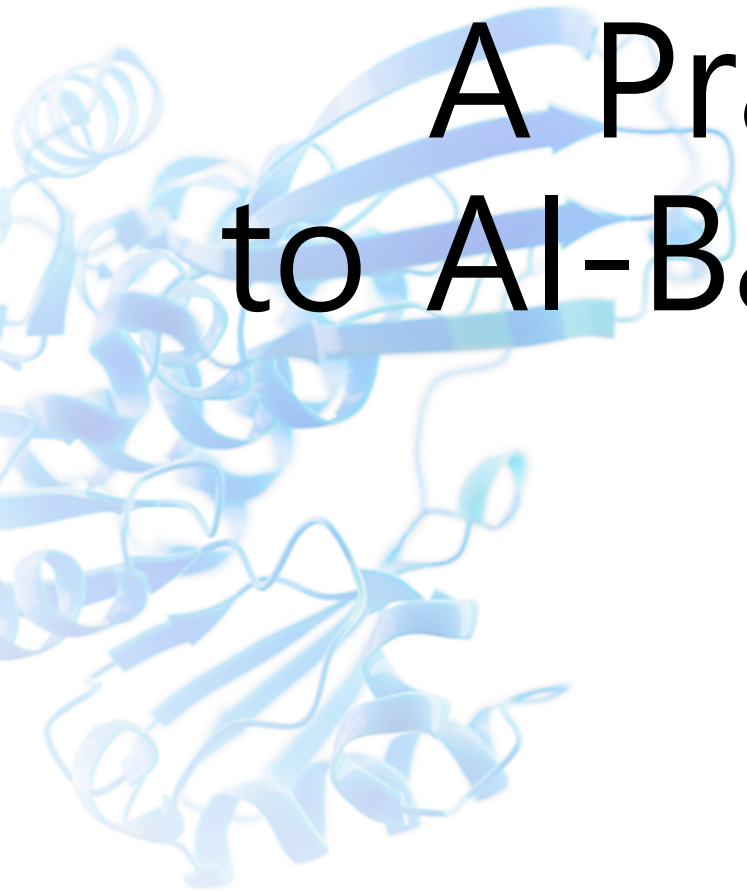
한국응용생명화학회
The Korean Society for Applied Biological Chemistry



