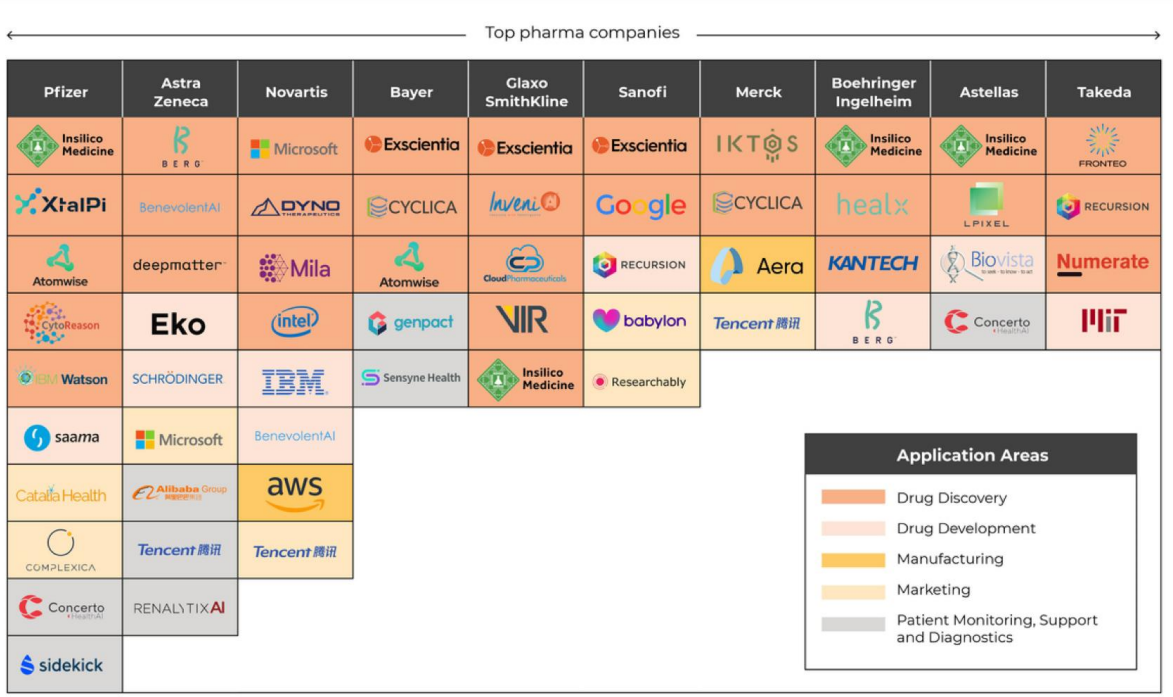
Nothing to declare

In Silico Drug Discovery

- Top pharmaceutical companies are collaborating with computational start-ups

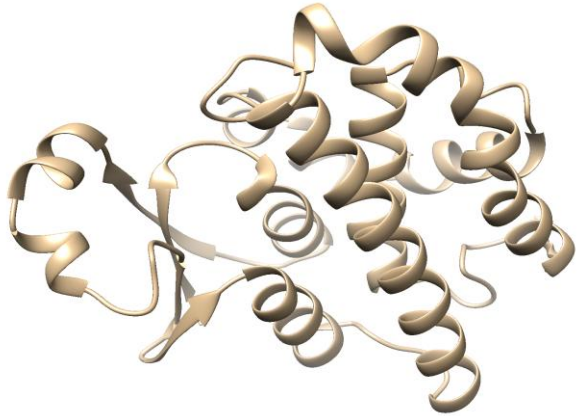






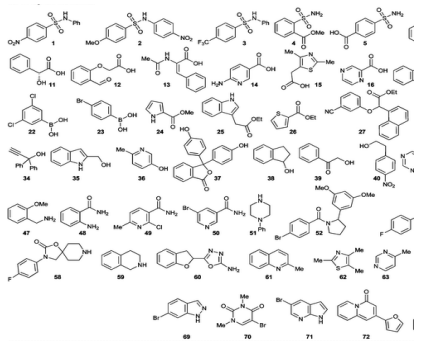
Structure-Based Virtual Screening

- Searching the binding pose of a ligand within a pocket
- Receptor structure is required
- Protein-ligand docking

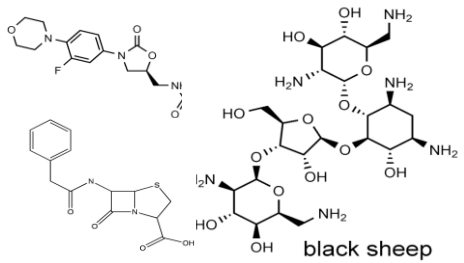


Target protein

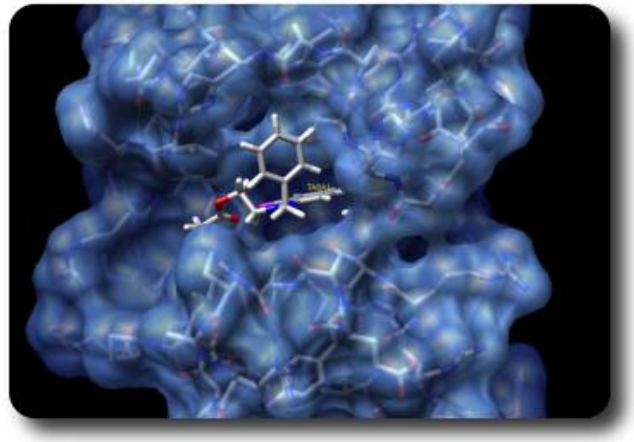
+



Chemical library



Binding ligands



$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

Docking of Millions

- Aim: Screening million scale library with a time scale of a week

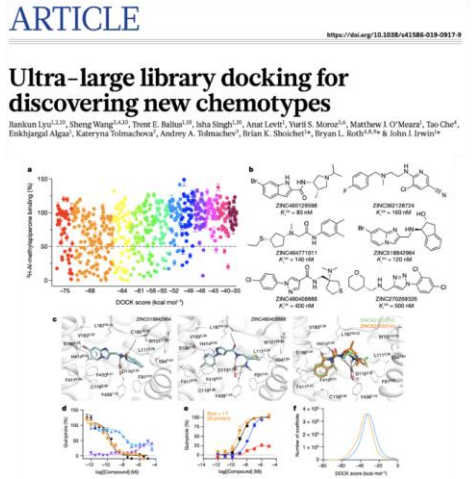
10^{63}
Drug-like small molecules in chemical space

10^{24}
Stars in the universe

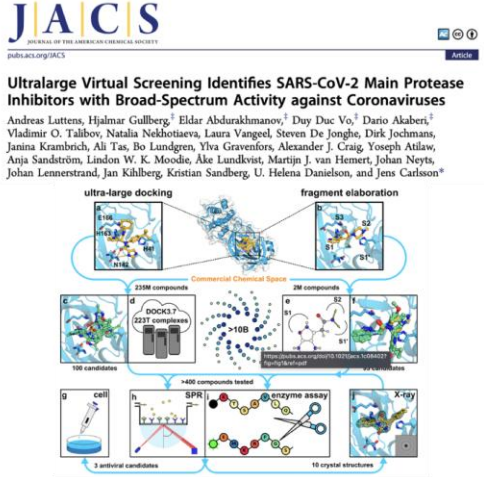
10^{20}
Potential compounds in Merck KGaA's Merck Accessible Inventory (MASSIV)

10^{4}
Small-molecule drugs

Protein	CDK2	EGFr	FGFr1	P38 MAP	SRC	VEGFr2	Average
# of Compounds	1500	1800	1800	1800	1800	2220	1816.7
AutoDock Vina (8 CPUs)	5.2 h	5.2 h	9.8 h	6.5 h	7.6 h	8.6 h	7.2 h
DOCK6 (1 CPU)	19.3 h	20.2 h	53.2 h	26.0 h	49.8 h	46.4 h	35.8 h

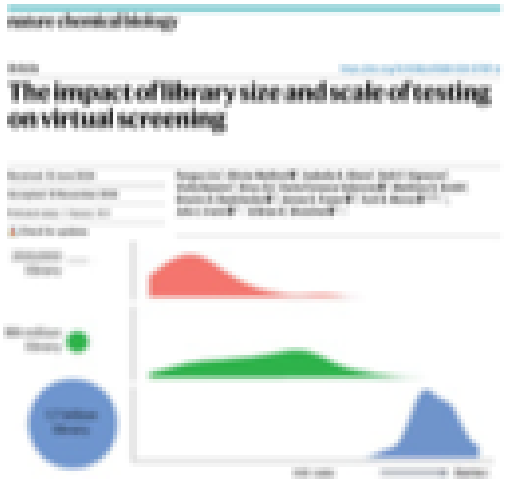

ARTICLE

Ultra-large library docking for discovering new chemotypes


J A C S

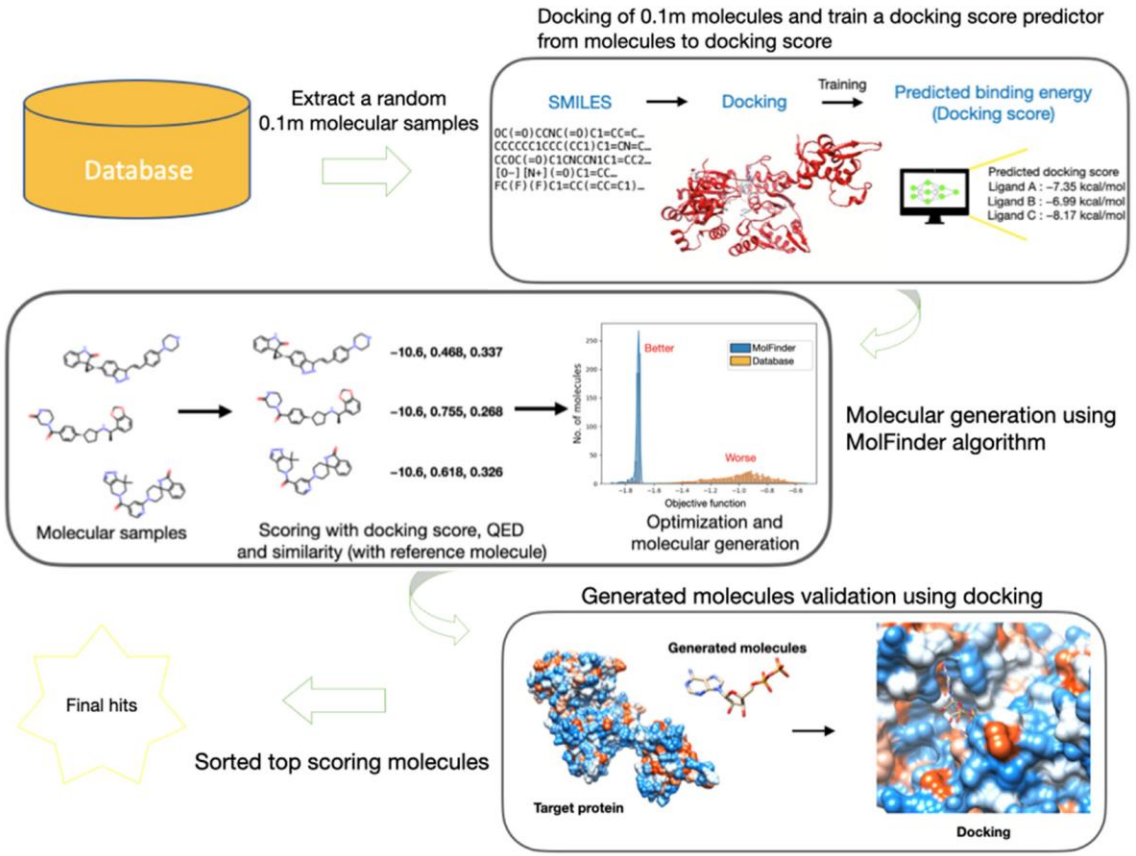
Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses

