DRUG DISCOVERY
: IRAK4 Ic50 ACTIVITY PREDICTION

내일은 제약왕팀
강성준 김성은 신주환 양은정 장은지

Contents

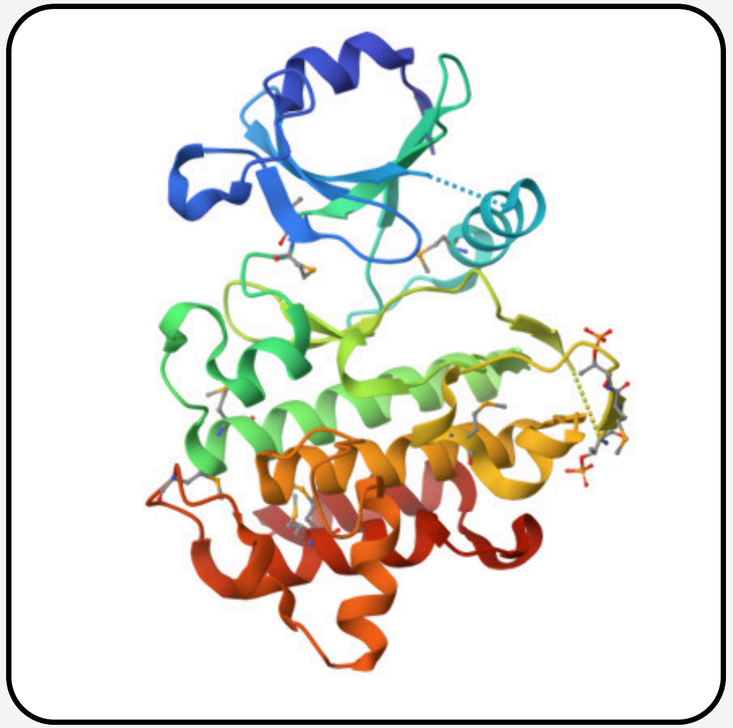
01 프로젝트 목표

02 데이터 수집 & 피쳐 추출
: Chemical & Physical & Statistical Features

03 모델 & 하이퍼파라미터
: MPNN, GATv2, XGboost

04 활용성
: Protein-Ligand Interaction & Mechanism study

Project Goal



IRAK4 kinase domain

화합물의 IRAK4 IC50 예측

Columns

물리화학적
특징

3D 구조

기존 화합물
유사도

Docking
Score

2D 구조
(그래프 임베딩)

Columns

물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

Columns

물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

물리 화학적 특징

Exact Mass
XLogP
Hydrogen Bond Donors
Hydrogen Bond Acceptors
Rotatable Bonds
TPSA
Heavy Atom Count

Complexity by RDkit
Fraction of sp3 Carbons by RDkit
Ring Count by RDkit
Aromatic Ring Count by RDkit
Heteroatom Count by RDkit
BalabanJ Index by RDkit
Covalently-Bonded Unit Count

