

- 구조 기반 가상 스크리닝

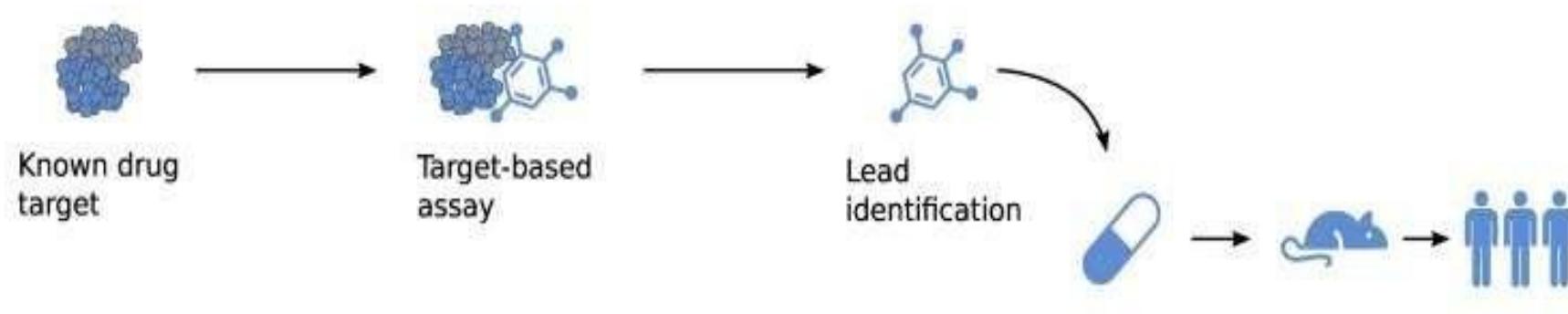
< Target-based drug discovery >

➤ select Target Protein & Library

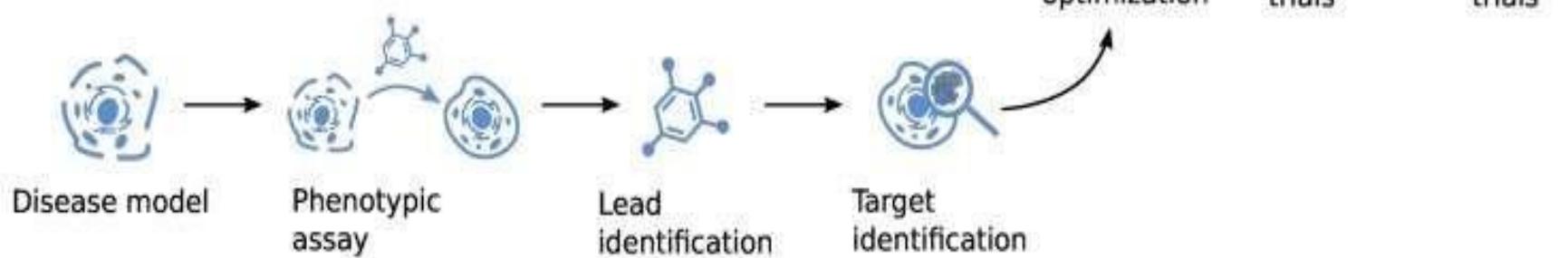
< Phenotypic drug discovery >

➤ select Model system & Library

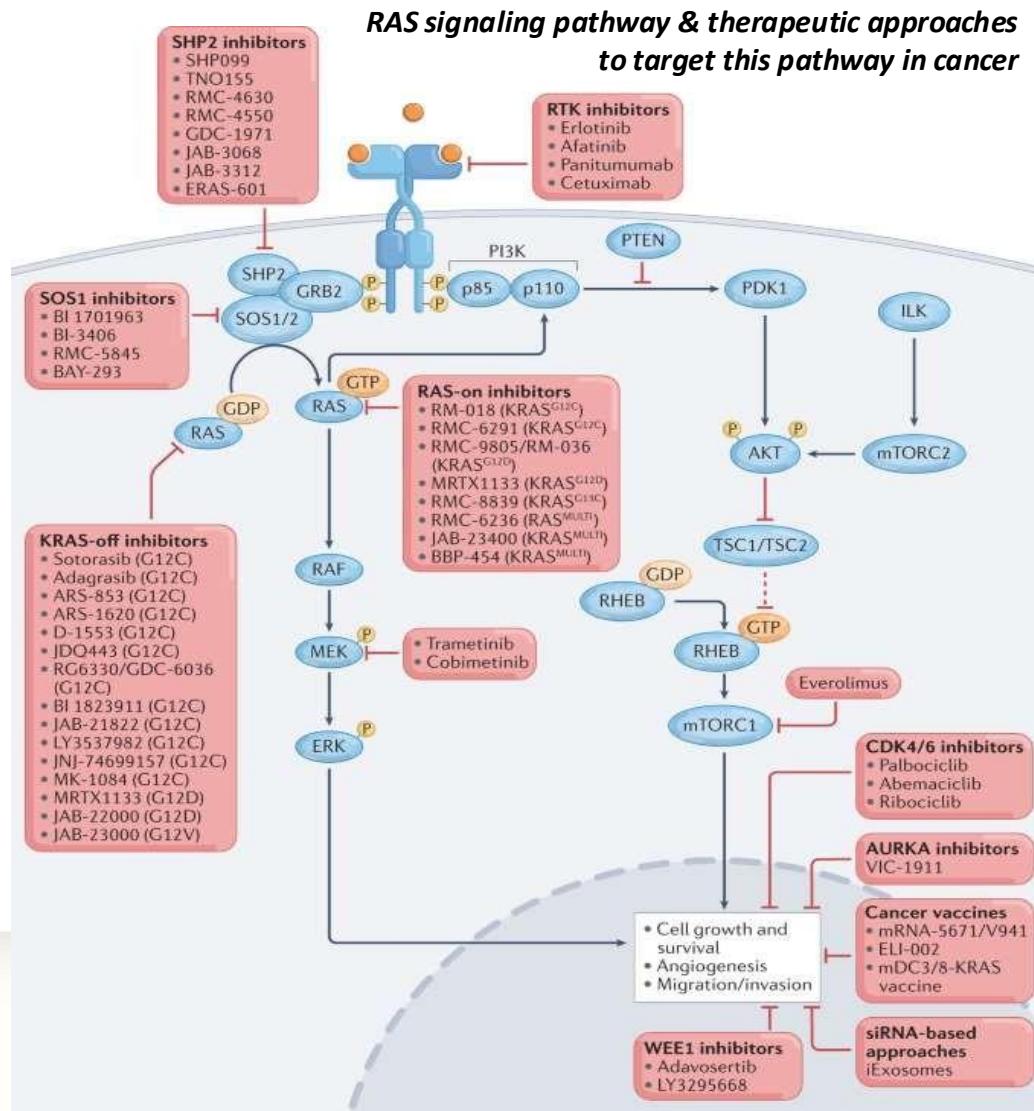
(A) Target-based drug discovery



(B) Phenotypic drug discovery



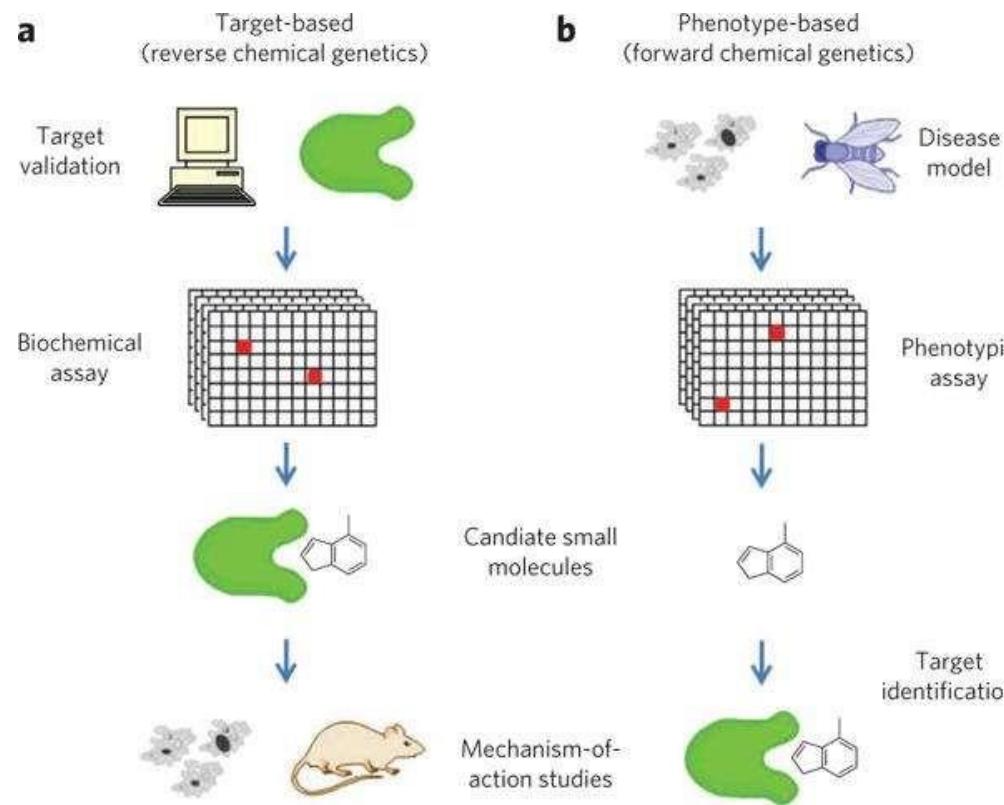
## &lt; Target-based drug discovery &gt;



## &lt; Phenotypic-based drug discovery &gt;

Table 1 | Phenotypic origins of approved drugs and clinical-phase compounds

Drug or clinical candidate	Structure	Indication	Phenotypic screening strategy	Mechanism of action	Development phase
Daclatasvir (modulators of NSSA are components of all anti-HCV drugs)	<chem>CC(=O)N1C2=C(C=C1C(=O)N3C4=C(C=C3C(=O)N5C6=C(C=C5C(=O)N7C8=C(C=C7C(=O)N9C=C8C9=O)C=C6)C=C4)C=C3)C=C2</chem>	Hepatitis C infection	Target-agnostic viral replication screen <sup>13</sup>	NSSA identified as molecular target; HCV replication inhibition; MoA unknown	Launched
Lumacaftor (component of Orkambi along with ivacaftor)	<chem>CC1=CC=C(C=C1C(=O)N2C3=C(C=C2C(=O)N4C5=C(C=C4C(=O)N6C7=C(C=C6C(=O)N8C9=C(C=C7C=C9)C=C5)C=C4)C=C3)C=C2)C=C1</chem>	Cystic fibrosis	Mechanism-agnostic cellular screen to enhance CFTR function <sup>14</sup>	Correctors enhance the folding and plasma membrane insertion of CFTR; novel MoA <sup>20</sup>	Launched
Lenalidomide	<chem>CC1=CC2=C(C=C1C(=O)N3C4=C(C=C2C(=O)N5C6=C(C=C4C(=O)N7C8=C(C=C6C(=O)N9C=C8C9=O)C=C5)C=C3)C=C2)C=C1</chem>	Multiple myeloma and other haematological malignancies	Functional cellular assays and off-label observational studies in patients <sup>15,16</sup>	Alters protein substrate specificity of E3 ubiquitin ligase Cereblon; novel target class and MoA <sup>21</sup>	Launched
Risdiplam	<chem>CC1=CC2=C(C=C1C(=O)N3C4=C(C=C2C(=O)N5C6=C(C=C4C(=O)N7C8=C(C=C6C(=O)N9C=C8C9=O)C=C5)C=C3)C=C2)C=C1</chem>	SMA	Mechanism-agnostic cellular assay to correct SMN2 pre-mRNA splicing <sup>22</sup>	Engagement and stabilization of SMN2 exon 7 and U1 snRNP complex; novel target class and MoA <sup>23-25</sup>	Launched
Clopidogrel (prodrug of active metabolite responsible for activity) <sup>26</sup>	<chem>CC1=CC2=C(C=C1C(=O)N3C4=C(C=C2C(=O)N5C6=C(C=C4C(=O)N7C8=C(C=C6C(=O)N9C=C8C9=O)C=C5)C=C3)C=C2)C=C1</chem>	Cardiovascular disease	Anti-platelet activity identified using a battery of in vivo and ex vivo rodent models screened to explore anti-inflammatory activity <sup>26</sup>	Active metabolite selectively and irreversibly blocks platelet P2Y <sub>12</sub> /ADP receptors <sup>20</sup>	Launched
SEP-363856	<chem>CC1=CC2=C(C=C1C(=O)N3C4=C(C=C2C(=O)N5C6=C(C=C4C(=O)N7C8=C(C=C6C(=O)N9C=C8C9=O)C=C5)C=C3)C=C2)C=C1</chem>	Schizophrenia, psychosis	Automated in vivo behavioural models, the 'SmartCube' system <sup>27,28,29</sup>	Positive phase II results mediated by novel non-dopamine GPCR mechanism; novel MoA <sup>30</sup>	Phase III (schizophrenia), phase II (psychosis)
Deucravacitinib	<chem>CC1=CC2=C(C=C1C(=O)N3C4=C(C=C2C(=O)N5C6=C(C=C4C(=O)N7C8=C(C=C6C(=O)N9C=C8C9=O)C=C5)C=C3)C=C2)C=C1</chem>	Psoriasis and other autoimmune conditions	Kinase biased compounds tested in cellular assay monitoring IL-23 signalling pathway <sup>31</sup>	Positive phase III results; novel MoA; allosteric inhibition of TYK2 kinase through catalytically inactive pseudo-kinase domain <sup>32</sup>	Phase III (psoriasis), phase II (other indications)



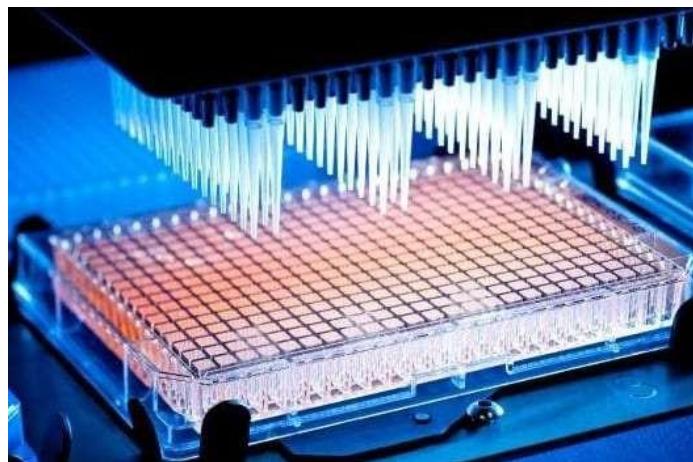
	High-throughput target-based screening	Phenotypic screening
Advantages	<ul style="list-style-type: none"> <li>+ Extremely quick assays (facilitated by automation)</li> <li>+ Low volumes required</li> </ul>	<ul style="list-style-type: none"> <li>+ Compounds are tested for actual function against chosen disease characteristics</li> <li>+ Identification of single biological target not required</li> </ul>
Challenges	<ul style="list-style-type: none"> <li>- High purity of protein is required</li> <li>- Focuses on only a single protein where multiple targets may be implicated</li> <li>- In general, information regarding therapeutic effect is not obtained</li> </ul>	<ul style="list-style-type: none"> <li>- Lower throughput</li> <li>- Generally more expensive per compound tested, especially if performed in animals</li> <li>- Assays require representative readouts that are relevant to the disease</li> </ul>

Trends in Chem. 2019, 1, 612-624

Trends In Chemistry

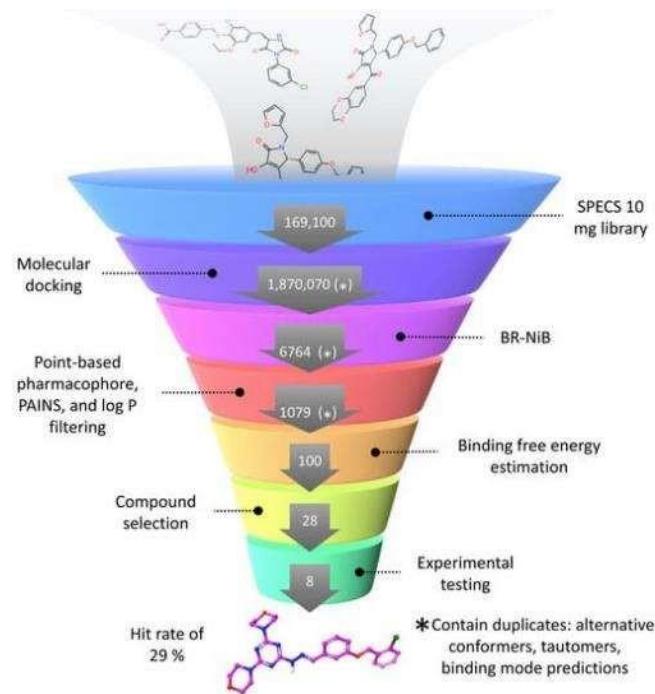
<https://europepmc.org/article/MED/23508189>

## High Throughput Screening (HTS)



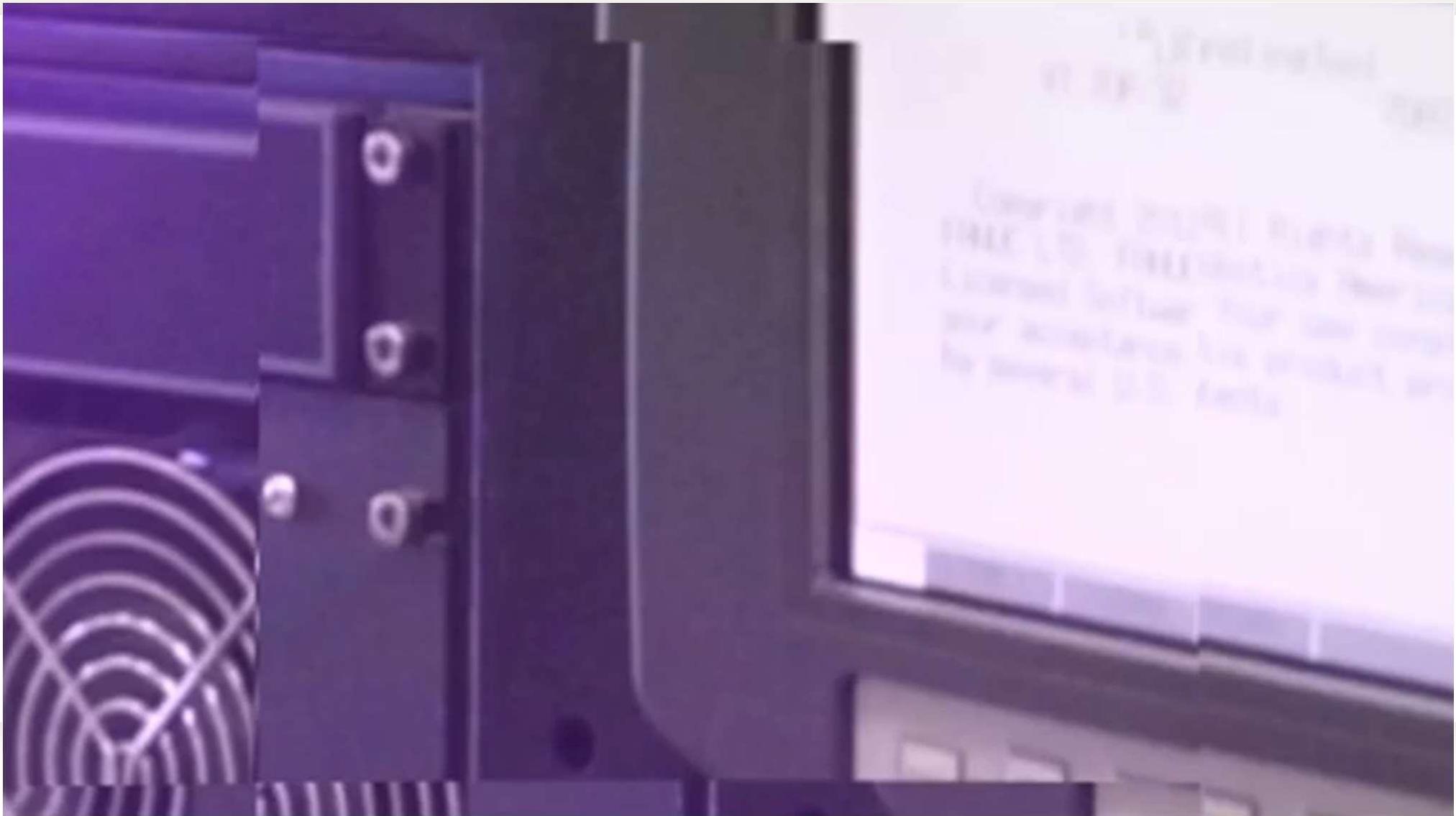
<https://www.youtube.com/watch?v=6SRC2zhe1zo>

## High Throughput Virtual Screening (HTVS)



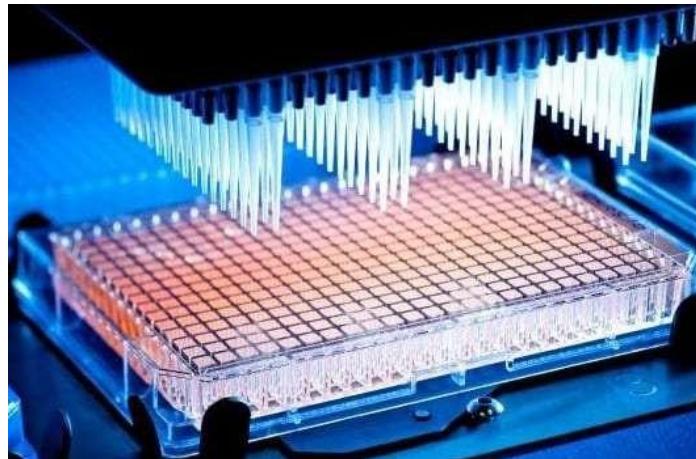
<https://www.mdpi.com/1420-3049/28/8/3420>

## High Throughput Screening (HTS)



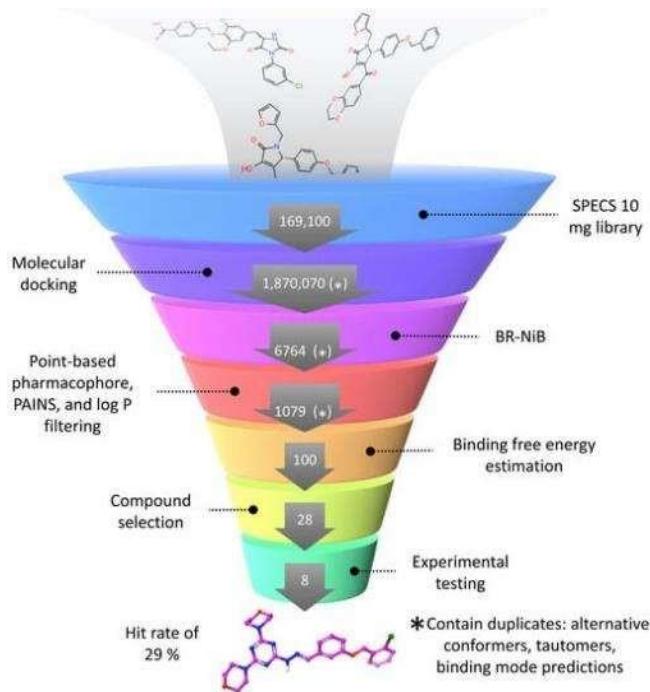
<https://www.youtube.com/watch?v=6SRC2zhe1zo>

## High Throughput Screening (HTS)



<https://www.youtube.com/watch?v=6SRC2zhe1zo>

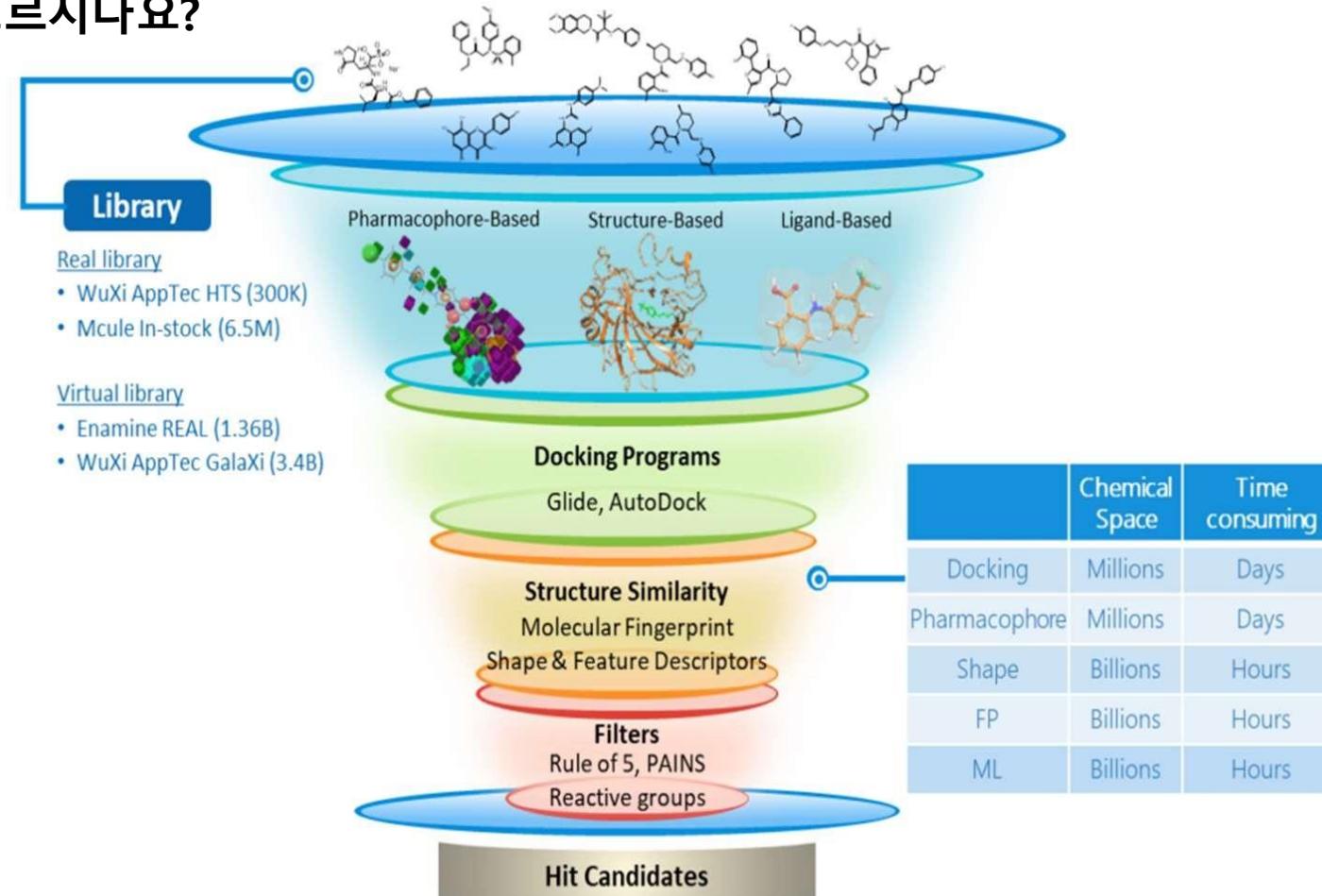
## High Throughput Virtual Screening (HTVS)



<https://www.mdpi.com/1420-3049/28/8/3420>

## • 가상 스크리닝 (Virtual Screening)

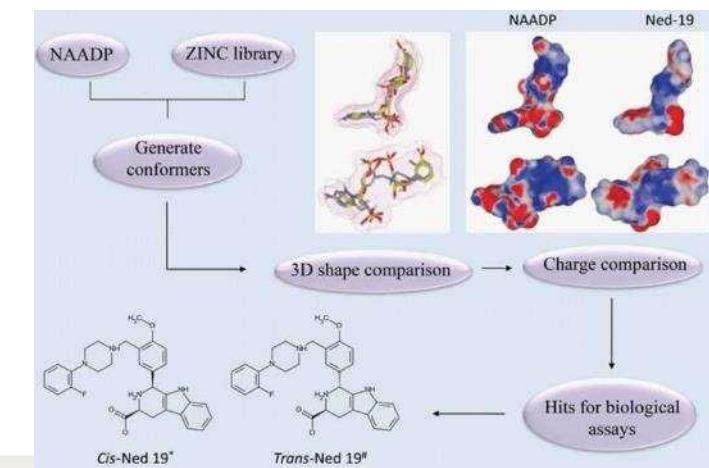
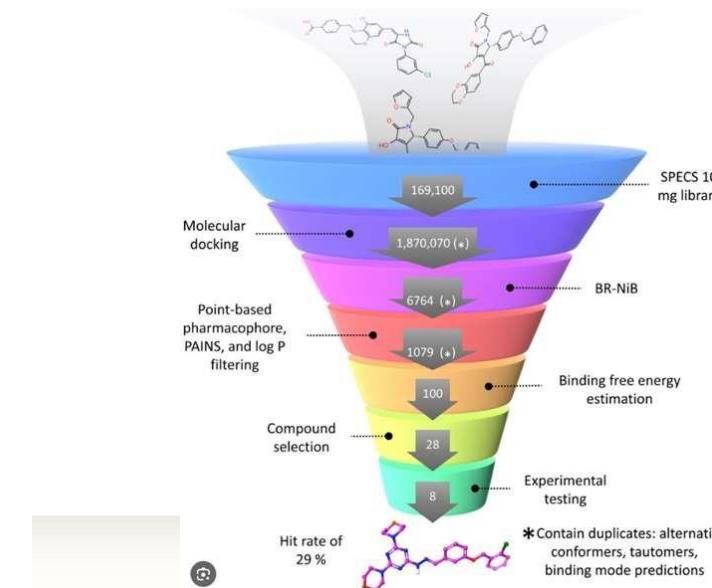
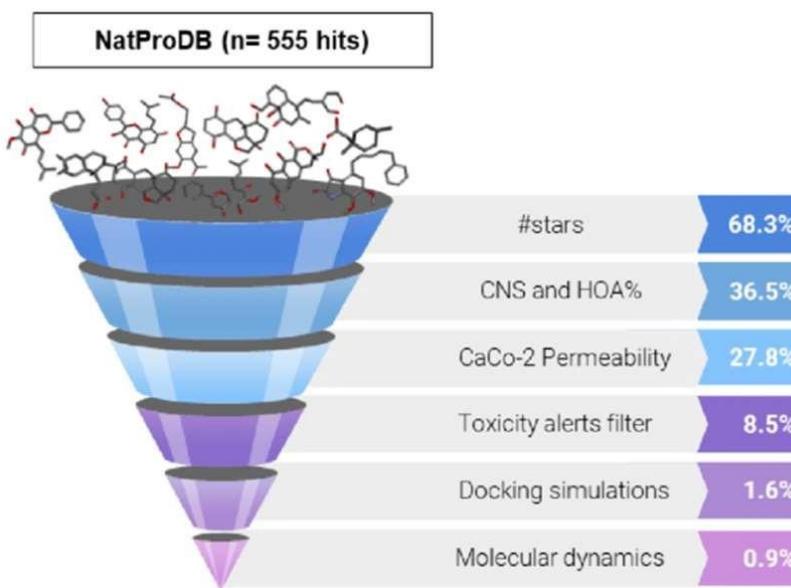
: 어떤 단어들이 떠오르시나요?



<https://wuxibiology.com/drug-discovery-services/hit-finding-and-screening-services/virtual-screening/>

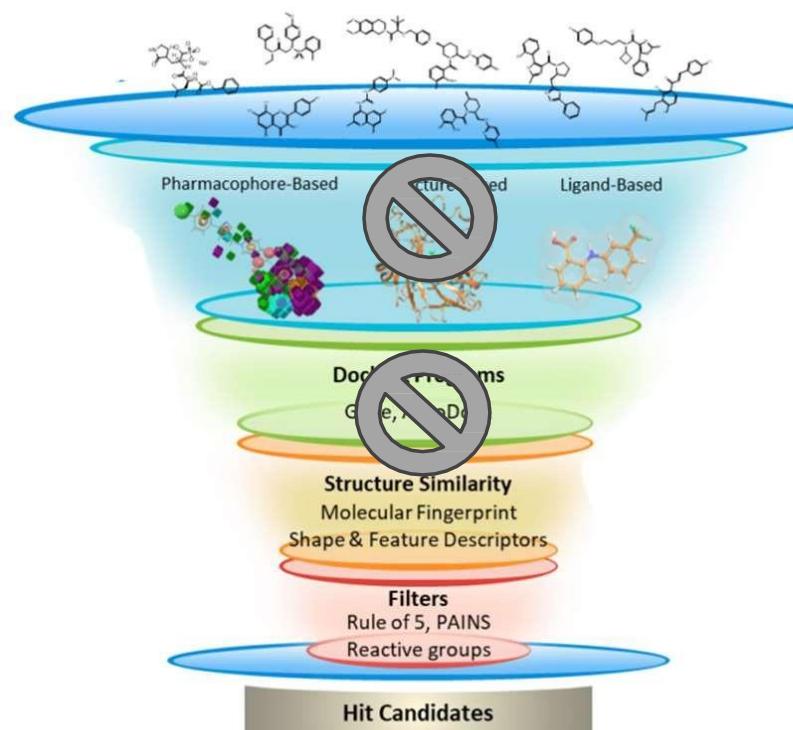
# • 가상 스크리닝 (Virtual Screening)

- ✓ In silico 상에서
- ✓ 활성을 보일 것으로 예상되는 화합물을 찾는 것
- ✓ 정해진 workflow는 없음 □ 연구자가 주어진 정보 & 목적 & assay capabilities에 맞게 설계



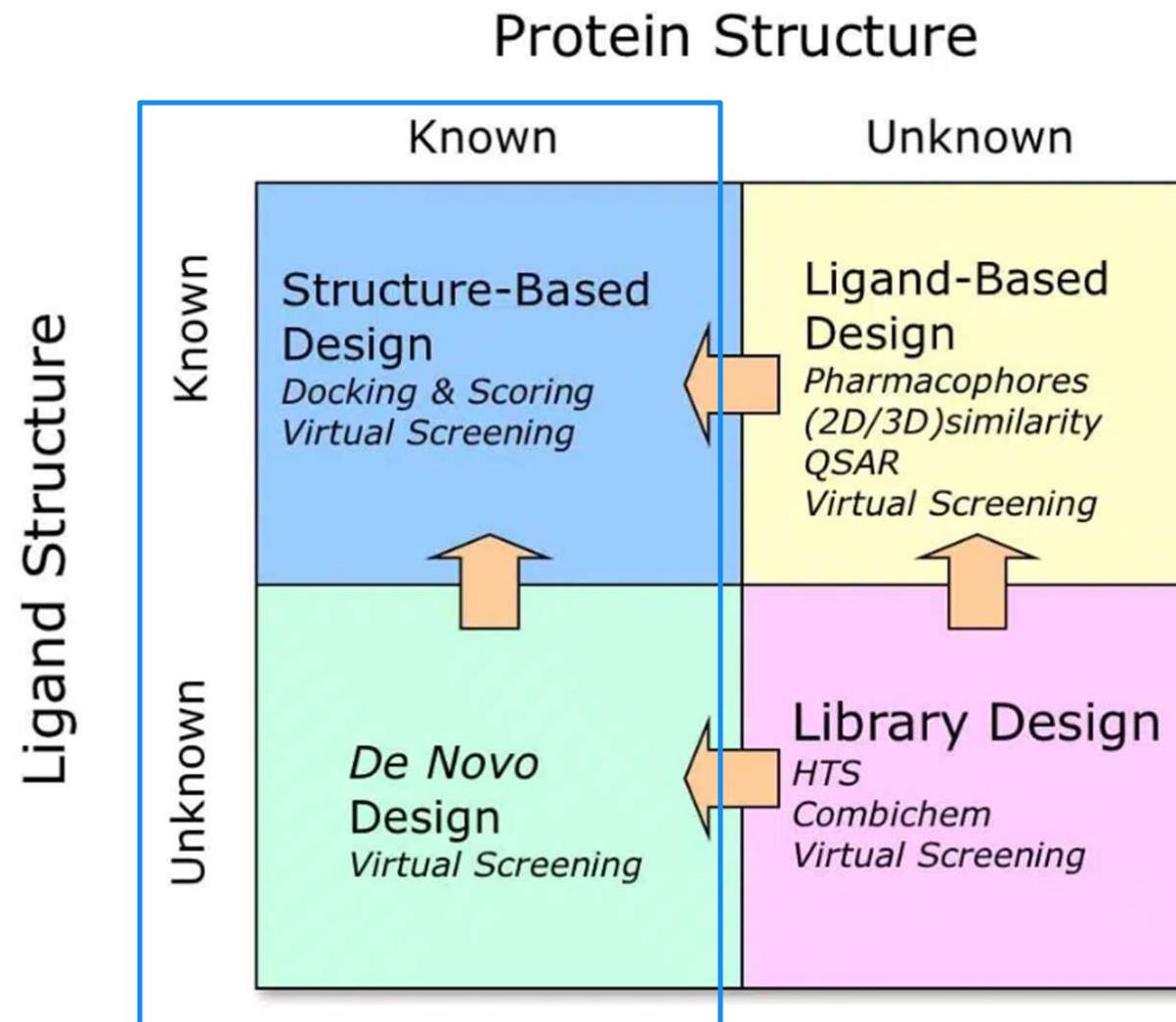
## &lt; 구조기반 가상 스크리닝 &gt;

- : 가상 스크리닝 과정에 타겟 단백질의 구조 정보 활용
- : 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것



## &lt; 리간드기반 가상 스크리닝 &gt;

- ✓ 타겟 단백질의 3차원 구조정보 없음
- ✓ Binding site이 알려져 있지 않음



## • 구조기반 가상 스크리닝

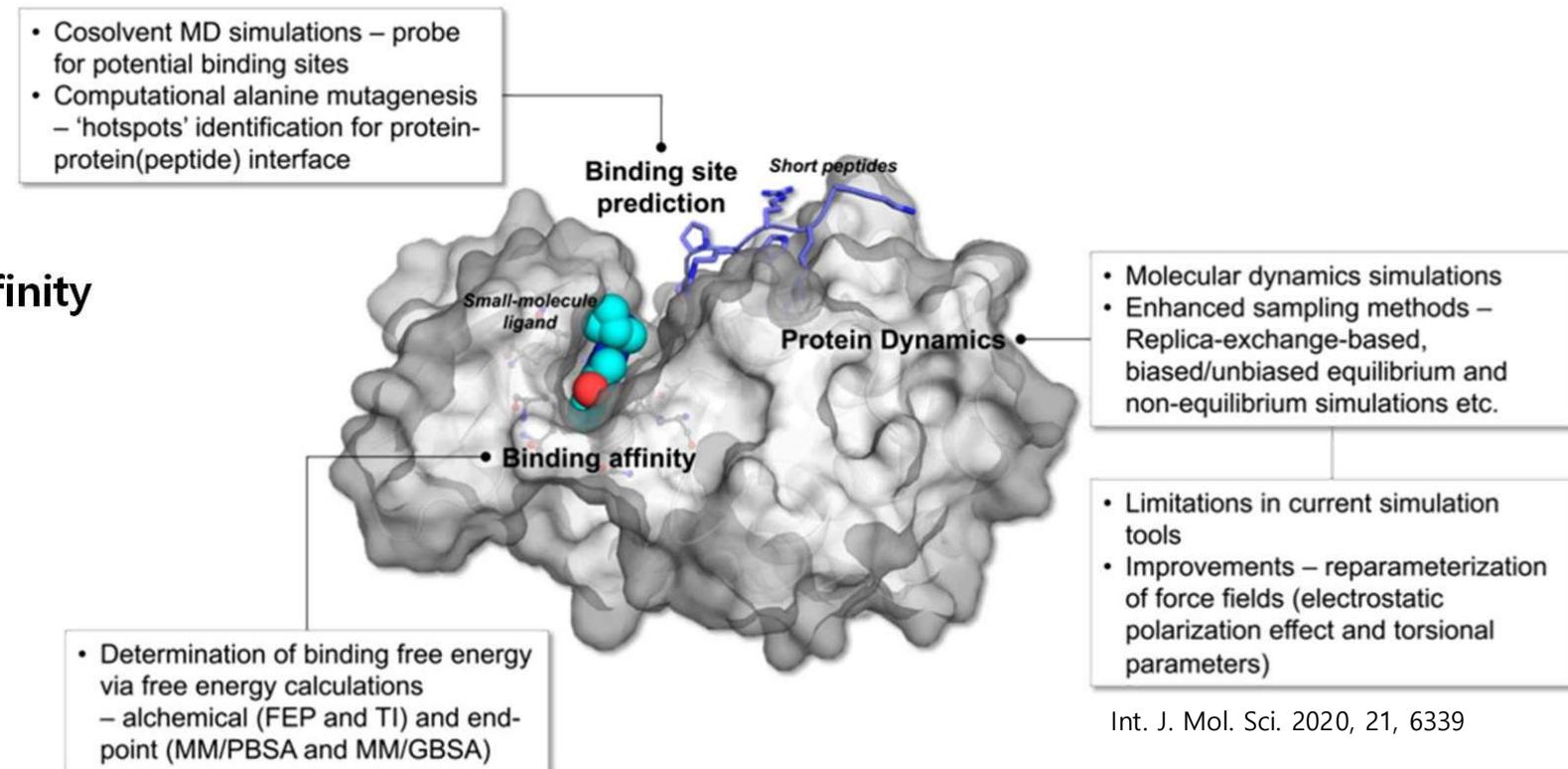
: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것

➤ Protein Structure

➤ Binding Site/Pocket

➤ Protein-ligand Interaction/Binding affinity

➤ Chemical Library



## • 구조기반 가상 **스크리닝**

: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것

- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket
- Protein-ligand Interaction/Binding affinity

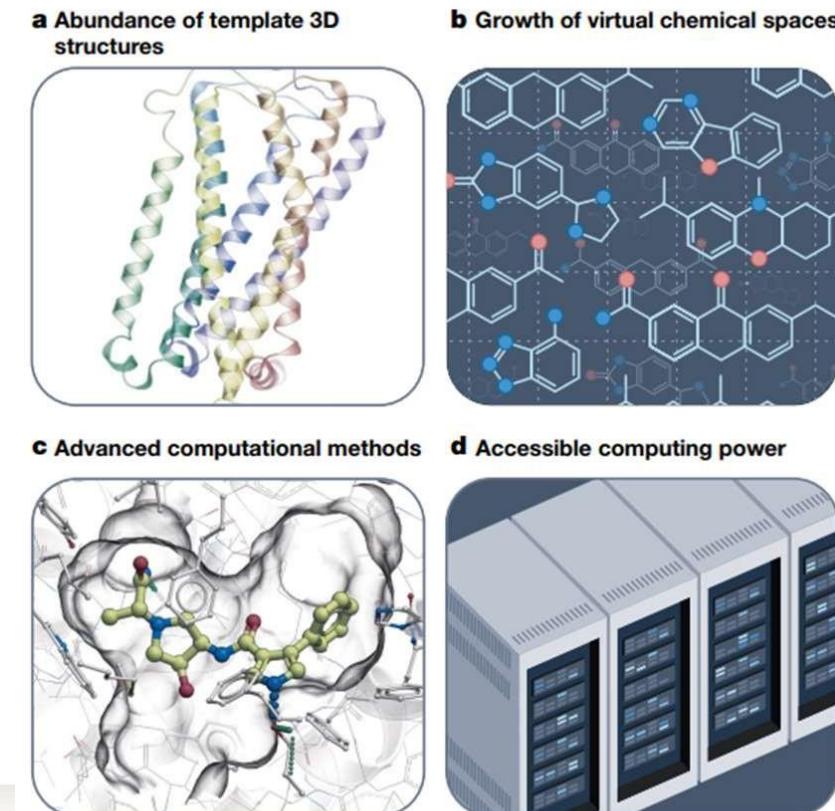


Fig. 1 | Key factors driving VLS technology breakthroughs for generation of high-quality hits and leads. **a**, More than 200,000 protein structures in the