A Practical Introduction to AI-Based Protein Structure Prediction

2025-11-28

한국과학기술연구원 천연물연구소
고 영 준 박사후연구원

KIST 한국과학기술연구원
Korea Institute of Science and Technology

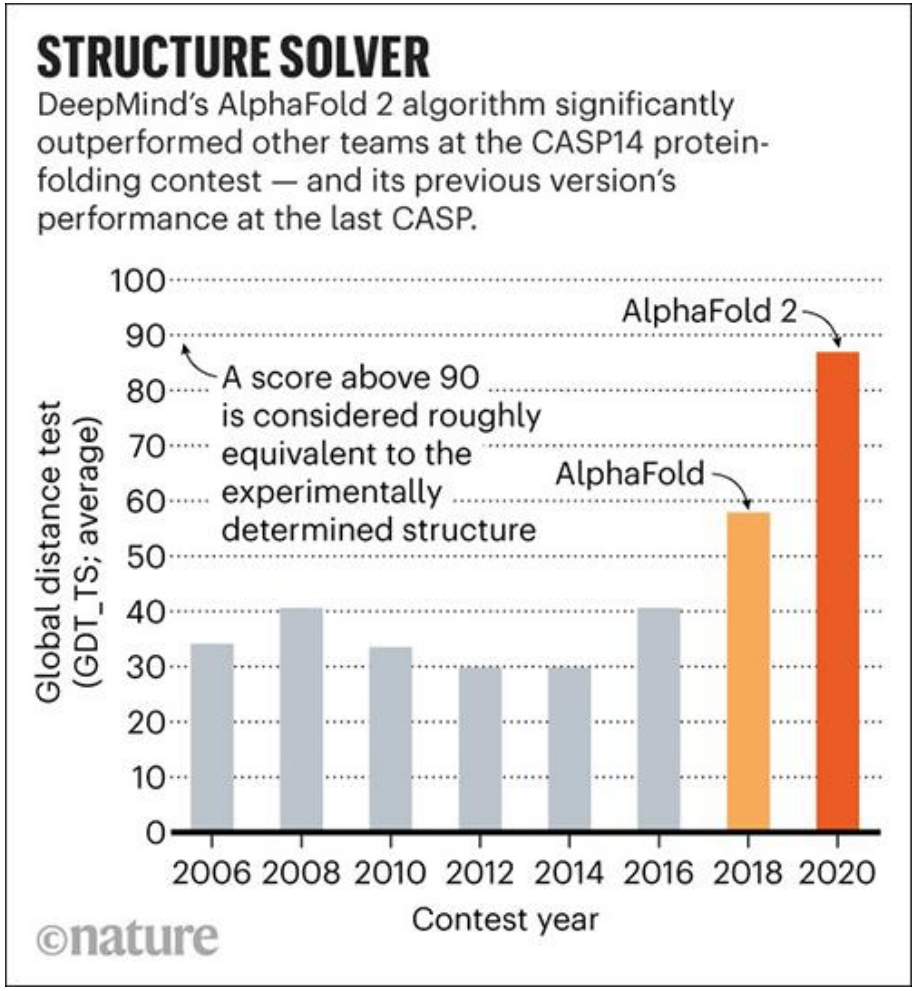
Contents

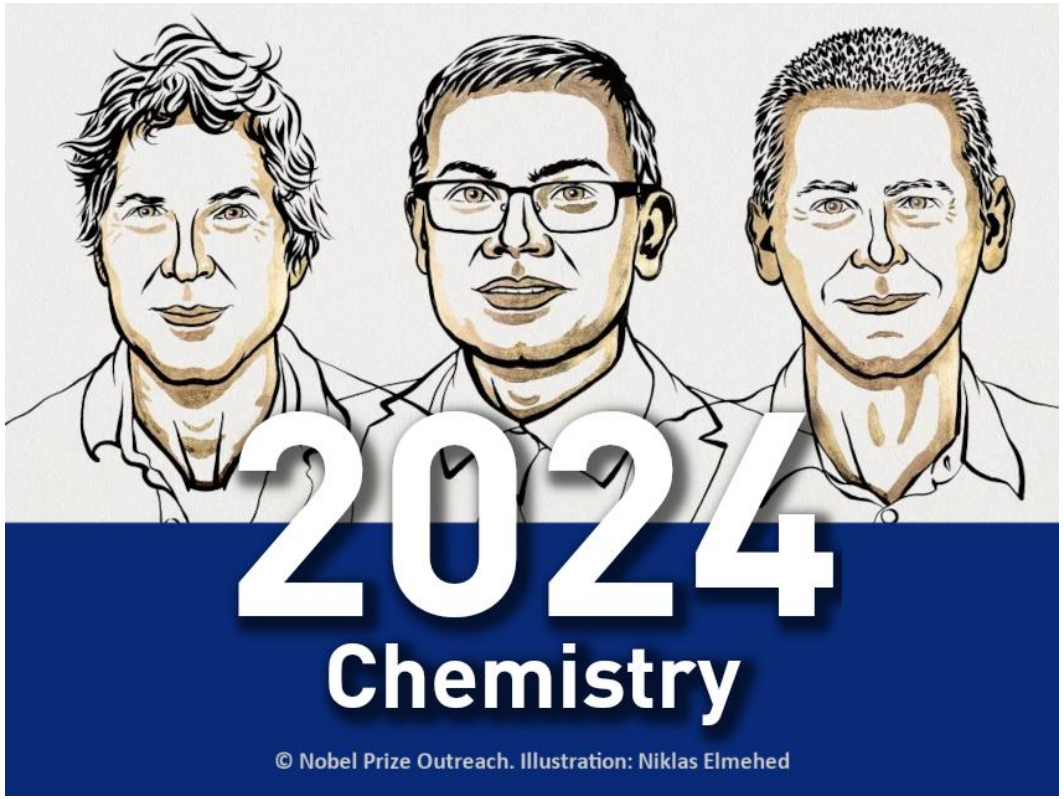
- Introduction
  - Overview of Alphafold
  - 연구 활용 사례 (Glutamine binding protein)

- Cloud-based 환경에서 단백질 구조예측 사용하기
  - Alphafold server 사용하기
  - Other Cloud models

- On-premises 환경에서 Alphafold3 사용하기
  - Docker로 Alphafold3 환경 구성하기
  - input json 파일 작성
  - docker 명령어 작성
  - Alphafold3 예제

Alphafold

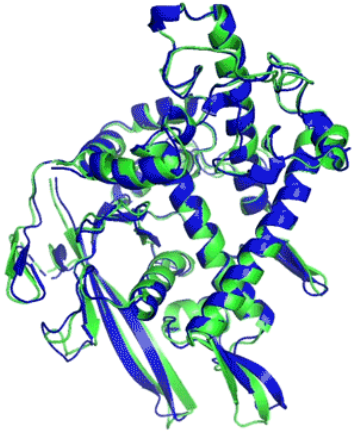




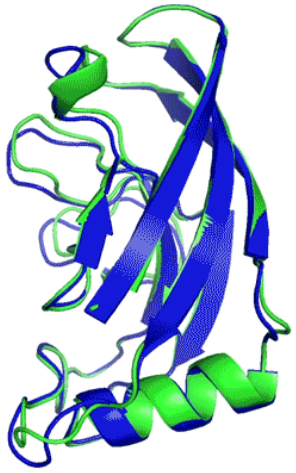
50년 이상의 난제 "단백질 폴딩 문제"
단백질 서열로 3D 구조 예측 어려움

Alphafold 3 등장 배경

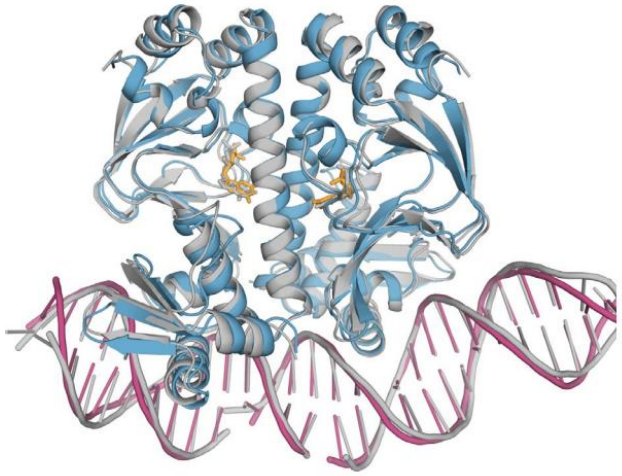
- Alphafold 2의 등장(2020), 단백질 단일 사슬 구조예측도가 혁신적으로 향상
- 단백질과 이종 분자(리간드, 핵산(DNA/RNA), 이온)와의 상호작용 구조 예측이라는 과제가 남아 있었음



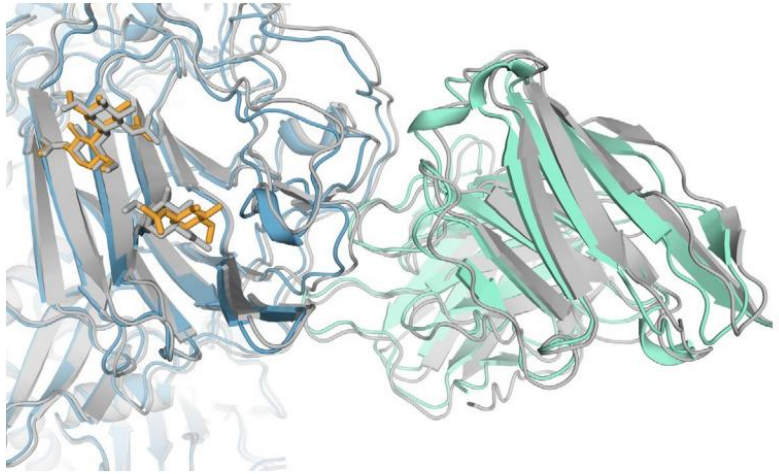
T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



