

AlphaFold2による
タンパク質の構造予測

清水謙多郎

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

時間がかかりますので、[課題1\(p.51\)](#)からすぐに始めて下さい。

2021 BREAKTHROUGH OF THE YEAR

Protein structures for all

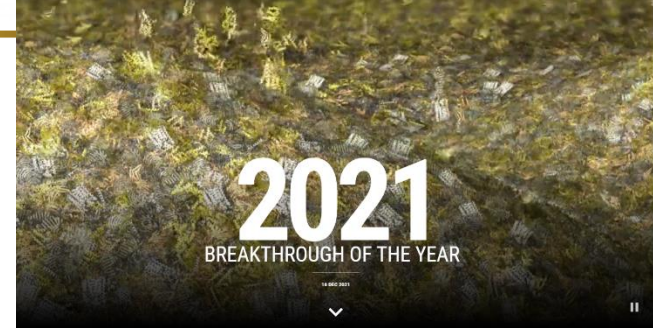
AI-powered predictions show proteins finding their shapes

BY ROBERT SERVICE

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



[Science 2021-12-16](#)

AlphaFold

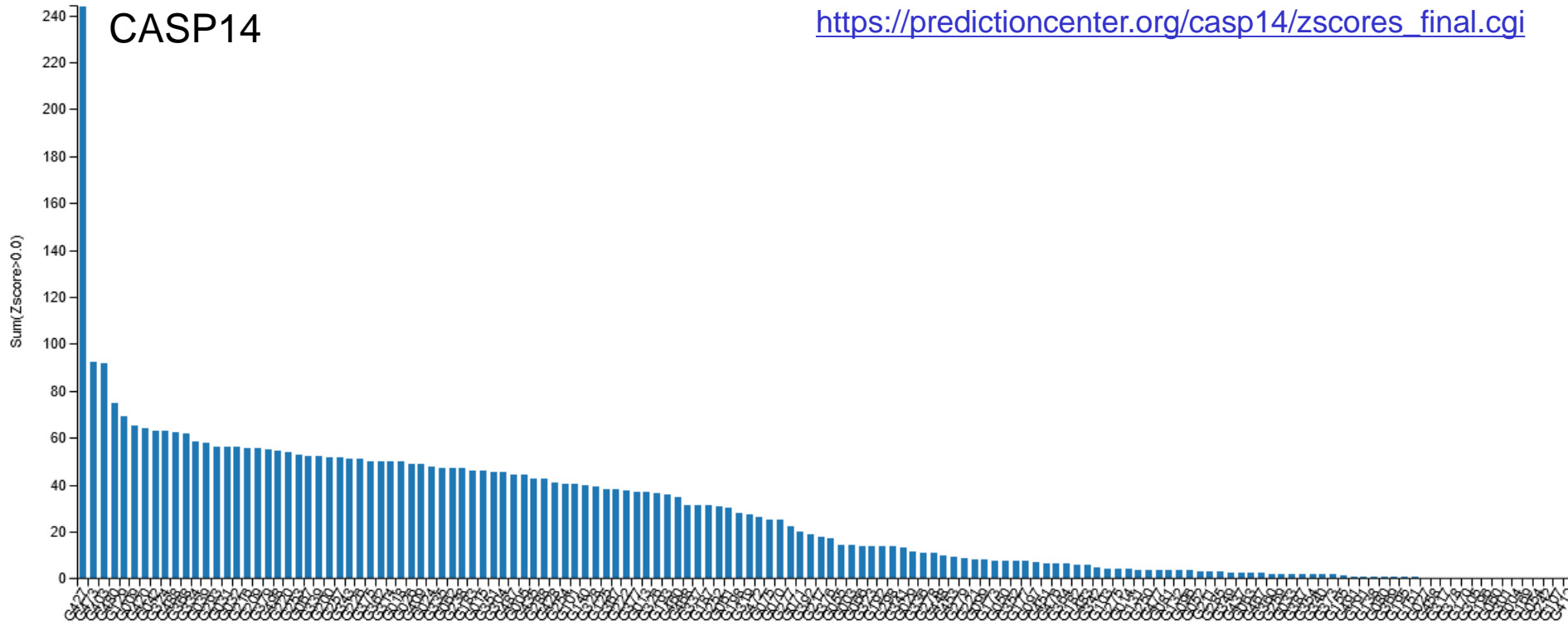
- DeepMind社によるタンパク質構造予測システム
 - DeepMind社: Googleおよびグループ企業の持株会社であるAlphabet 社の翼下にある人工知能開発会社
- CASP13で最も高い予測精度を達成
 - CASP (Critical Assessment of Structure Prediction): 1994年から2年ごとに開催されてきたタンパク質構造予測技術の客観的な評価に基づくコンペティション
- CASP14ではさらに予測精度を向上
 - CASP14の約3分の2のターゲットにおいてGDTが90以上のスコアを達成

[GDTについては後で説明します。](#)

