KSBi-BIML 2026

Bioinformatics & Machine Learning(BIML) Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

단백질 구조 예측
– 주형 기반 모델링부터 AlphaFold까지

신웅희 _ 고려대학교

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2026 워크샵을 목적으로
제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우
발생하는 **모든 법적 책임은 행위자 본인에게 있음**을 알립니다.

KSBi-BIML 2026

Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics &Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의가 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

강의개요

단백질 구조 예측
– 주형 기반 모델링부터 AlphaFold까지

생체 분자의 기능은 그 구조와 밀접한 관련을 가진다. 단백질의 기능 또한 그 구조로부터 밝힐 수 있다. 따라서 X-선 결정학, 극저온 전자 현미경등 다양한 실험 기법들이 단백질의 구조를 밝히기 위해 널리 사용되고 있다. 그러나 실험을 통해 모든 구조를 결정하기에는 어려움이 있고, 아미노산 서열 규명에 비하여 높은 난이도를 가지므로, 이에 따라 sequence-structure gap이 발생하고 있다.

본 강의에서는 아미노산 서열로부터 단백질의 3차원 구조를 예측하는 방법에 대하여 학습한다. 2010년대 중반부까지 널리 사용되었던 주형 기반 모델링의 원리와 웹서버를 이용한 작동법을 익힌다. 또한 최근의 동향인 심층학습기반 단백질 구조 예측에 대하여 학습하고, Colab을 이용한 간단한 예측 방법에 대하여 실습한다.

강의는 다음의 내용을 포함한다:

- 단백질 구조 예측 이론
- Swiss-Model등 주형 기반 모델링
- AlphaFold등의 심층학습 기반 단백질 구조 예측

* 교육생준비물:
  노트북

* 강의 난이도: 초급

* 강의: 신웅희교수 (고려대학교 의료정보학과)

Curriculum Vitae

Speaker Name: Woong-Hee Shin, Ph.D.

▶ **Personal Info**

Name	Woong-Hee Shin
Title	Associate Professor
Affiliation	Korea University

▶ **Contact Information**

Address	161, Jeungneung-Ro, Seongbuk-Gu, Seoul, 02708
Email	whshin@korea.ac.kr

Research Interest

Computer-aided drug discovery, Protein design

Educational Experience

2008	B.S. in Chemistry, Seoul National University, Korea
2014	Ph.D. in Chemistry, Seoul National University, Korea

Professional Experience

2014-2019	Post-doc research fellow, Purdue University, USA
2019-2023	Assistant Professor, Department of Chemistry Education, Sunchon National University
2023-	Associate Professor, Department of Biomedical Informatics, Korea University College of Medicine

Selected Publications (3 maximum)

1. Wonkyeong Jang and Woong-Hee Shin, CoBRA: Compound Binding Site Prediction using RNA Language Model, Briefings in Bioinformatics, In Press.
2. Jinung Song, Junsu Ha, Juyong Lee, Junsu Ko, and Woong-Hee Shin, Improving docking and virtual screening performance using AlphaFold2 multi-state modeling for kinases, Scientific Reports, 2024, 14, 25167.
3. Yiyu Hong, Junsu Ha, Chae-Jo Lim, Kwang-Seok Oh, Ramakrishana Chandrasekaran, Junsu Ko, Woong-Hee Shin, and Juyong Lee, Accurate Prediction of Protein-Ligand Intereactions by Combining Physical Energy Function and Graph-Neural Networks, Journal of Cheminformatics, 2024, 16, 121.

KSBI | 한국생명정보학회

KSBi-BIML 2026

Protein Structure Prediction

BIML 2026

Woong-Hee Shin

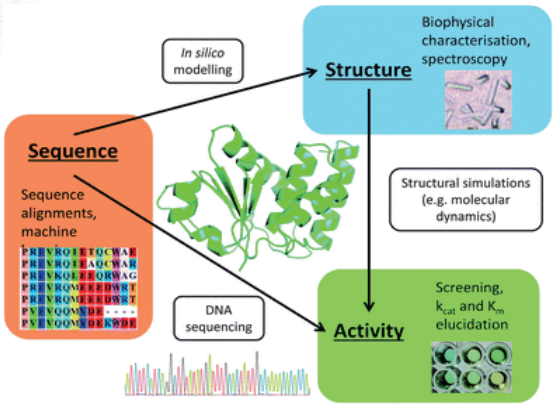
2

Lecture Goal

- Predict a protein 3D structure of your interest using web server and visualizing it

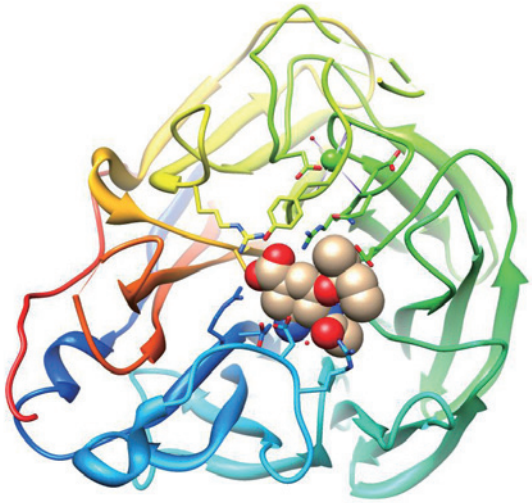
Need for Protein Structure Prediction

- Protein Structure-Function Relationship
- Similar structure, similar function



Need for Protein Structure Prediction

- Neuraminidase and Tamiflu

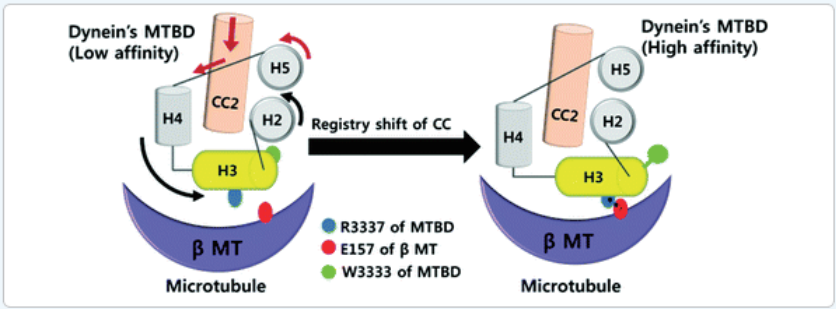


Need for Protein Structure Prediction

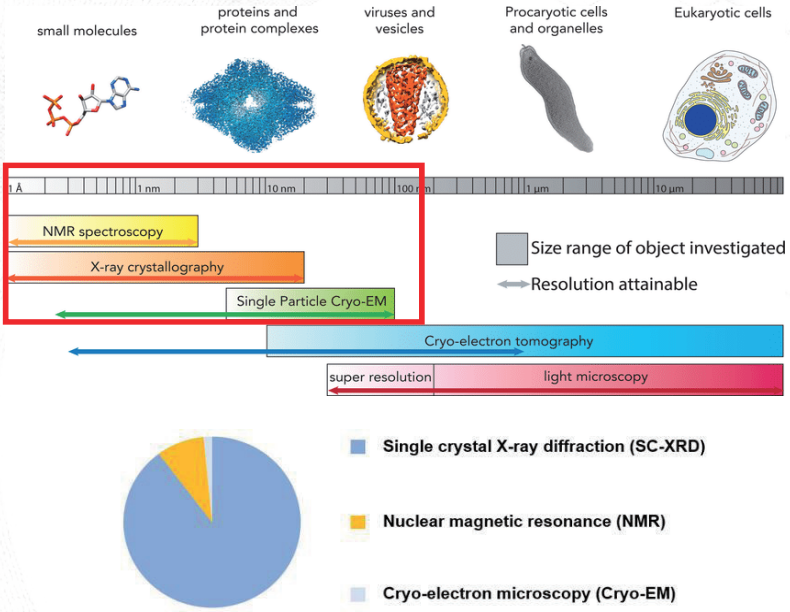
How Does a Registry Change in Dynein's Coiled-Coil Stalk Drive Binding of Dynein to Microtubules?

Junghyun Choi, Hahnbeom Park, and Chaok Seok*
Department of Chemistry, Seoul National University,
Seoul 151-747, Republic of Korea

Biochemistry, 2011, *50* (35), pp 7629-7636
DOI: 10.1021/bi200834k
Publication Date (Web): August 2, 2011



Experimental Techniques

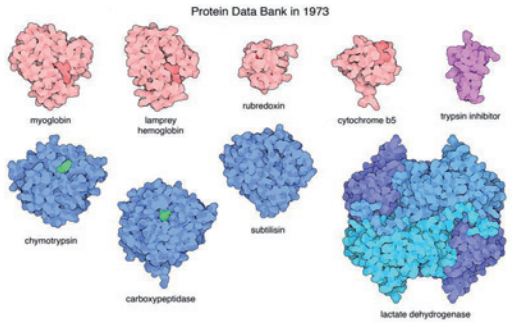


PDB (Protein Data Bank)

- Protein structure database
- rcsb.org

PDB

- Protein Data Bank
- Structural database of proteins
- Established in 1971 with 7 entries under the leadership of Walter Hamilton, Brookhaven National Laboratory
- In 1998, Research Collaborative for Structural Bioinformatics became responsible for managing the DB
- Worldwide PDB (wwPDB) was established to maintain the DB in 2003



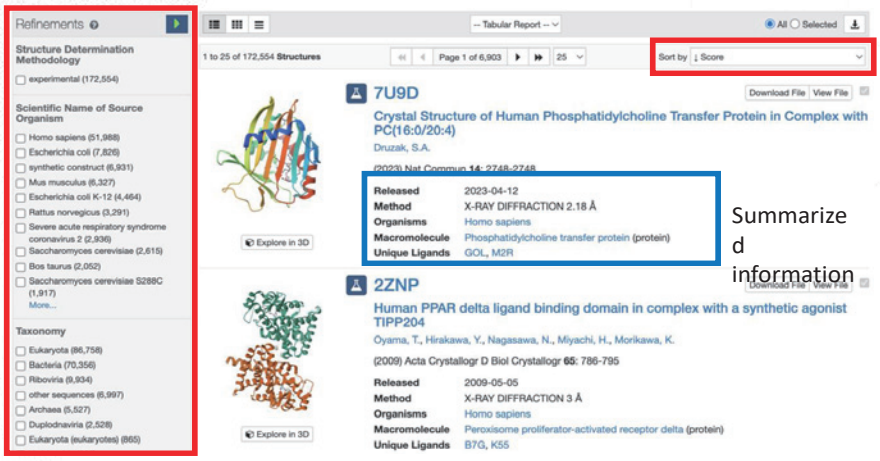
Searching a Specific Protein

- Type your protein name, PDB ID, or any keyword such as the name of experimentalist

Searching a Specific Protein

- Select a protein structure from the summarized information

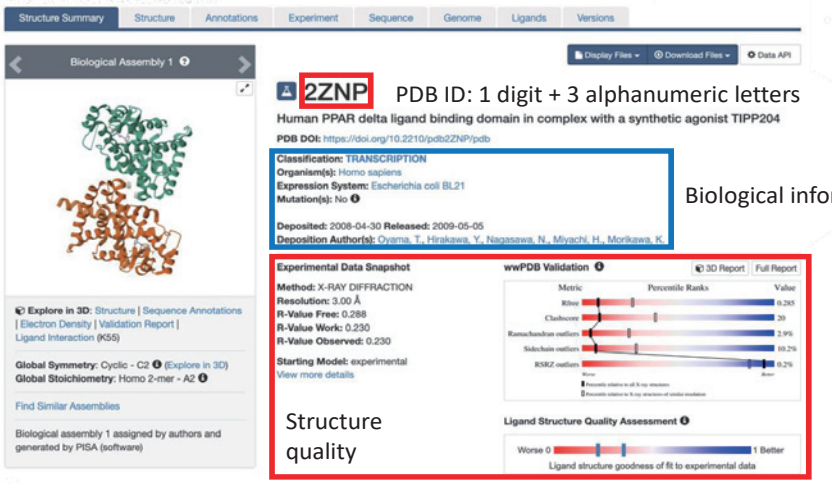
Refine structure by selecting conditions



Summarized information

Realign the results by changing sorting options such as resolution

Searching a Specific Protein



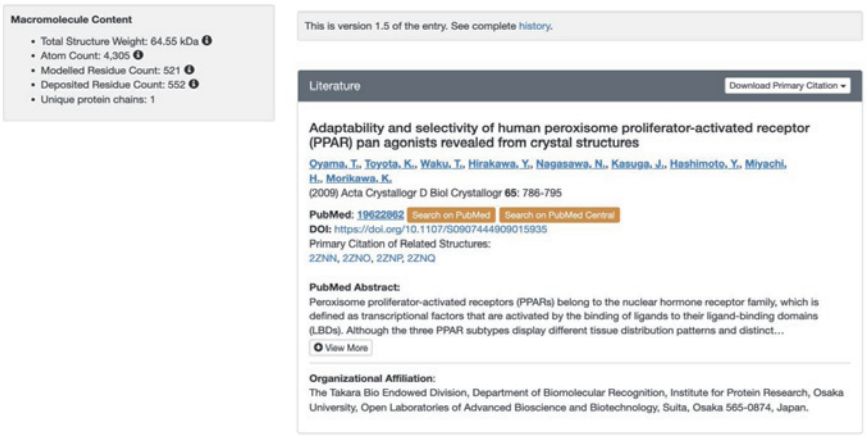
PDB ID: 1 digit + 3 alphanumeric letters

Biological information

Structure quality

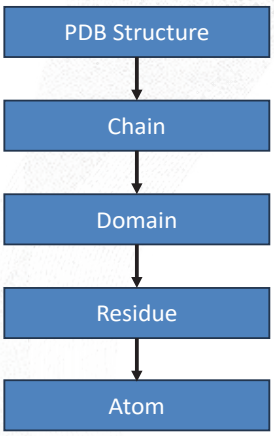
Searching a Specific Protein

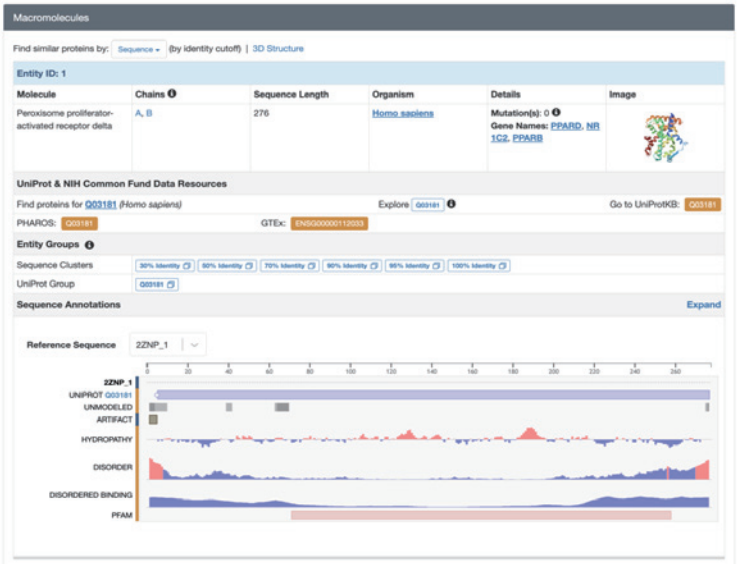
- Literature and molecular information



Searching a Specific Protein

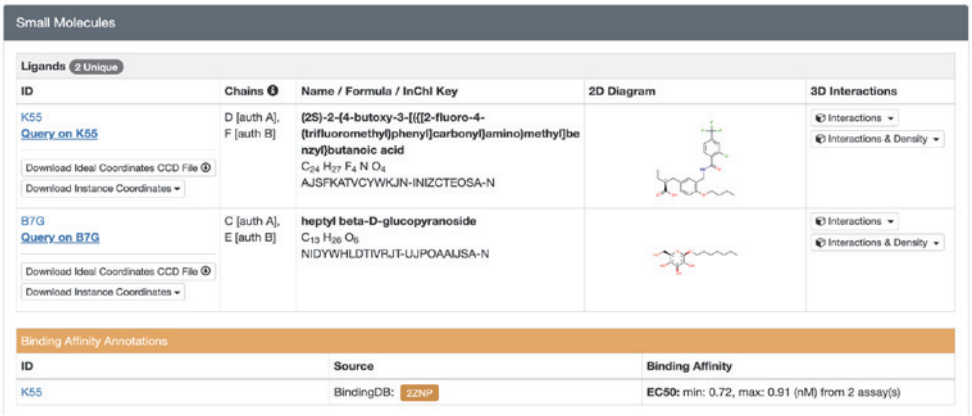
- Chain information





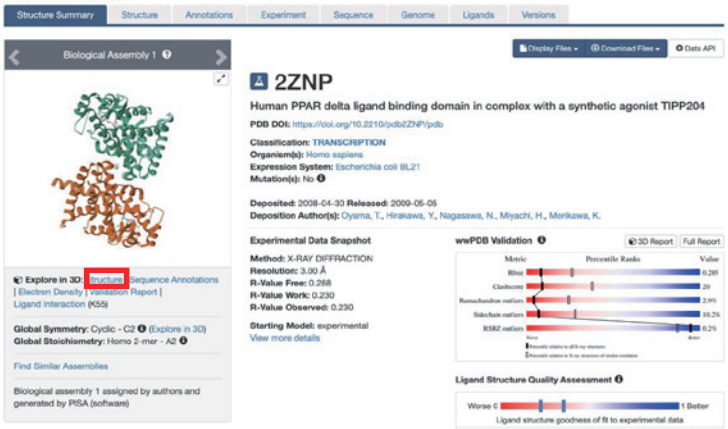
Searching a Specific Protein

- Hetero atom information
  - Noncanonical amino acid
  - Ligand
  - Ion
  - Etc.