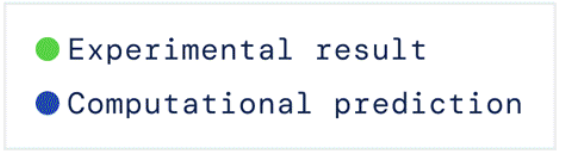
T1049 / 6y4f
93.3 GDT
(adhesin tip)



Bacterial CRP/FNR family transcriptional regulator protein bound to DNA and cGMP



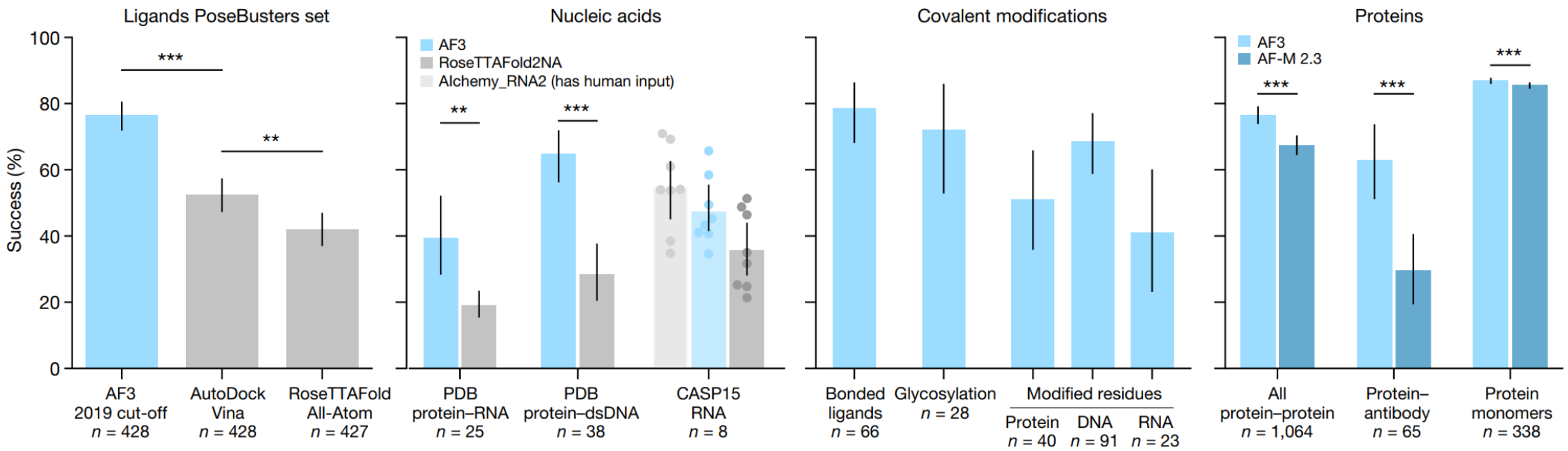
Human coronavirus OC43 spike protein, 4,665 residues, heavily glycosylated and bound by neutralizing antibodies



Abramson et al., *Nature* 630, 493–500 (2024).

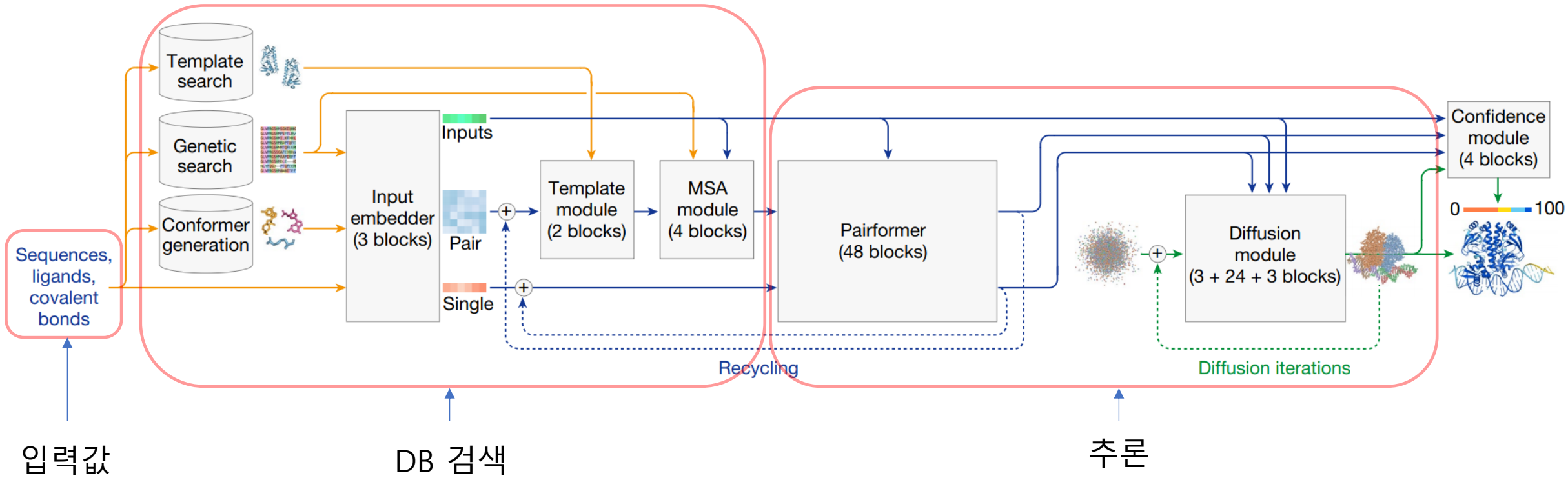
Alphafold 3의 성능 및 특징

- 예측 정확도 향상
- 단백질-단백질과 단백질-이종 분자(리간드, 핵산(DNA/RNA), 이온)간 복합체 모델링 가능
- Post-translational modification (PTM) 가능
- 항원-항체 상호작용 모델링 가능



Abramson et al., *Nature* 630, 493–500 (2024).

Alphafold 3의 기본 작동 구조



Abramson et al., *Nature* 630, 493–500 (2024).