AlphaFold2のCASPの成績

CASP14

https://predictioncenter.org/casp14/zscores_final.cgi



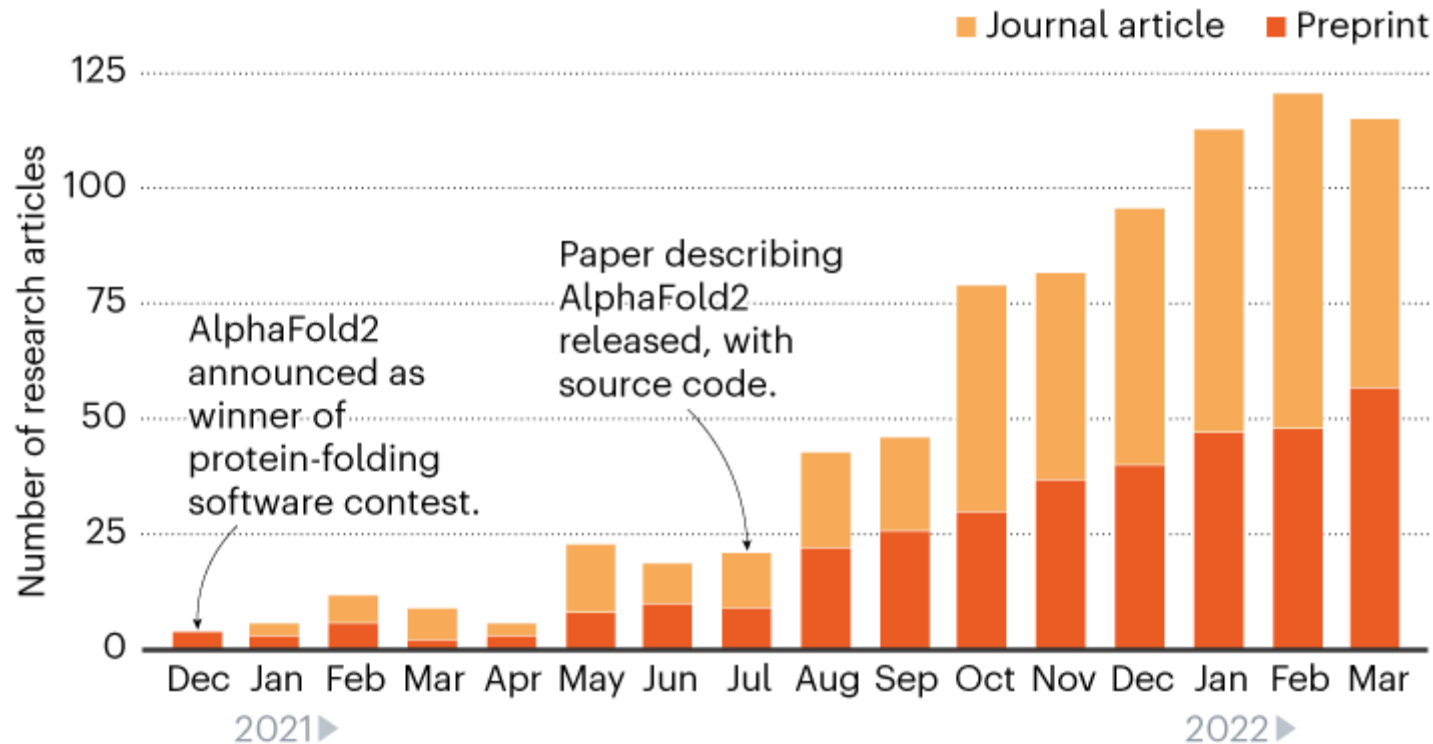
Groups

#	code	name	Domains Count	>-2.0	>-2.0	>-2.0	>-2.0	>0.0	>0.0	>0.0	>0.0
1	427	AlphaFold2	92	244.0217	1	2.6524	1	244.0217	1	2.6524	1
2	473	BAKER	92	90.8241	2	0.9872	2	92.1241	2	1.0013	2
3	403	BAKER-experimental	92	88.9672	3	0.9670	3	91.4731	3	0.9943	3
4	480	FEIG-R2	92	72.5351	4	0.7884	4	74.5627	4	0.8105	4
5	129	Zhang	92	67.9065	5	0.7381	5	68.8922	5	0.7488	5
6	009	tFold_human	92	61.2858	7	0.6661	8	65.2157	6	0.7089	7
7	420	MULTICOM	92	63.2689	6	0.6877	7	64.0531	7	0.6962	8
8	042	QUARK	92	60.0226	10	0.6524	11	62.9711	8	0.6845	9
9	324	Zhang-Server	92	60.8875	8	0.6618	9	62.9122	9	0.6838	10
10	488	tFold-IDT_human	92	57.6435	11	0.6266	12	62.0795	10	0.6748	11
11	368	tFold-CaT_human	92	60.5423	9	0.6581	10	61.8464	11	0.6722	12
12	334	FEIG-R3	92	48.4424	20	0.5265	23	58.5809	12	0.6367	13

AlphaFold2を引用している論文

ALPHAFOLD MANIA

The number of research papers and preprints citing the AlphaFold2 AI software has shot up since its source code was released in July 2021*.



*Nature analysis using Dimensions database; removing duplicate preprints and papers/R. Van Noorden, E. Callaway.

