

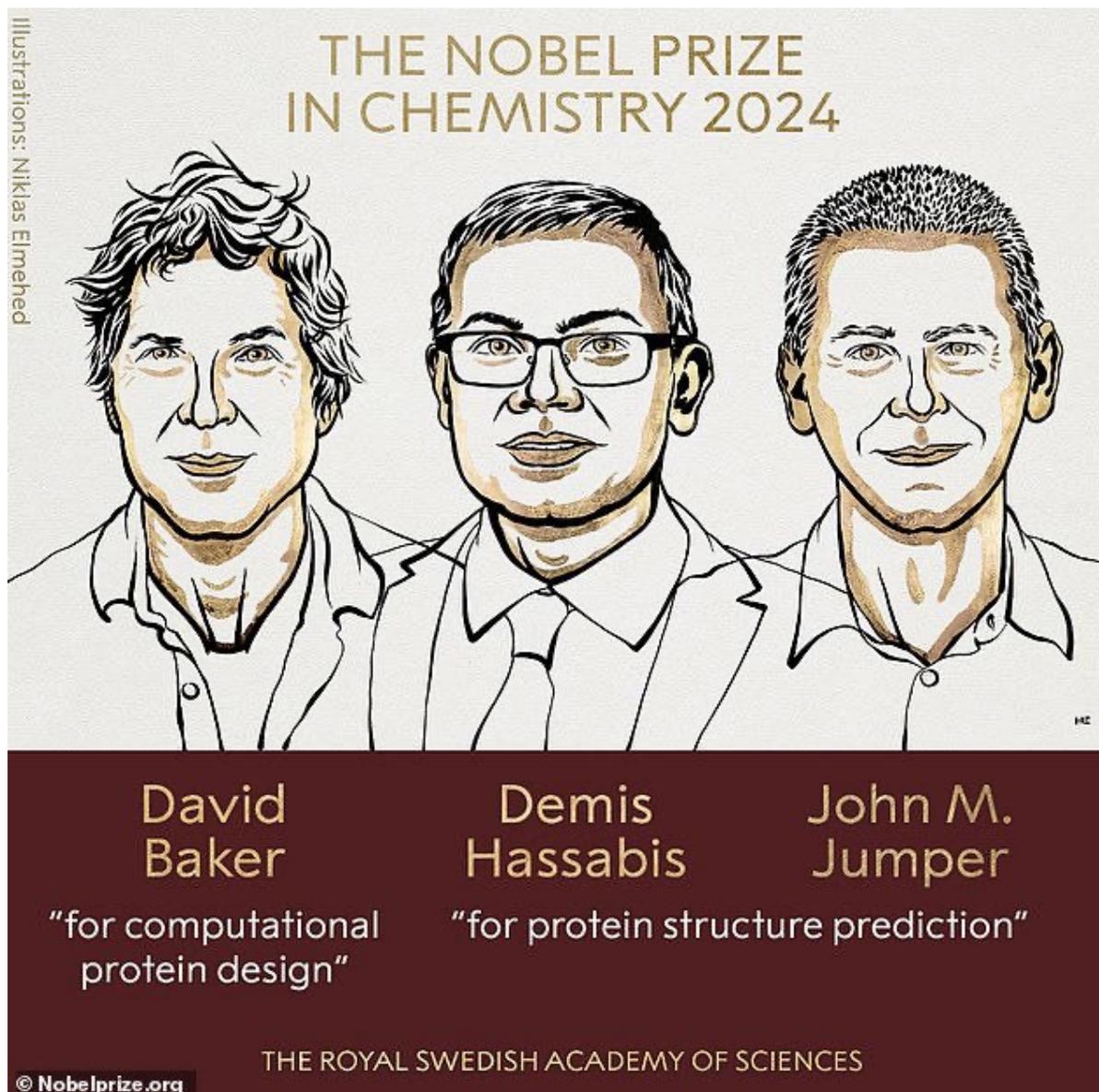
AI in Drug Discovery and Development



AI 신약팀, 대웅제약
신승우 (swshin017@daewoong.co.kr)
2024. 12. 03.



Nobel Prize in Chemistry 2024



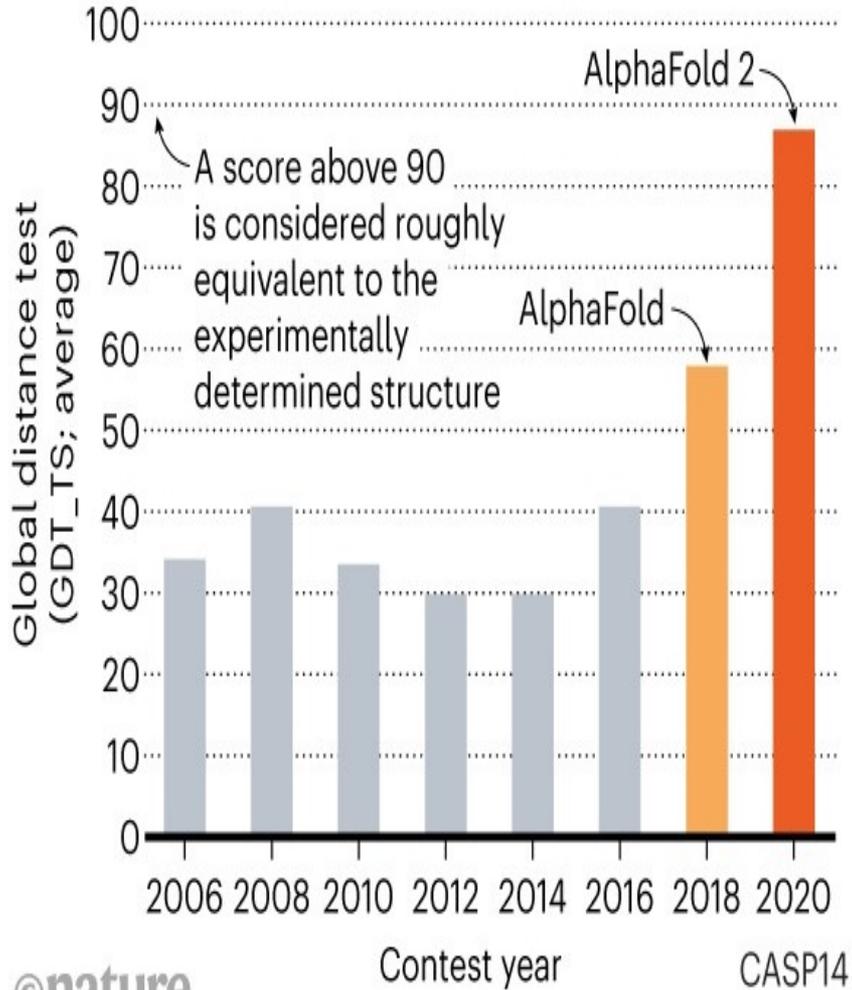
What is artificial intelligence?

Artificial Intelligence is the ability of a computer to perform tasks commonly associated with intelligent beings.