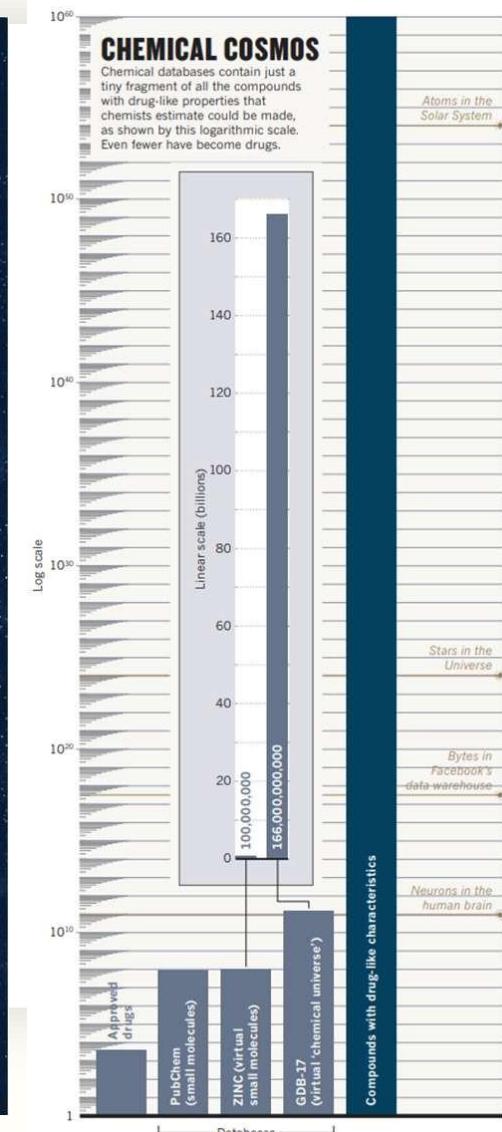
Nature, 2023, 616, 673

- 스크리닝

: 화합물을 찾는 것

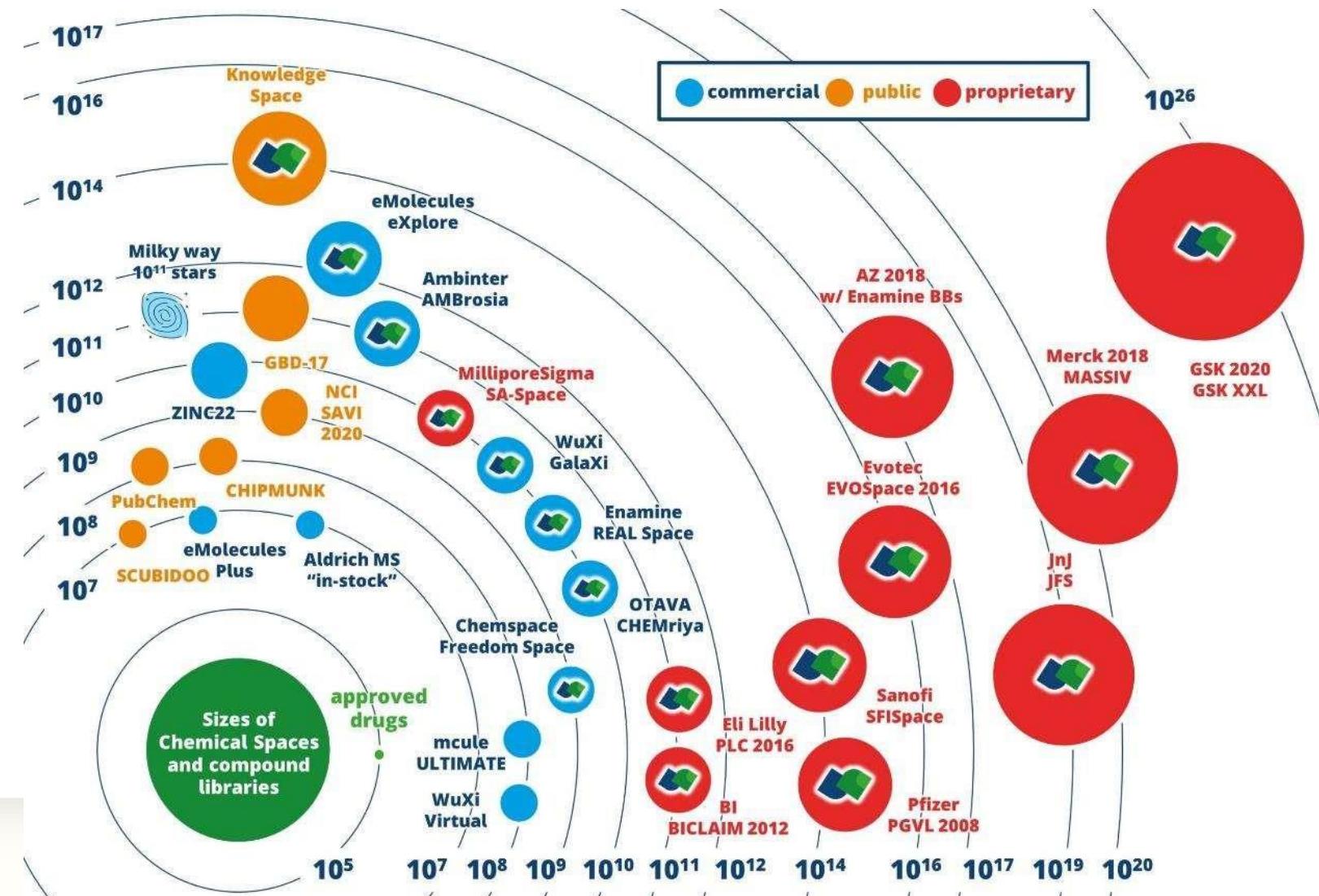


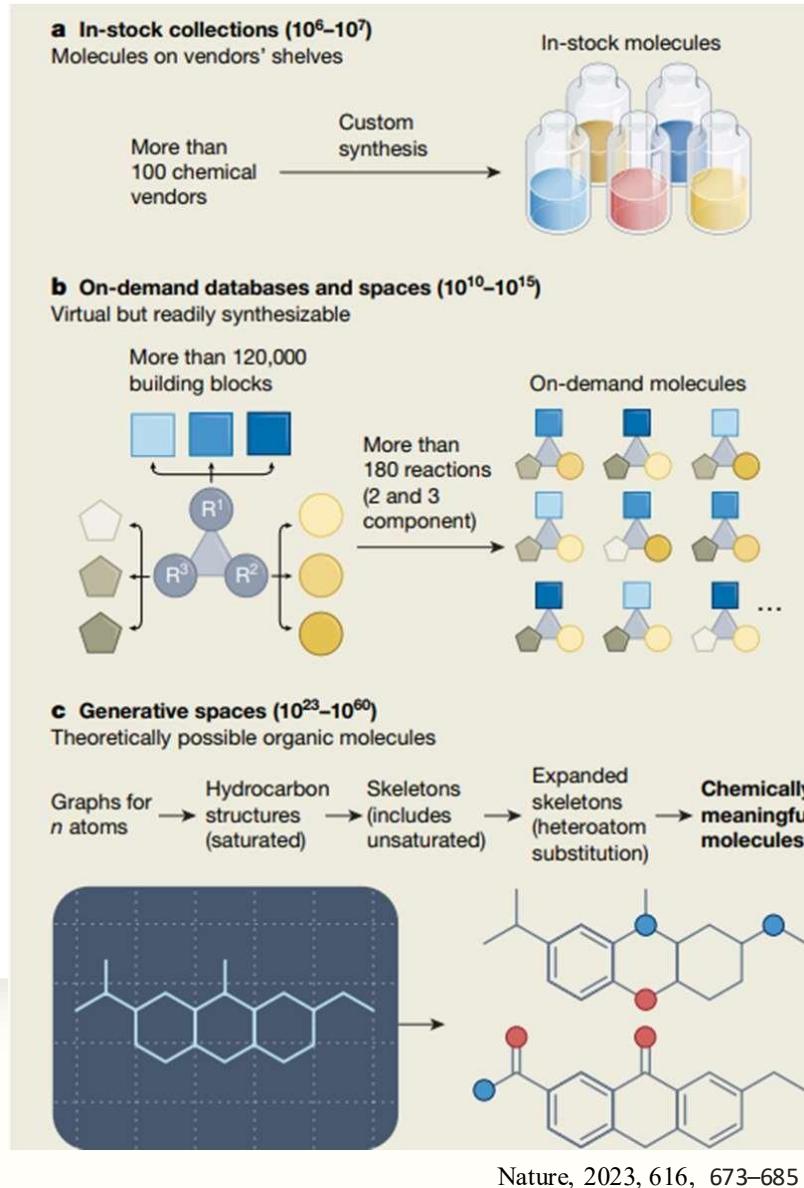
Nature, 2017, 549, 445



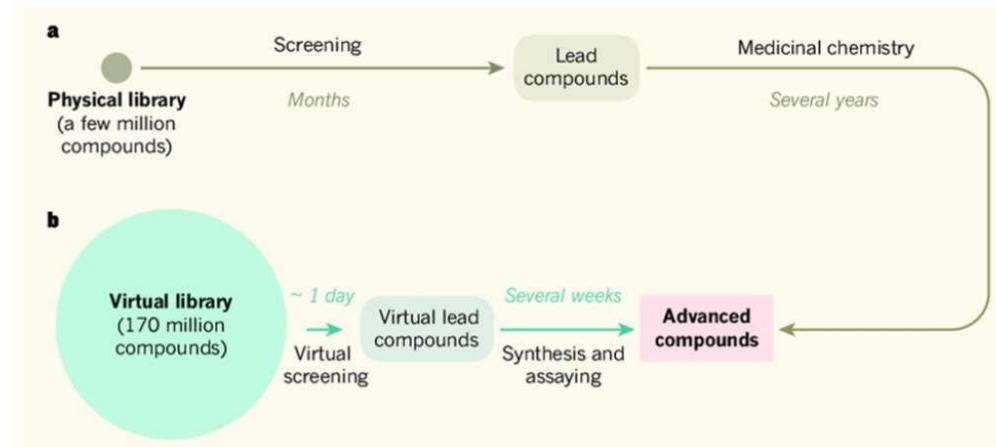
## • 스크리닝

: 화합물을 찾는 것

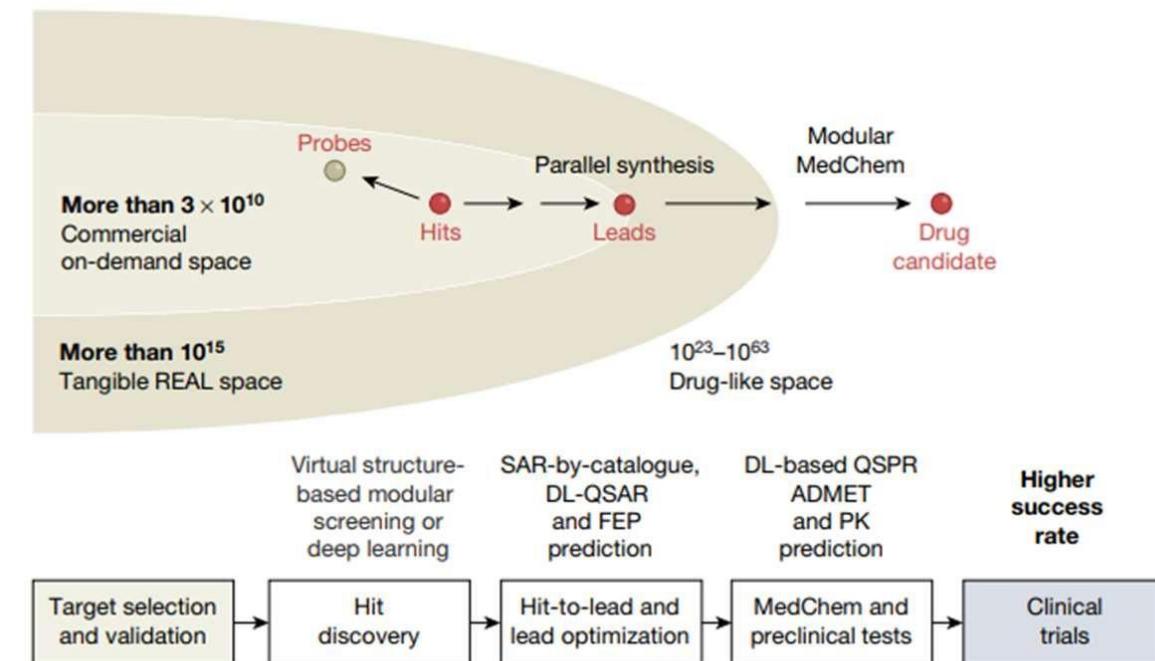




Type	Libraries	# of cpds	기간
<b>In-stock</b>	<b>KCB</b> ( <a href="https://chembank.org/">https://chembank.org/</a> )	<b>480,000</b>	<b>1~2 weeks</b>
	<b>MolPort</b> ( <a href="https://www.molport.com/shop/index">https://www.molport.com/shop/index</a> )	<b><math>4 \times 10^6</math></b>	<b>&lt; 4 weeks</b>
	<b>eMolecules</b> ( <a href="https://www.emolecules.com/">https://www.emolecules.com/</a> )	<b><math>2 \times 10^7</math></b>	<b>2, 4, 8 weeks</b>
	<b>mcule</b> ( <a href="https://mcule.com/">https://mcule.com/</a> )	<b><math>4 \times 10^7</math></b>	<b>-</b>
<b>On-demand</b>	Enamine REAL Database	<b><math>6 \times 10^9</math></b>	<b>3-4 weeks</b>
	Enamine REAL Space	<b><math>3.3 \times 10^{10}</math></b>	<b>3-4 weeks</b>
	Wuxi	<b><math>10^9</math></b>	<b>4, 8 weeks</b>
<b>Generative</b>	Using Deep Learning		



### Computationally driven timeline (2–12 months)



### Standard gene-to-lead discovery (4–6 years)

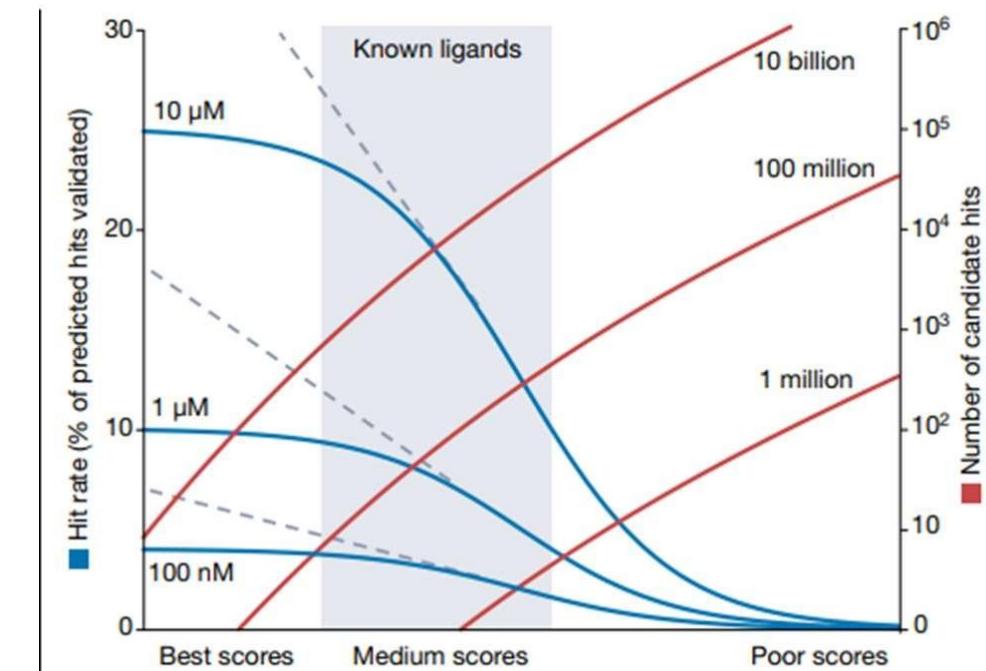


Nature, 2023, 616, 673–685

	HTS	Fragment-based ligand discovery	Gigascale DEL screening	Gigascale VLS
Initial library size	$10^5\text{--}10^7$	$10^3\text{--}10^5$	$10^{10}$	$10^{10}\text{--}10^{15}$
Hit rate (%)	0.01–0.5	1–5	0.01–0.5	10–40 <sup>a</sup>
Expected initial hit affinity	Weak (1–10 μM)	Very weak (100–1,000 μM) small fragments	Medium (0.1–10 μM)	Medium-high (0.01–10 μM)
Further steps to lead identifications	SAR by custom synthesis, QSAR-driven optimization	Merging or growing of fragments, structure-based and QSAR optimization	Label-free hit resynthesis, QSAR-driven optimization with custom synthesis	Extensive SAR-by-catalogue, structure-based and QSAR optimization
Expected number of custom syntheses to lead	500–1,000	500–1,000	200–500	0–50 (mostly on demand or easy parallel synthesis)
Composition of matter patentability	Hits are not novel, need modifications or scaffold hopping to achieve IP novelty	Fragment hits are not novel, require rational design to achieve IP novelty	Depends on the DNA-encoded library	Most hits are not previously synthesized and have IP novelty
Limitations	Modest library size, unknown binding mode, expensive equipment	Expensive NMR, X-ray and BIACORE equipment, many optimization steps	Many false positives, off-DNA resynthesis of hits needed	Computational resources (but reduced more than 1,000 times by modular VLS)

IP, intellectual property. <sup>a</sup>Fraction of predicted candidate hits that were confirmed experimentally.

Nature, 2023, 616, 673–685



**Fig. 2 | Benefits of a bigger chemical space.** The red curves in log scale

Type	Approach	Scalability	Applications	Requirements	Examples
Protein structure based	Fast empirical docking	$10^6\text{--}10^9$	Separate ligands from non-binders	High-resolution structures	DOCK <sup>54</sup> , GOLD <sup>149</sup> , AutoDock <sup>55</sup>
	Molecular mechanics based	$10^6\text{--}10^8$	Separate ligands from non-binders	High-resolution structures	ICM docking <sup>52</sup> , ROSETTALigand <sup>53</sup> , Glide <sup>56,57</sup>
	Flexible receptor docking	$10^3\text{--}10^5$	Separate ligands from non-binders	Medium-resolution structures	IFD-MD <sup>150</sup>
	Modular VLS	$10^{10}\text{--}10^{15}$	Separate ligands from non-binders	High-resolution structures	V-SYNTHES <sup>26</sup> , Chemical Space Docking <sup>151</sup>
	Free energy calculations	$10^2\text{--}10^3$	Affinity ranking	High-resolution structures	FEP <sup>+112</sup> , AB-FEP <sup>113,114</sup>
	QM/MM	$10^1\text{--}10^3$	Ion binding, transition state	High-resolution structures	Reviewed in ref. 152
Ligand based	2D/3D QSAR	Up to $10^8$	Screening and optimization	Ligand activity large datasets	AutoQSAR <sup>153</sup> , APF <sup>154</sup>
	3D pharmacophore and APF screening	Up to $10^{10}$	Screening	Ligand activity data	Reviewed in ref. 155, RIDE <sup>98</sup>
	ML/DL-QSAR	Up to $10^{10}$	Screening and affinity predictions	Ligand activity large datasets	Q.E.D <sup>78</sup> , LSTM-NN <sup>156</sup>
	Chemical space search	Up to $10^{26}$	Selection of analogues	Starting ligand (or ligands)	InfiniSee <sup>45</sup>
	QSPR-DL	Up to $10^{10}$	Predict solubility, lipophilicity, bioavailability, brain permeability, among others	Large datasets on ligand properties	AstraZeneca PK prediction <sup>73</sup> , prediction of oral bioavailability <sup>72-74</sup>
Hybrid	3D interaction fingerprints	Up to $10^{10}$	Improved docking and ligand selection	Data on ligand activity and protein-ligand 3D complexes	SIFT <sup>157</sup>
	3D/graph DL	$10^6\text{--}10^9$	Affinity prediction	Data on ligand activity and protein-ligand 3D complexes	Graph-CNN <sup>82,83</sup> , 3D-CNN <sup>84,85</sup>
	Dock/DL iterations	$10^8\text{--}10^{10}$	Separate ligands from non-binders	High-resolution structures	MolPal <sup>25</sup> , active learning <sup>110</sup> , deep docking <sup>111</sup>
	Dock to AI 3D protein models	$10^6\text{--}10^8$	Separate ligands from non-binders	Protein target sequence	AlphaFold <sup>99,100</sup> , RosettaFold <sup>101</sup>
	DL-based 3D score function	$10^6\text{--}10^8$	Separate ligands from non-binders	High-resolution structures	RT-CNN <sup>98</sup>

## • 구조기반 가상 스크리닝

: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아내는 것

- HTS vs. Virtual Screening
- Target-based vs. Phenotype-based
- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket
- Protein-ligand Interaction/Binding affinity

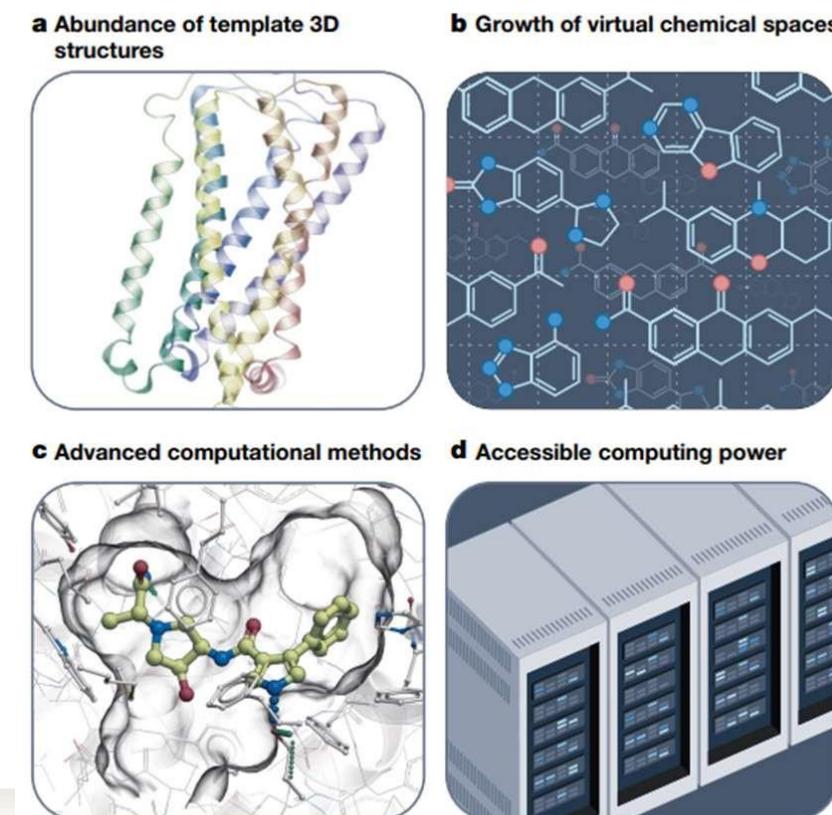
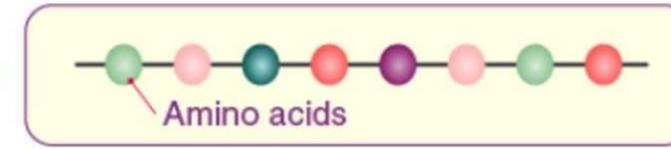
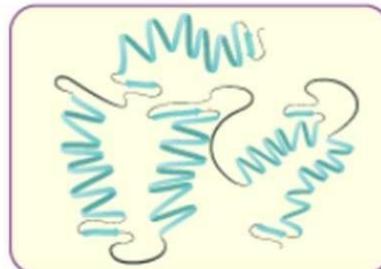
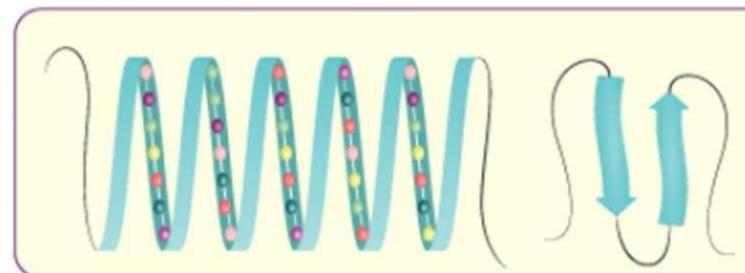


Fig. 1 | Key factors driving VLS technology breakthroughs for generation of high-quality hits and leads. a, More than 200,000 protein structures in the

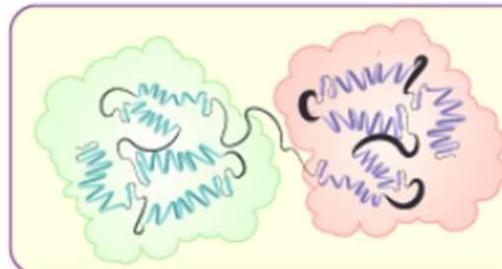
Nature, 2023, 616, 673

## Different Types of Protein Structure

Primary Structure

Secondary Structure ( $\alpha$ -Helix)

Tertiary Structure



Quaternary Structure

A	S
U	O
V	T
R	P

Alphabets



Primary Structure

Antibodies	Science
Useful	Routine
Vital	Tool
Rely	Practices

Words



Secondary Structure

Antibodies are a useful tool.  
Routine practices rely on antibodies.  
Science is vital.

Sentences



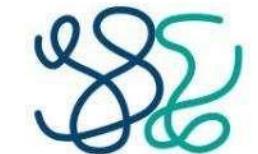
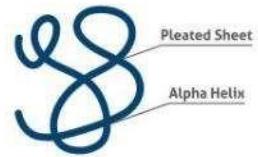
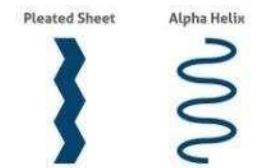
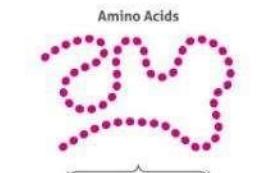
Tertiary Structure

Antibodies are a useful tool in the advancement of research. Routine practices such as Western Blot, Immunohistochemistry, Immunofluorescence, Immunoprecipitation and Flow Cytometry, rely on antibodies.

Paragraphs



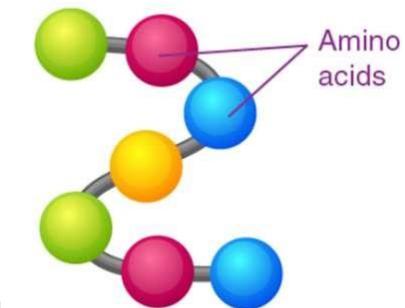
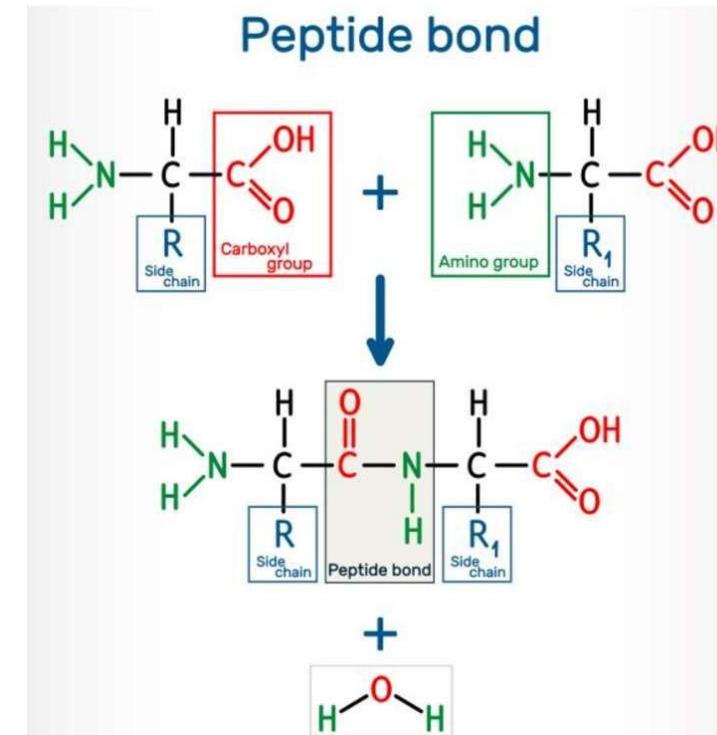
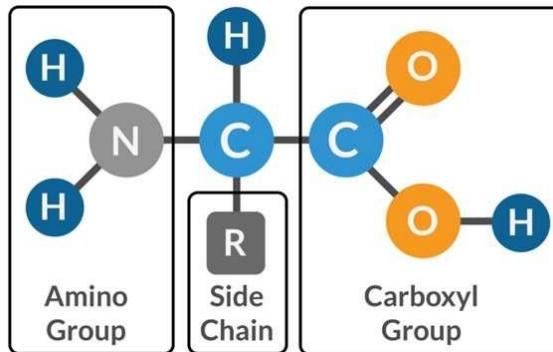
Quaternary Structure



<https://byjus.com/chemistry/protein-structure-and-levels-of-protein/>

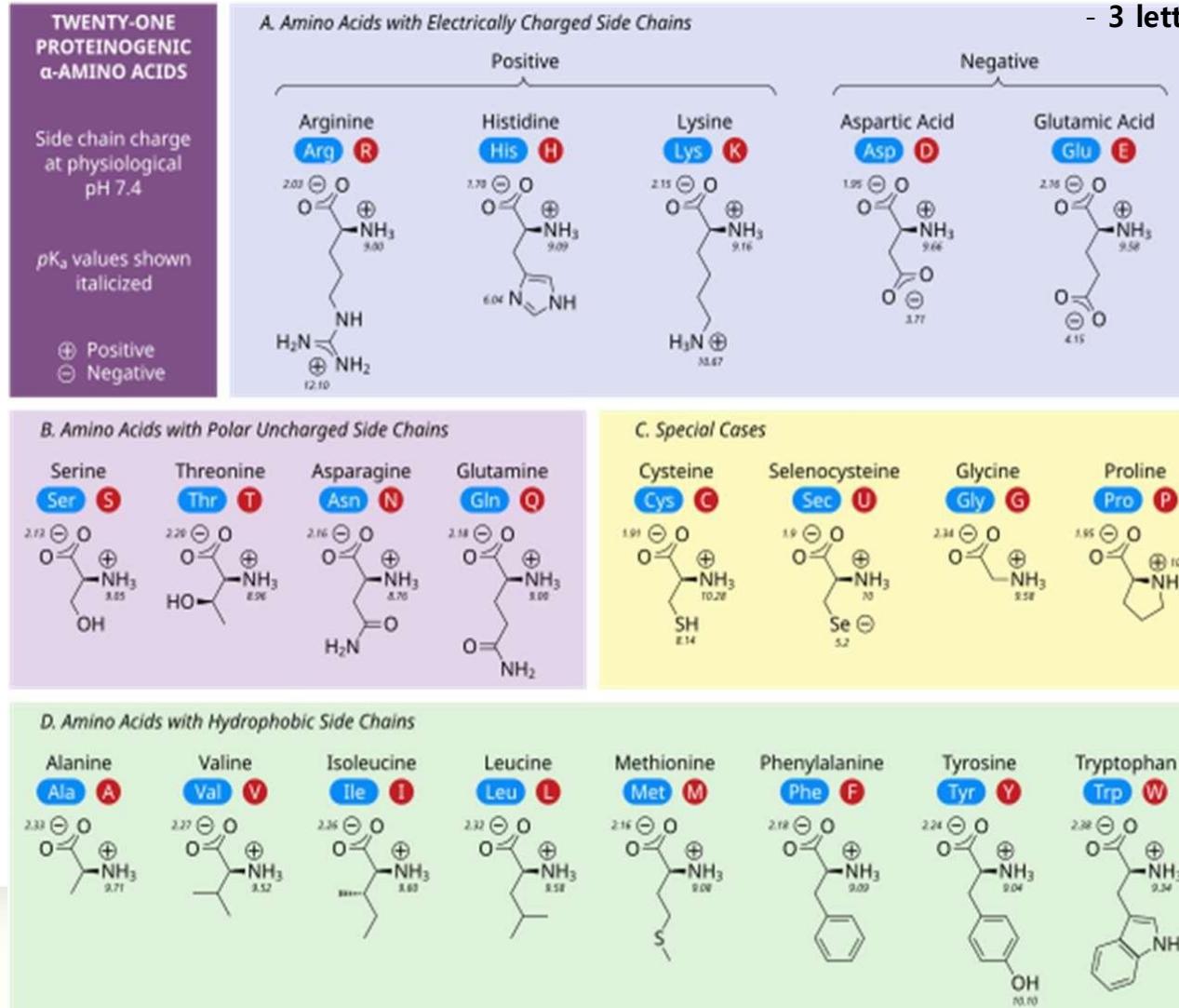
<https://www.ptglab.com/news/blog/the-complexity-of-proteins/>

- Primary Structure: Sequence



<https://www.reagent.co.uk/blog/what-are-amino-acids/>  
<https://biologydictionary.net/protein-structure/>

## • Primary Structure: Sequence



- 1 letter: R  
- 3 letter: Arg

## Three-/one-letter Amino Acid Codes

Name (not necessary):

Amino acid sequence:

(case insensitive, all symbols, except standard symbols of amino acids and stop-codons ("\*", "\*\*\*" and "end") are disregarded.

20  amino acids in one line;

capital letters (for one letter code);

print out the original sequence.

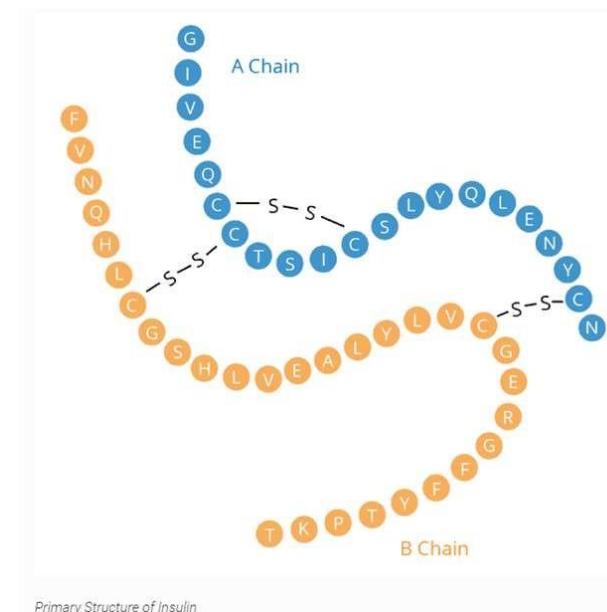
For the translation of nucleic acids use either of the "[Six-Frame Translation](#)" or "[Rare Codon Search](#)" tools.

[https://www.bioline.com/media/calculator/01\\_17.html](https://www.bioline.com/media/calculator/01_17.html)

## • Primary Structure: Sequence

: A primary protein is a simple, linear chain of amino acids (AKA a polypeptide chain).

### < Primary Structure of Insulin >

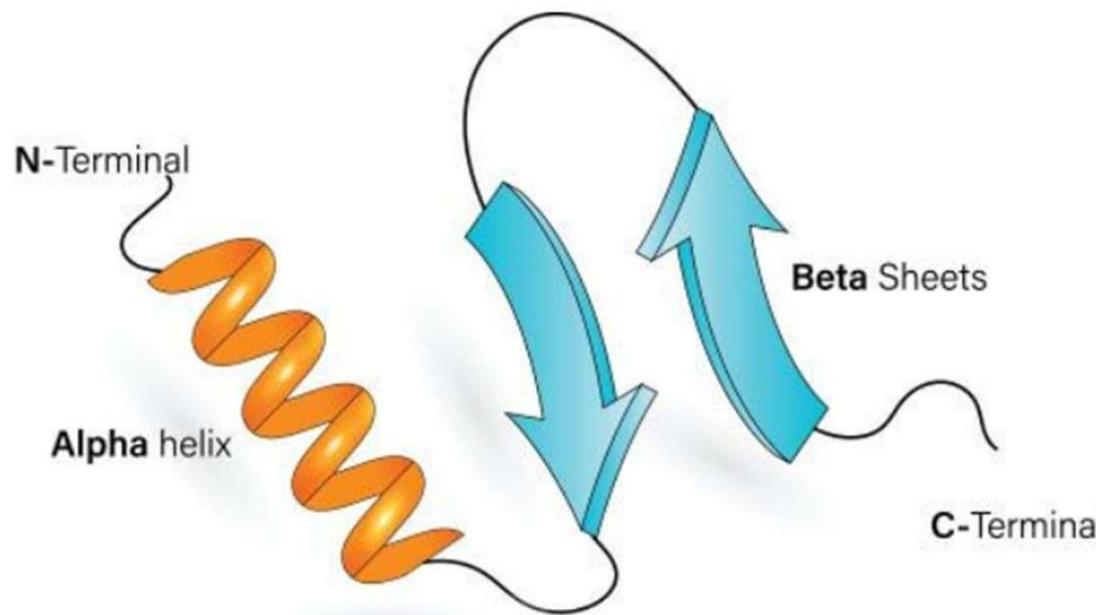


<https://www.connectedinmotion.ca/blog/crystals-and-chains-the-work-of-frederick-sanger-dorothy-hodgkin-in-understanding-the-structure-of-insulin/>

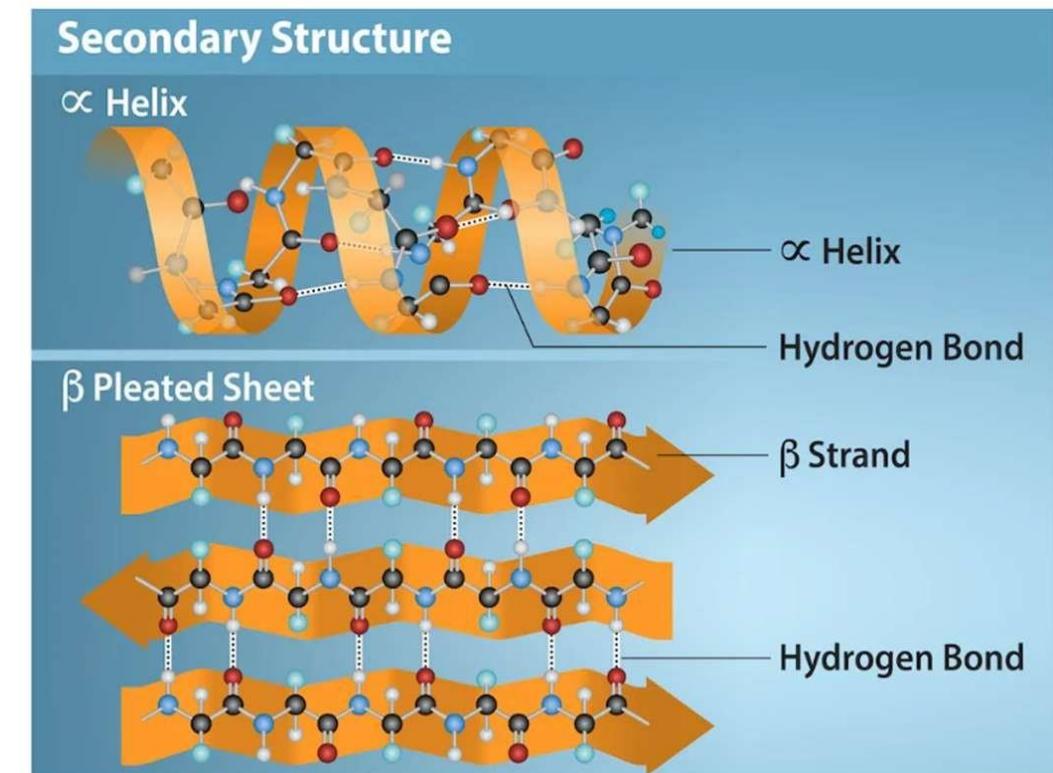
## • Secondary Structure: Sequence

The secondary protein structure is made by ~~hydrogen~~

There are two main types of secondary protein structures: the  $\alpha$ -helix and the  $\beta$ -pleated sheet.



<https://biologydictionary.net/protein-structure/>



<https://harpercollege.pressbooks.pub/chm100/chapter/secondary-tertiary-and-q-uaternary-structure-of-proteins/>

- **Tertiary Structure**

A protein isn't fully functional until it has a 3D shape.

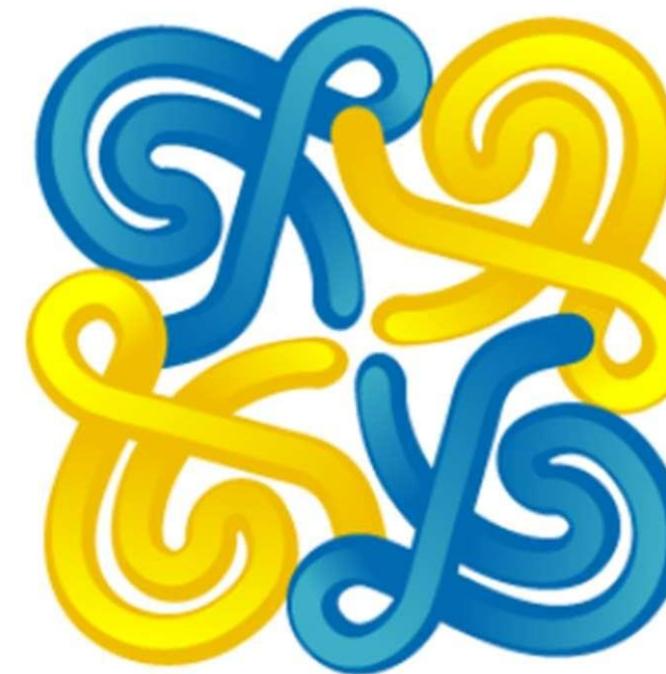
The 3D structure of a protein is referred to as its tertiary structure and is made by further folding of secondary proteins.



[https://biologydictionary.net/protein-structure//](https://biologydictionary.net/protein-structure/)

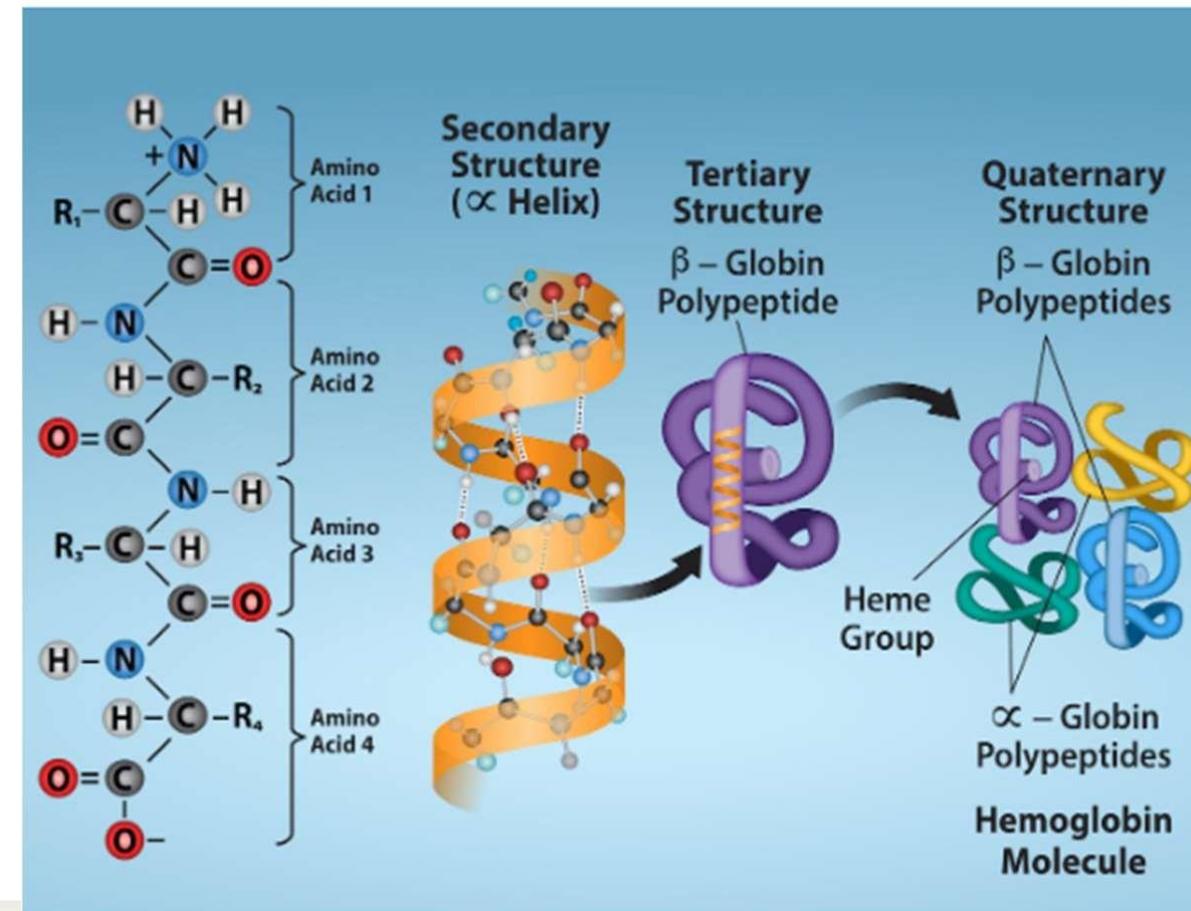
## Quaternary Structure

Many proteins are made of a single polypeptide chain and don't become any more complex than their tertiary structure. However, some proteins are made up of multiple polypeptide chains. When several polypeptide chains (AKA subunits) come together, they can form a structure known as a quaternary protein.