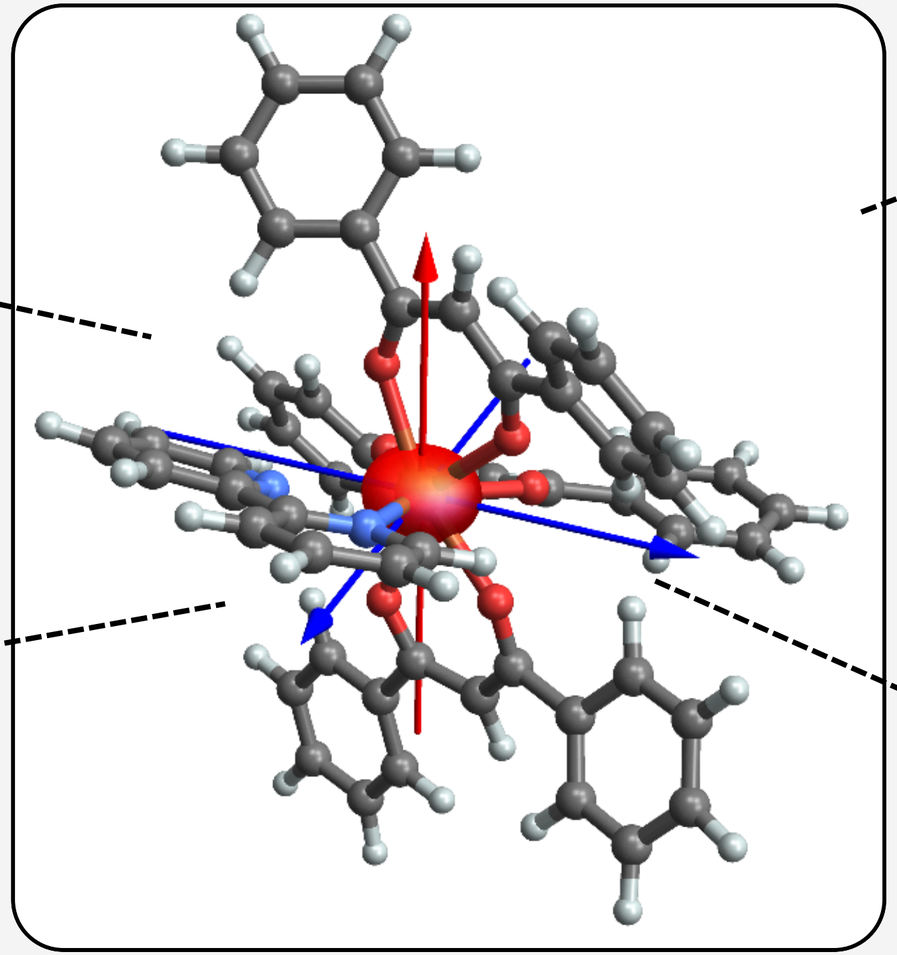
RDKit
Open-Source Cheminformatics and Machine Learning

PubChem

Columns

물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

3D 구조



Asphericity
(비구형도)

Spherocity
(구형성)

PMI
(주관성 모먼트)

NPR
(PMI 정규화)

Eccentricity
(이심률)

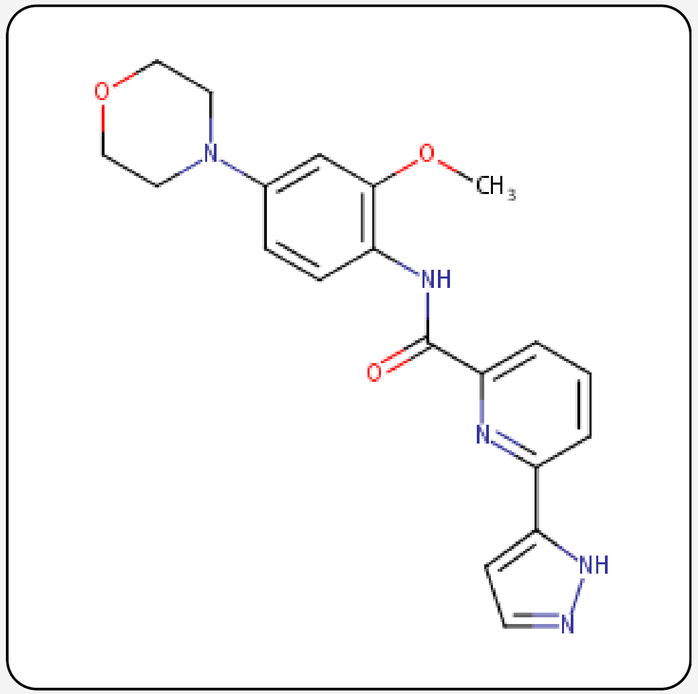
Columns

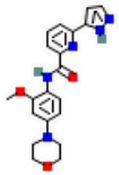
물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

기존 화합물 유사도

Tanimoto 유사도 (Tanimoto Similarity)

COC1=C(C=CC(=C1)N2CCOCC2)NC(=O)C3=CC=CC(=N3)C4=CC=NN4





CHEMBL256713 IRAK4 inhibitor 4b [PMID: 18474425]; N-[2-methoxy-4-(morpholin-4-yl)phenyl]-6-(1H-pyrazol-5-yl)pyridine-2-carboxamide; 1042673-20-0; IRAK4 inhibitor 4b; ...