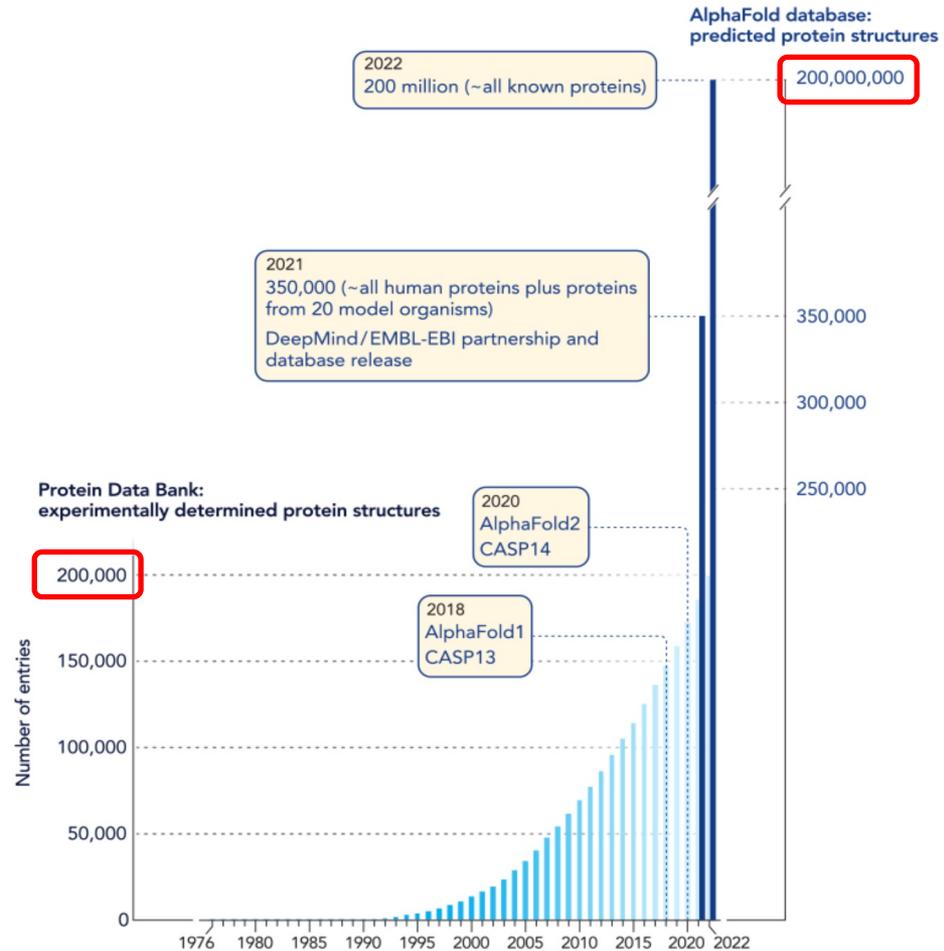
A AI role and performance

- **AI technology** is important because it enables **human capabilities** (understanding, reasoning, planning, communication and perception) to be undertaken by software increasingly **effectively, efficiently and at low cost**.
- General analytical tasks, including **finding patterns** in data, that have been performed by software for many years can also be **performed more effectively using AI**.

Example: AI power in protein structure prediction



©nature



<http://innovaformazione.net/alphafold-intelligenza-artificiale>

Paradigm shift in drug discovery

- **COMPUTER-AIDED DRUG DISCOVERY AND DEVELOPMENT (CADD D): *in silico*-chemico-biological approach (cited: 849)**

Kapetanovic IM. Computer-aided drug discovery and development (CADD): *in silico*-chemico-biological approach. Chem Biol Interact. 2008 Jan 30;171(2):165-76. doi: 10.1016/j.cbi.2006.12.006. Epub 2006 Dec 16. PMID: 17229415; PMCID: PMC2253724.

- **Computer-Aided Drug Design(CADD) Methods (cited: 580)**

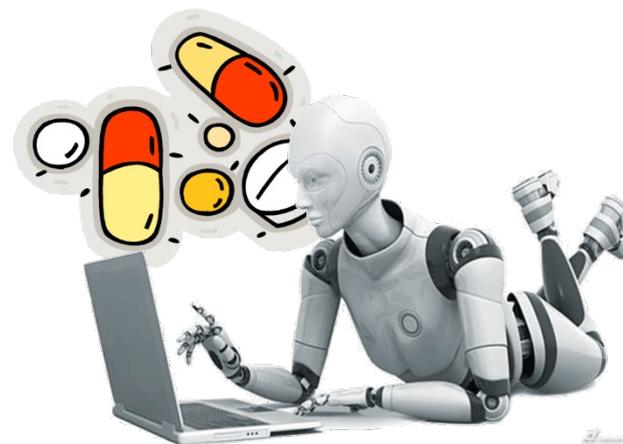
Yu W, MacKerell AD Jr. Computer-Aided Drug Design Methods. Methods Mol Biol. 2017;1520:85-106. doi: 10.1007/978-1-4939-6634-9_5. PMID: 27873247; PMCID: PMC5248982.

- **Artificial intelligence in drug discovery and development (cited: 1263)**

Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021 Jan;26(1):80-93. doi: 10.1016/j.drudis.2020.10.010. Epub 2020 Oct 21. PMID: 33099022; PMCID: PMC7577280.

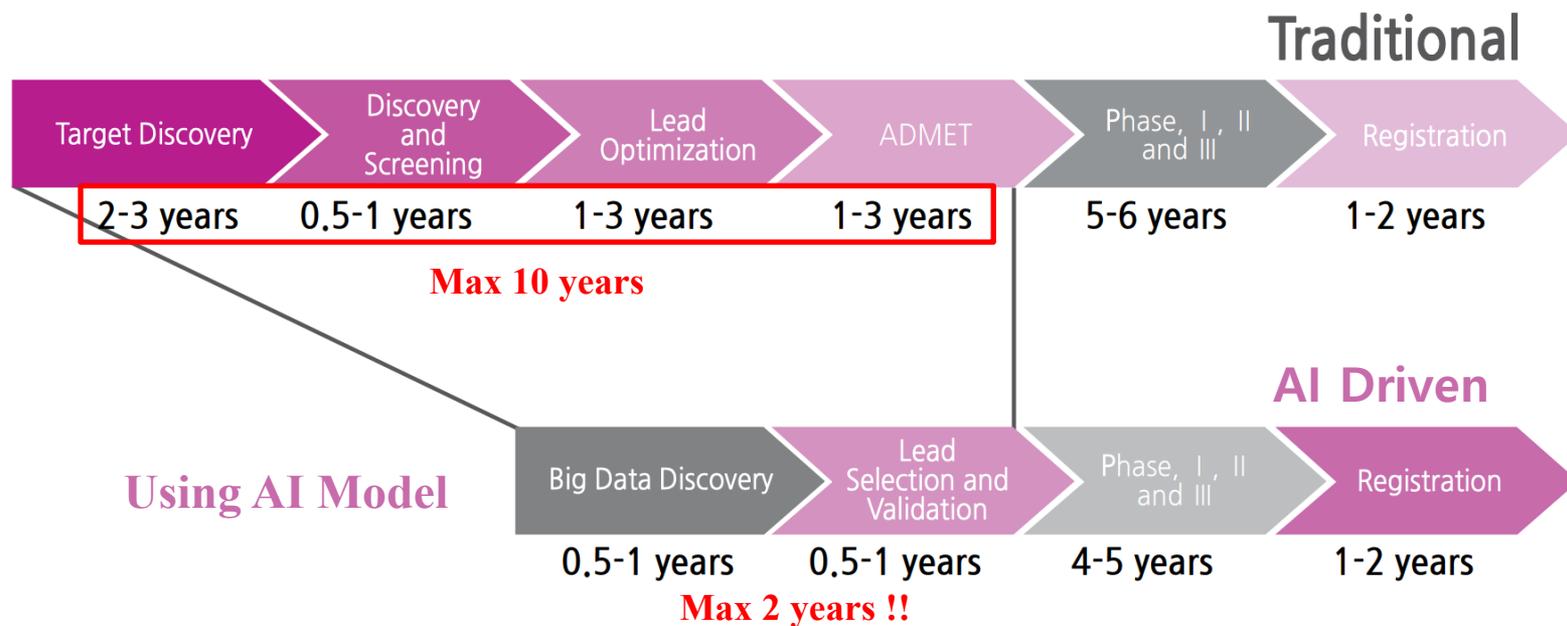


CADD



AADD

Why AI in New drug discovery ?



Cost : 2.2 billion (2~3조) → 500 million (6.5천억)

Time: Max 10 years → Max 2 years

And ?