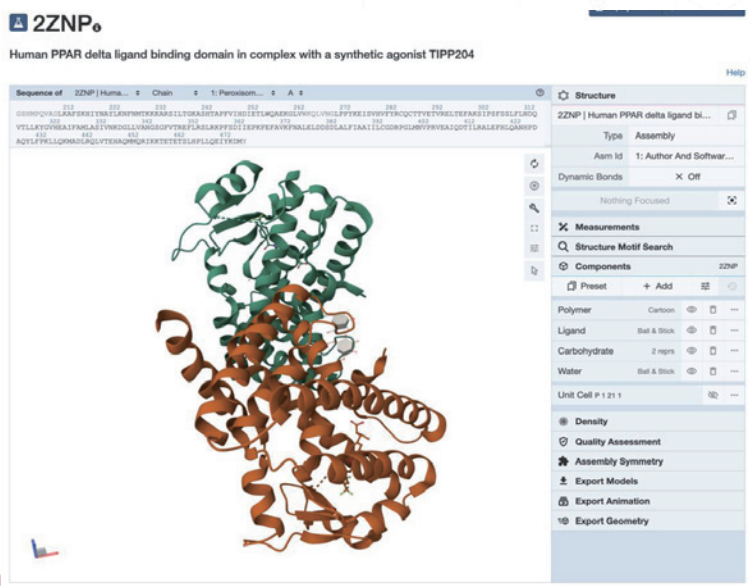
Visualizing a Protein Structure

- Click Explore in 3D: Structure



Visualizing a Protein Structure

- Holding left click: Rotating
- Holding right click: Translation
- Scrolling wheel: Zoom in / out



Visualizing a Protein Structure

Clicking specific residue focuses the viewer to the amino acid

Modify the representation of the viewer

PDB

- PDB file format
- HEADER: Protein name, PDB ID, Date deposited in PDB
- REMARK
  - Types of amino acids in the protein
  - Resolution
  - Organism
  - etc.

```
HEADER    TYROSINE-PROTEIN KINASE                 29-DEC-98   2SRC
TITLE     CRYSTAL STRUCTURE OF HUMAN TYROSINE-PROTEIN KINASE C-SRC,
TITLE    2 IN COMPLEX WITH AMP-PNP
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: TYROSINE-PROTEIN KINASE SRC;
COMPND    3 CHAIN: A;
COMPND    4 FRAGMENT: RESIDUES 86-536, CONTAINING SH2, SH3, KINASE 2
COMPND    5 DOMAINS AND C-TERMINAL TAIL;
COMPND    6 SYNONYM: C-SRC, P60-SRC;
COMPND    7 EC: 2.7.1.112;
COMPND    8 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE    3 ORGANISM_COMMON: HUMAN;
SOURCE    4 ORGANISM_TAXID: 9606;
SOURCE    5 EXPRESSION_SYSTEM: SPODOPTERA FRUGIPERDA;
SOURCE    6 EXPRESSION_SYSTEM_COMMON: FALL ARMYWORM;
SOURCE    7 EXPRESSION_SYSTEM_TAXID: 7108;
SOURCE    8 EXPRESSION_SYSTEM_VECTOR: BACULOVIRUS
KEYWDS    SRC, TYROSINE-PROTEIN KINASE, PHOSPHORYLATION, SH2, SH3,
KEYWDS   2 PHOSPHOTYROSINE, PROTO-ONCOGENE, PHOSPHOTRANSFERASE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    W.XU,A.DOSHI,M.LEI,M.J.ECK,S.C.HARRISON
REVDAT   3   24-FEB-09 2SRC    1       VERSN
REVDAT   2   01-APR-03 2SRC    1       JRNL
REVDAT   1   22-JUL-99 2SRC    0
JRNL        AUTH   W.XU,A.DOSHI,M.LEI,M.J.ECK,S.C.HARRISON
JRNL        TITL   CRYSTAL STRUCTURES OF C-SRC REVEAL FEATURES OF ITS
JRNL        TITL 2 AUTOINHIBITORY MECHANISM.
JRNL        REF    MOL.CELL                      V.   3   629 1999
JRNL        REFN                   ISSN 1097-2765
JRNL        PMID   10360179
JRNL        DOI    10.1016/S1097-2765(00)80056-1
REMARK   1
REMARK   2
REMARK   2 RESOLUTION.    1.50 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR
REMARK   3   AUTHORS     : BRUNGER
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 1.50
```

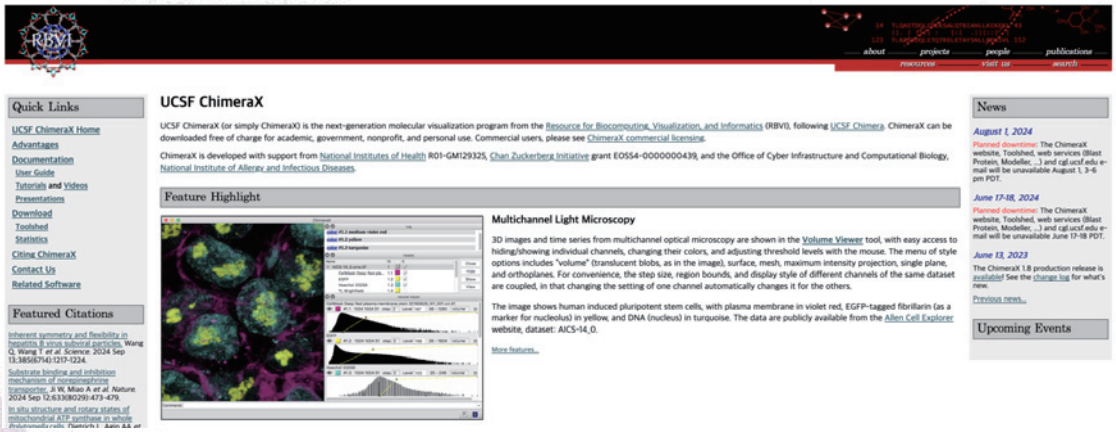
PDB

- ATOM(amino acid)/HETATM(non AA)
  - Atom information
  - Indices
  - Amino acid information
  - x,y,z coordinate

```
ATOM      1  N   THR A  84      33.954  62.808  55.198  1.00 38.76           N
ATOM      2  CA  THR A  84      32.735  62.188  54.583  1.00 39.76           C
ATOM      3  C   THR A  84      32.941  60.717  54.186  1.00 37.91           C
ATOM      4  O   THR A  84      32.021  60.068  53.678  1.00 36.13           O
ATOM      5  CB  THR A  84      32.285  62.973  53.319  1.00 41.34           C
ATOM      6  OG1 THR A  84      33.424  63.240  52.489  1.00 41.84           O
ATOM      7  CG2 THR A  84      31.616  64.288  53.711  1.00 40.28           C
ATOM      8  N   THR A  85      34.140  60.192  54.419  1.00 35.88           N
ATOM      9  CA  THR A  85      34.437  58.806  54.075  1.00 34.00           C
ATOM     10  C   THR A  85      34.001  57.804  55.154  1.00 33.76           C
ATOM     11  O   THR A  85      33.807  58.168  56.320  1.00 30.78           O
ATOM     12  CB  THR A  85      35.944  58.610  53.814  1.00 36.09           C
ATOM     13  OG1 THR A  85      36.671  58.783  55.036  1.00 39.72           O
ATOM     14  CG2 THR A  85      36.445  59.611  52.790  1.00 37.48           C
ATOM     15  N   PHE A  86      33.832  56.547  54.746  1.00 28.75           N
ATOM     16  CA  PHE A  86      33.449  55.480  55.665  1.00 25.99           C
ATOM     17  C   PHE A  86      34.625  54.522  55.755  1.00 24.12           C
ATOM     18  O   PHE A  86      35.606  54.655  55.013  1.00 23.08           O
ATOM     19  CB  PHE A  86      32.223  54.724  55.157  1.00 24.62           C
ATOM     20  CG  PHE A  86      30.915  55.310  55.598  1.00 23.17           C
ATOM     21  CD1 PHE A  86      30.463  56.515  55.066  1.00 26.21           C
ATOM     22  CD2 PHE A  86      30.108  54.636  56.508  1.00 25.23           C
ATOM     23  CE1 PHE A  86      29.218  57.039  55.432  1.00 24.64           C
ATOM     24  CE2 PHE A  86      28.858  55.150  56.885  1.00 26.97           C
ATOM     25  CZ  PHE A  86      28.414  56.352  56.342  1.00 28.54           C
ATOM     26  N   VAL A  87      34.533  53.566  56.672  1.00 23.62           N
ATOM     27  CA  VAL A  87      35.593  52.580  56.854  1.00 23.37           C
ATOM     28  C   VAL A  87      34.983  51.188  57.020  1.00 21.41           C
ATOM     29  O   VAL A  87      33.925  51.029  57.632  1.00 19.70           O
ATOM     30  CB  VAL A  87      36.471  52.926  58.087  1.00 25.01           C
ATOM     31  CG1 VAL A  87      35.672  52.736  59.377  1.00 25.81           C
```

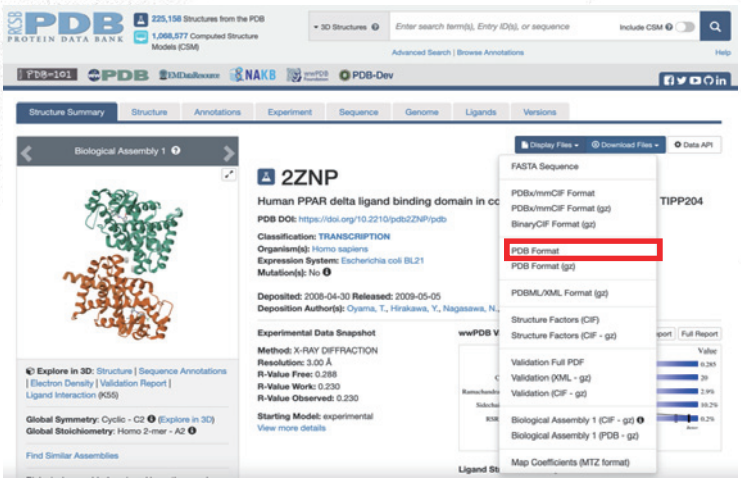
Visualizing a Protein Structure

- Use ChimeraX (https://www.cgl.ucsf.edu/chimerax/) or PyMOL (https://www.pymol.org/)



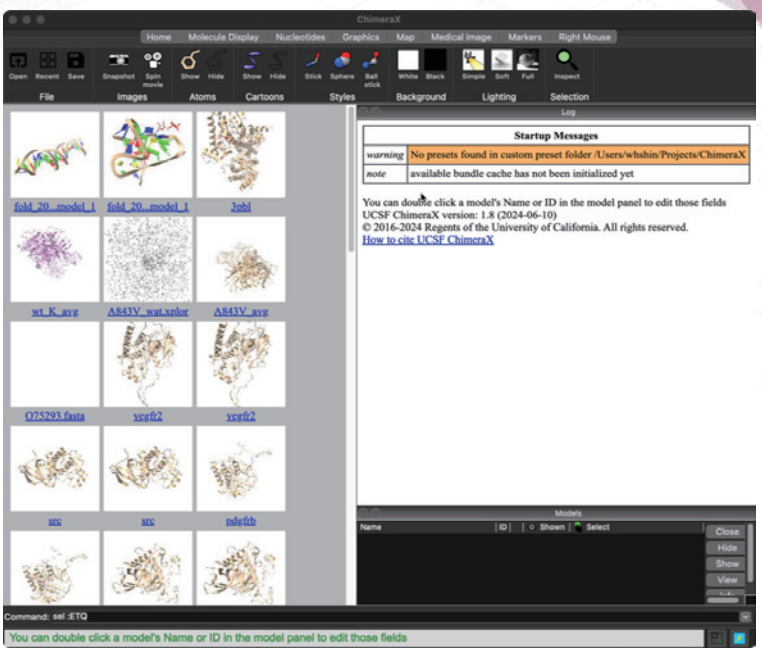
Visualizing a Protein Structure

- Download PDB file or take a note of the PDB ID



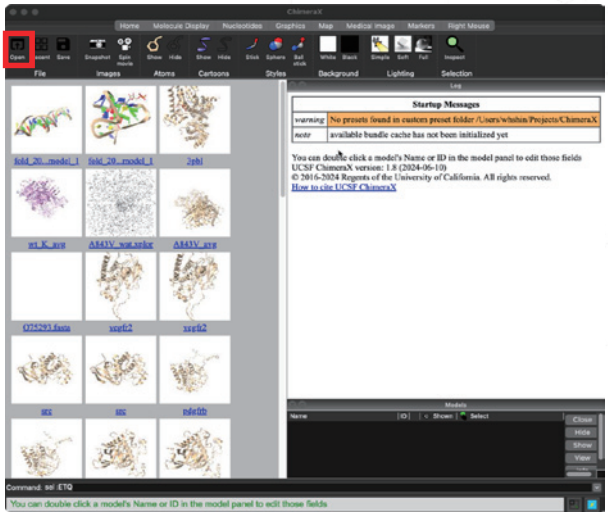
Visualizing a Protein Structure

- Execute ChimeraX



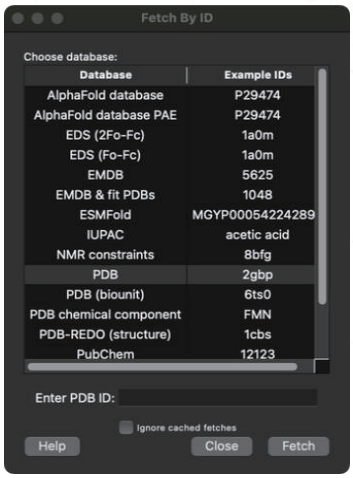
Visualizing a Protein Structure

- 1) If you have a PDB file
- Click Open and find the file



Visualizing a Protein Structure

- 2) If you know the PDB ID
- Click File > Fetch By ID, select PDB, and put your ID