<https://biologydictionary.net/protein-structure/> <https://byjus.com/chemistry/protein-structure-and-levels-of-protein/>

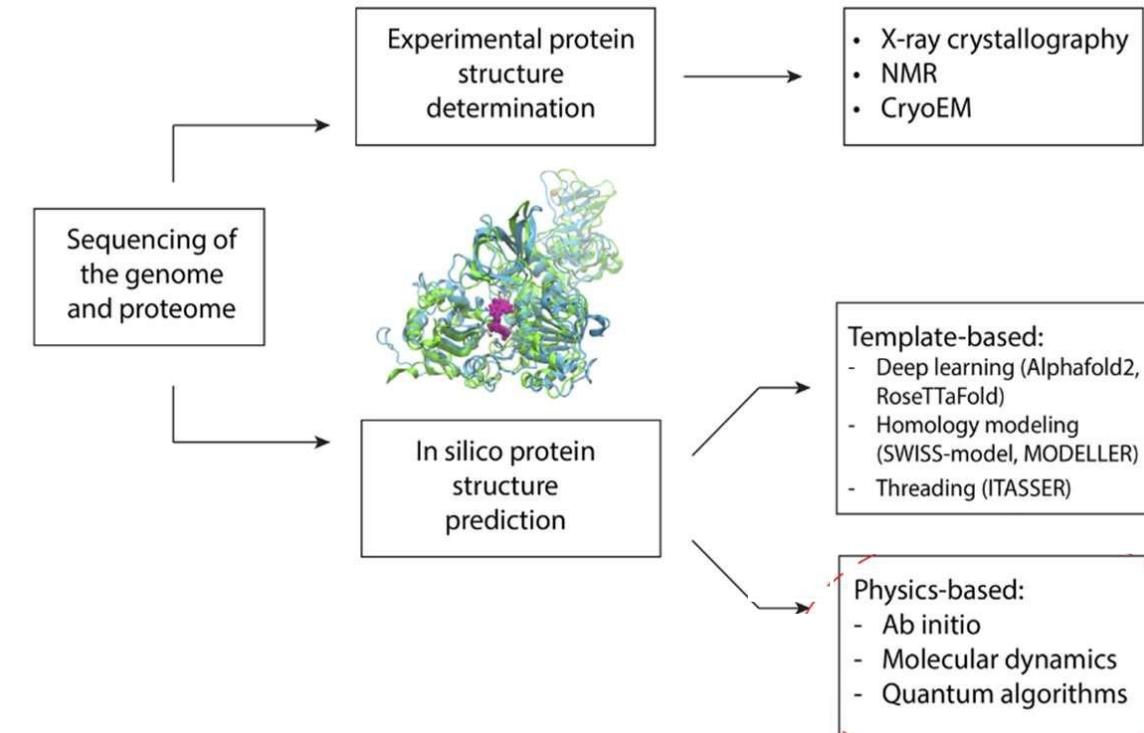
## Different Types of Protein Structure



[https://chem.libretexts.org/Courses/Roosevelt\\_University/General\\_Organic\\_and\\_Biochemistry\\_with\\_Problems\\_Case\\_Studies\\_and\\_Activities/14%3A\\_Pr.../14.04%3A\\_Secondary\\_Tertiary\\_and\\_Quaternary\\_Structure\\_of\\_Proteins](https://chem.libretexts.org/Courses/Roosevelt_University/General_Organic_and_Biochemistry_with_Problems_Case_Studies_and_Activities/14%3A_Pr.../14.04%3A_Secondary_Tertiary_and_Quaternary_Structure_of_Proteins)

## 1. Experimental

- 1) X-ray crystallography
- 2) Nuclear Magnetic Resonance
- 3) Electron Microscopy

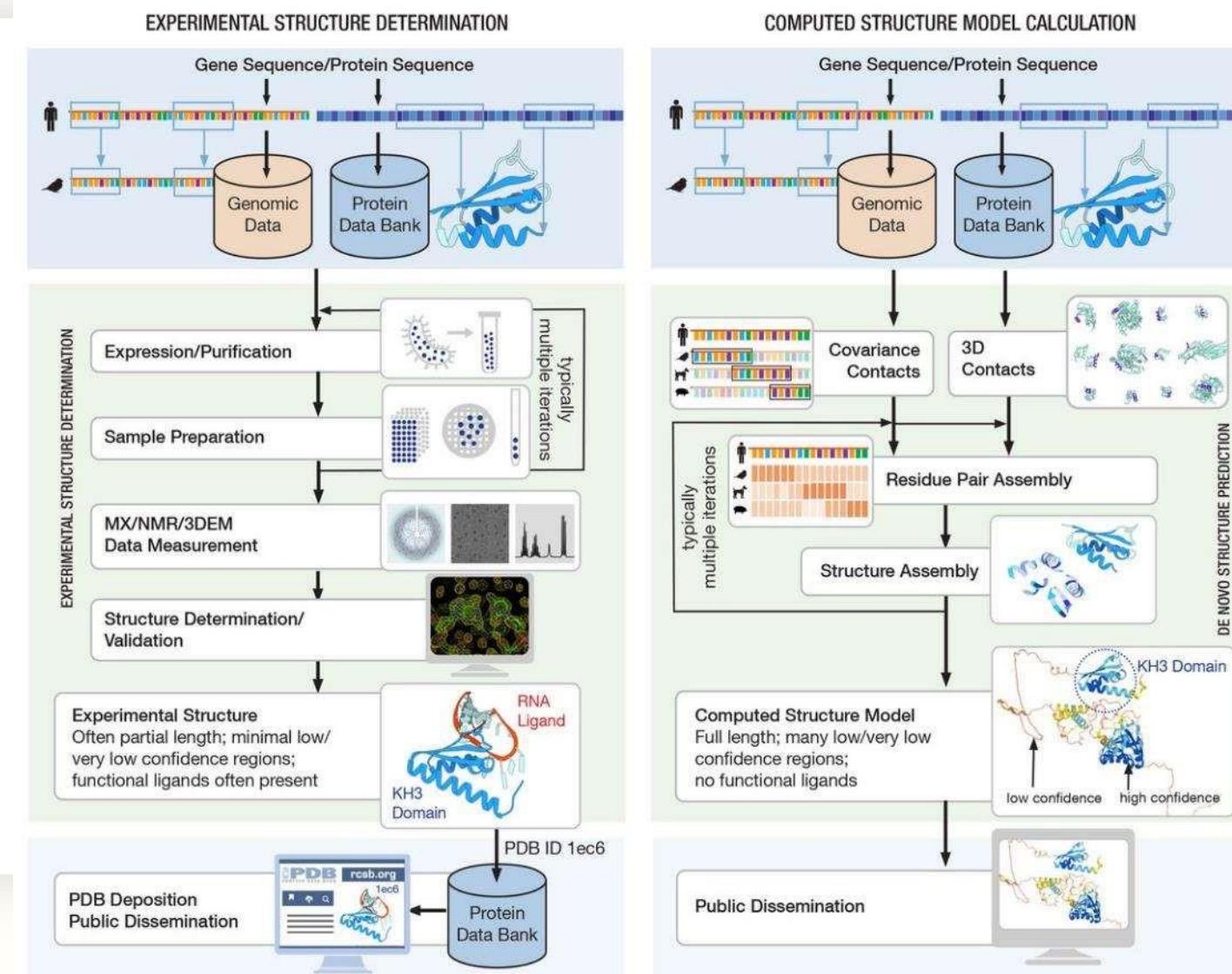


## 2. In silico

- 1) Homology Modeling  
: SWISS-model, MODELLER etc.
- 2) Deep learning  
AlphaFold, RoseTTAFold, ESMFold

### Protein Database

J. Chem. Theory Comput. 2024, 20, 3359-3378



- NMR Spectroscopy

: <https://www.youtube.com/watch?v=t8JIBwLsCGA&t=3s>

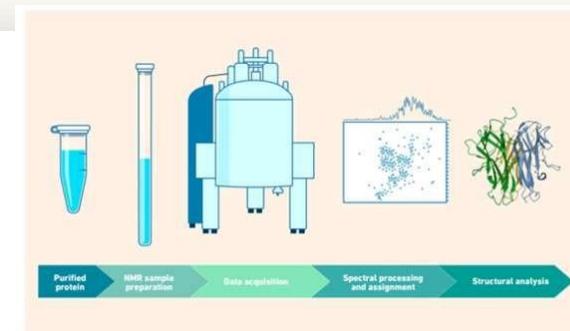


Figure 4: NMR spectroscopy workflow for protein structural determination. Credit: Technoloev Networks.

- X-ray crystallography

: <https://www.youtube.com/watch?v=RUok4O9oovQ&t=3s>

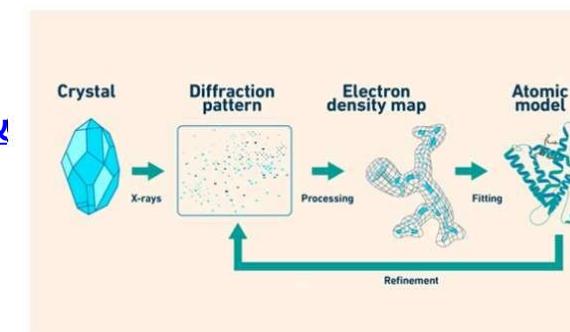


Figure 2: Steps for protein structure determination by X-ray crystallography. Credit: Technology Networks.

- Cryo-EM

: <https://www.youtube.com/watch?v=vLo7oqfRa74>

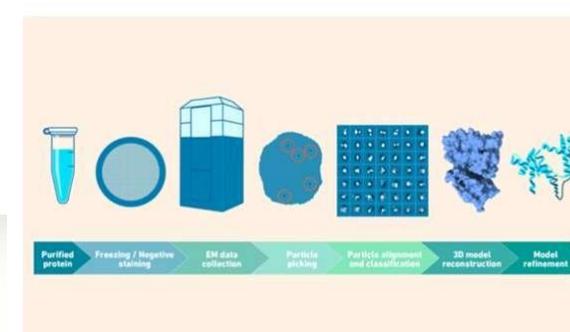


Figure 1: Workflow of protein structure elucidation by cryo-EM. Credit: Technology Networks.

<https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure#xray> <https://www.technologynetworks.com/analysis/articles/key-techniques-in-structural-biology-their-strengths-and-limitations-370666.html>

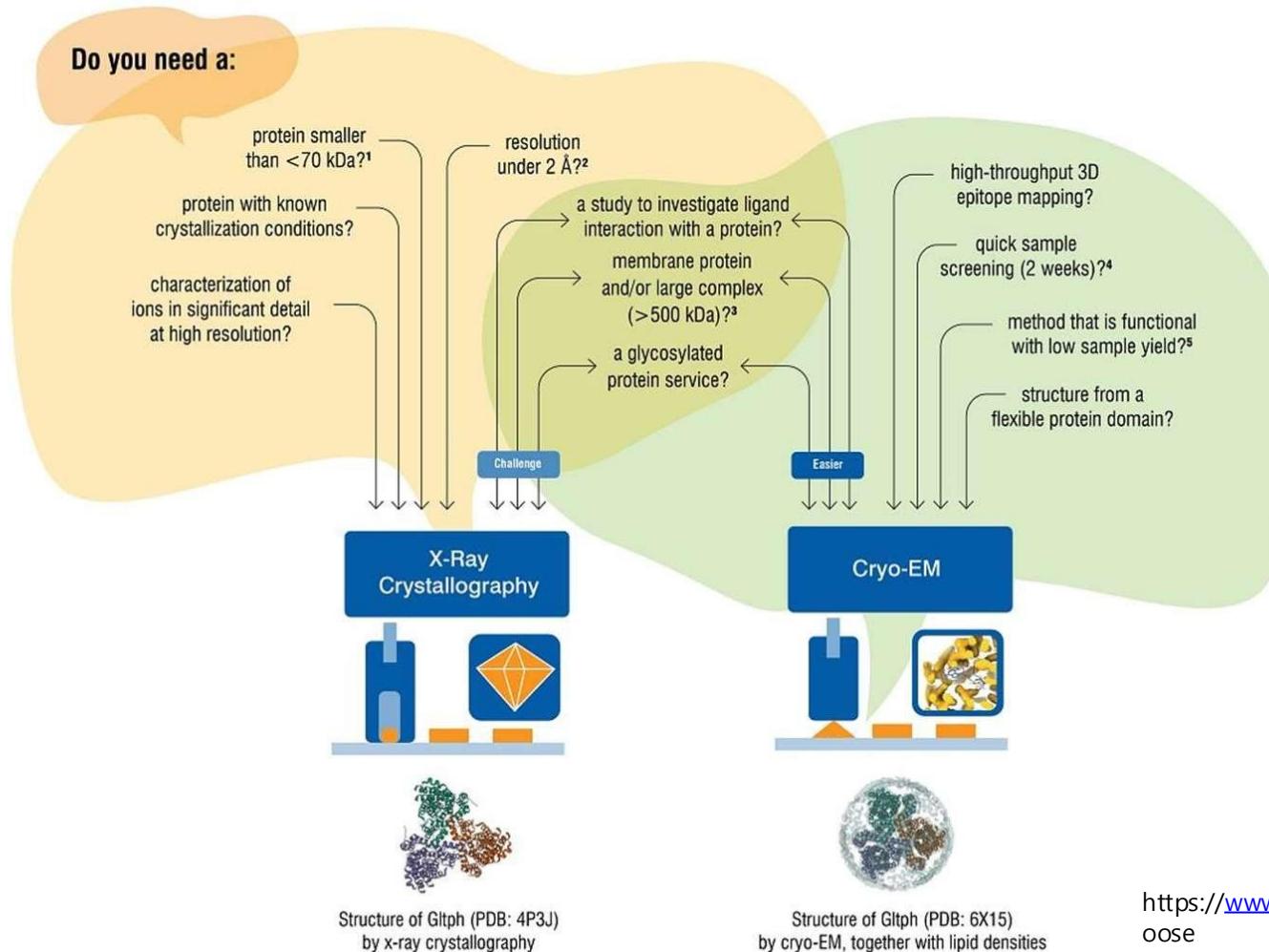
## < Comparison of different Methods for Protein Structure determination >

Method	Size (sample state)	Resolution Limits	Amounts	Advantages	Disadvantages
NMR	< 100 kDa (solution)	~ 3-4 Å	µmoles/milligrams	<ul style="list-style-type: none"> <li>High resolution</li> <li>3D structure in solution good <b>for dynamic study</b></li> </ul>	<ul style="list-style-type: none"> <li>Difficult sample prep</li> <li>High sample purity need</li> <li>Static crystalline state capture</li> <li>Only small proteins</li> </ul>
X-ray	Limited by crystal quality	< 1-3 Å	µmoles/milligrams	<ul style="list-style-type: none"> <li>Well developed</li> <li>High resolution</li> <li><b>Broad molecular weight range</b></li> <li><b>Atomic resolution</b></li> </ul>	<ul style="list-style-type: none"> <li>Difficult sample prep</li> <li>Static crystalline state</li> </ul>
Cryo-EM	< 100 kDa (verified ice)	Mostly > 3 Å	nanomoles/µgrams	<ul style="list-style-type: none"> <li><b>Simple sample preparation</b></li> <li><b>Structure in native state</b></li> <li><b>Small sample size needed</b></li> </ul>	<ul style="list-style-type: none"> <li>Lower resolution</li> <li>Works best for samples with high molecular weight</li> <li>Equipment can be expensive, but costs are decreasing</li> </ul>

[https://www.nlm.nih.gov/ncbi/workshops/ASCB\\_2023-04\\_3d-molecular-structures/background.html](https://www.nlm.nih.gov/ncbi/workshops/ASCB_2023-04_3d-molecular-structures/background.html) [https://www.researchgate.net/figure/Overview-of-different-Methods-used-for-protein-structure-determination\\_tbl1\\_353730158](https://www.researchgate.net/figure/Overview-of-different-Methods-used-for-protein-structure-determination_tbl1_353730158)

J. Mol. Biol. 2020, 432, 2973-2984

## Discover Which Method Best Suits Your Project



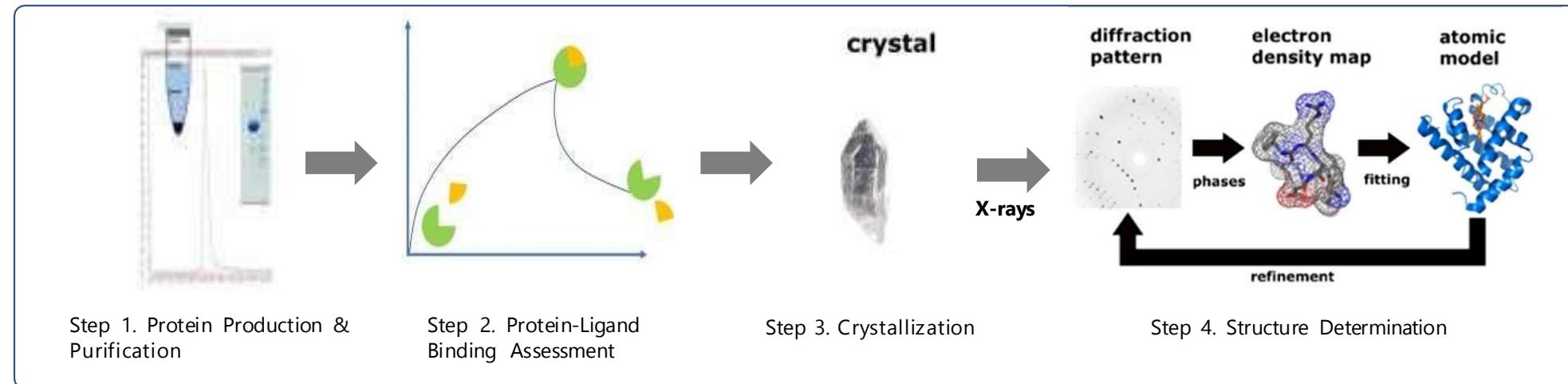
- 1) Small proteins are more suitable for x-ray crystallography since they can be difficult to discern by cryo-EM.
- 2) X-ray crystallography can obtain very high resolutions, although cryo-EM has had comparable results in recent years.
- 3) Large structures, complexes, and membrane proteins can be difficult to crystallize. Cryo-EM might therefore be an easier option.
- 4) EM (negative stain) offers rapid screening of samples to rule out aggregation and determine oligomerization, giving a visual overlook of how the sample behaves and what it contains.
- 5) Cryo-EM requires lower protein amounts than x-ray crystallography. This method might benefit samples with a yield of under 2 mg protein (<5-10 mg/ml).\*

<https://www.criver.com/resources/x-ray-crystallography-or-cryo-em-which-solution-should-you-choose>

# X-ray crystallography: CRO info & Tips

업체		SARomics	Proteros	Creative biostructure
사이트		스웨덴 <a href="https://www.saromics.com/">https://www.saromics.com/</a>	독일 <a href="https://www.proteros.com/">https://www.proteros.com/</a>	미국 <a href="https://www.creative-biostructure.com/">https://www.creative-biostructure.com/</a>
소요 시간	단백질 발현	7-12 weeks	8 weeks	16-18 weeks
	결정화&구조규명	4-12 weeks	6 weeks	11 weeks
	소요시간	<b>12-24 weeks</b>	<b>14 weeks</b>	<b>27-29 weeks</b>

## X-ray crystallography: CRO info & Tips



Task	비용 지급	Protein-Ligand
1. 단백질 확보	착수 시: \$#,###	단백질 확보 및 purification
	확보 시: \$#,###	
2. 단백질과 화합물 Binding	착수 시: \$#,###	DSF assay를 통해 결합확인
3. 단백질과 화합물의 3D co-crystal	착수 시: \$#,###	단백질과 화합물의 결정 구조 확보
	확보 시: \$\$#,###	3차원 구조 결정
	\$#,###	단백질과 화합물 1종의 결합구조 확보

# 03 구조기반 신약 후보물질 탐색

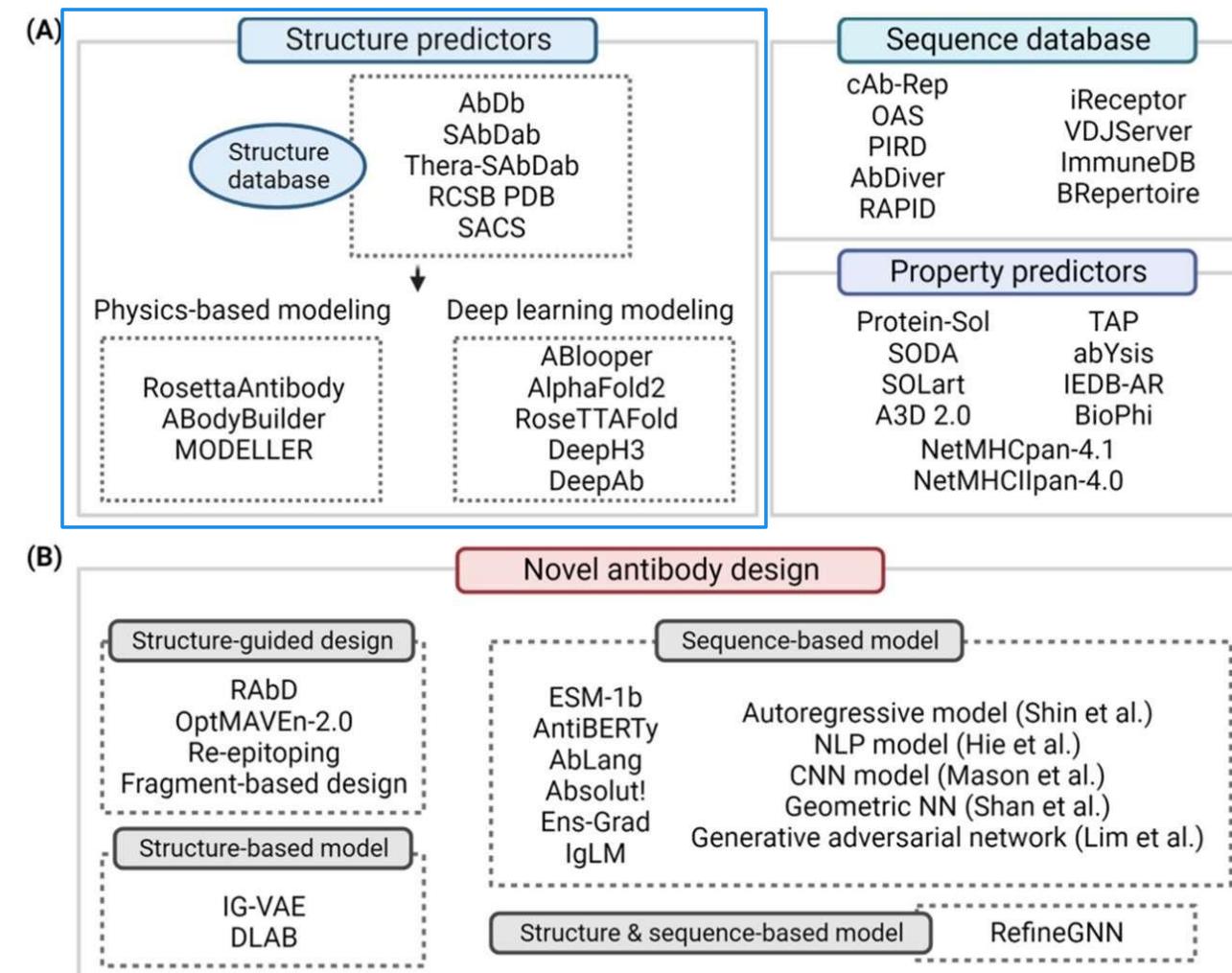
## 1. Experimental

- 1) X-ray crystallography
- 2) Nuclear Magnetic Resonance
- 3) Electron Microscopy

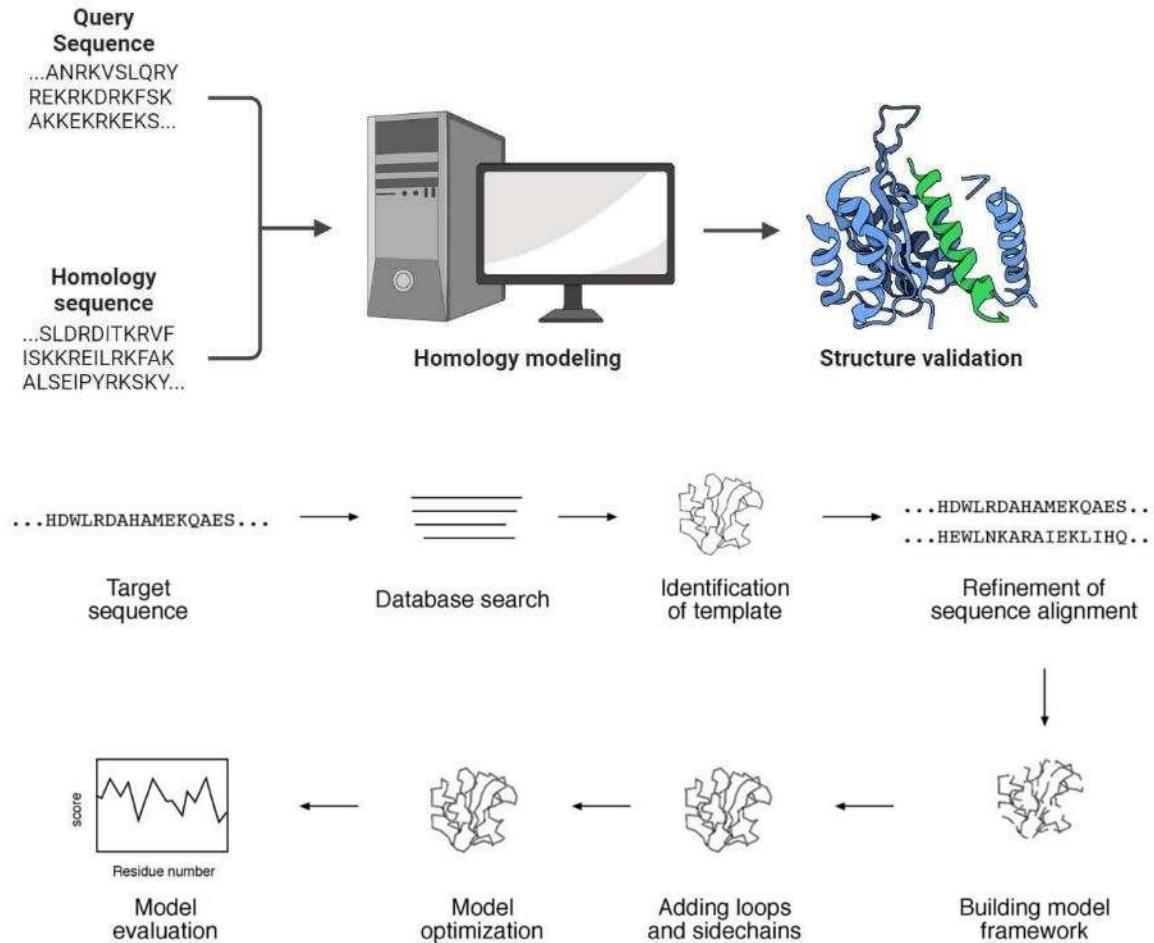
## 2. In silico

- 1) Homology Modeling  
: SWISS-model, MODELLER etc.
  - 2) Deep learning  
**AlphaFold, RoseTTAFold, ESMFold**
- Protein Database

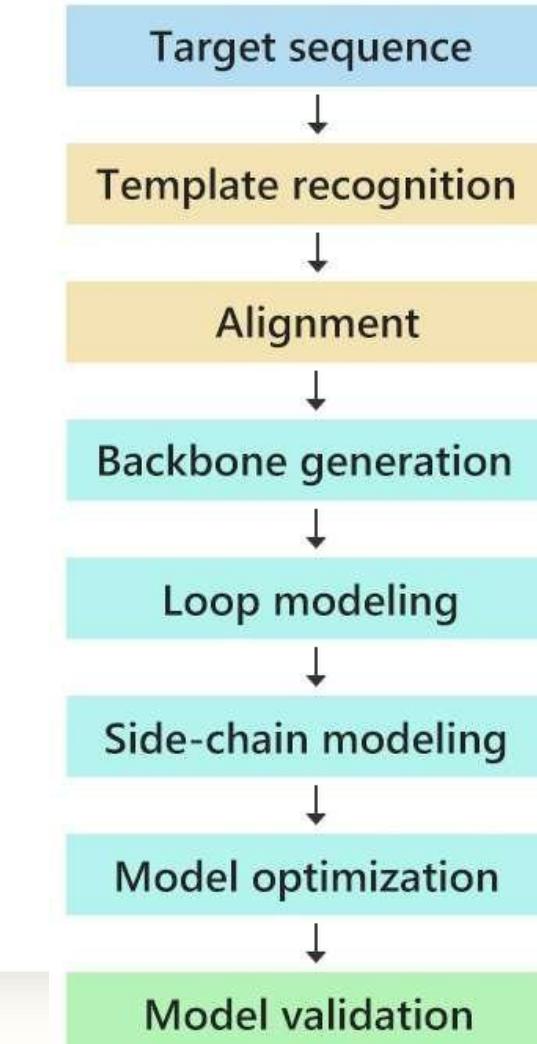
## < Recently developed computational methods for Protein & Antibody design >



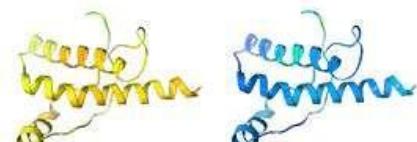
# Homology Modeling



<https://microbenotes.com/homology-modeling/>

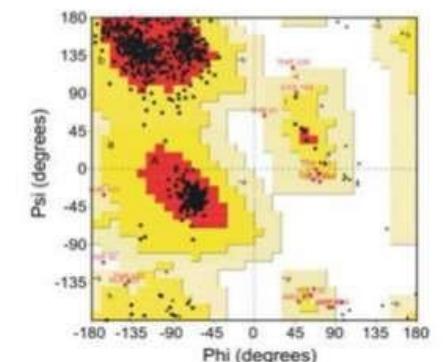
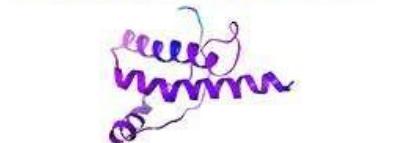


VSIDTMRAADVARAALQNLQNGAQMVNDFGTHRYTA.



VSIDTMRAADVARAALQNLQNGAQMVNDFGTHRYTA  
VSTDTMRAEVAKARIIONGQOMVNDFGTHRYTA  
VSI-TMRAEVAKARIIONGQOMVNE--THRYTI  
VSIDTMRAADVARAALQNLQNGAQMVNDFGTHRYSA

VSIDTMRAADVARAALQNLQNGAQMVNDFGTHRYTA  
VSTDTMRAEVAKARIIONGQOMVNDFGTHRYTA  
VSI-TMRAEVAKARIIONGQOMVNE--THRYTI  
VSIDTMRAADVARAALQNLQNGAQMVNDFGTHRYSA



# Multiple Sequence Alignment (MSA)

**Align**

Find a protein sequence by UniProt ID (e.g. P05067 or A4\_HUMAN or UPI0000000001) to align with the Clustal Omega program.

You can also paste a list of IDs.

UniProt IDs:

Enter multiple protein or nucleotide sequences, separated by a FASTA header. You may also load from a text file.

Protein or nucleotide sequences in FASTA format:

```
SP|Q00534|CDK6_HUMAN MEKDGLCRADQQYECVAEIGEGAYGKVFKA RD LKNGGRFVALKRV RV QTG--EEGMPLSTIREVA
SP|P11802|CDK4_HUMAN -----MATSRYEPVAEIGVGA YGT VYKARDPH-SGHFVALKSVRVPGGGGGGLPISTVREVA
SP|Q8IZL9|CDK20_HUMAN -----MDQYCILGRIGE GA HGVFKAKHVE-TGEIVALKKVALRLL--EDGFPNQALREIK
SP|P50613|CDK7_HUMAN -----MA LDVKSRAKRYEKLDFLGE GGQFATVYKARDKN-TNQIVAIKKIKLGHRS EAKDGINRTALREIK
SP|Q00535|CDK5_HUMAN -----MQKYEKLEKIGEGTYGTVFKA KNRE-THEIVALKRVR LDDD--DEGV PSSALREIC
SP|P24941|CDK2_HUMAN -----MENFQKVEKI GEGTYGVVYKARNKL-TGEVVALKKIRLDTE--TEGV PSTAIREIS
```

Name your Align job:

Advanced parameters

Output sequence order: from alignment Iterations: 0

Run Align

## Align results

Overview Trees Percent Identity Matrix Text Output Input Parameters API Request

Tools Download Add Resubmit

Phylogenetic tree

Tree type:  Phylogenetic tree  Guide tree Layout:  Horizontal  Circular Branch length:  Phylogram with aligned labels  Cladogram

	sp Q00534 CDK6_HUMAN	sp P11802 CDK4_HUMAN	sp Q8IZL9 CDK20_HUMAN	sp P50613 CDK7_HUMAN	sp Q00535 CDK5_HUMAN	sp P24941 CDK2_HUMAN
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
70.33%	70.33%	37.02%	37.67%	44.76%	44.98%	48.81%
37.38%	37.02%	100.00%	40.00%	46.05%	45.86%	48.81%
34.41%	37.67%	40.00%	100.00%	45.86%	45.86%	43.92%
46.85%	44.76%	46.05%	45.86%	100.00%	100.00%	59.79%
48.81%	44.98%	42.42%	43.92%	59.79%	59.79%	100.00%

## Align results

Overview Trees Percent Identity Matrix Text Output Input Parameters AI

Tools Download Add Resubmit

Percent Identity Matrix

	sp Q00534 CDK6_HUMAN	sp P11802 CDK4_HUMAN	sp Q8IZL9 CDK20_HUMAN	sp P50613 CDK7_HUMAN	sp Q00535 CDK5_HUMAN	sp P24941 CDK2_HUMAN
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
70.33%	70.33%	37.02%	37.67%	44.76%	44.98%	48.81%
37.38%	37.02%	100.00%	40.00%	46.05%	45.86%	42.42%
34.41%	37.67%	40.00%	100.00%	45.86%	45.86%	43.92%
46.85%	44.76%	46.05%	45.86%	100.00%	100.00%	59.79%
48.81%	44.98%	42.42%	43.92%	59.79%	59.79%	100.00%

UniProt BLAST Align Peptide search ID mapping SPARQL Tool results

Overview Trees Percent Identity Matrix Text Output Input Parameters API Request

Tools Download Add Resubmit

Highlight properties Select annotation View: Overview Wrapped

Q00534:Chain

```
MEKDGLCRADQQYECVAEIGEGAYGKVFKA RD LKNGGRFVALKRV RV QTG--EEGMPLSTIREVA
-----MATSRYEPVAEIGVGA YGT VYKARDPH-SGHFVALKSVRVPGGGGGGLPISTVREVA
-----MDQYCILGRIGE GA HGVFKAKHVE-TGEIVALKKVALRLL--EDGFPNQALREIK
-----MA LDVKSRAKRYEKLDFLGE GGQFATVYKARDKN-TNQIVAIKKIKLGHRS EAKDGINRTALREIK
-----MQKYEKLEKIGEGTYGTVFKA KNRE-THEIVALKRVR LDDD--DEGV PSSALREIC
-----MENFQKVEKI GEGTYGVVYKARNKL-TGEVVALKKIRLDTE--TEGV PSTAIREIS
```

Q00534:Chain

```
VLRHLETFEHPN-VVRLFDVCTVSRTDRET KLT LVFEHV DQDL TTYLDKVPEPGVP TETIKDMMFQ
LRRLEAFEHPN-VVRLMDVCATSRTDREIKVTLVFEHV DQDL RTYLDKAPPGLPAETIKDLMRQ
AQE---MEDNQYVYVQLKAMFP----HGGGVLA FE FMLS DAEVVRHAQ-RPLAQAVQKSYLQM
LHQE---LSHPN-IIGLLDAFG---HHSNISLV FDFM ET DLEVI KDN S-LVLT PSHIKAYMLM
LKE---LKHKN-IVRLHDVLH----SDKKLT LVFE FCDQDLKKYFDSCN-GDL DPEIVKSFLFQ
LKE---LNHPN-IVKLLDVIH----TENKLYLVFE FFLH QDLKKFMDASALTGIPPLIKSYLFQ
```

Q00534:Chain

```
LRLGLDFLHS HRVVHRDLKPQNI LVTSSCQIKLADFG LARIYSF--QMA L TSVVVTLWYRAPEVLL
FLRGLDFLHANCIVHRDLKPEN I LVTSGGTVKLADFG LARIYSF--QMA L TPV VVTLWYRAPEVLL
LKGVAFCHANNIHRDLKPANL I SASGQLKIA DFG LARV FSPDGSR LYTHQVATRWYRAPELLY
T LQGLEYLHQHWILHRDLKPNNL LDENGVLKLA DFG LAKSFGS-PNRAYTHQV VTR RWYRAPELLF
LKGFLGFCSRNVLHRDLKPQNL LINRCNELKLA DFG LARAFGIPVRC HSAEV VVTLWYRPPDVLF
LQGLAFC SHRVLHRDLKPQNL INTEGAIKLA DFG LARAFGV-PVRTYTHEV VTLWYRAPEI L
```

Q00534:Chain

```
QS-SYATPVDLWSVGCIFAE MFR-RKPLFRGSSDV DQLGK ILDVIGLPG EEDWPRDVA LP RQAF-
QS-TYATPVDMWSVGCIFAE MFR-RKPLFCGNSEADQ LGK IFDLIGLPP EDDWPRDVS LPRGAF-
GARQYDQGV DLWSVGCIMGELLN-GSP LFPG KNDIEQLCYVRLI L GTPNPQVWE LTEL PDYNKIS
GARMYGVG VDMWA VGCILAE LLL RFP LPGD SLDL QTRI FETL GTPT EEQWP DMCS L PDY VTFK
GAKLYSTS I DMWSA GCI FAE LANAGRP LPGD NVD DQL KRI FRL L GTPT EEQWP SMTKLP DYKYP
GCKYYSTAV D IWSL GC I FAE MVT RRAL FP GDSE I DQL FRI FRTL GT PDEV VWP GVT S MPD YKPSF
```

Q00534:Chain

```
-HSKSAQPIEKFT D IDELGKD L L KCLTFNPAKRISAYSALSHPYFQDLERCKENLDSHLPPSQN
-PPRGPRPVQS VVPEMEESGAQ L L EMLTFNPHKRISAFRALQHSY LHKDEGNPE-----
FKEQVPM PLEEVLP DVSPQ ALD L L QGFL L YPPHQR IA ASK ALL H QY F FTAPL PAH PSEL PIP - QR
SF-PGI-PLHHIFSAG D L L QGFL L FNPCARITATO QL KM KYFSNR PGPT PG CQL PRP - NC
MY-PATTSLVN VVPKLNATGRD L L QNLL K C N P V Q R ISAE E A L Q H P Y F S D F C P P -
PK-WARQDFSK VV PPL DED GRS L L SQMLHYDPNKRISAKA A LA HP F F QD VTKPVPHRL-----
```

T SELNTA-----

```
LGGPAPKAHPGP PHI HD FH VDRP LEES L L NPEL IRPF I FILEG
P VETLKE-QSNPA---LA I K R K R T E A L E Q G G L P K K L I F -
```

44

<https://swissmodel.expasy.org/>

# SWISS-MODEL

is a fully automated protein structure homology-modelling server, accessible via the [Expasy web server](#).  
The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

[Start Modelling](#)

**Start A New Modelling Project**

Paste your target sequence(s) or UniProtKB AC here  
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

+ Upload Target Sequence File... [Validate](#)

Project Title: Untitled Project  
Email: Optional

[Search For Templates](#) [Build Model](#)

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding articles.

You are currently not logged in - to take advantage of the workspace, please [log in](#) or [create an account](#).

(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projects here.)