<https://www.nature.com/articles/s41587-019-0224-x#:~:text=02%20September%202019-,Deep%20learning%20enables%20rapid%20identification%20of%20potent%20DDR1%20kinase%20inhibitors,-Alex%20Zavoronkov%2C>

Leading pharmaceutical companies and their association with AI Company



New drug discovery using AI : Ex. 1

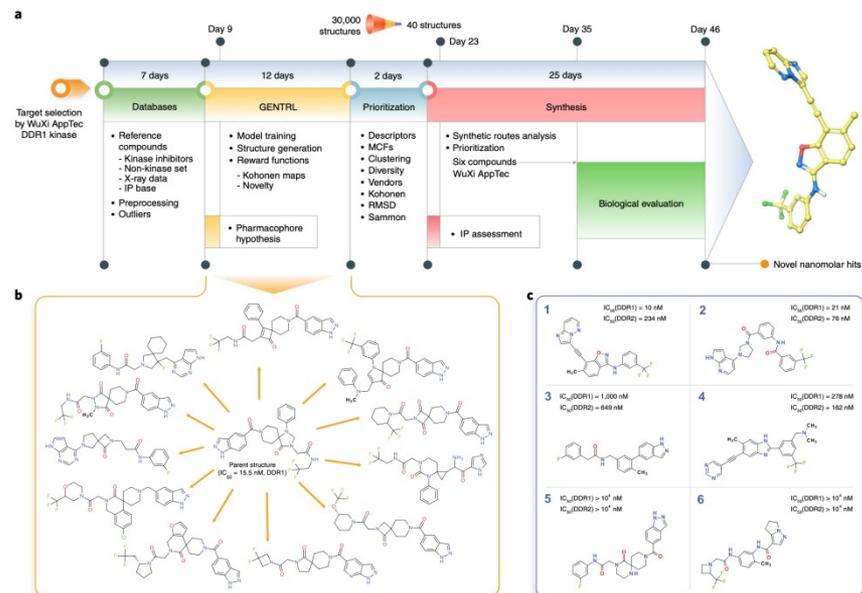
BRIEF COMMUNICATION

<https://doi.org/10.1038/s41587-019-0224-x>

nature
biotechnology

Deep learning enables rapid identification of potent DDR1 kinase inhibitors

Alex Zhavoronkov^{1*}, Yan A. Ivanenkov¹, Alex Aliper¹, Mark S. Veselov¹, Vladimir A. Aladinskiy¹, Anastasiya V. Aladinskaya¹, Victor A. Terentiev¹, Daniil A. Polykovskiy¹, Maksim D. Kuznetsov¹, Arip Asadulaev¹, Yury Volkov¹, Artem Zholus¹, Rim R. Shayakhmetov¹, Alexander Zhebrak¹, Lidiya I. Minaeva¹, Bogdan A. Zagribelnyy¹, Lennart H. Lee², Richard Soll², David Madge², Li Xing², Tao Guo² and Alán Aspuru-Guzik^{3,4,5,6}



Ex.:10 to 15 years for a drug, → **INSILICO MEDICINE**, 46 days to in vitro

<https://www.nature.com/articles/s41587-019-0224-x>

New drug discovery using AI : Ex. 2

< 주요 AI 설계 약물의 임상시험 현황 >

후보물질	개발사	임상단계	적응증
REC-2282	Recursion	2/3(4Q24결과)	신경섬유종증 제2형
REC-994	Recursion	2(3Q24결과)	대뇌 해면상 기형
REC-4881	Recursion	2(1Q25결과)	가족성 선종성 폴립증
INS018_055	Insilico Medicine	2	특발성 폐섬유화증
BEN-2293	BenevolentAI	2a 실패	아토피 피부염
EXS-21546	Exscientia	2a 실패	R/R 신세포암, 비소세포폐암
RLY-4008	Relay therapeutics	2	FGFR2 과발현 담관암
EXS-4318	Exscientia	1	염증성 자가면역 질환
BEN-8744	BenevolentAI	1	궤양성 대장염
REC-3599	Recursion	1	GM2 강글리오사이드증
REC-3964	Recursion	1상 종료	클로스트리디움 디피실 장염
미정	Recursion	전임상	HRD-음성 난소암

End-to-End 생성형 AI가 설계한 약물의 임상 2상 결과가 곧 발표될 예정 (2024)

Nature Medicine, Inside the nascent industry of AI-designed drugs, 2023.6, Nature Biotechnology, A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models, 2024.3, LG 경영연구원, AI 신약개발 선도기업 리커전의 성공전략, 2024.7

New drug discovery process

Human in the loop !!

1. 가상탐색(VS)