Alphafold의 연구 활용 사례

International Journal of Biological Macromolecules 323 (2025) 147072

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

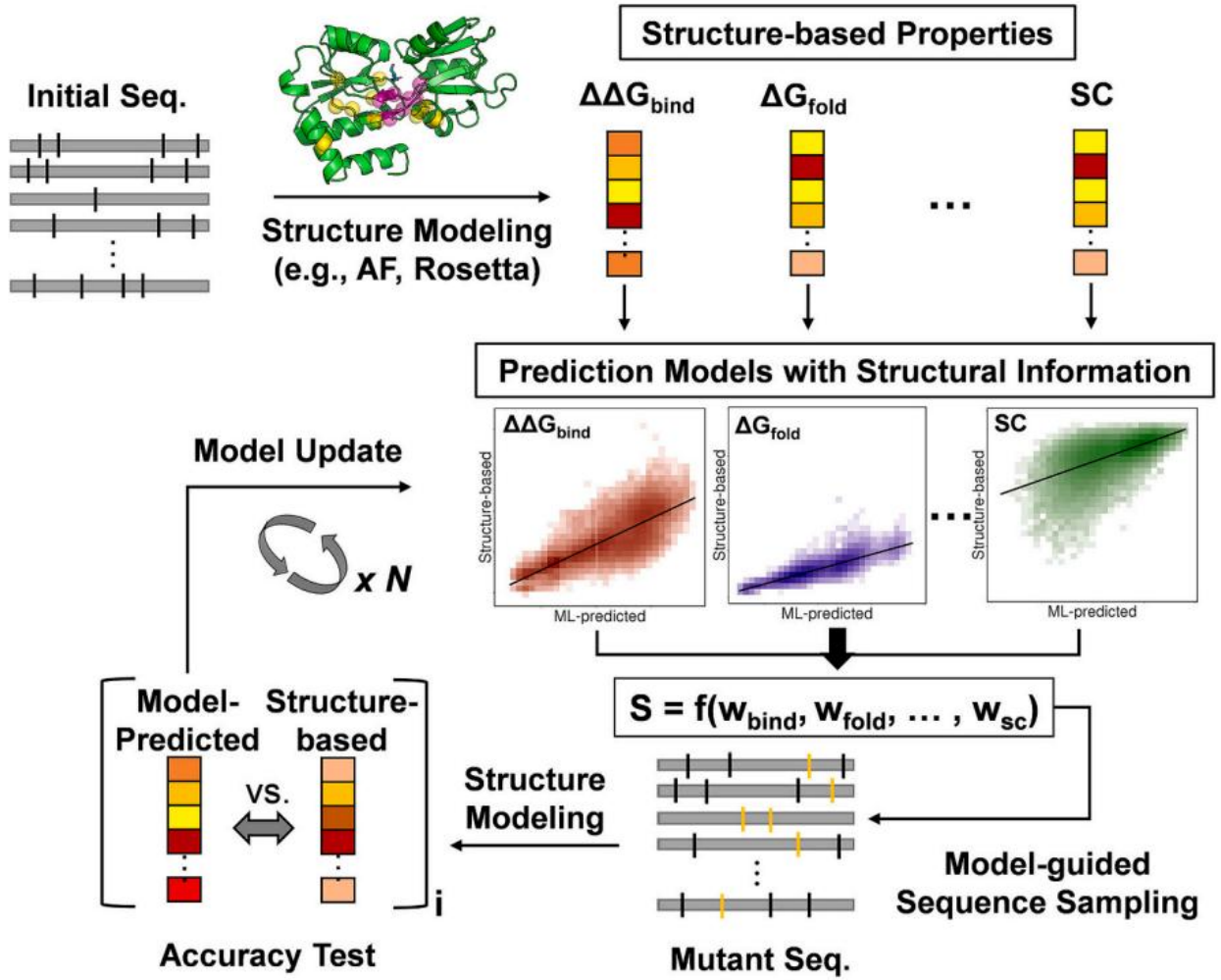
Check for updates

Enhancing protein structural properties through model-guided sequence optimization

Young-Joon Ko [a,1], Dohyeon Kim [a,b,1], Charuvaka Muvva [a,1], Won-Kyu Lee [d], Moon-Hyeong Seo [c,*], Keunwan Park [a,*]

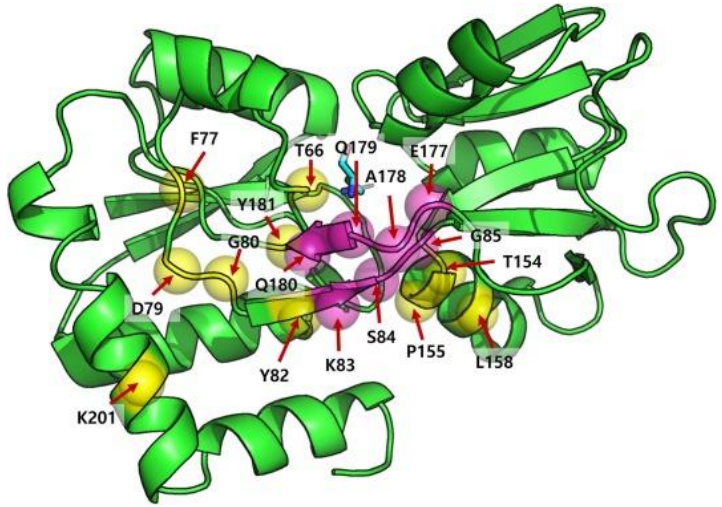
[a] Center for Natural Product Systems Biology, Korea Institute of Science and Technology, Gangneung, 25451, Republic of Korea
[b] Department of Bioinformatics and Life Science, Soongsil University, Seoul, 06978, Republic of Korea
[c] Center for Natural Product Efficacy Optimization, Korea Institute of Science and Technology, Gangneung, 25451, Republic of Korea
[d] New Drug Development Center, Osong Medical Innovation Foundation, Cheongju 28160, Republic of Korea

Alphafold의 연구 활용 사례



Overview of the sequence optimization workflow for improving protein structural properties