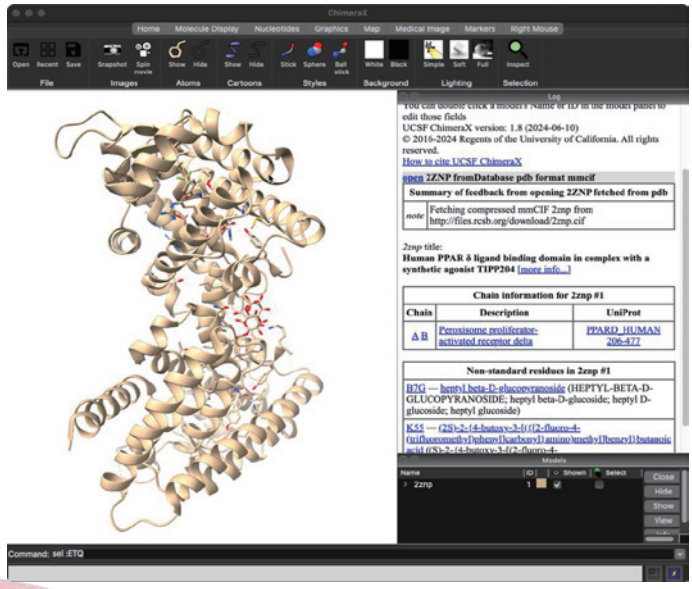


Visualizing a Protein Structure

Holding left click: Rotating

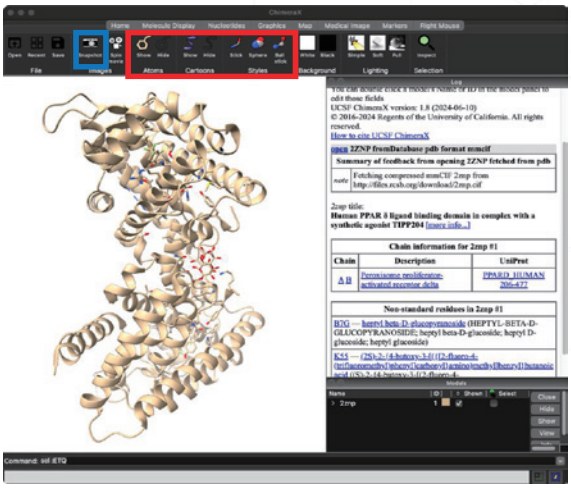
Holding right click: Translation

Scrolling wheel: Zoom in / out



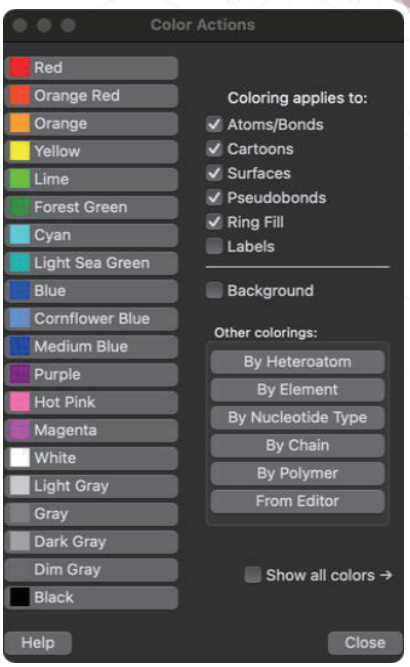
Visualizing a Protein Structure

- Change the representation style using a panel
- Capture the current screen by clicking the camera ion



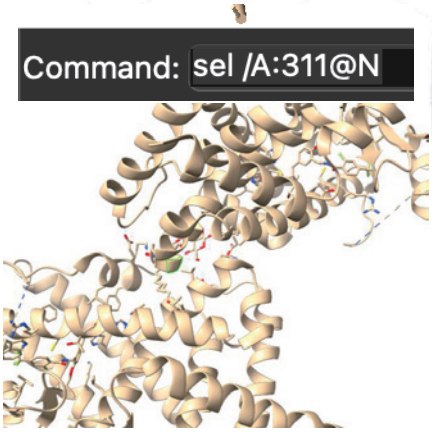
Visualizing a Protein Structure

- Actions > Color can change the color of the atom, residue, structure, and background

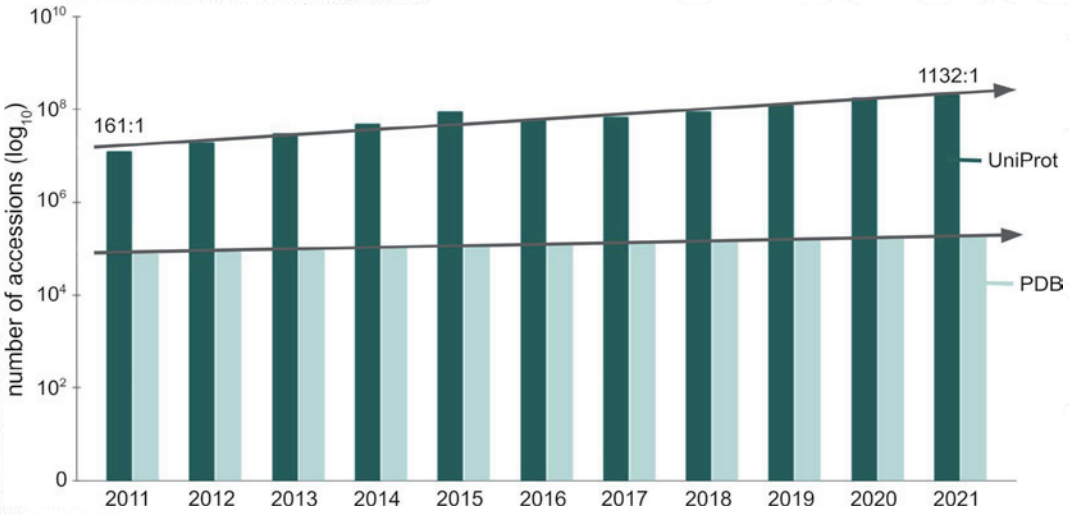


Visualizing a Protein Structure

- Select a specific residue by Ctrl + click
- Or if you know the residue number or name type sel :[residue] at the command line
- Identifiers
  - # model
  - / chain
  - : residue
  - @ atom
- Action > Focus zooms on the selected atoms

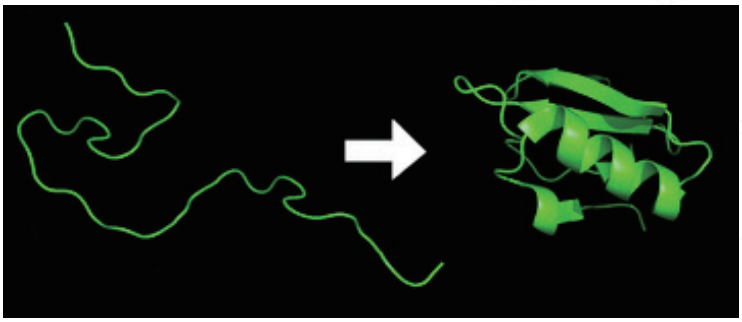


Need for Protein Structure Prediction



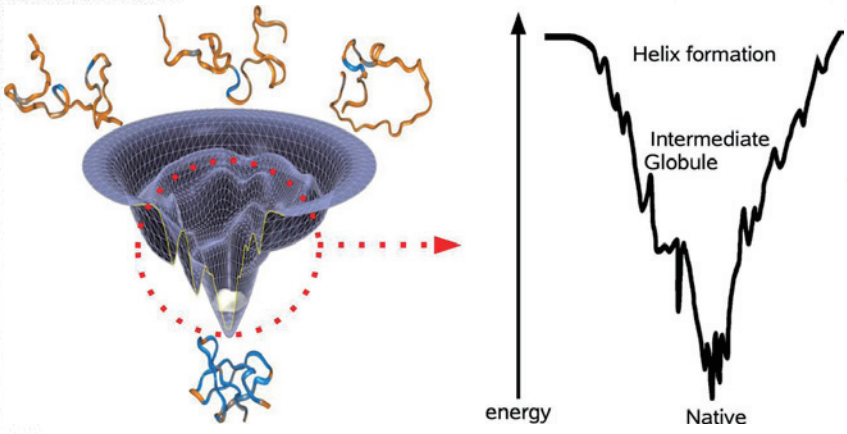
Protein Folding Problem

- Predict 3D structure of a protein from a amino acid sequence



Levinthal's Paradox

- Nature favors the lowest energy state

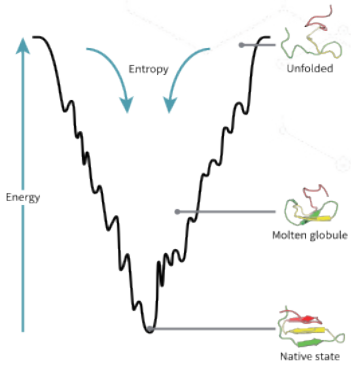


Levinthal's Paradox

- How to find the lowest energy conformation?
- If the protein is composed of 100 residues,
- Total number of cases if the three states allowed for a single residue backbone
- 3^{100}
- Assuming that the protein can search 10^{12} number of possible states in one second,
- $5X10^{35}$ seconds = $2X10^{28}$ years

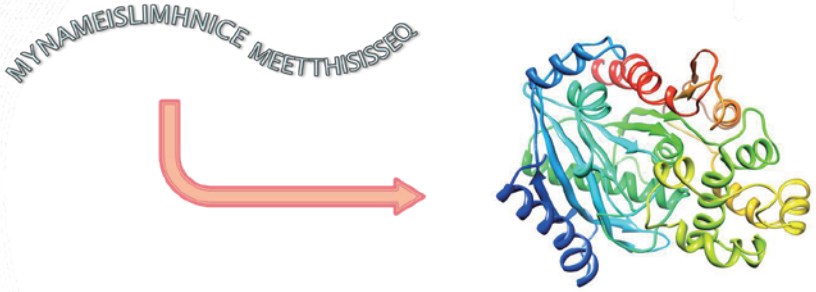
Levinthal's Paradox

- Time scale for protein folding
- ~ ns, ps
- For a protein with 100 residues, protein folding occurs in one step
- Nature follows a specific pathway to fold the protein, not a random search
- Folding funnel



Protein Structure Prediction

- Let's predict a protein structure if it is not available
- But as you can see in Levinthal's paradox, it is too much inefficient that search all the possible conformational states

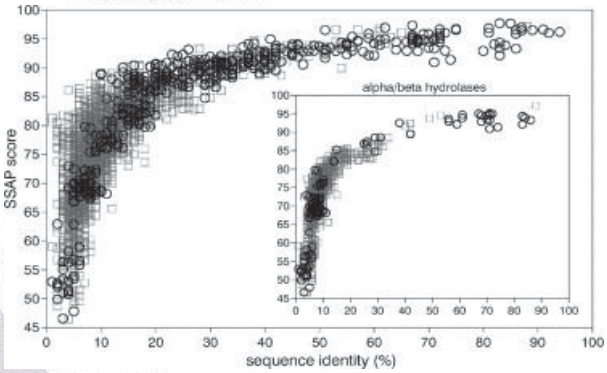


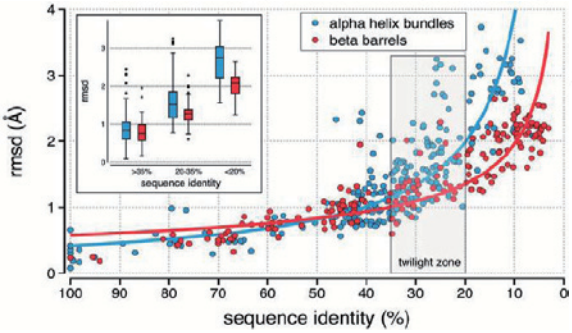
Template-based Modeling

- Template? It's a cast for a shape

Template-based Modeling

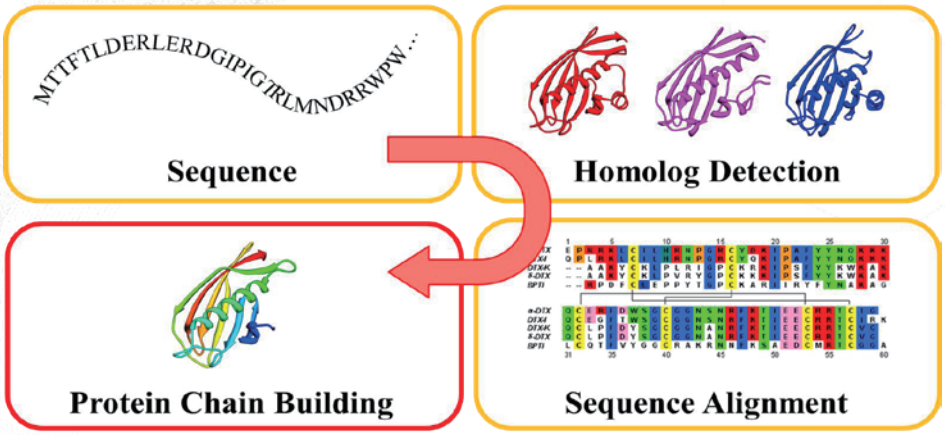
- Basic idea: similar sequence → similar structure
- Lowest bound: 20-25%





Template-based Modeling

- Flow of TBM

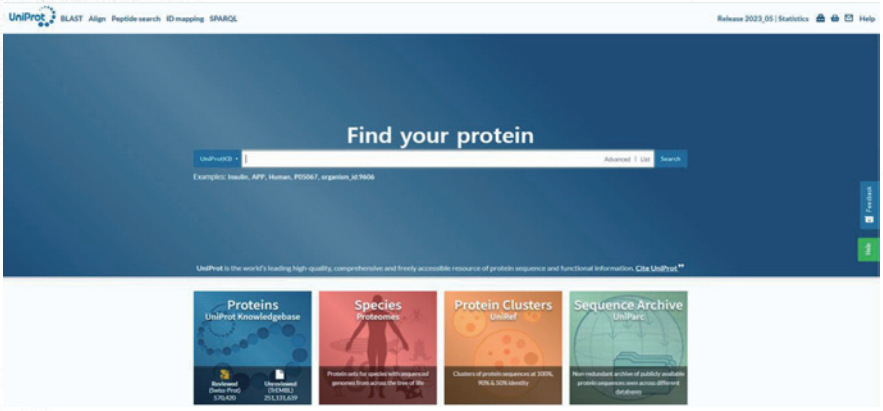


Template-based Modeling

- Template Detection
  - Find proteins with structures that are similar with our target proteins
- Sequence Alignment
  - Detect the similar environment of templates to our sequence
- Model Building
  - Build 3D structure from template information

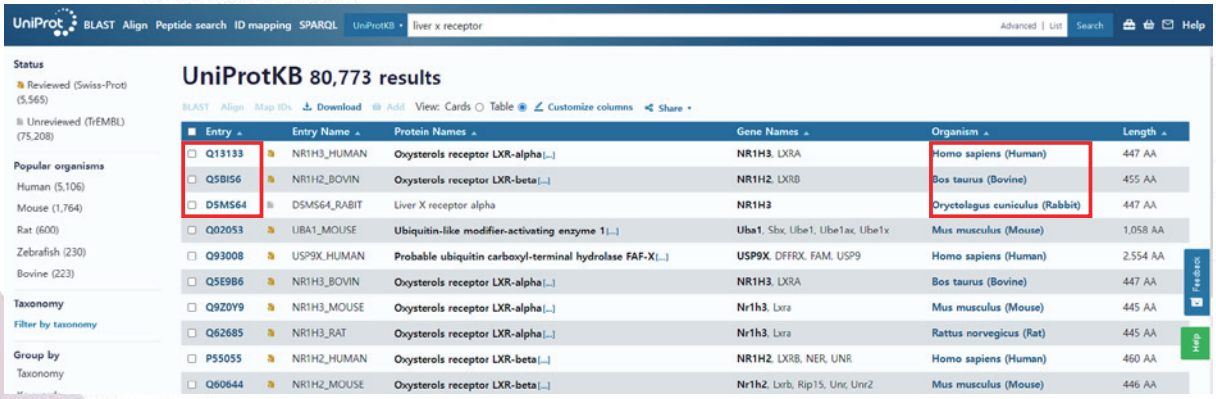
Where to Find My Sequence?