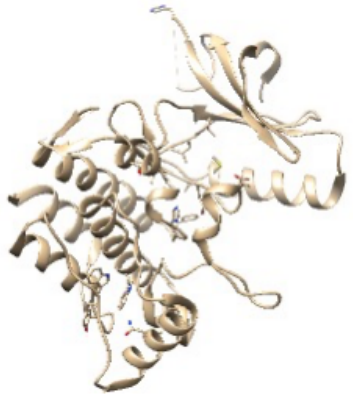
Andreas Luttens, Hjalmar Gullberg, Eldar Abdurakhmanov, Duy Duc Vo, Dario Akaberi, Vladimir O. Talibov, Natalia Nekhotiaeva, Laura Vangeel, Steven De Jonghe, Dirk Jochmans, Janina Krambrich, Ali Tas, Bo Lundgren, Ylva Gravenfors, Alexander J. Craig, Yoseph Atilaw, Anja Sandström, Lindon W. K. Moodie, Åke Lundkvist, Martijn J. van Hemert, Johan Neyts, Johan Lennerstrand, Jan Kihlberg, Kristian Sandberg, U. Helena Danielson, and Jens Carlsson*

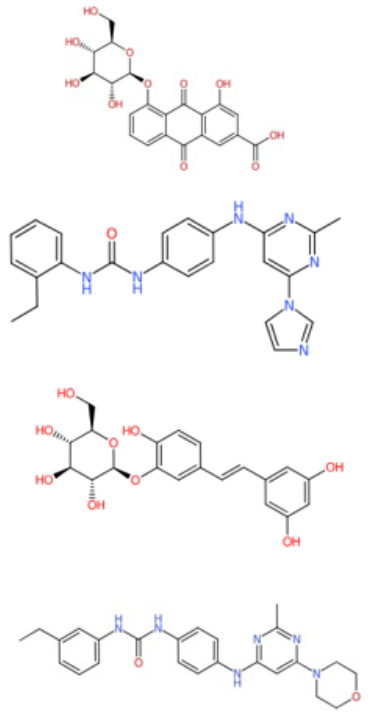

The impact of library size and scale of testing on virtual screening

Time for screening million scale library?

V-Dock







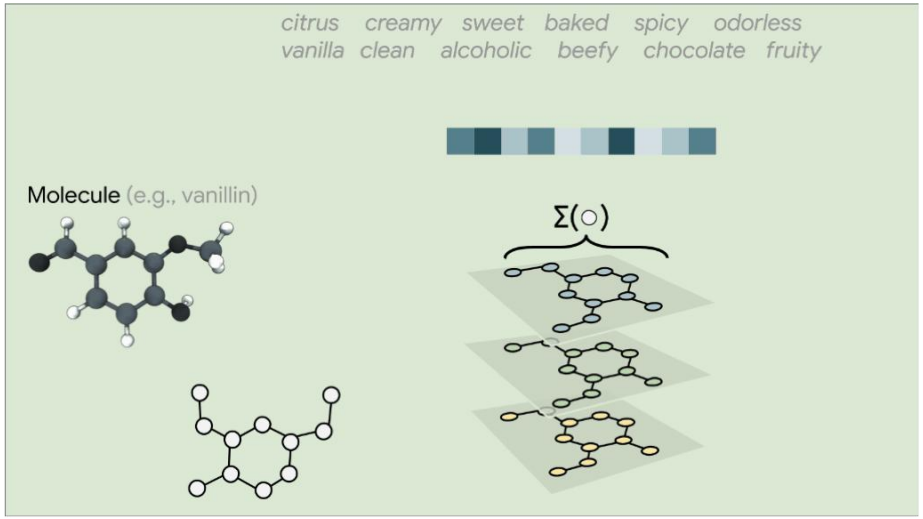
-7.65 kcal/mol

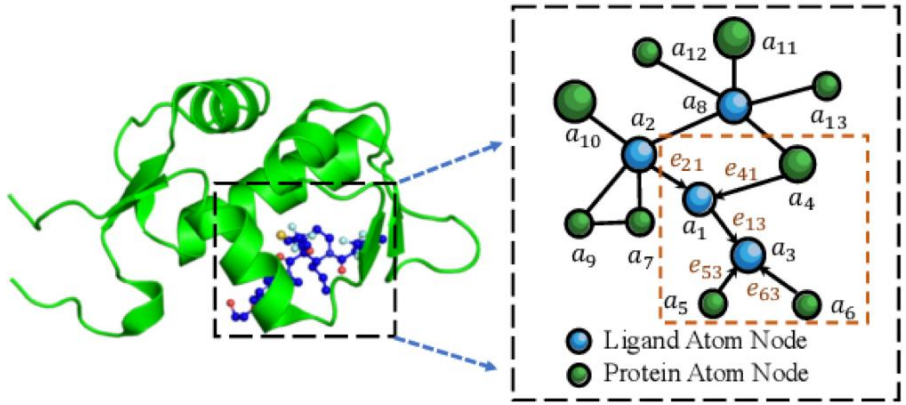
-13.86 kcal/mol

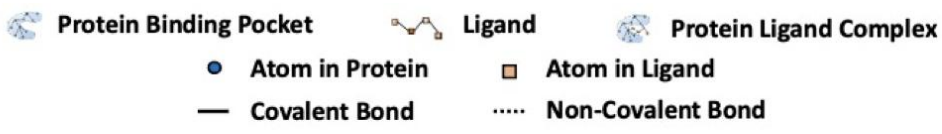
-8.17 kcal/mol

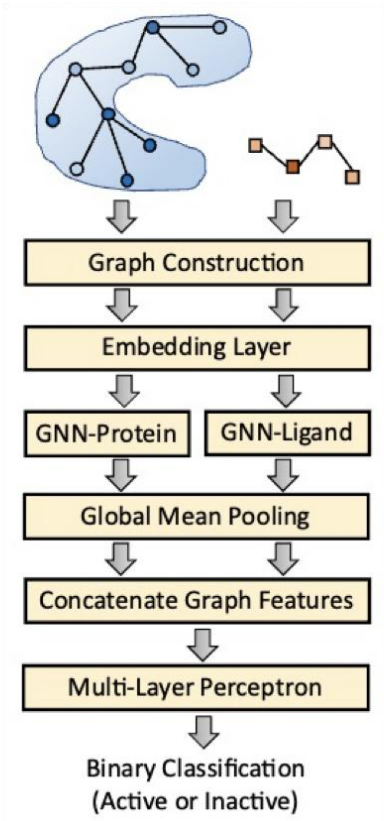
??? kcal/mol

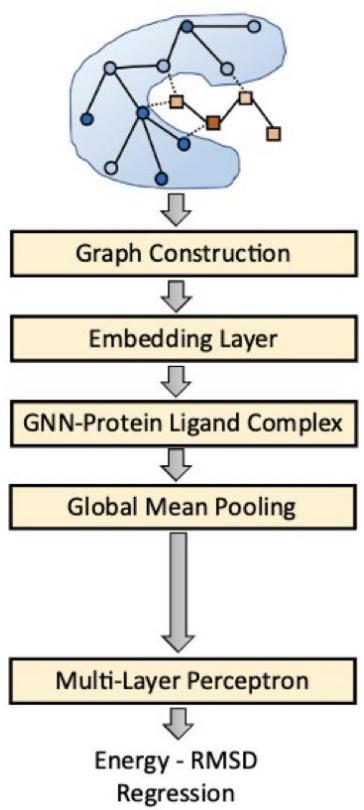
AK-Score2


citrus creamy sweet baked spicy odorless
vanilla clean alcoholic beefy chocolate fruity
Molecule (e.g., vanillin)
Σ(○)

