情報科学IV 11

タンパク質の構造モデリング

清水謙多郎

shimizuk@fc.jwu.ac.jp

アクチンと熱ショックプロテイン

- アクチン
  - 細胞骨格の主要構成要素であり、細胞の形状や運動、分裂などに関与
- 熱ショックプロテイン（特にHSP70）
  - 細胞が高温や酸化ストレスなどにさらされた際に、変性したタンパク質の正しいフォールディングを助ける
  - どちらも、ATPが結合した状態 → ATP加水分解後のADP結合状態に変わることで、タンパク質の立体構造が変化し、機能が切り替わる
- 「配列は似ていないのに構造が似ている」典型的な例として挙げられる

DALIの検索

http://ekhidna2.biocenter.helsinki.fi/dali/

PDB search

Compare query structure against Protein Data Bank.

STEP 1 - Enter your query protein structure

Structures may be specified by concatenating the PDB identifier (4 characters) and a chain identifier (1 character) or, alternatively, you may upload a PDB file.

1atnA OR upload file 参照... ファイルが選択されていません。

PDB ID: アクチン1atnのチェインAを入力

STEP 2 - Optional data

You may leave an e-mail address for notification when the job has finished. The job title is used as subject heading in the e-mail.

actinA

ジョブ名

shimizu@bi.a.u-tokyo.ac.jp

メールアドレスを入力

STEP 3 - Submit your job

Submit Clear

「submit」をクリック

If the same structure has been submitted recently, you will be redirected to the result page of the previous instance.

前回、実行したときの画面

DALIの検索結果の例（1）

通知メール

Dali job finished: actin

LH Liisa Holm <luholm@sans-17.biocenter.helsinki.fi>
宛先 shimizu@bi.a.u-tokyo.ac.jp

Dali results in http://ekhidna2.biocenter.helsinki.fi/barcosel/tmp//1atnA/. Results are deleted after one week. This is an automatically gen

リンク先を表示

Results: 1atnA

Chain: 1atnA

- Matches against PDB25. Correlation matrix
- Matches against PDB50
- Matches against PDB90
- Matches against full PDB
- Download matches against PDB25
- Download matches against PDB50
- Download matches against PDB90
- Download matches against full PDB

Results will be deleted after one week.

DALIの検索結果の例（2）

Results: 1atnA

Chain: 1atnA

- Matches against PDB25. Correlation matrix
- Matches against PDB50
- Matches against PDB90
- Matches against full PDB
- Download matches against PDB25
- Download matches against PDB50
- Download matches against PDB90
- Download matches against full PDB

Results will be deleted after one week.

非常に多数のヒットがあるPDBの全チェインが対象
骨格筋ミオシン IIなどの分子が検索結果に表れているが、アクチンを複合体として含んでいるためヒットした

熱ショックプロテインは1000位以内ではヒットしない

Results: 1atnA

Query: 1atnA

MOLECULE: ACTIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimpo
Each neighbour has links to pairwise structural alignment with the query structure, and to the

Structural Alignment | Expand gaps | 3D Superimposition (PV) | SANS | PANZ | Pfam | Reset Selection

Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1atn-A	62.9	0.0	372	372	100	PDB	MOLECULE: ACTIN;
2:	1o1f-V	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
3:	1o1c-2	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
4:	2w49-O	62.0	0.1	372	372	100	PDB	MOLECULE: TROPONIN C, SKELETAL MUSCLE;
5:	1o1b-Y	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
6:	1m8q-W	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
7:	1o1b-8	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
8:	1o1c-3	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
9:	1o1g-6	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
10:	1o1c-9	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
11:	1o1e-4	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
12:	2w4u-D	62.0	0.1	372	372	100	PDB	MOLECULE: TROPONIN C, SKELETAL MUSCLE;
13:	1o19-6	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
14:	1o1b-1	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
15:	1o1b-7	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
16:	1o1f-5	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
17:	1o1d-3	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
18:	1o1e-3	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
19:	1m8q-1	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
20:	1o19-1	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;
21:	2w4u-J	62.0	0.1	372	372	100	PDB	MOLECULE: TROPONIN C, SKELETAL MUSCLE;
22:	2w49-G	62.0	0.1	372	372	100	PDB	MOLECULE: TROPONIN C, SKELETAL MUSCLE;
23:	1o1e-7	62.0	0.1	372	372	100	PDB	MOLECULE: SKELETAL MUSCLE MYOSIN II;

DALIの検索結果の例（3）

Results: 1atnA

Chain: 1atnA

- Matches against PDB25. Correlation matrix
- Matches against PDB50
- Matches against PDB90
- Matches against full PDB
- Download matches against PDB25
- Download matches against PDB50
- Download matches against PDB90
- Download matches against full PDB

Results will be deleted after one week.

PDB50は、50%配列一致度のカットオフで冗長性を取り除いたデータセット
類似の配列は削減される

熱ショックプロテインは23位と31位でヒット

Results: 1atnA

Query: 1atnA

MOLECULE: ACTIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The lis
Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format

Structural Alignment | Expand gaps | 3D Superimposition (PV) | SANS | PANZ | Pfam | Reset Selection

Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	4pl7-A	50.4	1.2	346	382	82	PDB	MOLECULE: ACTIN,THYMOSIN BETA-4;
2:	8e9b-B	44.2	3.2	365	387	48	PDB	MOLECULE: ACTIN-RELATED PROTEIN 3;
3:	9c57-L	42.4	2.0	354	404	38	PDB	MOLECULE: RUVB-LIKE 1;
4:	8a5o-W	41.6	2.3	358	431	31	PDB	MOLECULE: CHROMATIN-REMODELING ATPASE INO80;
5:	8a5q-W	41.3	2.2	354	433	32	PDB	MOLECULE: INO80 ATPASE;
6:	7zvw-G	40.1	2.4	359	422	27	PDB	MOLECULE: TRANSCRIPTION-ASSOCIATED PROTEIN;
7:	8oof-J	40.1	3.0	358	447	27	PDB	MOLECULE: CHROMATIN-REMODELING COMPLEX SUBUNIT IES6;
8:	6fml-J	38.1	3.0	357	503	27	PDB	MOLECULE: RUVB-LIKE HELICASE;
9:	8av6-J	38.1	3.0	357	647	27	PDB	MOLECULE: RUVB-LIKE HELICASE;
10:	5tgc-D	38.1	2.6	351	428	22	PDB	MOLECULE: ACTIN-RELATED PROTEIN 7;
11:	6gej-R	37.2	2.7	348	411	26	PDB	MOLECULE: VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN 72;
12:	4bql-A	37.2	2.2	346	436	18	PDB	MOLECULE: ACTIN/ACTIN FAMILY PROTEIN;
13:	7z8i-F	36.9	2.0	276	279	58	PDB	MOLECULE: ARP1 ACTIN RELATED PROTEIN 1 HOMOLOG A;
14:	6etx-H	36.5	1.9	317	326	31	PDB	MOLECULE: RUVB-LIKE 1;
15:	4fo0-A	36.3	2.3	357	502	22	PDB	MOLECULE: ACTIN-RELATED PROTEIN 8;
16:	6znl-J	35.6	3.0	329	379	29	PDB	MOLECULE: ARP1 ACTIN RELATED PROTEIN 1 HOMOLOG A;
17:	6igm-G	33.3	3.0	352	366	28	PDB	MOLECULE: RUVB-LIKE 1;
18:	4czm-A	31.6	3.2	308	336	12	PDB	MOLECULE: ROD SHAPE-DETERMINING PROTEIN MREB;
19:	7aqk-b	31.5	3.4	366	387	49	PDB	MOLECULE: ACTIN-RELATED PROTEIN 2, ARP2;
20:	3wee-A	28.7	3.9	331	428	18	PDB	MOLECULE: ACTIN-LIKE PROTEIN ARP9;
21:	8a5a-U	28.6	2.8	352	690	21	PDB	MOLECULE: CHROMATIN-REMODELING ATPASE INO80;
22:	7aqk-a	27.2	3.3	354	418	37	PDB	MOLECULE: ACTIN-RELATED PROTEIN 2, ARP2;
23:	3c7n-A	23.5	3.6	296	648	11	PDB	MOLECULE: HEAT SHOCK PROTEIN HOMOLOG SSE1;
24:	2fsk-B	22.7	3.0	273	324	15	PDB	MOLECULE: HYPOTHETICAL PROTEIN TA0583;
25:	7krw-A	22.6	3.5	290	608	15	PDB	MOLECULE: CHAPERONE PROTEIN DNAK FUSED WITH SUBSTRATE PEPTI
26:	5ec0-A	21.2	3.2	285	379	11	PDB	MOLECULE: ALP7A;
27:	6izr-1	21.2	3.9	277	349	13	PDB	MOLECULE: PUTATIVE PLASMID SEGREGATION PROTEIN PARM;
28:	4a2b-A	20.5	3.2	257	386	12	PDB	MOLECULE: CELL DIVISION PROTEIN FTSA, PUTATIVE;
29:	8x1i-B	20.2	3.8	276	369	13	PDB	MOLECULE: PARM PRESENT OF GENOME OF DESUFITOBACTERIUM HAFNI
30:	7q6g-A	20.0	4.2	269	388	12	PDB	MOLECULE: CELL DIVISION PROTEIN FTSA;
31:	4j8f-A	19.8	3.3	296	551	14	PDB	MOLECULE: HEAT SHOCK 70 KDA PROTEIN 1A/1B, HSC70-INTERACTIN
32:	3js6-A	19.2	3.9	265	321	14	PDB	MOLECULE: UNCHARACTERIZED PARM PROTEIN;

DALIの結果の例（4）

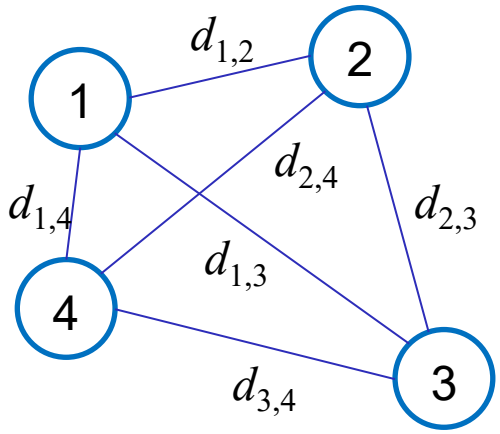
4j8f Aチェイン（HEAT SHOCK 70 KDA PROTEIN 1A/1B, HSC70-INTERACTIN ）
DALI PDB50で31位のヒット
配列一致度: 14% RMSD: 3.3Å 551残基中296残基でアラインメント

```
 No 31: Query=1atnA Sbjct=4j8fA Z-score=19.8
back to top

DSSP  --LLLLllLEEEEELLLEEEEEELL--LLLLLE------EEELLEEEEllleeellllLL
Query --DEDEttALVCDNGSGLVKAGFAG--DDAPRA------VFPSIVGRPrhqgvmvgmgQK   50
ident         |   | |      |          |        || |
Sbjct fmAKAA--AIGIDLGTTYSCVGVFQhgKVEIIAndqgnrTTPSYVAFT----------DT   48
DSSP  llLLLL--LEEEEELLLEEEEEEEllEEEELLllllllLEELLEEEL----------LL


DSSP  LLEELHHHHHLH--HHEEEE-LLEE-----------------------------------
Query DSYVGDEAQSKR--GILTLK-YPIE-----------------------------------   72
ident     || |
Sbjct ERLIGDAAKNQValNPQNTVfDAKRligrkfgdpvvqsdmkhwpfqvindgdkpkvqvsy  108
DSSP  LEEELHHHHHLHhhLHHHEElHHHHhlllllllhhhhhhhhllllleeeellleeeeeee


DSSP  ---lleelLHHHHHHHHHHHHHlLLLL--lhHHLLEEEEELLLLLHHHHHHHHHHHHHhHL
Query ---xgiitNWDDMEKIWHHTFYnELRV--apEEHPTLLTEAPLNPKANREKMTQIMFeTF  127
ident                                        |         |
Sbjct kgetkafyPEEISSMVLTKMKE-IAEAylgyPVTNAVITVPAYFNDSQRQATKDAGV-IA  166
DSSP  lleeeeelHHHHHHHHHHHHHH-HHHHhhllLLLEEEEEELLLLLHHHHHHHHHHHH-HL


DSSP  LLLeEEEEEHHHHHHHHLLLL-------LEEEEEELLLLEEEEEEEL----LEELhhHLE
Query NVPaMYVAIQAVLSLYASGRT-------TGIVLDSGDGVTHNVPIYE----GYALphAIM  176
ident                 | |              | | |                    |
Sbjct GLNvLRIINEPTAAAIAYGLDrtgkgerNVLIFDLGGGTFDVSILTIddgiFEVK--ATA  224
DSSP  LLEeEEEEEHHHHHHHHLLLLlllllleEEEEEELLLLEEEEEEEEelleEEEE--EEE


DSSP  EE-LLLHHHHHHHHHHHHHLLL----------lllLLHHHHHHHHHHHHHHLLllllhh
Query RL-DLAGRDLTDYLMKILTERG----------ysfVTTAEREIVRDIKEKLCYvaldfen  225
ident     | | |     |      |                     |       |   |
Sbjct GDtHLGGEDFDNRLVNHFVEEFkrkhkkdisqnkrAVRRLRTACERAKRTLSS-------  277
DSSP  EElLLLHHHHHHHHHHHHHHHHHhhhhlllllllhhHHHHHHHHHHHHHHHHHH-------
```