- 8억 ligands library
- AI 기반 가상탐색
- AI을 이용한 ligand 생성

4. ADME/T 예측

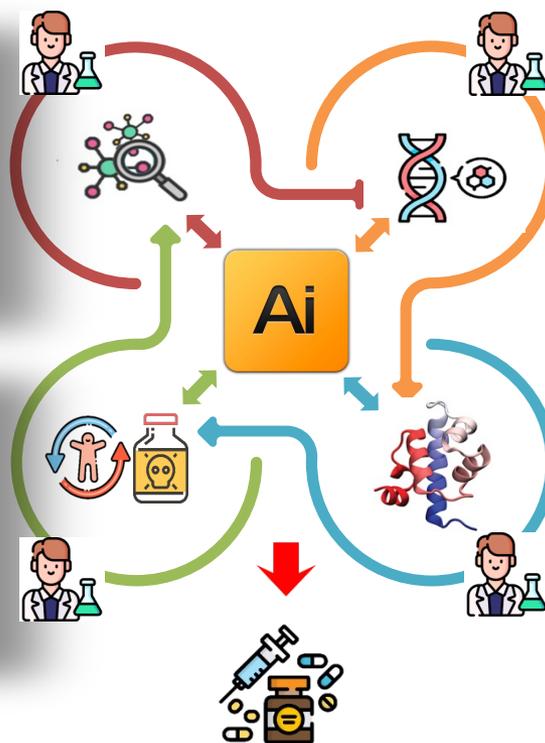
- 약물의 독성예측
- 흡수/퍼짐/대사/배출 예측

2. 도킹 시뮬레이션

- 바인딩 가능성 확인
- 최적 결합 Pose 도출
- Hit discovery

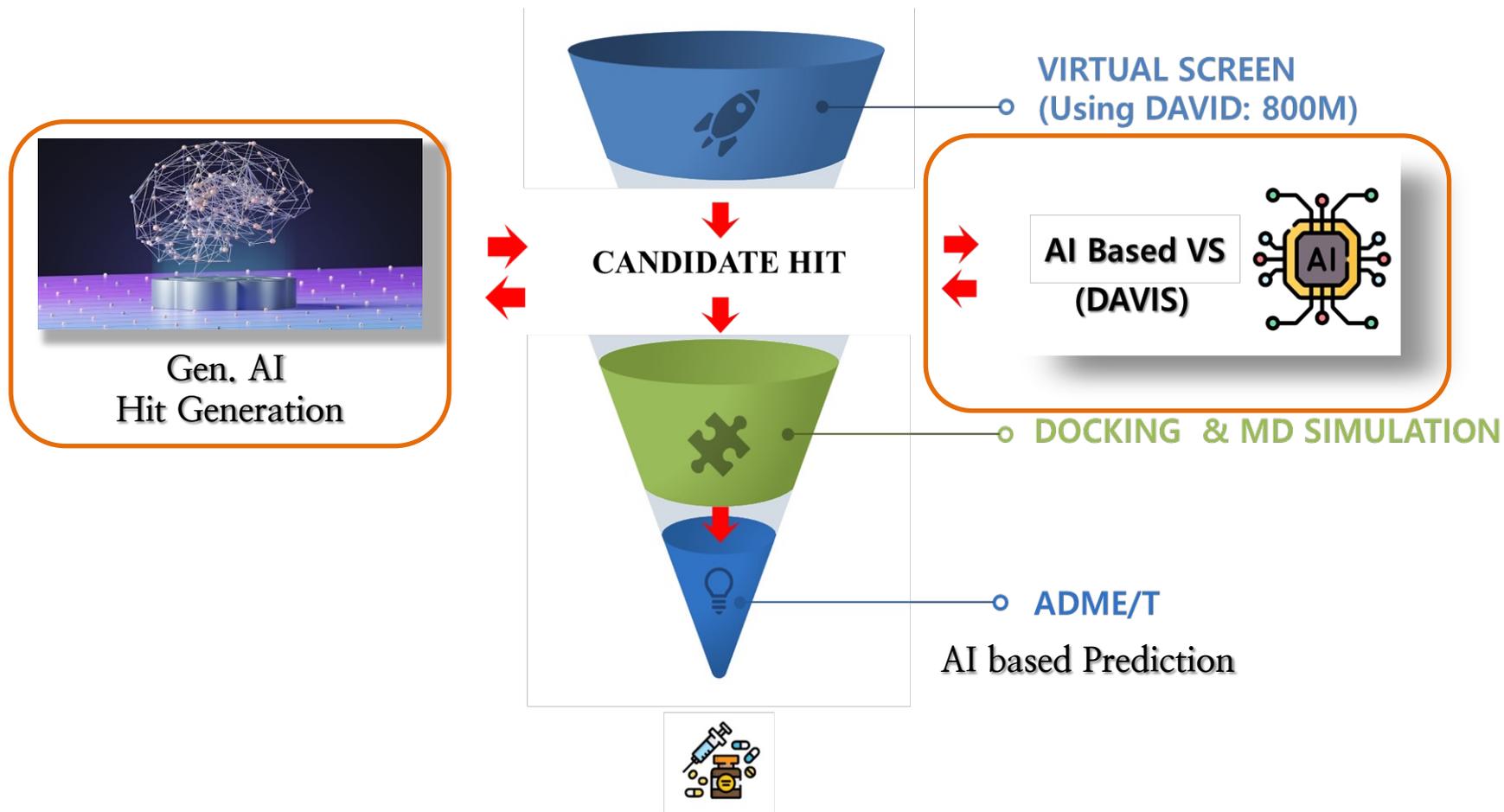
3. 분자 동역학

- Quantum MD Simulation
- Drug의 바인딩 안정성 검증



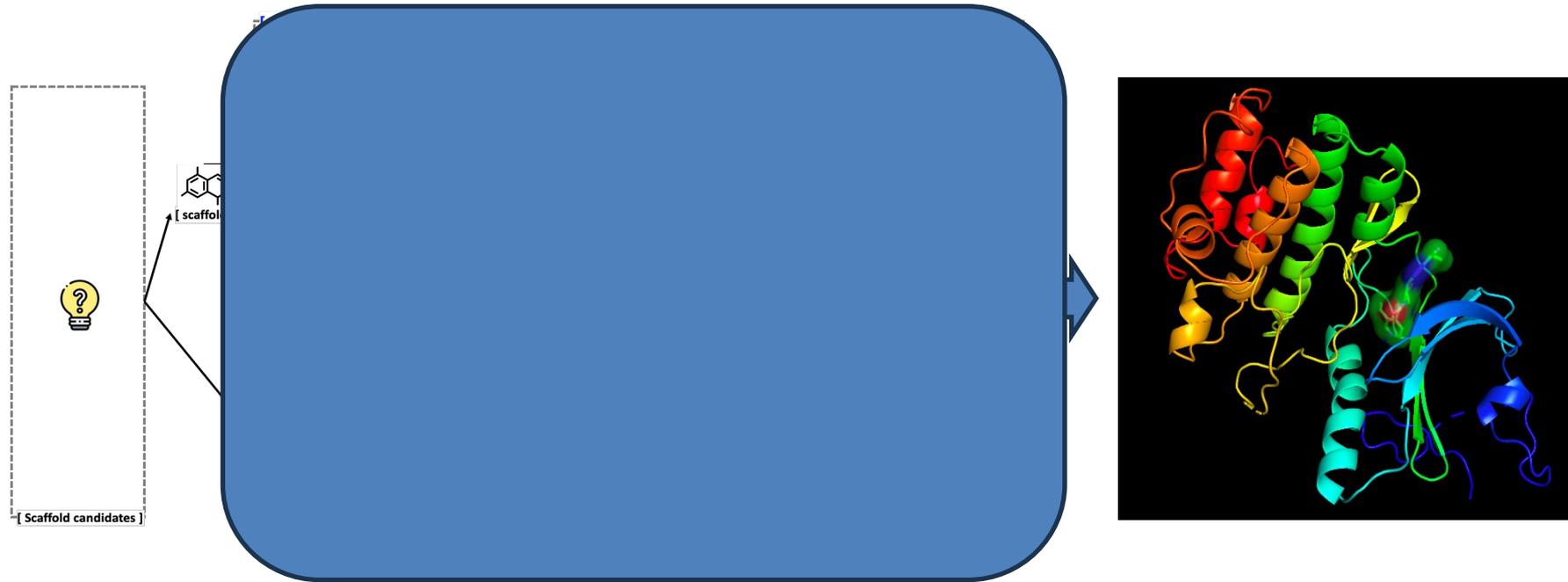
1. Virtual Screening (VS)

DaeWoong Basic strategy 1 : AI-based virtual screening



1. AI-based advanced virtual screening (DAIVS)

o DAIVS(AI based advanced virtual screening)을 이용한 가상 탐색

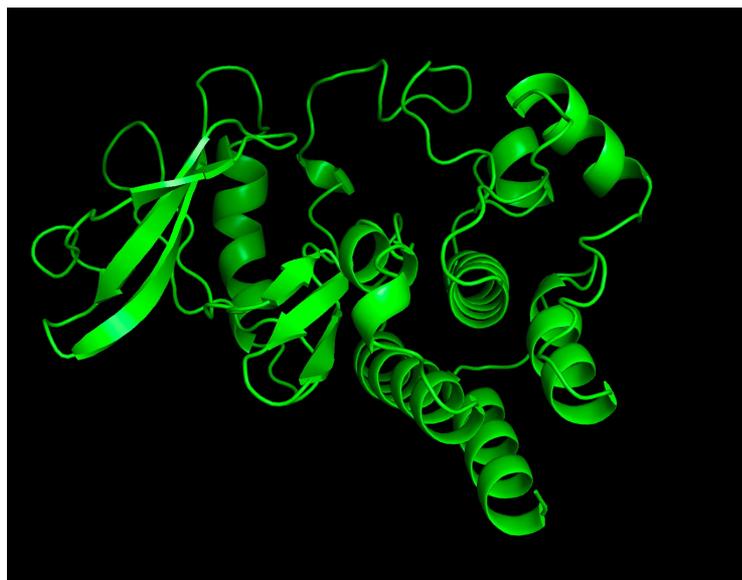


DAIVS 개념도

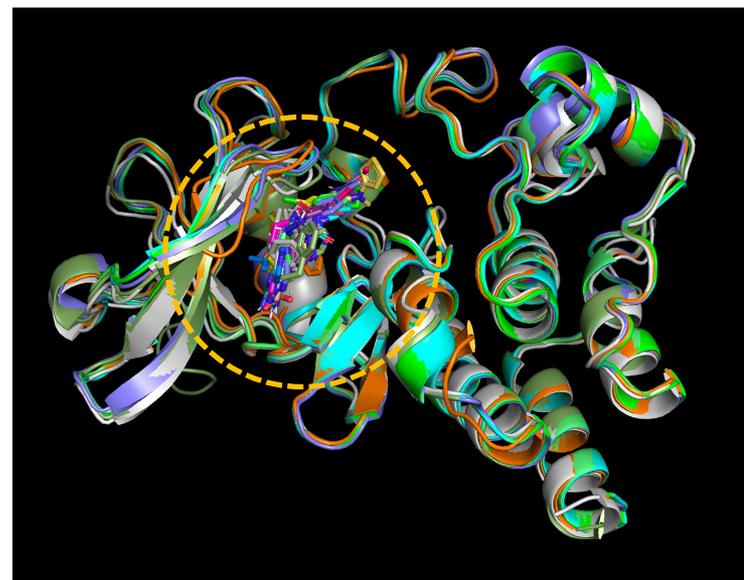
Hit 발굴 예

2. Virtual screening with **Gen. AI Model 1**

- Generative AI in Drug Discovery(*de novo generation*)
 - Generative Adversarial Networks (GANs)
 - Variational Autoencoders (VAEs)
 - Reinforcement Learning (RL)
- To generate **novel molecular structures** with **desired properties**, such as high binding affinity to a target protein or low toxicity.