Mutation sites selected in glutamine-binding protein



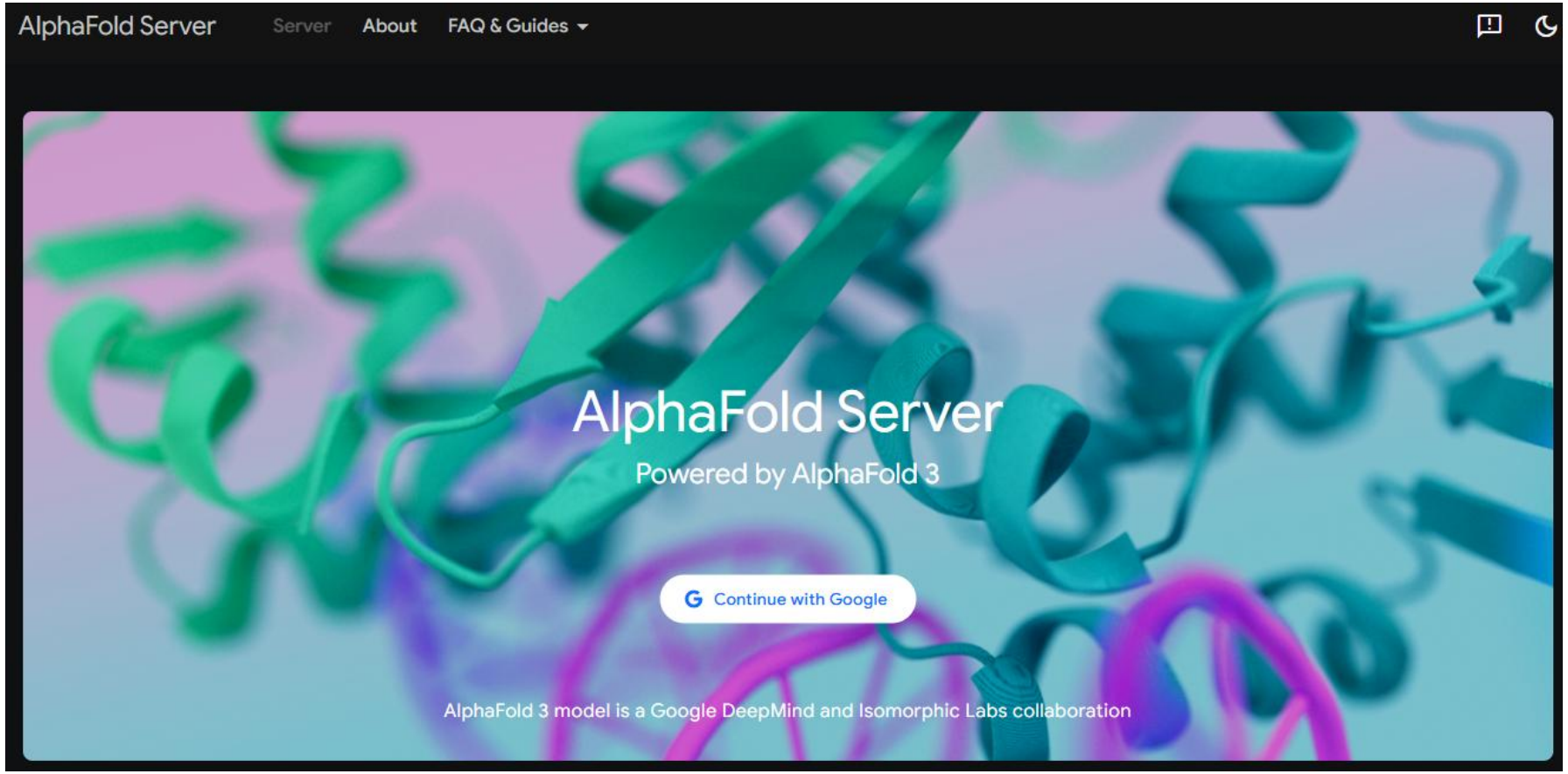
Potential structural determinants and additional scoring metrics associated with the observed thermal stability (T_m) values.

Scores	Description	Pearson's r	*p*-value
SMINA score (AF3-SMINA)	SMINA local minimization score for AF3 -predicted model	−0.814	0.002
Total Score (AF3-Rosetta)	Rosetta total score for relaxed AF3-pre dicted model	−0.687	0.014
omega (AF3-Rosetta)	Rosetta omega term for relaxed AF3-pr edicted model	−0.637	0.024
Normalized interface energy (AF3-Rosetta)	Rosetta interface energy for relaxed AF 3-predicted model	−0.583	0.038
Total Score (Rosetta, binding site fixed)	Rosetta total score for relaxed Rosetta model structure fixing binding site geo metry	−0.527	0.059
ProGen2 score (Sequence)	ProGen2 sequence likelihood score	0.508	0.067
fa_atr (AF3-Rosetta)	Rosetta Lennard-Jones attractive potent ial for relaxed AF3-predicted model	−0.445	0.099