Compound CID: 44449334

MF: $C_{20}H_{21}N_5O_3$ MW: 379.4g/mol

IUPAC Name: N-(2-methoxy-4-morpholin-4-ylphenyl)-6-(1H-pyrazol-5-yl)pyridine-2-carboxamide

Isomeric SMILES: COC1=C(C=CC(=C1)N2CCOCC2)NC(=O)C3=CC=CC(=N3)C4=CC=NN4

InChIKey: RAFFLDOJXQAJPF-UHFFFAOYSA-N

InChI: InChI=1S/C20H21N5O3/c1-27-19-13-14(25-9-11-28-12-10-25)5-6-17(19)23-20(26)18-4-2-3-15(22-18)16-7-8-21-24-16/h2-8,13H,9-12H2,1H3,(H,21,24)(H,23,26)

Create Date: 2009-11-19

Columns

물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

Docking Score

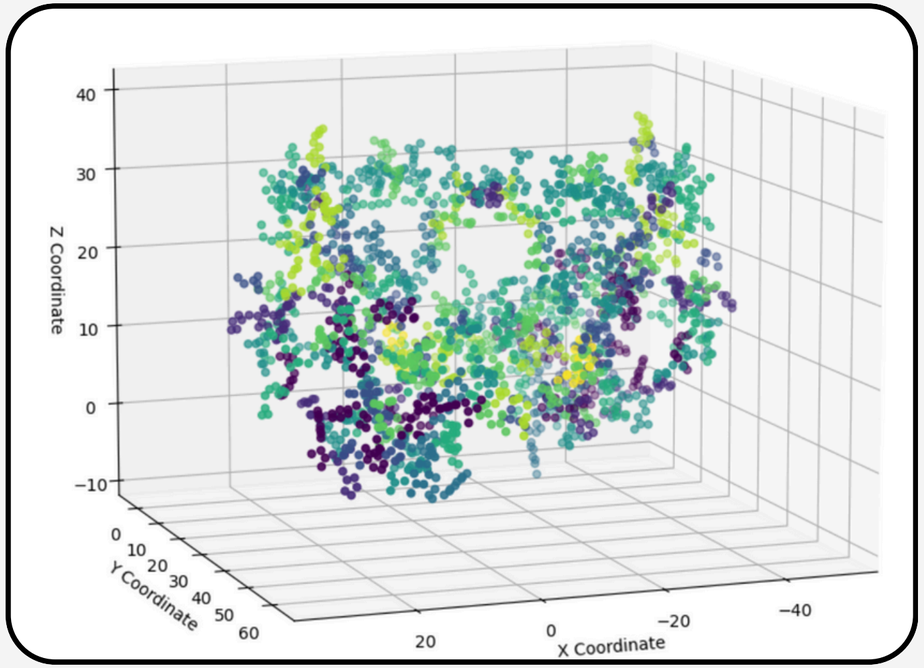
https://github.com/DeltaGroupNJUPT/Vina-GPU-2.0