純粋に構造ベースの
アラインメント

DALIの手法

- 距離行列: タンパク質アミノ酸残基の間の距離を2次元行列で表す



	1	2	3	4
1	0	$d_{1,2}$	$d_{1,3}$	$d_{1,4}$
2	$d_{1,2}$	0	$d_{2,3}$	$d_{2,4}$
3	$d_{1,3}$	$d_{2,3}$	0	$d_{3,4}$
4	$d_{1,4}$	$d_{2,4}$	$d_{3,4}$	0

$L \times L$の距離行列

- 2つの構造の類似部分を$L \times L$の部分構造を単位に探す

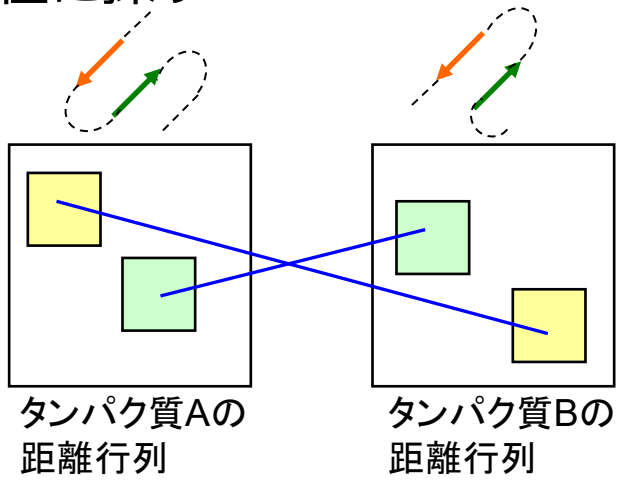
- $L \times L$の部分構造間の類似度スコア

$$S = \sum_{i=1}^{L}\sum_{j=1}^{L}\phi(i,j)$$

対応するアミノ酸残基間の
距離の和 ＝ 類似度

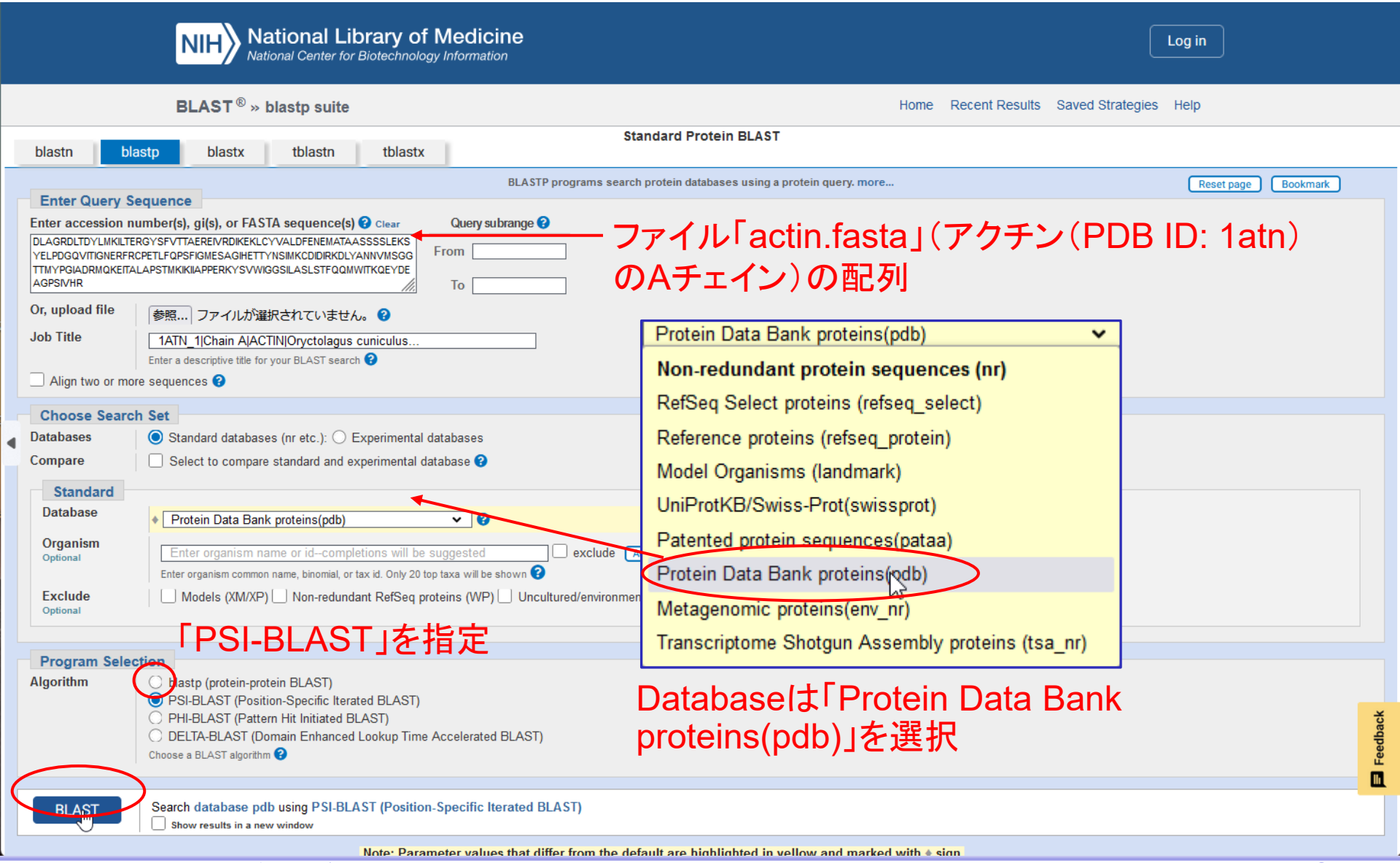
$$\phi^R(i,j) = \theta^R - \left|d_{ij}^A - d_{ij}^B\right|$$

$$\phi^E(i,j) = \begin{cases} \left(\theta^E - \dfrac{\left|d_{ij}^A - d_{ij}^B\right|}{d_{ij}^*}\right)\exp(-d_{ij}^*/\alpha^2) & i \neq j \\ \theta^E & i = j \end{cases}$$



- 類似度の高い部分構造を（ギャップを入れ）つなげてアラインメント

PSI-BLASTの利用（1）



ファイル「actin.fasta」（アクチン（PDB ID: 1atn）のAチェイン）の配列

「PSI-BLAST」を指定

Databaseは「Protein Data Bank proteins(pdb)」を選択

PSI-BLASTの利用（2）

繰り返し1回目の結果

Descriptions | Graphic Summary | Alignments | Taxonomy

Number of sequences 500 Run

さらに繰り返しをする場合は、「Run」をクリック

Sequences producing significant alignments Download Select columns Show 500

118 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Sequences with E-value BETTER than threshold

select all 117 sequences selected PSI-BLAST iteration 1

アクチンが多数検索される

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
ISOMETRICALLY CONTRACTING INSECT ASYNCHRONOUS FLIGHT MUSCLE [Oryctolagus cun...	Oryctolagus cu...	777	777	99%	0.0	100.00%	372	2W49_D	☑		
The X-Ray Crystal Structure Of The Complex Between Rabbit Skeletal Muscle Actin And Latruncul...	Oryctolagus cu...	776	776	99%	0.0	100.00%	377	1IJJ_A	☑		
gelsolin G4-G6/actin complex [Oryctolagus cuniculus]	Oryctolagus cu...	776	776	99%	0.0	100.00%	375	1H1V_A	☑		
Structure of the Phactr1 RPEL domain and RPEL motif directed assemblies with G-actin reveal the...	Oryctolagus cu...	776	776	99%	0.0	100.00%	376	4B1U_B	☑		
Structure of the Phactr1 RPEL-2 domain bound to actin [Oryctolagus cuniculus]	Oryctolagus cu...	775	775	99%	0.0	99.73%	376	4B1W_B	☑		
Crystal structure of actin in complex with lobophorolide [Oryctolagus cuniculus]	Oryctolagus cu...	775	775	99%	0.0	100.00%	371	3M6G_A	☑		
Chain A, Actin, alpha skeletal muscle [Mus musculus]	Mus musculus	774	774	99%	0.0	100.00%	374	7NEP_A	☑		
Structure of bare actin filament [Oryctolagus cuniculus]	Oryctolagus cu...	774	774	99%	0.0	100.00%	373	6BNO_A	☑		
Atomic structure of the actin:DNASE I complex [Oryctolagus cuniculus]	Oryctolagus cu...	773	773	99%	0.0	99.73%	373	1ATN_A	☑		
Chain A, ALPHA ACTIN [Oryctolagus cuniculus]	Oryctolagus cu...	773	773	99%	0.0	99.73%	377	1EQY_A	☑		
Cryo-EM structure of actin filament in the presence of phosphate [Oryctolagus cuniculus]	Oryctolagus cu...	773	773	99%	0.0	99.73%	376	3G37_O	☑		
Uncomplexed Actin [Oryctolagus cuniculus]	Oryctolagus cu...	773	773	99%	0.0	99.73%	375	1J6Z_A	☑		
Rigor myosin X co-complexed with an actin filament [Oryctolagus cuniculus]	Oryctolagus cu...	772	772	99%	0.0	99.73%	375	5KG8_B	☑		
Crystal structure of an actin monomer in complex with a chimeric peptide of Cordon-Bleu WH2 mu...	Oryctolagus cu...	771	771	99%	0.0	99.46%	377	6JH8_A	☑		
Mical-oxidized Actin complex with Gelsolin Segment 1 [Oryctolagus cuniculus]	Oryctolagus cu...	771	771	99%	0.0	99.46%	377	5UBO_A	☑		
Ternary Complex of Ac-Alpha-Actin with Profilin and Arp2/3 Complex [Oryctolagus cuniculus]	Oryctolagus cu...	771	771	99%	0.0	99.73%	375	6NAS_A	☑		
Chain A, Actin, alpha cardiac muscle 1 [Sus scrofa]	Sus scrofa	770	770	99%	0.0	98.92%	377	7JH7_A	☑		

E Value≧0.005の検索結果にも熱ショックタンパク質は登場しない

PSI-BLASTの利用（3）

Sequences with E-value BETTER than threshold

select all 163 sequences selected　Skip to the first new sequence　PSI-BLAST iteration 2

繰り返し2回目

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
Chain A, Actin-1 [Acanthamoeba castellanii]	Acanthamoeba castellanii	606	606	99%	0.0	92.45%	375	4EFH_A	✓	✓	
Chain A, Actin, alpha skeletal muscle [Oryctolagus cuniculus]	Oryctolagus cuniculus	606	606	99%	0.0	100.00%	377	7U8K_A	✓	✓	
Chain B, ACTIN, ALPHA SKELETAL MUSCLE [Oryctolagus cuniculus]	Oryctolagus cuniculus	606	606	99%	0.0	100.00%	376	4B1U_B	✓	✓	
Chain A, ACTIN, ALPHA SKELETAL MUSCLE [Oryctolagus cuniculus]	Oryctolagus cuniculus	606	606	99%	0.0	100.00%	377	1IJJ_A	✓	✓	
Chain R, Actin-related protein 5 [Saccharomyces cerevisiae S288C]	Saccharomyces cerevisiae S288C	209	336	98%	8e-61	24.90%	755	8EU9_R	✓	✓	
Chain H, Actin-related protein 5 [Homo sapiens]	Homo sapiens	187	326	97%	2e-53	26.27%	607	6HTS_H	✓	✓	
Chain B, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyces cerevisiae S288C	172	172	81%	7e-49	19.51%	439	4I6M_B	✓	✓	
Chain A, Actin-related protein [Homo sapiens]	Homo sapiens	164	222	94%	2e-44	17.11%	593	4FO0_A	✓	✓	
Chain B, Actin-like protein ARP9 [Saccharomyces cerevisiae]	Saccharomyces cerevisiae	157	157	81%	3e-43	17.98%	467	5TGC_B	✓		✓
Chain U, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyces cerevisiae S288C	157	157	81%	5e-43	17.98%	467	6TDA_U	✓		✓
Chain A, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyces cerevisiae S288C	151	151	81%	7e-41	17.98%	476	3WEE_A	✓		✓
Chain U, Actin-related protein 8 [Thermochaetoides thermophila]	Thermochaetoides thermophila	89.7	89.7	71%	7e-19	14.83%	747	8A5D_U	✓		✓
Chain A, ACTIN/ACTIN FAMILY PROTEIN [Pyrobaculum calidifontis JCM 1...	Pyrobaculum calidifontis JCM 11...	80.1	80.1	90%	8e-16	19.90%	432	4CJ7_A	✓		✓
Chain A, ACTIN/ACTIN FAMILY PROTEIN [Pyrobaculum calidifontis]	Pyrobaculum calidifontis	79.7	79.7	90%	1e-15	19.90%	455	4BQL_A	✓		✓
Chain A, Actin-like ATPase [Paramagnetospirillum magneticum AMB-1]	Paramagnetospirillum magnetic...	61.2	61.2	90%	8e-10	15.90%	347	5JYG_A	✓		✓
Chain A, Actin-like ATPase [Paramagnetospirillum magneticum AMB-1]	Paramagnetospirillum magnetic...	60.4	60.4	90%	2e-09	15.90%	347	5LJV_A	✓		✓
Chain A, Actin-like ATPase [Paramagnetospirillum magneticum AMB-1]	Paramagnetospirillum magnetic...	60.1	60.1	90%	2e-09	15.90%	347	5LJW_A	✓		✓
Chain C, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	45.4	45.4	6%	2e-06	95.65%	23	6ICV_C	✓	✓	
Chain Y, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	43.9	43.9	6%	7e-06	91.30%	23	6V63_Y	✓	✓	