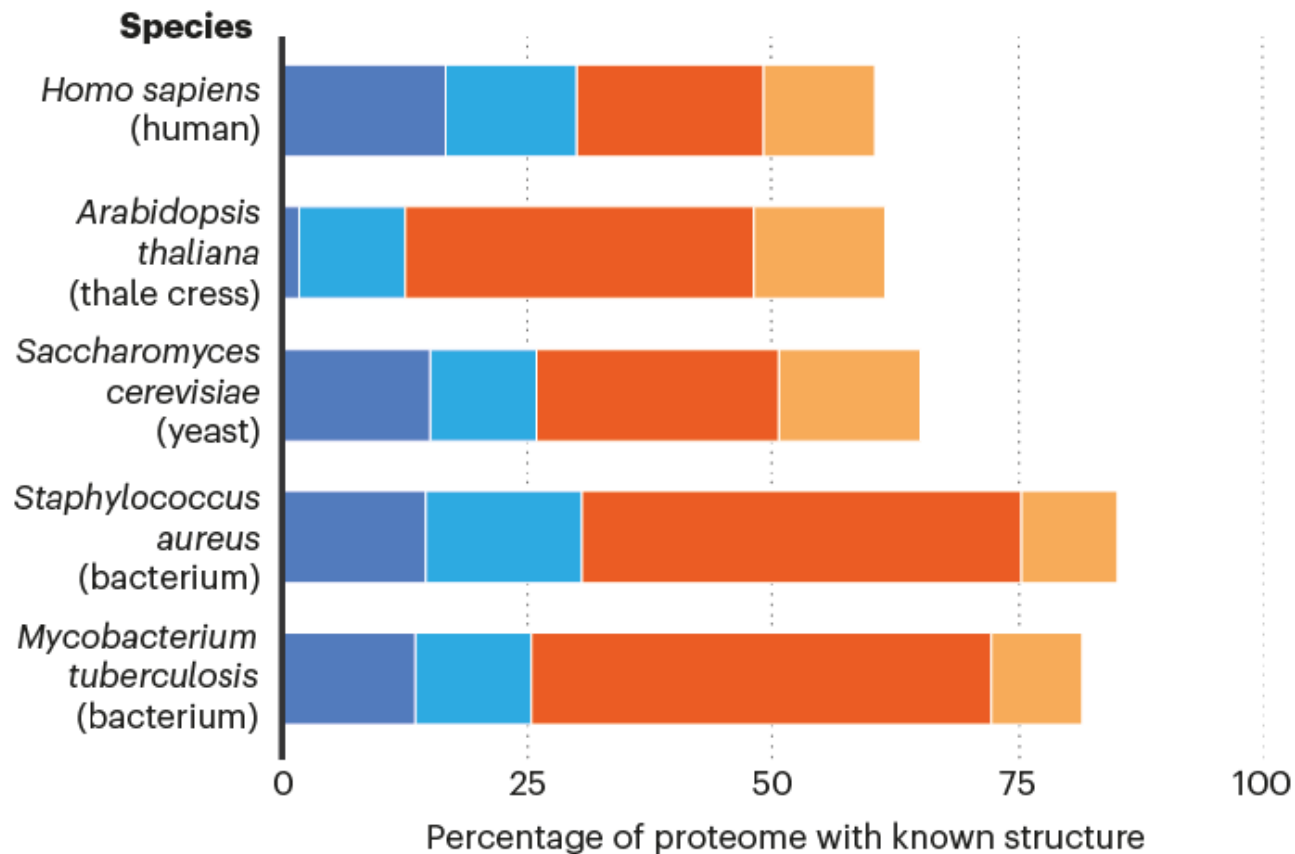
©nature

[Callaway, E, et al. What's next for AlphaFold and the AI protein-folding revolution, Nature New Feature, 13 April 2022](#)

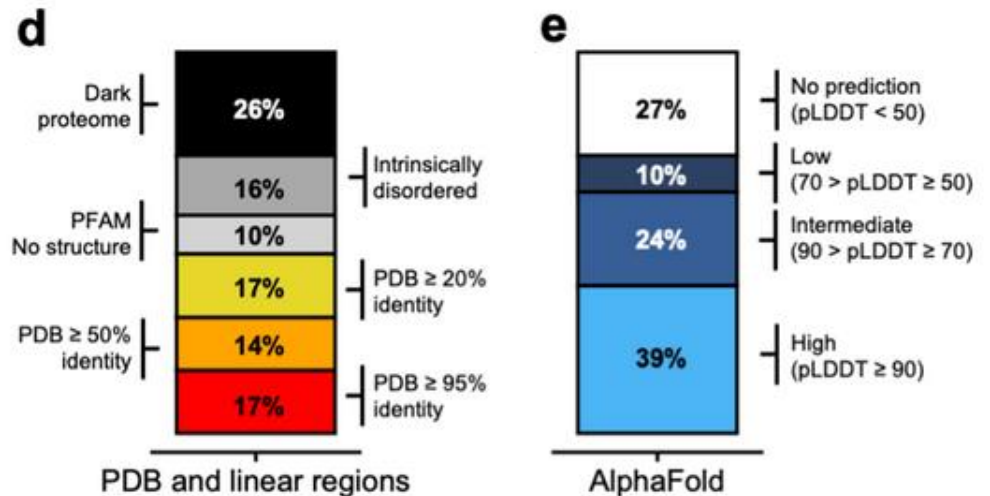
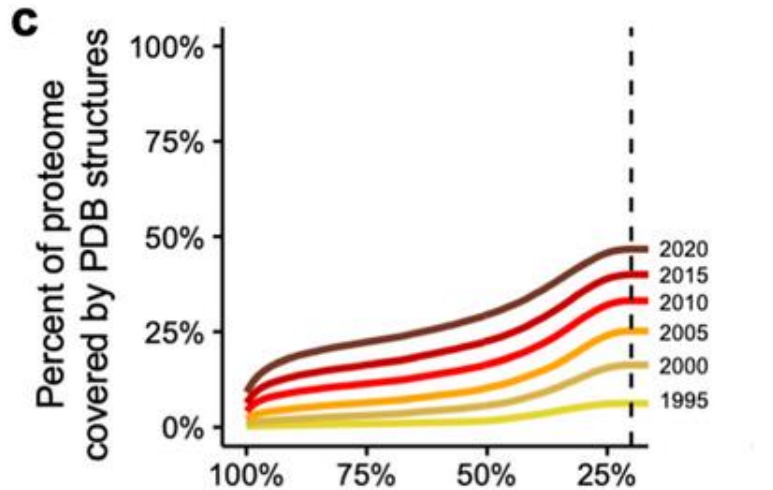
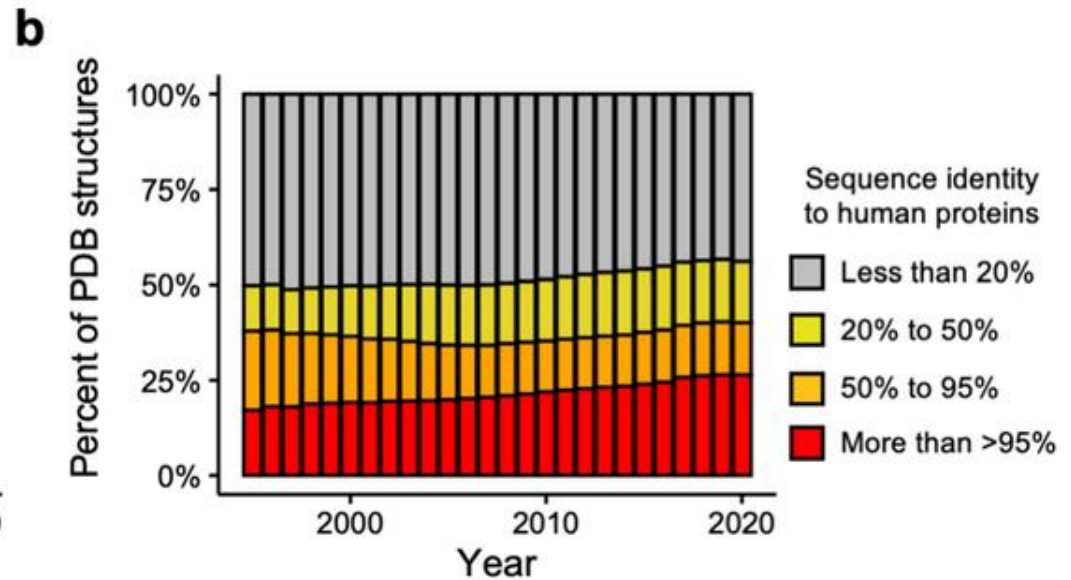
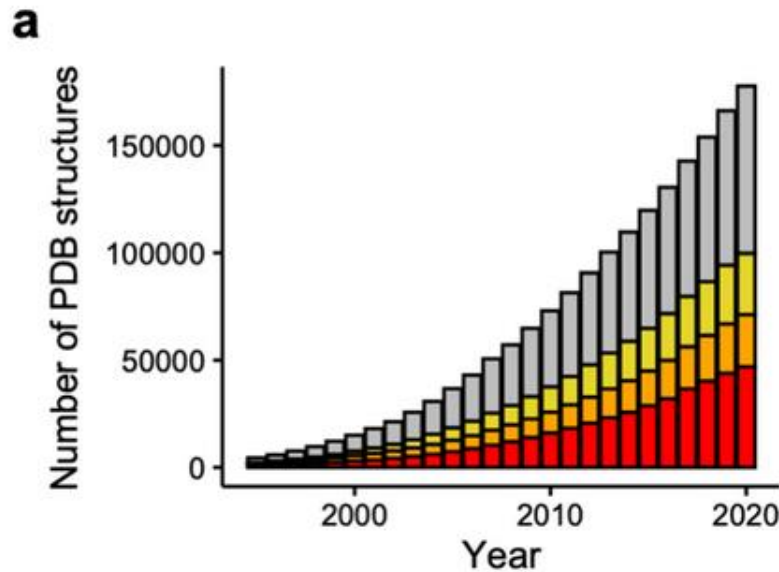
構造が既知のタンパク質の割合