**Modelling Projects In Session**

Untitled Project [X](#)  
463 residues. Running  
Created: today  
(Expires: Next week) [X](#)

**SWISS-MODEL**

All Projects CDK7\_HUMAN P50613 Cyclin-dependent kinase 7 Created: today at 00:34

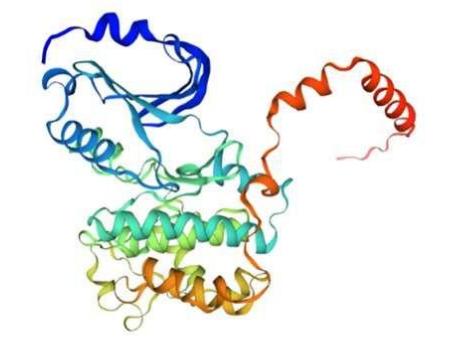
[Summary](#) [Templates 50](#) [Models](#) [Project Data](#)

**Template Results**

	Sort	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
<input checked="" type="checkbox"/> A0A2Y9PY83_1.A Cyclin-dependent kinase 7	AlphaFold DB model of A0A2Y9PY83_DELLE (gene: CDK7; organism: Delphinapterus leucas (Beluga whale))	0.83	-	97.11	AlphaFold v2	monomer ✓	None	
<input type="checkbox"/> 7egb_1.i Cyclin-dependent kinase 7	TFIID-based holo PIC on SCP promoter	0.82	-	100.00	EM	hetero-49-mer ▲	17 x ZN <sup>2+</sup> , 1 x SF4 <sup>2+</sup> , 1 x MG <sup>2+</sup>	
<input type="checkbox"/> 1ua2_1.A Cell division protein kinase 7	Crystal Structure of Human CDK7	0.76	-	100.00	X-ray, 3.0 Å	monomer ✓	1 x ATP <sup>2-</sup>	
<input type="checkbox"/> 1ua2_3.A Cell division protein kinase 7	Crystal Structure of Human CDK7	0.76	-	100.00	X-ray, 3.0 Å	monomer ✓	1 x ATP <sup>2-</sup>	
<input type="checkbox"/> 1ua2_2.A Cell division protein kinase 7	Crystal Structure of Human CDK7	0.75	-	100.00	X-ray, 3.0 Å	monomer ✓	1 x ATP <sup>2-</sup>	
<input type="checkbox"/> 8pyr_1.A Cyclin-dependent kinase 7	Crystal structure of the dual T-loop phosphorylated Cdk7/CycH/Matt complex	0.75	-	100.00	X-ray, 2.1 Å	hetero-tetramer ▲	None	
<input type="checkbox"/> 1ua2_4.A Cell division protein kinase 7	Crystal Structure of Human CDK7	0.75	-	100.00	X-ray, 3.0 Å	monomer ✓	1 x ATP <sup>2-</sup>	
<input type="checkbox"/> 7egc_1.i Cyclin-dependent kinase 7	p53-bound TFIID-based holo PIC on HDM2 promoter	0.74	-	100.00	EM	hetero-49-mer ▲	16 x ZN <sup>2+</sup> , 1 x SF4 <sup>2+</sup> , 1 x MG <sup>2+</sup>	
<input type="checkbox"/> 6091_13 Cyclin-dependent kinase 7	Human holo-PIC in the closed state	0.68	-	100.00	EM	hetero-31-mer ▲	2 x MG <sup>2+</sup> , 17 x ZN <sup>2+</sup> , 1 x SF4 <sup>2+</sup>	
<input type="checkbox"/> 2lw6_1.A CELL DIVISION PROTEIN KINASE 2	STRUCTURE OF HUMAN THR160-PHOSPHO CDK2-CYCLIN A COMPLEXED WITH A BISANILINOPYRIMIDINE INHIBITOR	0.65	-	45.42	X-ray, 2.3 Å	hetero-dimer ▲	1 x QO2 <sup>2-</sup> , 1 x SGM <sup>2+</sup> , 1 x MG <sup>2+</sup>	

**Build Models**

[Clear Selection](#)



A0A2Y9PY83\_1.A

Cartoon ▲ □ △ ▶ ▲ □ △ ▶ ▲ □ △ ▶ ▲ □ △ ▶

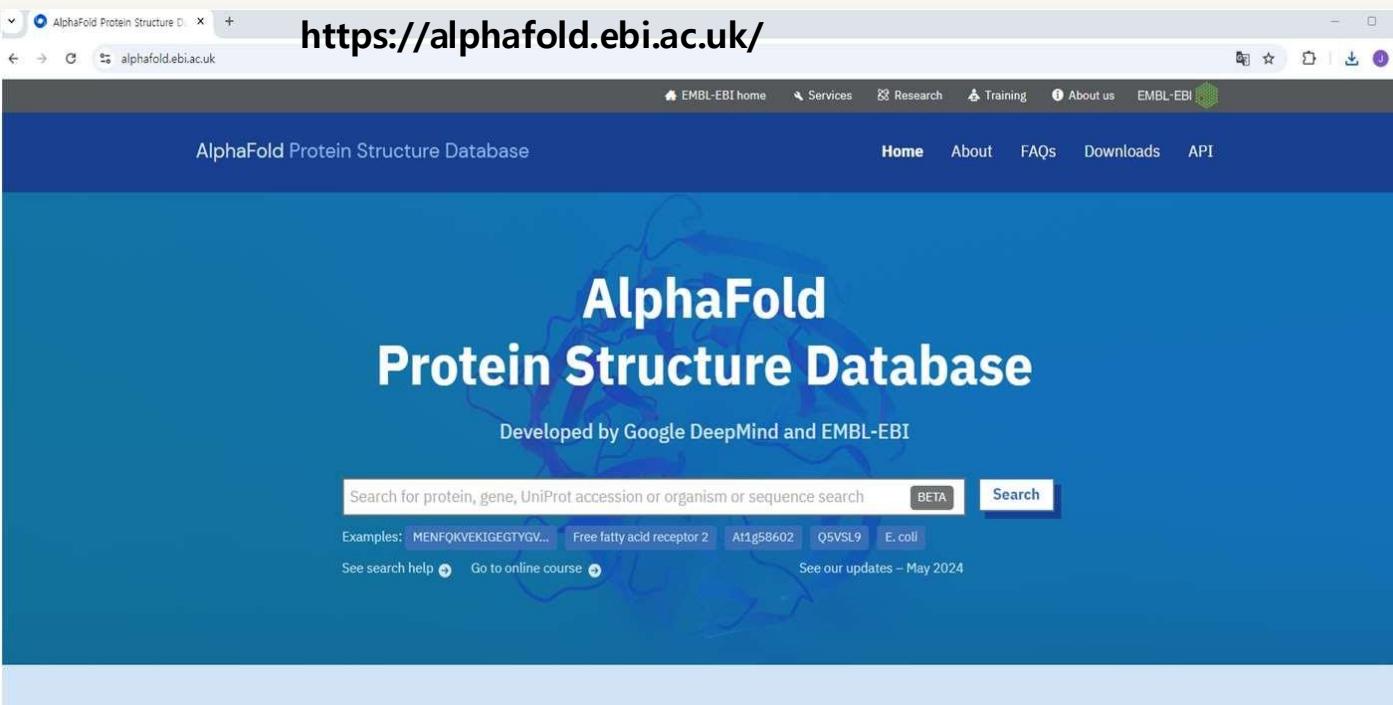
# Deep learning

## < Summary of these four protein structure prediction models >

	Model	Speed	Accuracy [44]	Use of MSA	Strength
<i>Google DeepMind</i>	AlphaFold2	Requires high-powered and high-capacity computing resources	AlphaFold2 attains a mean GDT-TS score of 73.06.	Yes, leverages MSA for rich evolutionary context	High accuracy
	AlphaFold3				
<i>Meta</i>	ESMFold	6× faster than a single AlphaFold2 model.	ESMFold attains a mean GDT-TS score of 61.62.	No, predicts structures from a single sequence	Fast prediction speed
	RoseTTAFold	Vary depending on the specific protein and computational resources, compared to AlphaFold2.	In over 80% of cases, RoseTTAFold's performance was lower than ESMFold, with the latter achieving a higher mean GDT-TS score.	Yes, uses MSAs	predicting protein complexes with RNA or DNA
<i>non-profit AI research consortium</i>	RoseTTAFold All-Atom (RFAA)				
	OpenFold	Slightly faster than AlphaFold2 [45].		Yes, uses MSAs	Allows for application-specific training

Biomolecules 2024, 14, 339

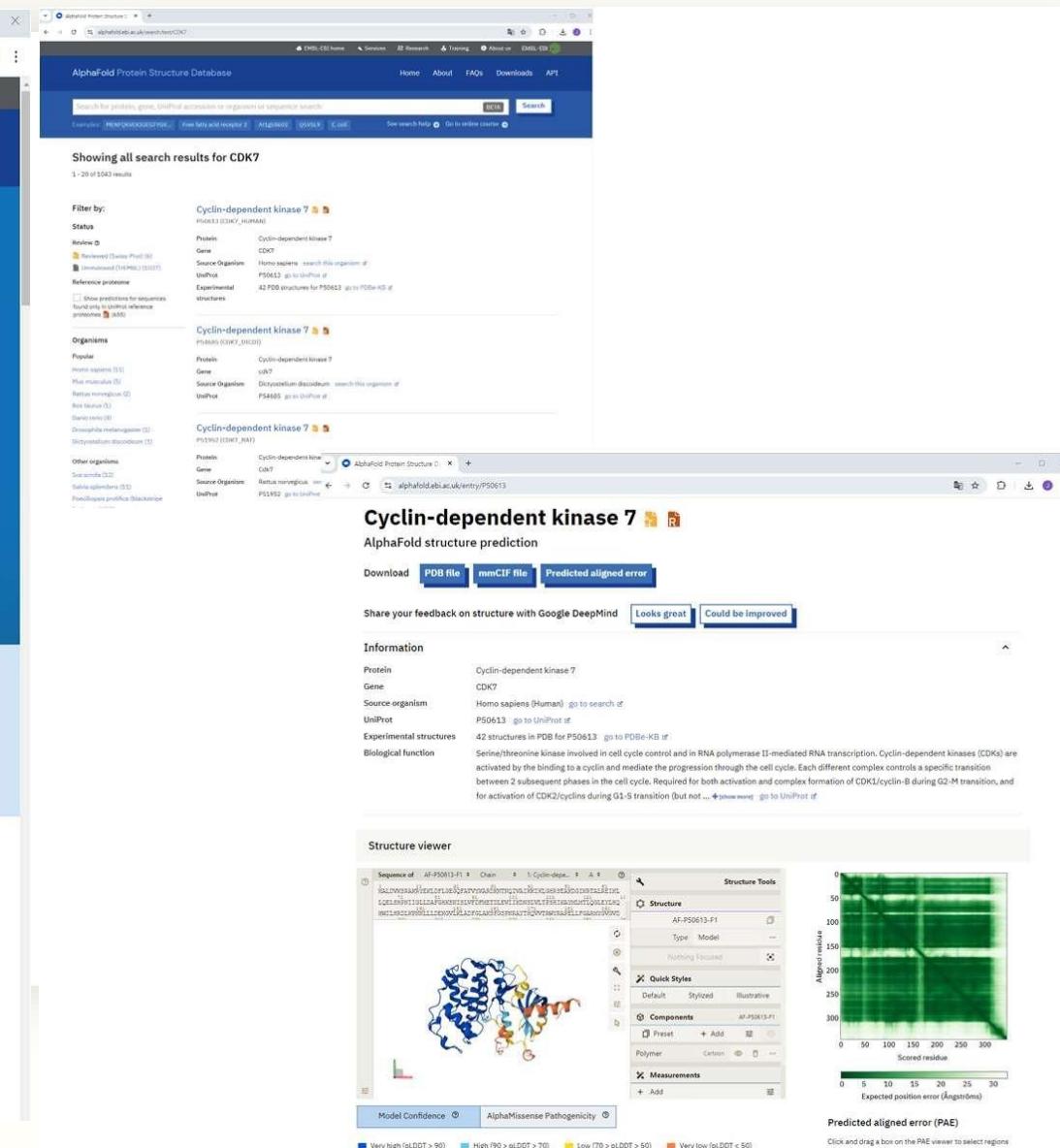
<https://alphafold.ebi.ac.uk/>



The homepage of the AlphaFold Protein Structure Database. It features a large banner with the text "AlphaFold Protein Structure Database" and "Developed by Google DeepMind and EMBL-EBI". Below the banner is a search bar with placeholder text "Search for protein, gene, UniProt accession or organism or sequence search" and a "BETA" button. Below the search bar are examples of search terms: "MENFQKVEKIGEGTYGV...", "Free fatty acid receptor 2", "At1g58602", "Q5VSL9", and "E.coli". There is also a link to "See our updates – May 2024". At the bottom of the page, it states: "AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research."

## Background

AlphaFold is an AI system developed by Google DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves

The screenshot shows the AlphaFold Protein Structure Database interface. On the left, a sidebar lists search results for "CDK7". The main area displays a detailed view of the "Cyclin-dependent kinase 7" entry. The top navigation bar includes links for Home, About, FAQs, Downloads, and API.

**Showing all search results for CDK7**

**Filter by:**

- Status: Review (Reviewed (2467) | Unreviewed (145862))
- Reference protein: P50613 (CDK7\_HUMAN)
- Proteins: Cyclin-dependent kinase 7
- Source Organism: Homo sapiens (search this organism)
- UniProt: P50613 (go to UniProt)
- Experimental structures: 42 PDB structures for P50613 (go to PDBe-KB)

**Organisms:**

- Popular: Homo sapiens (1), Mus musculus (1), Rattus norvegicus (1), Bos taurus (1), Danio rerio (1), Drosophila melanogaster (1), Dicotyledonaceae (1)
- Other organisms: Saccharomyces cerevisiae (1), Salvia splendens (1), Pseudoplatystoma brachyops (1)

**Cyclin-dependent kinase 7**

**AlphaFold structure prediction**

**Download:** PDB file, mmCIF file, Predicted aligned error

**Share your feedback on structure with Google DeepMind:** Looks great, Could be improved

**Information:**

- Protein: Cyclin-dependent kinase 7
- Gene: CDK7
- Source organism: Homo sapiens (Human) (go to search)
- UniProt: P50613 (go to UniProt)
- Experimental structures: 42 structures in PDB for P50613 (go to PDBe-KB)
- Biological function: Serine/threonine kinase involved in cell cycle control and in RNA polymerase II-mediated RNA transcription. Cyclin-dependent kinases (CDKs) are activated by the binding to a cyclin and mediate the progression through the cell cycle. Each different complex controls a specific transition between 2 subsequent phases in the cell cycle. Required for both activation and complex formation of CDK1/cyclin-B during G2-M transition, and for activation of CDK2/cycline during G1-S transition (but not ...). (more)

**Structure viewer:**

The structure viewer displays the predicted 3D structure of the protein. It includes a ribbon model, a sequence viewer, and various analysis tools. A color scale at the bottom indicates the "Predicted aligned error (PAE)" in Ångströms, ranging from 0 to 30.

## • UniProtKB(UniProt Knowledgebase), <https://www.uniprot.org/>

UniProtKB는 단백질에 관한 기능 정보를 수집하기 위한 central hub이며, 정확하고 일관성있는 풍부한 annotation이 붙어 있다. 각 UniProtKB 항목에 필수 핵심 데이터(주로 the amino acid sequence, protein name or description, taxonomic data and citation information)를 캡처할뿐만 아니라 가능한 많은 annotation 정보가 추가된다.

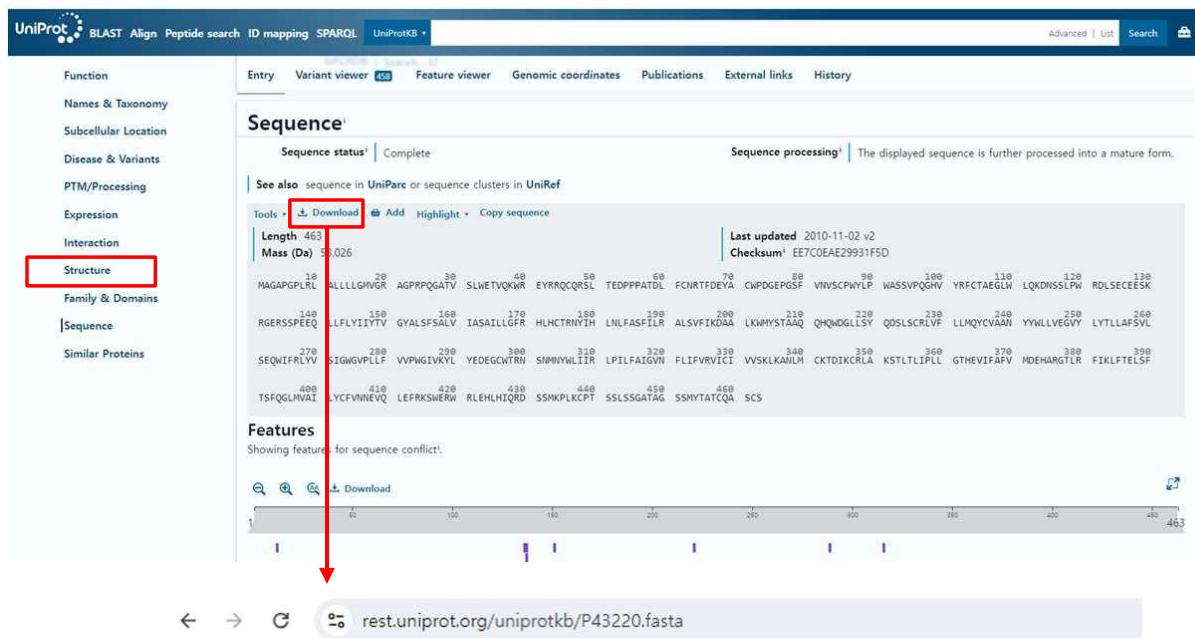


This screenshot shows the UniProtKB search results for the query "GLP1". The results page displays 1,825 entries. The left sidebar includes links for Status (Reviewed: 173, Unreviewed: 1,652), Popular organisms (Human: 53, Mouse: 42, Rat: 36, Bovine: 33, Zebrafish: 18), Taxonomy (Filter by taxonomy), Group by (Taxonomy, Keywords, Gene Ontology, Enzyme Class), and Proteins with (3D structure: 45, Active site: 11). The main content area lists protein entries with columns for Entry, Entry Name, Protein Names, Gene Names, Organism, and Length. Examples include Q1WG82 (ZGLP1\_MOUSE), P13508 (GLP1\_CAEEL), and P43220 (GLP1R\_HUMAN).

This screenshot shows the detailed view for the protein P43220 · GLP1R\_HUMAN. The top right shows basic information: Protein (Glucagon-like peptide 1 receptor), Gene (GLP1R), Status (UniProtKB reviewed (Swiss-Prot)), Organism (Homo sapiens (Human)), and Annotation score (463). The bottom right shows a sequence alignment with GLP1, showing a conservation scale from 10 to 400. The detailed view includes sections for Function, Expression, Interaction, Structure, Family & Domains, Sequence, and Similar Proteins. Key details include: Function (G-protein coupled receptor for glucagon-like peptide 1 (GLP-1)), Expression (Homo sapiens (Human)), Interaction (Ligand binding triggers activation of a signaling cascade that leads to the activation of adenylyl cyclase and increased intracellular cAMP levels), and Features (Selective recognition of glucagon-like peptide over glucagon is determined by residues located at the C-terminal end of the glucagon-like peptide).

- UniProtKB(UniProt Knowledgebase), <https://www.uniprot.org/>

UniProtKB는 단백질에 관한 기능 정보를 수집하기 위한 central hub이며, 정확하고 일관성있는 풍부한 annotation이 붙어 있다. 각 UniProtKB 항목에 필수 핵심 데이터(주로 the amino acid sequence, protein name or description, taxonomic data and citation information)를 캡처할뿐만 아니라 가능한 많은 annotation 정보가 추가된다.



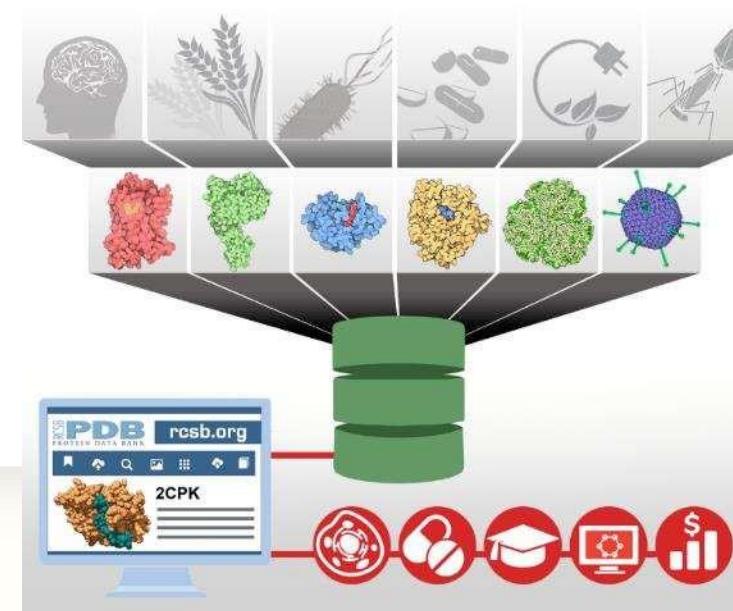
The screenshot shows the UniProtKB sequence page for P43220. The page has a left sidebar with various links like Function, Names & Taxonomy, Subcellular Location, Disease & Variants, PTM/Processing, Expression, Interaction, Structure (which is highlighted with a red box), Family & Domains, Sequence, and Similar Proteins. The main content area shows the protein sequence starting with MAGAPGPLRLR, followed by a detailed view of the sequence with numbered amino acids (18 to 463). It includes sections for Sequence status (Complete), Sequence processing (further processed into a mature form), Tools (with a Download button highlighted with a red box), and Features. Below the sequence is a timeline showing the submission history from 10 to 30 submissions. At the bottom, there's a download link and a browser address bar showing 'rest.uniprot.org/uniprotkb/P43220.fasta'.



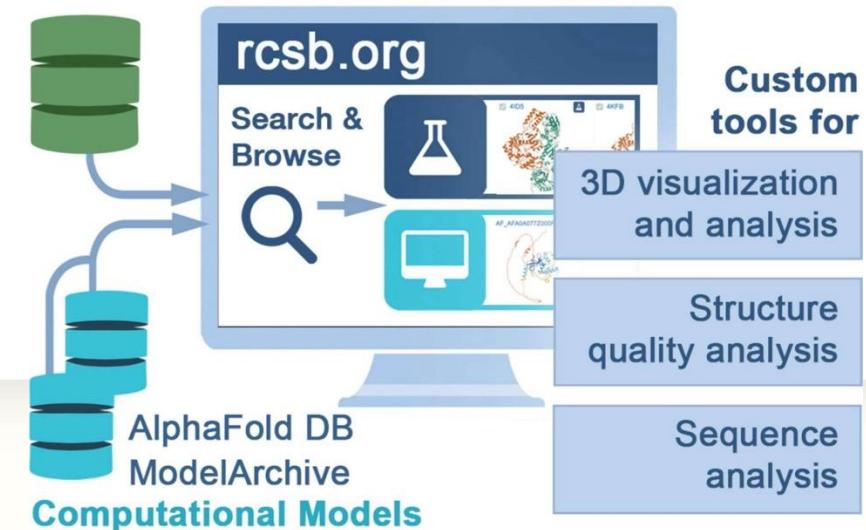
```
>sp|P43220|GLP1R_HUMAN Glucagon-like peptide 1 receptor OS=Homo sapiens OX=9606 GN=GLP1R PE=1 SV=2
MAGAPGPLRLRALLLGMVGRAGPPQGATVSLWETVQK#REYRQQCQLTEDPPPATDL
FCNRTFDEYK#PDGEPGSFYNSCP#VLP#ASSVPQGHVYRFCTAEGL#IQLKDQSSLW#
RDLSECEESKRGERSSPEEQLLFLYIYTGVGVALSFSALVIAASATLLGFRHLHCTRNYIH
LNLFASFILRLASVFICDAALK#WSTAQQHQWDGLLSVQDSLSCRLVFLLMQCVAAAN
YYVLLVEGVLYTLLAFSVLSEQ#IFRLVVS1G#GVPLLFWVP#GIVKVLVEDEGCWTRN
SNMNYWLIIIRLP1LFA1GNFLTF1FVRV1C1VWSKLKANLMCKTDIKCRLAKSTLTLPLL
GTHEVIAFVMDHARGTLRIKLFTELSFTSFQQLMVIALYCFVNNEYQLEFRKSWEW#I
RLEHLHICDSSMKPLKCPTSSLSGATAGSSMTATCQASC8
```

- RCSB PDB, <https://www.rcsb.org>

the US data center for the global Protein Data Bank (PDB) archive of 3D structure data for large biological molecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology.



### Experimental Structures Protein Data Bank



## • RCSB PDB, <https://www.rcsb.org>

the US data center for the global Protein Data Bank (PDB) archive of 3D structure data for large biological molecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology.



RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

Search Summary This query matches 51 Structures

Refinements 0

Structure Determination Methodology

- experimental (51)

Scientific Name of Source Organism

- Homo sapiens (51)
- other synthetic construct (18)
- Bos taurus (14)
- Lami glama (14)
- Rattus norvegicus (14)
- Escherichia coli (5)
- Mus musculus (5)
- Tetratrichuris trichuris (4)
- Heloderma suspectum (4)
- Clostridium perfringens (1)
- More...

Taxonomy

- Eukaryota (51)
- other sequences (18)
- Bacteria (6)
- Duplodnaviria (5)
- Eukaryote (eukaryotes) (2)

Experimental Method

- ELECTRON MICROSCOPY (34)
- X-RAY DIFFRACTION (17)

Polymer Entity Type

- Protein (51)

Refinement Resolution (Å)

- 1.5 - 2.0 (3)
- 2.0 - 2.5 (11)
- 2.5 - 3.0 (18)
- 3.0 - 3.5 (14)
- 3.5 - 4.0 (3)
- 4.0 - 4.5 (2)

1 to 25 of 51 Structures

Page 1 of 3 ► 25 ▾ Sort by Score

**6GB1** Crystal structure of the GLP1 receptor ECD with Peptide 11 Schreuder, H.A., Liesum, A. (2018) J Med Chem 61: 5580-5593. Released 2018-06-20 Method X-RAY DIFFRACTION 2.73 Å Organisms Homo sapiens synthetic construct Macromolecule Glucagon-like peptide 1 receptor (protein) Peptide 11 (protein) Unique Ligands HEZ, SO4

**5OTT** Extracellular domain of GLP-1 receptor in complex with exendin-4 variant Gly2Hcs/Thr5Hcs Mortensen, S. (2018) Biochemistry 57: 4148-4154. Released 2018-07-04 Method X-RAY DIFFRACTION 1.92 Å Organisms Heloderma suspectum Homo sapiens Macromolecule Exendin-4 (protein) Glucagon-like peptide 1 receptor (protein)

**7LCJ** PF 06882961 bound to the glucagon-like peptide-1 receptor (GLP-1R);Gs complex Belousoff, M.J., Johnson, R.M., Drulyte, I., Yu, L., Kotecha, A., Danev, R., Wootten, D., Zhang, X., Sexton, P.M. (2021) Structure 29: 963. Released 2021-01-20 Method ELECTRON MICROSCOPY 2.82 Å Organisms Homo sapiens Macromolecule Glucagon-like peptide 1 receptor (protein)

Download File View File

Download File View File

Download File View File

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

224,004 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

Advanced Search | Browse Annotations Help

PDB-101 CPDB EMDataResource NAKB wwPDB Foundation PDB-Dev

Facebook YouTube LinkedIn

Structure Summary Structure Annotations Experiment Sequence Genome Versions

**7LCK** PF 06882961 bound to the glucagon-like peptide-1 receptor (GLP-1R) PDB DOI: <https://doi.org/10.2210/pdb7LCK/pdb> EM Map EMD-23276: EMDB EMDataResource Classification: MEMBRANE PROTEIN Organism(s): Homo sapiens Expression System: Trichoplusia ni Mutation(s): No Membrane Protein: Yes PDB | PDBTM | MemProTM | msmsmc

Deposited: 2021-01-11 Released: 2021-01-20 Deposited Author(s): Belousoff, M.J., Johnson, R.M., Drulyte, I., Yu, L., Kotecha, A., Danev, R., Wootten, D., Zhang, X., Sexton, P.M. Funding Organization(s): Australian Research Council (ARC), National Health and Medical Research Council (NHMRC), Australia, Japan Science and Technology

Experimental Data Snapshot

wwPDB Validation 3D Report Full Report

Method: ELECTRON MICROSCOPY Resolution: 3.24 Å Aggregation State: PARTICLE Reconstruction Method: SINGLE PARTICLE

Global Symmetry: Asymmetric - C1 Global Stoichiometry: Monomer - A1

Find Similar Assemblies Biological assembly 1 assigned by authors

Biological Assembly Evidence: gel filtration

Metric ClassScore Percentile Ranks Value Ramachandran outliers 0 0 0 Sidechain outliers 0 0 0

This is version 1.2 of the entry. See complete history.

Literature Download Primary Citation

Evolving cryo-EM structural approaches for GPCR drug discovery.

- RCSB PDB, <https://www.rcsb.org>

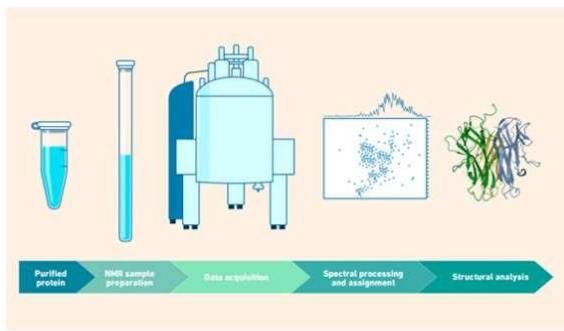
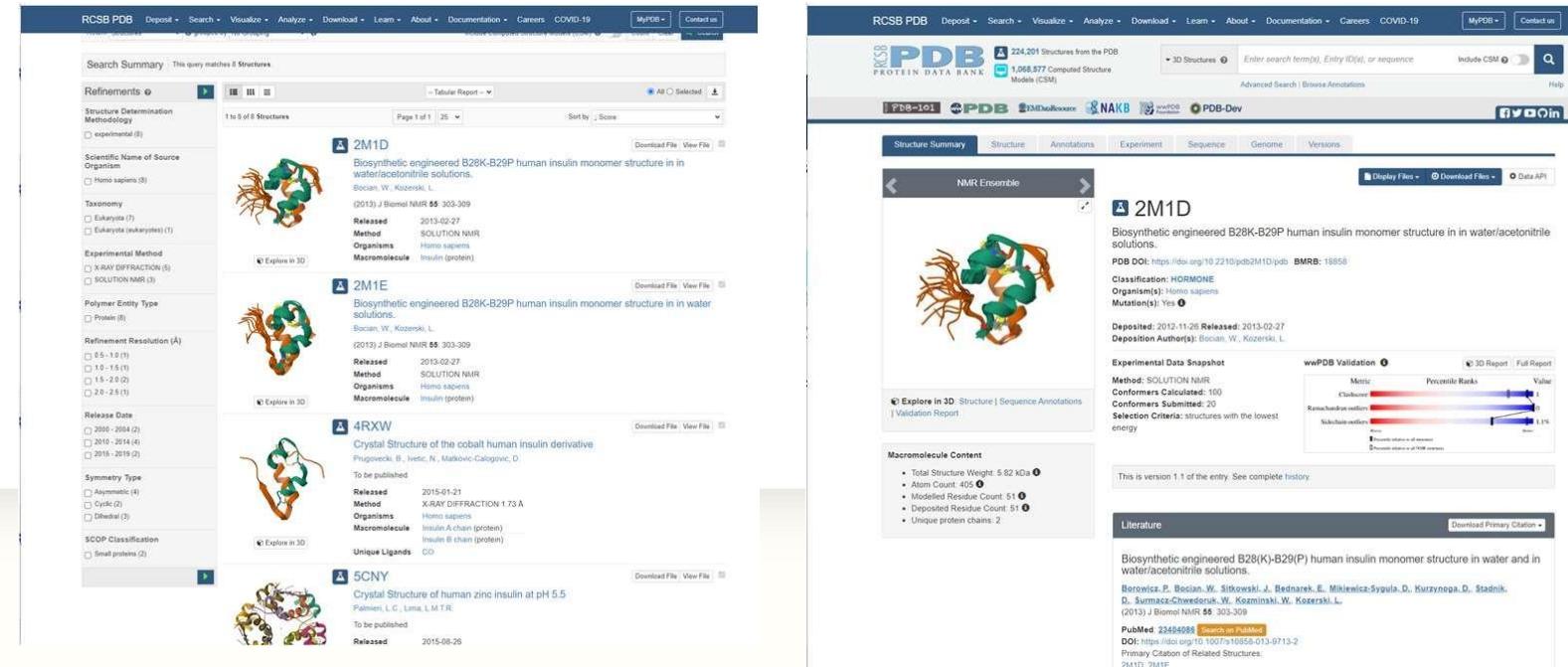


Figure 4: NMR spectroscopy workflow for protein structural determination. Credit: Technologe Networks.

Method	Size (sample state)	Resolution Limits	Amounts	Advantages	Disadvantages
NMR	< 100 kDa (solution)	~ 3-4 Å	µmoles/milligrams	<ul style="list-style-type: none"> <li>High resolution</li> <li>3D structure in solution good <b>for dynamic study</b></li> </ul>	<ul style="list-style-type: none"> <li>Difficult sample prep</li> <li>High sample purity need</li> <li>Static crystalline state capture</li> <li>Only small proteins</li> </ul>

## PDB >> human insulin



Search Summary: This query matches 8 Structures

Refinements: Structure Determination Methodology (Experimental), Scientific Name of Source Organism (Homo sapiens), Taxonomy (Eukaryota, Eukaryota (eukaryotes)), Experimental Method (X-RAY DIFFRACTION, SOLUTION NMR), Polymer Entity Type (Protein), Refinement Resolution (Å) (0.5 - 1.0), Release Date (2000 - 2004, 2010 - 2014, 2015 - 2019), Symmetry Type (Asymmetric, Cyclic, Dihedral), SCOP Classification (Small proteins).

Structure Summary: 2M1D, 2M1E, 4RXW, 5CNY

2M1D: Biosynthetic engineered B28K-B29P human insulin monomer structure in water/acetonitrile solutions. Bocian, W., Kożerski, L. (2013) J Biomol NMR 55: 303-309. Released: 2013-02-27. Method: SOLUTION NMR. Organisms: Homo sapiens. Macromolecule: Insulin (protein). Explore in 3D.

2M1E: Biosynthetic engineered B28K-B29P human insulin monomer structure in water. Bocian, W., Kożerski, L. (2013) J Biomol NMR 55: 303-309. Released: 2013-02-27. Method: SOLUTION NMR. Organisms: Homo sapiens. Macromolecule: Insulin (protein). Explore in 3D.

4RXW: Crystal Structure of the cobalt human insulin derivative. Prugocki, B., Nedic, N., Maljkovic-Calegovic, D. To be published. Released: 2015-01-21. Method: X-RAY DIFFRACTION 1.73 Å. Organisms: Homo sapiens. Macromolecule: Insulin A chain (protein), Insulin B chain (protein). Unique Ligands: CO. Explore in 3D.

5CNY: Crystal Structure of human zinc insulin at pH 5.5. Palme, L.C., Lima, L.M.T.R. To be published. Released: 2015-08-26.

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

224,201 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

Advanced Search | Browse Annotations Help

PDB-101 PDB-Dev PDB-NAKB UniPDB PDB-DB

Display Files Download Files Data API

Structure Summary Structure Annotations Experiment Sequence Genome Versions

NMR Ensemble

2M1D Biosynthetic engineered B28K-B29P human insulin monomer structure in water/acetonitrile solutions. PDB DOI: https://doi.org/10.2210/pdb2M1D/pdb. BMRB: 1895. Classification: HORMONE. Organism(s): Homo sapiens. Mutation(s): Yes. Deposited: 2012-11-26. Released: 2013-02-27. Deposition Authority: Bocian, W., Kożerski, L.

wwPDB Validation 3D Report Full Report

Method: SOLUTION NMR. Conformers Calculated: 100. Conformers Submitted: 20. Selection Criteria: structures with the lowest energy.

Macromolecule Content: Total Structure Weight: 5.62 kDa. Atom Count: 403. Modelled Residue Count: 51. Deposited Residue Count: 51. Unique Protein Chains: 2.

Literature Download Primary Citation

This is version 1.1 of the entry. See complete history.

Berowicz, P., Bocian, W., Słotkowski, J., Bednarek, E., Mikiewicz-Sygula, D., Kurzynska, D., Stadnik, D., Sumarcz-Chwedoruk, W., Kozimski, W., Kożerski, L. (2013) J Biomol NMR 55: 303-309. PubMed: 23464894. PMID: 23464894. DOI: https://doi.org/10.1007/s10658-013-9713-2. Primary Citation of Related Structures: 2M1D, 2M1E.

- RCSB PDB, <https://www.rcsb.org>

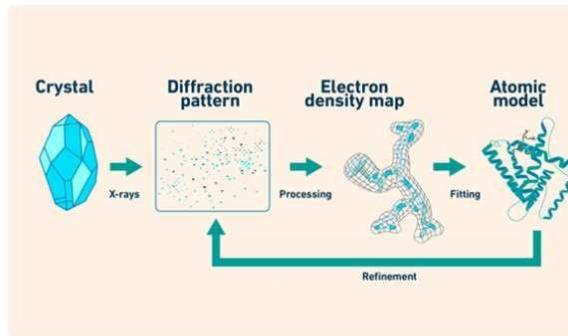
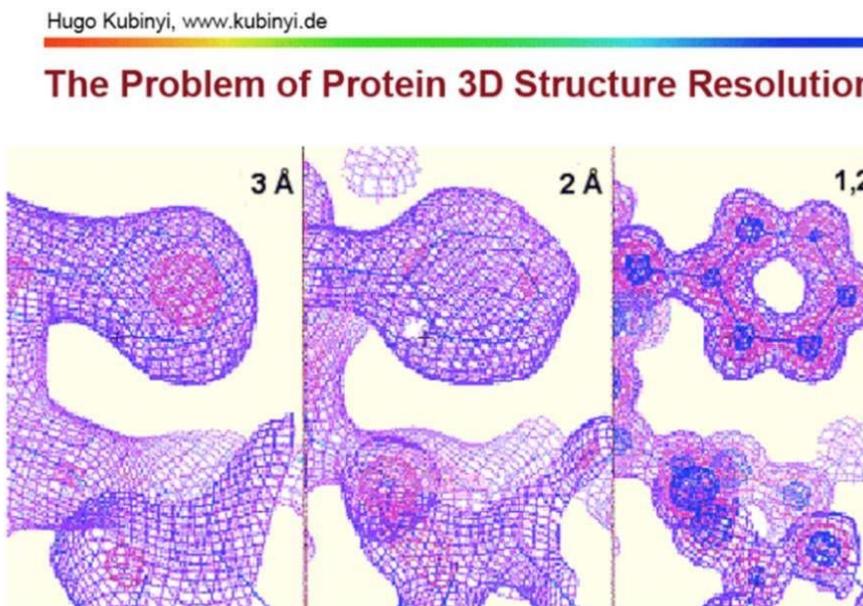


Figure 2: Steps for protein structure determination by X-ray crystallography. Credit: Technology Networks.

Method	Size (sample state)	Resolution Limits	Amounts	Advantages	Disadvantages
X-ray	Limited by crystal quality	< 1-3 Å	µmoles/milligrams	<ul style="list-style-type: none"> <li>Well developed</li> <li>High resolution</li> <li><b>Broad molecular weight range</b></li> <li><b>Atomic resolution</b></li> </ul>	<ul style="list-style-type: none"> <li>Difficult sample prep</li> <li>Static crystalline state</li> </ul>

- ✓ Species: human, mouse etc.
- ✓ Method: X-ray, Cryo-EM etc.
- ✓ Resolution: < 3 Å
- ✓ Mutation: check
- ✓ Chain: A, B, E, F etc.
- ✓ Bound ligand



- RCSB PDB, <https://www.rcsb.org>

## PDB >> CDK7

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

Refinements 

Structure Determination Methodology  experimental (67)

Scientific Name of Source Organism  Homo sapiens (65)  Sus scrofa (20)  synthetic construct (11)  Amanita phalloides (8)  Human mastadenovirus C (6)  Xenopus laevis (4)  unidentified adenovirus (2)  Lama glama (1)  Thermochaetoides thermophila (1)  Thermochaetoides thermophila DSM 1495 (1) More...

Taxonomy  Eukaryota (67)  other sequences (11)  Varidnaviria (8)

Experimental Method  ELECTRON MICROSCOPY (48)  X-RAY DIFFRACTION (19)

Polymer Entity Type  Protein (67)  DNA (24)  RNA (8)

Refinement Resolution (Å)  1.5 - 2.0 (16)  2.0 - 2.5 (13)  2.5 - 3.0 (9)  3.0 - 3.5 (4)  3.5 - 4.0 (3)  4.0 - 4.5 (7)  > 4.5 (15)

Release Date  1995 - 1999 (1)  2000 - 2004 (1)  2015 - 2019 (6)  2020 - 2024 (59)

1 to 25 of 67 Structures Page 1 of 3 Sort by Score All Selected  

**1UA2** Crystal Structure of Human CDK7  
Lolli, G., Lowe, E.D., Brown, N.R., Johnson, L.N.  
(2004) Structure 12: 2067-2079  
Released 2004-12-07  
Method X-RAY DIFFRACTION 3.02 Å  
Organisms Homo sapiens  
Macromolecule Cell division protein kinase 7 (protein)  
Unique Ligands ATP

**8R99** A soakable crystal form of human CDK7 in complex with AMP-PNP  
Mukherjee, M., Cleasby, A.  
(2024) Structure 32: 1040-1048.e3  
Released 2024-05-29  
Method X-RAY DIFFRACTION 1.81 Å  
Organisms Homo sapiens  
Macromolecule Cyclin-dependent kinase 7 (protein)  
Unique Ligands ANP, MG

**8R9A** A soakable crystal form of human CDK7 in complex with AMP-PNP  
Mukherjee, M., Cleasby, A.  
(2024) Structure 32: 1040-1048.e3  
Released 2024-05-29  
Method X-RAY DIFFRACTION 1.71 Å  
Organisms Homo sapiens  
Macromolecule Cyclin-dependent kinase 7 (protein)  
Unique Ligands I74, PEG

**8R9O** A soakable crystal form of human CDK7 in complex with AMP-PNP  
Mukherjee, M., Cleasby, A.  
(2024) Structure 32: 1040-1048.e3  
Released 2024-05-29  
Method X-RAY DIFFRACTION 2.22 Å  
Organisms Homo sapiens  
Macromolecule Cyclin-dependent kinase 7 (protein)  
Unique Ligands YN3

## • 구조기반 가상 스크리닝

: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것

- Target-based vs. Phenotype-based
- HTS vs. HTVS
- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket
- Protein-ligand Interaction/Binding affinity
- Structure-based Virtual Screening: Case Study

# AI Tools for Scientific Research

Category	Tools	Web
번역	DeepL  - Copy & past - Pdf upload	<a href="https://www.deepl.com/ko/translator">https://www.deepl.com/ko/translator</a>
논문 이해	SciSpace 	<a href="https://scispace.com/">https://scispace.com/</a>
Find the best science, faster.	Consensus  GoatStack.AI  Papeítalk.io 	<a href="https://consensus.app/search/">https://consensus.app/search/</a> <a href="https://goatstack.ai/">https://goatstack.ai/</a> <a href="https://papertalk.io/">https://papertalk.io/</a>
Science Figures	Bio Render 	<a href="https://www.biorender.com/">https://www.biorender.com/</a>
	Research Rabbit 	<a href="https://www.researchrabbit.ai/">https://www.researchrabbit.ai/</a>

### ▪ First-in-class

: 기존 제약계에서는 마땅한 치료제가 없는 질병을 고치는 세계 최초 혁신신약

- **New scaffold:** 새로운 구조의 물질로 특허 확보가 용이함.

### ▪ Best-in-class

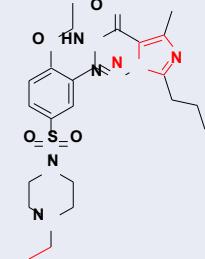
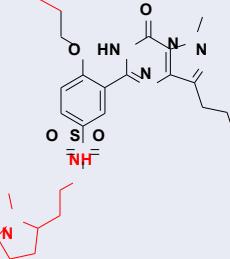
: 같은 치료기전을 가진 치료제 중에서 가장 우수한 효과를 자랑하는 경쟁 신약

- **Me-too:** 선행약물의 화학구조를 일부 바꿔 새로운 약으로 출시한 신약, 화학구조가 크게 다르지 않기에 효능도 비슷하게 나타나지만, **투자 금액과 개발 리스크가 줄어들고 개발 기간이 짧다는 장점이 있다.**

무엇보다 혁신신약 대비 효능과 안전성이 높아지거나 부작용이 감소한 Best in class가 개발될 수 있음.

- **New scaffold:** 새로운 구조의 물질로 특허 확보가 용이함.