黄色: 前の繰り返しで、しきい値以下のスコアだったもの

E Value ≥ 0.005の検索結果にも熱ショックタンパク質は登場しない

PSI-BLASTの利用（4）

Sequences with E-value BETTER than threshold

select all 221 sequences selected　Skip to the first new sequence　PSI-BLAST iteration 3

繰り返し3回目

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
Chain A, Actin-1 [Acanthamoeba castellanii]	Acanthamoeb...	590	590	99%	0.0	92.45%	375	4EFH_A	✓	✓	
Chain A, cardiac actin [Sus scrofa]	Sus scrofa	584	584	99%	0.0	98.92%	375	7TIJ_A	✓	✓	
Chain B, Actin-like protein ARP9 [Saccharomyces cerevisiae]	Saccharomyc...	180	180	82%	8e-52	18.11%	467	5TGC_B	✓	✓	
Chain U, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyc...	179	179	82%	3e-51	18.11%	467	6TDA_U	✓	✓	
Chain A, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyc...	173	173	82%	5e-49	18.11%	476	3WEE_A	✓	✓	
Chain M, Cell shape-determining protein MreB [Geobacillus stearothermophilus]	Geobacillus s...	65.4	65.4	92%	4e-11	16.38%	340	7ZPT_M	✓		✓
Chain M, Cell shape-determining protein MreB [Geobacillus stearothermophilus ATCC 7953]	Geobacillus s...	65.4	65.4	92%	4e-11	16.38%	347	8AZG_M	✓		✓
Chain C, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	45.8	45.8	6%	1e-06	95.65%	23	6ICV_C	✓	✓	
Chain Y, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	43.8	43.8	6%	6e-06	91.30%	23	6V63_Y	✓	✓	
Chain Y, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	43.5	43.5	6%	8e-06	91.30%	23	6V62_Y	✓	✓	
Chain E, Actin, cytoplasmic 1 [Homo sapiens]	Homo sapiens	43.5	43.5	6%	1e-05	91.30%	23	6ICT_E	✓	✓	
Chain Y, Actin, cytoplasmic 2 [Homo sapiens]	Homo sapiens	42.7	42.7	6%	2e-05	91.30%	23	6WK1_Y	✓	✓	
Chain A, Heat shock 70 kDa protein 1-like [Homo sapiens]	Homo sapiens	47.3	47.3	71%	3e-05	15.96%	408	3GDQ_A	✓		✓
Chain A, HEAT SHOCK COGNATE 71 KDA PROTEIN [Homo sapiens]	Homo sapiens	44.2	44.2	50%	3e-04	16.42%	386	5AQM_A	✓		✓
Chain A, HEAT SHOCK-RELATED 70KDA PROTEIN 2 [Homo sapiens]	Homo sapiens	43.5	43.5	57%	5e-04	17.47%	387	5FPD_A	✓		✓
Chain A, Heat shock-related 70 kDa protein 2 [Homo sapiens]	Homo sapiens	43.5	43.5	57%	5e-04	17.47%	404	3I33_A	✓		✓
Chain A, HEAT-SHOCK COGNATE 70 kD PROTEIN [Bos taurus]	Bos taurus	43.5	43.5	74%	5e-04	15.81%	386	1NGA_A	✓		✓
Chain A, HEAT-SHOCK COGNATE 70 kD PROTEIN [Bos taurus]	Bos taurus	43.5	43.5	50%	5e-04	16.92%	386	1NGB_A	✓		✓
Chain A, HEAT-SHOCK COGNATE 70 KD PROTEIN [Bos taurus]	Bos taurus	43.5	43.5	50%	5e-04	16.92%	386	1ATR_A	✓		✓
Chain A, HEAT SHOCK-RELATED 70 KDA PROTEIN 2 [Homo sapiens]	Homo sapiens	43.5	43.5	57%	6e-04	17.47%	640	5FPN_A	✓		✓
Chain B, Actin, gamma-enteric smooth muscle [Homo sapiens]	Homo sapiens	38.5	38.5	5%	6e-04	100.00%	22	6JAT_B	✓	✓	

黄色：前の繰り返しで、しきい値以下のスコアだったもの

Mreb: 原核生物の細胞骨格形成タンパク質（アクチン様タンパク質）

多数の熱ショックタンパク質がヒット

PSI-BLASTの利用（5）

Sequences with E-value BETTER than threshold

select all 259 sequences selected　Skip to the first new sequence　PSI-BLAST iteration 4

繰り返し4回目

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
Chain A, Actin-1 [Acanthamoeba castellanii]	Acanthamoeb...	571	571	99%	0.0	92.45%	375	4EFH_A	✓	✓	
Chain A, Actin, alpha cardiac muscle 1 [Sus scrofa]	Sus scrofa	565	565	99%	0.0	98.92%	377	7JH7_A	✓	✓	
Chain A, HEAT SHOCK 70 KDA PROTEIN 1A [Homo sapiens]	Homo sapiens	187	187	83%	2e-55	14.29%	394	5AQW_A	✓		✓
Chain A, Heat shock 70 kDa protein 1 [Homo sapiens]	Homo sapiens	187	187	83%	3e-55	14.29%	409	3JXU_A	✓		✓
Chain A, Heat shock 70 kDa protein 1A [Homo sapiens]	Homo sapiens	187	187	83%	4e-55	14.29%	388	5BN9_A	✓		✓
Chain A, Heat shock 70kDa protein 1A variant [Homo sapiens]	Homo sapiens	186	186	83%	6e-55	14.29%	378	6ZYI_A	✓		✓
Chain A, Heat shock 70kDa protein 1A variant [Homo sapiens]	Homo sapiens	186	186	83%	6e-55	14.29%	382	4IO8_A	✓		✓
Chain R, Actin-related protein 5 [Saccharomyces cerevisiae S288C]	Saccharomyc...	193	331	98%	8e-55	24.90%	755	8EU9_R	✓	✓	
Chain A, Heat shock 70 kDa protein 1B [Homo sapiens]	Homo sapiens	186	186	83%	9e-55	14.29%	386	7F4Z_A	✓		✓
Chain R, Actin-related protein 5 [Saccharomyces cerevisiae S288C]	Saccharomyc...	193	331	98%	9e-55	24.90%	744	8ETS_R	✓	✓	
Chain A, Heat shock 70 kDa protein 1A [Homo sapiens]	Homo sapiens	186	186	83%	1e-54	14.29%	388	5BPM_A	✓		✓
Chain A, HEAT SHOCK-RELATED 70 KDA PROTEIN 2 [Homo sapiens]	Homo sapiens	190	190	83%	3e-54	15.20%	640	5FPN_A	✓	✓	
Chain A, Heat shock-related 70 kDa protein 2 [Homo sapiens]	Homo sapiens	184	184	83%	3e-54	14.59%	387	4FSV_A	✓	✓	
Chain B, Heat shock 70 kDa protein 1 [Homo sapiens]	Homo sapiens	184	184	83%	5e-54	13.68%	382	3D2E_B	✓		✓
Chain A, Heat shock 70 kDa protein 1A/1B, Hsc70-interacting protein [synthetic construct]	synthetic cons...	187	187	83%	1e-53	14.29%	561	4J8F_A	✓		✓
Chain A, ACTIN/ACTIN FAMILY PROTEIN [Pyrobaculum calidifontis JCM 11548]	Pyrobaculum ...	184	184	90%	1e-53	19.41%	432	4CJ7_A	✓	✓	
Chain A, ACTIN/ACTIN FAMILY PROTEIN [Pyrobaculum calidifontis]	Pyrobaculum ...	184	184	90%	3e-53	19.41%	455	4BQL_A	✓	✓	
Chain C, Heat shock 70 kDa protein 1A [Homo sapiens]	Homo sapiens	186	186	83%	7e-53	14.29%	641	7KW7_C	✓		✓
Chain B, Actin-like protein ARP9 [Saccharomyces cerevisiae S288C]	Saccharomyc...	182	182	82%	1e-52	19.64%	439	4I6M_B	✓	✓	
Chain A, Heat shock 70 kDa protein 6 [Homo sapiens]	Homo sapiens	177	177	83%	3e-51	13.41%	403	3FE1_A	✓		✓

さらに多数の熱ショックタンパク質がヒット

PDB ID: 4j8fが新たにヒット

PSI-BLASTの結果の例

4j8f Aチェイン（Mreb From Thermotoga Maritima）
PSI-BLASTの4回目の繰り返しで登場

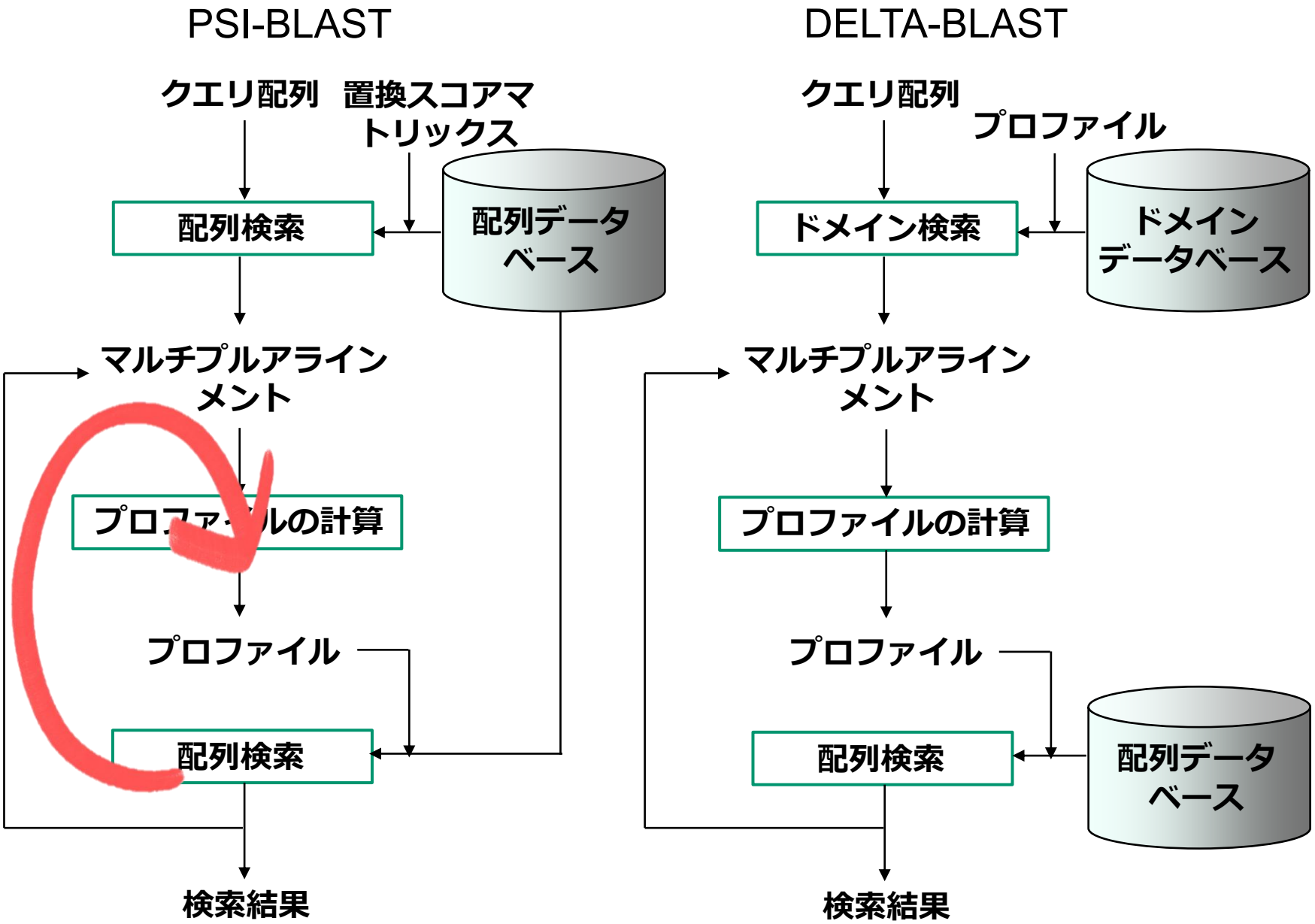
```
Chain A, Heat shock 70 kDa protein 1A/1B, Hsc70-interacting protein [synthetic construct]
Sequence ID: 4J8F_ALength: 561Number of Matches: 1
Related Information
Structure-3D structure displays
Range 1: 84 to 386GenPeptGraphics
Next Match
Previous Match
Alignment statistics for match #1 Score     Expect     Method     Identities Positives  Gaps
                         187 bits(474) 1e-53 Composition-based stats. 47/329(14%) 102/329(31%)
42/329(12%)

Query  38   RPRHQGVMVGMGQKDSYVGDEAQSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRV  97
            +     V   M      V ++    +  ++ K   +          M      T   E+
Sbjct  84   KFGDPVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEISSMVL----TKMKEIAE  139

Query  98   APEEHPT--LLTEAPLNPKANREKMTQIMFETFNVPAMYVAIQAVLSLYASGRT------  149
            A   +P    +   P     ++ + T+       +  + +  +   +  A G
Sbjct  140  AYLGYPVTNAVITVPAYFNDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGE  199

Query  150  -TGIVLDSGDGV--THNVPIYEG-YALPHAIMRLDLAGRDLTDYLMKILTE---RGYSFV  202
               ++ D G G      + I +G + +        L G D  + L+    E   R +
Sbjct  200  RNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHKKD  259

Query  203  TTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFR-CPET  261
             +  +  VR ++           +      +S  ++  +E  D    +I   RF
Sbjct  260  ISQNKRAVRRLRTACERAKRTLSSS---TQASLEIDSLFEGIDFYT-SITRARFEELCSD  315

Query  262  LFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNVMSGGTTMYPGIADRMQKEITAL  321
            LF+ +          E    ++    +D  K    + V+ GG+T  P +   +Q
Sbjct  316  LFRSTL---------EPVEKALRDAKLD--KAQIHDLVLVGGSTRIPKVQKLLQDFFNGR  364

Query  322  APSTMKIKIIAPPERKYSVWIGGSILASL  350
                  +     P+   +V  G ++ A++
Sbjct  365  -----DLNKSINPDE--AVAYGAAVQAAI  386
```