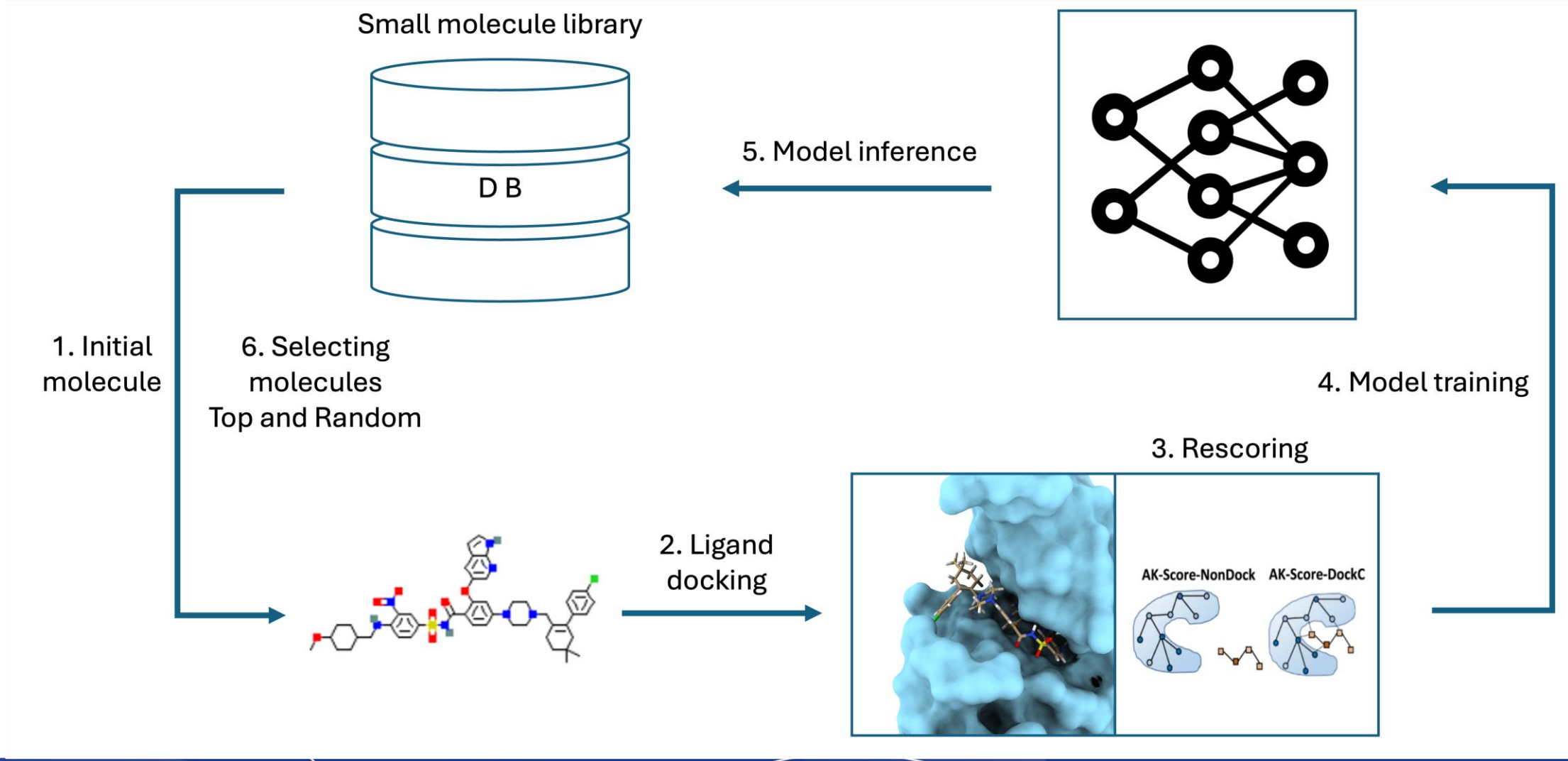

Ligand Atom Node
Protein Atom Node


Protein Binding Pocket
Ligand
Protein Ligand Complex
Atom in Protein
Atom in Ligand
Covalent Bond
Non-Covalent Bond


Graph Construction
Embedding Layer
GNN-Protein
GNN-Ligand
Global Mean Pooling
Concatenate Graph Features
Multi-Layer Perceptron
Binary Classification
(Active or Inactive)


Graph Construction
Embedding Layer
GNN-Protein Ligand Complex
Global Mean Pooling
Multi-Layer Perceptron
Energy - RMSD
Regression

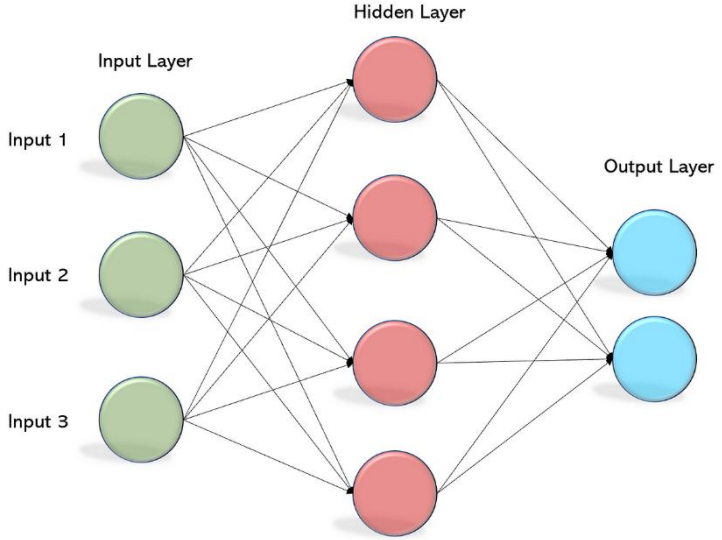
DoM: Iterative V-Dock



Benchmark

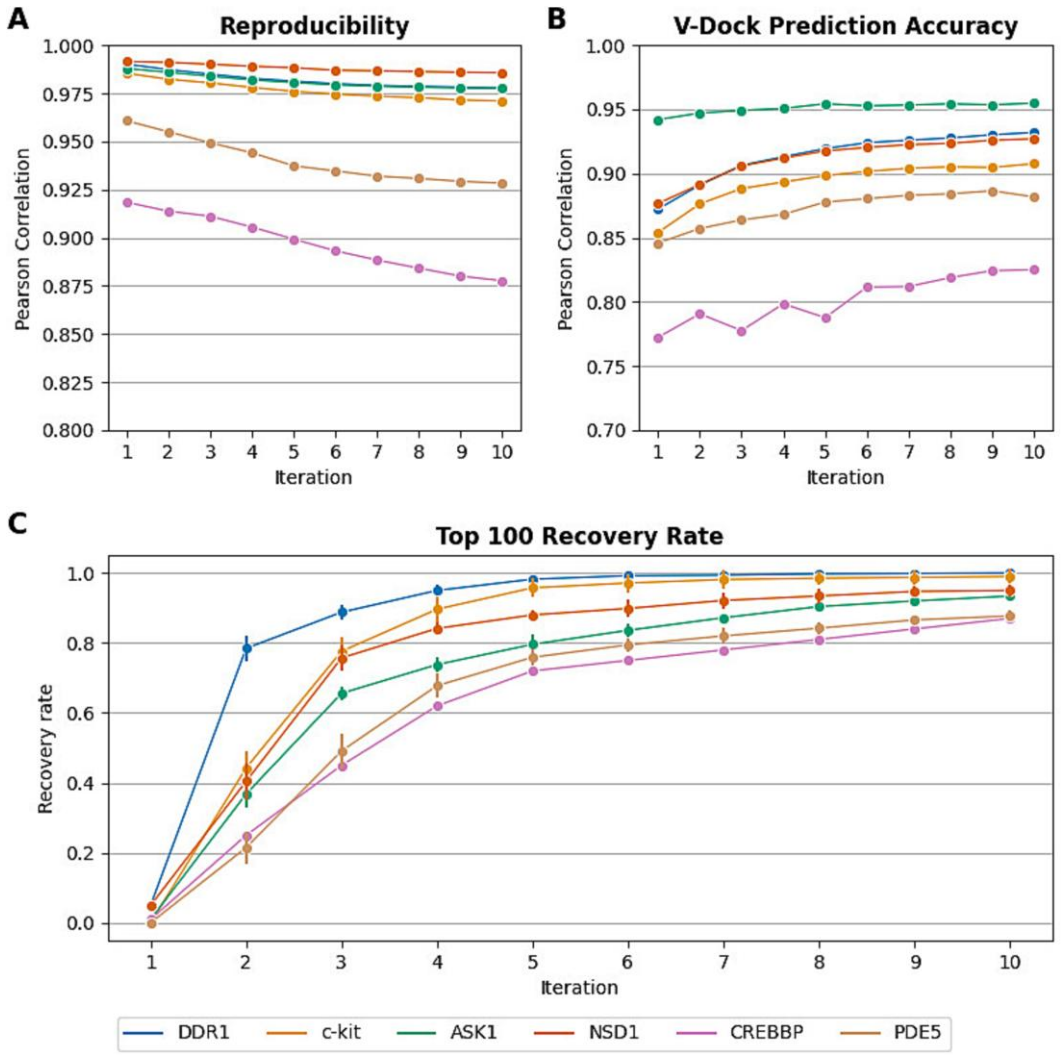
- Compare DoM result with virtual screening data

- Mcule in stock library (5M compounds)
- Learning algorithm: Multi-Layer Perceptron with 1386 features
- Initial compounds: 142k clustered from Mcule in stock library



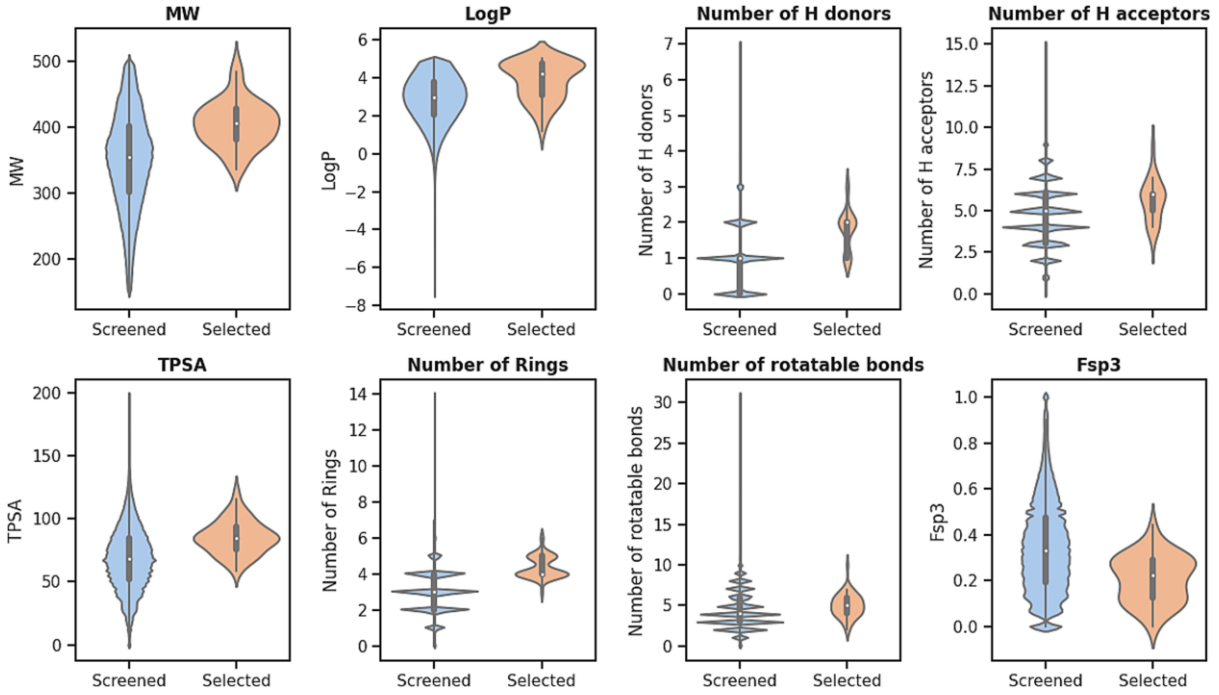
Target	Family	PDB ID
DDR1	Kinase	4BKJ
C-KIT	Kinase	1T46
ASK1	Kinase	6OYT
NSD1	Histone methyltransferase	6KQQ
CREBBP	Bromodomain	5NU3
PDE5	Phosphodiesterase	2H42