Autodockvina

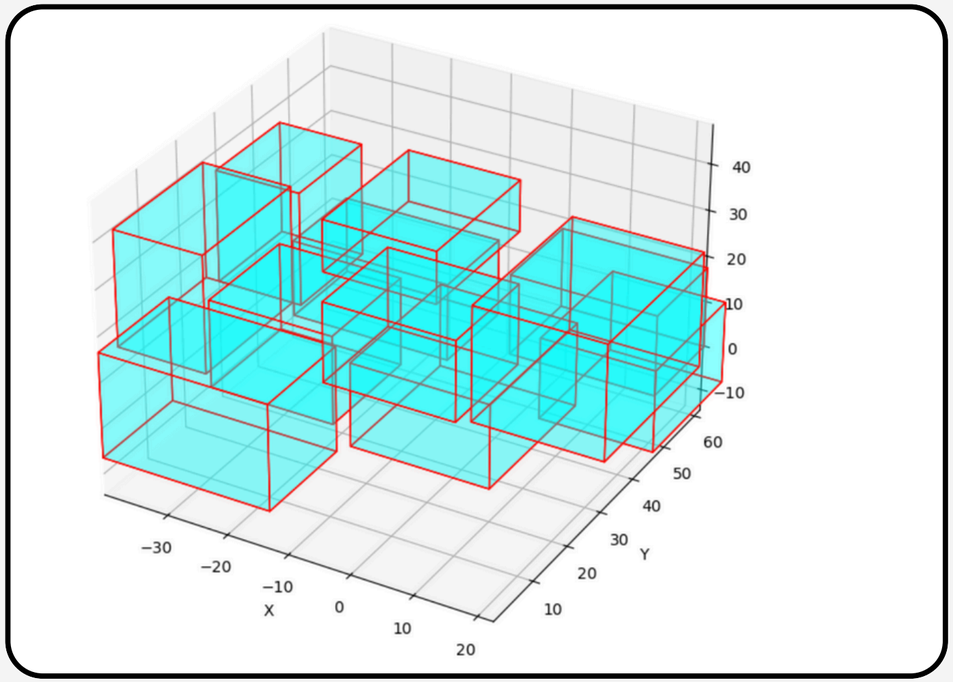
결합에 중요한 residue 별로 Cluster

11곳의 결합 부위(site) 선정

	Residue	Atom	Chain	Residue Number	X	Y	Z
0	PHE	N	A	165	27.431	46.212	2.758
1	PHE	CA	A	165	26.233	46.966	3.235
2	PHE	C	A	165	26.087	48.330	2.568
3	PHE	O	A	165	26.810	48.640	1.625
4	PHE	CB	A	165	26.280	47.125	4.758
...	...	...	...	...	...	...	...
2147	GLU	CB	B	456	-30.210	50.463	11.256
2148	GLU	CG	B	456	-31.684	50.814	11.055
2149	GLU	CD	B	456	-32.437	50.990	12.371
2150	GLU	OE1	B	456	-31.868	51.569	13.323
2151	GLU	OE2	B	456	-33.606	50.555	12.447



사용 잔기: S, D, E, W, Y, H, K, R, F



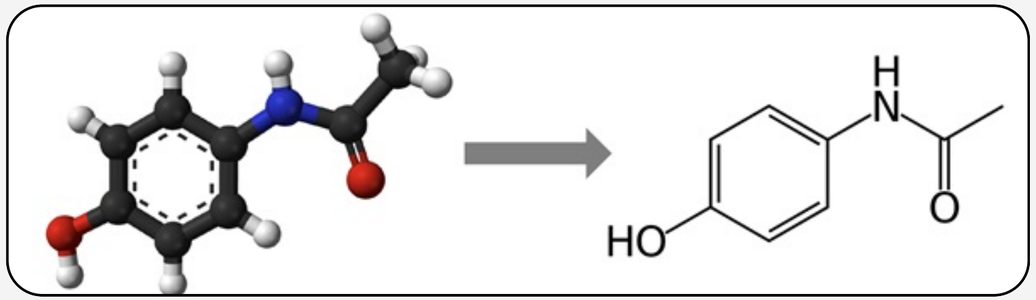
<3D clusters and Residue boxes>

Columns

Stacking !

물리 화학적 특징	3D 구조	기존 화합물 유사도	Docking Score	2D 구조 (Graph)
Exact Mass XLogP . . . TPSA	Asphericity Eccentricity Spherocity PMI NPR1 NPR2	Similarity	Cluster_0 Cluster_1 . . . Cluster_8	mpnn_predict (model ensemble)

2D 구조
(Graph)