발기부전 치료제로 사용되는 PDE5 저해제

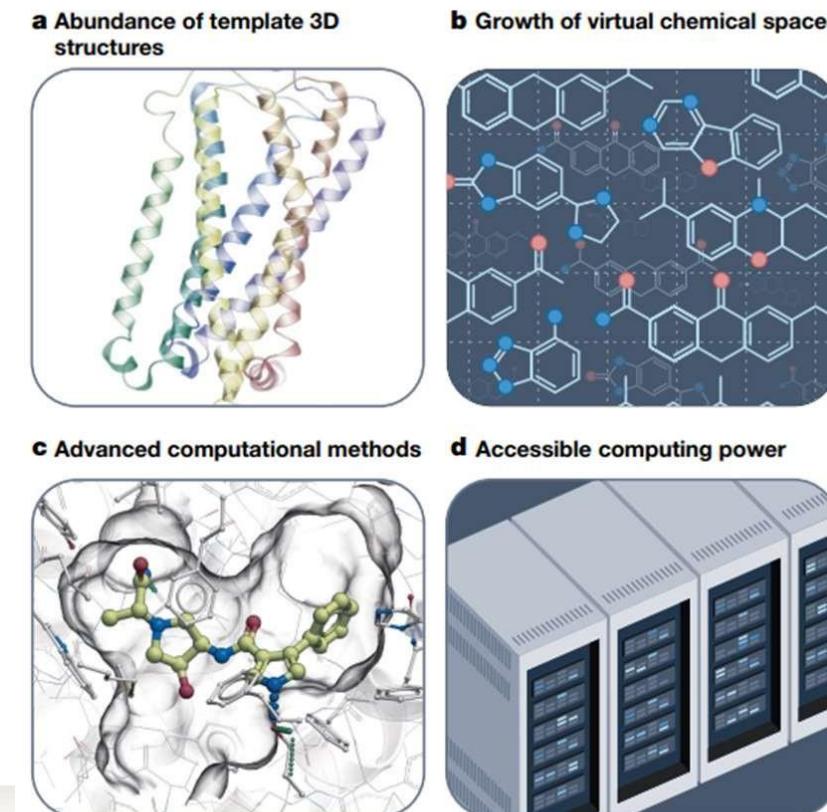
	First-in-Class	Best-in-Class	
구조		 	
성분명	Sildenafil	Vardenafil	Udenafil
상품명	Viagra	Pfizer 1998	Levitra Zydena 등
개발사		Bayer	아에스티
FDA 허가	1시간	2003	2005
발현시간	4시간	16분	16분
지속시간		4.5시간	24시간

<http://m.pharmstock.co.kr/news/articleView.html?idxno=7039>

## • 구조기반 가상 스크리닝

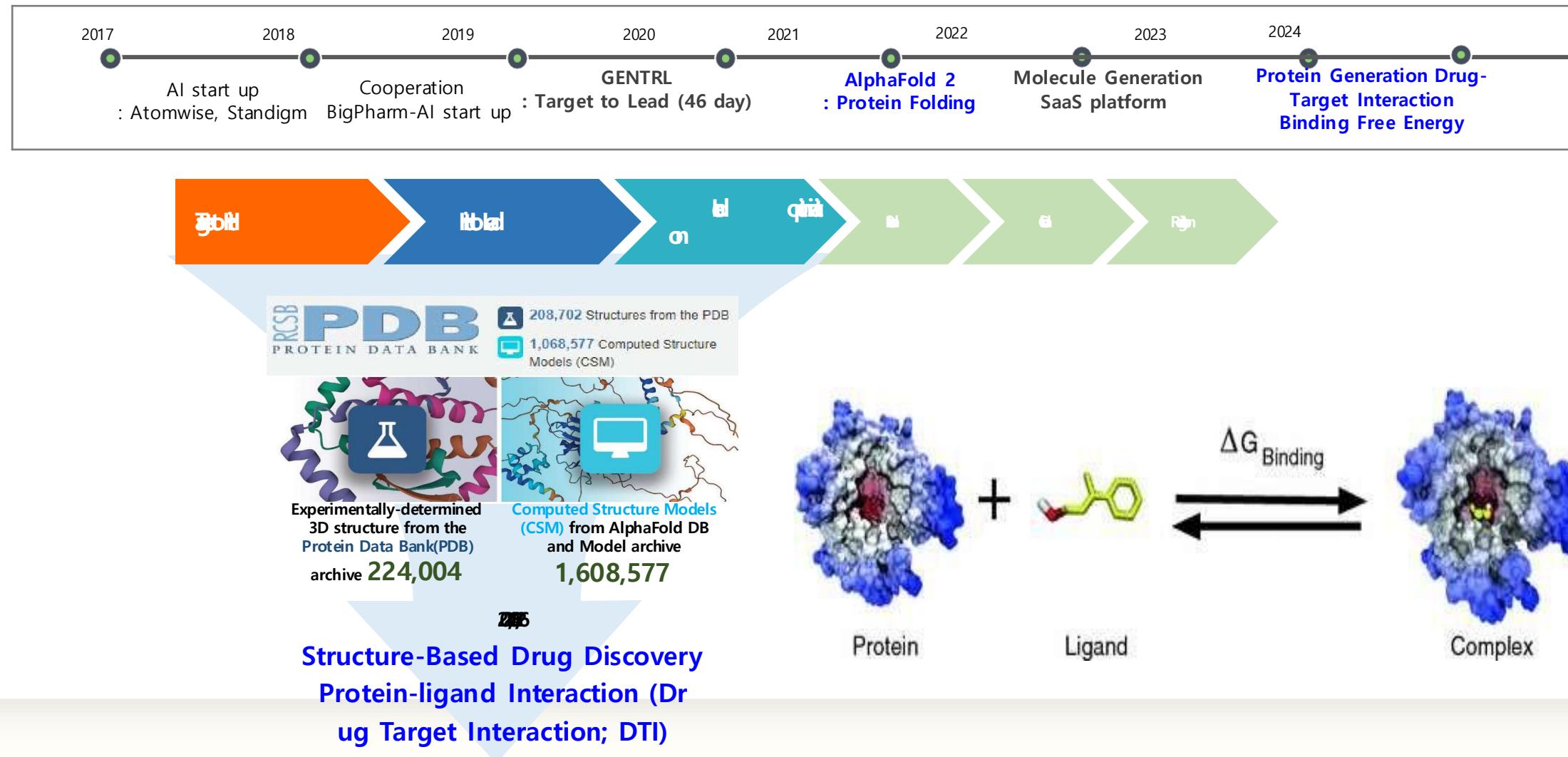
: 질환을 일으키는 특정 타겟 단백질의 **binding site**에 결합하여 활성을 보이는 화합물을 찾아 내는 것

- HTS vs. Virtual Screening
- Target-based vs. Phenotype-based
- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket: 단백질 **flexibility & structure selection**
- Protein-ligand Interaction/Binding affinity

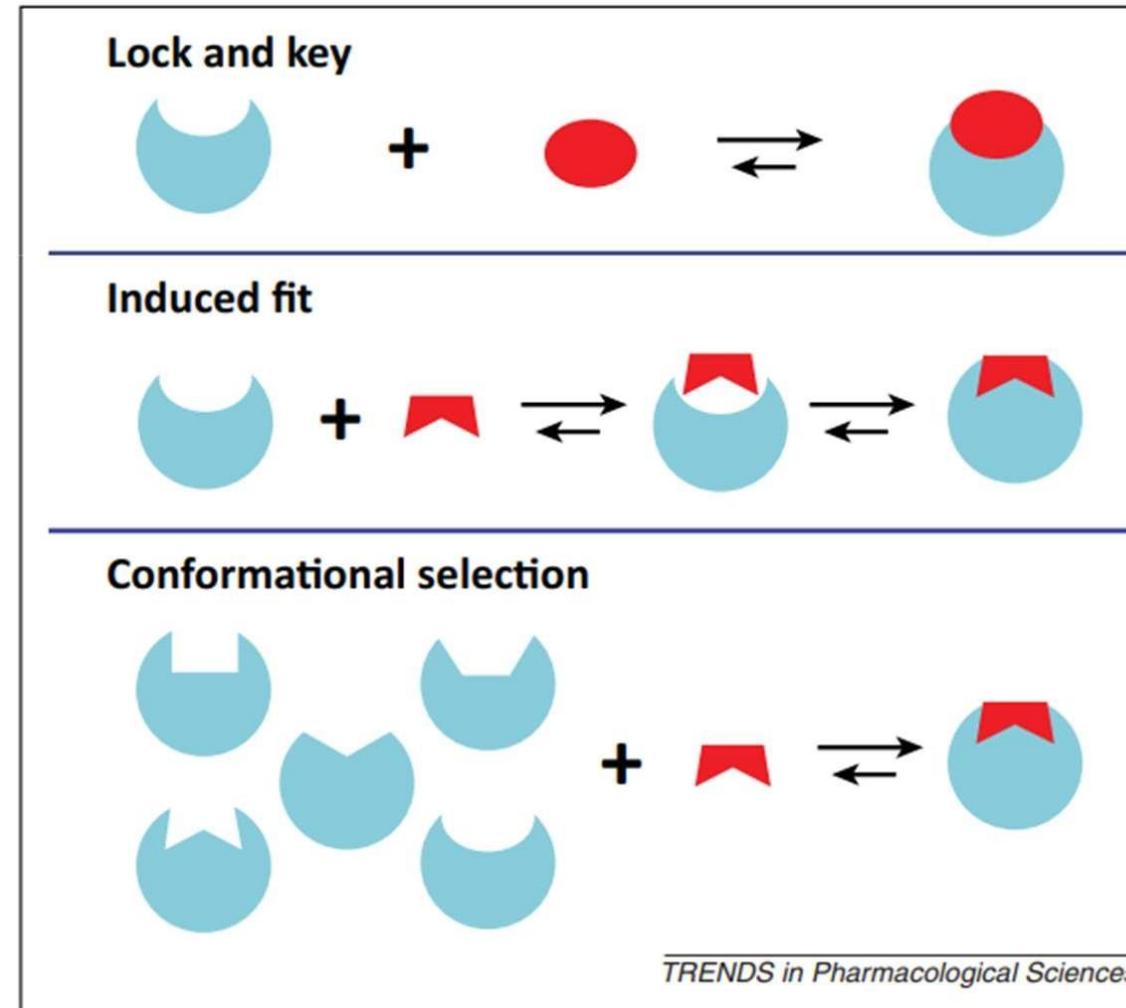


**Fig. 1 | Key factors driving VLS technology breakthroughs for generation of high-quality hits and leads.** a, More than 200,000 protein structures in the

Nature, 2023, 616, 673

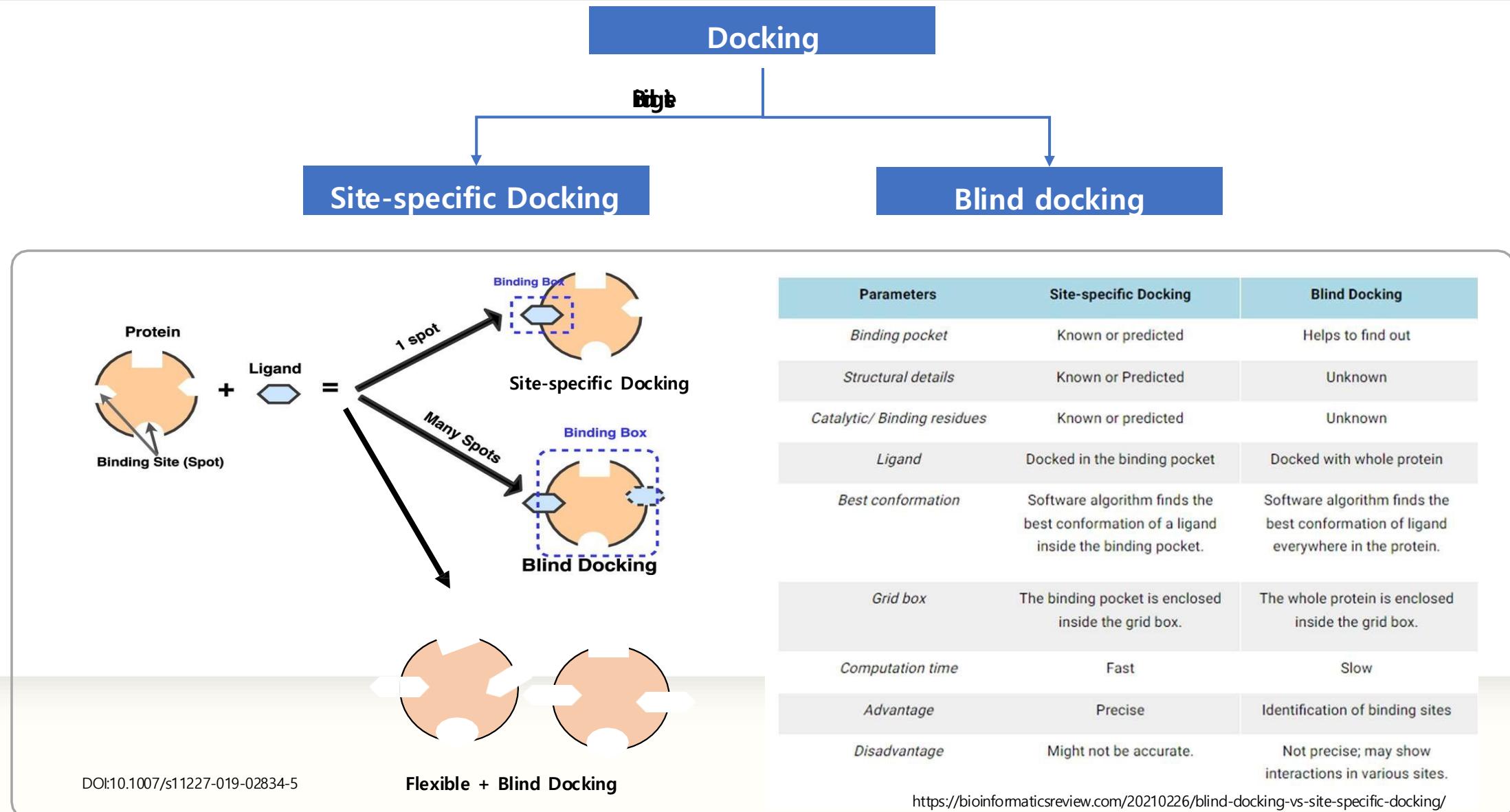


## &lt; Docking mode &gt;

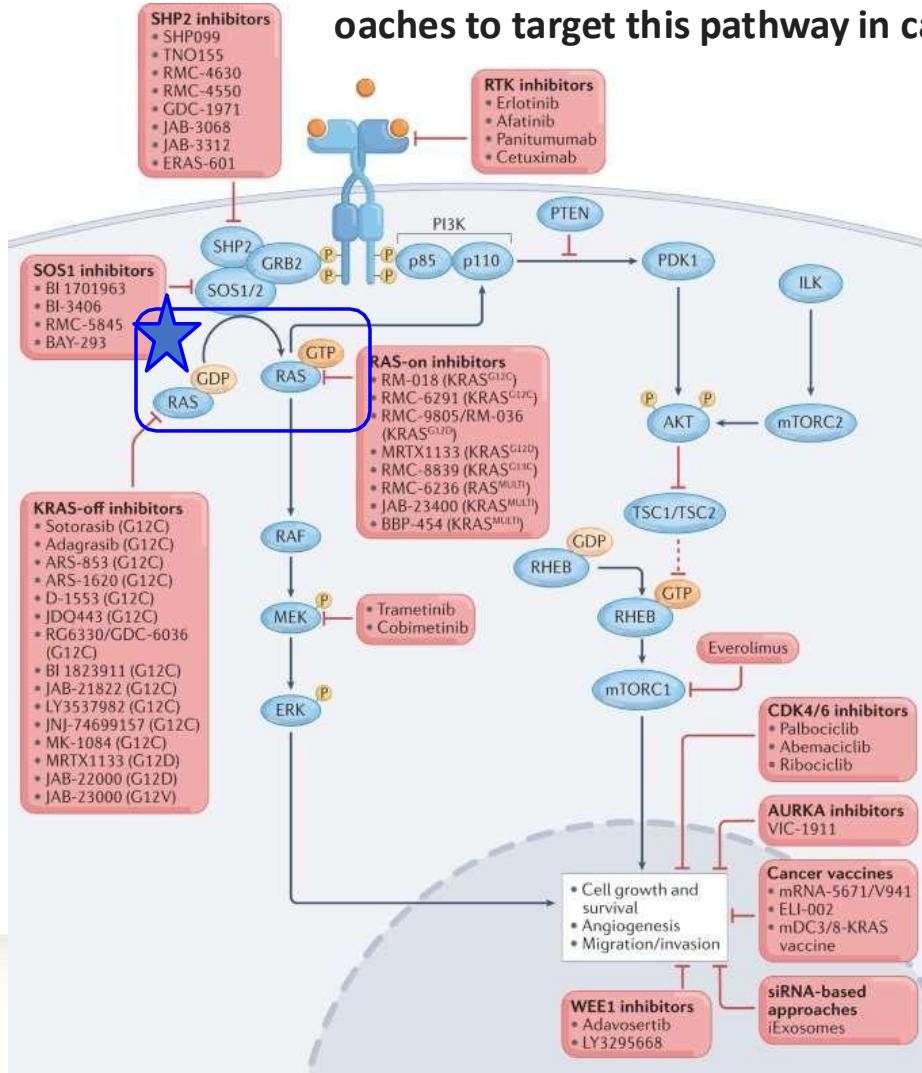


Trend in Pharm. Sci. 2015, 32, 78-95

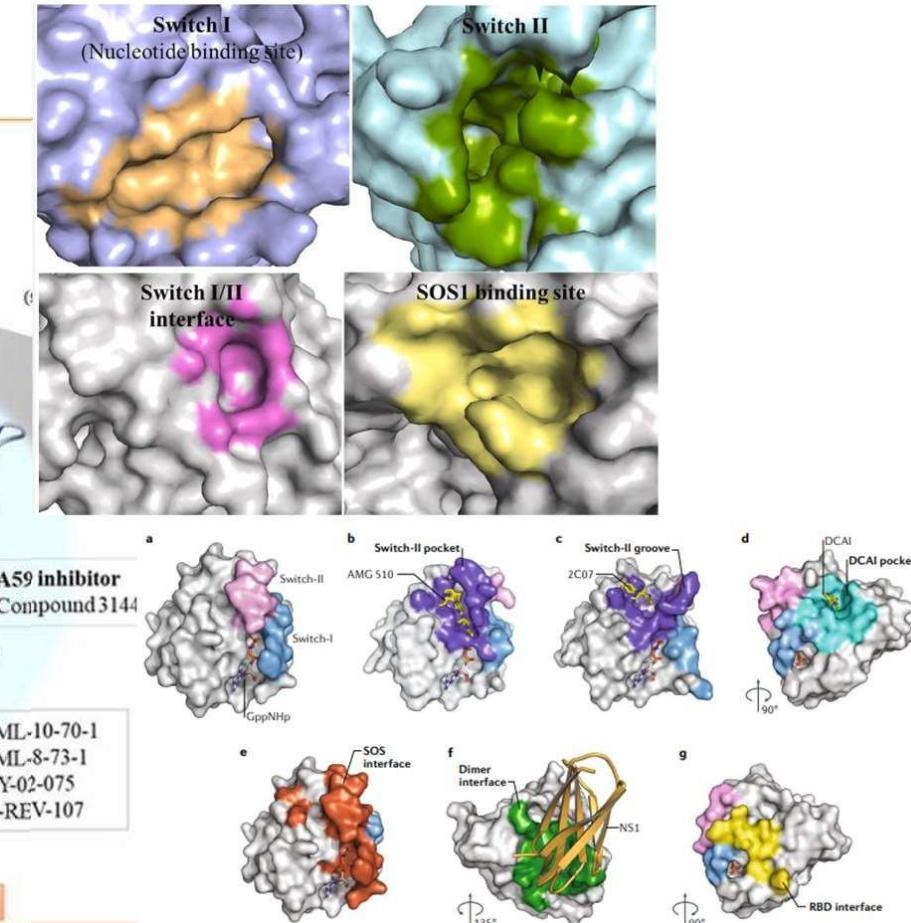
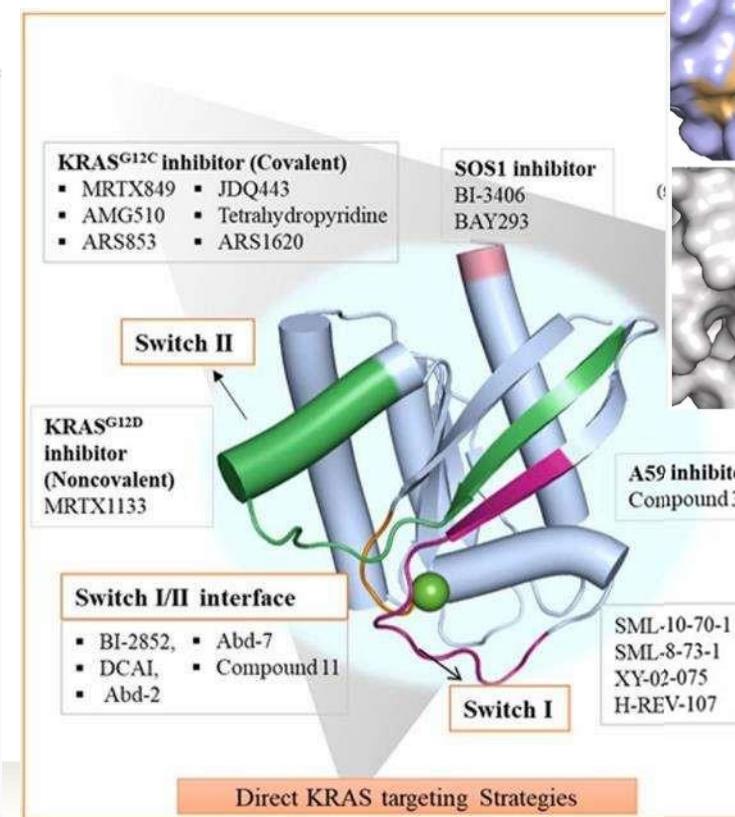
# Blind docking vs Site-specific docking



## RAS signaling pathway and therapeutic approaches to target this pathway in cancer.



## Direct KRAS targeting Strategies

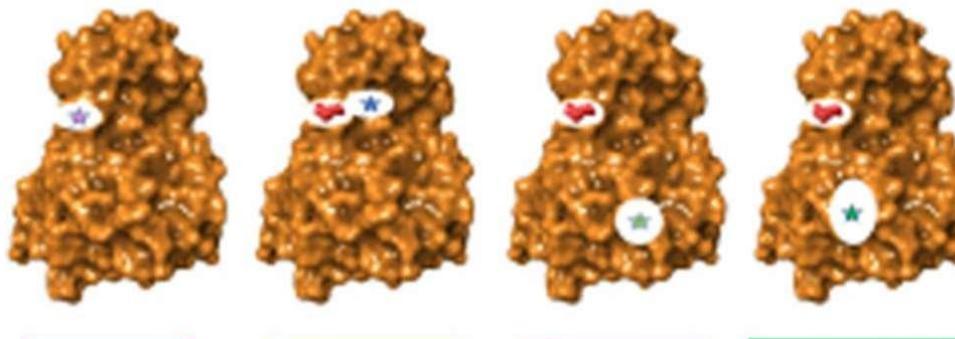
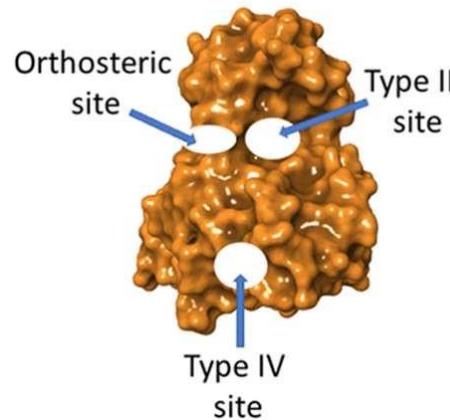


Eur. J. Med. Chem. 2024, 277, 116771  
Nat. Rev. Drug Discov. 2020, 19, 533-552

- 리간드 binding site이 알려져 있지 있을 때: Mutation study data, Reported co-crystal structure

□ Site specific Docking

< Kinase Binding Site >



Type I, II

Type III

Type IV

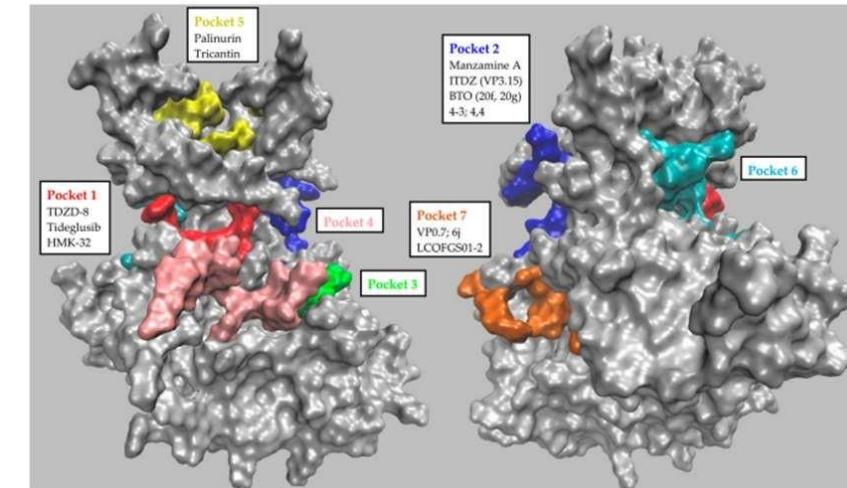
Type VI  
pseudokinase

Table 1. Key residues of the main GSK-3 $\beta$  pockets (indicated as P1–P7), as described in Palomo et al., 2011 [37]. The residue's numbering is also reported according to PDB 1JJC, chain B.

GSK-3 $\beta$ Pocket	Pocket Key Residues
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	85 97 113 134 135 138 186 188 199 200 Lys Glu Asp Tyr Val Thr Asn Leu Cys Asp 563 575 591 612 613 616 664 666 677 678
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	67 89 94 93 95 96 97 180 205 Phe Gln Lys Phe Asn Arg Glu Arg Lys 545 567 572 571 573 574 575 658 683
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	215 220 223 229 Ser Arg Arg Phe 693 698 701 707
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	140 144 148 185 219 220 221 222 249 Tyr Arg Arg Gln Ser Arg Tyr Tyr Glu 618 622 626 663 697 698 699 700 727
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	26 38 56 71 86 119 Met Thr Tyr Tyr Lys Ser 504 516 534 549 564 597
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	80 111 113 133 135 190 197 Glu Arg Arg Asp Val Asp Lys 558 589 591 611 613 668 675
Numbering of Palomo et al. Residue's identity 1JJC_B numbering	173 178 207 209 211 234 235 236 330 369 His Cys Leu Arg Glu Tyr Thr Ser Thr Ser 651 656 685 687 689 712 713 714 808 847

List of abbreviations: Lys (Lysine); Glu (Glutamate); Asp (Aspartate); Tyr (Tyrosine); Val (Valine); Thr (Threonine); Asn (Asparagine); Leu (Leucine); Cys (Cysteine); Phe (Phenylalanine); Gln (Glutamine); Ser (Serine); Met (Methionine); Arg (Arginine); His (Histidine).

Int. J. Mol. Sci. 2023, 24, 7541

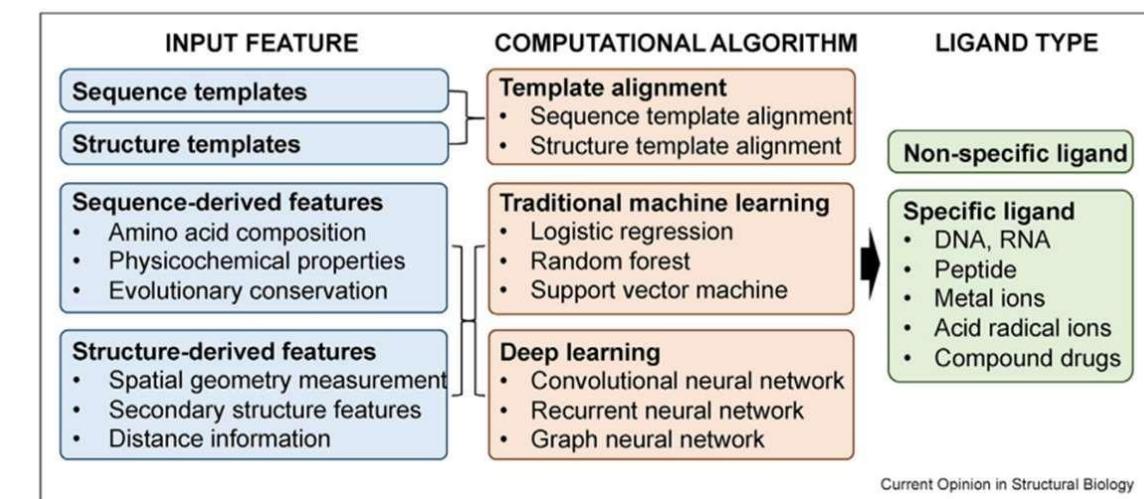
## • 리간드 binding site이 알려져 있지 않을 때: Binding site prediction

Table 1

Some popular LBS prediction methods published in recent five years.

Method	Year	Feature	Algorithm <sup>a</sup>	Ligand type
PepBind [13]	2018	sequence	Template , SVM	Peptide
P2Rank [14]	2018	structure	RF	Non-specific
COACH-D [15]	2018	sequence, structure	Template , SVM	Non-specific
DeepCSeqSite [16]	2019	sequence	CNN	Ca <sup>2+</sup> , Zn <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Fe <sup>3+</sup> , Na <sup>+</sup> , SO <sub>4</sub> <sup>2-</sup> , PO <sub>4</sub> <sup>3-</sup> , ATP, ADP, FMN, GDP, HEME, NAD
DNApred [17]	2019	sequence	SVM	DNA
NucBind [18]	2019	sequence	Template, SVM	DNA, RNA
NucleicNet [19]	2019	structure	CNN	RNA
DELIA [20]	2020	sequence, structure	CNN, biLSTM	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , ATP, HEME
DeepPocket [21]	2021	structure	CNN	Non-specific
DeepSurf [22]	2021	structure	CNN	Non-specific
GraphBind [23]	2021	structure	GNN	DNA, RNA, Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , ATP, HEME
3DLigandSite [24]	2022	structure	Template	Non-specific
ScanNet [25]	2022	structure	Attention	Protein
GraphSite [26]	2022	structure	Attention	DNA
BindWeb [27]	2022	sequence, structure	CNN, biLSTM, GNN	DNA, RNA, Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , ATP, HEME
SiteRadar [28]	2023	structure	GNN	Non-specific
LigBind [29]	2023	structure	GNN	Over 1000 specific ligands
PeSTo [30]	2023	structure	Attention	Protein
NABind [31]	2023	structure	Template, GNN	DNA, RNA

<sup>a</sup> RF: random forests [49]. SVM: support vector machines [50]. Template: template alignment. CNN: convolutional neural networks [51]. biLSTM: bidirectional long short-term memory networks [52]. GNN: graph neural networks [53]. Attention: attention mechanism-based models, such as Transformer [54].



LBS prediction methods can be generally categorized by input features, computational algorithms, and ligand types.

<https://ftmap.bu.edu/login.php>

**FTMap**  
computational solvent mapping

**Login**

Sign Up Examples Help Papers Contact

**Important Considerations**

-Please submit no more than 20 jobs at a time

**Welcome to the FTMap Server!**

FTMap is for non-commercial use only.

**Login**

Username:   
Password:

[Use FTMap without your own account](#)

Need an Account?  
[Sign Up](#)

Forgot Password?  
[Reset Password](#)

Vajda Lab | ABC Group  
Boston University | Stony Brook University

**FTMap**  
computational solvent mapping

**Important Considerations**

-Please submit no more than 20 jobs at a time

[sign out](#)

**Map**

Jobs submitted via our guest account are publicly accessible.  
Please [create an account](#) if you wish to keep your results private.

Job Name:   
Protein  
PDB ID:   
[Upload PDB](#) ?  
Chains:   
Whitespace separate desired chains. Leave chains blank to use all chains.

**Advanced Options ?**

Protein Mask:  선택된... 없음  
 PPI Mode (for binding hot spots on protein protein interfaces)  
 Has Nucleic Acid

FTMap is for non-commercial use only. By submitting this form, you agree not to use FTMap for commercial work.

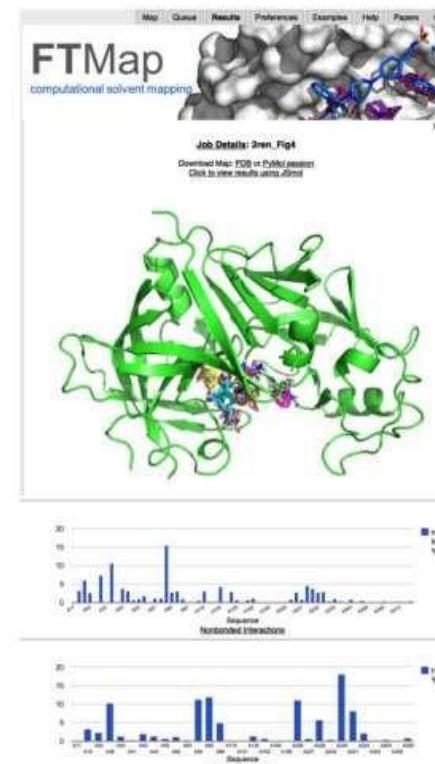
**Results**

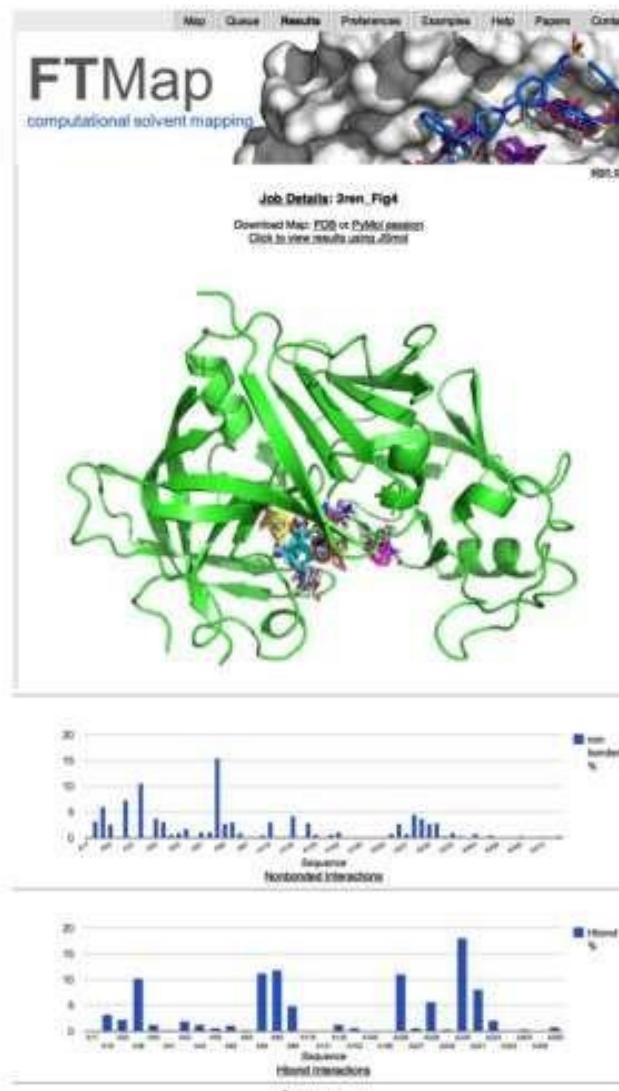
Job Details: 2ren\_Fig4  
Download Map: PDF or Pymol script  
Click to view results using Jmol

Sequence Nonbonded Interactions

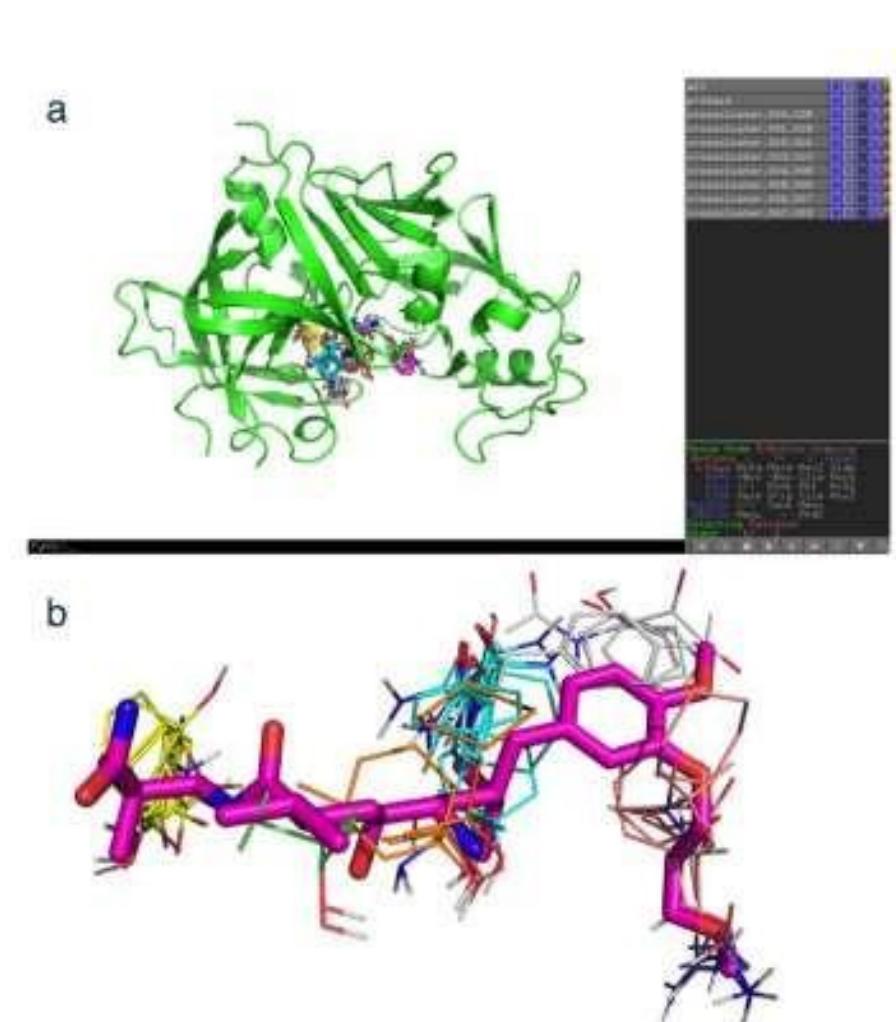
Distance Bonded Interactions

Poser summary





Nat. Protocols. 2015, 10, 733-755



<https://ftmap.bu.edu/login.php>

## Expanding FTMap for Fragment-Based Identification of Pharmacophore Regions in Ligand Binding Sites

E-FTMap (beta) SUBMIT MY JOBS

PAPERS EXAMPLES HELP CONTACT

Login  
Sign Up

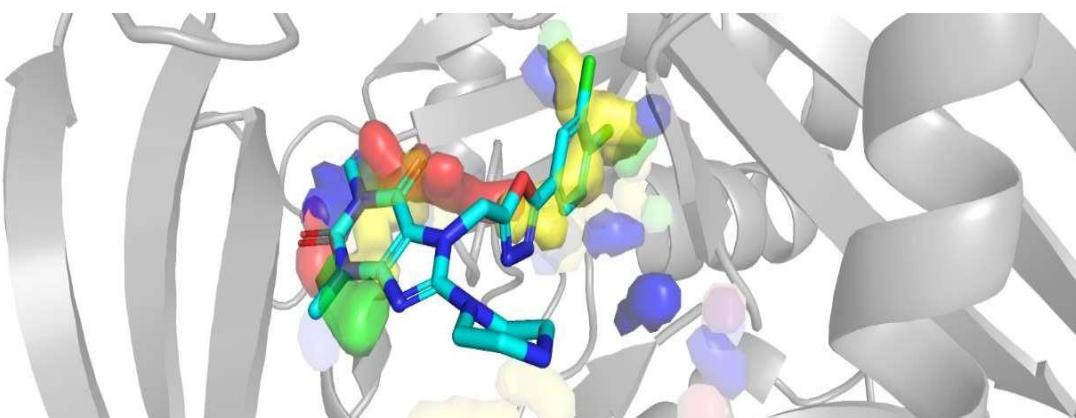
[https://eftmap.bu.edu.](https://eftmap.bu.edu)

### Welcome to E-FTMap!

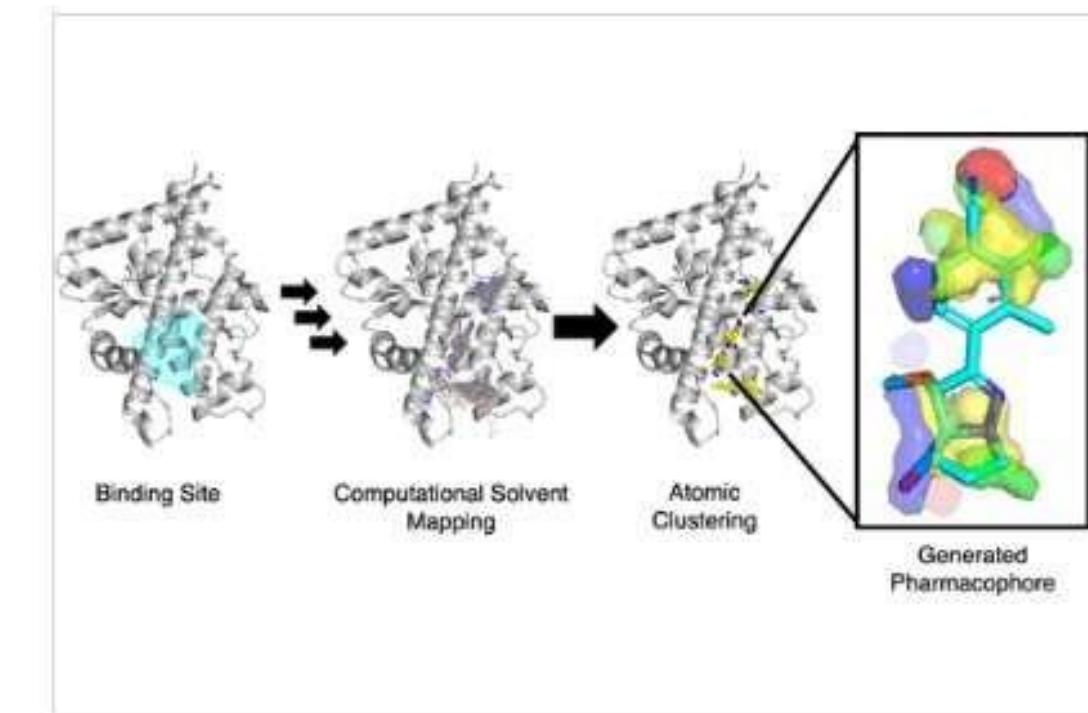
E-FTMap is an automated protocol that applies computational solvent mapping to identify important interactions ligand within binding sites. E-FTMap is based on the [FTMap algorithm](#).

For assistance with submitting jobs, please refer to the [Help](#) page.

Please [login](#) to use E-FTMap.



Developed by the [Vajda Lab at Boston University](#).  
Copyright © 2024. All rights reserved. Non-commercial use only.