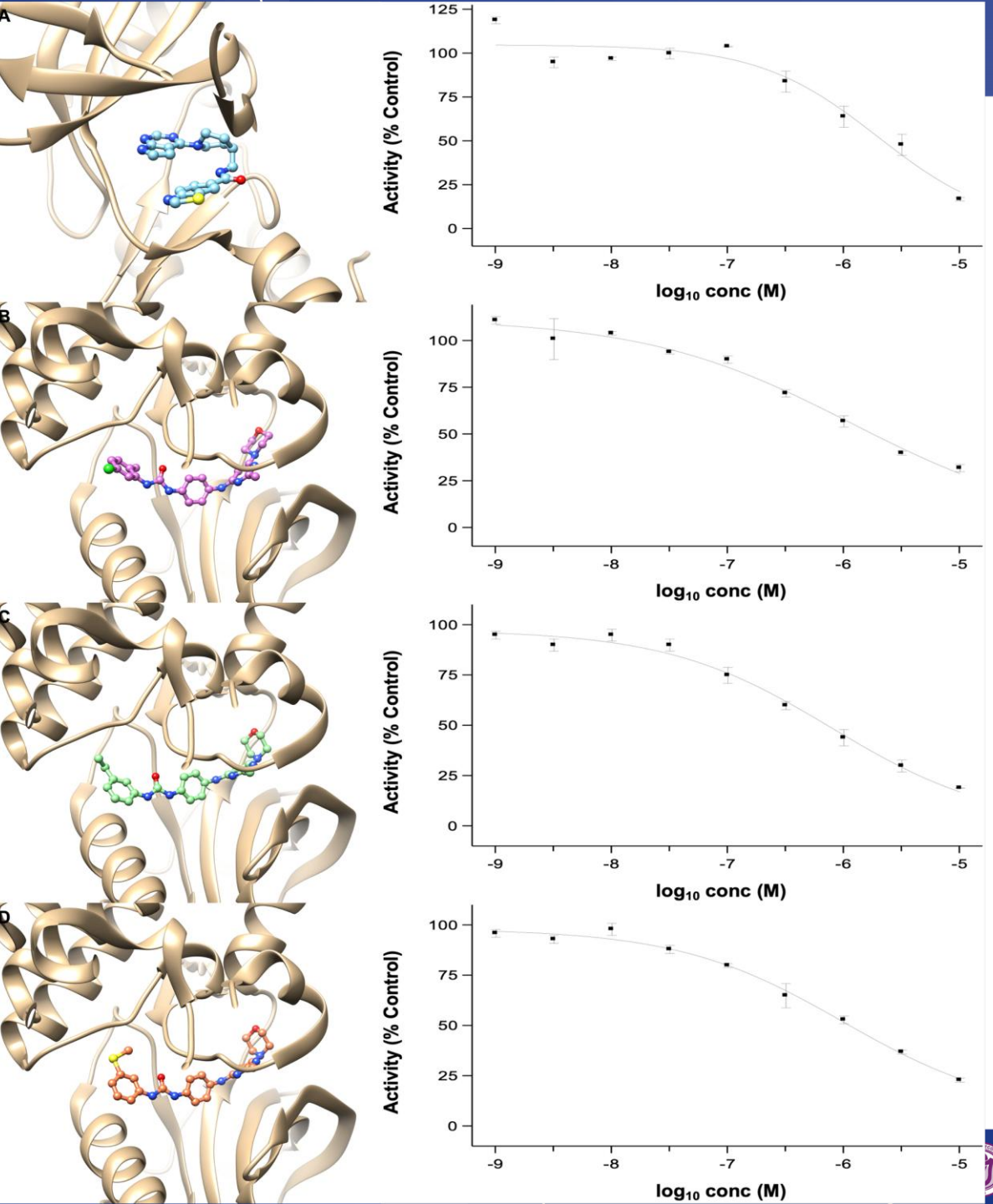
Benchmark



Average computation time: 4.2%
2616 h (full screening) → 109 h (DoM)

Average docking runs: 5.4%
5M (full screening) → 0.27M (DoM)



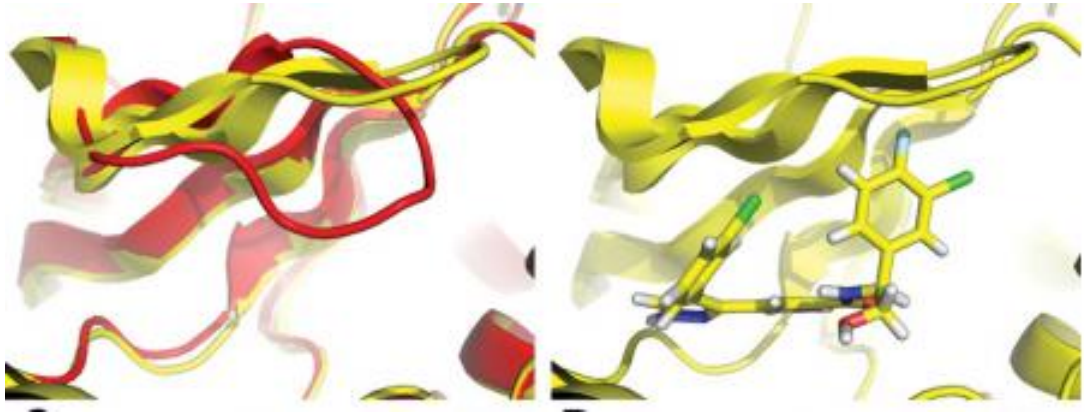


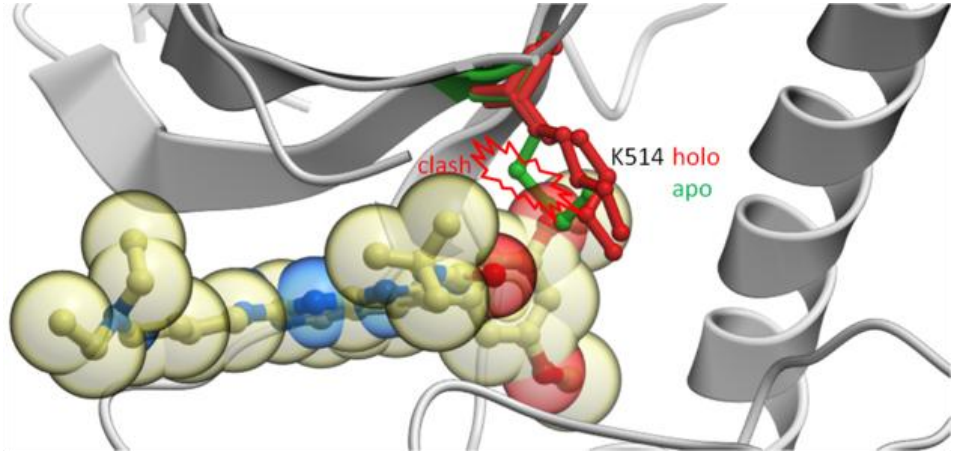
- Target: ASK1
  - 27 molecules purchased
  - 1 molecules showed activity < 50%
  - Lowest IC_{50} = 1.9 μM

- Target: DDR1
  - 31 molecules purchased
  - 4 compounds with activity < 50%
  - Lowest IC_{50} = 788 nM

Hurdles in Structure-Based Virtual Screening

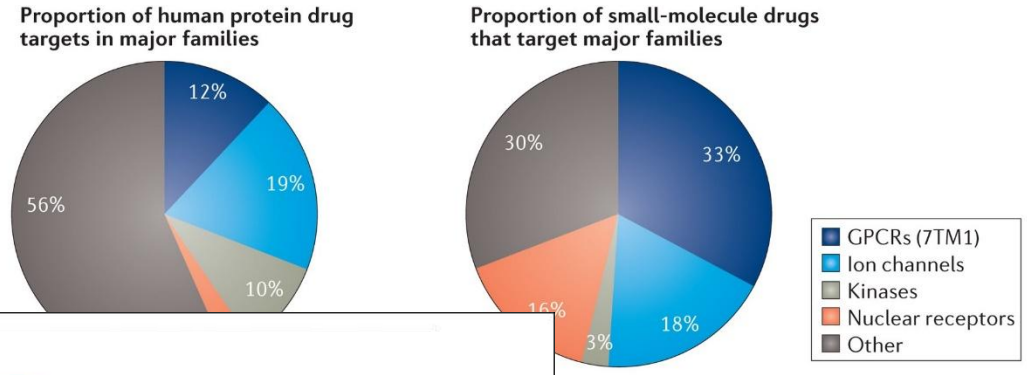
- Cannot handle conformational change of a receptor properly
  - Performance on homology modelled structure is not as high as on holo structure (Hao *et al.*, J. Chem. Inf. Model., 2009)
- Considering receptor flexibility requires more time and computational resources