- Search UniProt with a protein name (uniprot.org)

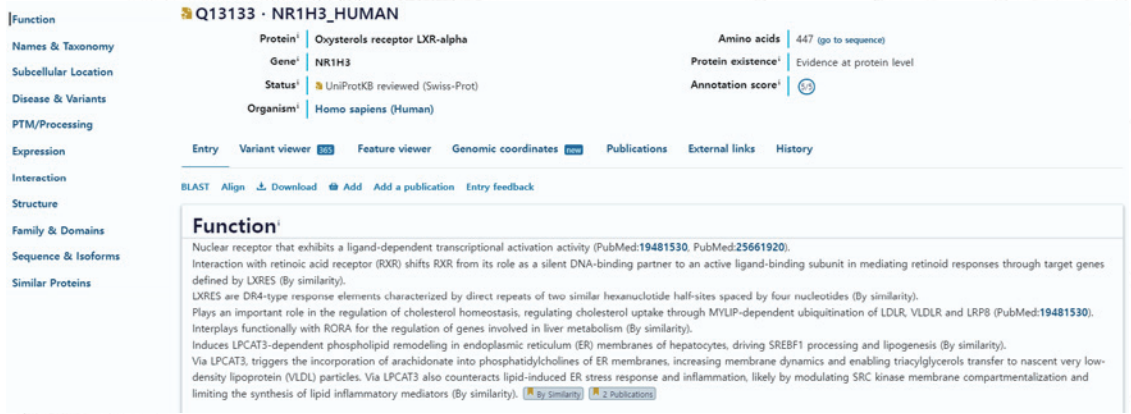


Where to Find My Sequence?

- Click Entry in the result page (Beware of Organism)

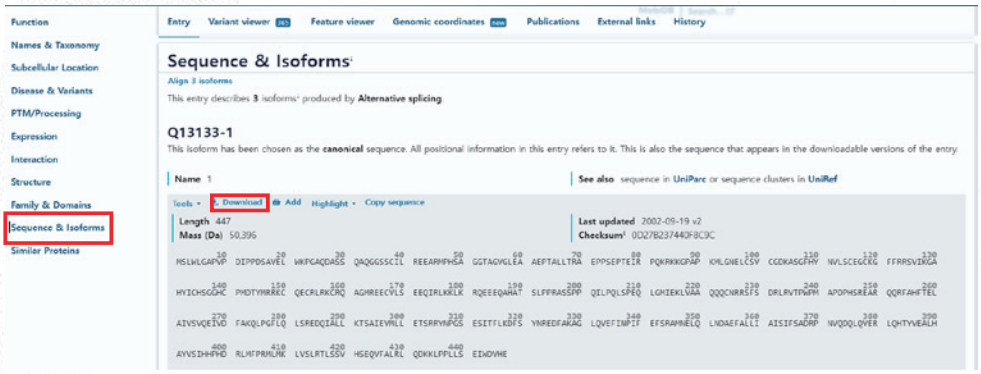


Where to Find My Sequence?



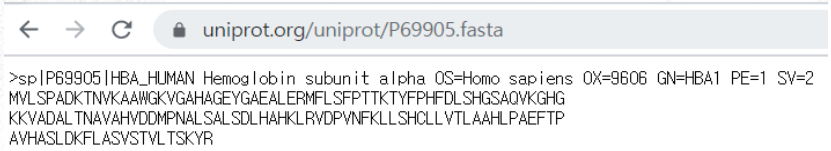
Where to Find My Sequence?

- Click Sequence on the left side and click FASTA on the main page



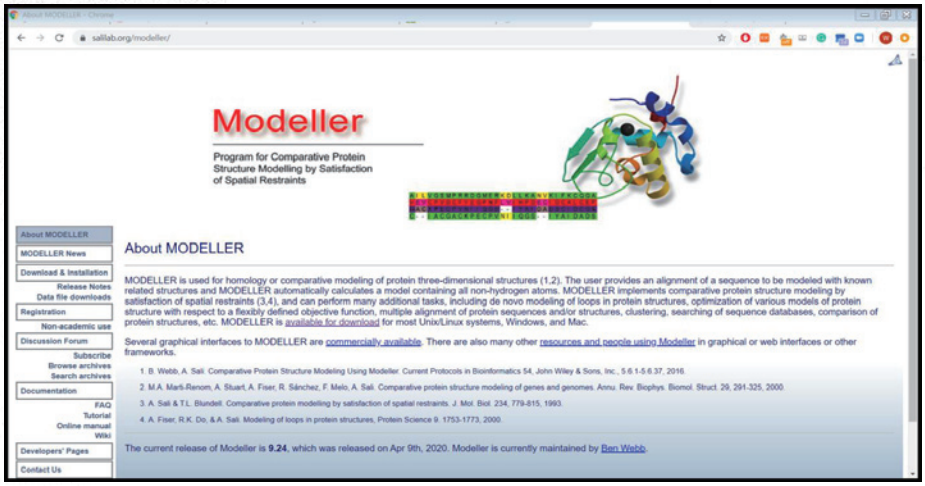
Where to Find My Sequence?

- Copy the text or right-click and select save as



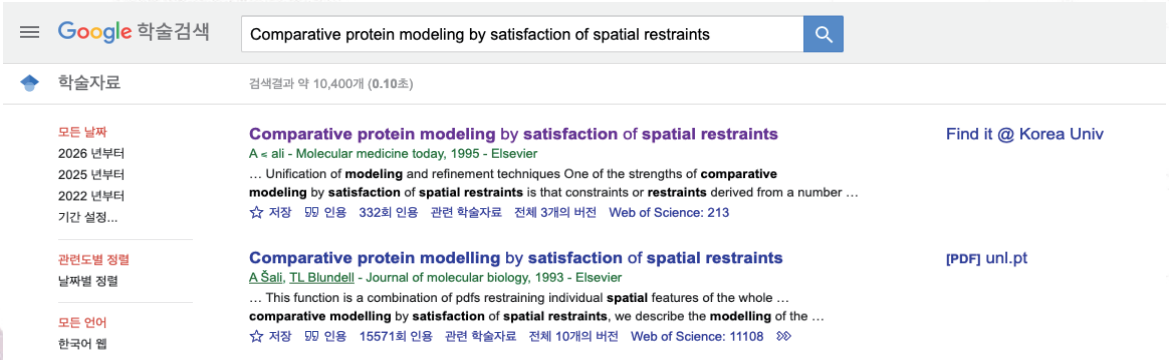
Template-based Modeling

- Modeller (salilab.org/modeller)



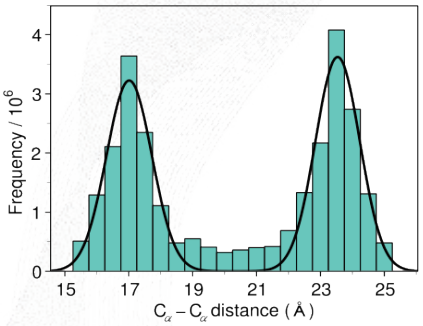
Template-based Modeling

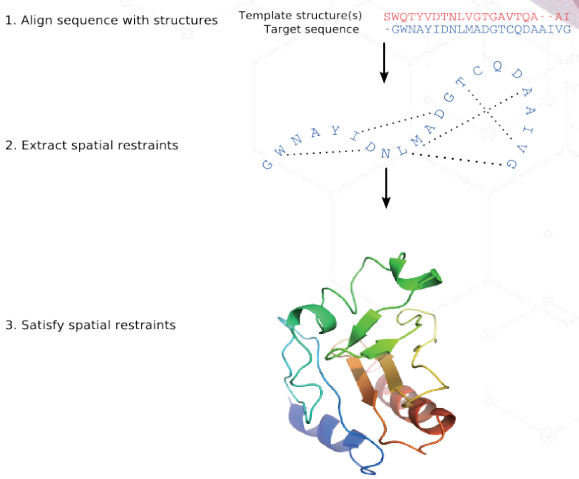
- Modeller



Template-based Modeling

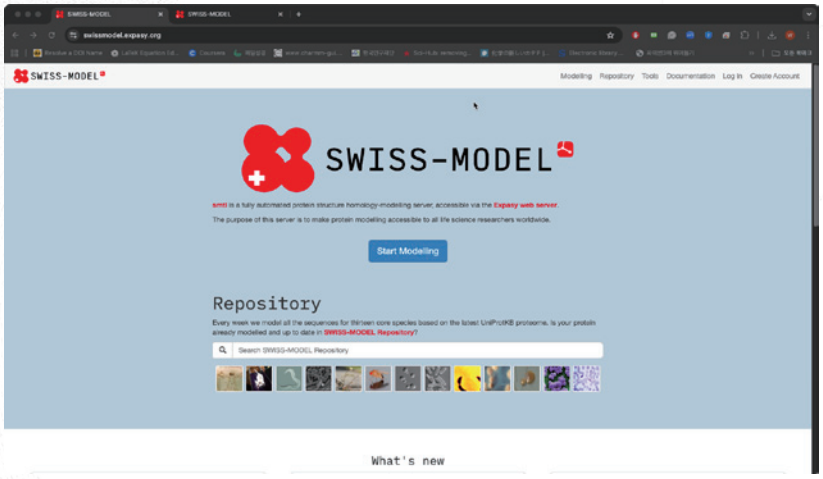
- How Modeller works
- Spatial restraints





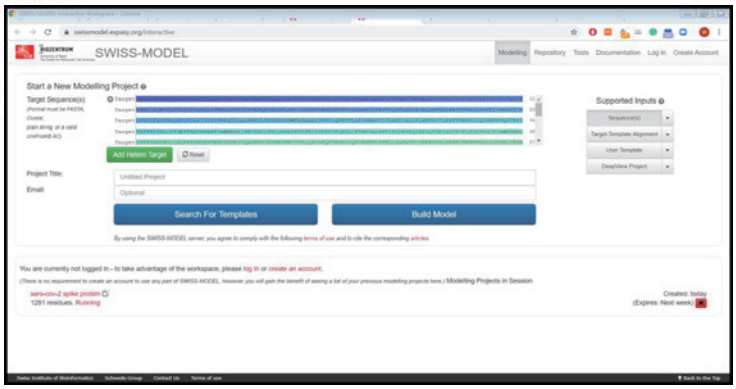
SWISS-MODEL

- https://swissmodel.expasy.org/



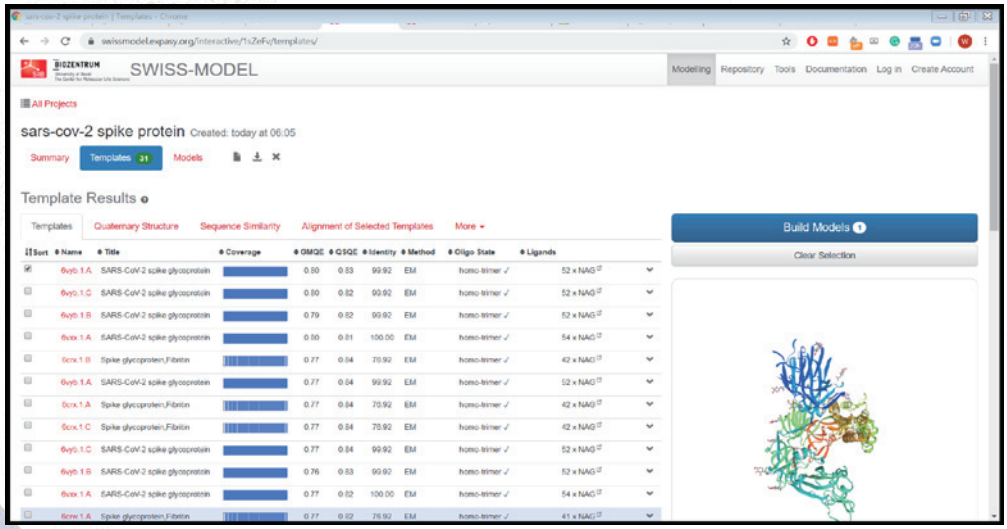
SWISS-MODEL

- Click Start Modeling
- Paste your sequence or upload a text file that contains the sequence



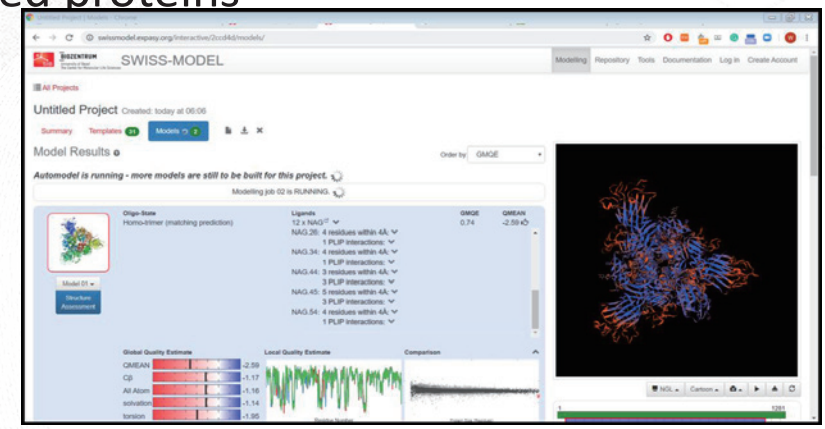
SWISS-MODEL

- Can see the template by clicking Search For Templates.
- The templates are identified by HHBlits



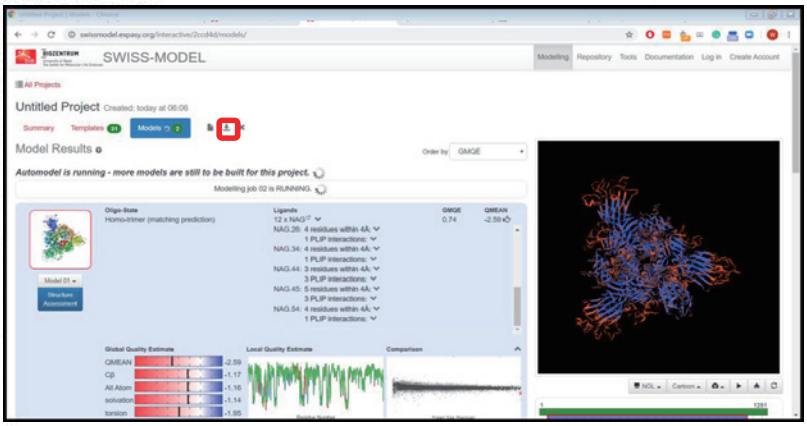
SWISS-MODEL

- Select templates with clicking check boxes on the left and click Build Models to predict structure from the selected proteins