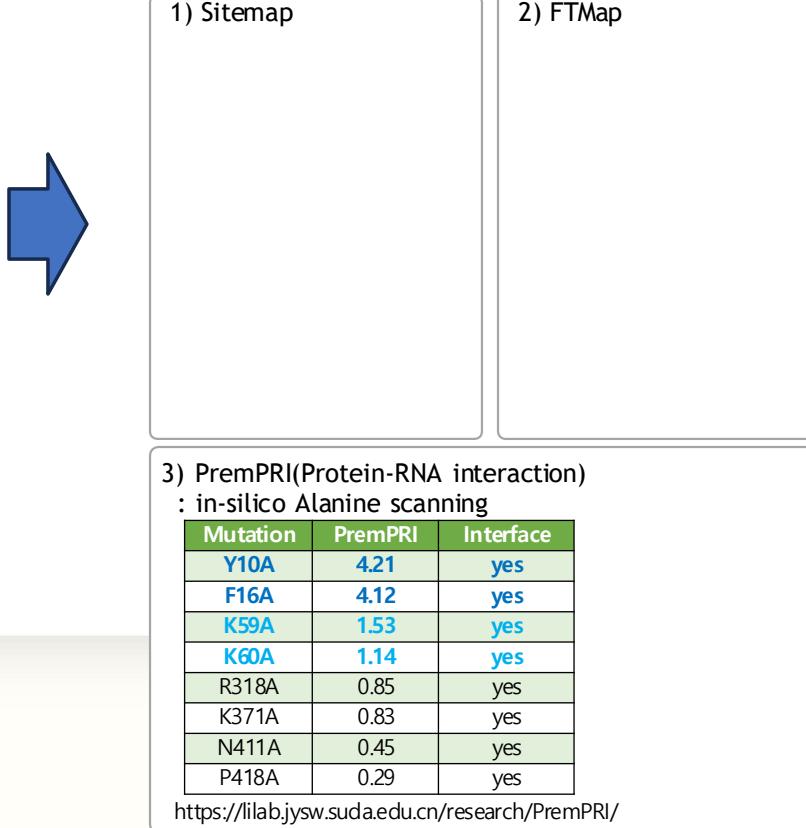


## • Case study: Protein-RNA binding inhibition

<Protein-RNA complex structure>

- Pocket prediction:

- 1) Sitemap(Schrodinger)
- 2) FTMap
- 3) PremPRI(Protein-RNA interaction)
- 4) Literature analysis

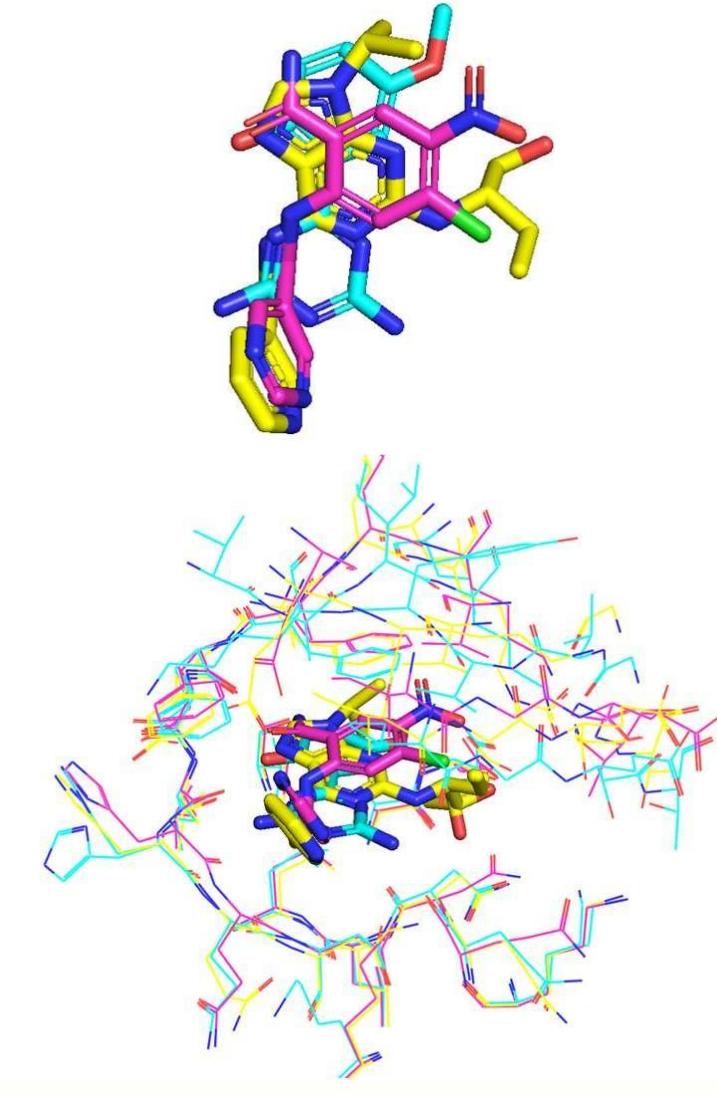
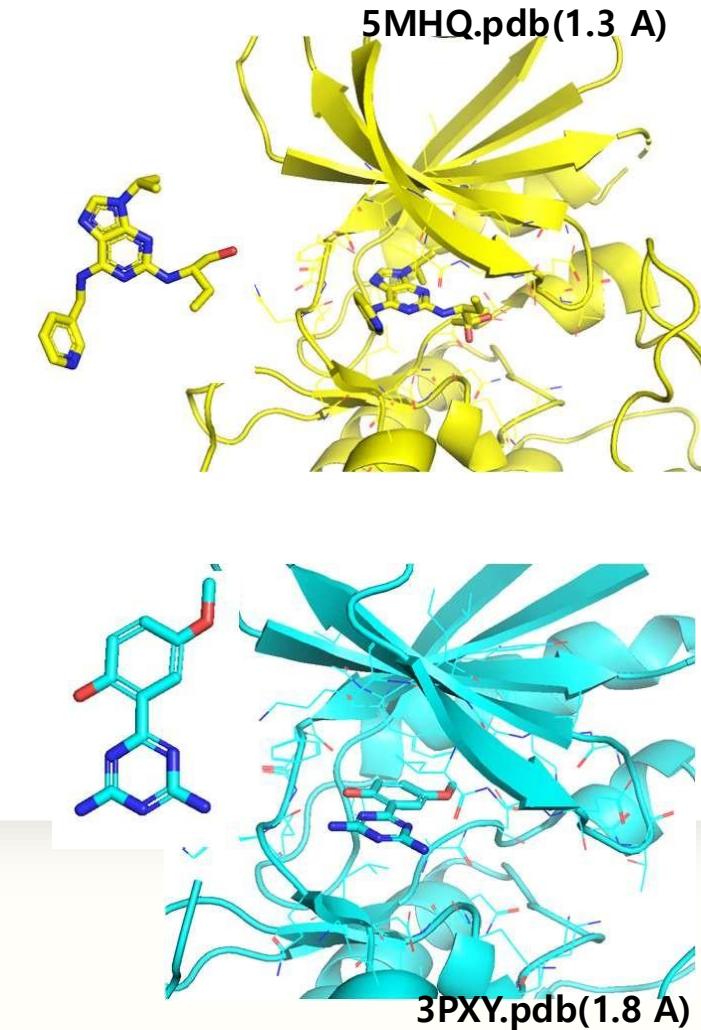
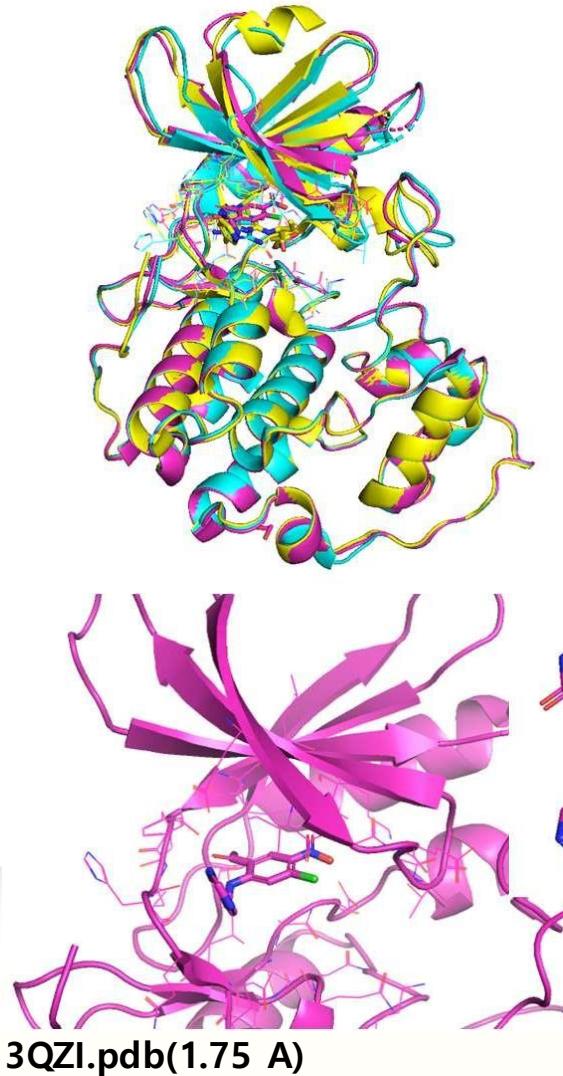


(1abc.pdb)

- 2개의 binding site이 예측됨
- Site 1 & site 2에 대한 구조기반 가상탐색 수행

- Binding Site Flexibility

- 동일한 human CDK2라도 결합한 ligand에 따라 binding pocket이 flexible함.
- PDB 구조 선택 중요



## • 구조기반 가상 스크리닝

: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것

- HTS vs. Virtual Screening
- Target-based vs. Phenotype-based
- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket: 단백질 flexibility & structure selection
- Protein-ligand Interaction/Binding affinity

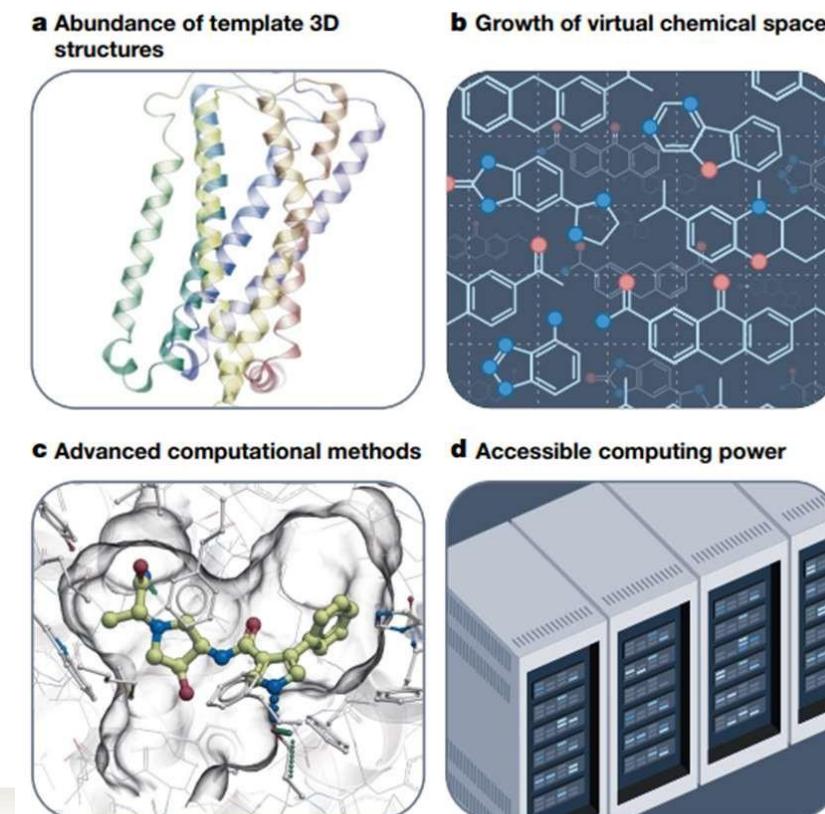
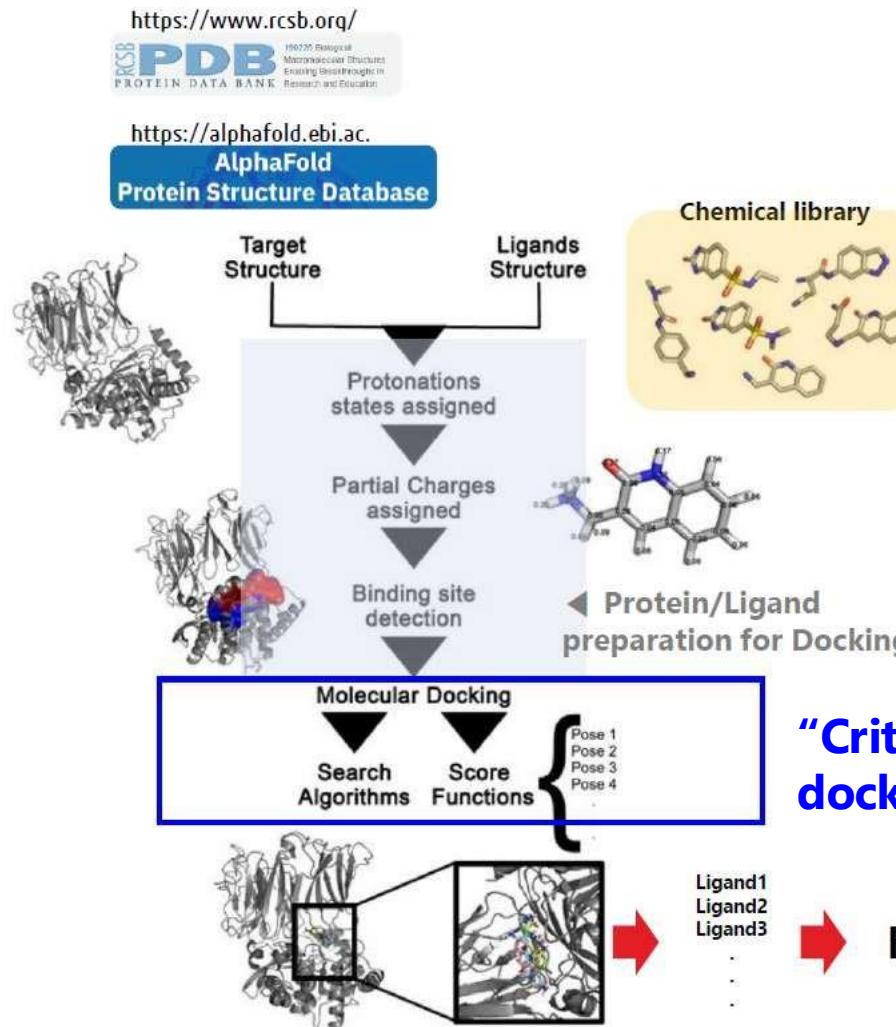


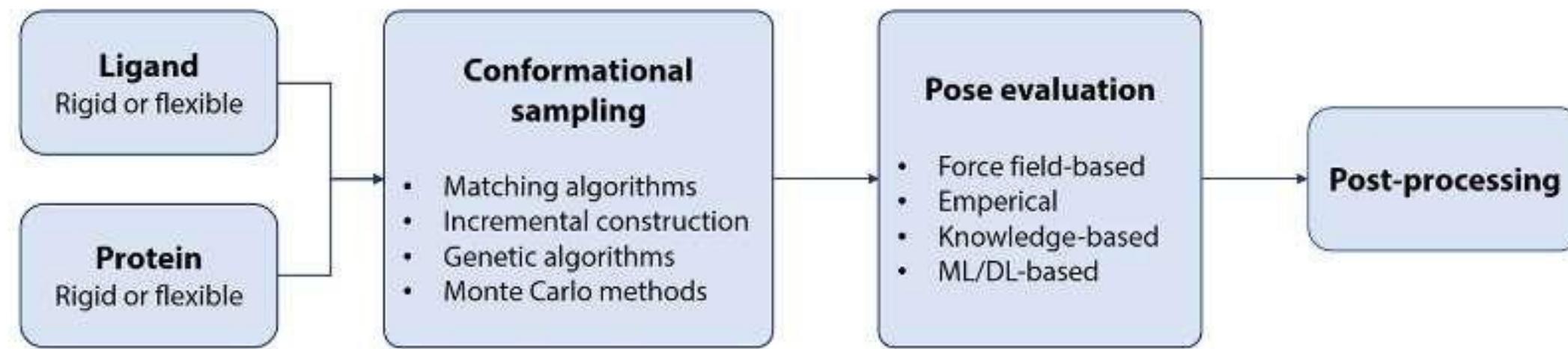
Fig. 1 | Key factors driving VLS technology breakthroughs for generation of high-quality hits and leads. a, More than 200,000 protein structures in the

Nature, 2023, 616, 673



- ✓ **Docking:** 타겟 단백질의 binding site에서 화합물(chemical/ligand)이 어떤 binding mode(=pose)를 갖는지?를 예측하는 방법
- ✓ **Docking based Virtual Screening:** 예측한 binding mode에서 protein-ligand interaction에 해당하는 score를 기준으로 물질선별
  - 실제 활성평가를 통해 active 여부를 확인
  - 후보물질(hit) 발굴에 활용

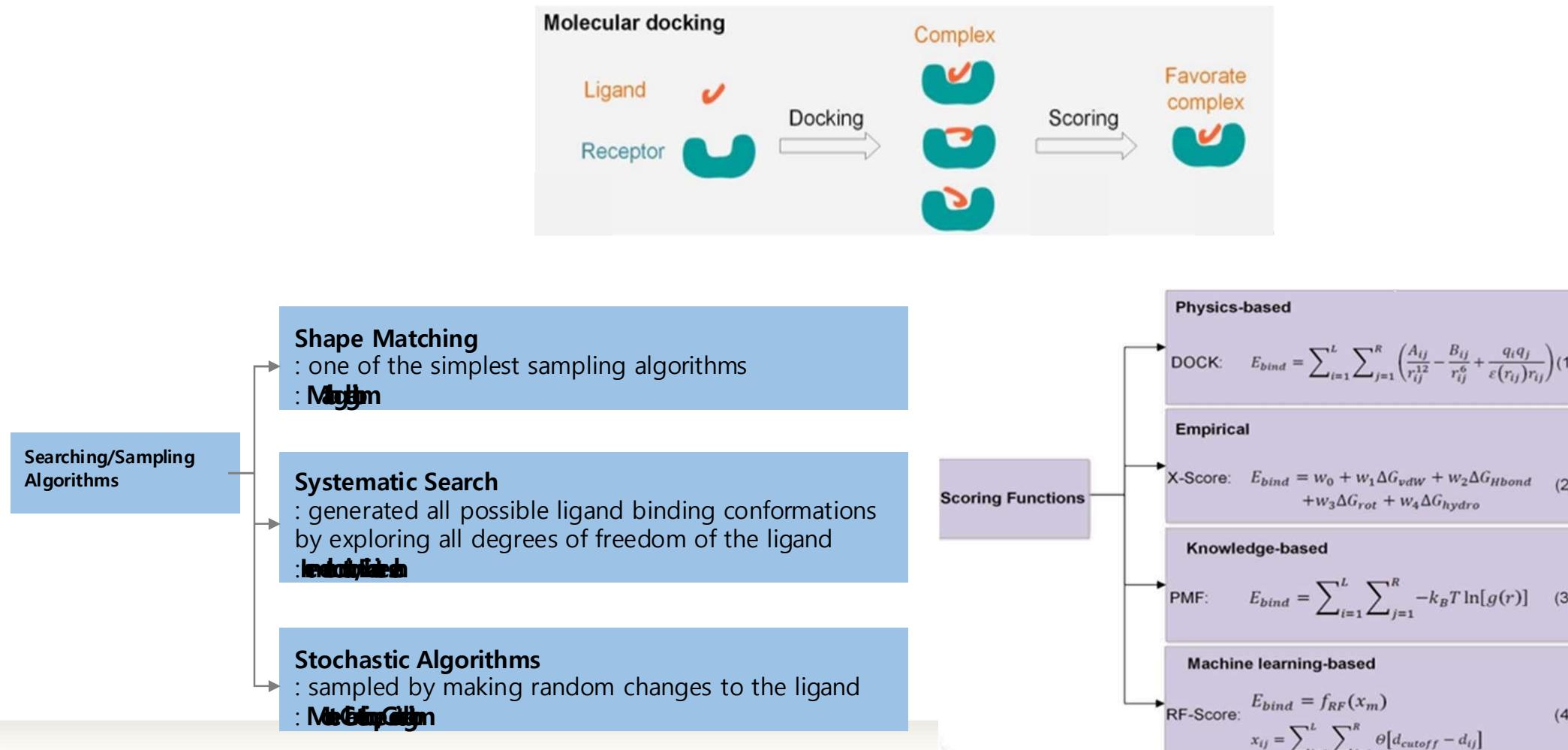
**“Critical components of docking process”**



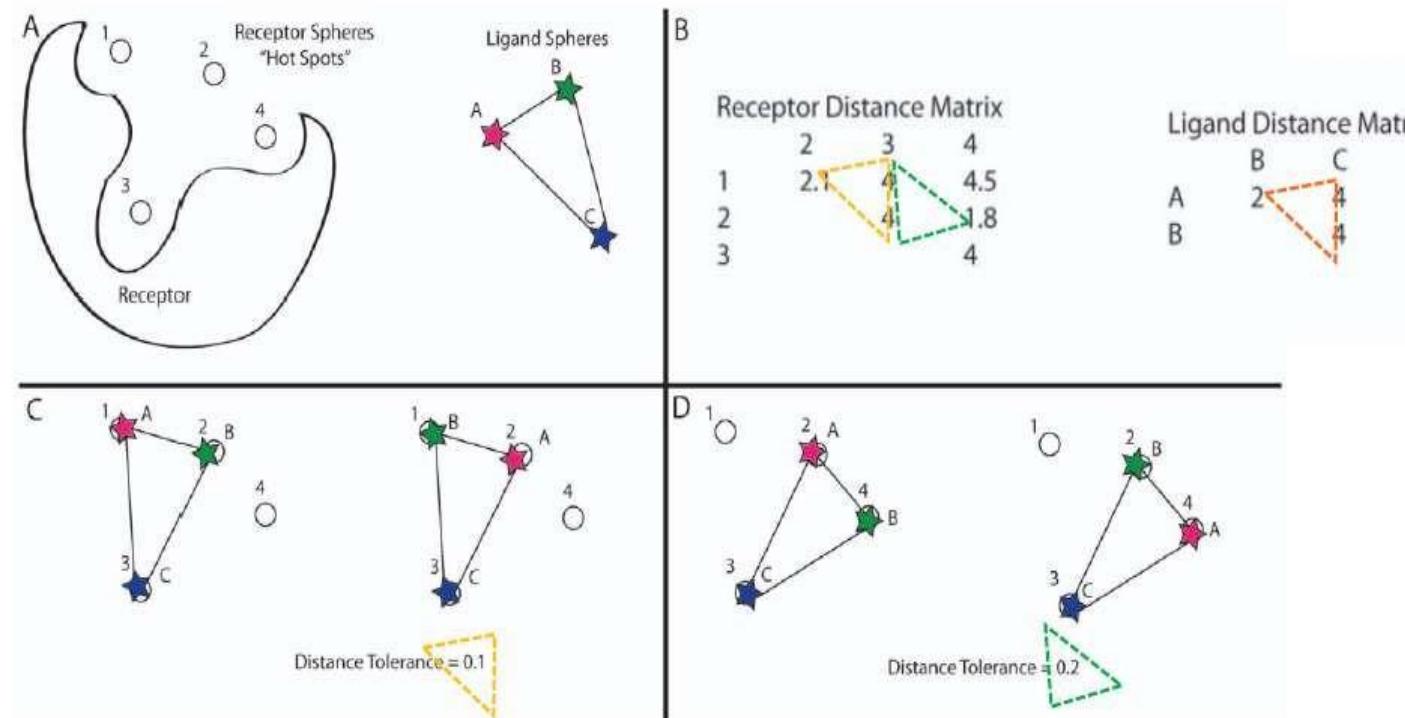
**Fig. 2:** Molecular docking workflow, Michele Wichmann and David Schaller

[https://projects.volkamerlab.org/teachopencadd/talktutorials/T015\\_protein\\_ligand\\_docking.html](https://projects.volkamerlab.org/teachopencadd/talktutorials/T015_protein_ligand_docking.html)

# AI 기반 신약 후보물질 탐색



- Searching/Sampling Algorithms (1) Matching algorithm
- Docking methods: LibDock, LIDAEUS, PhDOCK, Ph4DOCK, Q-fit, SANDOCK



Illustrating the matching sphere orientational **matching algorithm**.

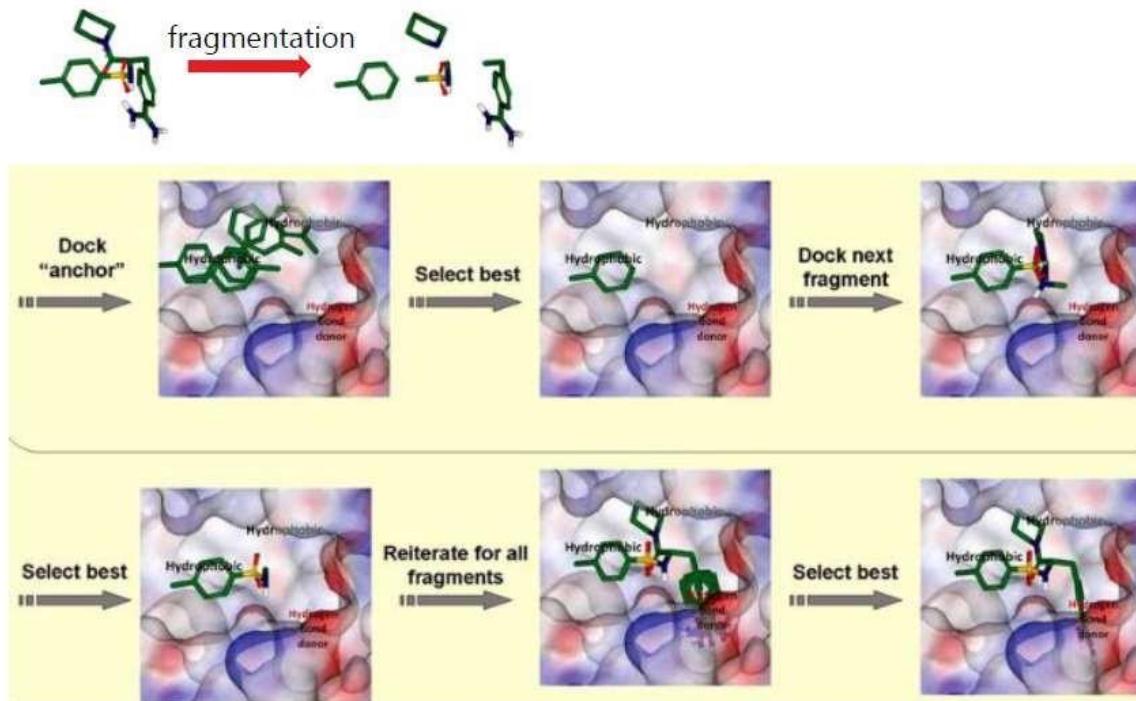
A) Receptor with 4 **matching spheres** shown as circles and a ligand with 3 spheres shown as stars.

B) The distance matrices constructed from these **spheres** are show in the upper right.

C) The 2 possible orientational matches of the ligand spheres (as stars) onto the receptor spheres **with a distance tolerance of 0.1**(assuming 3 matching nodes are used, in 3D this is usually 4).

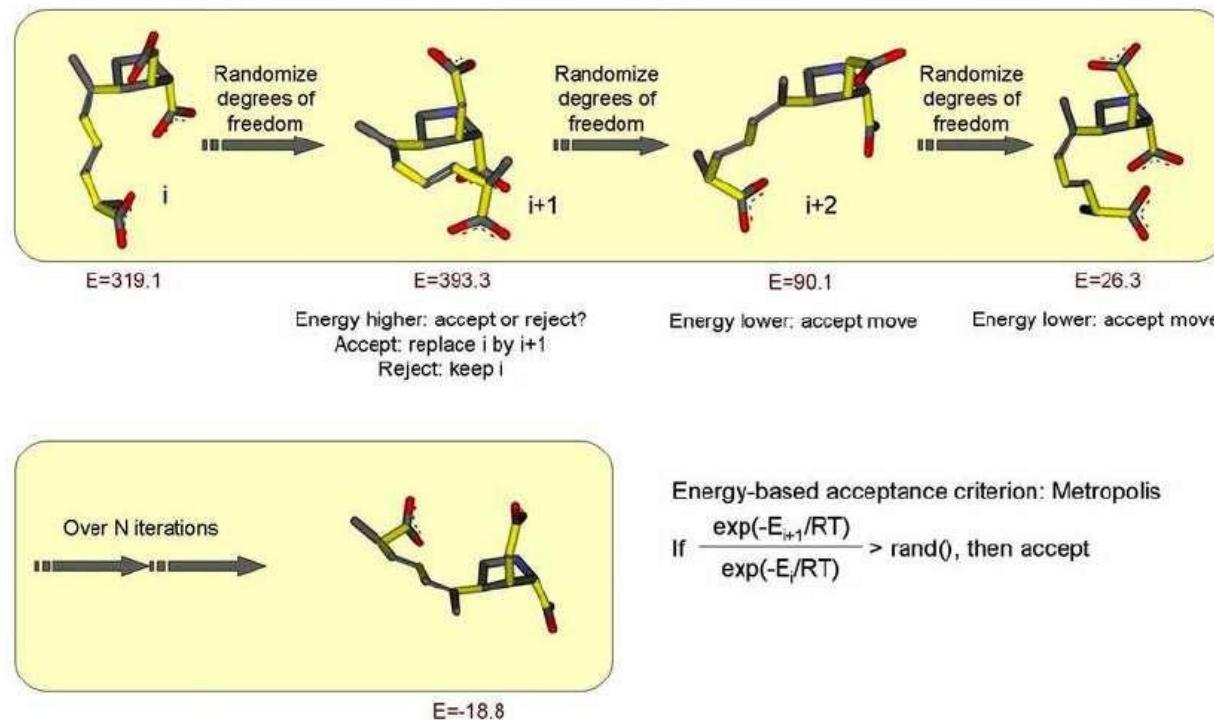
D) The additional two orientations produced when the distance tolerance is raised to 0.2.

- Searching/Sampling Algorithms (2) Incremental construction
- Docking methods: DOCK 4.0, FlexX, Hammerhead, SLIDE, eHiTS, SKELGEN, ProPose, PatchDock, MacDock, FLOG



The ligand is fragmented from rotatable bonds into various segments. One of the segments is anchored to the receptor surface. Once the anchor has been established, the next step is to add each of the fragment step by step. Ideally, those fragments are added first which have a greater chance of showing interactions like hydrogen bonding since they are directional in nature and are responsible for specificity of the ligand. Once a particular fragment is added, the poses with the least energies are considered for the next iteration, making the algorithm extremely fast and robust.

- Searching/Sampling Algorithms (3) Monte Carlo technique (MC)
- Docking methods: DockVision1.0.3, FDS, GlamDock, ICM, MCDOCK, PRODOCK, QXP, ROSETTALIGAND, RiboDock, Yucca, AutoDock



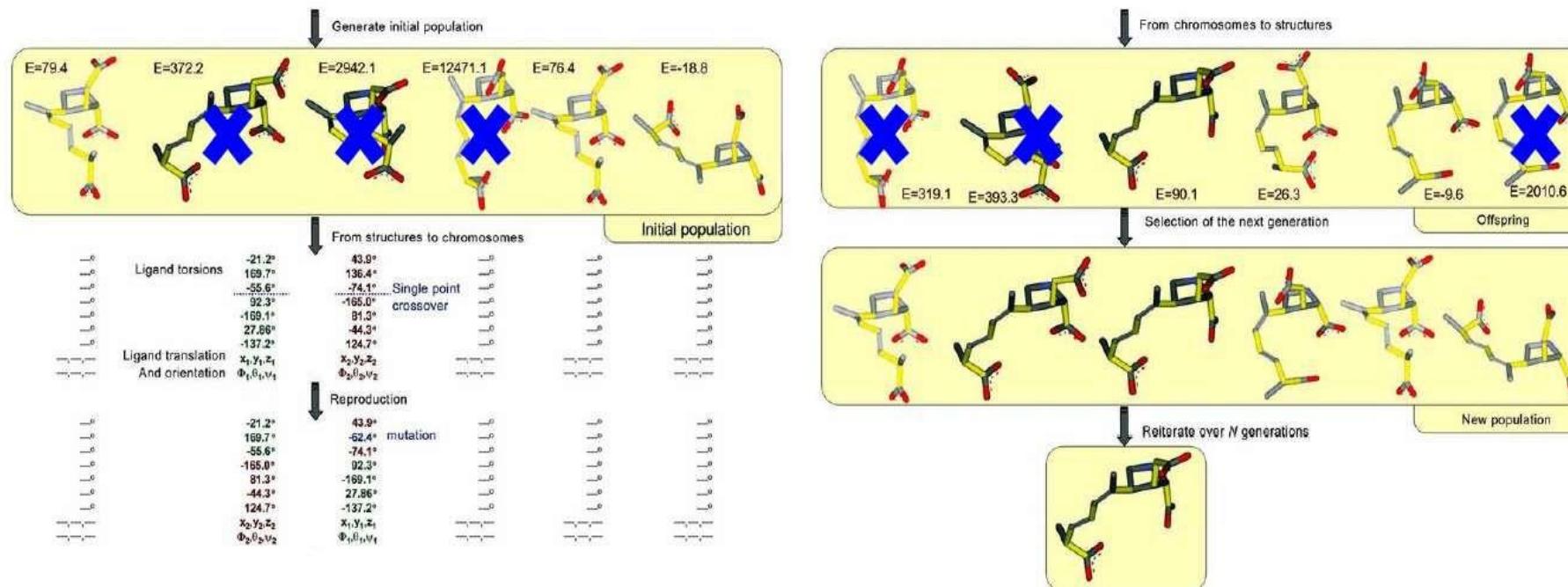
A ligand is modified **gradually using bond rotation and translation or rotation of the entire ligand**.

That conformation is then evaluated at the binding site **based on energy calculation** using molecular mechanics and **is then rejected or accepted for the next iteration based on Boltzmann's probability constant**.

Acceptance or rejection of the conformation is a function of the change in energy with respect to a parameter T, which can be physically interpreted as temperature (simulated annealing).

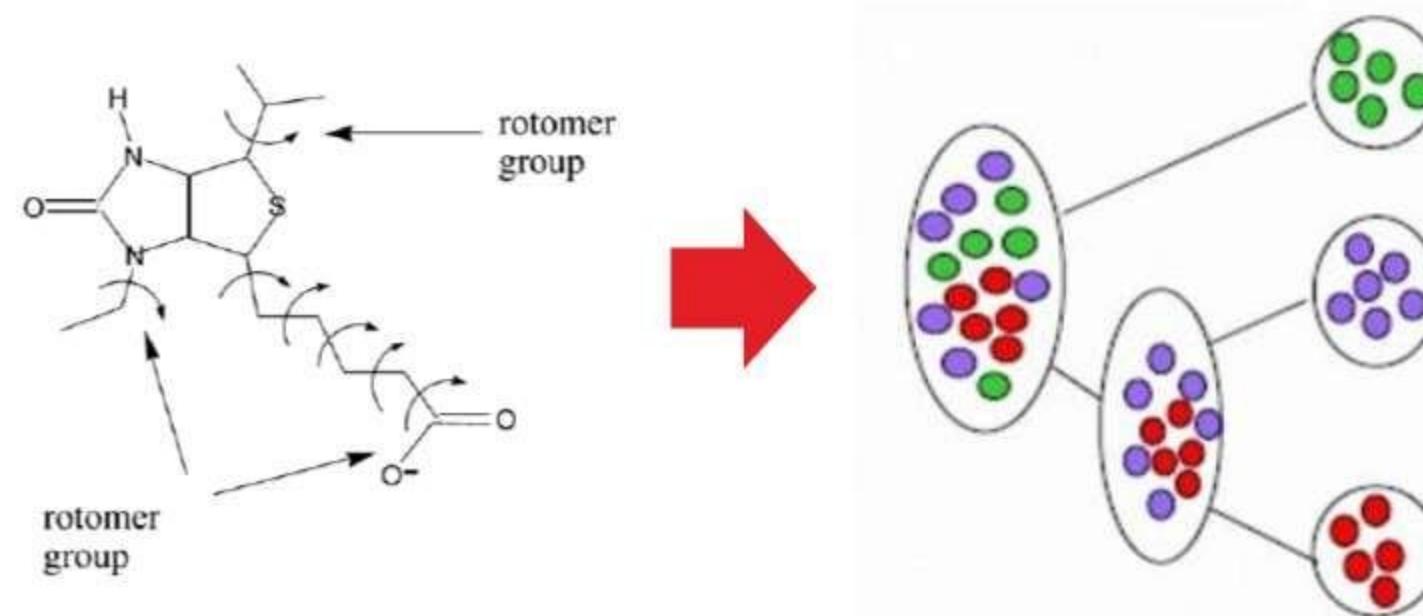
An interesting spin-off of the MC approach is the Tabu search, which maintains a record of the search space of the binding site which has already been visited and thus ensures that the binding site is explored to the maximum.

- Searching/Sampling Algorithms (4) Genetic Algorithm(GA)
- Docking methods: Autodock4.0, DARWIN, DIVALI, FITTED, FLIPDock, GAMBLER, GAsDock, GOLD 3.1, PSI-DOCK



- These are much inspired by the *Darwin's Theory of Evolution*.
- GA maintains a **population of ligands with an associated fitness determined by the scoring function**.
- The GA alters the ligands of the population **by mutation or crossover**.
- GA is basically used to find **the global minima**.

- Searching/Sampling Algorithms (5) Exhaustive search & Hierarchical method
- Docking methods: Glide



- In this approach, **the low energy conformations of the ligand are pre-generated and aligned.**
- The populations of the pre-generated ligand conformations **are merged into a hierarchy such that similar conformations** are positioned adjacent to each other within the hierarchy.
- Afterwards, on carrying out rotation or translation of the ligand, the docking program will make use of **this hierarchical data structure and thus minimize the outcomes.**
- Let us understand with a simple example—if an atom near the rigid center of the ligand is found to clash with the protein in given rotation/translation, then this approach **can reject all of conformations lying below in the hierarchy** to that of the conformation under scrutiny, because the descendants must contain the same clash as well .

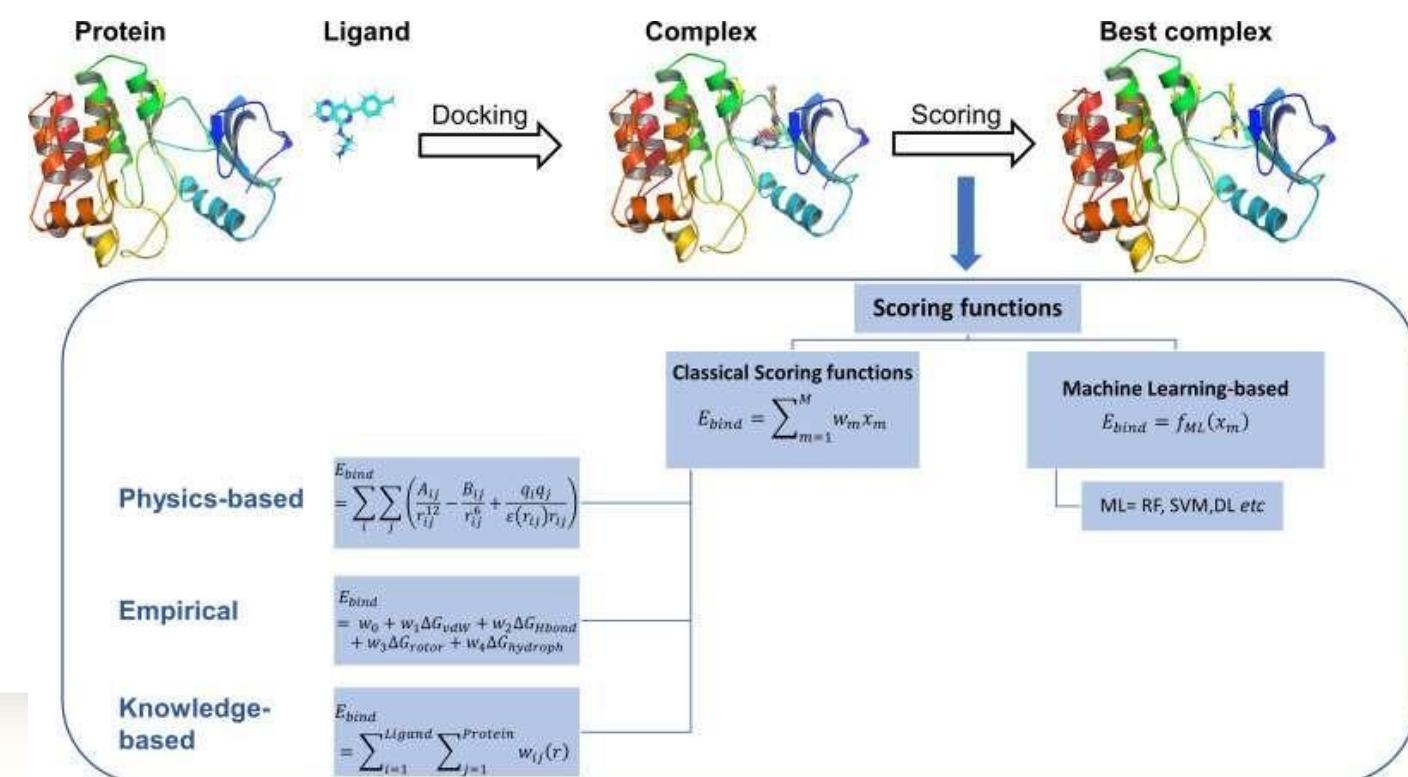
## &lt; Searching/Sampling Algorithms &gt;

Searching/Sampling algorithm	Docking methods	Cons.
<b>▼ Shape Matching:</b> one of the simplest sampling algorithms		
<b>Matching algorithm</b>	LibDock, LIDAEUS, PhDOCK, Ph4DOCK, Q-fit, SANDOCK	Fixed ligand conformation Ligand's initial conformation 중요
<b>▼ Systematic Search:</b> normally used for <b>flexible-ligand docking</b> , which generate all possible ligand binding conformations by exploring all degrees of freedom of the ligand		
<b>Incremental construction</b>	DOCK 4.0, FlexX, Hammerhead, SLIDE, eHiTS, SKELGEN, ProPose, PatchDock, MacDock, FLOG	Unfavorable conformation
<b>Exhaustive search &amp; Hierarchical method</b>	Glide	
<b>▼ Stochastic Algorithms:</b> ligand binding orientations and conformations are sampled by making random changes to the ligand at each step in both the conformational space and the translational/rotational space of the ligand, respectively. The random change will be accepted or rejected according to a probabilistic criterion.		
<b>Monte Carlo technique</b>	DockVision 1.0.3, FDS, GlamDock, ICM, MCDOCK, PRODOCK, QXP, ROSETTALIGAND, RiboDock, Yucca, AutoDock	Convergence or not Ligand's initial conformation 중요
<b>Genetic algorithm</b>	Autodock 4.0, DARWIN, DIVALI, FITTED, FLIPDock, GAMBLER, GAsDock, GOLD 3.1, PSI-DOCK	

- Scoring Functions

Searching Algorithms generate possible pose, which  
are ranked by Scoring function

*Int. J. Mol. Sci.* 2019, 20, 4574



- Scoring function (1) Force Field functions

- Be developed ***based on physical atomic interactions*** like van der Waals interactions, electrostatic interactions and bond lengths, bond angles and torsions.

□ Force field functions and parameters are usually ***derived from both experimental data and ab initio quantum mechanical calculations*** according to the principles of physics.

$$E = \sum_i \sum_j \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right)$$

- $r_{ij}$  stands for the distance between protein atom i and ligand atom j,
- $A_{ij}$  and  $B_{ij}$  are the van der Waal parameters,
- $q_i$  and  $q_j$  are the atomic charges
- $\epsilon(r_{ij})$  is the distance dependent dielectric constant.

□ **It does not consider an important solvent effect** that charged groups favor aqueous environments whereas non-polar groups tend to stay in non-aqueous environments, commonly referred to as the desolvation effect.

- Scoring function (2) Empirical Scoring function

- The binding energies of a complex can be approximated **by a sum of individual uncorrelated terms.**

- The coefficients of the various terms** involved in calculation of binding energy are obtained from regression analysis **using experimentally determined** binding energies or potentially from X-ray structural information.
- Empirical functions have simpler energy terms to evaluate when compared to force field scoring functions and thus are **much faster in binding score calculations.**

$$\Delta G_{bind} = \Delta G_O + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic int.} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} |A_{lipo}| \\ + \Delta G_{rot} NROT.$$

✓ Because of many different energy terms,  
**→ the problem of unknowingly double counting of certain energy terms**

- $\Delta G_O$ : binding energy independent of protein interactions
- $\Delta G_{hb}$ : contribution to binding energy from hydrogen bonds
- $\Delta G_{ionic}$ : contribution to binding energy from unperturbed ionic interactions
- $\Delta G_{lipo}$ : contribution to binding energy through lipophilic interactions
- $A_{lipo}$ : lipophilic contact surface between the protein and the ligand
- **$\Delta G_{rot}$ : the loss of binding energy due to freezing of internal degrees of freedom in the ligand** → **solute entropy terms**
- $NROT$ : number of rotatable bonds
- **$f(\Delta R, \Delta \alpha)$  is a penalty function** that accounts for large deviations from ideal hydrogen bond and salt bridge geometry.