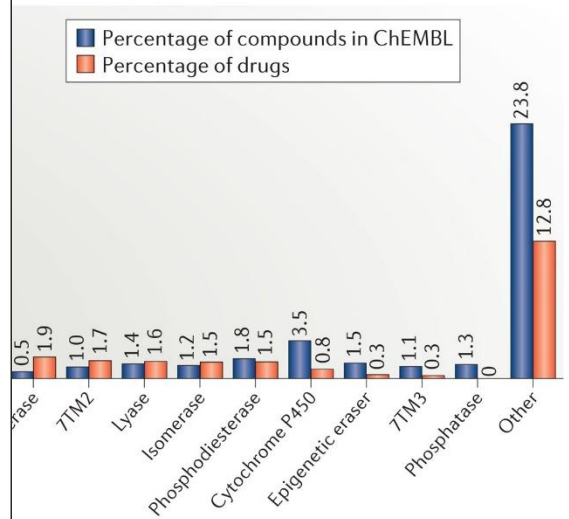


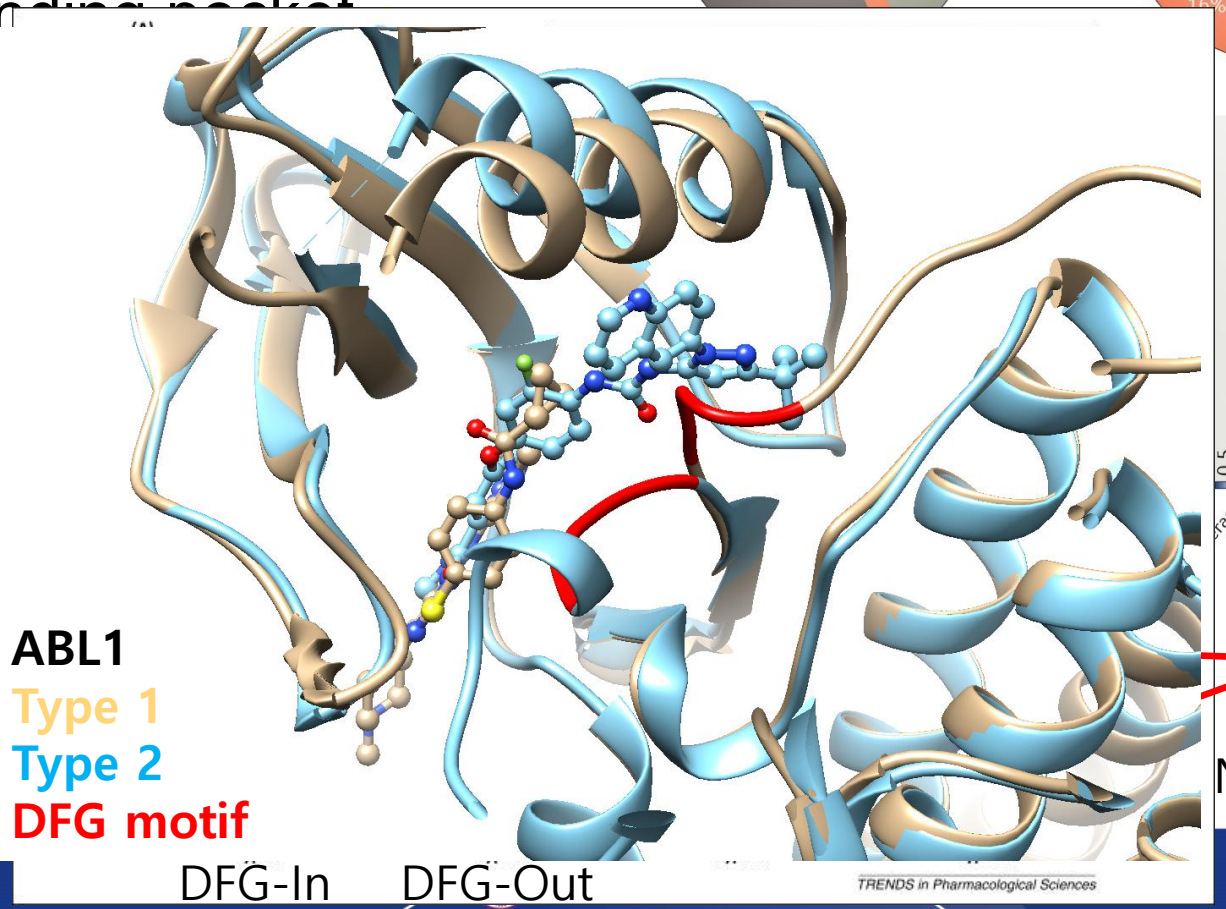


Kinase: Major Drug Targets

- 56% of targeted proteins
- Target of 30% of small molecules
- Loop flexibility in bi[illegible] pocket







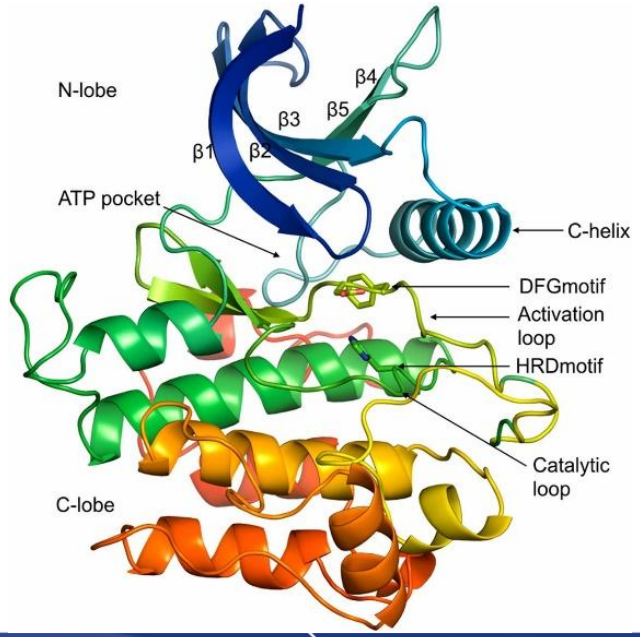
Allosteric site

Wu et al. Nat Rev Drug Discov 2017

Trends Pharmacol Sci 2015

Kinase: Structural Changes

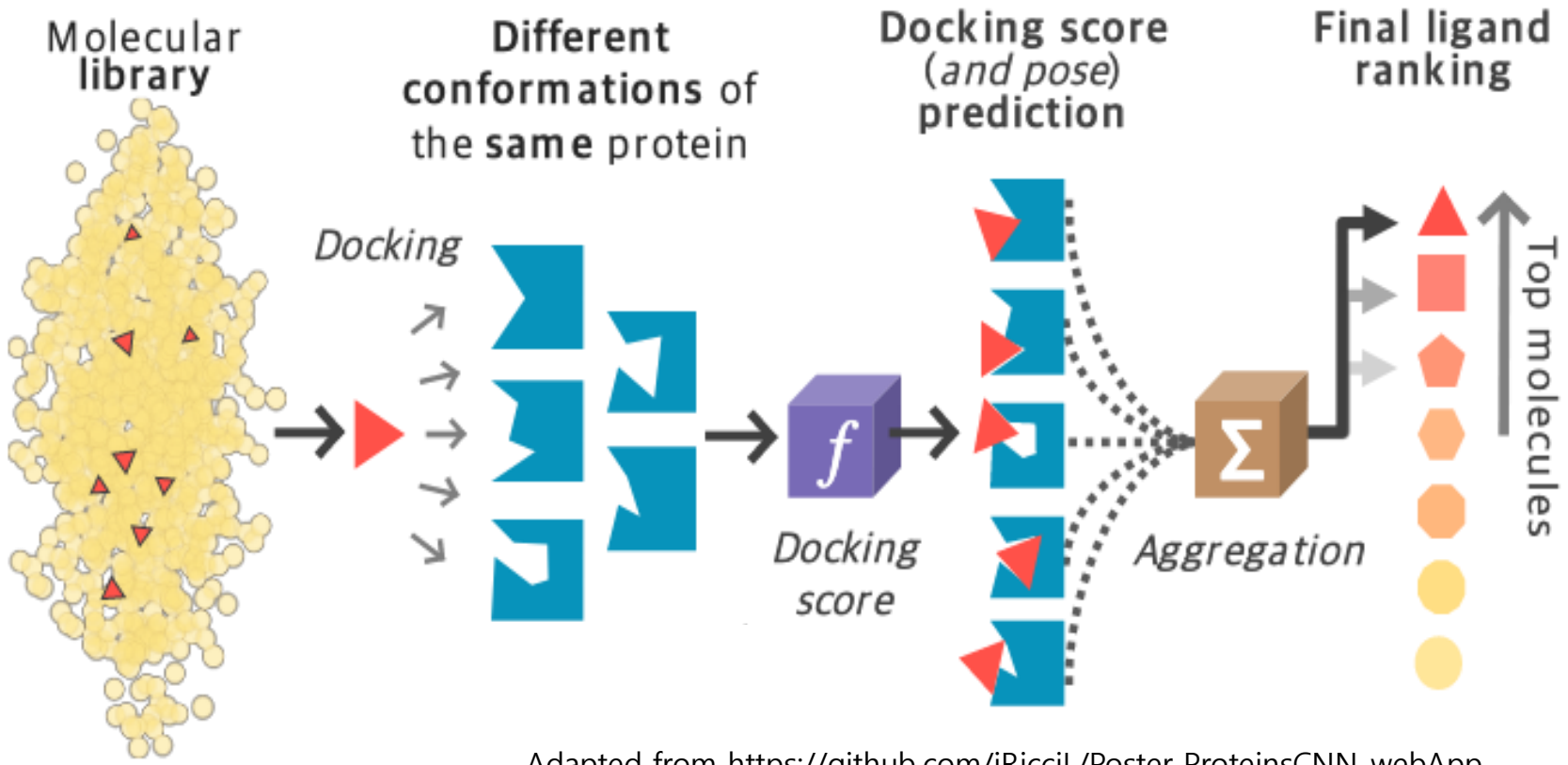
- KinCoRe (Modi and Dunbrack PNAS 2019)
  - Classifying kinase active site structure into 12 categories
  - Spatial group: DFGin, DFGinter, DFGout
  - Dihedral cluster: (φ,ψ) of X-DFG (residue before conserved Asp), DFG-Asp, DFG-Phe and side chain dihedral (χ1) of DFG-Phe.



Class	KinCoRe Notation
Activator	DFGin_BLAminus
Inhibitor	DFGin_BLBplus, DFGin_ABAminus
Type-II Binding	DFGout_BBAminus