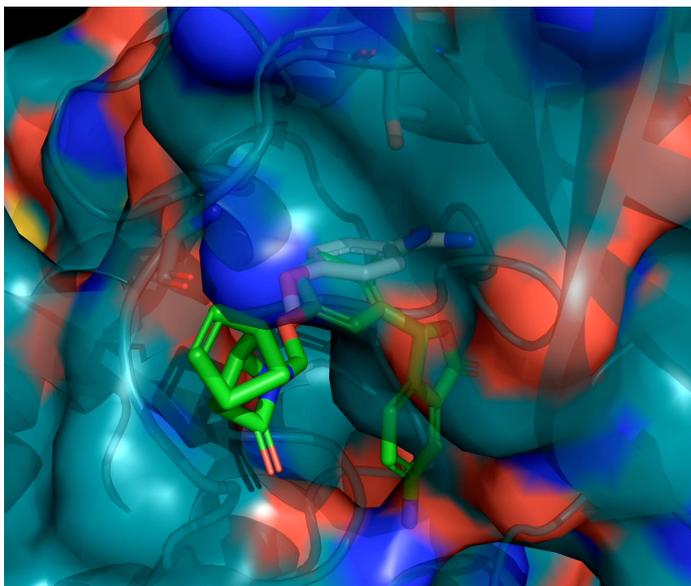
Target Protein



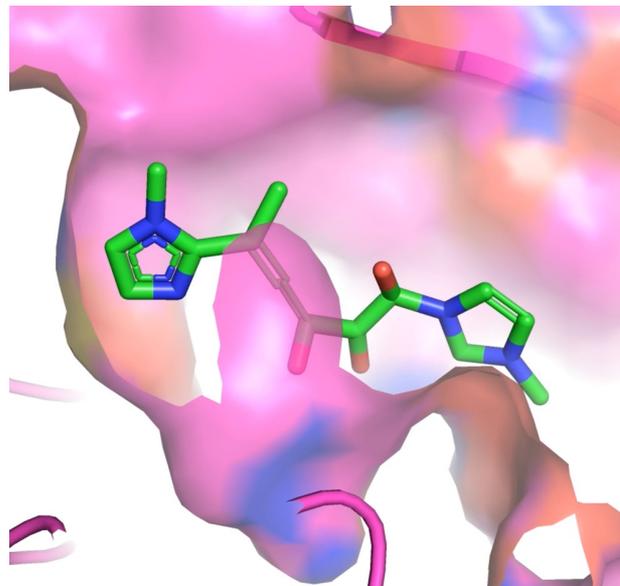
de novo generation 결과

2. Virtual screening with Gen. AI Model 2 (DAIFRGS)

- Using Gen. AI
 - Given base fragment([FBDD](#): Fragment-Based Drug Discovery)
 - Growing or Linking the base fragment(s)



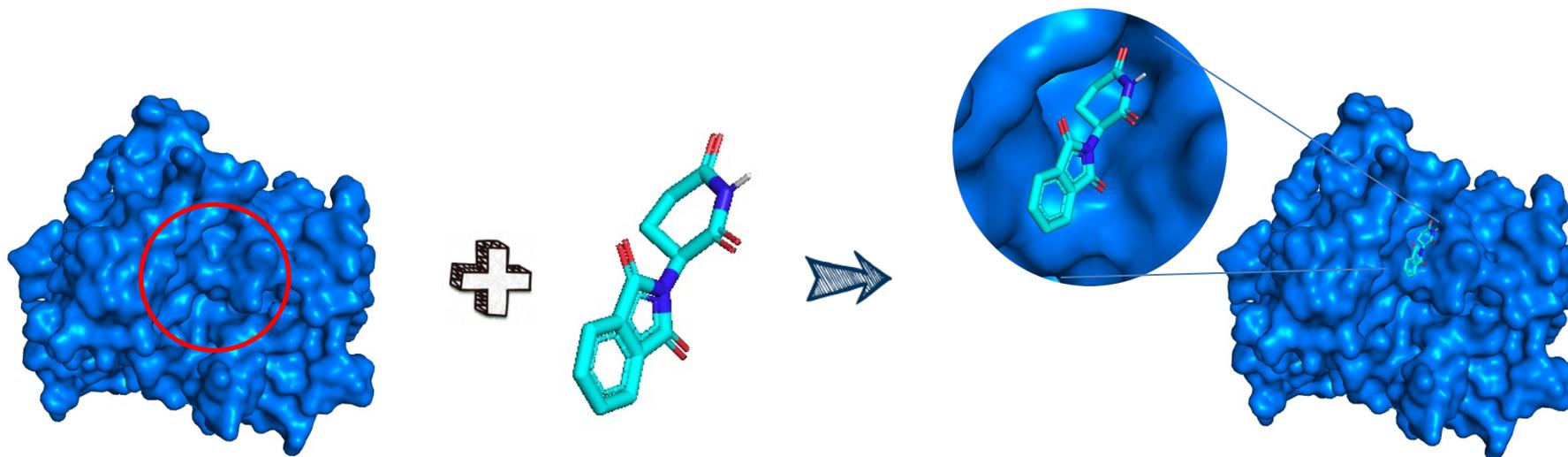
Growing example



Merging example

2. Docking Simulation

Docking Simulation



Protein & Binding site (pocket ?) ligand

Complex

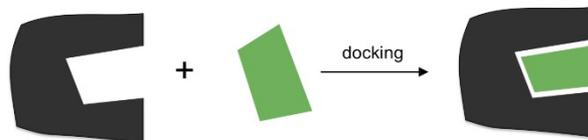
1. The **Best Pose (binding mode)**

- We can see which parts are important for binding
- We can suggest change to improve the affinity
- Avoid change that will clash with the protein

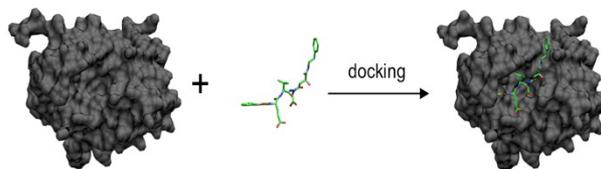
2. The **binding affinity or Score**

Docking approaches

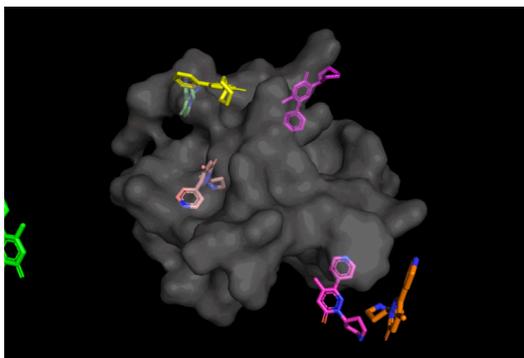
- One approach uses a matching technique that describes the protein and the ligand as complementary **surfaces**



- The second approach simulates the actual docking process in which the ligand-protein pairwise interaction **energies** are calculated: Vina-GPU