PSI-BLASTとは

- **PSI-BLAST**:
  Position-Specific Iterated BLAST
  位置依存スコア　　繰り返し

- タンパク質配列のホモロジー検索ツールで、より高感度に関係する配列を検索できるようにBLASTを拡張して設計された
  - 遠縁の配列を高感度に見つけだす
  - 構造データベースに対して検索を実行すれば、類似構造をもつタンパク質の配列が検索される

PSI-BLASTとDELTA-BLAST



タンパク質の構造モデリング

- タンパク質の構造は重要
  - 配列よりも構造の方がより多くの情報をもつ
  - 創薬や病気の治療、酵素や機能性食品の開発などに役立つ
- 実験でタンパク質の構造を決定するには時間と労力を必要とする
- そのため、コンピュータによるモデリングが期待されている

AIを用いたタンパク質の構造モデリングやタンパク質のデザインが利用されている

AIを使った構造モデリング

- AlphaFold2
  - Googleの関連企業であるDeepMind社による
  - タンパク質の構造モデリングで画期的な性能
  - オープンソース（プログラムを公開）
- AlphaFold2で予測した2億個以上のタンパク質の構造をデータベースで公開
  - https://alphafold.ebi.ac.uk/
- 最新版はAlphaFold3
  - 拡散モデルを使用
  - タンパク質とDNA、リガンドとの相互作用の予測
- 簡易版のGoogle ColabFold（AlphaFold2ベース）は、Googleのアカウントがあれば簡単に利用できる
  - https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb

https://www.nobelprize.org/prizes/chemistry/2024/popular-information/

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

The Nobel Prize in Chemistry 2024

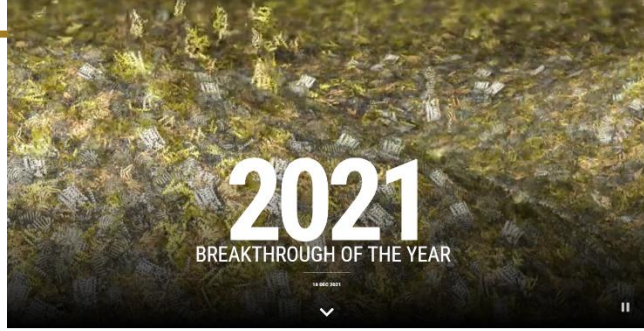
Chemists have long dreamed of fully understanding and mastering the chemical tools of life – proteins. This dream is now within reach. **Demis Hassabis** and **John Jumper** have successfully utilised artificial intelligence to predict the structure of almost all known proteins. **David Baker** has learned how to master life's building blocks and create entirely new proteins. The potential of their discoveries is enormous.

They have revealed proteins' secrets through computing and artificial intelligence

How is the exuberant chemistry of life possible? The answer to this question is the existence of proteins, which can be described as brilliant chemical tools. They are generally built from 20 amino acids that can be combined in endless ways. Using the information stored in DNA as a blueprint, the amino acids are linked together in our cells to form long strings.

Then the magic of proteins happens: the string of amino acids twists and folds into a distinct – sometimes unique – three-dimensional structure (Figure 1). This structure is what gives proteins their function. Some become chemical building blocks that can create muscles, horns or feathers, while others may become hormones or antibodies. Many of them form enzymes, which drive life's chemical reactions with astounding precision. The proteins that sit on the surfaces of cells are also important, and function as communication channels between the cell and its surroundings.

2021 BREAKTHROUGH OF THE YEAR



Protein structures for all

AI-powered predictions show proteins finding their shapes

BY ROBERT SERVICE

Science 2021-12-16

In his 1972 Nobel Prize acceptance speech, American biochemist Christian Anfinsen laid out a vision: One day it would be possible, he said, to predict the 3D structure of any protein merely from its sequence of amino acid building blocks. With hundreds of thousands of proteins in the human body alone, such an advance would have vast applications, offering insights into basic biology and revealing promising new drug targets. Now, after nearly 50 years, researchers have shown that artificial intelligence (AI)-driven software can churn out accurate protein structures by the thousands—an advance that realizes Anfinsen's dream and is *Science*'s 2021 Breakthrough of the Year.

Protein structures could once be determined only through painstaking lab analyses. But they can now be calculated, quickly, for tens of thousands of proteins, and for complexes of interacting proteins. "This is a sea change for structural biology," says Gaetano Montelione, a structural biologist at Rensselaer Polytechnic Institute. David Baker, a University of Washington, Seattle, computational biochemist who led one of the prediction projects, adds that with the bounty of readily available structures, "All areas of computational and molecular biology will be transformed."

Proteins are biology's workhorses. They contract our muscles, convert food into cellular energy, ferry oxygen in our blood, and fight microbial invaders. Yet despite their varied talents, all proteins start out with the same basic form: a linear chain of up to 20 different kinds of amino acids, strung together in a sequence encoded in our DNA. After being assembled in cellular factories called ribosomes, each chain folds into a unique, exquisitely complex 3D shape. Those shapes, which determine how proteins interact

ホモロジーモデリングによる予測

ターゲットのアミノ酸配列

```
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAGKPRWDWQPEPVN...
```

↓

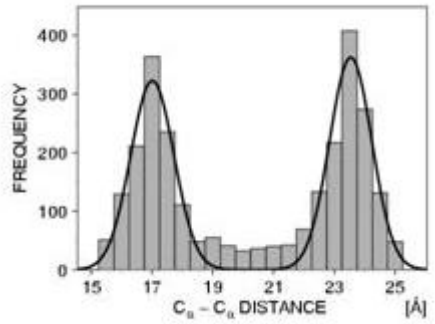
ターゲットとテンプレートのアラインメント

```
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAGKPRWDWQPEPVN...
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDTMLGAVVFGHEQQQVVIQAINDLVKEAGKPRWDWQPEAVN...
...QYVLNPTQEELKSSKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQIVIQNINDLVKEAGKPRWDWQPEAVN...
```

配列類似性をもとに、構造既知のテンプレートを探索

↓

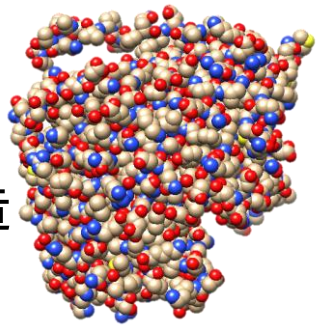
構造特徴の抽出
Cα原子間距離、二面角の分布など
二次構造、溶媒露出度など



テンプレートの構造群およびデータベースの既知の構造から拘束条件を求め、ターゲットに適用、
様々な拘束条件を合わせて、最小化

モデル構造の構築

→ 3D構造



コンタクトマップを用いた予測

ターゲットのアミノ酸配列

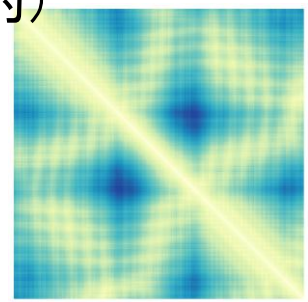
```
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAGKPRWDWQPEPVN...
```

高感度の配列類似性探索

ターゲットの類似配列の多数のアラインメント

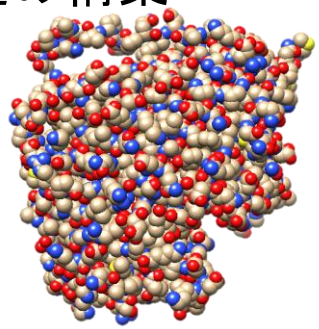
```
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAGKPRWDWQPEPVN...
...QYVLNPTQDELKESKLDLVVAGTEAAVLMVESEAELLSEDTMLGAVVFGHEQQQVVIQAINDLVKEAGKPRWDWQPEAVN...
...QYVLNPTQEELKSSKLDLVVAGTEAAVLMVESEAELLSEDQMLGAVVFGHEQQQIVIQNINDLVKEAGKPRWDWQPEAVN...
...QYVLNPTSDELKESKLDLVVAGTKGAVLMVESEAELLSEDQMLGAVVFGHDQQQIVIDNINALVAEAGKPRWDWQPEAVN...
...QYVLNPTTDELKESRLDLVVAGTAGAVLMVESEADILSEEQMLGAVVFGHEQQQVVIENINALVAEAGKPKWDWQAEPVN...
...QYVLNPTVDELKISKLDLVVAGTAGAVLMVESEADLLSEEQMLGAVVFGHEQQQVVIENINALVAEVGKEKWDWAPEPIN...
...QLVLNPSEKELKQSRLDLVVAGTDNAVLMVESEAQILTEEEMLAAVVFGHDQQQAVIKAINEFAAEVATPAWEWVAPAEN...
...QYILNPNVNEIKSSSLDLIISGTEESVLMVEAEANMLTEEQIINAINYGHEQQKIVIKNIEHFAKKVKIPVWEQCLYPVN...
...EYLLNPSLDELKDSALDLVVAGTRDAVLMVESEAQELPESVMLGAVLHGHQAMQVAIQAIAEFIQEAGGAKWEWEPPTVN...
.....
.....
```