Traditional Methods to Target Multiple Structures

- Ensemble: docking ligands to set of target protein structures

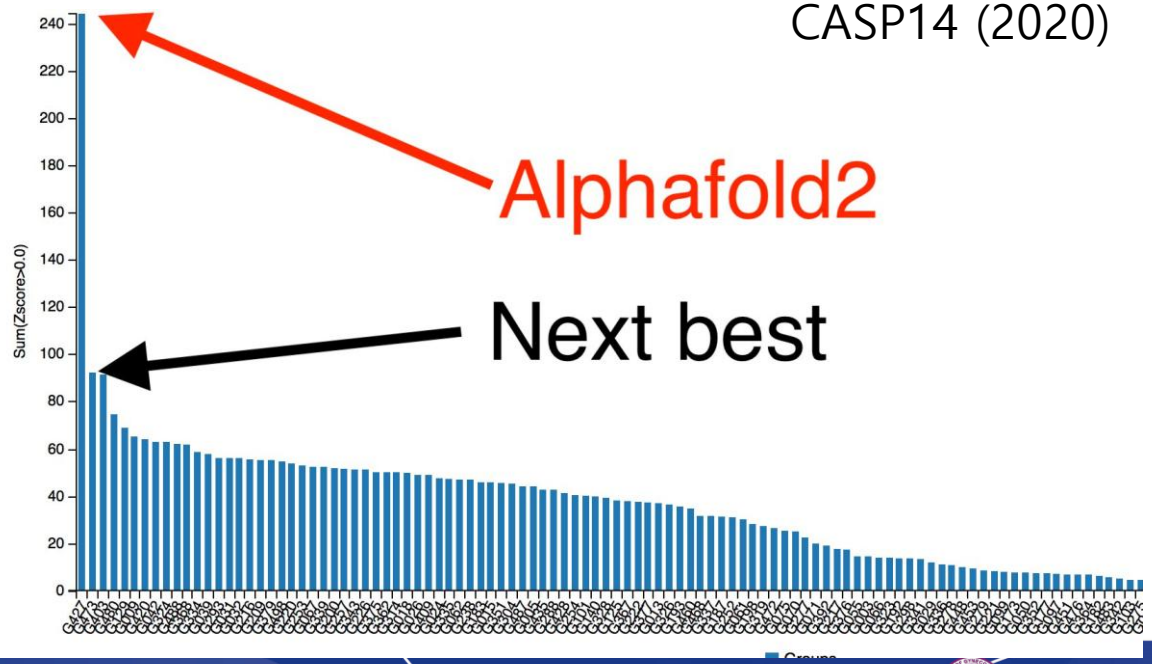


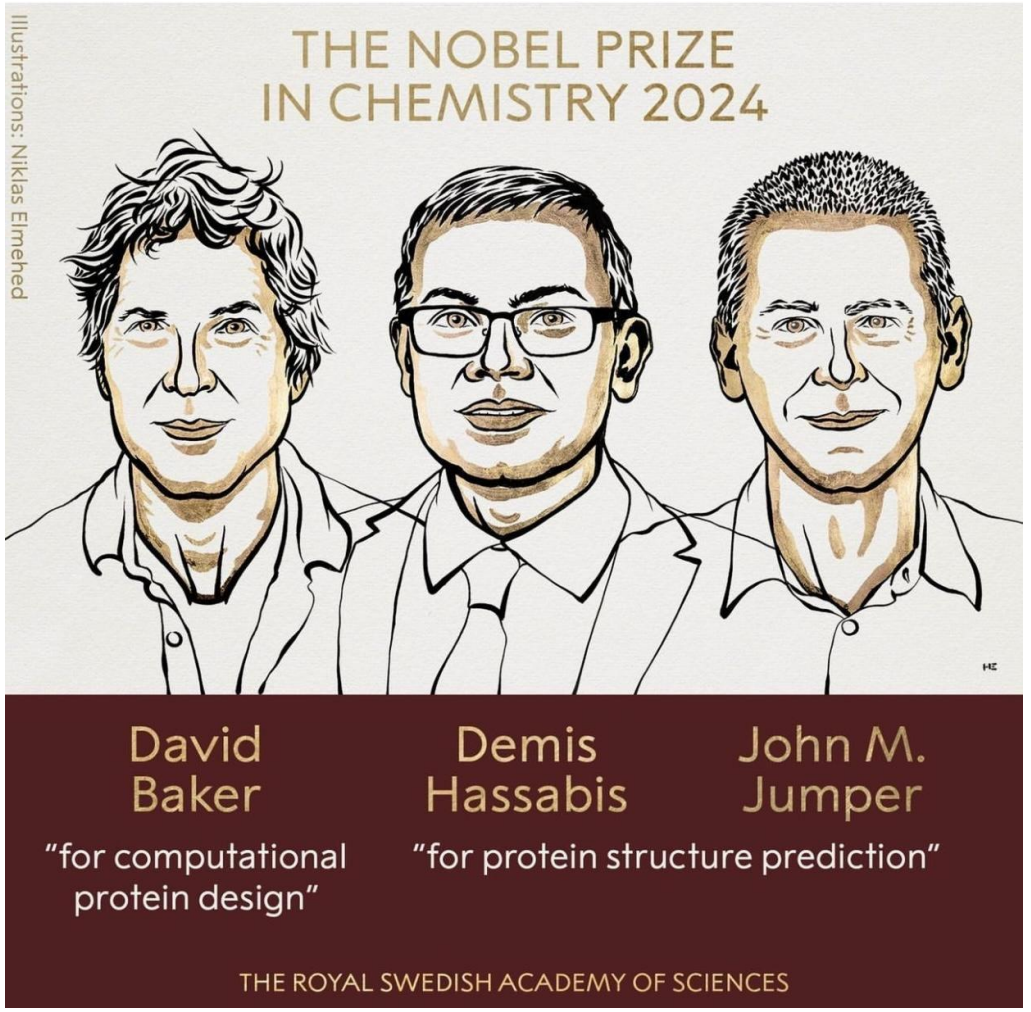
Adapted from https://github.com/jRicciL/Poster-ProteinsCNN-webApp

Deep Learning Methods to Generate Multiple Conformation

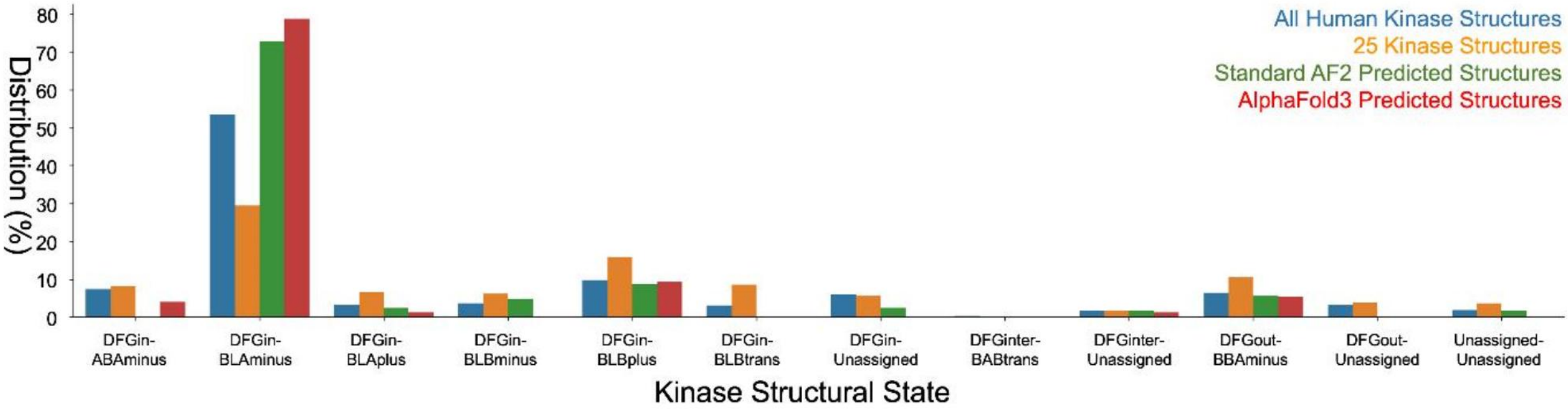
- Dominates structure prediction field
- AlphaFold3 (DeepMind)
- RoseTTAFold (D. Baker)
- ESMFold (Meta)

CASP14 (2020)





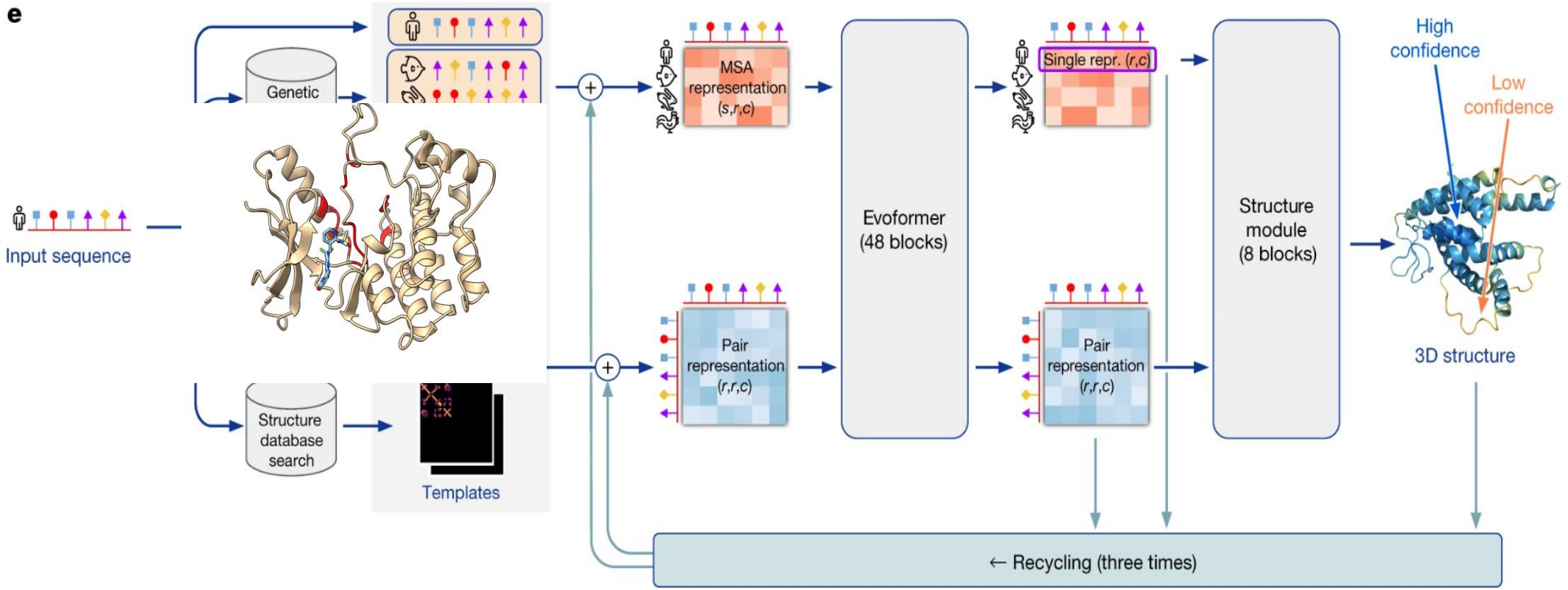
Biases in Kinase Structure Space



Song et al. Sci. Rep. (2024) 14:25167

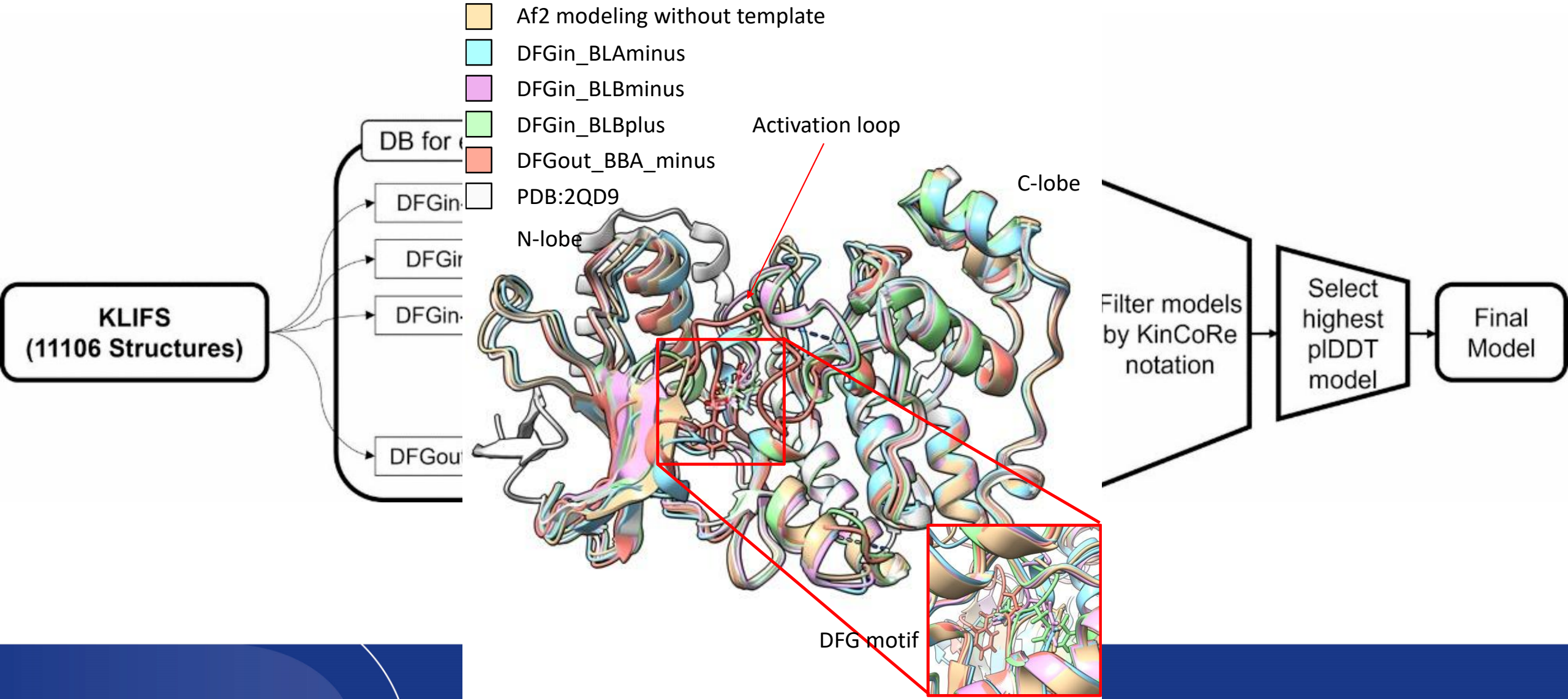
Multi-State Modeling to Get the Conformation of Interest