- The third approach uses **AI DiffDock (Protein-Rigid, Ligand-Flexible)**

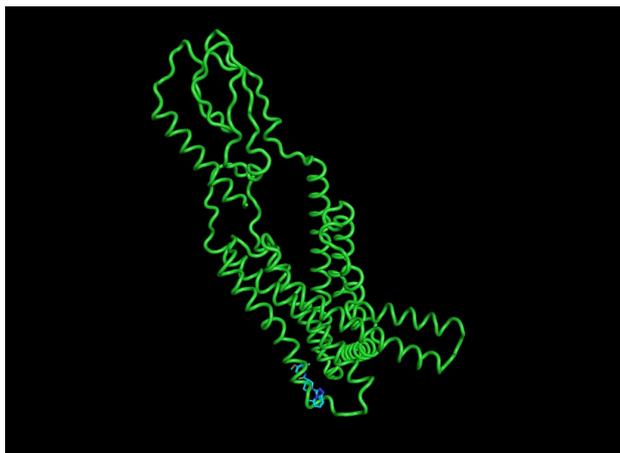
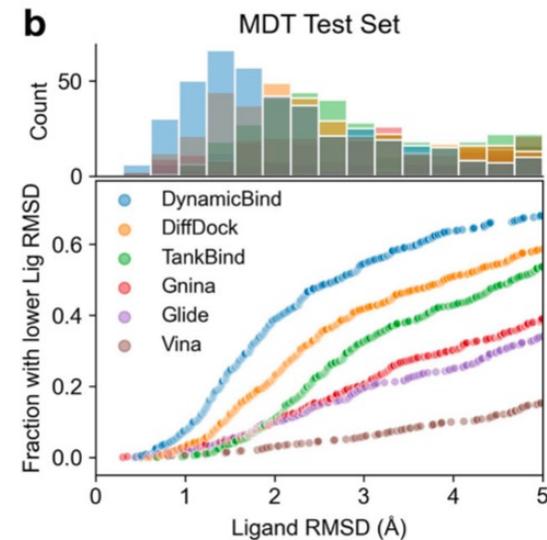
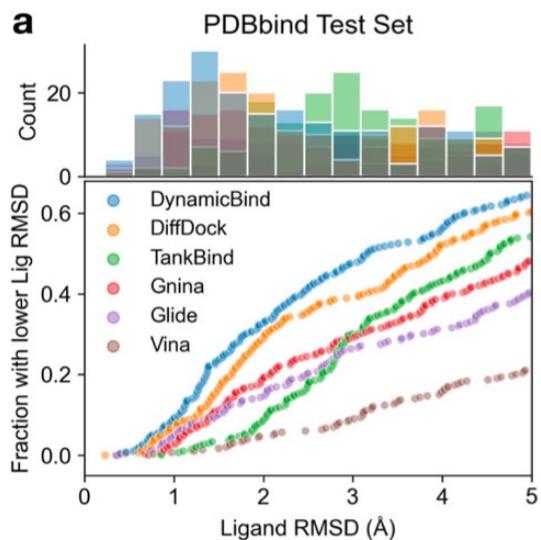
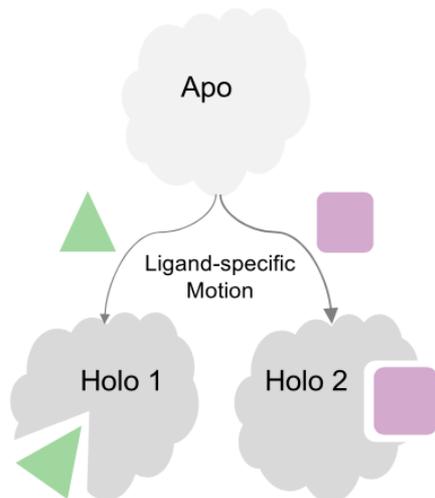


Method	PDBBind		DOCKGEN-full		DOCKGEN-clusters		Average Runtime (s)
	%<2Å	Med.	%<2Å	Med.	%<2Å	Med.	
SMINA	18.7	7.1	7.9	13.8	2.4	16.4	126*
SMINA (EX. 64)	25.4	5.5	10.6	13.5	4.7	14.7	347*
P2RANK+SMINA	20.4	4.3	7.9	14.1	1.2	16.4	126*
GNINA	22.9	7.7	14.3	15.2	9.4	14.5	127
GNINA (EX. 64)	32.1	4.2	17.5	8.1	11.8	6.2	348
P2RANK+GNINA	28.8	4.9	13.8	16.2	4.7	15.3	127
EQUIBIND	5.5	6.2	0.0	13.3	0.0	13.3	0.04
TANKBIND	20.4	4.0	0.5	11.6	0.0	11.1	0.7
DIFFDOCK (10)	35.0	3.6	7.1	6.8	6.1	6.0	10
DIFFDOCK (40)	38.2	3.3	6.0	7.3	3.7	6.7	40
DIFFDOCK-L [†] (10)	43.0	2.8	22.6	4.3	27.6	3.7	25
DIFFDOCK-S + C.B. [†] (10)	-	-	-	-	24.0	3.8	2.8

[Submitted on 4 Oct 2022 (v1), last revised 11 Feb 2023 (this version, v2)]

Docking approaches

- 4th approach uses **AI DynamicBind(Protein-Flexible, Ligand-Flexible)**



→ **Project X5: Find a candidate hit using bulk docking!**

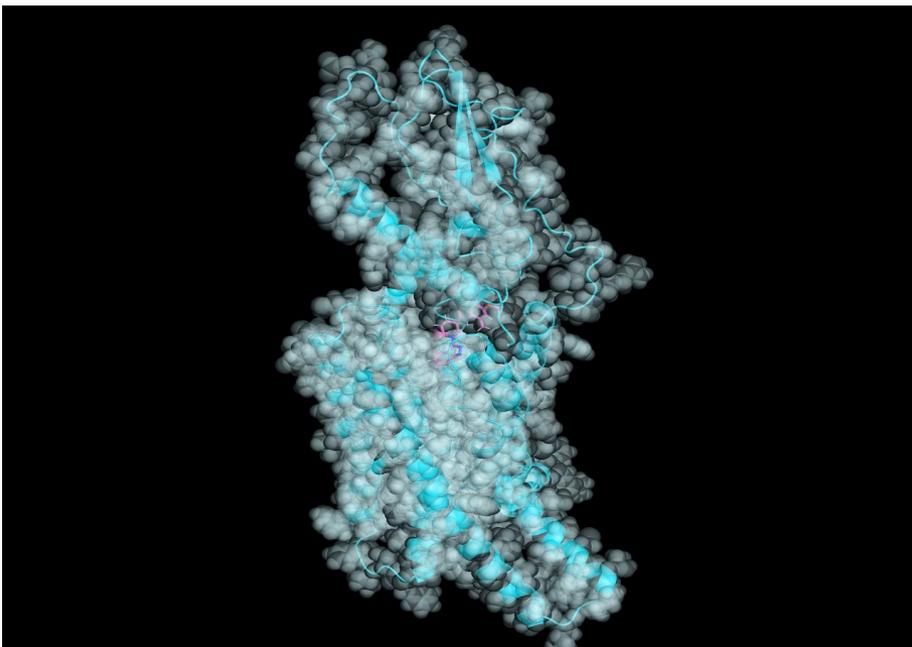
Lu, W., Zhang, J., Huang, W. *et al.* DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nat Commun* **15**, 1071 (2024). <https://doi.org/10.1038/s41467-024-45461-2>

Molecular Dynamics Simulation

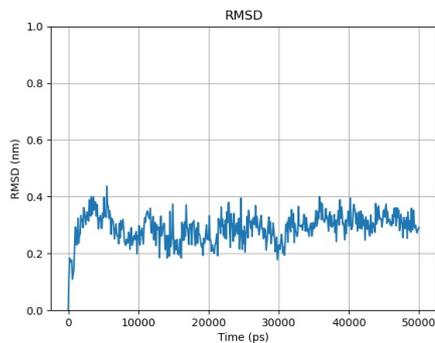
What is molecular dynamics (MD)?

- **Numerical method for studying many-particle systems** such as molecules, clusters, and even macroscopic systems such as gases, liquids and solids
- Used extensively in materials science, chemical physics, and biophysics/biochemistry
- In Drug discovery, it shows **the stability** between protein and ligand (hit, candidate drug)