コンタクトマップ（残基間距離制約）



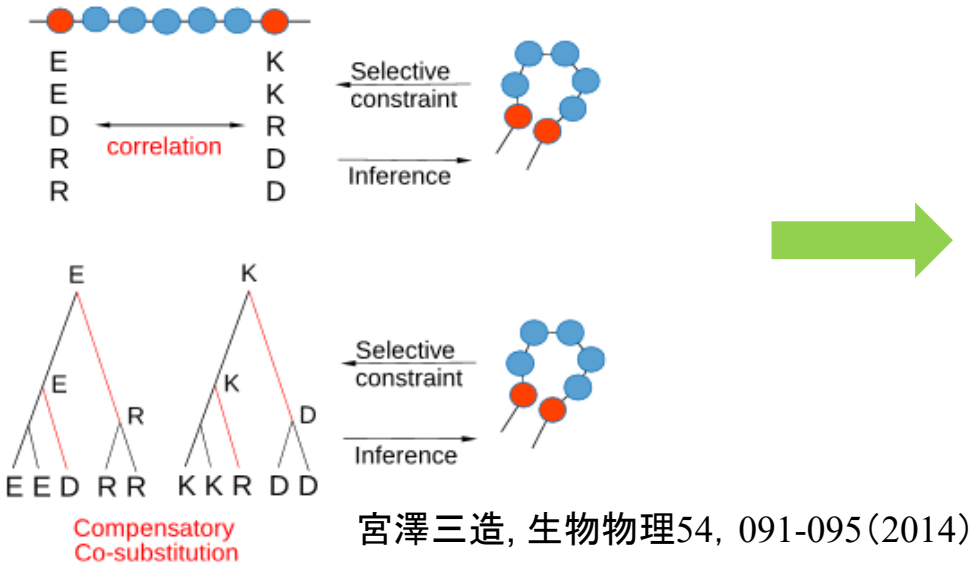
モデル構造の構築

3D構造



共進化情報

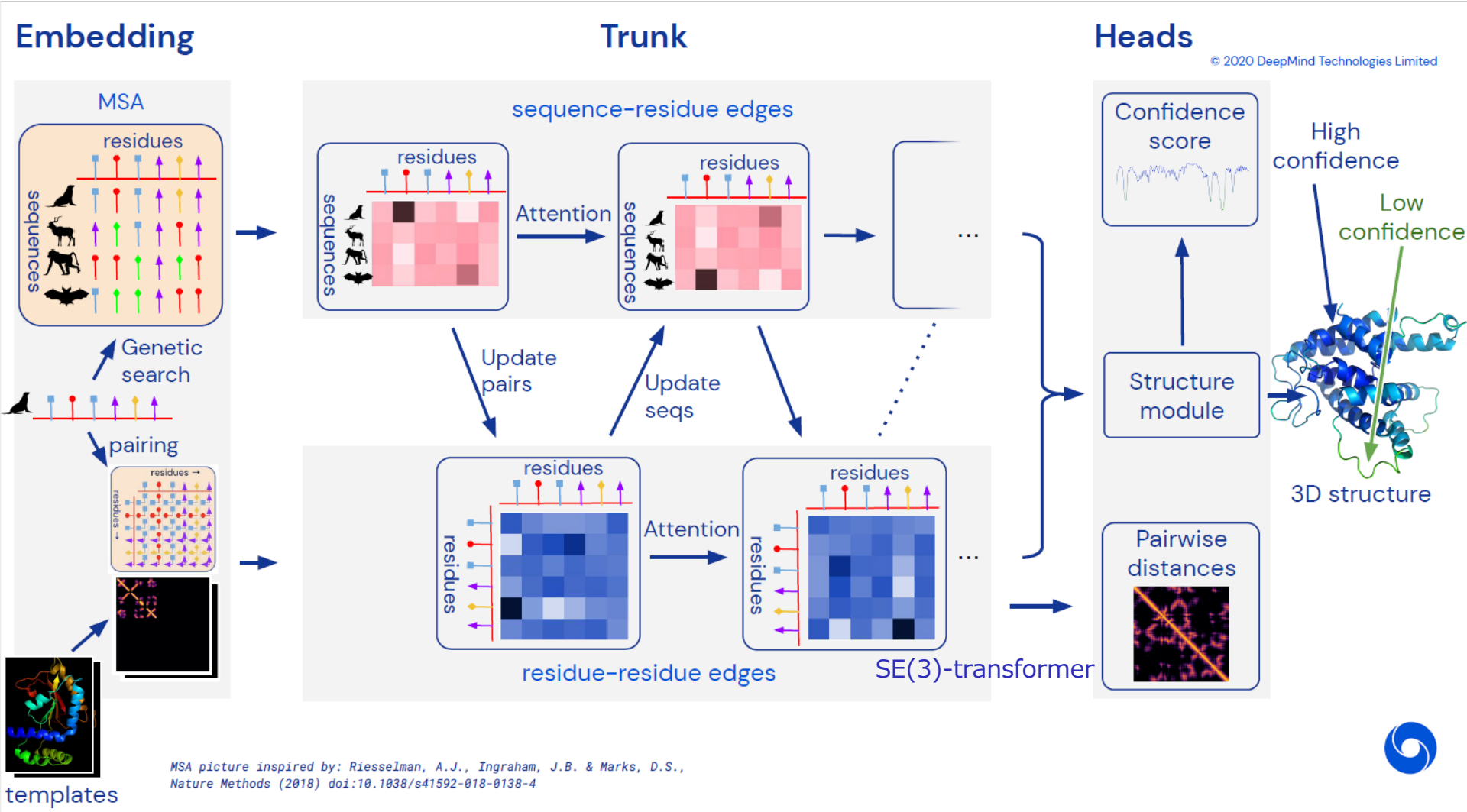
- 多数のタンパク質の配列のマルチプルアラインメント（MSA, Multiple Sequence Alignment）
  - 進化の間で行われてきたアミノ酸置換の知識の集積
  - 共変異: タンパク質を構成するアミノ酸残基のうち、複数の位置のアミノ酸がともに置換する現象
    - 一般的には相互作用によりともに進化したということで共進化とも呼ばれる
  - 立体構造において近接する残基間の相互作用に起因 → 共変異をもとにコンタクト予測を行う



深層学習により特徴量を抽出
コンタクト予測を行う

宮澤三造, 生物物理54, 091-095(2014)

AlphaFold2の手法



https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

ColabFoldによる構造予測（1）

- ColabFoldのページにアクセス
- Web検索で「ColabFold」で検索してもよい

AlphaFold2.ipynb

ファイル　編集　表示　挿入　ランタイム　ツール　ヘルプ

共有

＋ コード　＋ テキスト　ドライブにコピー　接続済み　Gemini

ColabFold v1.5.5: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.4, v1.5.1, v1.5.2, v1.5.3-patch

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. *Nature Methods*, 2022

Input protein sequence(s), then hit Runtime -> Run all

[1]

query_sequence: PIAQIHILEGRSDEQKETLIREVSEAISRSLDAPLTSVRVIITEMAKGHFGIGGELASK

- Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: test

num_relax: 0

- specify how many of the top ranked structures to relax using amber

template_mode: none

- none = no template information is used. pdb100 = detect templates in pdb100 (see notes). custom - upload and search own templates (PDB or mmCIF format, see

ColabFoldによる構造予測（2）

AlphaFold2.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 変更は保存されませんでした

共有

+ コード + テキスト ドライブにコピー

接続 T4 ハイメモリ Gemini

ColabFold v1.5.5: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using AlphaFold2 and Alphafold2-multimer. Sequence alignments/templates are generated through MMseqs2 and HHsearch. For more details, see bottom of the notebook, checkout the ColabFold GitHub and read our manuscript. Old versions: v1.4, v1.5.1, v1.5.2, v1.5.3-patch

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. *Nature Methods*, 2022

配列をそのまま入力（FASTA形式などではない）
複合体を予測するときは、配列A:配列Bのように「:」で区切って指定

ターゲットの配列を入力（配列の部分だけ入力する）
flu-na.fasta

Input protein sequence(s), then hit Runtime -> Run all

query_sequence: ¡FEMIWDPNGWTETDSSFSVRQDIVAITDWSGY SGSFVQHPELTGLDCMRPCFWVELIRGQPKENTIWTSGSSISFCGVNSDTVGWSWPDGAELPFSI

- Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: flu-na

ジョブ名を入力（ここでは、「flu-na」とした）
結果のファイルの名前などに反映される

num_relax: 0

- specify how many of the top ranked structures to relax using amber

template_mode: none

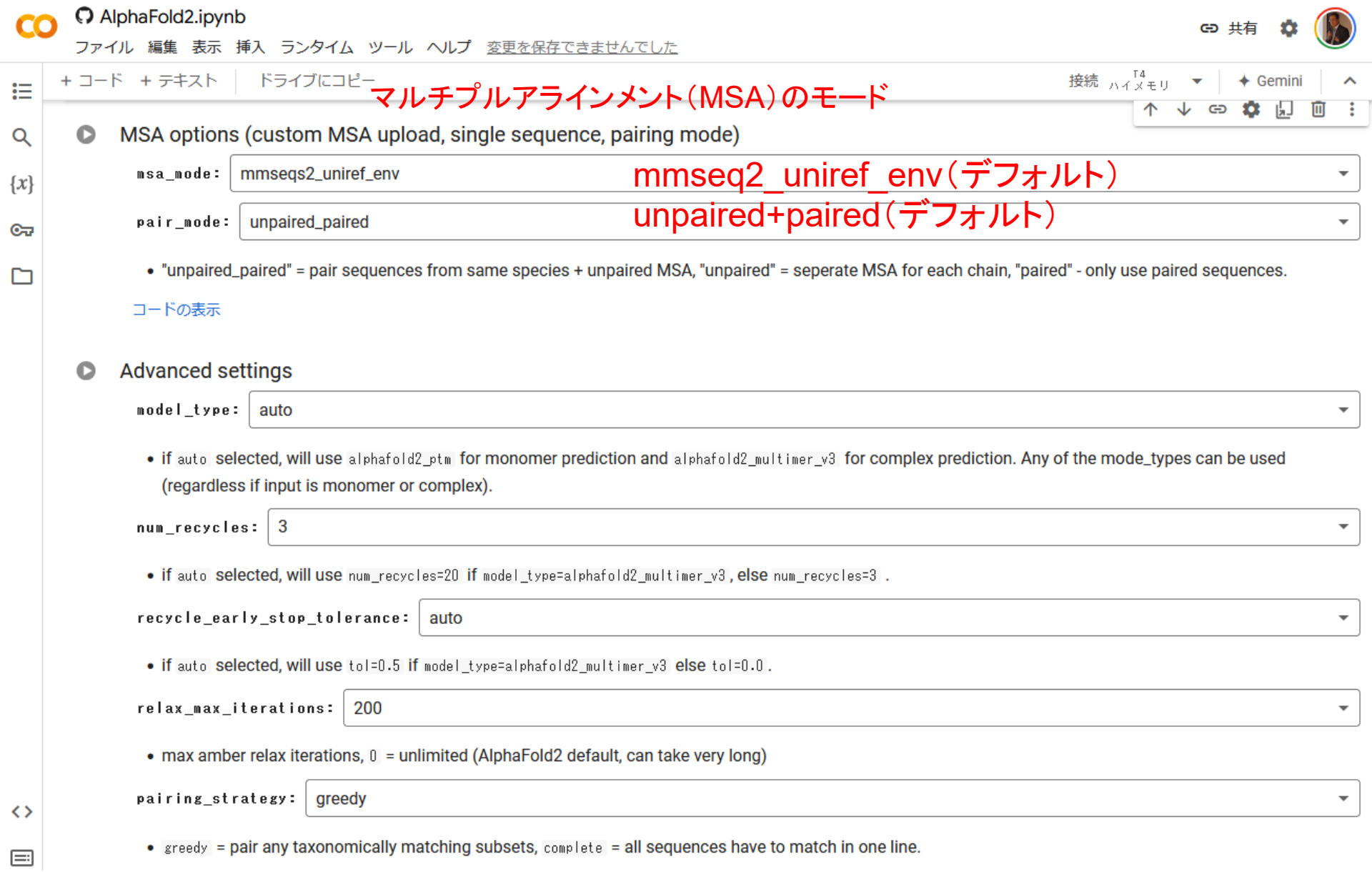
- none = no template information is used. pdb100 = detect templates in pdb100 (see notes). custom - upload and search own templates (PDB or mmCIF format, see notes)