- Replacing MSA part to follow specific template conformation



Jumper et al. Nature (2021) 596:583-589

Multi-State Modeling to Get the Structure of Interest

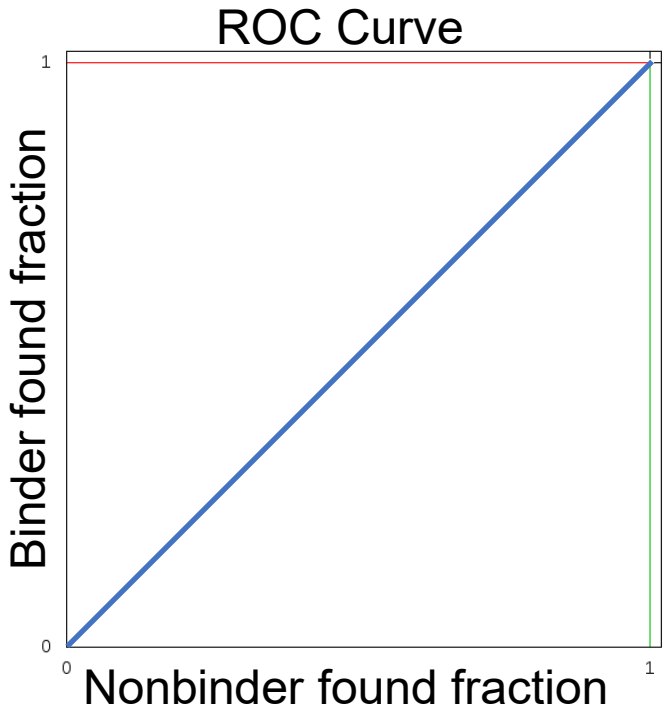


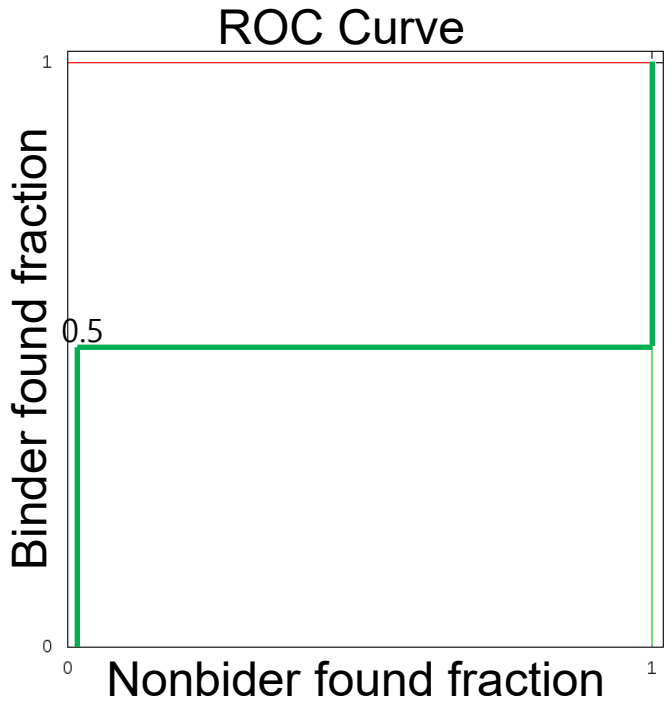
Benchmarking AF2 Multi-State Modeling

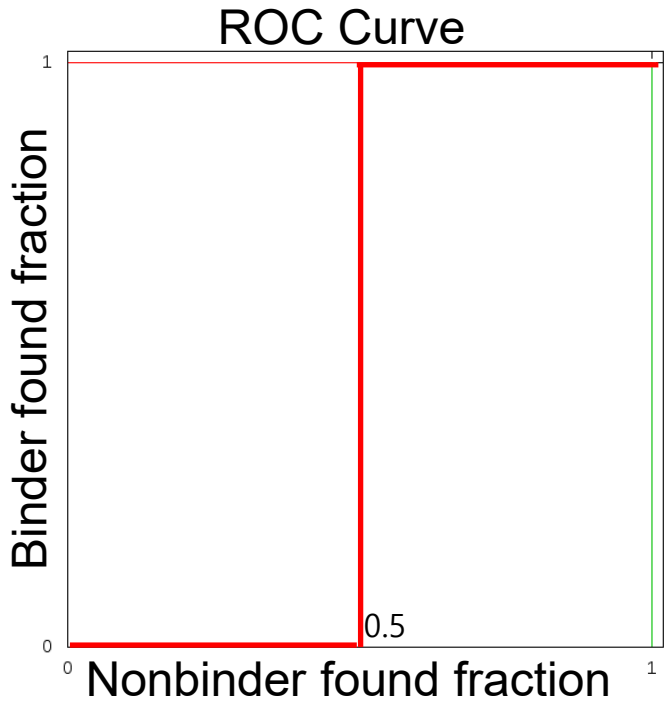
- 25 Kinase from DUD-E set
  - Protein crystal structure
  - Active compounds bind to the target kinase
  - Decoy compounds have similar chemical features to the actives but do not bind to the protein
- Docking using AutoDock-GPU
- Virtual Screening Benchmark
  - Enrichment factor
  - AUC
  - BEDROC

Screening Evaluation Metric: BEDROC

- AUC cannot discriminate three cases below
- BEDROC puts additional weight to the early recognition







MSM Virtual Screening Benchmark

- Categorizing kinases based on ligand diversity
  - Based on the dissimilarity of active compounds
  - The multi-state model performs better than the standard AF2 model when the active compounds are diverse

Diversity (# Targets)	EF 1%				AUC				BEDROC			
	MSM	AF2	AF3	X-tal	MSM	AF2	AF3	X-tal	MSM	AF2	AF3	X-tal
All Targets	**8.2**	6.6	6.9	7.8	0.661	0.659	0.664	**0.674**	0.183	0.183	0.184	**0.213**
> 0.7 (14)	**6.4**	4.6	4.8	5.3	0.648	0.641	0.647	**0.667**	0.154	0.146	0.151	**0.175**
< 0.7 (11)	10.6	9.1	9.5	**10.9**	0.678	0.681	**0.685**	0.679	0.220	0.229	0.228	**0.245**

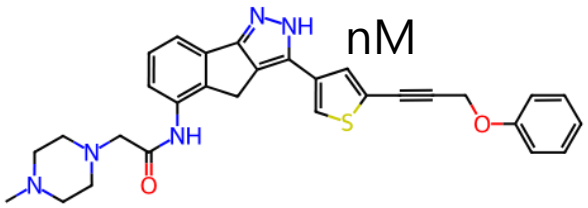
Song et al. Sci. Rep. (2024) 14:25167

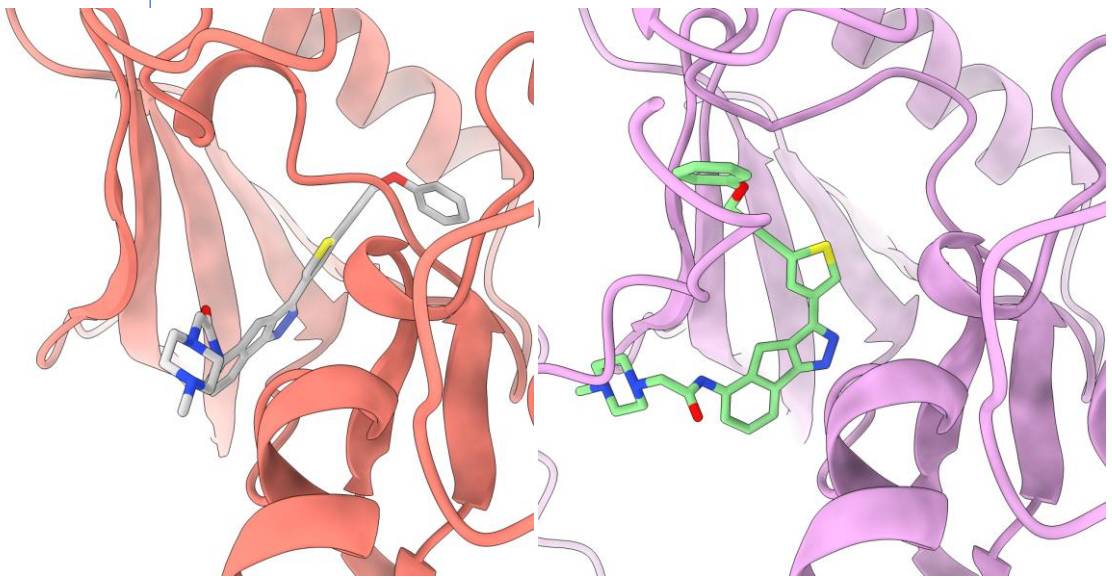
Example: CSF1R

	EF 1 %	AUC	BEDROC	Div. of Top 1 %
MSM	5.4	0.643	0.161	0.73
Standard AF	3.6	0.597	0.134	0.23