Source of knowledge about proteome

- High-quality experimental structures in the PDB*
- Structural knowledge derived from related proteins in the PDB*
- Knowledge from AlphaFold models only (high confidence)
- Knowledge from AlphaFold models only (intermediate confidence)



ヒトのタンパク質のモデリング

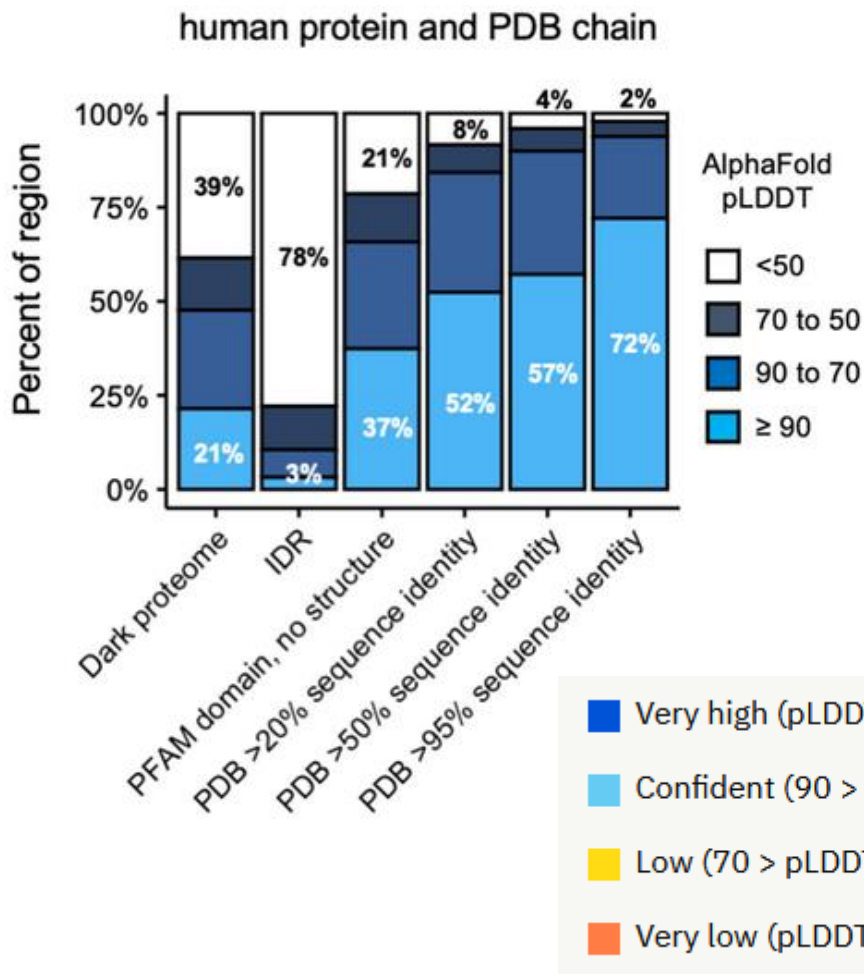


Minimum sequence identity between human protein and PDB chain

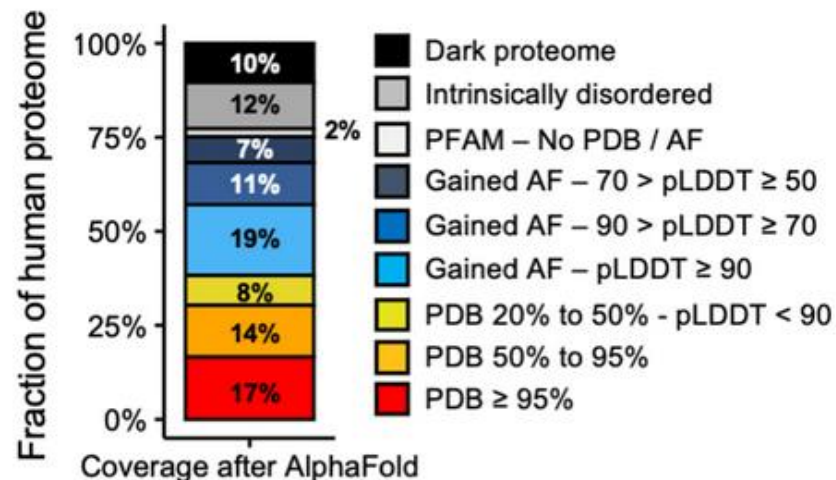
E Porta-Pardo, et al. The structural coverage of the human proteome before and after AlphaFold. PLoS Comput Biol. 2022 18:e1009818.