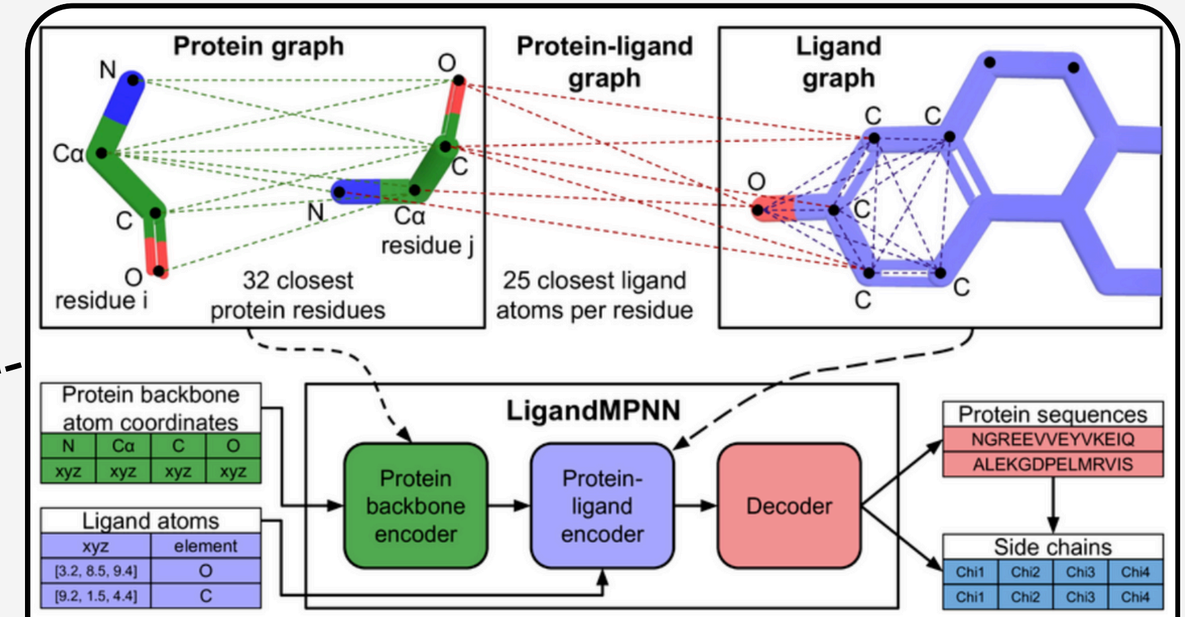


GATv2	vs	MPNN
노드 간 중요도 반영		메시지 전달 구조
복잡한 관계 학습		상호작용 학습
<u>대규모 데이터 요구</u>		<u>일반화 가능성 높음</u>

Model

“MPNN”

“XGboost”

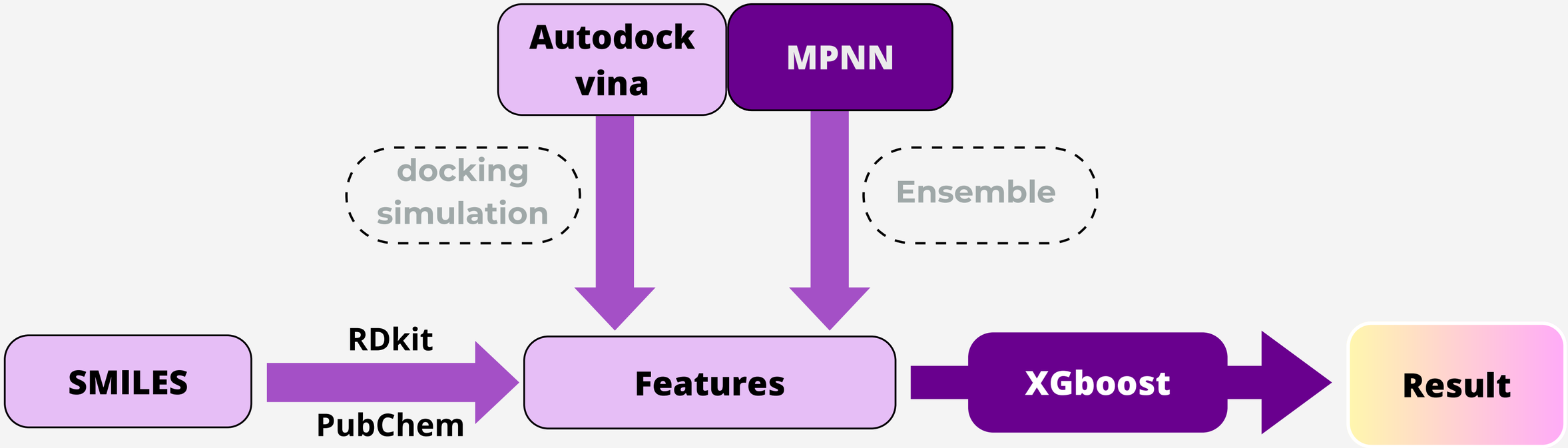


The LigandMPNN model. LigandMPNN operates on three different graphs. First, a protein-only graph with residues as nodes and 25 distances between N, Cα, C, O, and virtual

OPTUNA

hyperparameter searching

Overall process



Generality

Protein-Ligand 상호작용에 범용적 적용 가능

Kinase → Other proteins

Ligands → Other Chemicals

생리적 메커니즘 발견 가능

분포가 다른 Train & Test 학습 및 예측 가능

THANK YOU