コードの表示

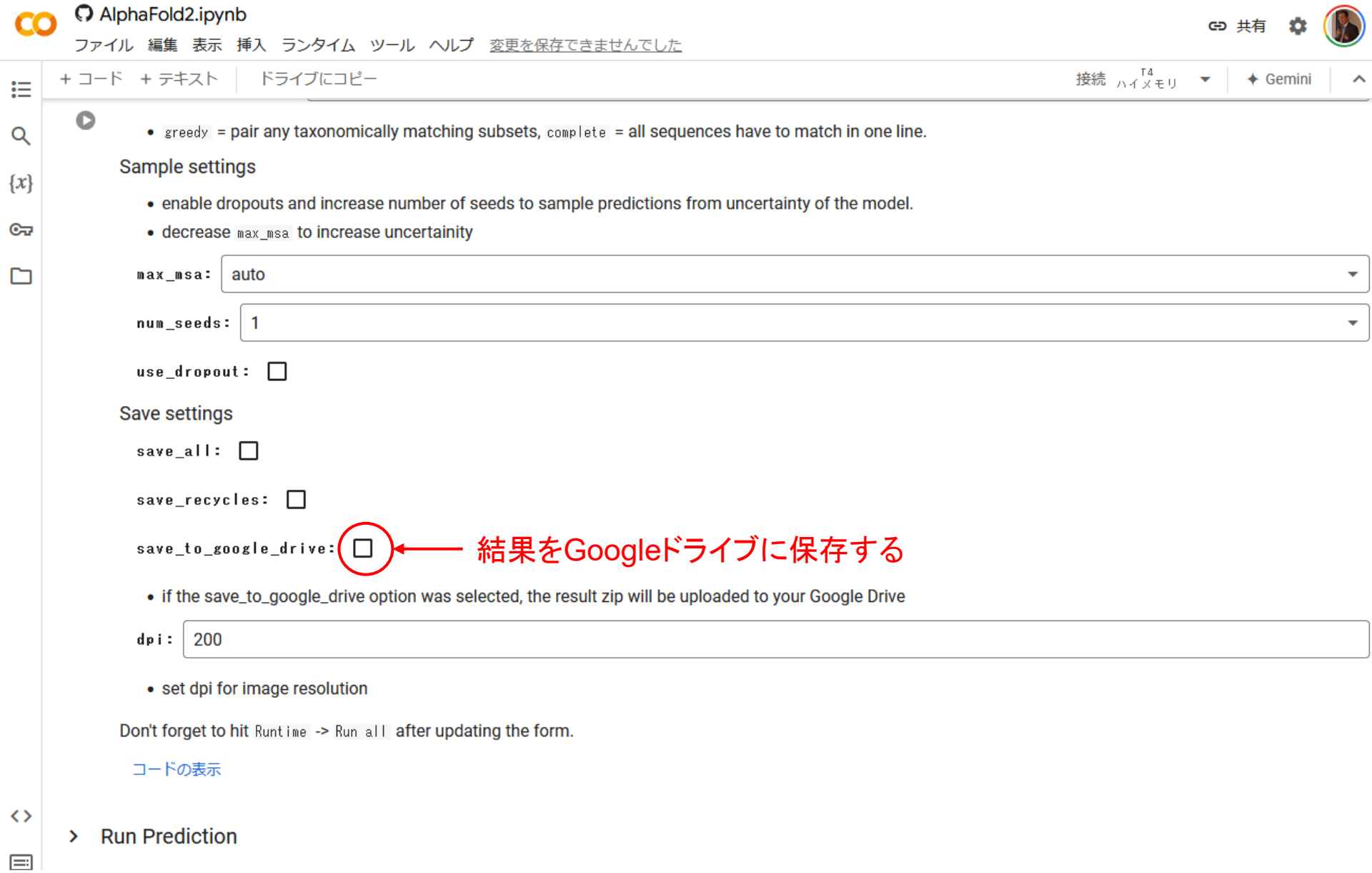
Install dependencies

コードの表示

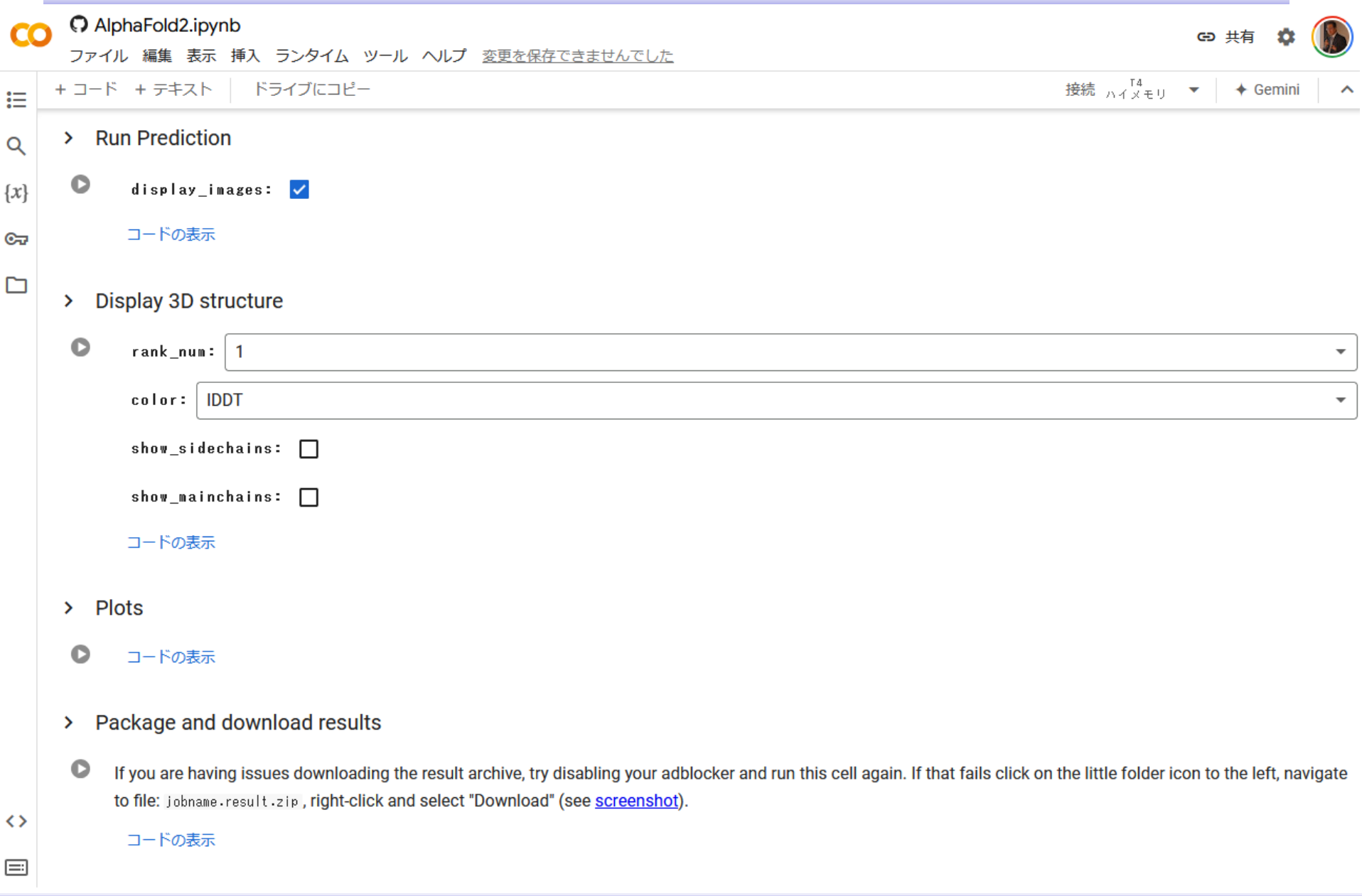
ColabFoldによる構造予測（3）



ColabFoldによる構造予測（4）



ColabFoldによる構造予測（5）



ColabFoldによる構造予測（6）

AlphaFold2.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 変更を保存できませんでした

共有

+ コード + テキスト

接続 T4 ハイメモリ Gemini

すべてのセルを実行 Ctrl+F9
より前のセルを実行 Ctrl+F8
現在のセルを実行 Ctrl+Enter
選択範囲を実行 Ctrl+Shift+Enter
以降のセルを実行 Ctrl+F10
実行を中断 Ctrl+M I
セッションを再起動する Ctrl+M .
セッションを再起動してすべて実行する
ランタイムを接続解除して削除
ランタイムのタイプを変更
セッションの管理
リソースを表示
ランタイムログの表示

ColabFold v1.5.5

Easy to use protein structu … afold2-multimer. Sequence alignments/templates are … details, see bottom of the notebook, checkout the Co … v1.4, v1.5.1, v1.5.2, v1.5.3-patch

Mirdita M, Schütze K, Moriv … Making protein folding accessible to all. *Nature M*

Input protein sequ

query_sequence: …SFVQHPELTGLDCMRPCFWVELIRGQPKENTIWTSGSSISFCGVNSDTVGWSWPDGAELPFSI

• Use : to speci … supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: flu-…

num_relax: 0

• specify how many of the top ranked structures to relax using amber

template_mode: none

• `none` = no template information is used. `pdb100` = detect templates in pdb100 (see notes). `custom` - upload and search own templates (PDB or mmCIF format, see notes)

コードの表示

Install dependencies

ColabFoldによる構造予測（7）

AlphaFold2.ipynb

ファイル　編集　表示　挿入　ランタイム　ツール　ヘルプ　変更を保存できませんでした

共有

＋ コード　＋ テキスト　ドライブにコピー

T4　RAM　ディスク　Gemini

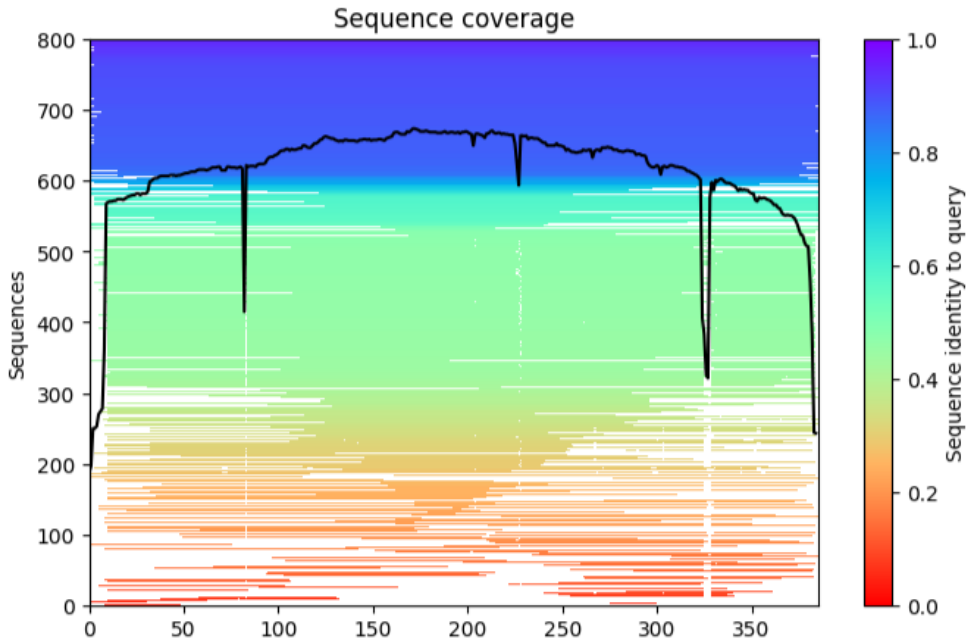
0秒 [4]　コードの表示

Run Prediction

16分 [5]　display_images: ☑

コードの表示

```
Downloading alphafold2_ptm weights to .: 100%|██████████| 3.47G/3.47G [01:37<00:00, 38.1MB/s]
2024-09-09 07:47:00,126 Running on GPU
2024-09-09 07:47:00,496 Found 5 citations for tools or databases
2024-09-09 07:47:00,497 Query 1/1: fluna_906ba (length 385)
COMPLETE: 100%|██████████| 150/150 [elapsed: 00:02 remaining: 00:00]
```



ColabFoldによる構造予測（8）

AlphaFold2.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 変更を保存できませんでした

共有

＋ コード ＋ テキスト ドライブにコピー

T4 RAM ディスク Gemini

[4] コードの表示

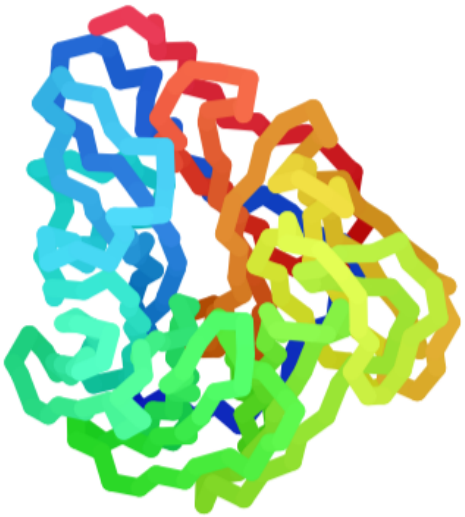
Run Prediction

display_images:

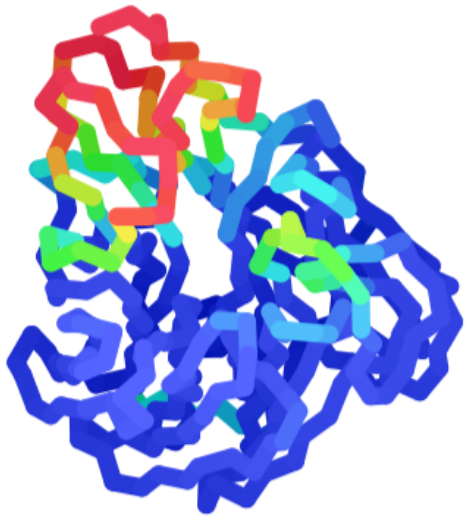
コードの表示

```
2024-09-09 07:47:04,091 Setting max_seq=512, max_extra_seq=289
2024-09-09 07:48:16,352 alphafold2_ptm_model_1_seed_000 recycle=0 pLDDT=78.7 pTM=0.829
2024-09-09 07:48:59,215 alphafold2_ptm_model_1_seed_000 recycle=1 pLDDT=85.1 pTM=0.886 tol=1.48
2024-09-09 07:49:41,437 alphafold2_ptm_model_1_seed_000 recycle=2 pLDDT=86.2 pTM=0.893 tol=0.603
2024-09-09 07:50:24,035 alphafold2_ptm_model_1_seed_000 recycle=3 pLDDT=87.2 pTM=0.899 tol=0.501
2024-09-09 07:50:24,036 alphafold2_ptm_model_1_seed_000 took 187.5s (3 recycles)
```

colored by N→C



colored by pLDDT



pLDDT値:
「比較する分子の対応する原子間の距離が0.5 Å, 1 Å, 2 Å, 4 Å以内にある割合の平均」の予測値

ColabFoldによる構造予測（9）

Display 3D structure

[6]

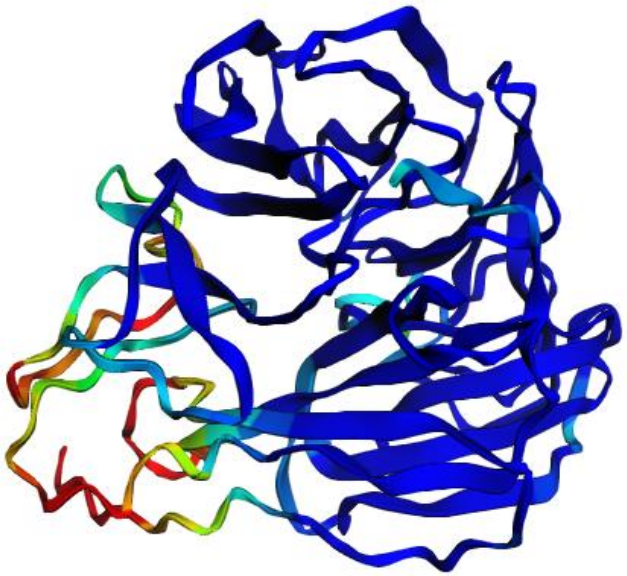
rank_num: 1

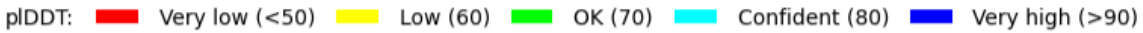
color: IDDT

show_sidechains:

show_mainchains:

コードの表示





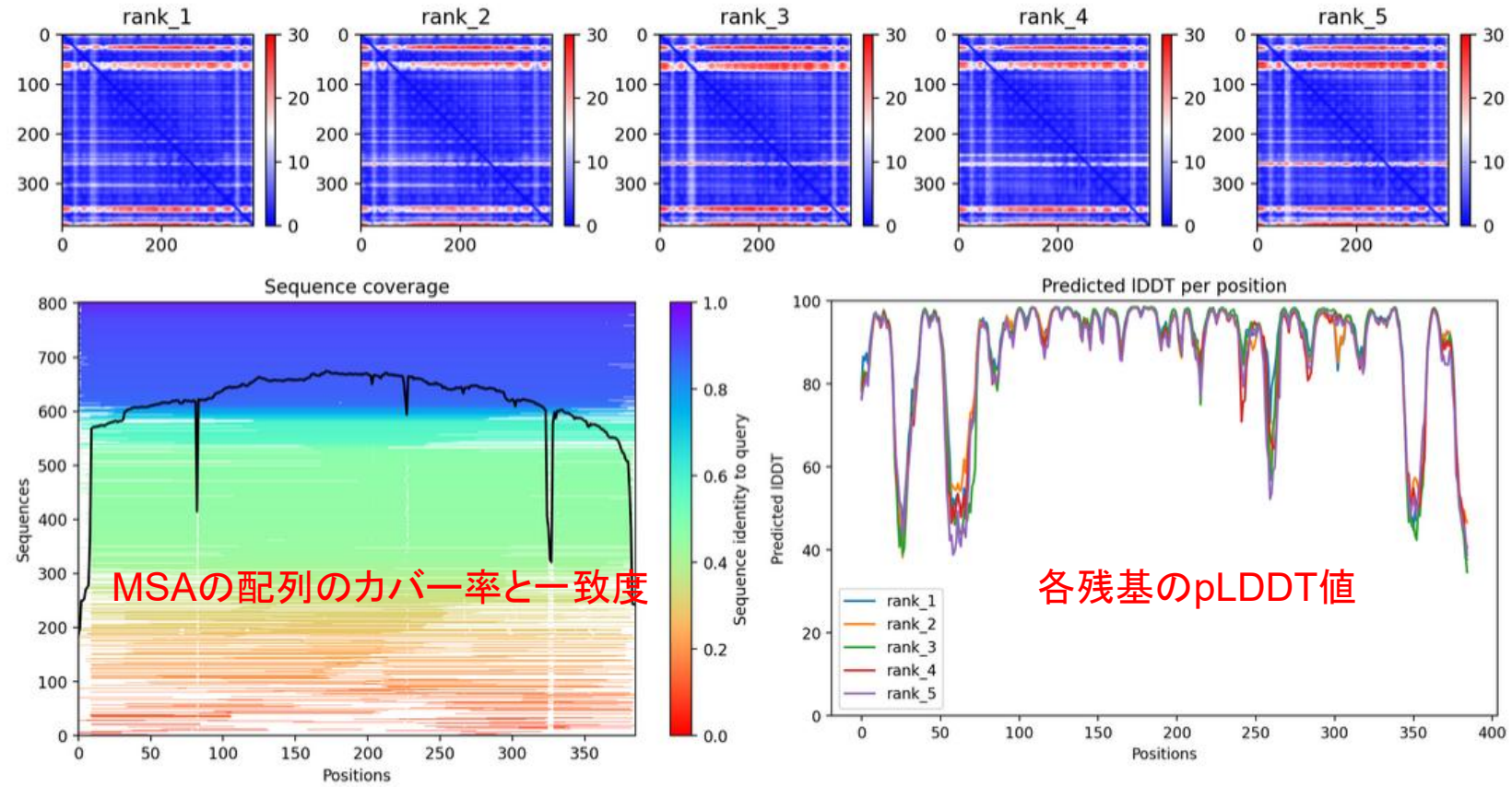
ColabFoldによる構造予測（10）

Plots

コードの表示

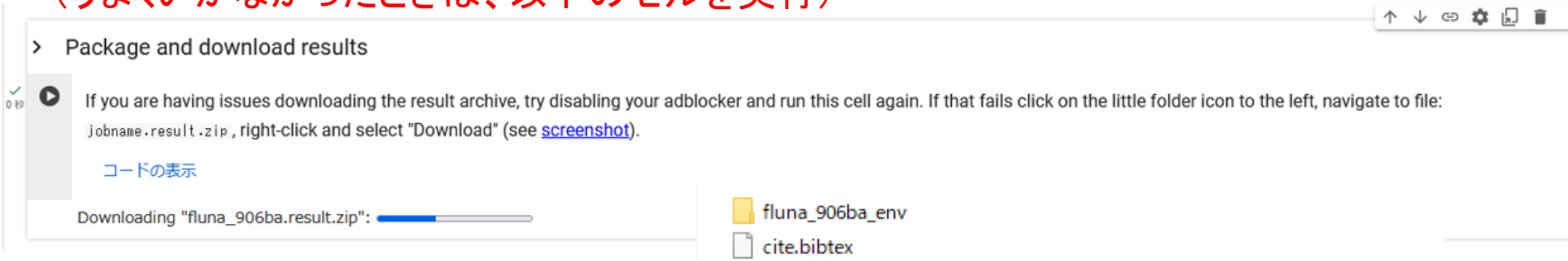
PAE（Predicted Aligned Error）: 2つの構造がある残基yでアラインされたときの残基xの予測位置の誤差（Å）

PAEの分布



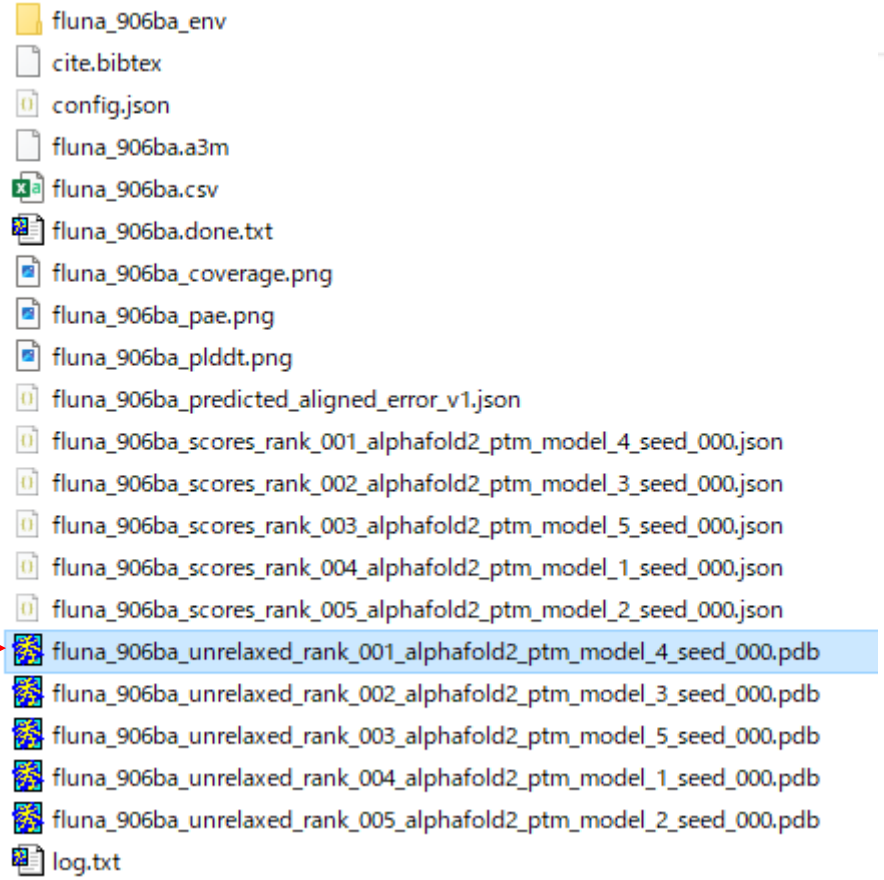
ColabFoldによる構造予測（11）

ダウンロードのフォルダにzipファイルが生成される
（うまくいかなかったときは、以下のセルを実行）



fluna_906ba.result.zipを展開

rank_001のPDBファイルを取得



AIを使った構造モデリング

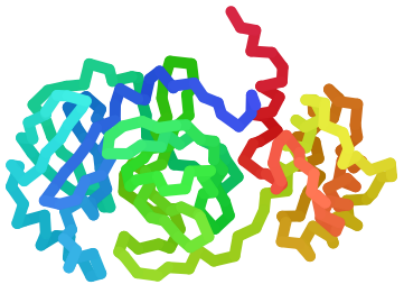
SARS-CoV-2メインプロテアーゼ

rcsb_pdb_6LU7.fasta

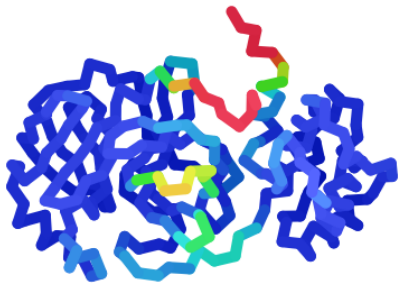
Aチェインのみを使用すること

```
2023-08-03 02:56:25.371 Setting max_seq=512, max_extra_seq=94
2023-08-03 02:57:19.697 alphafold2_ptm_model_1_seed_000 recycle=0 pLDDT=83.6 pTM=0.748
2023-08-03 02:57:46.629 alphafold2_ptm_model_1_seed_000 recycle=1 pLDDT=88.1 pTM=0.812 tol=1.18
2023-08-03 02:58:13.703 alphafold2_ptm_model_1_seed_000 recycle=2 pLDDT=88.8 pTM=0.815 tol=0.645
2023-08-03 02:58:40.346 alphafold2_ptm_model_1_seed_000 recycle=3 pLDDT=89.7 pTM=0.829 tol=0.3
2023-08-03 02:58:40.350 alphafold2_ptm_model_1_seed_000 took 122.7s (3 recycles)
```

colored by N→C

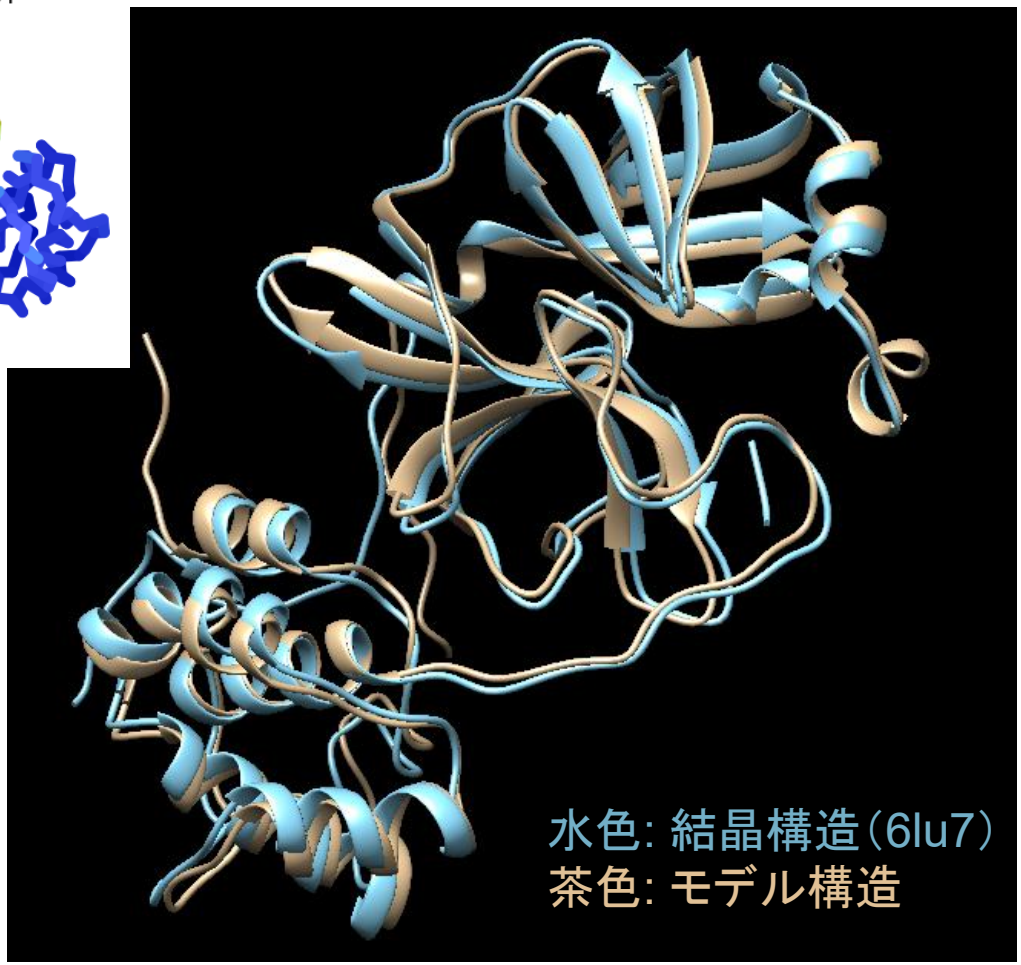


colored by pLDDT



以下は一例（数値は参考まで）

結晶構造（実験によって決定された構造）とモデル構造のとの差は、一致部分は.1.205Å（266残基）、全体は3.019Å（306残基）

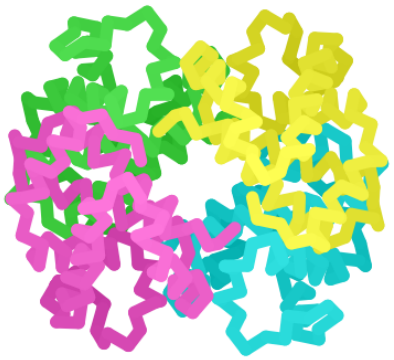


AIを使った構造モデリング

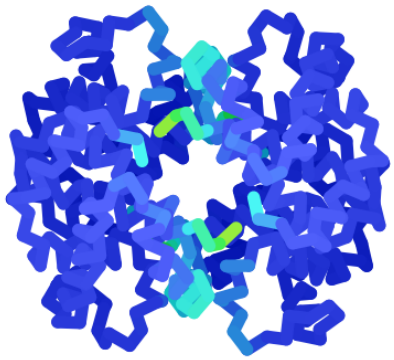
ヘモグロビン(4つの鎖まとめて)

```
2023-08-03 04:06:52,745 Setting max_seq=508, max_extra_seq=2048
2023-08-03 04:08:58,227 alphafold2_multimer_v3_model_1_seed_000 recycle=0 pLDDT=91.4 pTM=0.879 ipTM=0.854
2023-08-03 04:10:28,448 alphafold2_multimer_v3_model_1_seed_000 recycle=1 pLDDT=93.6 pTM=0.911 ipTM=0.893 tol=0.836
2023-08-03 04:11:58,605 alphafold2_multimer_v3_model_1_seed_000 recycle=2 pLDDT=93.4 pTM=0.9 ipTM=0.878 tol=0.0891
2023-08-03 04:11:58,607 alphafold2_multimer_v3_model_1_seed_000 took 300.9s (2 recycles)
```