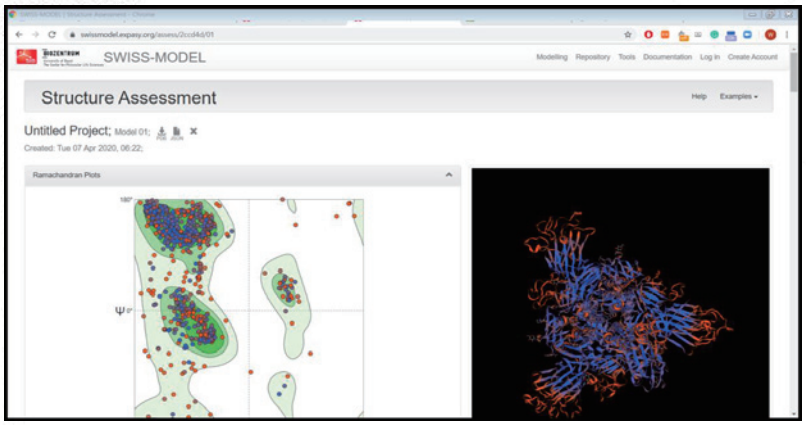
SWISS-MODEL

- Download the result by selecting the downward arrow



SWISS-MODEL

- Can check the quality of the predicted structure by selecting Structure Assessment on the left side



SWISS-MODEL

- MolProbity Score
- How protein-like is the structure?
- The lower, the better

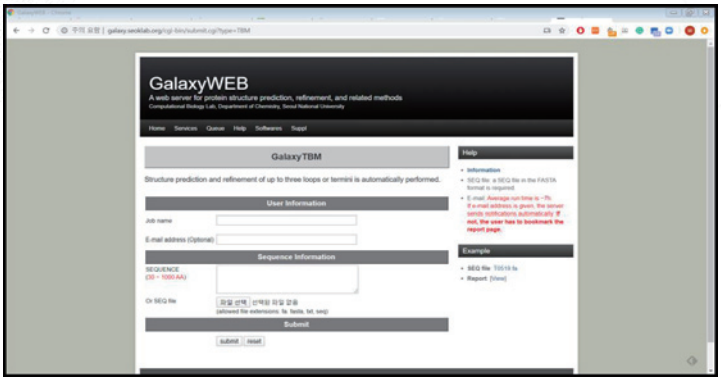
MolProbity Results

MolProbity Score	1.36	
Clash Score	1.17	(A253 ASN-_3 NAG), (C350 ASN-_44 NAG), (C362 ASN-_45 NAG), (C484 GLU-_3 NAG), (A350 ASN-_6 NAG), (A359 GLU-A375 LYS)
Ramachandran Favoured	90.71%	
Ramachandran Outliers	2.09%	ARG, B131 SER, B142 ALA, C874 PHE, B661 VAL, C277 TRP, C142 ALA, B439 ASP, A253 LEU, A186 THR, A199 GLU, A204 ASN, B200 GLY, A117 SER, C652 TRP, B93 ASN, B100 ASN, C708 SER, C87 ILE, A650 PRO, C845 VAL, A270 PRO, A264 HIS, C262 ALA, C274 SER, B873 LYS, A858 ASP, C270 PRO, B656 SER, C118 ASN, A268 LEU, A181 SER, C509 PHE, A645 ALA, A133 THR, C706 VAL, B436 LYS, C650 PRO, A646 ASP, B869 ILE, B192 GLN, A864 ALA, C508 TYR, A273 SER, C205 PHE, B864 ALA, B264 HIS, A183 ASN, C89 VAL, B507 CYS, B85 HIS, C269 THR, C497 THR
Rotamer Outliers	0.86%	B479 ASN, B355 CYS, C163 TYR, C649 THR, C1137 ASP, A887 GLU, B214 LYS, B1137 ASP, B926 ASN, C513 SER, A232 VAL, C173 GLU, C402 SER, C440 TYR, B141 ASN, A1103 ASP, A489 THR, B622 ASN, A141 ASN, C650 PRO, C232 VAL, A247 ASP, B247 ASP, C858 ASP, B646 ASP
C-Beta Deviations	27	C141 ASN, C87 ILE, C93 ASN, B475 PHE, A206 LYS, B103 LEU, C263 LEU, B100 ASN, C474 LEU, B181 SER, B1005 PRO, A277 TRP, C262 ALA, B869 ILE, C657 THR, B93 ASN, C1005 PRO, C1060 ASP, A217 ASP, A764 ASP, C849 ASP, B764 ASP, A864 ALA, B473 ARG, A1005 PRO, C121 ARG, C873 LYS
Bad Bonds	16 / 27009	C362 ASN-_45 NAG, C350 ASN-_44 NAG, A253 ASN-_3 NAG, A1093 ASN-_16 NAG, B1093 ASN-_34 NAG, C1093 ASN-_54 NAG, A736 ASN-_12 NAG, B362 ASN-_26 NAG, A362 ASN-_7 NAG, B301 ASN-_24 NAG, A350 ASN-_6 NAG, _44 NAG, _26 NAG, _23 NAG, _24 NAG
Bad Angles	188 / 36825	(B362 ASN-_26 NAG), (C269 THR-C270 PRO), (A539 ALA-A540 PRO), (A699 SER-A700 PRO), (B509 PHE-B510 PRO), (B192 GLN-B193 PRO), (C497 THR-C498 PRO), C849 ASP, (A253 ASN-_3 NAG), (A100 ASN-A101 PRO), (A645 ALA-A646 ASP), C858 ASP, (A1005 PRO-A1006 PRO), B100 ASN, (B1005 PRO-B1006 PRO), (B100 ASN-B101 PRO), (C350 ASN-_44 NAG), C440 TYR, A85 HIS, (B1161 GLN-B1162 PRO), C197 ASP, C118 ASN, A206 LYS, B361 PHE, (A1161 GLN-A1162 PRO), (C362 ASN-_45 NAG), B505 PHE, (C1161 GLN-C1162 PRO), (C192 GLN-C193 PRO), (B301 ASN-_24
Twisted Non-Proline	2 / 3198	(B99 ASP-B100 ASN), (C648 LEU-C649 THR)

Results obtained using MolProbity version 4.4

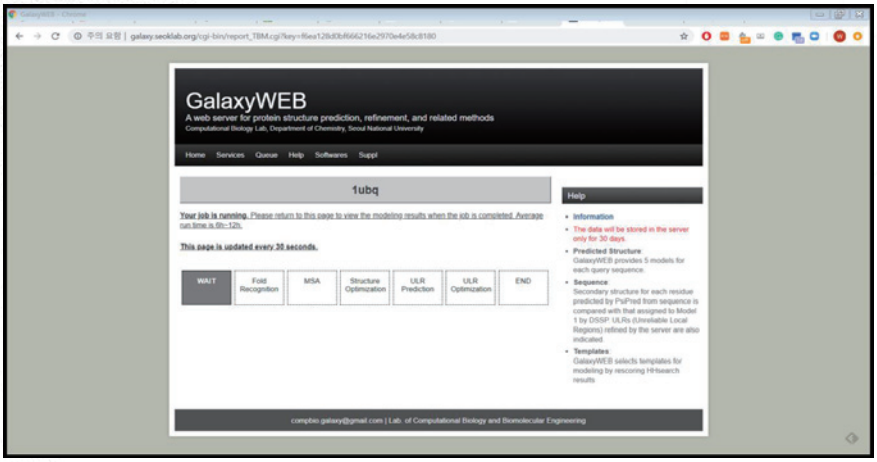
Template-based Modeling

- GalaxyTBM
- http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TBM



Template-based Modeling

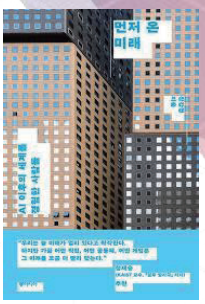
- After submitting your sequence…



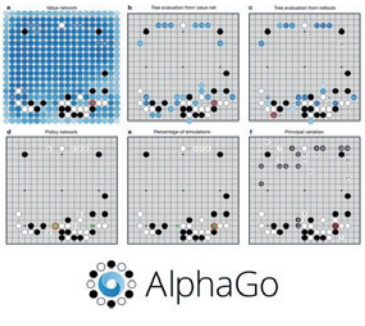
Template-based Modeling

Recent Method using Deep Learning

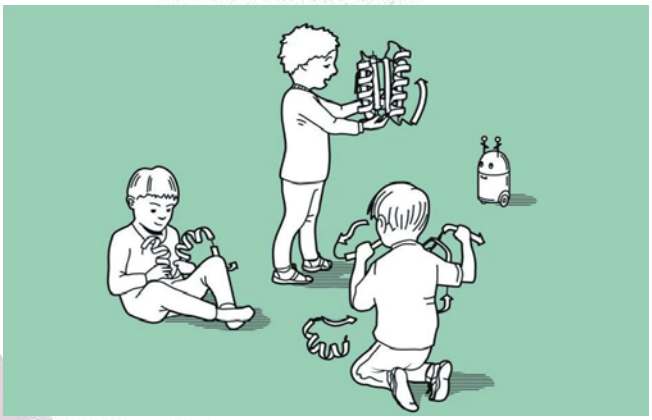
인공지능과 심층학습: AlphaGo



- AlphaGo: 기보 + 강화학습
- AlphaGo Zero: 규칙 + 강화학습
- AlphaZero: 일반화 (장기, 체스…)



Protein Structure Prediction: Deep-Learning Revolution



https://www.nobelprize.org/all-nobel-prizes-2024/

"for computational protein design"

David Baker

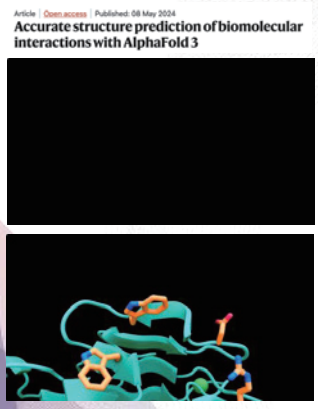
"for protein structure prediction"

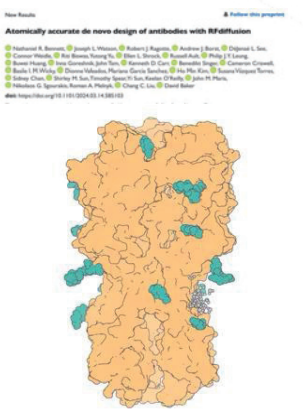
Demis Hassabis

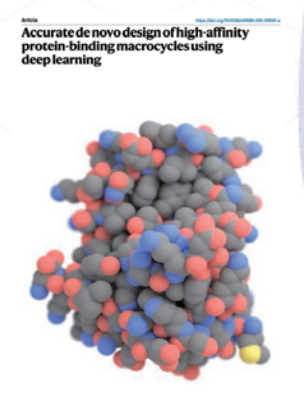
John Jumper

Protein Structure Prediction: Deep-Learning Revolution

AlphaFold3 (2024)	RFdiffusion (2023)	RFantibody (2024)	RFpeptide (2025)
Complex Modeling	Protein Design	Antibody Design	Macrocyclic Peptide Design

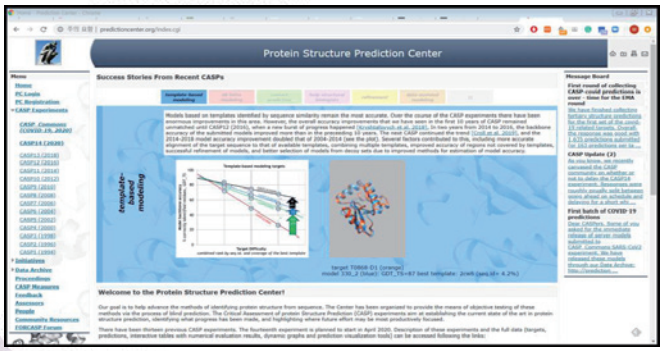


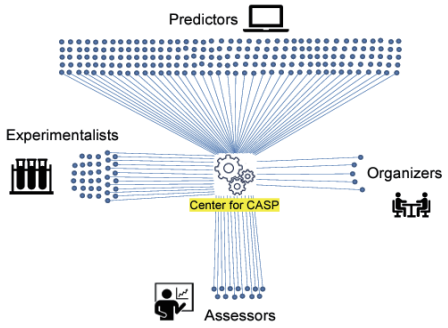




CASP

- Biannual structure prediction contest, blinded





CASP

Prediction Categories of CASP16

Modeling categories

CASP16 categories are as follows:

- **Single Proteins and Domains**
  As in previous CASPs, the accuracy of single proteins and where appropriate single protein domains will be assessed, using the established metrics. The major emphasis is now on the fine-grained accuracy of models, whether limitations related to sequence alignment depth and target size are surmounted, and whether interdomain relationships are accurately captured. There is also interest how well the many new deep learning methods perform, including those using large language models.
- **Protein Complexes**
  As in recent CASPs, the ability of current methods to correctly model subunit-subunit and protein-protein interactions will be assessed. We will again work in close collaboration with our CAPRI partners. There was enormous progress in this category in the last CASP, but accuracy was not yet as high as for single proteins, so there is substantial room for a further advance. New in this CASP is the option of predicting stoichiometry. Where possible, targets will initially be released without that information, models collected, followed by re-release with that data provided.
- **Accuracy Estimation**
  Members of the community will again be invited to submit accuracy estimates for multimeric complexes and inter-subunit interfaces provided by others. There is no longer a category for general methods of estimating single protein structure accuracy, since in recent CASPs estimates provided by model builders have been consistently more reliable. However, there will be an emphasis on the reliability of accuracy estimates provided with submitted structures, both overall and at the individual amino acid level. Note that all accuracy estimates are in plddt units, not Angstroms.
- **Nucleic acid (NA) structures and complexes**
  An RNA structure category was introduced in the previous CASP and the results were interesting and provocative. In particular, it appeared that deep learning methods were not yet as effective as more traditional ones for this type of macromolecule. Has that now changed? This CASP we expect to include RNA and DNA single structures and complexes, and complexes of these with proteins.
- **Protein - organic ligand complexes**
  The last round of CASP included this category for the first time. Results indicated that, as with RNA structure, deep learning methods were not yet competitive with more traditional approaches. So there is considerable interest in whether that has now changed. In addition to ligands integral to protein targets, we expect to have several target sets related to drug design.
- **Macromolecular conformational ensembles**
  Following the success of deep-learning methods for single structures, it is increasingly important to assess methods for predicting structure ensembles, and CASP included this category for the first time in 2022. While it was clear deep learning methods have considerable potential for generating ensembles, the best procedures are still hotly debated with many new papers appearing. In CASP16, we expect to have a variety of targets for both protein and RNA ensembles.
- **Integrative modeling**
  Deep learning methods combined with sparse experimental data such as SAXS and chemical crosslinking are now being used extensively to obtain the structure of large marcomolecular complexes. To assess effectiveness of these approaches, CASP is reintroducing this category of modeling, provided appropriate targets will be available.