Alphafold 3 사용하기

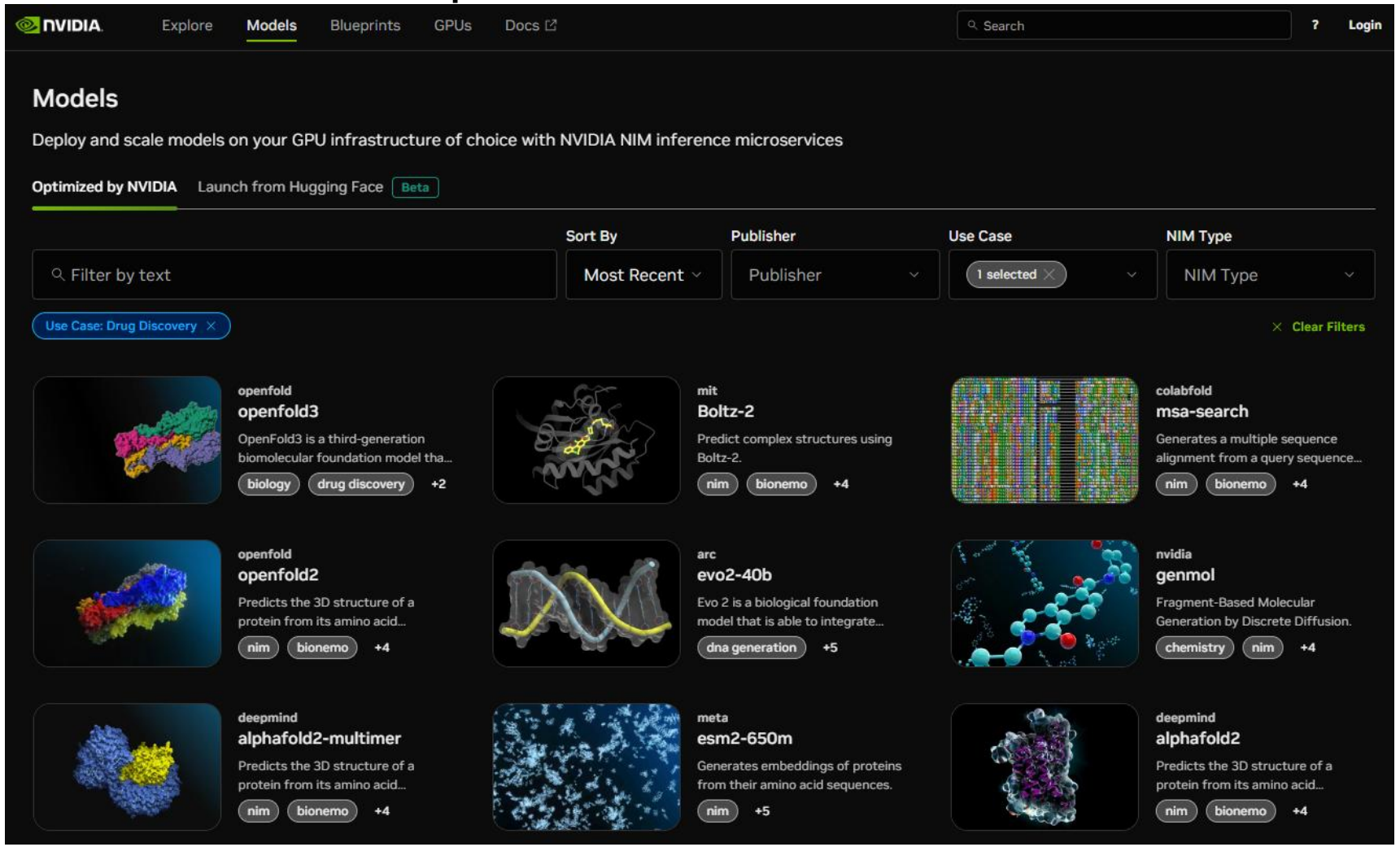
Cloud 환경에서 단백질 구조예측 사용하기

- Alphafold server : https://alphafoldserver.com/



Cloud 환경에서 단백질 구조예측 사용하기

- NVIDIA models : https://build.nvidia.com/models



Alphafold3 사용하기
- On-premises 환경

On-premises 환경에서 Alphafold3 사용하기

- Alphafold3 github URL
  - https://github.com/google-deepmind/alphafold3

- 권장사양
  - GPU : RTX 3090 or 4090 VRAM 8-12GB
  - CUDA : 12.6
  - Modern Multi-core CPU
  - System RAM : 32GB
  - HDD for DB : 630GB

가상화 환경

- **도커(Docker)**? - 컨테이너 기반 가상화 플랫폼
  - 응용 프로그램과 그 종속성을 격리된 환경인 **컨테이너**로 패키징하여 실행하는 기술
  - 응용 프로그램을 **서로 다른 환경**에서도 일관되게 실행
  - **가볍고 빠르며 확장성이** 좋아서 개발 및 **배포** 프로세스를 **간소화**하는 데 사용된다.

On-premises 환경에서 Alphafold3 사용하기

준비물

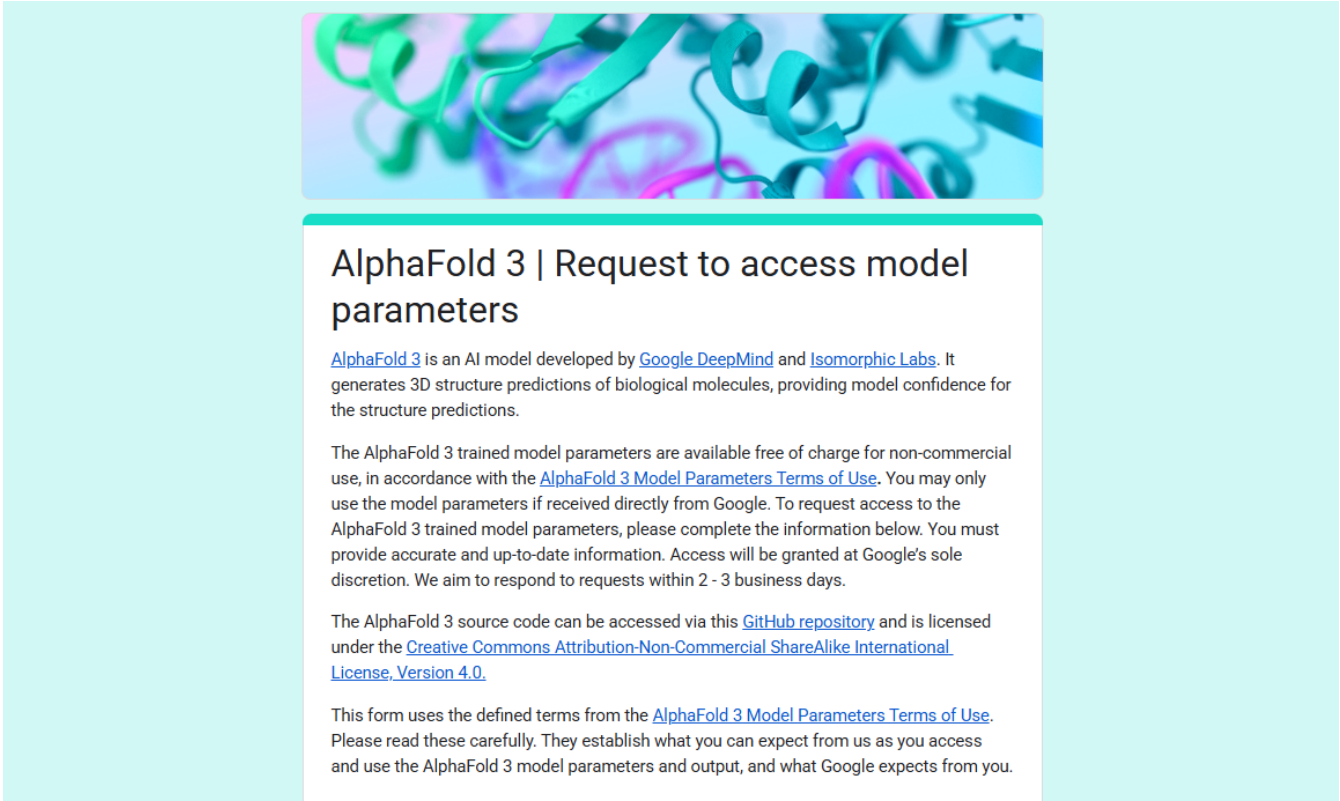
- Database
  - BFD, PDB, Mgnify, UniProt, UniRef, Rfam, NT, RNACentral 등
- Model parameter
  - Apply form link : https://forms.gle/svvpY4u2jsHEwWYS6
- Docker images
  - Alphafold3 실행환경과 알고리즘이 담겨있는 가상화 이미지
- JSON 파일
  - 서열, smile, PTM, MSA 등 예측을 위한 정보가 담겨있는 파일
- Docker command
  - DB 경로, Model parameter 경로, 작업 경로, 결과물 경로를 지정하여 Docker 환경에서 Alphafold3를 실행하는 명령어