colored by chain



colored by pLDDT

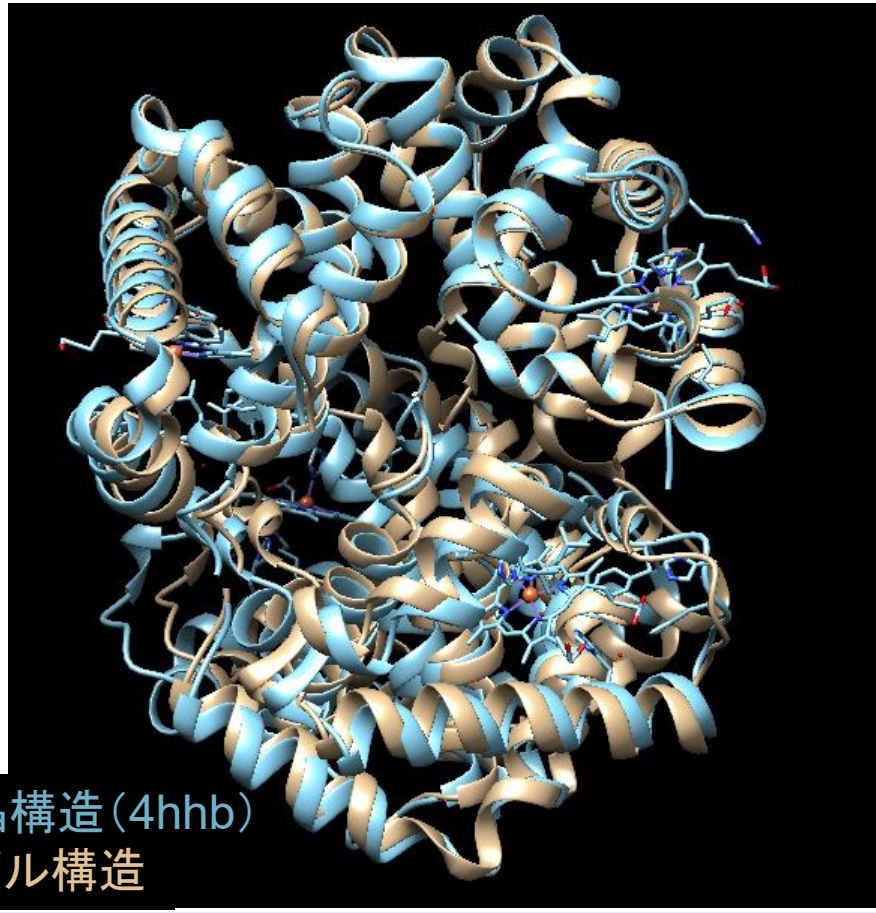


rcsb_pdb_4HHB.fasta

全チェインを使用

ColabFoldに入力するときは、

全チェインを「:」で区切って入力



水色: 結晶構造(4hhb)

茶色: モデル構造

以下は一例(数値は参考まで)

結晶構造(実験によって決定された構造)

とモデル構造のとの差は、一致部分は

0.437Å(145残基)

まとめ課題

インフルエンザH1N1のノイラミニダーゼのモデル構造とタミフルをドッキングし、実際の複合体構造と比較してみよう

ノイラミニダーゼの配列は、flu-na.fastaにある

これをColabFoldに入力し、構造をモデリングする

タミフルは薬剤の一般名で、正式名は、オセルタミビルリン酸塩（Oseltamivir Phosphate）

PubChemで、タミフルの正式名で検索し、構造を取得する

モデル構造とタミフルをSwissDockを用いてドッキングする

ドッキングした結果と実際の複合体構造（PDB ID: 3CL0）をChimeraX（他のソフトでもよい）を用いて比較する

モデル構造をAutoDock（SwissDockのページの右側のメニュー）に入力するとエラーになる可能性があるが、その場合は、薬剤化合物が結合していないノイラミニダーゼの構造を用いてドッキングしてもよい

モデル構造に対してAttracting Cavities（左側）を適用した結果

Query

Ligand	OP(O)(O)=O.CCOC(=O)C1=C[C@@H](OC(CC)CC)[C@H](NC(C)=O)[C@@H](N)C1
Target	fluna_906ba_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb
Method	Attracting Cavities 2.0
Date	December 16, 2024, 9:52 am UTC

Parameters:

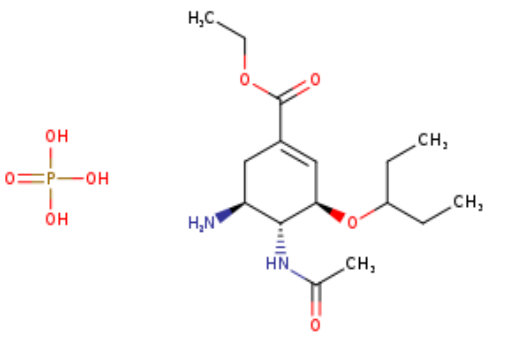
Box center:	-7 - 0 - -7	Sampling exhaustivity:	medium	Number of RIC:	1
Box size:	25 - 25 - 25	Cavity prioritization:	buried		

If you publish these results, please, cite the following papers:

Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, Zoete V. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Res.* **2024**

Röhrig UF, Goullieux M, Bugnon M, Zoete V. Attracting Cavities 2.0: improving the flexibility and robustness for small-molecule docking. *J. Chem. Inf. Model.*, **2023**

Ligand



1nn2の構造に対してAutoDock（右側）を適用した結果　→ 現在、正常に動作しない

Query

Ligand	OP(O)(O)=O.CCOC(=O)C1=C[C@@H](OC(CC)CC)[C@H](NC(C)=O)[C@@H](N)C1
Target	1nn2_modified.pdb
Method	AutoDock Vina
Date	December 16, 2024, 10:06 am UTC

Parameters:

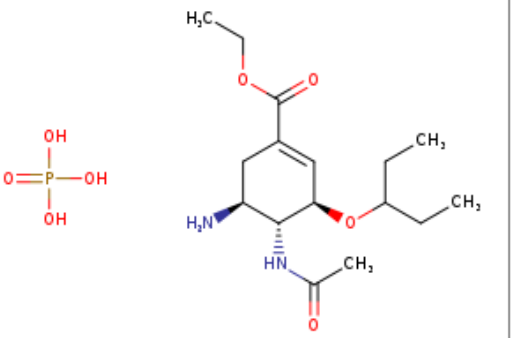
Box center:	91 - 93 - 66	Sampling exhaustivity:	4
Box size:	24 - 24 - 24		

If you publish these results, please, cite the following papers:

Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, Zoete V. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Res.* **2024**

Eberhardt J, Santos-Martins D, Tillack AF, Forli S.. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.*, **2021**

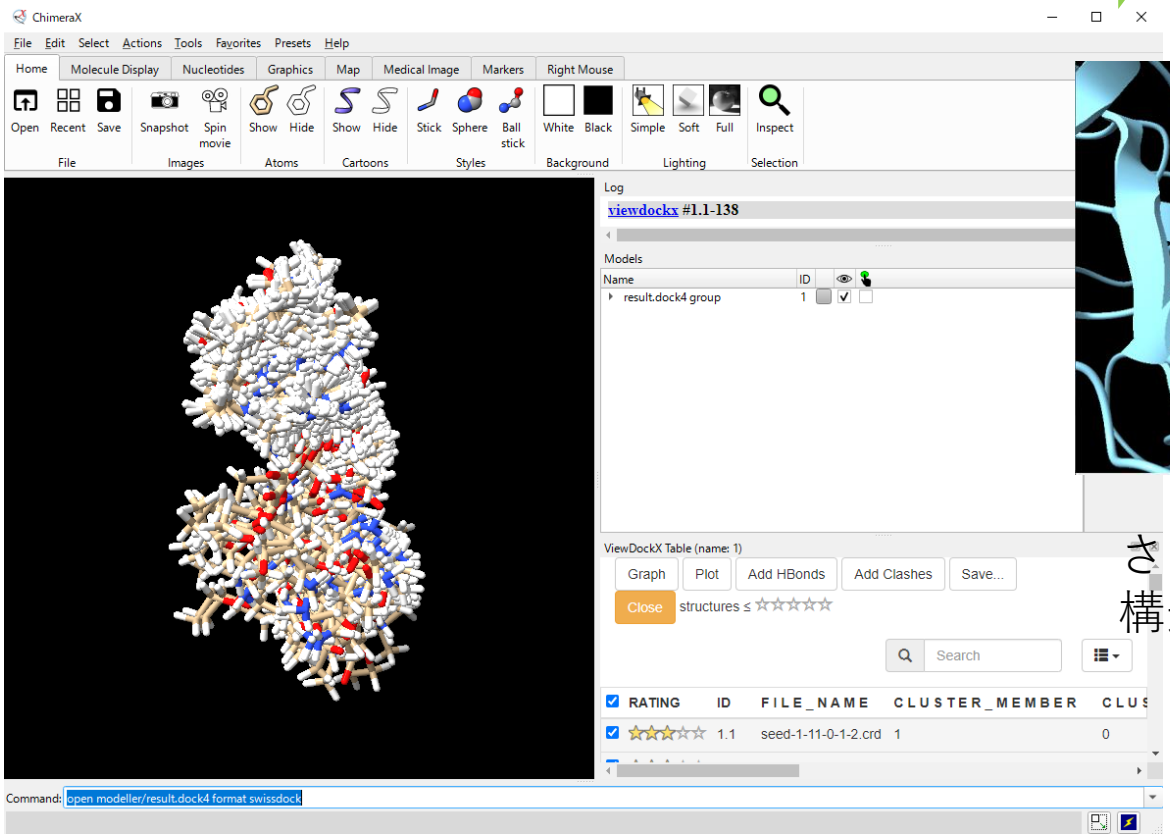
Ligand



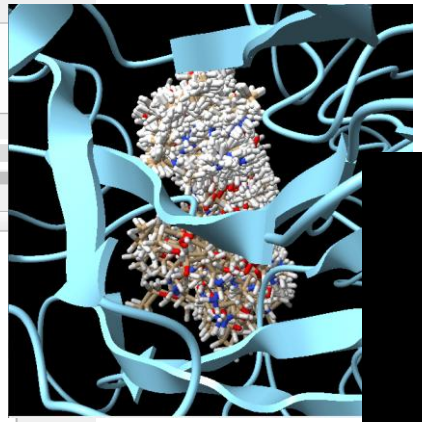
ドッキング結果のChimeraXによる表示

ChimeraXのコマンドで、「open result.dock4 format swissdock」と入力
ただし、result.dock4 のディレクトリに注意（pwdコマンドで、ChimeraXの作業用ディレクトリを取得し、そこにresult.dock4を置く）

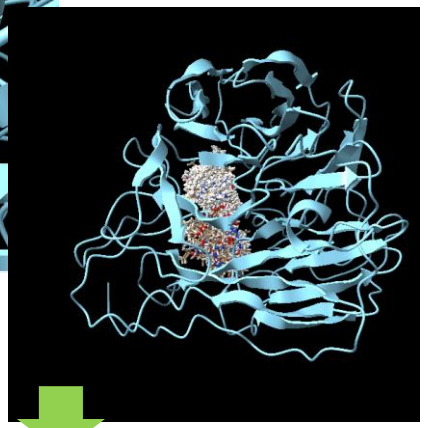
リガンドの複数の候補（ポーズ）が表示される



「File」→「Open」で receptor.pdbを指定



ズームアウト



さらにモデル構造を絞る

実際の複合体構造（PDB ID: 3CL0）をFetch by IDで読み込み、Matchmakerで構造比較）