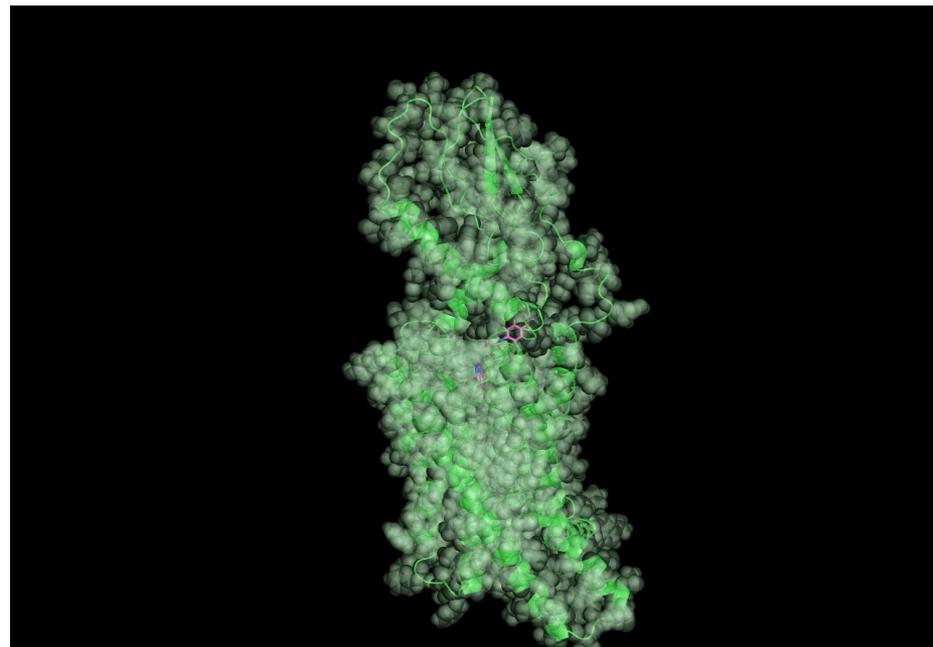
Molecular dynamics (MD) – Example



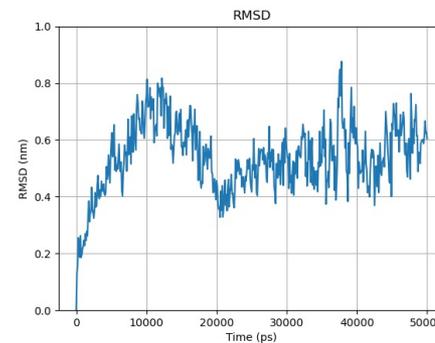
[Fig.1 : XX - 103606 **R-form** RMSD during 200 ns MD Simulation]



[Fig. 1: XX - 103606 **R-form** RMSD during 50 ns]



[Fig. 2 : XX - 103606 **S-form** RMSD during 200 ns MD Simulation]



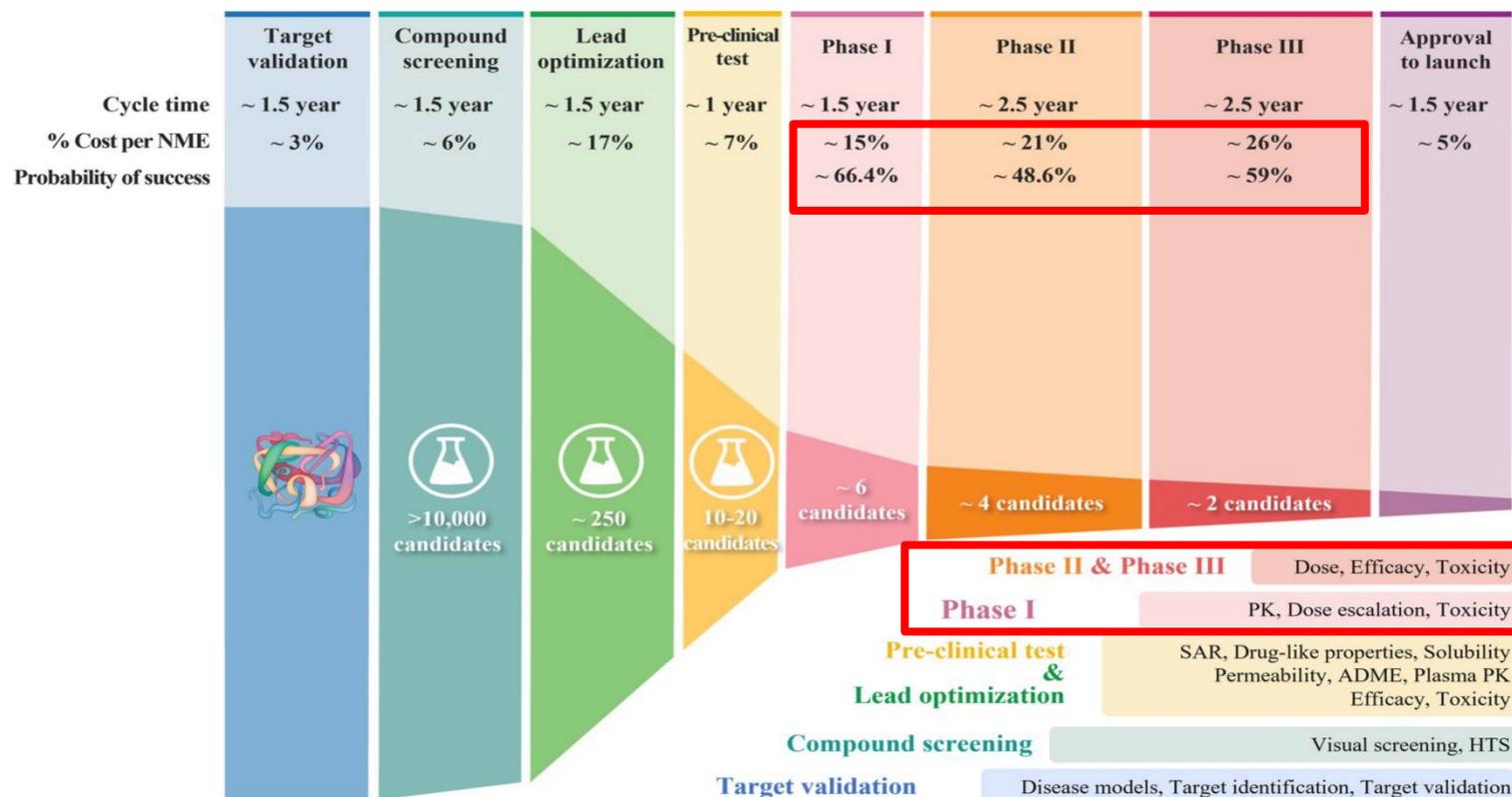
[Fig. 2 : XX - 103606 **S-form** RMSD during 50 ns]

4. ADME/T Prediction

Why drugs fail ?

○ Phase I,II,III : 90%

→ PK, Dose escalation, Toxicity, Efficacy !



DAISY-ADME/T Prediction



ADAPT: Advanced Daewoong ADME/T Prediction Tool

Compound ID:

SMILES:

Predict

Need Help?

Mail to Webmaster: mhs0511@daewoong.co.kr

Go to Home



ADME/T Property Prediction Result

Result file generated. Download it here:

Download Result

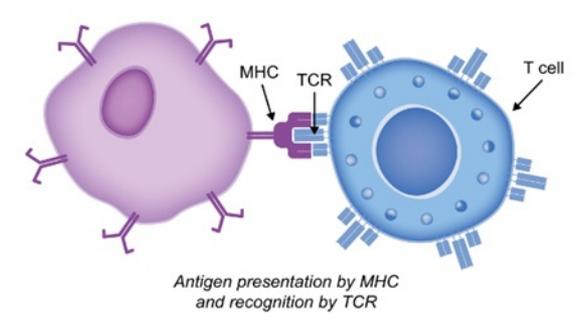
Back

ID	Test
Smiles	<chem>O=C(O)c1ccc(N[C@H]2OC[C@@H](O)[C@H](O)[C@H]2O)cc1</chem>
Query_time	2024-04-23-04-41-02
bioavailability	Inactive
probability for bioavailability prediction	0.5499
PGP	Inactive
probability for PGP prediction	0.75
HIA	Inactive
probability for HIA prediction	0.6031
Lipophilicity (logD at pH7.4)	0.5686
Caco-2 (logPapp)	-3.7241
Solubility (logS at room temp)	-1.5012
BBB	Inactive
probability for BBB prediction	0.5175
VDss(The degree of a drug's conc. in body tissue compared to conc. in blood)	3.3455
PPBR (Plasma protein binding rate %)	0.8887
CYP1A2	Inactive
probability for CYP1A2 prediction	0.6759
CYP2A6	Inactive

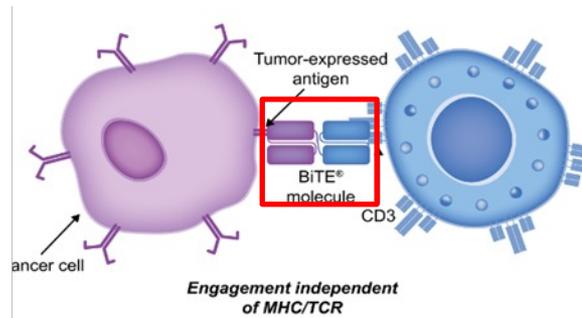
- ADME/T prediction using AI model (19/24 endpoints: **14!, TDC**)
- <http://192.168.250.189/adme.php>
- Metabolism prediction Upgrading
 - Active/inactive → IC50

mRNA-Bispecific T cell engager(BiTE)