- Scoring function (3) Knowledge based scoring functions

- Be derived from the structural information embedded in ***experimentally determined atomic structures***
- The functions use ***statistical analysis on crystal structures of complexes*** to obtain the interatomic contact frequencies between the protein and the ligand ***based on the presumption that stronger an interaction is, the greater the frequency of its occurrence will be.***

$$w(r) = -k_B T \ln [g(r)], g(r) = (r)\rho(r)/\rho^*(r)$$

- $k_B$  is the Boltzmann constant,
- $T$  is the absolute temperature of the system,
- $\rho(r)$  is the number density of the protein-ligand atom at distance  $r$ ,
- $\rho^*(r)$  is the pair density in the reference state where interatomic interactions are zero and
- $g(r)$  is pair distribution function

- the accuracy of predicting the reference state and **underrepresentation of interactions with halogens and metals** are the major hurdles

## • Docking Program

Software	Posing	Scoring	Availability				
Vina	Iterated Local Search + BFGS Local Optimiser Lamarckian Genetic Algorithm, Genetic Algorithm or Simulated Annealing	Empirical/ Knowledge-Based	Free (Apache License)	GOLD	Genetic Algorithm	Physics-based (GoldScore), Empirical (ChemScore, ChemPLP) and Knowledge-based (ASP)	Commercial
AutoDock4	Differential Evolution (Alternatively Simplex Evolution and Iterated Simplex)	Semiempirical	Free (GNU License)	GEMDOCK	Generic Evolutionary Algorithm	Empirical (includes pharmacophore potential)	Free (for non-commercial research)
Molegro/MolDock	Monte Carlo stochastic sampling + local optimisation	Semiempirical	Commercial	Dock6	Anchor-and-grow incremental construction	Physics-based (several other options)	Academic License
Smina	Ant Colony Optimisation	Empirical (customisable)	Free (GNU License)	GAsDock	Entropy-based multi-population genetic algorithm	Physics-based	*
Plants	Biased Probability Monte Carlo + Local Optimisation	Empirical	Academic License	FlexX	Fragment-Based Pattern-recognition (Pose Clustering) + Incremental Growth	Empirical	Commercial
ICM	Systematic search + Optimisation (XP mode also uses anchor-and-grow)	Physics-Based	Commercial	Fred	Conformer generation + Systematic rigid body search	Empirical (defaults to Chemgauss3)	Commercial
Glide	Fragmentation and alignment to idealised molecule (Protomol) + BFGS optimisation	Empirical	Commercial	DockThor	Steady-state genetic algorithm (with Dynamic Modified Restricted Tournament Selection method)	Physics-based + Empirical	Free (Webserver)

#10 of docking programs: LigandFit, Glide, GOLD, MOE Dock, Surflex-Dock, AutoDock, AutoDockVina, LeDock, rDock, UCSF DOCK

- Diverse set of 2002 protein–ligand complexes from the PDBbindDB (version 2014)
- Prediction accuracy of sampling power and scoring power
- RMSD between the docked binding pose and the native binding pose
- Pearson's correlation coefficient( $r_p$ ) and Spearman's ranking coefficient ( $r_s$ ) between the docking scores and experimental binding data

Docking program	1 <sup>st</sup> ranked pose		#20 top of rank에 포함된 correct pose	
	Regular organic molecule		Peptide or peptide mimic	
	Top scored pose	Best pose	Top scored pose	Best pose
AutoDock (LGA)	0.378	0.559	0.216	0.324
AutoDock (PSO)	0.477	0.686	0.331	0.439
AutoDock Vina	0.485	0.726	0.384	0.597
LeDock	0.574	0.808	0.352	0.465
rDock	0.503	0.763	0.283	0.465
UCSF DOCK	0.445	0.591	0.340	0.415
LigandFit	0.479	0.689	0.267	0.504
Glide (SP)	0.544	0.754	0.403	0.547
Glide (XP)	0.584	0.666	0.403	0.484
GOLD	0.599	0.726	0.371	0.472
MOE Dock	0.457	0.612	0.195	0.245
Surflex-Dock	0.533	0.800	0.440	0.673

- 1<sup>st</sup> pose가 correct pose는 아니다.
- Chemical에 대해서 적어도 correct pose가 포함되게 다수의 pose를 살펴봐야 한다.
- Searching algorithm OK, Scoring function not good!!

Docking program	Correlation coefficient	Top scored pose	Best pose
AutoDock (LGA)	$r_p^a$	$0.433 \pm 0.009^c$	$0.404 \pm 0.009$
	$r_s^b$	$0.477 \pm 0.008$	$0.450 \pm 0.009$
AutoDock (PSO)	$r_p$	$0.492 \pm 0.008$	$0.466 \pm 0.008$
	$r_s$	$0.534 \pm 0.007$	$0.513 \pm 0.008$
AutoDock Vina	$r_p$	$0.564 \pm 0.008$	$0.569 \pm 0.008$
	$r_s$	$0.580 \pm 0.008$	$0.584 \pm 0.008$
LeDock	$r_p$	$0.442 \pm 0.009$	$0.463 \pm 0.009$
	$r_s$	$0.462 \pm 0.010$	$0.486 \pm 0.009$
rDock	$r_p$	$-0.015 \pm 0.011$	$-0.021 \pm 0.011$
	$r_s$	$-0.017 \pm 0.011$	$-0.005 \pm 0.011$
UCSF DOCK	$r_p$	$0.291 \pm 0.010$	$0.276 \pm 0.011$
	$r_s$	$0.331 \pm 0.011$	$0.323 \pm 0.011$
LigandFit	$r_p$	$-0.132 \pm 0.011$	$-0.105 \pm 0.011$
	$r_s$	$-0.221 \pm 0.012$	$-0.192 \pm 0.012$
Glide (SP)	$r_p$	$0.444 \pm 0.008$	$0.402 \pm 0.009$
	$r_s$	$0.473 \pm 0.009$	$0.419 \pm 0.010$
Glide (XP)	$r_p$	$0.367 \pm 0.010$	$0.356 \pm 0.010$
	$r_s$	$0.389 \pm 0.010$	$0.374 \pm 0.010$
GOLD	$r_p$	$-0.500 \pm 0.008$	$-0.494 \pm 0.008$
	$r_s$	$-0.515 \pm 0.008$	$-0.513 \pm 0.008$
MOE Dock	$r_p$	$0.564 \pm 0.008$	$0.411 \pm 0.009$
	$r_s$	$0.589 \pm 0.009$	$0.457 \pm 0.009$
Surflex-Dock	$r_p$	$-0.340 \pm 0.009$	$-0.350 \pm 0.009$
	$r_s$	$-0.370 \pm 0.009$	$-0.382 \pm 0.009$

← Scoring functions not good^^;;

<sup>a</sup>  $r_p$  represents Pearson's correlation coefficient. <sup>b</sup>  $r_s$  represents Spearman's ranking coefficient. <sup>c</sup> The standard error was estimated by randomly sampling 80% of the tested dataset 100 repeats.

- Docking의 장점

- Ligand Binding Pose
- Balance between time/computational cost and accuracy  virtual screening

- Docking의 한계

- ✓ Receptor flexibility
- ✓ Desolvation effects
- ✓ Correlation between docking score and binding affinity/activity



### Binding Affinity Prediction

#### ➤ End-point Methods

- Molecular Mechanics Poisson-Boltzmann Surface Area (MMPSA)
- Molecular Mechanics Generalized Born Surface Area (MMGBSA)
- Linear Interaction Energies (LIE)

#### ➤ Alchemical Transfer Methods

- Free Energy Perturbation (FEP)
- Thermodynamic Integration (TI)

- : MM-PBSA is an endpoint method that calculates binding free energy without consideration of any intermediate state.
- : Sample conformations only with the free and bound states and get binding free energy by taking a difference between these states
- : Splits the binding free energy into molecular mechanics terms and solvation energies to be calculated separately. The basic principle is to calculate the difference between the binding free energy of two solvated molecules in the bound and unbound states or to compare the free energy of different solvation conformations of the same molecule. Taking the binding of receptor or protein and small ligand molecule as an example, the calculation principle will be roughly described.
- : Good balance between accuracy and computational efficiency
  - Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)
  - Molecular Mechanics Generalized Born Surface Area (MMGBSA)
  - Linear interaction energies (LIE)

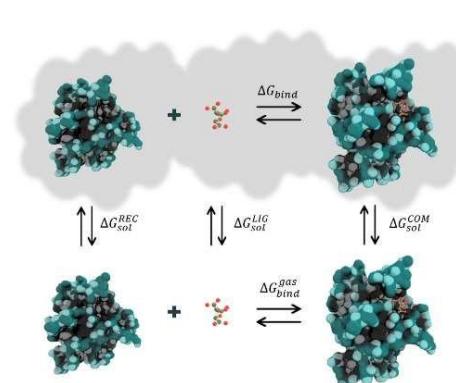
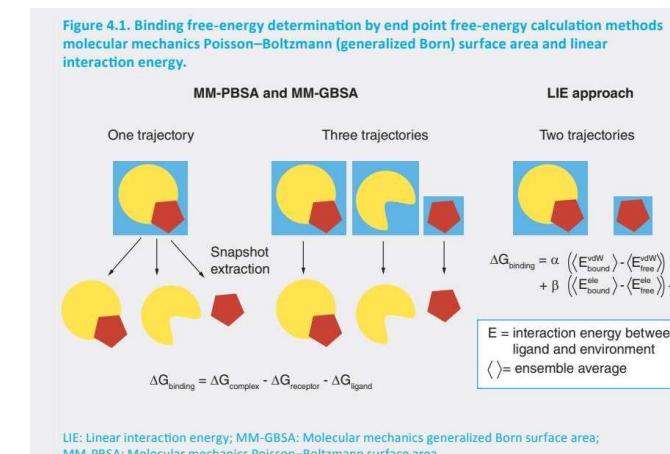


Figure 1. Thermodynamic cycle for binding free energy calculations

[https://valdes-tresanco-ms.github.io/gmx\\_MMPSA/dev/introduction/](https://valdes-tresanco-ms.github.io/gmx_MMPSA/dev/introduction/)



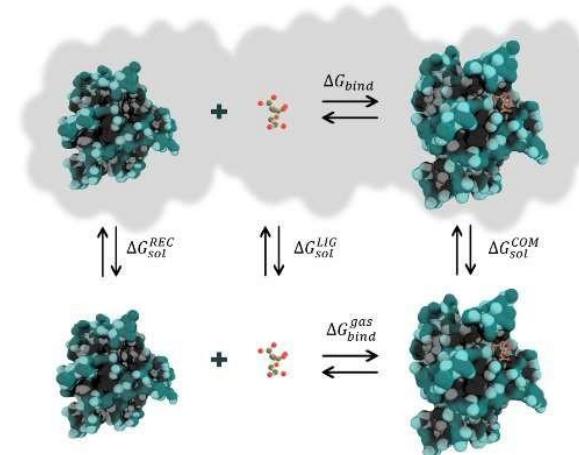


Figure 1. Thermodynamic cycle for binding free energy calculations

J. Chem. Theory Comput. 2021, 17, 6281–6291

The MM/PB(GB)SA method can be used for calculating binding free energies of noncovalently bound complexes.

The free binding energy for a complex can be estimated as follows:

$$\Delta G_{\text{bind}} = \langle G_{\text{COM}} \rangle - \langle G_{\text{REC}} \rangle - \langle G_{\text{LIG}} \rangle$$

where each term to the right in the equation is given by

$$\langle G_x \rangle = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - \langle TS \rangle$$

In turn,  $\Delta G_{\text{bind}}$  can also be represented as

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S$$

The  $\Delta H$  can be decomposed into different terms:

$$\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{sol}}$$

where

$$\begin{aligned} \Delta E_{\text{MM}} &= \Delta E_{\text{bonded}} + \Delta E_{\text{nonbonded}} \\ &= (\Delta E_{\text{bond}} + \Delta E_{\text{angle}} + \Delta E_{\text{dihedral}}) + (\Delta E_{\text{ele}} + \Delta E_{\text{vdW}}) \end{aligned}$$

and

$$\Delta G_{\text{sol}} = \Delta G_{\text{polar}} + \Delta G_{\text{non-polar}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{non-polar}}$$

where

$$\Delta G_{\text{NP-polar}} = N P_{\text{TENSION}} \times \Delta \text{SASA} + N P_{\text{OFFSET}}$$

or

$$\begin{aligned} \Delta G_{\text{non-polar}} &= \Delta G_{\text{disp}} + \Delta G_{\text{cavity}} \\ &= \Delta G_{\text{disp}} + (\text{CAVITY}_{\text{TENSION}} \times \Delta \text{SASA} + \text{CAVITY}_{\text{OFFSET}}) \end{aligned}$$

# MM/PB(GB)SA Binding Free Energy

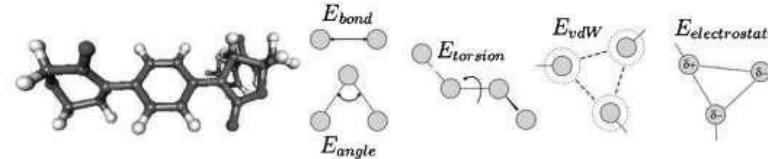
## Methods Corner

DOI: 10.1002/minf.201100135

### Free Energy Calculations by the Molecular Mechanics Poisson – Boltzmann Surface Area Method

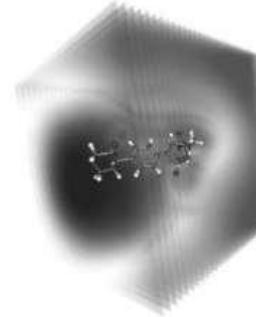
Nadine Homeyer<sup>[a]</sup> and Holger Gohlke<sup>[a]</sup>

**Molecular Mechanics**

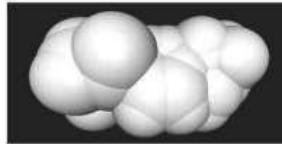


$$E_{MM} = \sum_{bonds} E_{bond} + \sum_{angles} E_{angle} + \sum_{torsions} E_{torsion} + \sum_{atoms}^{atoms} E_{vdW} + \sum_{i \neq j}^{atoms} E_{Electrostatic}$$

**Poisson Boltzmann**

$$-\nabla \cdot \epsilon(x) \nabla \phi(x) + \kappa^2(x) \phi(x) = 4\pi\rho_f(x)$$


**Surface Area**



$$G_{solv}^{SA} = \gamma SA + b$$

$$G_{Molecule} = E_{MM} + G_{solv}^{PB} + G_{solv}^{SA} - TS$$

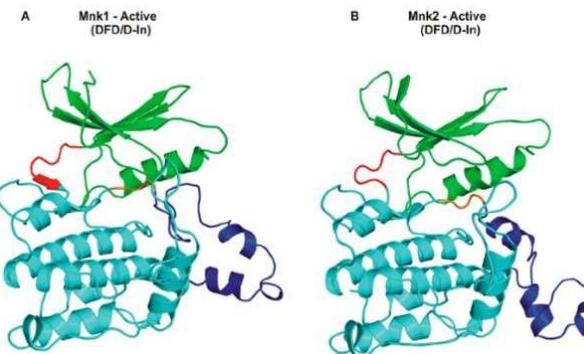
Holger Gohlke is a Professor of Pharmaceutical and Medicinal Chemistry at the Heinrich-Heine-University, Düsseldorf. His research aims at understanding and predicting receptor-ligand interactions and the modulation of biological processes by pharmacologically relevant molecules. His group develops and applies methods at the interface of computational pharmaceutical and biophysical chemistry and molecular bioinformatics.



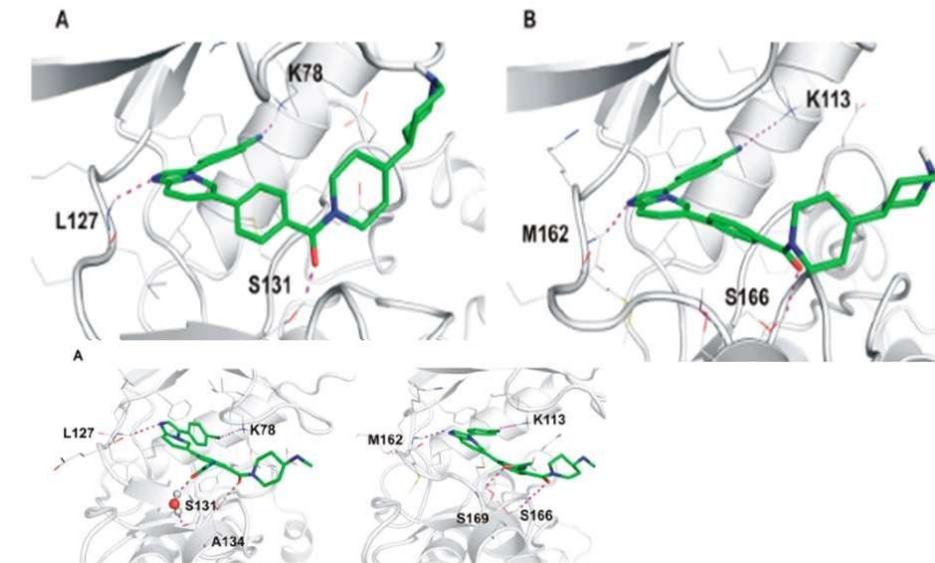
<https://cpclab.uni-duesseldorf.de/>

## ➤ Mnk1/2 Dual Inhibitors (Cancer)

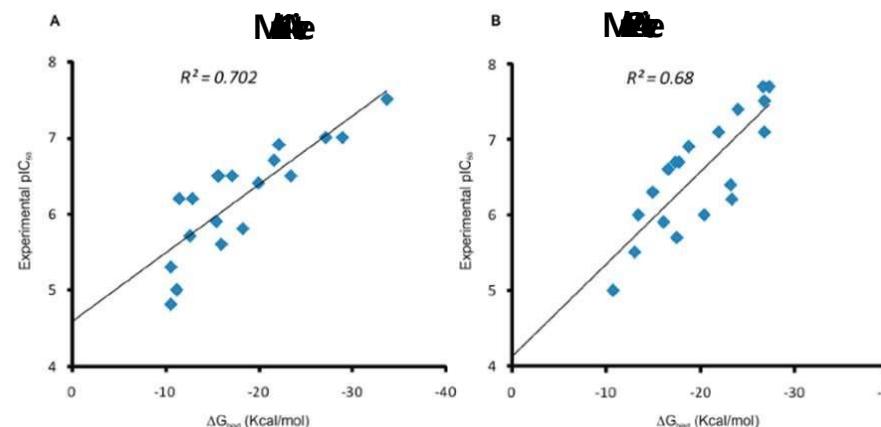
### <Homology modeling for Active forms>



### <Docking Study>



### <MMPBSA for 25 cpds>

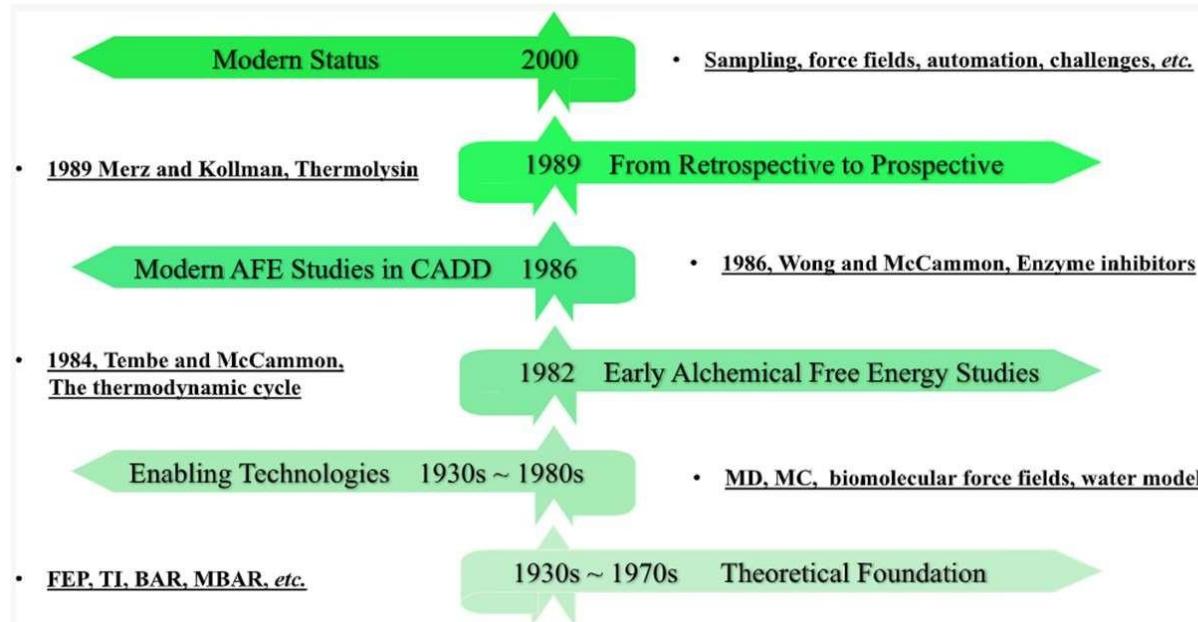


Biochemistry 2015, 54, 32–46

compd	$\Delta E_{\text{elec}}$	$\Delta E_{\text{vdw}}$	$\Delta G_{\text{SA}}^b$	$\Delta G_{\text{PB}}$	$\Delta H_{(\text{PB})}^c$	$T\Delta S^d$	$\Delta G_{\text{pred(PB)}}^e$	pIC <sub>50</sub> <sup>f</sup>
1	-14.77	-49.35	-5.88	37.46	-32.54	-10.26	-22.28	6.99
2	-18.01	-49.99	-6.2	37.5	-36.69	-15.00	-21.69	6.69
4	-26.66	-59.09	-6.75	51.59	-40.92	-13.73	-27.19	7.04
5	-12.71	-56.02	-6.49	34.69	-40.53	-11.5	-29.03	7.00
6	-14.77	-53.54	-6.37	40.33	-34.35	-17.09	-17.26	6.48
7	-14.29	-49.79	-5.89	38.06	-31.91	-19.26	-12.65	5.75
8	-18.87	-54.27	-6.27	42.86	-36.55	-14.99	-21.56	7.64
9	-25.94	-56.32	-6.28	47.5	-41.04	-7.23	-33.81	7.54

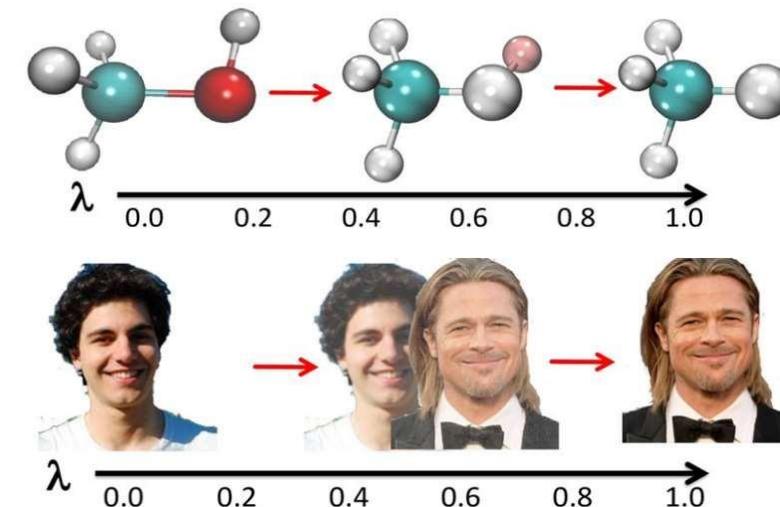
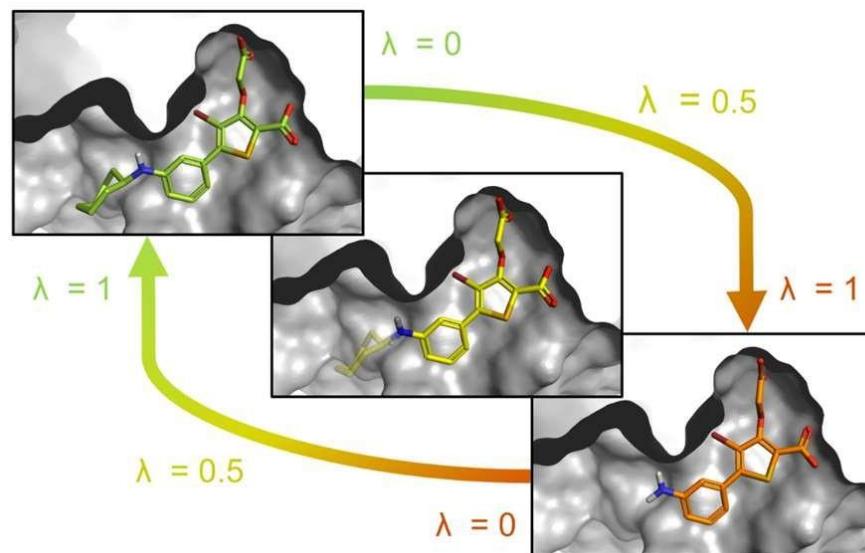
<sup>a</sup>Mean energies are in kcal/mol. <sup>b</sup> $\Delta G_{\text{SA}} = \gamma \times \text{SASA} + \beta$ ;  $\gamma = 0.00542 \text{ kcal/mol A}^{-2}$ ;  $\beta = 0.92 \text{ kcal/mol}$ . <sup>c</sup> $\Delta H_{(\text{PB})} = \Delta E_{\text{elec}} + \Delta E_{\text{vdw}} + \Delta G_{\text{SA}} + \Delta G_{\text{PB}}$ .

<sup>d</sup> $T\Delta S$  = entropy changes. <sup>e</sup> $\Delta G_{\text{pred(PB)}}$  (calculated binding free energy by MMPBSA method) =  $\Delta H_{(\text{PB})} - T\Delta S$ . <sup>f</sup>pIC<sub>50</sub> =  $(-\log \text{IC50})$ .



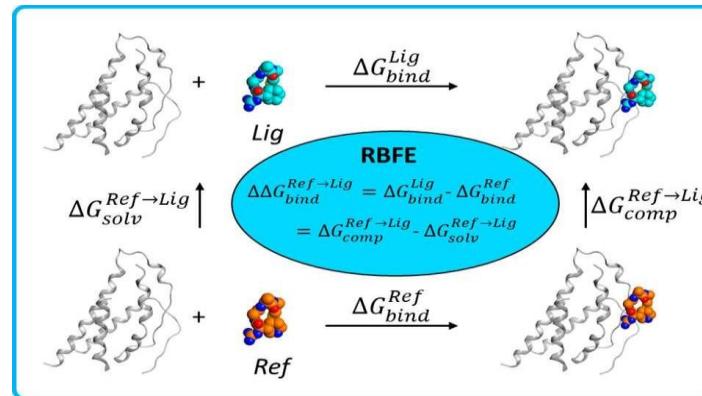
<https://pubs.acs.org/toc/jcisd8/60/11>

- Predict the relative binding affinity changes ( $\Delta\Delta G$ ) within a **congeneric ligand series**
- This is achieved by non-physical ('**alchemical**') **transformations**, in which a molecule (A) is gradually converted into a **structurally related molecule (B)** through a number of discrete steps, the so-called  $\lambda$  windows.
- The ligand simulated in each window can be thought of as an alchemical (i.e., hybrid) molecule consisting of a  $1-\lambda$  fraction of A and a  $\lambda$  fraction of B.

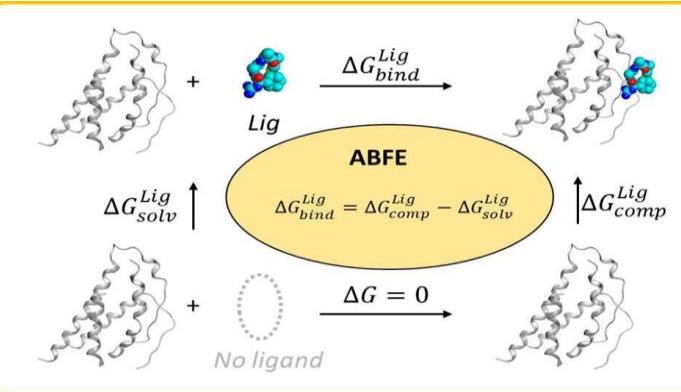


<https://www.cresset-group.com/about/news/flare-alchemical-fep/>

## &lt; Relative Binding Free Energy &gt;



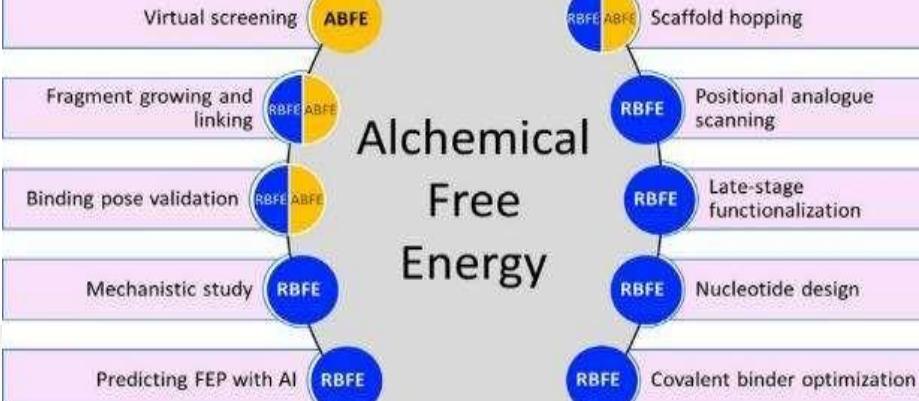
## &lt; Absolute Binding Free Energy &gt;



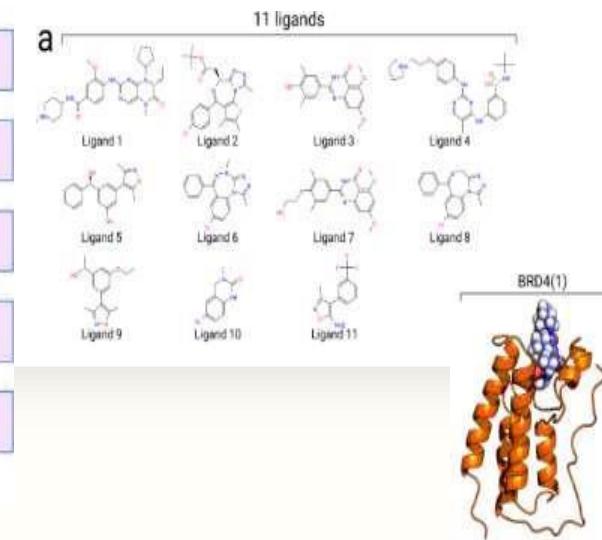
## congeneric ligand series

Tyk2 PDBID: 4GIH

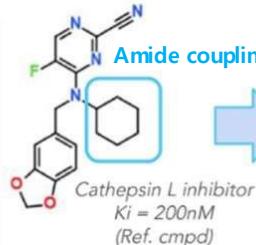
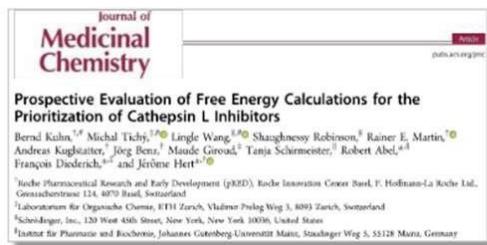
	$\Delta G_{exp} / \text{kcal mol}^{-1}$
1	-9.54
2	-10.94
3	-8.98
4	-11.31
5	-9.21
6	-8.26
7	-10.91
8	-7.75
9	-9.56
10	-7.42
11	-11.28
12	-9.00
13	-9.70
14	-11.70
15	-9.78
16	-10.53

Alchemical  
Free  
Energy

## Diverse ligand series



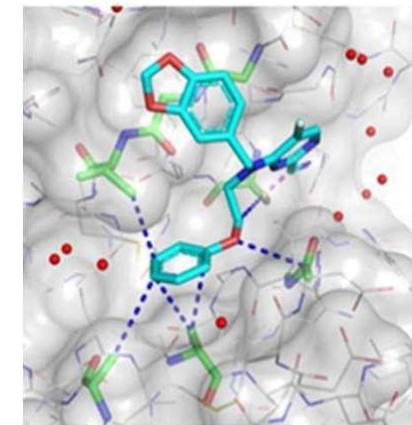
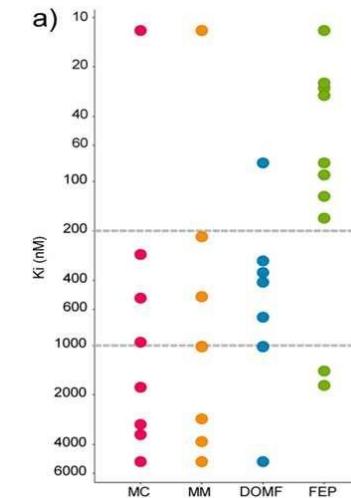
## • Free Energy Calculations (FEP+)



3,325 ideas

Potency optimization  
with new S2 pocket  
substituents

- Collaboration with Roche
- Head-to-head comparison of 4 prioritization processes
- Selection of 10 cmpds each
- Experimental testing



Co-crystal structure of 35 in complex with hCatL

FEP+		Med Chem		SBDD (Any tool except FEP+)		Docking and filtering	
Cmpd. ID	Exp $K_i$ (nM)	Cmpd. ID	Exp $K_i$ (nM)	Cmpd. ID	Exp $K_i$ (nM)	Cmpd. ID	Exp $K_i$ (nM)
3	12	3	12	3	12	22	77
37	25	11	279	14	217	29	304
31	27	9	515	16	505	28	358
33	30	7	952	13	1010	26	411
35	77	6	1800	20	1020	23	671
34	91	4	3020	17	2790	13	1010
30	123	8 (cis)	3500	15	3860	20	1020
38	167	5	>5100	18	>5100	24	5100
36	1430	8 (trans)	>5100	19	>5100	25	>5100
32	1750	10	>5100	21	>5100	27	>5100

8/10 tighter binding  
than the reference

1/10 tighter binding  
than the reference

1/10 tighter binding  
than the reference

1/10 tighter binding  
than the reference

# AI 기반 신약 후보물질 탐색



- Drug Discovery: Time & Cost & Accuracy
- Structure-Based Drug Discovery
- Protein-Ligand Interactions

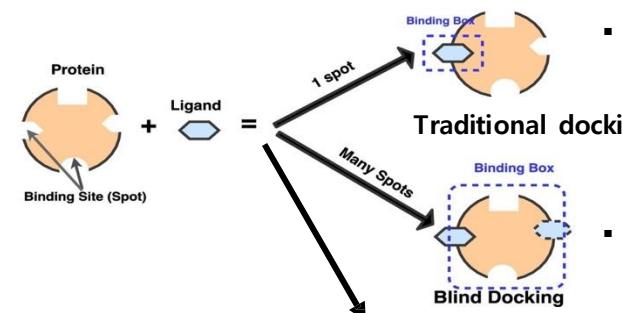
	Time	Cost	Accuracy	Application	Stage
Docking	😊	😊	😐 😕	Binding Pose Virtual Screening	Target to Hit Hit to Lead
End-Point (MMPBSA/GBSP)	😐	😐	😊 😐	Binding Affinity Scaffold hopping	Target to Hit Hit to Lead
Alchemical Transformation (FEP)	😕	😕	😊	Binding Affinity Mechanistic study	Target to Hit Hit to Lead Lead optimization

## ▪ Blind Docking (When the target site is not known~)

- refer to docking a ligand to the whole surface of a protein without any prior knowledge of the target pocket.

□ AlphaFold(google), RoseTTAFold, OpenFold, ESMFold(meta)에서 Protein Folding 구조 예측

□ Binding site을 알지 못하는 상태에서 단백질과 리간드의 결합에 대한 연구 가속화



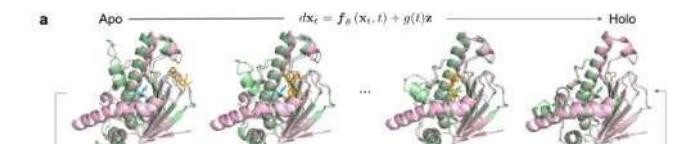
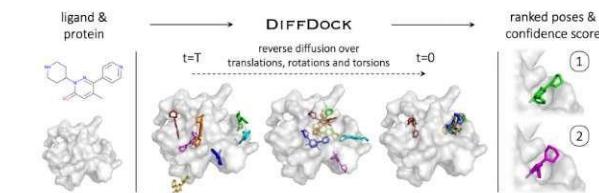
DOI:10.1007/s11227-019-02834-5

- Binding site 지정 O  
: Autodock, Glide, LeDock etc.

- Binding site 지정 X  
: DiffDock, TANKBind, E3Bind, etc.

- Binding site 지정 X, Protein flexibility w/DL  
: DynamicBind, etc.  
: GNINA(binding site 지정필요)

### Flexible + Blind Docking



## &lt; Blind Docking &gt;

- DiffDock (<https://arxiv.org/abs/2210.01776>)

- Input: Protein  PDB file  
Ligand  SMILES

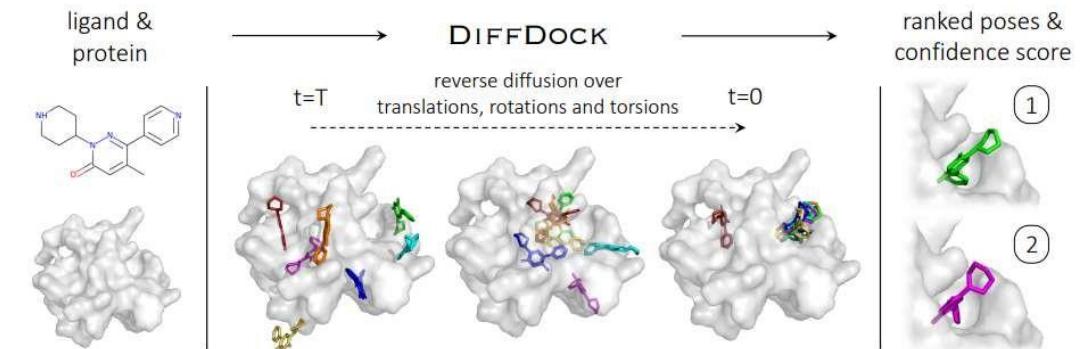
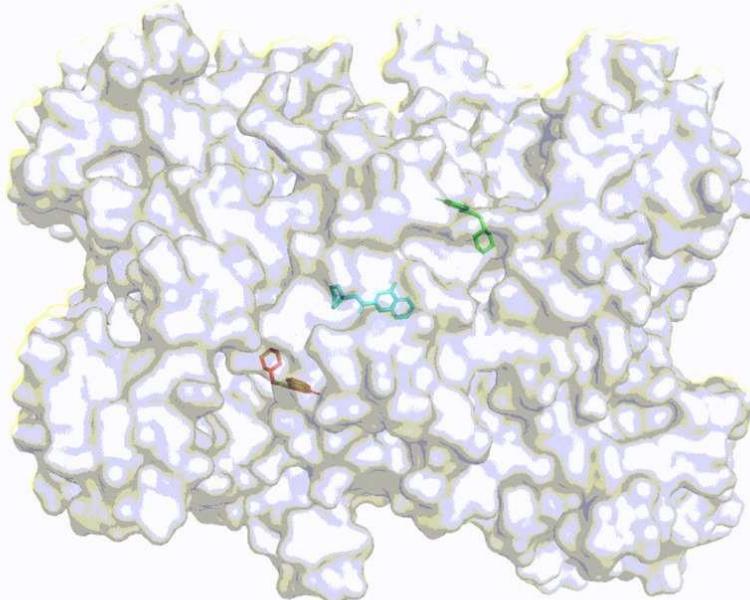


Figure 1: Overview of DIFFDOCK. *Left*: The model takes as input the separate ligand and protein structures. *Center*: Randomly sampled initial poses are denoised via a reverse diffusion over translational, rotational, and torsional degrees of freedom. *Right*: The sampled poses are ranked by the confidence model to produce a final prediction and confidence score.

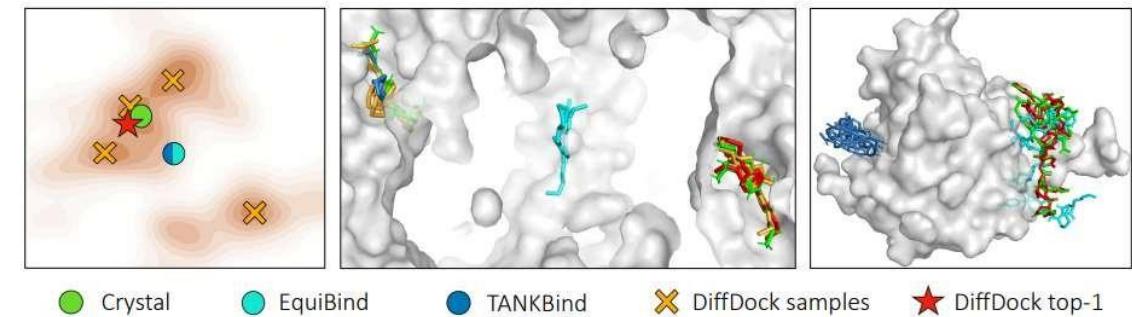
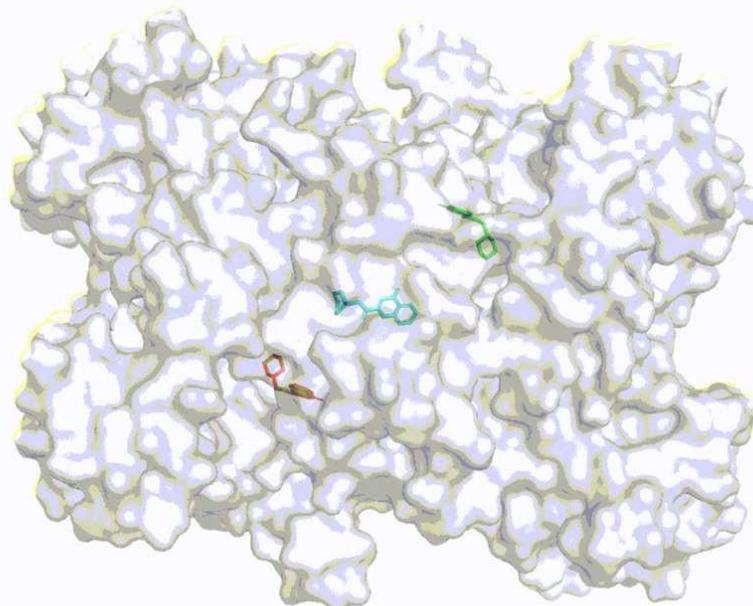


Figure 2: “DIFFDOCK top-1” refers to the sample with the highest confidence. “DIFFDOCK sam-

## &lt; Blind Docking &gt;

- DiffDock (<https://arxiv.org/abs/2210.01776>)

- Input: Protein  PDB file  
 Ligand  SMILES



HuggingFace 를 통해 test 가능

<https://huggingface.co/spaces/reginabarzilaygroup/DiffDock-Web>

huggingface.co/spaces/reginabarzilaygroup/DiffDock-Web

Spaces reginabarzilaygroup DiffDock-Web like 26 Running on T4

DiffDock Web

Run DiffDock for a single protein and ligand. We have provided the most important inputs as UI elements.

**Input**

**Protein**

Input PDB ID  
PDB Code or upload file below

Input PDB File  
파일을 끌어 놓으세요  
- 또는 -  
클릭해서 업로드하기

**Ligand**

SMILES string  
Provide SMILES input or upload mol2/sdf file below

Input Ligand  
파일을 끌어 놓으세요  
- 또는 -  
클릭해서 업로드하기

Samples Per Complex  
10

**Configuration (Optional)**

Configuration file to be passed to `inference.py`. If this is provided, it must supply all necessary arguments. If not provided, the `default configuration` will be used.