CASP

- Targets of CASP16

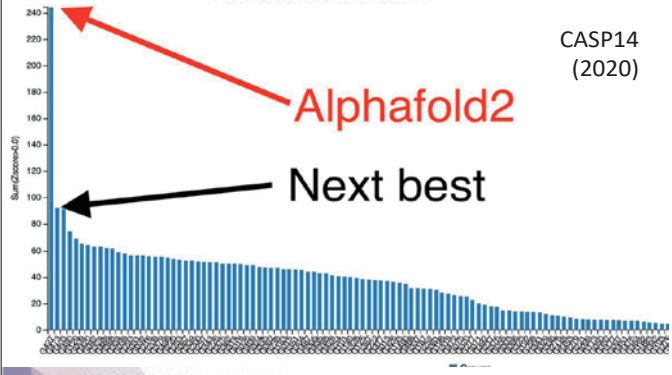
All targets | Mono/Homo. -mers | Multimers | RNA | Ligand

All groups | Server

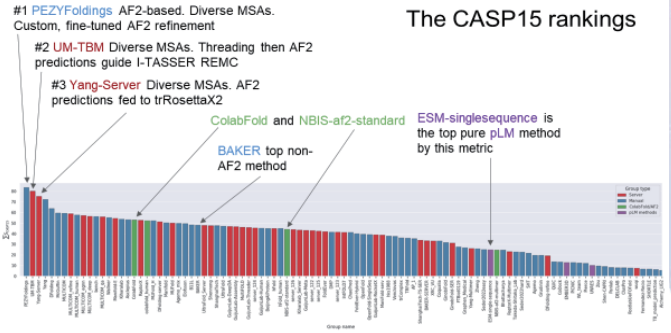
#	Target	Type	Res	Stoichiom.	Entry Date	Server Expiration	Ligand Deadline	Human Expiration	QA Prediction	Description
1.	T1201 *	All groups	210	A2	2024-05-01	2024-05-04	2024-05-15	2024-05-15	from: 2024-05-15 to: 2024-05-18	Q9GZX9
2.	H1202 *	All groups	190	A2B2	2024-05-02	2024-05-05	2024-05-16	2024-05-14	from: 2024-05-16 to: 2024-05-19	Q9GZX9_plus_P43026
3.	R1203	NucA	134	R1	2024-05-02	2024-05-05	2024-05-14	2024-05-14	-	Rev response element stem-loop II
4.	H1204 *	All groups	429	A2B2C2	2024-05-03	2024-05-06	2024-05-15	2024-05-17	from: 2024-05-17 to: 2024-05-20	Human hemoglobin in complex with nanobody
5.	L1000	Ligand	147	A1	2024-05-03	2024-05-03	2024-05-13	2024-05-13	-	Chymase
6.	L2000	Ligand	103	A1	2024-05-03	2024-05-03	2024-05-13	2024-05-13	-	Cathepsin G
7.	L3000	Ligand	346	A1	2024-05-03	2024-05-03	2024-05-13	2024-05-13	-	Achtheim
8.	L4000	Ligand	306	A2	2024-05-03	2024-05-04	2024-05-14	2024-05-15	-	MPro
9.	R1205	NucA	59	R1	2024-05-03	2024-05-06	2024-05-17	2024-05-17	-	xrRNA
10.	T0206	All groups	237	UNK	2024-05-06	2024-05-09	2024-05-20	2024-05-20	-	Porcine astrovirus 4 capsid spike
11.	T1207	All groups	144	A1	2024-05-06	2024-05-09	2024-05-20	2024-05-20	-	Q9YLS4, 2A protein part of viral polyprotein
12.	H0208	All groups	646	UNK	2024-05-07	2024-05-10	2024-05-21	2024-05-21	-	dahAB
13.	T0208x1	All groups	328	A1	2024-05-07	2024-05-10	2024-05-21	2024-05-21	-	dahAB subunit 1
14.	T0208x2	All groups	318	A1	2024-05-07	2024-05-10	2024-05-21	2024-05-21	-	dahAB subunit 2
15.	H0215	All groups	369	UNK	2024-05-08	2024-05-11	2024-05-22	2024-05-22	-	mNeonGreen with bound nanobody
16.	M1209	Prot-NucA	520	A1B1R1	2024-05-08	2024-05-11	2024-05-20	2024-05-22	-	FAB - HIV-1 RRE stem loop II
17.	R1209	NucA	72	R1	2024-05-08	2024-05-11	2024-05-22	2024-05-22	-	FAB - HIV-1 RRE stem loop II RNA component
18.	T1210	All groups	1770	A1	2024-05-09	2024-05-12	2024-05-30	2024-05-30	-	Q66NR5
19.	M1211	Prot-NucA	282	A2R1	2024-05-10	2024-05-13	2024-05-24	2024-05-24	-	CVB3 cloverleaf RNA in complex with 3C protease
20.	R1211	NucA	90	A1	2024-05-10	2024-05-13	2024-05-24	2024-05-24	-	CVB3 cloverleaf RNA in complex with 3C protease
21.	T1214	All groups /Ligand	677	A1	2024-05-10	2024-05-13	2024-05-22	2024-05-24	-	YncC (NP_415968)
22.	M1212	Prot-NucA	809	A1R1V1W1	2024-05-13	2024-05-16	2024-05-22	2024-05-03	-	Fanzor2 ternary structure
23.	R1212	NucA	147	A1R1V1W1	2024-05-13	2024-05-16	2024-05-03	2024-05-03	-	Fanzor2 ternary structure
24.	T1213	All groups	464	A1	2024-05-13	2024-05-16	2024-05-03	2024-05-03	-	Fanzor2 ternary structure
25.	H0213	All groups	1635	UNK	2024-05-14	2024-05-17	2024-05-28	2024-05-23	-	DP71_eIF2_PP1_complex
26.	M1216	Prot-NucA	1106	A4B4C4D4	2024-05-15	2024-05-18	2024-05-29	2024-05-29	-	Tetrameric SPARSAgRNA-target ssDNA complex
27.	T2201	All groups	210	A2	2024-05-15	2024-05-18	2024-05-28	2024-05-27	-	Q9GZX9

CASP: AlphaFold Revolution

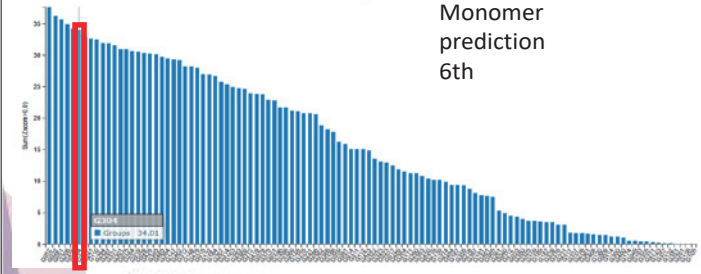
- AlphaFold2 from DeepMind dominates!



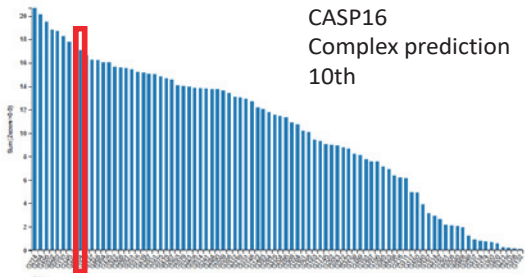
CASP14
(2020)



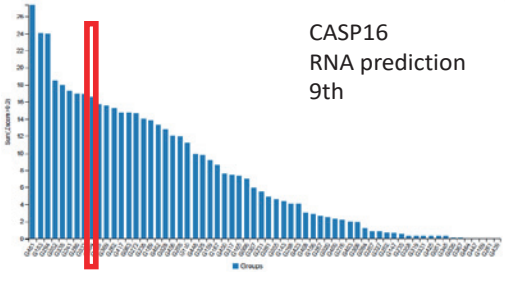
CASP: AlphaFold Revolution



CASP16
Monomer
prediction
6th



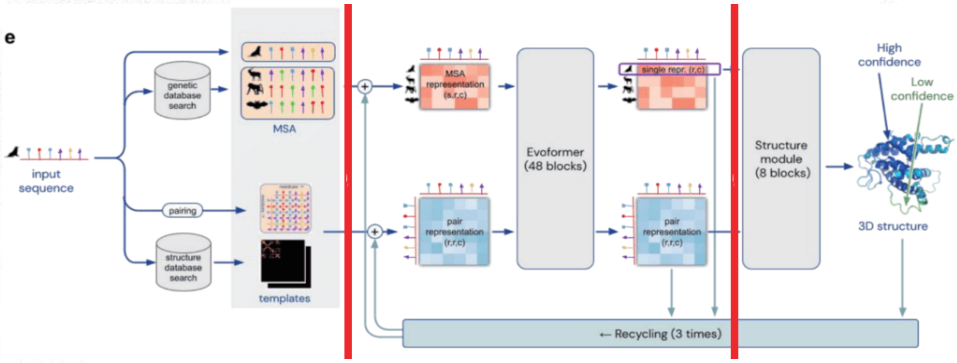
CASP16
Complex prediction
10th



CASP16
RNA prediction
9th

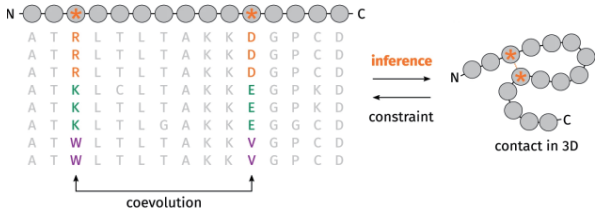
Flow of AlphaFold2

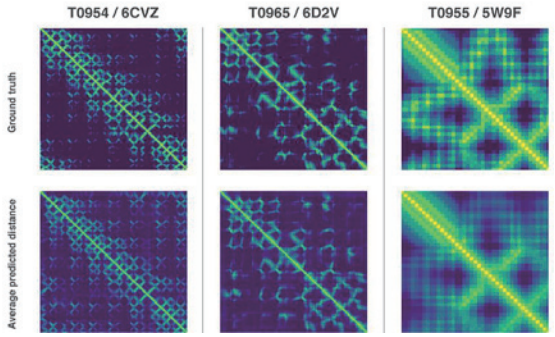
- How AF2 works



Getting Information from MSA

- Coevolution and contact map: deep learning with structural database





Variants of AF2

New Results

Follow this preprint:

Assessing Structures and Conformational Ensembles of Apo and Holo Protein States Using Randomized Alanine Sequence Scanning Combined with Shallow Subsampling in AlphaFold2 : Insights and Lessons from Predictions of Functional Allosteric Conformations

Nishank Raisinghani, Vedant Parikh, Brandon Foley, Gennady Verkhivker

doi: https://doi.org/10.1101/2024.11.04.621947

AlphaFold-Metainference: Prediction of Structural Ensembles of Disordered Proteins

Z. Faidon Brotzakis, Shengyu Zhang, Mhd Hussein Murtada, Michele Vendruscolo

doi: https://doi.org/10.1101/2024.11.09.622758

PROTEINS

STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

RESEARCH ARTICLE | Open Access

Multi-state modeling of G-protein coupled receptors at experimental accuracy

Lim Heo, Michael Feig

First published: 05 May 2022 | https://doi.org/10.1002/prot.26382 | Citations: 29

JOURNAL ARTICLE

MoDAFold: a strategy for predicting the structure of missense mutant protein based on AlphaFold2 and molecular dynamics

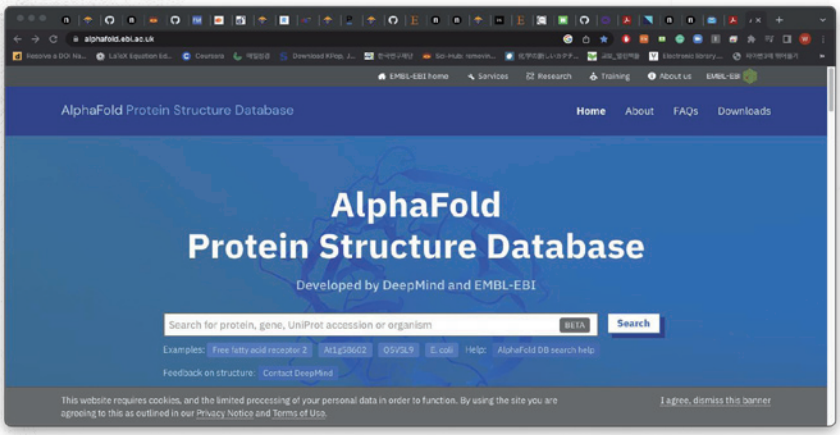
Lingyan Zheng, Shuiyang Shi, Xiuna Sun, Mingkun Lu, Yang Liao, Sisi Zhu, Hongning Zhang, Ziqi Pan, Pan Fang, Zhenyu Zeng ... Show more

Author Notes

Briefings in Bioinformatics, Volume 25, Issue 2, March 2024, bbae006, https://doi.org/10.1093/bib/bbae006

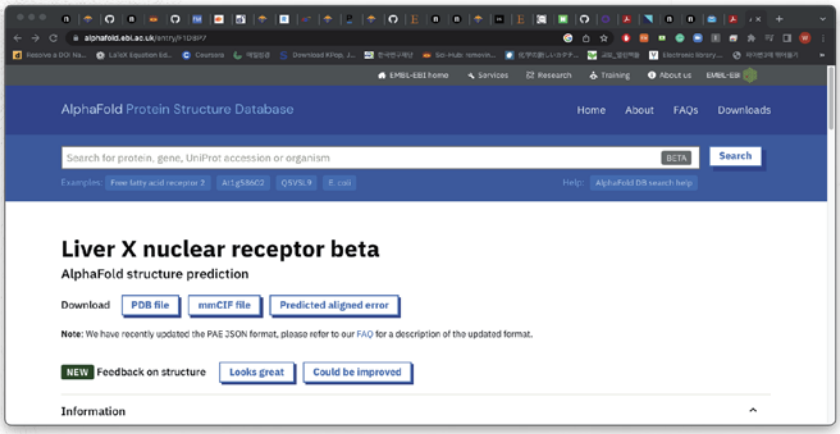
AF2 Structure Database

- EBI-AF2 database (https://alphafold.ebi.ac.uk/)



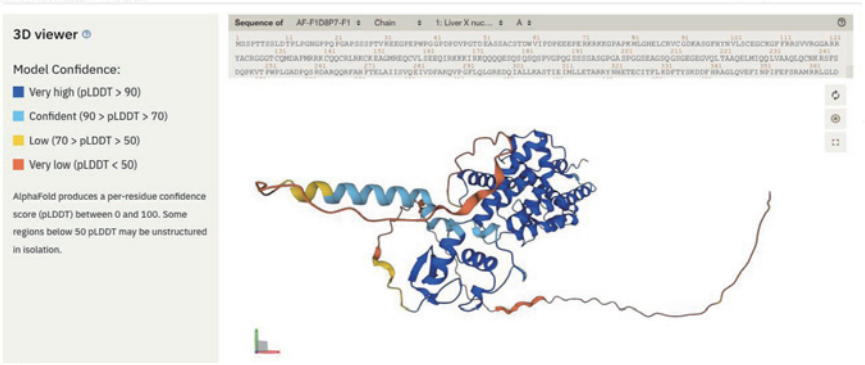
AF2 Structure Database

- Search with protein name or UniProt AC



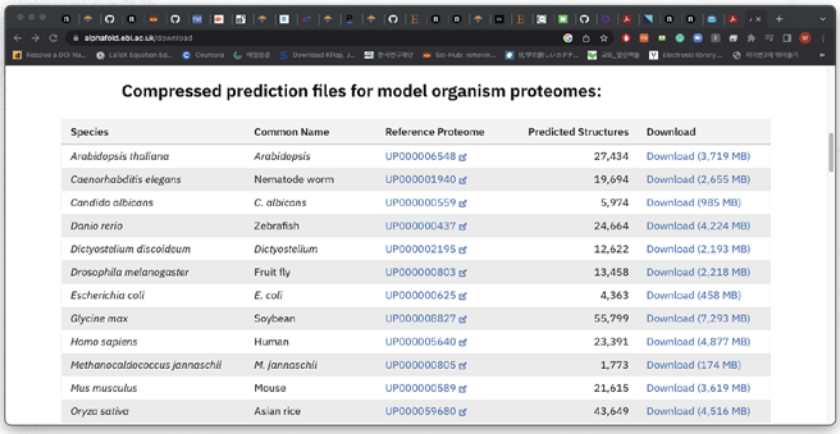
AF2 Structure Database

- pLDDT shows the reliability of a residue
- Can see a detailed structure with your mouse