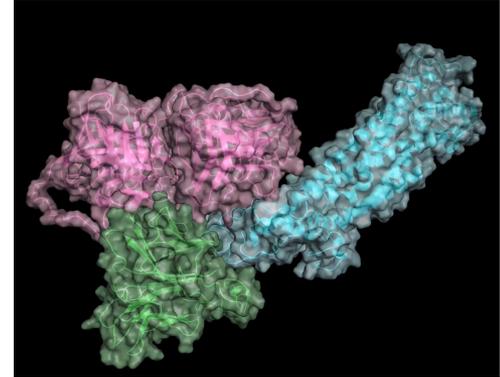
mRNA-Bispecific T cell engager(BiTE)



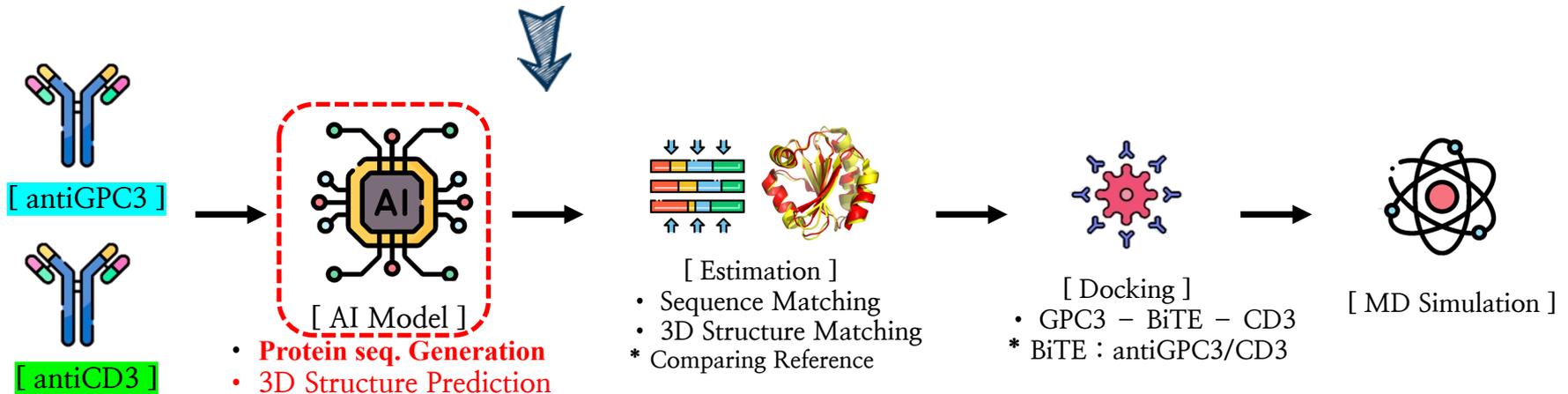
Normal T cell response



Bispecific T cell engager



To make humanized antibodies: **about 6 months.**

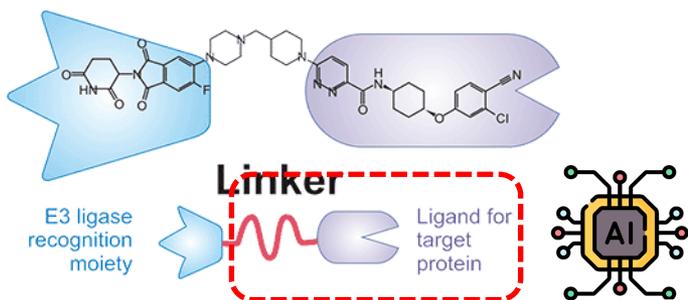


To make **humanized antigen using AI: less 1 min.**

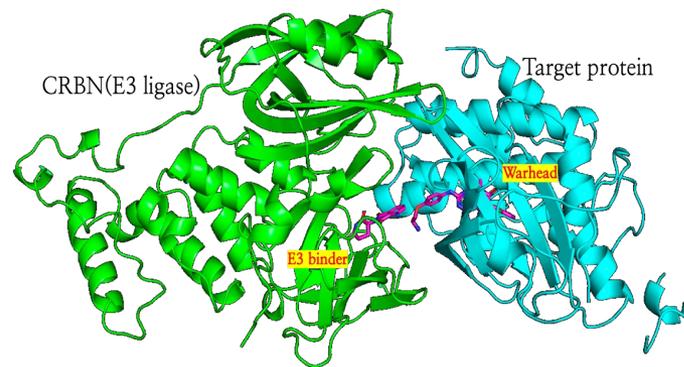
PROTAC in silico assay

PROTAC in silico assay

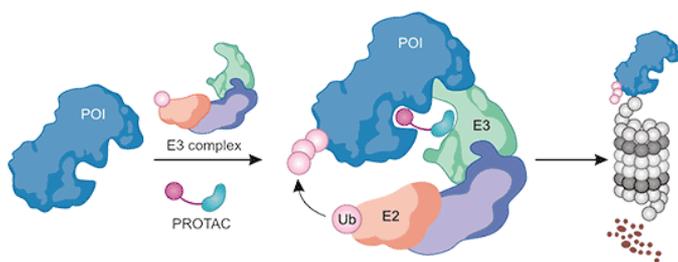
: A small molecule composed of two active domain-binding molecules and a linker
Removing specific unwanted proteins → **Cancer**



The Structure of PROTACs

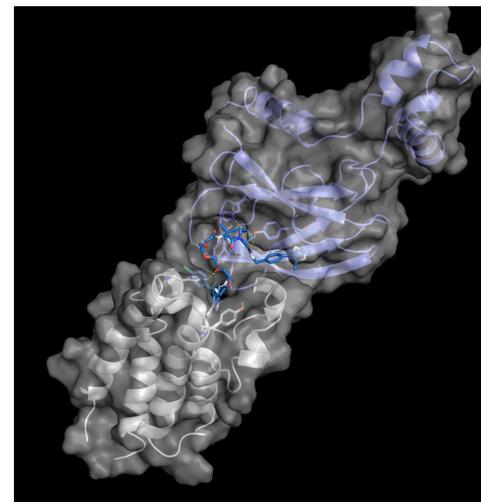


The Modeling of PROTACs



“Target protein-PROTAC-E3 ubiquitin ligase” ternary complex formation process

Target



The Simulation of PROTACs

E3

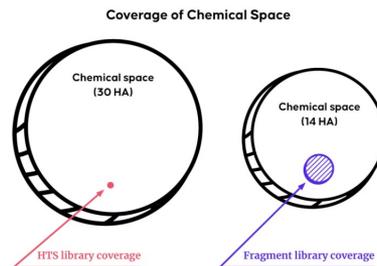
감사합니다.



대웅제약
www.daewoong.co.kr

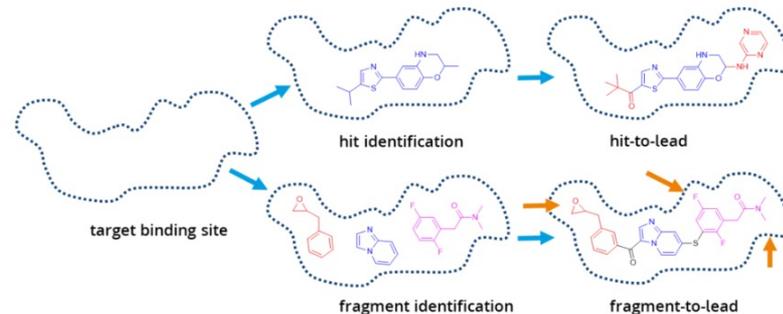
* Fragment-Based Drug Discovery(FBDD)

- Why is Fragment-Based Drug Discovery Useful?
→ Chemical space: 10^{60} ! (compound libraris: 10^{5-6})



- Key concepts in fragment-based drug discovery

Fragments: Perfect Binders



- Successful example

→ Over 50 molecules in clinical trials and 6 FDA-approved drugs