On-premises 환경에서 Alphafold3 사용하기

- Database
  - Link : https://storage.googleapis.com/alphafold-databases/v3.0
  - BFD, PDB, Mgnify, UniProt, UniRef, Rfam, NT, RNACentral 등 데이터베이스
    - pdb_2022_09_28_mmcif_files.tar.zst
    - mgy_clusters_2022_05.fa
    - bfd-first_non_consensus_sequences.fasta
    - uniref90_2022_05.fa
    - uniprot_all_2021_04.fa
    - pdb_seqres_2022_09_28.fasta
    - rnacentral_active_seq_id_90_cov_80_linclust.fasta
    - nt_rna_2023_02_23_clust_seq_id_90_cov_80_rep_seq.fasta
    - rfam_14_9_clust_seq_id_90_cov_80_rep_seq.fasta
  - Database download
    - shell> bash fetch_databases.sh <download path>

On-premises 환경에서 Alphafold3 사용하기

- Model parameter
  - Alphafold3 model parameter
  - Apply form link : https://forms.gle/svvpY4u2jsHEwWYS6
  - Form 작성후 심사후 download



AlphaFold 3 | Request to access model parameters

AlphaFold 3 is an AI model developed by Google DeepMind and Isomorphic Labs. It generates 3D structure predictions of biological molecules, providing model confidence for the structure predictions.

The AlphaFold 3 trained model parameters are available free of charge for non-commercial use, in accordance with the AlphaFold 3 Model Parameters Terms of Use. You may only use the model parameters if received directly from Google. To request access to the AlphaFold 3 trained model parameters, please complete the information below. You must provide accurate and up-to-date information. Access will be granted at Google's sole discretion. We aim to respond to requests within 2 - 3 business days.

The AlphaFold 3 source code can be accessed via this GitHub repository and is licensed under the Creative Commons Attribution-Non-Commercial ShareAlike International License, Version 4.0.

This form uses the defined terms from the AlphaFold 3 Model Parameters Terms of Use. Please read these carefully. They establish what you can expect from us as you access and use the AlphaFold 3 model parameters and output, and what Google expects from you.

On-premises 환경에서 Alphafold3 사용하기

- JSON 파일
  - Sequence (Protein, DNA, RNA), Ligand, bonds 등 예측을 위한 정보가 담겨있는 파일
  - https://github.com/google-deepmind/alphafold3/blob/main/docs/input.md

The top-level structure of the input JSON is:

```
{
  "name": "Job name goes here",
  "modelSeeds": [1, 2],  # At least one seed required.
  "sequences": [
    {"protein": {...}},
    {"rna": {...}},
    {"dna": {...}},
    {"ligand": {...}}
  ],
  "bondedAtomPairs": [...],  # Optional.
  "userCCD": "...",  # Optional, mutually exclusive with userCCDPath.
  "userCCDPath": "...",  # Optional, mutually exclusive with userCCD.
  "dialect": "alphafold3",  # Required.
  "version": 4  # Required.
}
```

On-premises 환경에서 Alphafold3 사용하기

- Docker command
  - DB 경로, Model parameter 경로, 작업 경로, 결과물 경로를 지정하여 Docker 환경에서 Alphafold3를 실행하는 명령어

```
docker run -it \
    --volume $HOME/af_input:/root/af_input \
    --volume $HOME/af_output:/root/af_output \
    --volume <MODEL_PARAMETERS_DIR>:/root/models \
    --volume <DB_DIR>:/root/public_databases \
    --gpus all \
    alphafold3 \
    python run_alphafold.py \
    --json_path=/root/af_input/fold_input.json \
    --model_dir=/root/models \
    --output_dir=/root/af_output
```

Docker로 Alphafold3 환경 구성하기

- Docker로 Alphafold3 환경 구성하기
  - Docker image 생성
    - shell> git clone https://github.com/google-deepmind/alphafold3.git
    - shell> cd alphafold3
    - shell> docker build -t alphafold3 -f docker/Dockerfile .