AF2 Structure Database

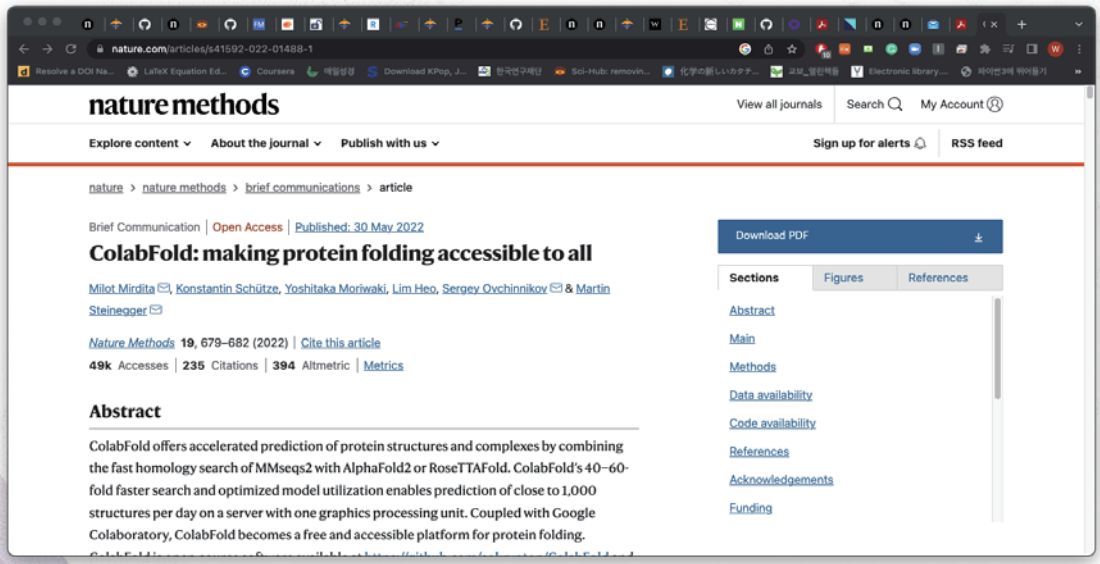
- If you are working on the proteome, get all proteins of an organism by cl icking download



Compressed prediction files for model organism proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
Arabidopsis thaliana	*Arabidopsis*	UP000006548	27,434	Download (3,719 MB)
Caenorhabditis elegans	Nematode worm	UP000001940	19,694	Download (2,655 MB)
Candida albicans	*C. albicans*	UP000000559	5,974	Download (985 MB)
Danio rerio	Zebrafish	UP000000437	24,664	Download (4,224 MB)
Dictyostelium discoideum	*Dictyostelium*	UP000002195	12,622	Download (2,193 MB)
Drosophila melanogaster	Fruit fly	UP000000803	13,458	Download (2,218 MB)
Escherichia coli	*E. coli*	UP000000625	4,363	Download (458 MB)
Glycine max	Soybean	UP000008827	55,799	Download (7,293 MB)
Homo sapiens	Human	UP000005640	23,391	Download (4,877 MB)
Methanocaldococcus jannaschii	*M. jannaschii*	UP000000805	1,773	Download (174 MB)
Mus musculus	Mouse	UP000000589	21,615	Download (3,619 MB)
Oryza sativa	Asian rice	UP000059680	43,649	Download (4,516 MB)

ColabFold: AF2 Opened to Public



nature methods

View all journals | Search | My Account

Explore content | About the journal | Publish with us | Sign up for alerts | RSS feed

nature > nature methods > brief communications > article

Brief Communication | Open Access | Published: 30 May 2022

ColabFold: making protein folding accessible to all

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov & Martin Steinegger

Nature Methods 19, 679–682 (2022) | Cite this article

49k Accesses | 235 Citations | 394 Altmetric | Metrics

Download PDF

Sections | Figures | References

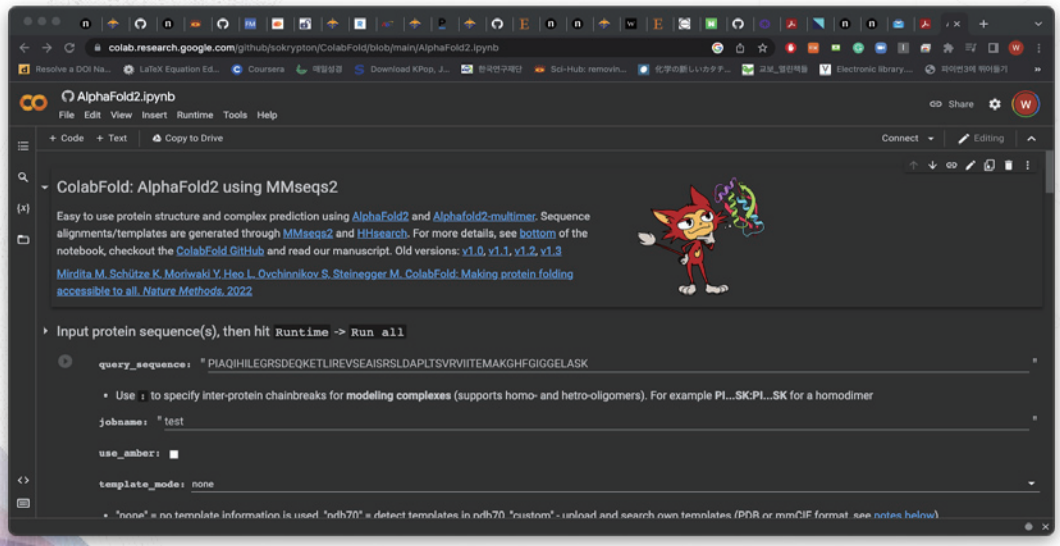
Abstract
Main
Methods
Data availability
Code availability
References
Acknowledgements
Funding

Abstract

ColabFold offers accelerated prediction of protein structures and complexes by combining the fast homology search of MMseqs2 with AlphaFold2 or RoseTTAFold. ColabFold's 40–60-fold faster search and optimized model utilization enables prediction of close to 1,000 structures per day on a server with one graphics processing unit. Coupled with Google Colaboratory, ColabFold becomes a free and accessible platform for protein folding.

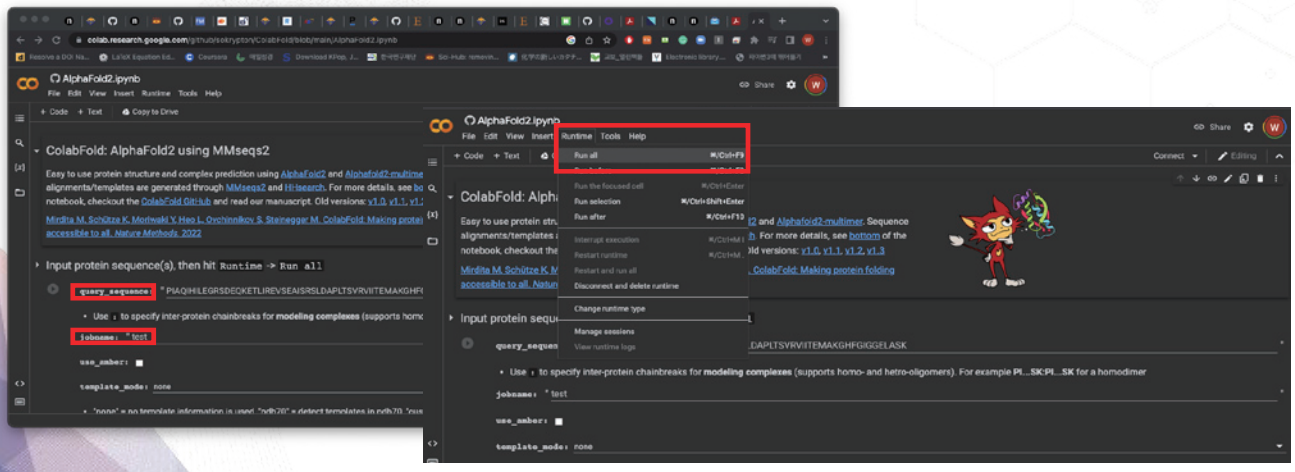
ColabFold

- https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb



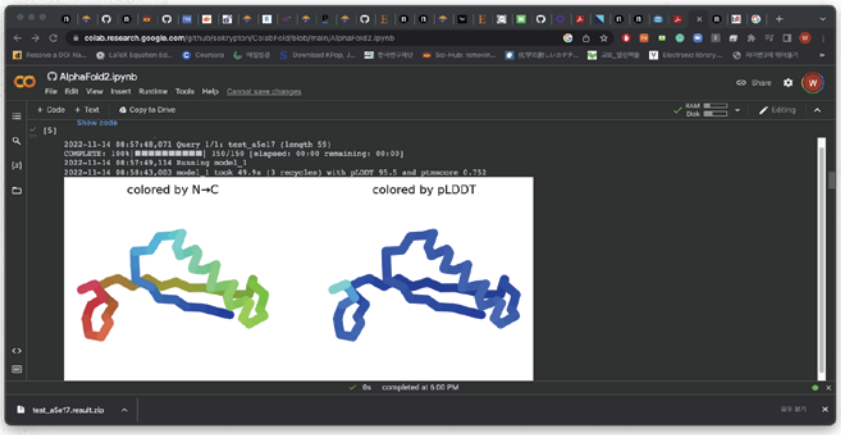
ColabFold

- Put sequence and jobname, click Runtime > Run all



ColabFold

- Five predicted models are zipped and downloaded automatically



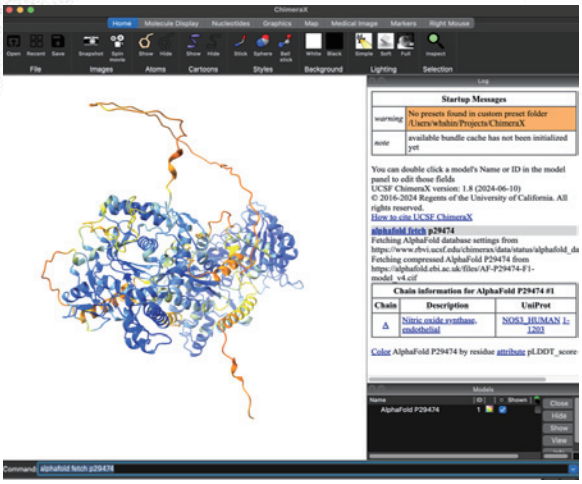
Modeling using ChimeraX

- ChimeraX supports ColabFold or retrieving AlphaFold DB
- Get models from DB

```
Usage: alphafold fetch uniprot-id [ alignTo chain-spec [ trim true | false ]] [ colorConfidence true | false ] [ ignoreCache true | false ] [ pae true | false ] [ version 1 | 2 | 3 | 4 ]
Usage: alphafold match sequence [ search true | false ] [ trim true | false ] [ colorConfidence true | false ] [ ignoreCache true | false ] [ pae true | false ]
Usage: alphafold search sequence [ matrix similarity-matrix ] [ cutoff evalue ] [ maxSequences M ] [ version 1 | 2 | 3 | 4 ]
```

Modeling using ChimeraX

- Result of *alphafold fetch p29474*



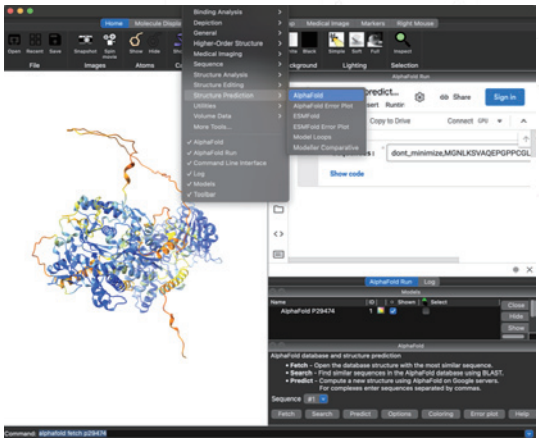
Modeling using ChimeraX

- *Tools > Structure prediction > AlphaFold* opens a colab page

Or

- Type in command line:

alphafold predict sequence



Comparing Structures using ChimeraX

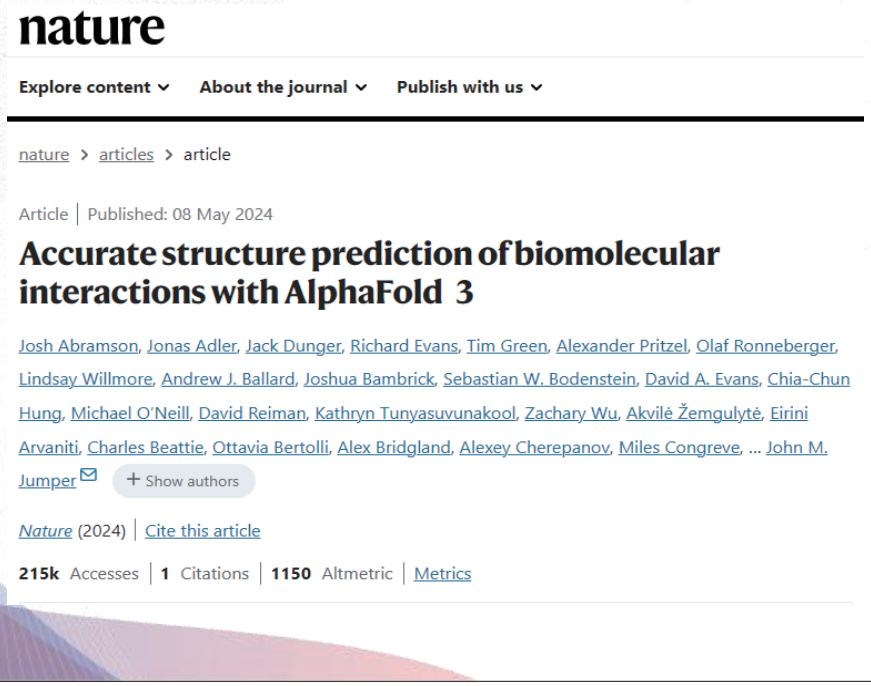
- Tools > Structure Analysis > Matchmaker

Or

- mmaker #model1 to #model2

AlphaFold3

- AlphaFold3 published in Nature!