ヒトのタンパク質のモデリング

f



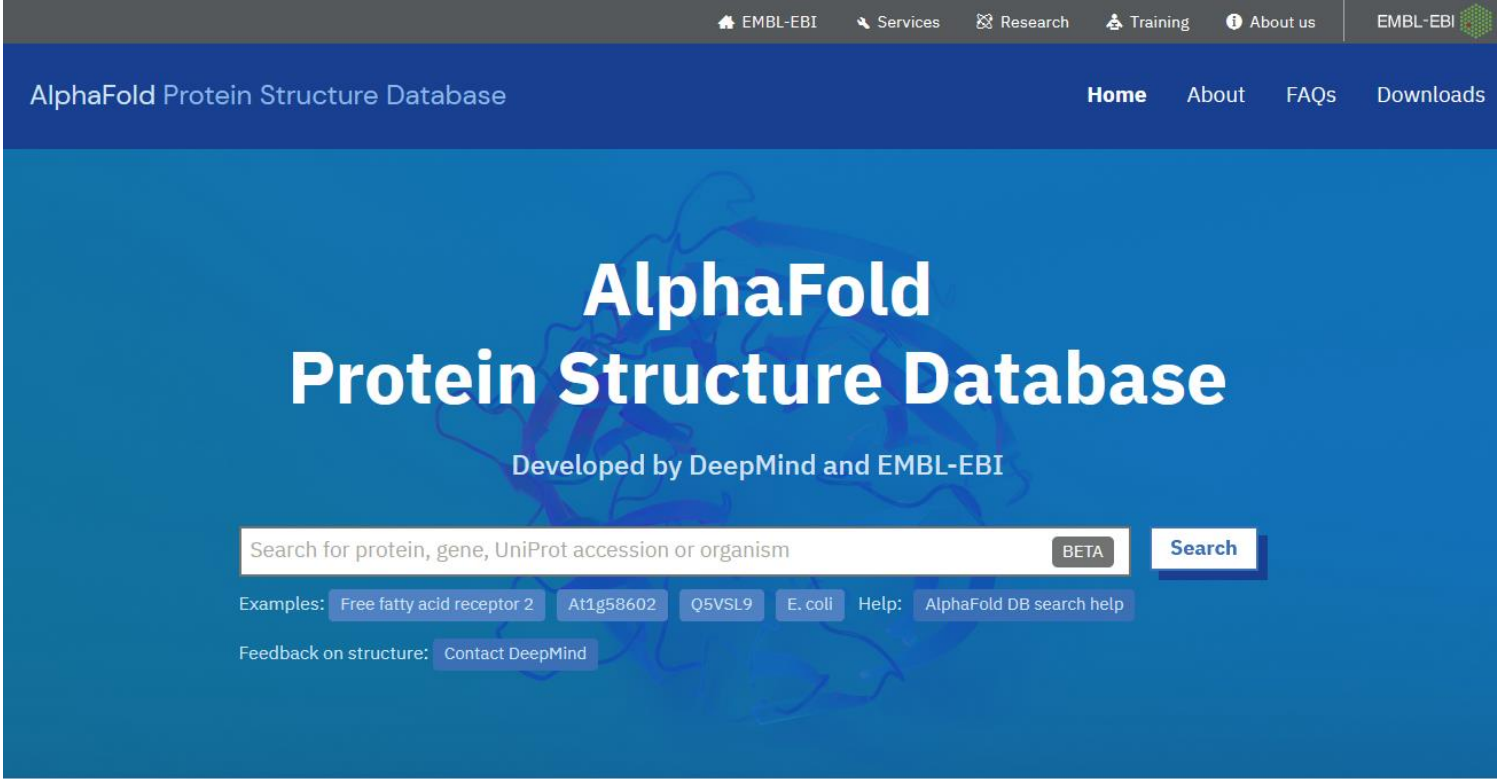
g



pLDDTについては後で説明します。

AlphaFold Protein Structure Database

- AlphaFoldが予測したタンパク質構造のデータベース
- <https://alphafold.ebi.ac.uk/>



AlphaFold Protein Structure Database

Home About FAQs Downloads

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) Help: [AlphaFold DB search help](#)

Feedback on structure: [Contact DeepMind](#)

AlphaFold DB provides open access to 992,316 protein structure predictions for the human proteome and other key proteins of interest, to accelerate scientific research.

AlphaFold Protein Structure Database

gibberellin receptor|

「gibberellin receptor」と入力

BETA

Search

Showing all search results for gibberellin receptor ジベレリン受容体

1 - 20 of 13284 results

Filter by:

Organism

- Mus musculus (2766)
- Rattus norvegicus (2325)
- Homo sapiens (1762)
- Danio rerio (1281)
- Caenorhabditis elegans (758)
- Drosophila melanogaster (365)
- Zea mays (343)
- Arabidopsis thaliana (304)
- Bos taurus (277)
- Schistosoma mansoni (175)
- Trichuris trichiura (147)
- Gallus gallus (127)
- Sus scrofa (125)
- Xenopus laevis (124)
- Bos taurus (124)

Gibberellin receptor GID1

Q6L545 (GID1_ORYSJ)

Protein Gibberellin receptor GID1

Gene GID1

Source Organism Oryza sativa subsp. japonica [search this organism](#)

UniProt Q6L545 [go to UniProt](#)

PDBe-KB 2 PDB structures for Q6L545 [go to PDBe-KB](#)

UniProtKBへのリンク

PDB構造が存在するもの

Gibberellin receptor GID1L2

A0A1D6ERI3 (A0A1D6ERI3_MAIZE)

Protein Gibberellin receptor GID1L2

Gene ZEAMMB73_Zm00001d005909

Source Organism Zea mays [search this organism](#)

UniProt A0A1D6ERI3 [go to UniProt](#)

PDB構造が存在しないもの

Gibberellin receptor GID1L2

A0A1D6F547 (A0A1D6F547_MAIZE)

AlphaFold Protein Structure Database

EMBL-EBI

Services

Research

Training

About us

EMBL-EBI



AlphaFold Protein Structure Database

Home

About

FAQs

Downloads

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#)

Help: [AlphaFold DB search help](#)

Gibberellin receptor GID1A

AlphaFold structure prediction

Download [PDB file](#) [mmCIF file](#) [Predicted aligned error](#)

NEW Feedback on structure

[Looks great](#)

[Could be improved](#)

Google Formsによる
フィードバック

Information

Protein	Gibberellin receptor GID1A
Gene	GID1A
Source organism	Arabidopsis thaliana (Mouse-ear cress) go to search
UniProt	Q9MAA7 go to UniProt
Experimental structures	2 structures in PDB for Q9MAA7 go to PDBe-KB
Biological function	Functions as soluble gibberellin (GA) receptor. GA is an essential hormone that regulates growth and development in plants. Binds with high affinity the biologically active gibberellin GA4, but has no affinity for the biologically inactive GAs. In response to GA, interacts with specific DELLA proteins, known as repressors of GA-induced growth, and targets them for degradation via proteasome. Seems to be required for GA signaling that controls root growth, seed