ChEMBL245377	Rank
MSM	10
Standard AF2	167

IC50: 45 nM



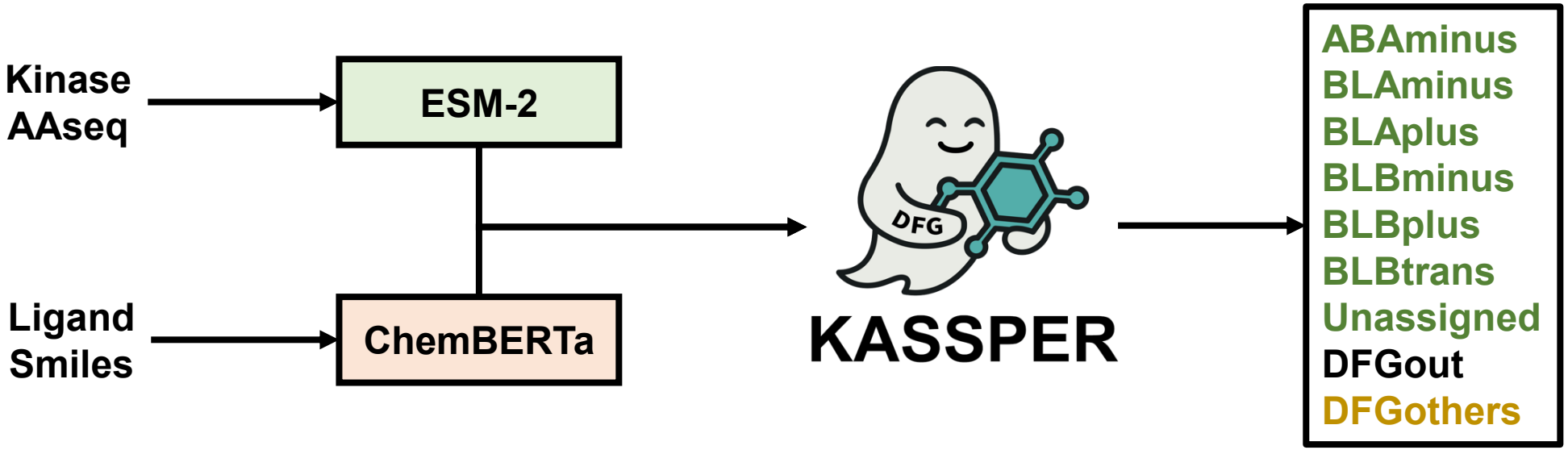


MSM Standard AF2

Song et al. Sci. Rep. (2024) 14:25167

Predicting Conformational State using Language Model

- KASSPer (Kinase Active Site Structure Predictor)



Predicting Conformational State using Language Model

- Human kinase structure dataset
  - With co-crystalized ligands
  - 7673 structures
  - Training : Validation : Test = 70 : 15 : 15

- External validation set
  - Mouse kinase structures co-crystalized with ligands
  - 303 structures

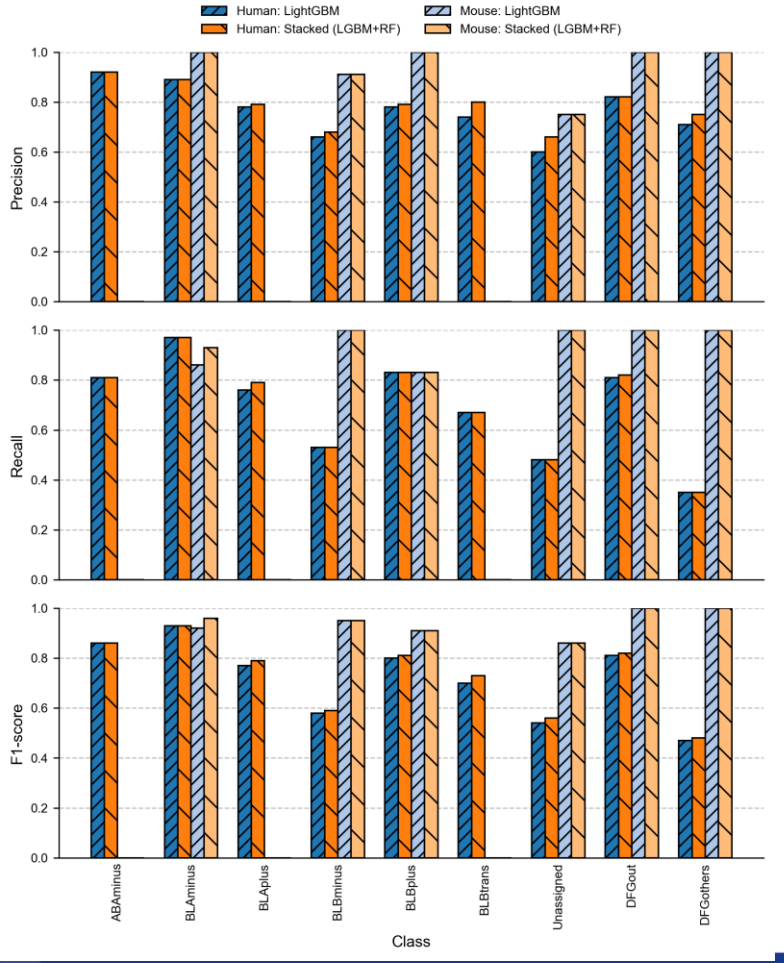
- Annotate structural state using KinCoRe

Conformational State Prediction Benchmark

- LightGBM performed best among tested models for human structures

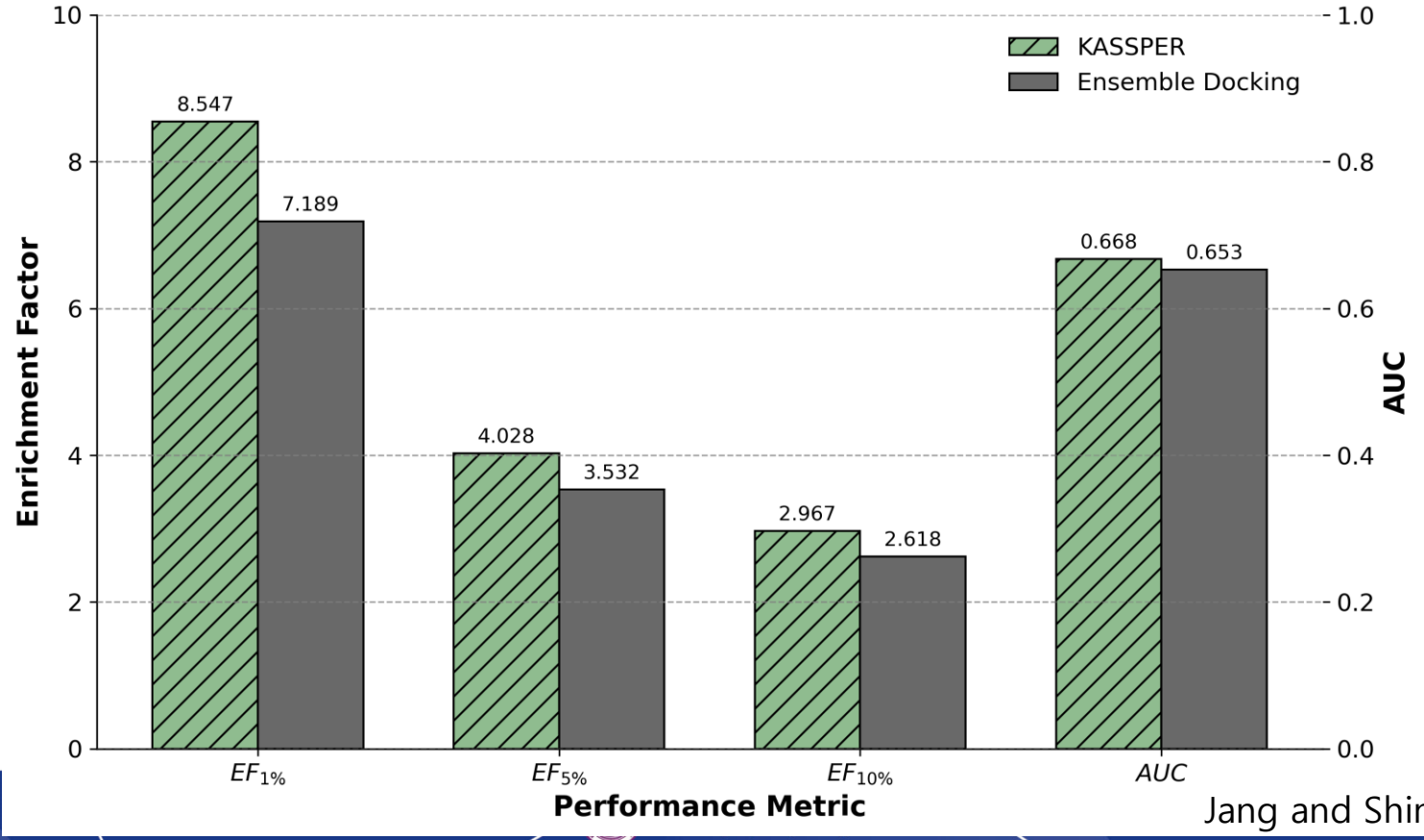
MCC: Single Model				
XGBoost	LightGBM	Random Forest	SVM	Cross Attention
0.796	**0.819**	0.798	0.511	0.771

MCC: Stacking Ensemble		
XGBoost + LightGBM	LightGBM + RF	XGBoost + RF
0.811	0.818	0.806



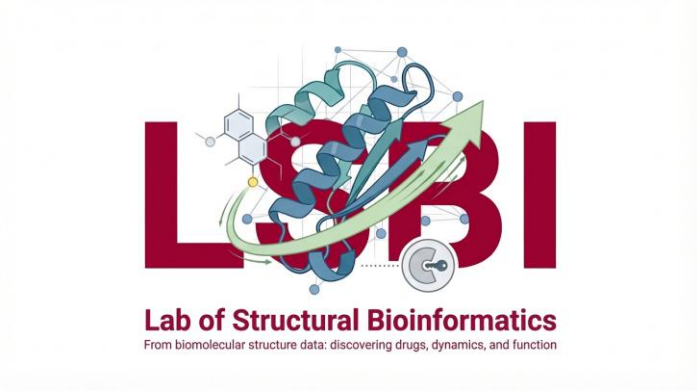
Application to Virtual Screening

- 25 kinases of DUD-E
- Predicting conformational state and docking the ligand to the state before docking improves for all metrics



Jang and Shin, Bioinformatics, In Revision

Acknowledgement



Graduate Students
Wonkyeong Jang
Jinhyeok Yoo
Seyee Kim
Minyoung Baek
Suyeon Lee
Jae Young Shin
Chaehyun Park

Collaborators

Prof. Daisuke Kihara (Purdue)
Prof. Jer-Yen Yang (China Med. U.)

Dr. Junsu Ko (Arontier)
Dr. Yiyu Hong (Arontier)
Mr. Junsu Ha (Arontier)
Prof. Juyong Lee (SNU)
Mr. Jinung Song (SNU)

Prof. Yoonjoo Choi (JNU)
Prof. Yeongju Lee (PNU)
Prof. Changha Hwang (Dankook U.)
Prof. Minseok Choi (KyungHee U.)

Prof. Hoon Kim (SCNU)
Prof. Jeong Ho Park (Hanbat U.)



2026년
대한부인종양학회
제31차 추계 심포지엄

2026.10. 23. (금) - 24. (토)

정부세종컨벤션센터 / 세종 코트야드 바이 메리어트

대한부인종양학회
Korean Society of Gynecologic Oncology