- < Blind Docking & Protein Flexibility>
- DynamicBind (Nat. Comm. 2024, 15, 1071)
  - Input: Protein  PDB file  
Ligand  SMILES

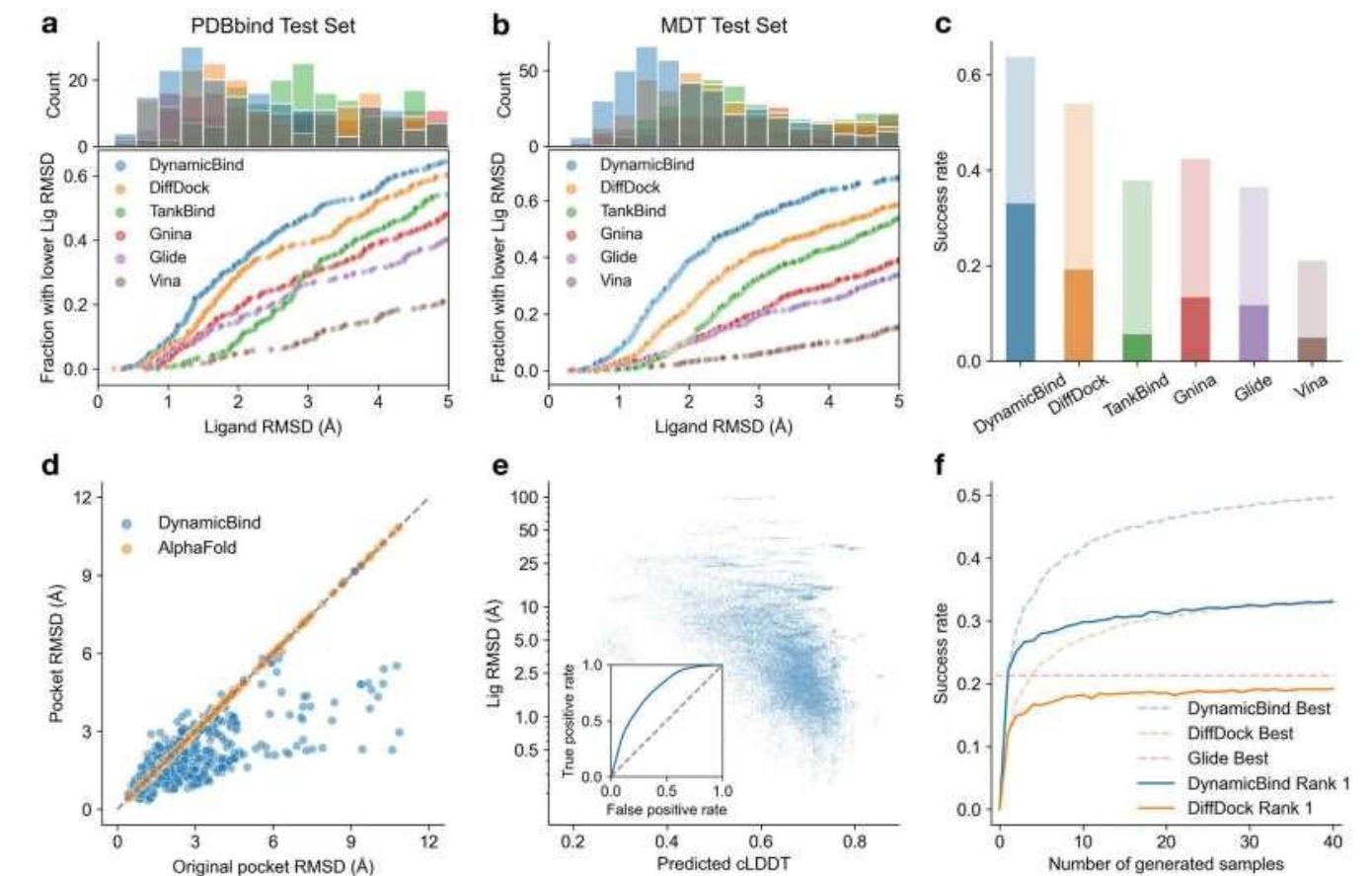
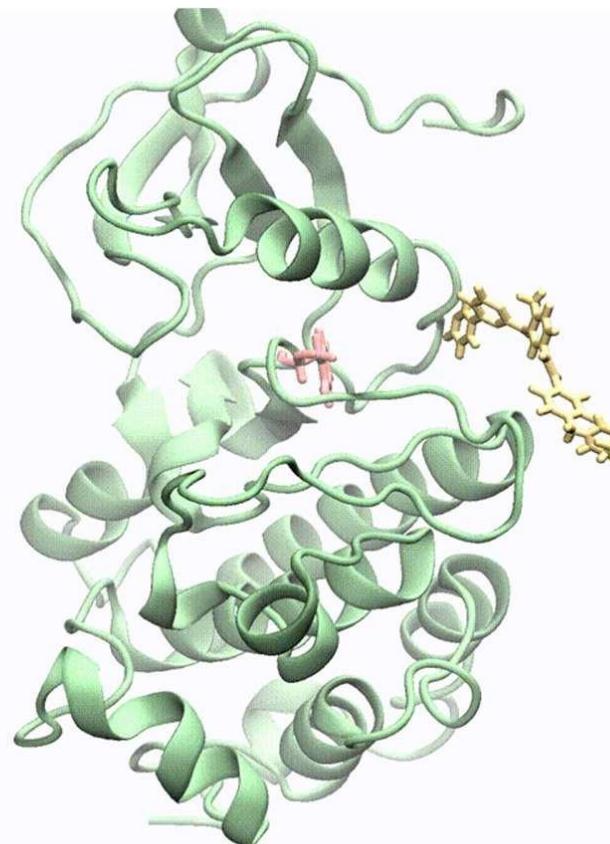
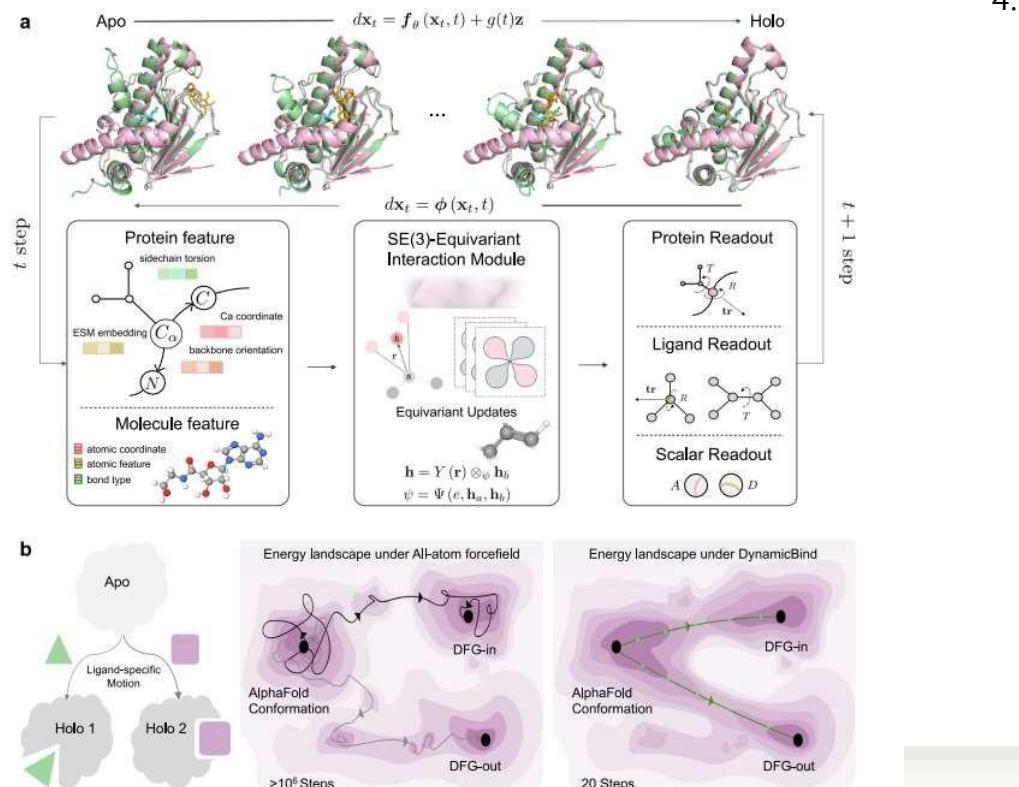


Fig. 2 | Benchmark results overview. a, b DynamicBind outperforms other

(cLDCT) score predicted by DynamicBind correlates well with the ligand RMSD and

## < Blind Docking & Protein Flexibility>

- DynamicBind (Nat. Comm. 2024, 15, 1071)
- Input: Protein  PDB file  
Ligand  SMILES

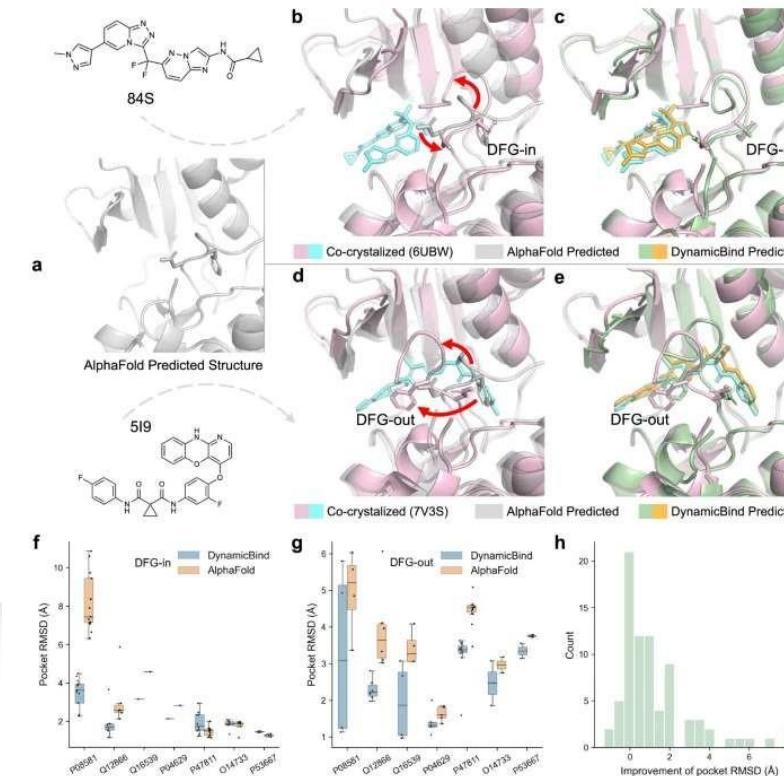


**Fig. 1 | Overview of DynamicBind model.** **a** The holo state is represented in pink, the initial apo and the model-predicted conformation in green. The native ligand is depicted in cyan, and the predicted ligand shown in orange. The model accepts as input both the features and the current conformation of the protein and ligand. The output readouts include the predicted updates: global translation and rotation for both the ligand and each protein residue, the rotation of torsional angles for the ligands and chi angles for the protein residues, and two prediction modules (binding affinity, A and confidence score, D). During the training phase, the model is

designed to learn the transformation from the apo-like conformation into the holo conformation. During inference, the model iteratively updates the initial input structure twenty times. **b** A schematic figure shows that our model could predict the two different holo conformations when the protein binds with two different ligands. Our model could predict the bounded protein conformation within 20 steps, while millions of steps of all-atom MD simulations are needed to find the same bounded state.

1. input: PDB, SMILES
2. **Randomly places the ligand** whose seed conformation is generated using RDKit, around the protein
3. the model gradually translates and rotates the ligand while adjusting its internal torsional angles
4. After the initial five steps where only the ligand conformation is changed, the model then simultaneously translates and rotates the protein residues, while modifying the side-chain chi angles, in the remaining steps.

**Fig. 3: DynamicBind captures ligand-specific protein conformational changes.**



## &lt; PoseBench&gt;

- <https://arxiv.org/pdf/2405.14108>
- <https://github.com/BioinfoMachineLearning/PoseBench>

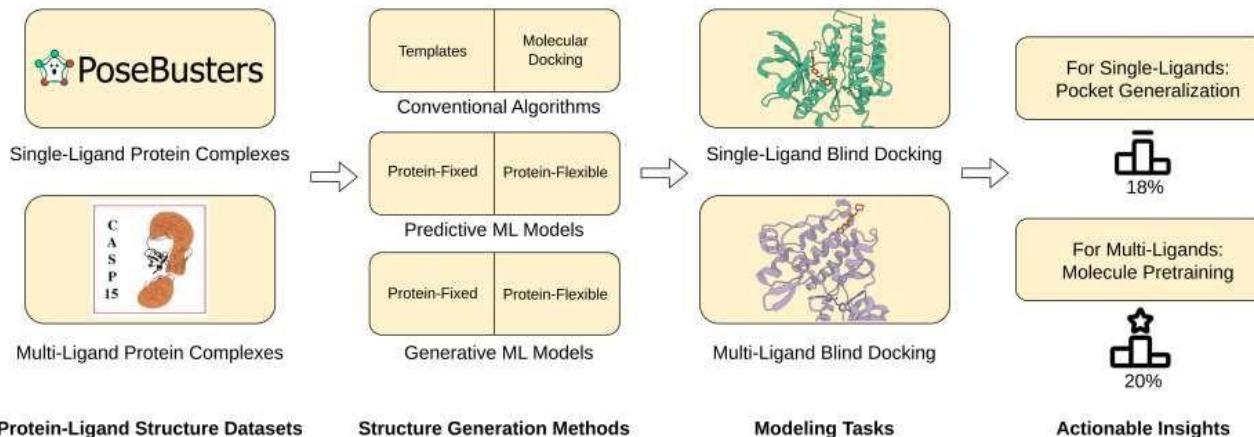


Figure 1: Overview of POSEBENCH, our comprehensive benchmark for *practical* ML modeling of single and multi-ligand protein complex structures in the context of apo (predicted) protein structures without known binding pockets (i.e., blind docking).

- **FABind**
- **RoseTTAFold-All-Atom**
- **DiffDock-L**
- **DynamicBind**
- **NeuralPLexer**

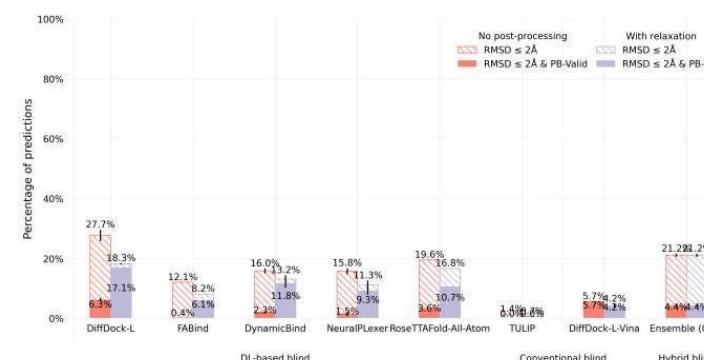
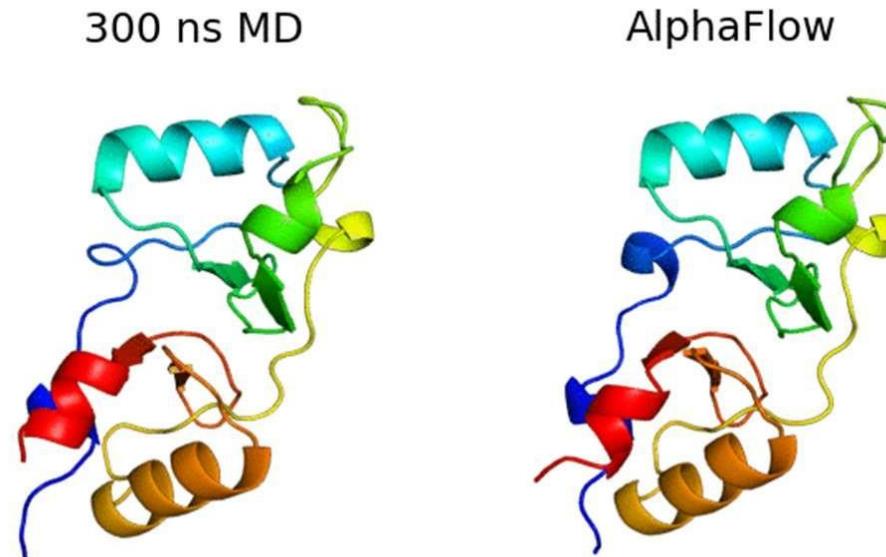


Figure 3: PoseBusters dataset results for successful single-ligand docking with relaxation.



- ✓ MD를 대체하면서 시간을 줄일 수 있는 방법연구
  - Multiple conformation
- ✓ Flow & Diffusion

When further trained on ensembles from all-atom MD, our method accurately captures conformational flexibility, positional distributions, and higher order ensemble observables for unseen proteins. Moreover, our method can diversify a static PDB structure with faster wall-clock convergence to certain equilibrium properties than replicate MD trajectories, demonstrating its potential as a proxy for expensive physics-based simulations.

<https://github.com/bjing2016/alphaflow>.

## • 구조기반 가상 스크리닝

: 질환을 일으키는 특정 타겟 단백질의 binding site에 결합하여 활성을 보이는 화합물을 찾아 내는 것

- HTS vs. Virtual Screening
- Target-based vs. Phenotype-based
- Chemical Library
- Protein Structure: 단백질 구조 규명 & 데이터베이스
- Binding Site/Pocket: 단백질 flexibility & structure selection
- Protein-ligand Interaction/Binding affinity
- Case Study

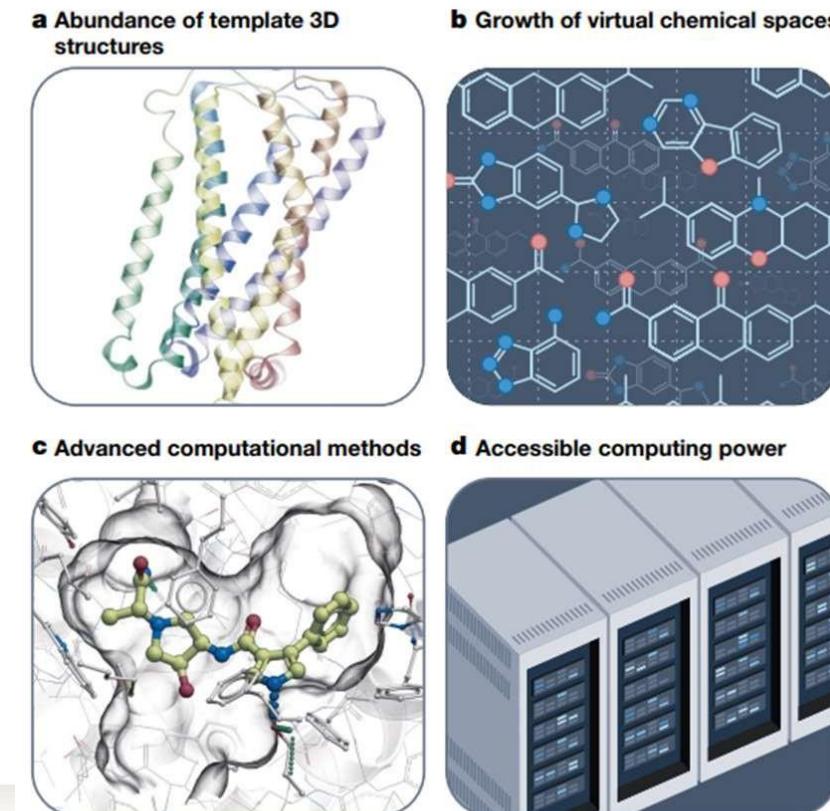
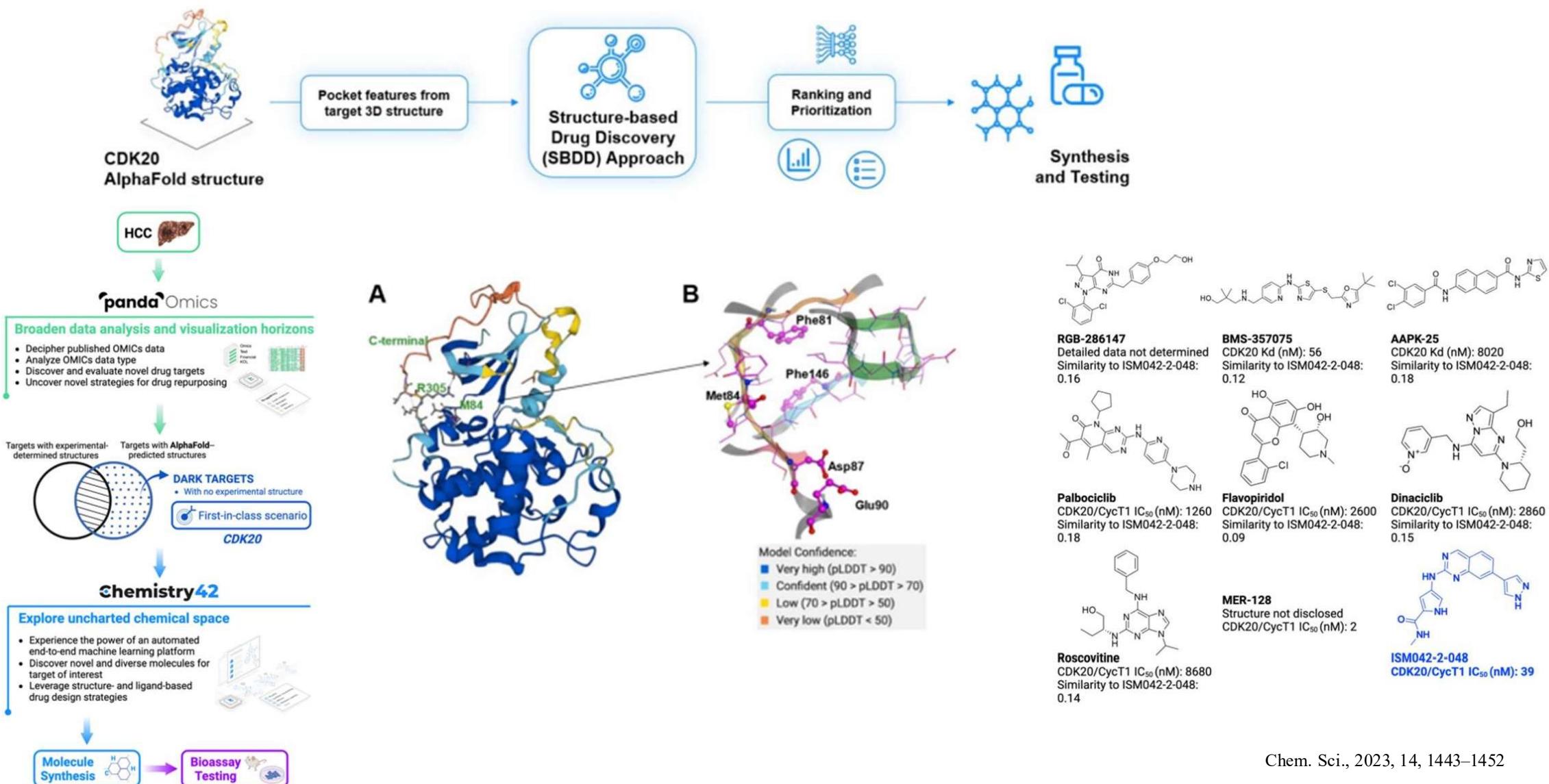
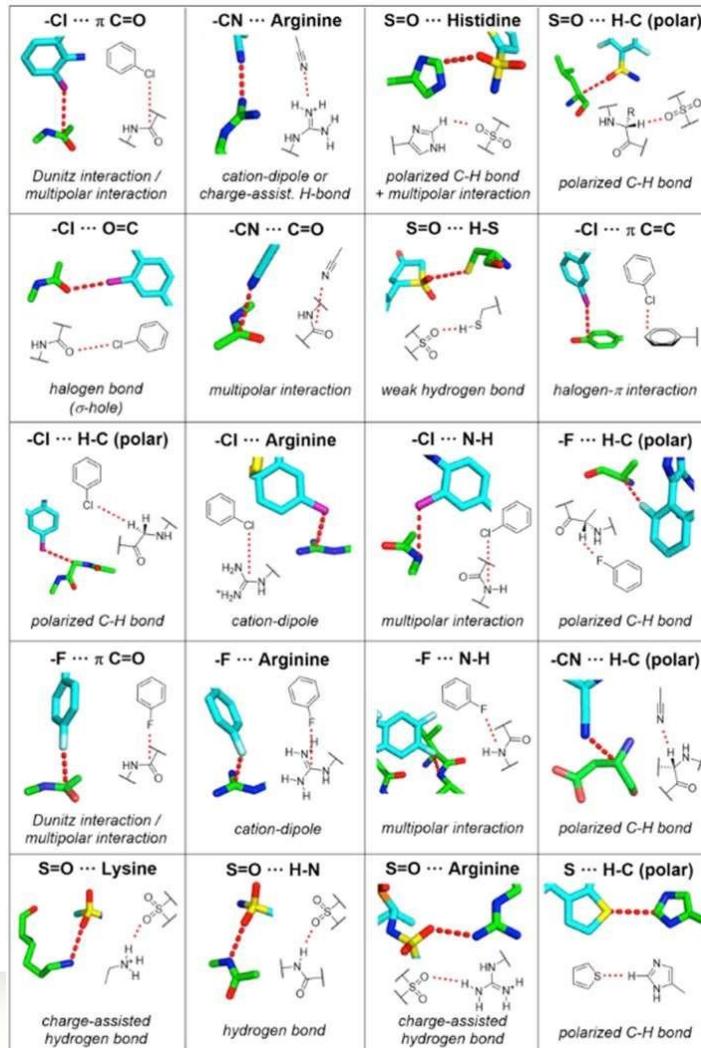


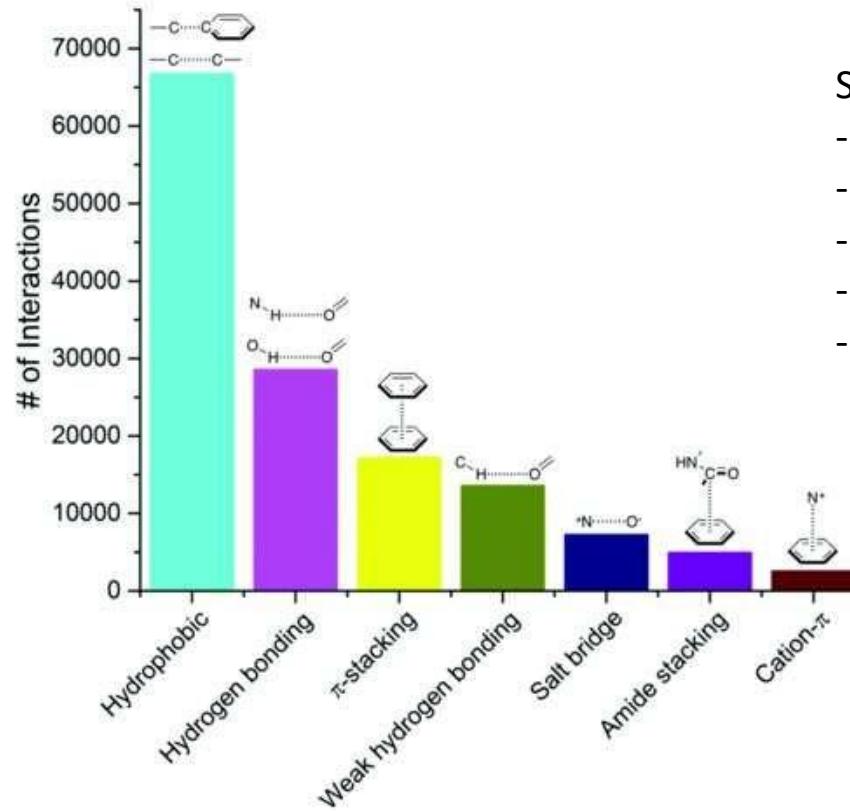
Fig. 1 | Key factors driving VLS technology breakthroughs for generation of high-quality hits and leads. **a**, More than 200,000 protein structures in the

Nature, 2023, 616, 673





### < Frequency of non covalent-bonding >

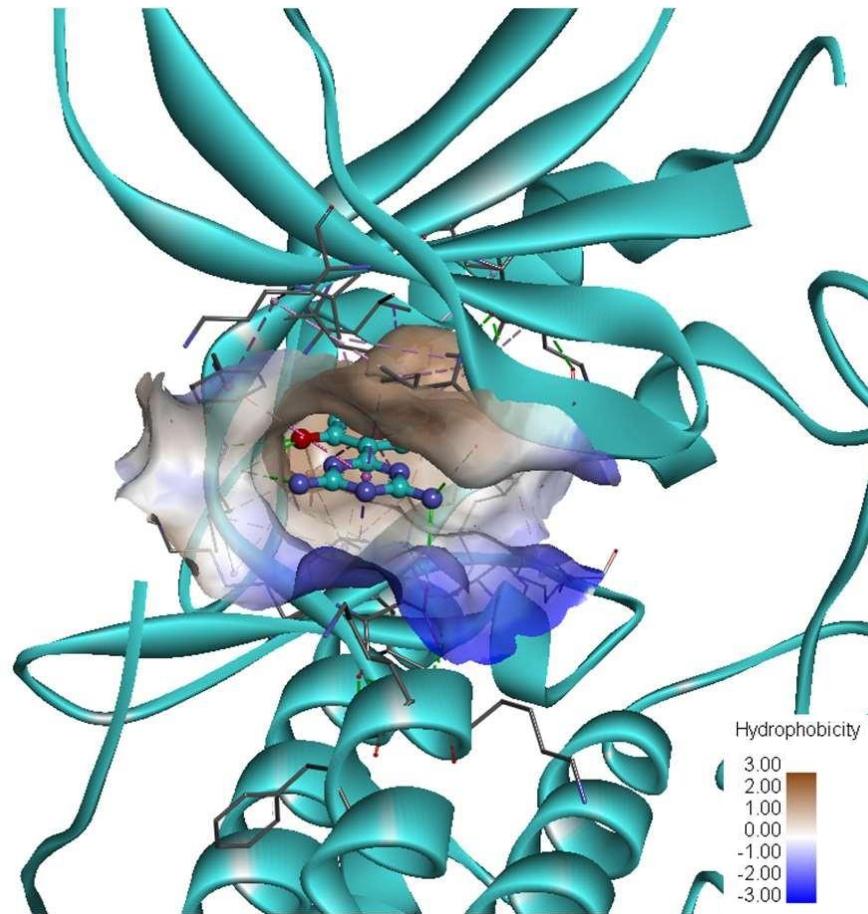


#### Strength >>

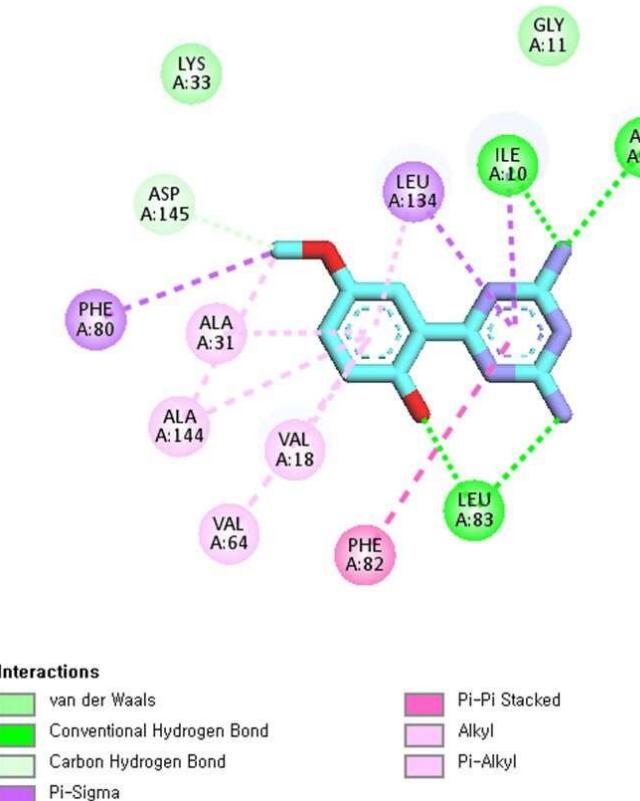
- covalent single bond: 80-100 Kcal/mol
- Salt Bridge ~2 kcal/mol
- H-bond ~ 1 kcal/mol
- Hydrophobic ~ 0.7 kcal/mol
- Aromatic 1 ~ 3 kcal/mol

The most common interactions being hydrophobic, hydrogen bonds and pi-stacking.

## &lt; Interaction Diagram &gt;



3PXY.pdb(1.8 Å)



# The latest on Isomophic Labs' plans to use AI models to revamp drug discovery

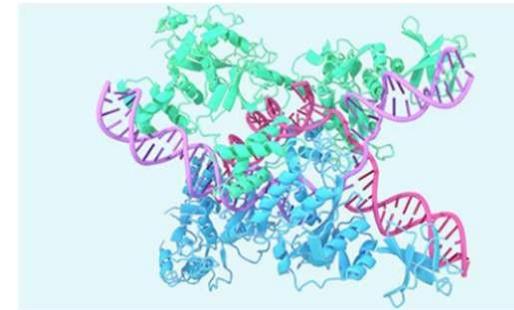
DeepMind & Isomophic Lab's

By Brian Buntz | August 30, 2024

- **Target identification:** AI models can be used to analyze large amounts of biological data to identify the specific proteins involved in a disease process. That translates into more precise targets for drug development.
- **Drug-target interaction prediction:** By understanding how potential drugs bind to and interact with target proteins at a molecular level, AI can help predict the efficacy and specificity of drug candidates. That can curb the reliance on costly and time-consuming experimental screening.
- **Drug efficacy and safety prediction:** In addition, AI models learn to predict the potential effectiveness and side effects of drug candidates based on their chemical structure and predicted interactions within the body. That can potentially streamline the drug development pipeline, which is one of the core promises of AI in the space.

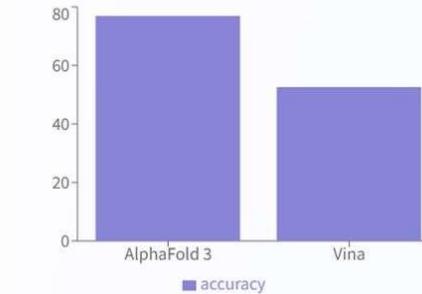
*Nature* volume 630, pages493–500 (2024)

<https://www.drugdiscoverytrends.com/the-latest-on-isomophic-labs-plans-to-use-ai-models-to-revamp-drug-discovery/>



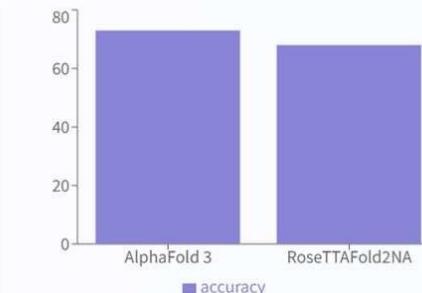
## Protein-Ligand Interaction Accuracy

Authors of the May 2024 *Nature* paper (link below) reported that AlphaFold 3 significantly outperforms Vina (a popular molecular docking program) in predicting protein-ligand interactions:



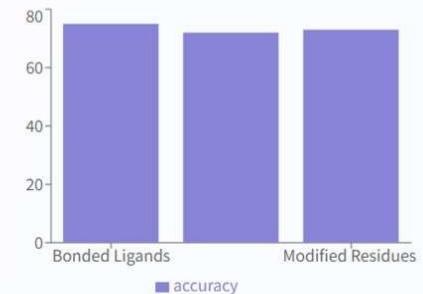
## Protein-Nucleic Acid Interaction Accuracy

Comparison of AlphaFold 3 and RoseTTAFold2NA (a machine learning method for predicting protein-nucleic acid complexes) in predicting protein-nucleic acid interactions:



## Covalent Modifications Accuracy

AlphaFold 3's accuracy in predicting various covalent modifications:

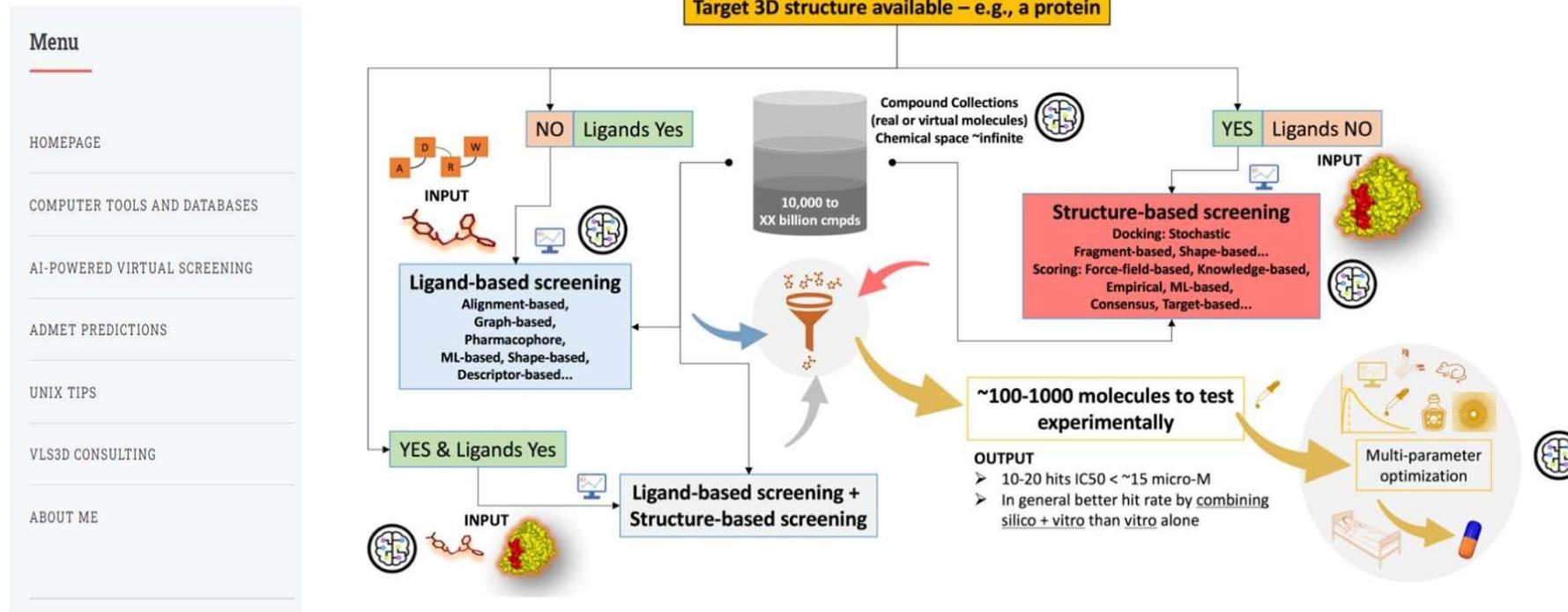


## Antibody-Antigen Prediction Improvement

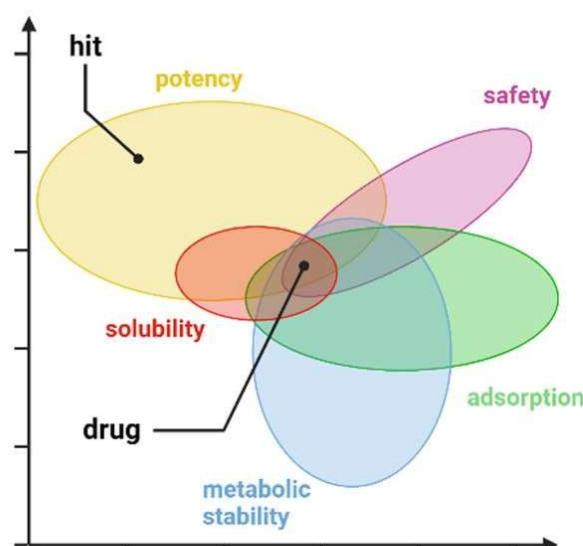
AlphaFold 3 achieved approximately 69% accuracy in antibody-antigen interface prediction ( $\text{DockQ} > 0.23$ ), compared to 55% for AlphaFold-Multimer v2.3, representing a significant improvement in this critical area of biomolecular interaction prediction.

Category	Program	Comment	Web
3D Structure Viewer	Discovery Studio Viewer (BIOVIA)	무료	<a href="https://discover.3ds.com/discovery-studio-visualizer-download">https://discover.3ds.com/discovery-studio-visualizer-download</a>
molecular visualization system	PyMOL	무료/유료	<a href="https://www.pymol.org/">https://www.pymol.org/</a>
Data Visualization and Analysis with Chemical Intelligence	DataWarrior	무료	<a href="https://openmolecules.org/datawarrior/">https://openmolecules.org/datawarrior/</a>
Web-based Drug discovery tools	VLS3D.com	무료	<a href="https://www.vls3d.com/ai-powered-VS.html">https://www.vls3d.com/ai-powered-VS.html</a>

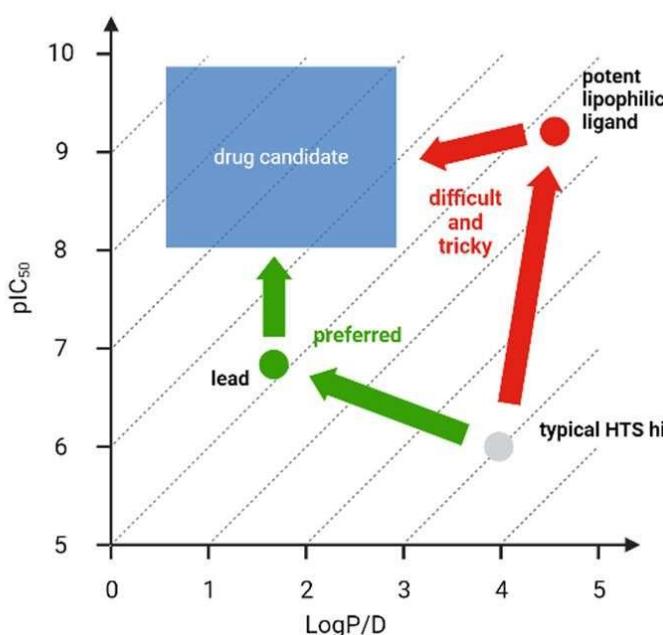
## VLS3D.COM directory of tools & databases collected over 20 years



<https://www.vls3d.com/ai-powered-VS.html>

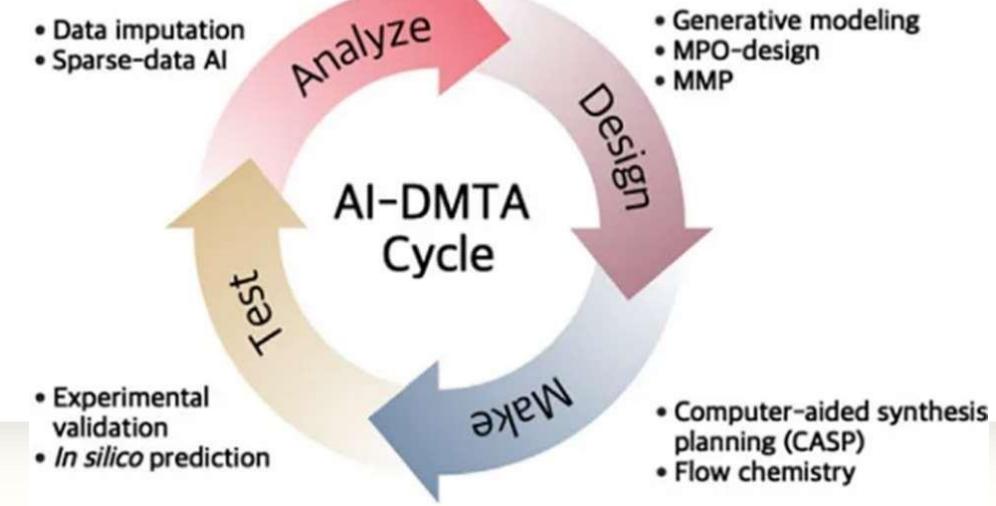


<https://www.efmc.info/hit-to-lead>



- Data imputation
- Sparse-data AI

- Experimental validation
- *In silico* prediction



<https://medium.com/@oleksiigavr/enhancing-preclinical-drug-discovery-with-artificial-intelligence-article-review-67385879284c>