AlphaFold Protein Structure Database

3D viewer

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Views

Predicted aligned error

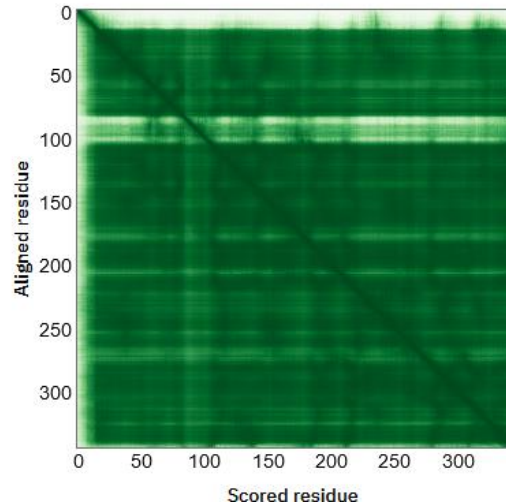
Sequence Features (coming soon)

```
Sequence of AF-Q9MAA7... 1: Gibberellin... A
1  MAASDEVNLI  ESRTVVPLNT  WVLISNFKVA  YNILRRPDGT  FNRHLAEYLD  RKVTTANANP  V  DGVFSFDVLI  DRRINLLSR  VYRPAADQEQ  PPSILDLEK  PVDGDIVF  VILFFHGG  SFAHSSAN
21  SAIYDITL  CRRLVGLCK  CVVSVNYRA  PENPYPCAY  D  DGNIALNWN  V  SRSWLKSK  KD  SKVHI  FLAGD  SSGGNIA  HNVALRAGE  SGIDVLGN  ILLNPM  FGGNERT  ESEKSLD  GKYFVT  VRDRDWY
41  WKAFLPE  GEDREH  PACNPF  SPRG  KSLEGV  SFPKSL  VVVAGL  DL  IRD  WQLAY  AE  GLKKAG  QEVK  LMHLE  KATVG  FYLLP  NNNHF  HNV  MDEI  SAF  VNAEC
```

pLDDTの値による色づけ



PAEの評価結果



PAEについては後で説明します。

AlphaFold Protein Structure Database

- Function
- Names & Taxonomy
- Subcellular Location
- Phenotypes
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar Proteins

Structure UniProtKB Q9MAA7 GID1A_ARATHの構造の表示



既知のPDB構造 (2ZSH) が表示される

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
--Select--		--Select--				
PDB	2ZSH	X-ray	1.80 Å	A	1-344	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	2ZSI	X-ray	1.80 Å	A	1-344	PDBe · RCSB-PDB · PDBj · PDBsum
AlphaFold	AF-Q9MAA7-F1	Predicted			1-345	AlphaFold

Features

Showing features for helixⁱ, beta strandⁱ, turnⁱ.



AlphaFoldをクリックすると、そのモデル構造が表示される

AlphaFold Protein Structure Database

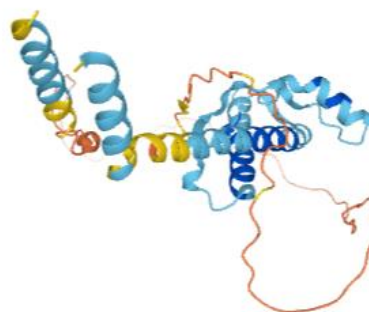
3D viewer

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured

```
Sequence of AF-P04156-F1 1: Major prio... A
1  MANLGCNMLVLFVATWSDLG LCKKRPKPGGWNITGGSRYPG QGSPGGNRYPPQGGGGWGQP HGGGGWQPHGGGGWQPHGGGGWQ QGGGTHSQNNKPSKPKTNMK HMAGAAAAGAVVG
194  GLGGYMLGSAMSRPIIH FGSYEDRYV RENMHRYPNQVYYRPMDEYS NQNNFVHDCVNITIKQHVTI TTTKGENFTE TDVQMMERVVEQMCITQYER ESQAYYQRGSSMVLFPSPV ILLISFL
261  IFLIVG
```



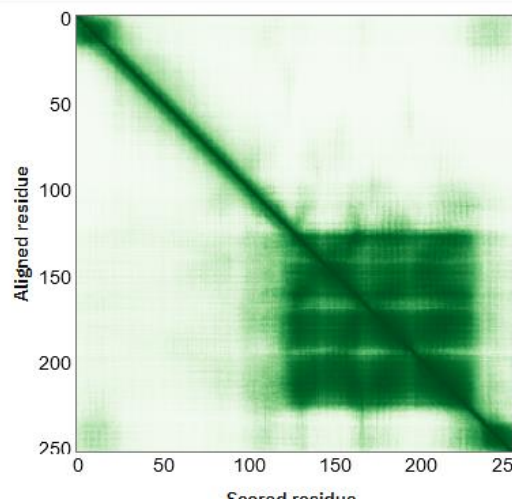
ヒトのプリオンタンパク質 (Major prion protein) の例

Major prion protein

P04156 (PRIO_HUMAN)

Protein	Major prion protein
Gene	PRNP
Source Organism	Homo sapiens search this organism
UniProt	P04156 go to UniProt
PDBe-KB	62 PDB structures for P04156 go to PDBe-KB

Sequence Features (coming soon)



AlphaFold Protein Structure Database

- |Function
- Names & Taxonomy
- Subcellular Location
- Disease & Drugs
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence & Isoform
- Similar Proteins

P04156 · PRIO_HUMAN

Major prion protein · Homo sapiens (Human) · Gene: PRNP (ALTPRP, PRIP, PRP) · 253 amino acids · Evidence at protein level · Annotation score: 5/5

ヒトのプリオンタンパク質のUniProtKBのページ



Entry Feature viewer Publications External links History

<https://www.uniprot.org/uniprot/P04156>

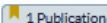
BLAST Align  Download  Add Add a publication Entry feedback


Functionⁱ

Its primary physiological function is unclear. May play a role in neuronal development and synaptic plasticity. May be required for neuronal myelin sheath maintenance. May promote myelin homeostasis through acting as an agonist for ADGRG6 receptor. May play a role in iron uptake and iron homeostasis. Soluble oligomers are toxic to cultured neuroblastoma cells and induce apoptosis (in vitro) (By similarity).