  - Docker images pulling
    - shell> docker pull suppak/alphafold3:clean

Alphafold3 실행을 위한 json 파일 작성하기

- On-premises 환경에서 Alphafold3 사용하기
  - https://github.com/suppakoko/alphafold3_class/examples

- af3cli – json file 작성 도구
  - https://github.com/SLx64/af3cli
    - shell> git clone https://github.com/SLx64/af3cli.git

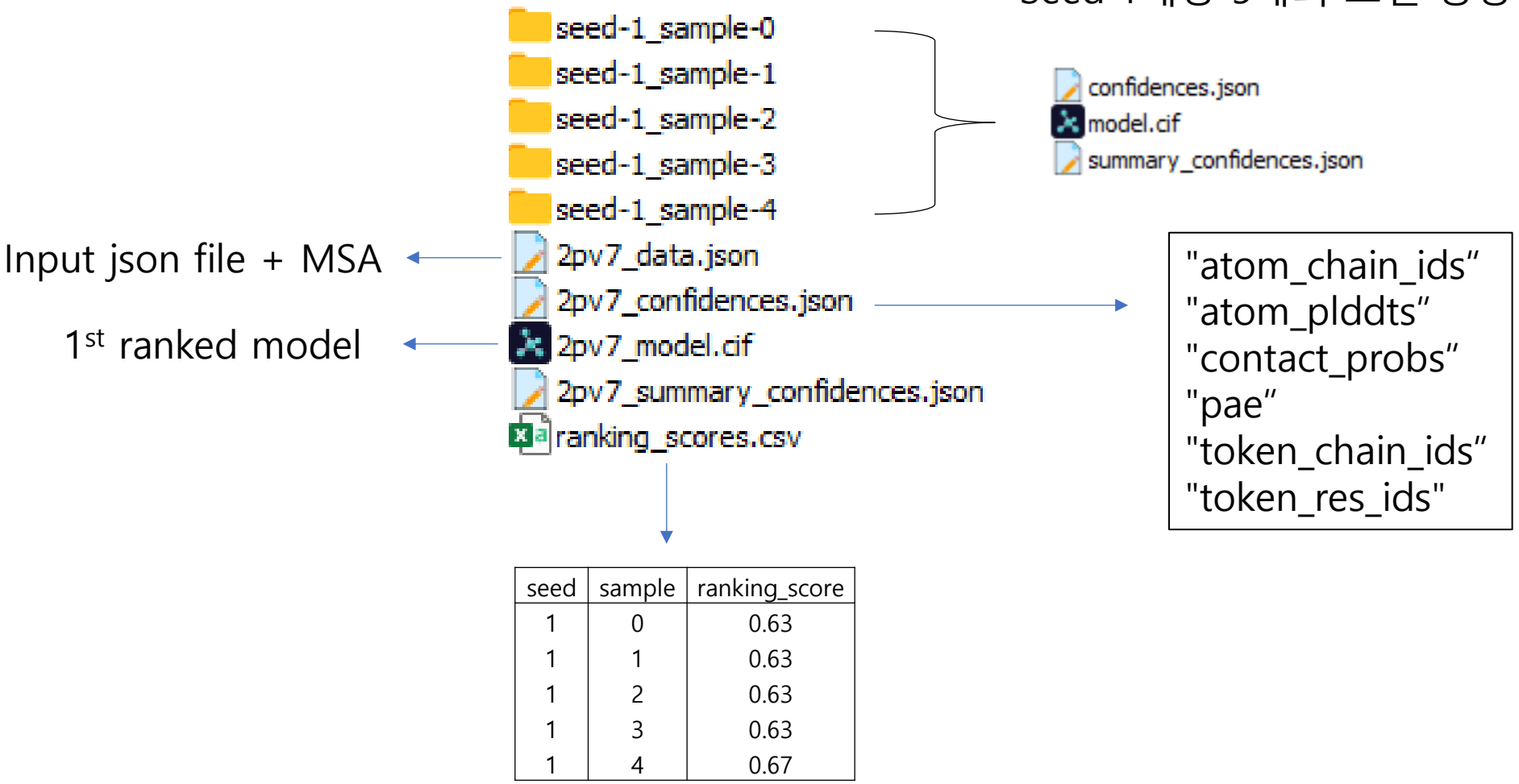
Alphafold3 실행을 위한 docker 명령어 작성하기

- On-premises 환경에서 Alphafold3 사용하기

  - https://github.com/suppakoko/alphafold3_class/blob/main/Docker/docker_command_example.txt

Alphafold3 실행 결과

- root/af_output/



seed	sample	ranking_score
1	0	0.63
1	1	0.63
1	2	0.63
1	3	0.63
1	4	0.67

Alphafold3 실행 결과

- summary_confidences.json

	A	B
A	0.76	1.04
B	1.03	0.76


```
"chain_iptm": [ 0.63, 0.63 ],
"chain_pair_iptm": [
 [ 0.68, 0.63 ],
 [ 0.63, 0.68 ]],
"chain_pair_pae_min": [
 [ 0.76, 1.04 ],
 [ 1.03, 0.76 ]],
"chain_ptm": [ 0.68, 0.68 ],
"fraction_disordered": 0.09,
"has_clash": 0.0,
"iptm": 0.63,
"ptm": 0.62,
"ranking_score": 0.67
```


pLDDT (Predicted Local Distance Difference Test)
"Is this loop or helix **shaped correctly**?"

> 90 : Very high confidence (side-chains are reliable).
70–90 : High confidence (backbone is reliable).
50–70 : Low confidence (caution needed).
< 50 : Very low confidence (likely intrinsically disordered or unstructured).

ipTM (Interface Predicted Template Modeling Score)
"Do these two proteins **actually bind** like this?"

> 0.8: High confidence (very likely a real binder with correct pose).
0.6–0.8: Moderate confidence (likely interacts, but exact orientation might vary).
< 0.5: Low confidence

PAE (Predicted Aligned Error)
"Are these **two domains actually sitting next to each other**, or are they floating apart?"

Scale: 0 to ~30 Å.
< 5 Å (Dark Blue/Green): The two residues are locked together rigidly
> 15 Å (White/Red): The relative position is uncertain

pTM (Predicted Template Modeling Score)
"Is the **overall shape** of this protein correct?"

> 0.7: High confidence (the fold is likely correct).
> 0.5: Correct topology (the general shape is right).
< 0.5: Failed prediction (random coil or incorrect fold).

Alphafold3를 이용한 예제

- On-premises 환경에서 Alphafold3 실제 사용하기

감사합니다

Q & A