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Published: 08 May 2024

Accurate structure prediction of biomolecular interactions with AlphaFold 3

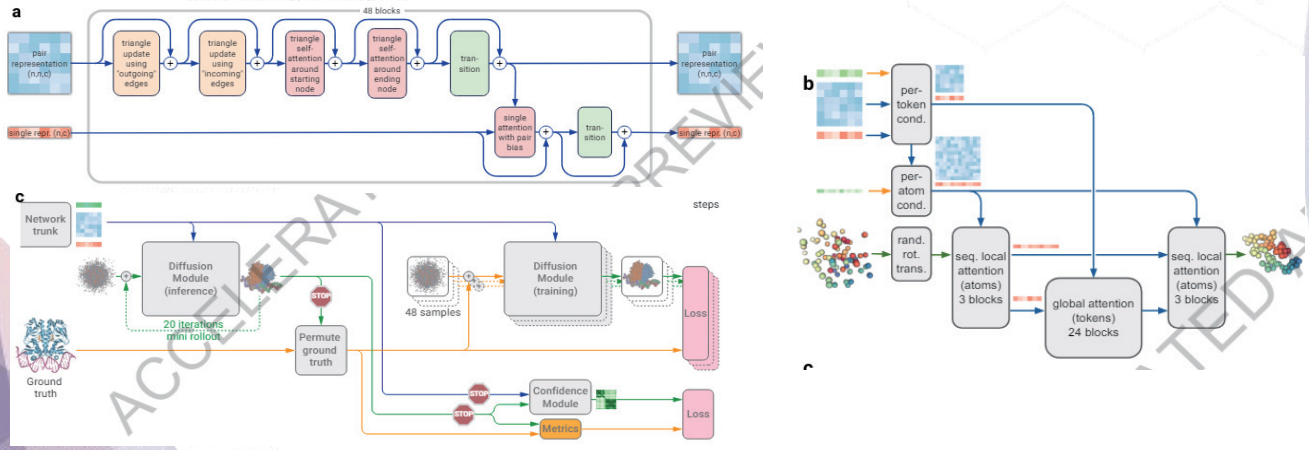
Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, ... John M. Jumper + Show authors

Nature (2024) | Cite this article

215k Accesses | **1** Citations | **1150** Altmetric | Metrics

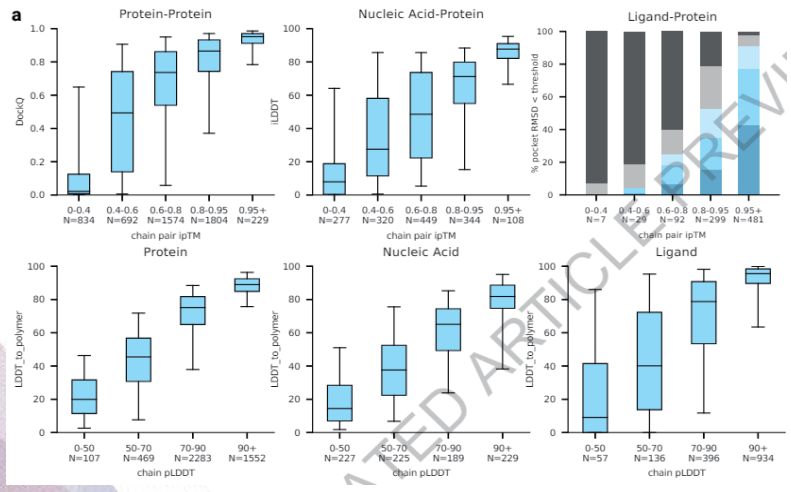
AlphaFold3

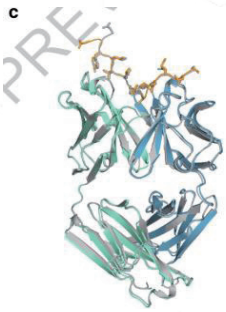
- Adopts diffusion model for structure prediction

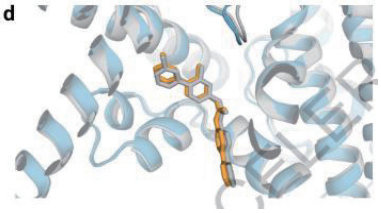


AlphaFold3

- Now it can deal with complex structures

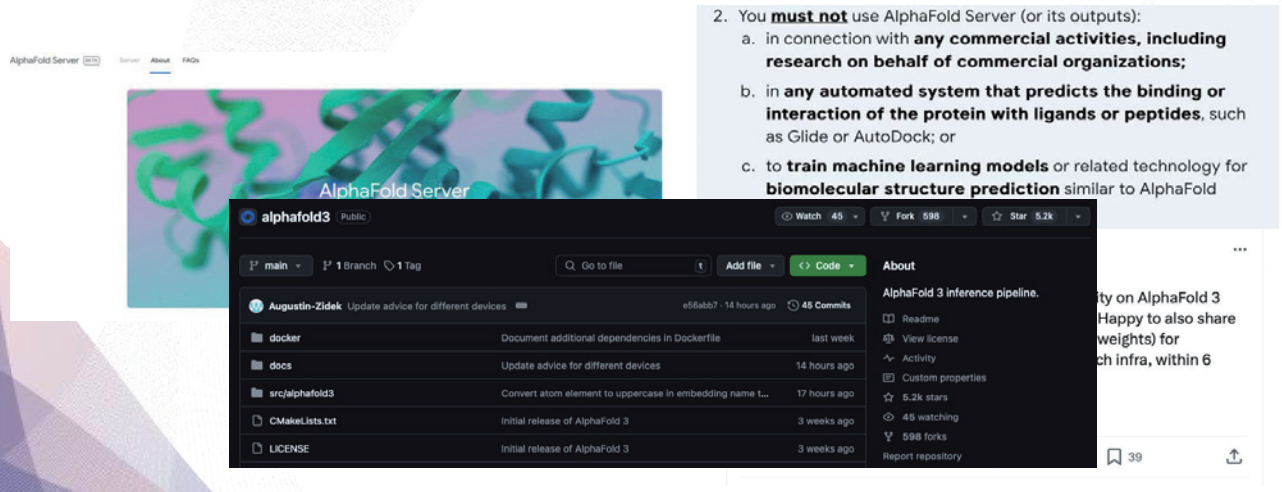






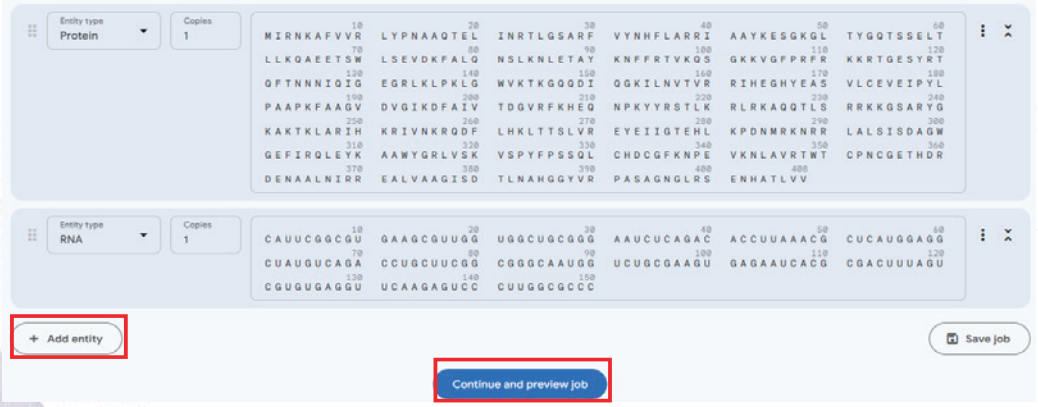
AlphaFold3

- AlphaFold3 webserver and policies (https://alphafoldserver.com/)

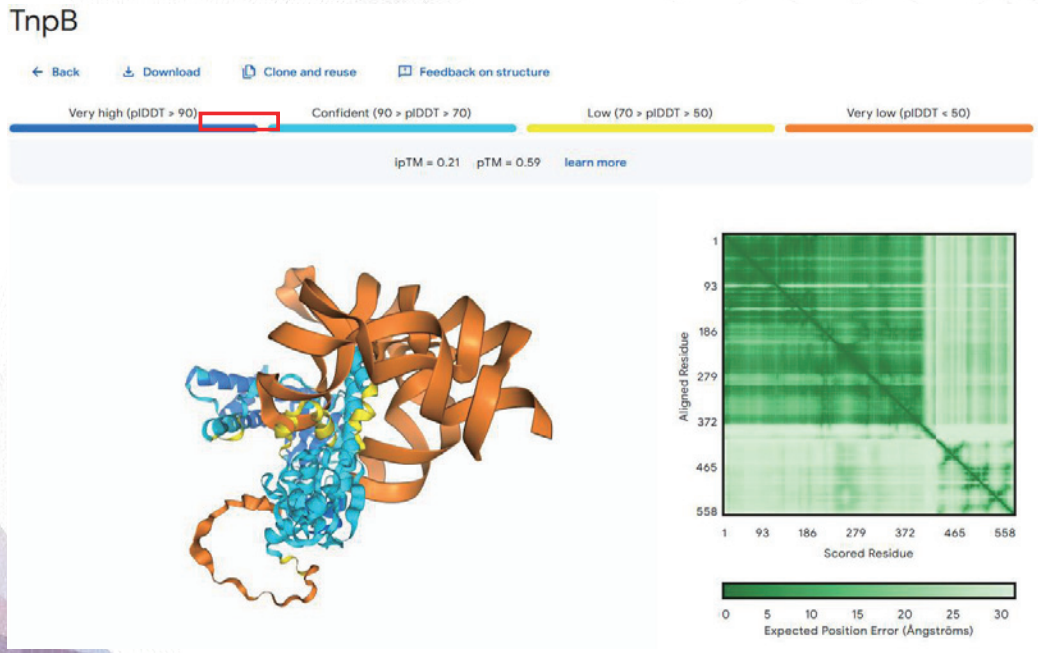


AF3 Webserver

- Select the molecule type, put your target sequence
- To model the complex, add entity



AF3 Webserver

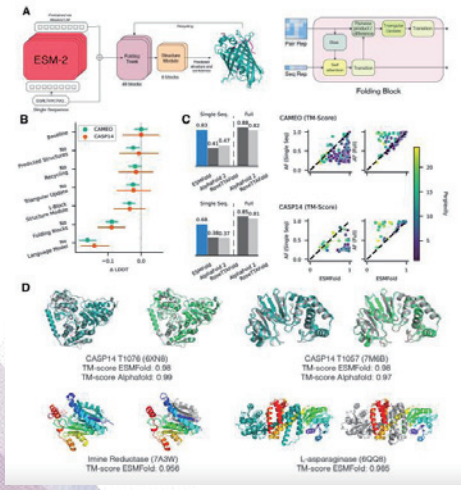


AF3 Webserver

- Job limitations
- 30 jobs / day
- 5000 tokens / day
  - One amino acid, one base = 1 token
  - One atom for small molecule = 1 token

ESMFold

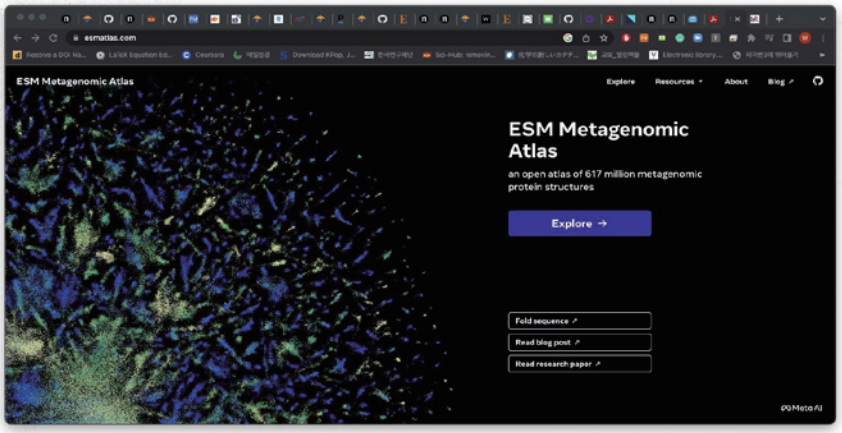
- by Meta



Meta's network, called ESMFold, isn't quite as accurate as AlphaFold, Rives' team reported earlier this year[2], but it is about 60 times faster at predicting structures for short sequences, he says. "What this means is that we can scale structure prediction to much larger databases."

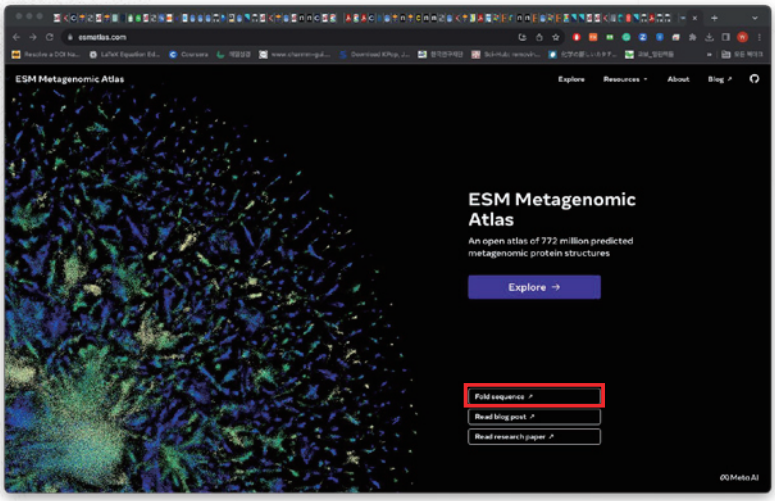
ESMFold

- Structure database (esmatlas.com)



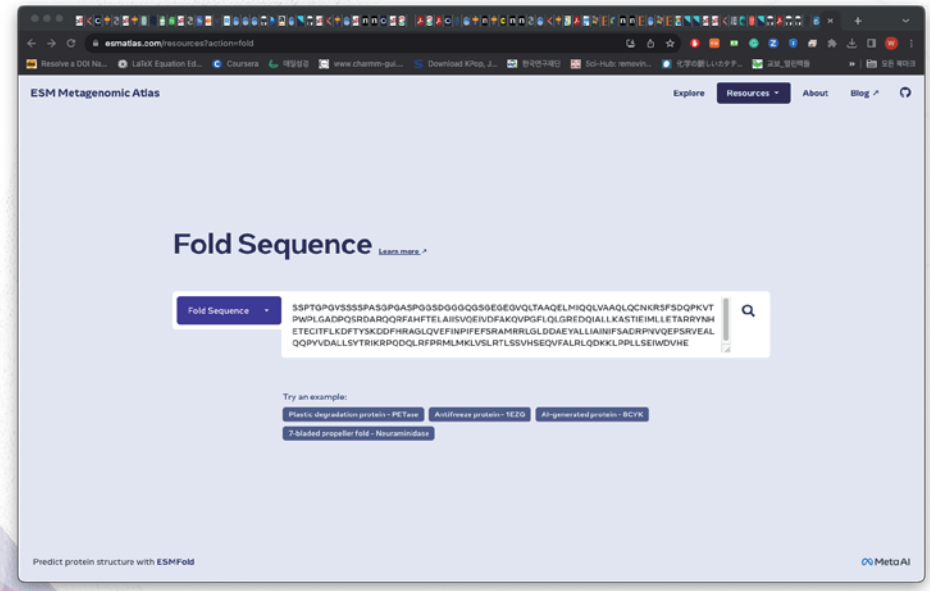
ESMFold

- Structure prediction using ESMFold

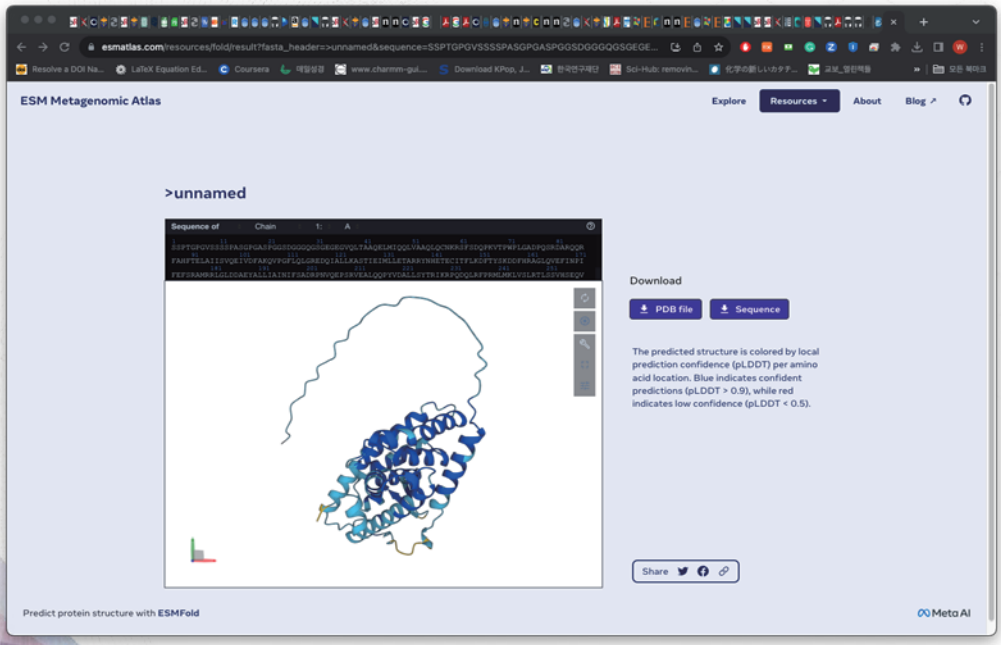


ESMFold

- Paste sequence and click the magnifying glass

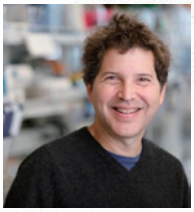


ESMFold



RoseTTAFold

- D. Baker of UW



RoseTTAFold

- RoseTTAFold-AA