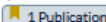
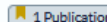
Association with GPC1 (via its heparan sulfate chains) targets PRNP to lipid rafts. Also provides Cu(2+) or Zn(2+) for the ascorbate-mediated GPC1 deaminase degradation of its heparan sulfate side chains (By similarity).  3 Publications 

Miscellaneous

This protein is produced by a bicistronic gene which also produces the alternative prion protein/AltPrP (AC AC F7VJQ1) from an overlapping reading frame. 

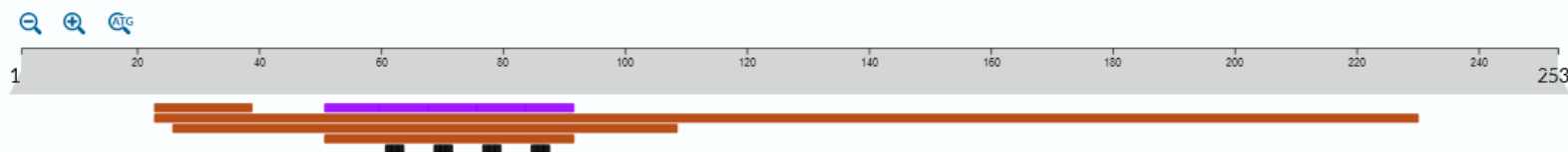
The alternative prion protein/AltPrP (AC AC F7VJQ1) and PRNP have no apparent direct functional relation since a mutation that removes the start codon of the AltPrP has no apparent effect on the biology of PRNP. In mouse and hamster, the alternative initiation AUG codon is absent and is replaced by a GUG codon. 

Caution

 An isoform was shown to be localized to both the cytoplasm and the nucleus and to be sumoylated with SUMO1 (PubMed:19059915). The article has later been withdrawn by the authors  1 Publication 

Features

Showing features for repeatⁱ, regionⁱ, metal bindingⁱ.



TYPE	ID	POSITION(S)	DESCRIPTION
-- Select --			

AlphaFold Protein Structure Database

Function

Names & Taxonomy

Subcellular Location

Disease & Drugs

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence & Isoform

Similar Proteins



TYPE	ID	POSITION(S)	DESCRIPTION	
-- Select --				
▶ Repeat		51-59	1	BLAST
▶ Repeat		60-67	2	BLAST
▶ Repeat		68-75	3	BLAST
▶ Repeat		76-83	4	BLAST
▶ Repeat		84-91	5	BLAST
▶ Region		23-230	Interaction with GRB2, ERI3 and SYN1	By Similarity BLAST
▶ Region		23-38	Interaction with ADGRG6	By Similarity BLAST
▶ Region		26-108	Disordered	1 Automatic Annotation BLAST
▶ Region		51-91	5 X 8 AA tandem repeats of P-H-G-G-W-G-Q	BLAST

GO Annotationsⁱ

Slimming set:

generic



Cell color indicative of number of GO terms

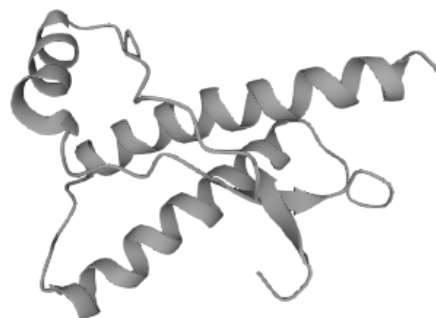
ASPECT	TERM	
Cellular Component	anchored component of external side of plasma membrane	1 Publication NAS:ARUK-UCL











AlphaFold Protein Structure Database

- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Drugs
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence & Isoform
- Similar Proteins

Structureⁱ

Model 1 / 20



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
-- Select --		-- Select --				
PDB	1E1G	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1E1J	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1E1P	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1E1S	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1E1U	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1E1W	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1FKC	NMR		A	90-231	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1FO7	NMR		A	90-231	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1H0L	NMR		A	121-230	PDBe · RCSB-PDB · PDBj · PDBsum 
PDB	1HJM	NMR		A	125-228	PDBe · RCSB-PDB · PDBj · PDBsum 

Features

Showing features for beta strand¹, turn¹, helix¹.

AlphaFold Protein Structure Database

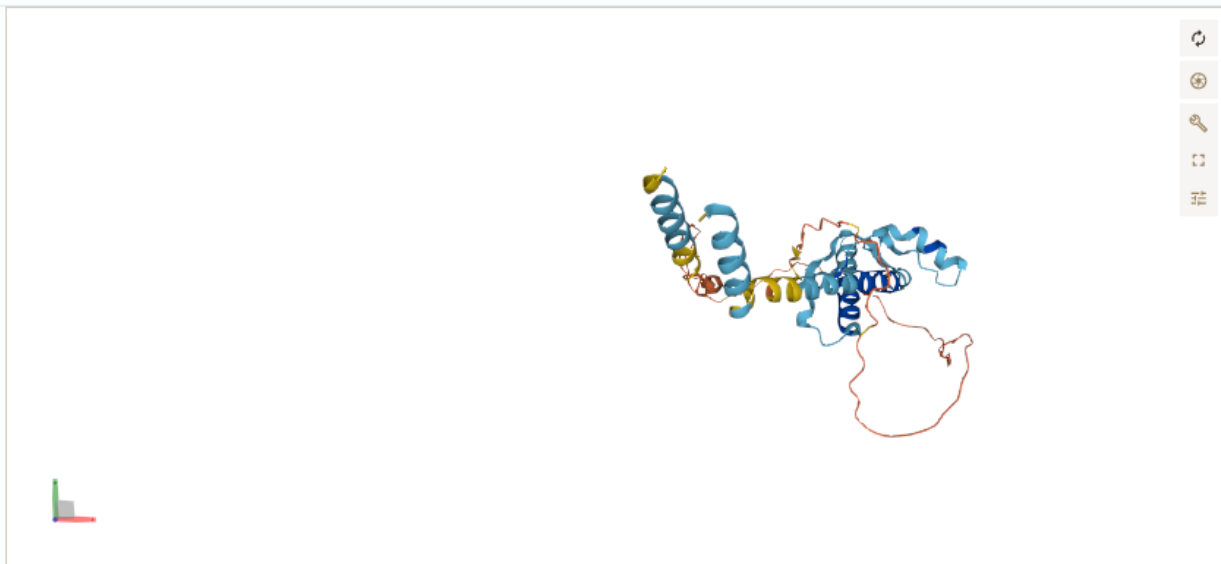
- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Drugs
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence & Isoform
- Similar Proteins

Structureⁱ

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
PDB	5YJ5	NMR		A	91-231	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6DU9	X-ray	2.33 Å	A	90-230	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6LNI	EM	2.70 Å	A/B/C/D/E/F/G/H/I/J	23-231	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6PQ5	X-ray	1.50 Å	A/B	113-118	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6PQA	X-ray	1.46 Å	A	119-124	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6SUZ	X-ray	2.50 Å	A	125-223	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6SV2	X-ray	2.30 Å	A	119-231	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	6UUR	EM	3.50 Å	A/B/C/D/E/F/G/H/I/J	94-178	PDBe · RCSB-PDB · PDBj · PDBsum
PDB	7DWV	EM	3.07 Å	A/B/C/D/E/F	23-231	PDBe · RCSB-PDB · PDBj · PDBsum
AlphaFold	AF-P04156-F1	Predicted			1-253	AlphaFold

Features

Showing features for beta strand¹, turn¹, helix¹.

AlphaFoldの構造は全長をカバーしているが、
N末側の天然変性領域の部分の信頼性は低い

講義の内容

1. タンパク質構造予測の技術
2. AlphaFold2、ColabFold
3. ノイラミニダーゼの構造予測 (課題1)
4. プリオンタンパク質の構造予測 (課題2)
5. Ras-Raf複合体の構造予測 (課題3)

【参考資料】 Swiss-Model、HHpred、Phyre2、
二次構造予測、天然変性予測など

実習の資料について

- 「バイオインフォマティクス 実習」で検索

– <https://lecture.ecc.u-tokyo.ac.jp/~ashimizu/>

ID: bioinfo
パスワード: 5455

Google search results for "バイオインフォマティクス 実習". The search term is circled in red. The first result is from <https://lecture.ecc.u-tokyo.ac.jp/~ashimizu/> and is circled in red. The title of the result is "バイオインフォマティクス実習 資料 (清水謙多郎)". Below the title, there is a snippet of text: "最新の「バイオインフォマティクス実習」の資料は、こちらにアップして...".

Content from the lecture page. The title "バイオインフォマティクス実習 資料" is circled in red. Below it, there is a paragraph: "以下の資料は、平成20年度以前の「バイオインフォマティクス実習」の資料です。最新版の「バイオインフォマティクス実習」の資料は、こちらにアップしています。詳しい内容は、講義のスライド資料をご覧ください。". Below this, there are several links: "生物データベース概論", "ゲノム情報解析とプログラミング", "ホモロジー検索", "タンパク質の配列から機能を予測する", "Rを用いた統計データ処理の基礎", "Rを用いたバイオインフォマティクス解析", and "タンパク質構造のバイオインフォマティクス". The last link is circled in red. A green callout bubble contains the ID and password information.

ColabFoldを用いて、以下のタンパク質の構造予測を実行せよ。

1. H1N1ノイラミニダーゼ
2. ヒトのプリオンタンパク質
3. Ras-Raf複合体

それぞれについて、ランク1位およびランク2位のモデル構造のグラフィックス画像、pLDDT値、結晶構造とのRMSD値を提出せよ。これらの回答は、PowerPointもしくはWordのファイルに貼り付けてITC-LMSの課題として提出すること。

出席の確認と課題の提出

- 出席は、課題の提出に代えます。
- 課題(解答のファイル)は、ITC-LMSから提出して下さい。
- 授業は20:30に終了します。
- 課題の締め切りは、5月9日(月)24:00です。

Assignments

Predict the structures of the following proteins using ColabFold.

1. H1N1 neuraminidase
2. Human prion protein
3. Ras-Raf complex

For rank 1 and rank 2 model structures, submit graphics images of the models, the pLDDT values, and the RMSD values with respect to the crystal structure. The answers should be uploaded to ITC-LMS as a PowerPoint or Word file.

Attendance check and assignments submission

- Attendance is taken by completing the in-class assignments.
- Take in the assignments (an answer file) to the ITC-LMS site.
- The class will end at 20:30.
- You should submit the assignments at 24:00 on May 9.

タンパク質の立体構造予測手法

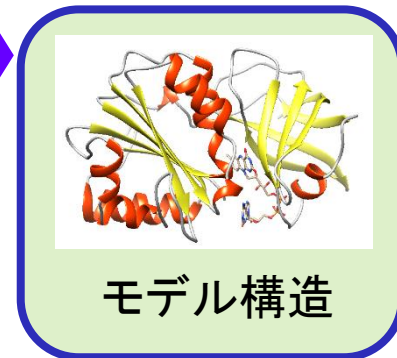
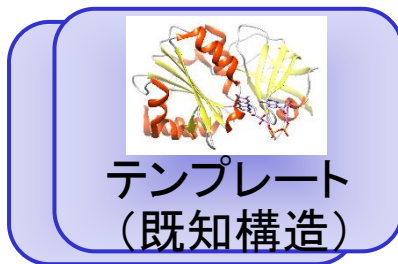
入力

予測手法

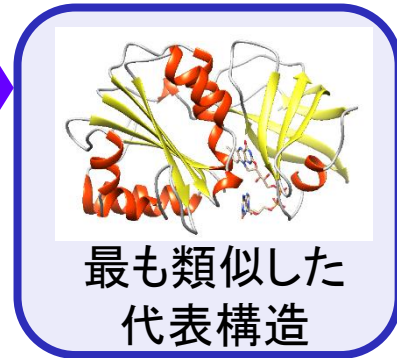
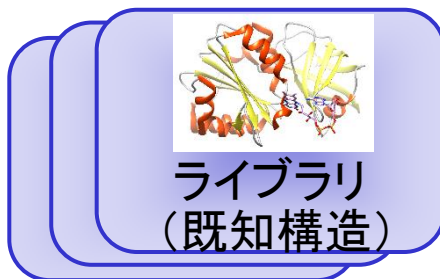
出力

アミノ酸配列
FYGELVCAGFRCLC...
(予測対象のタンパク質)

ホモロジーモデリング

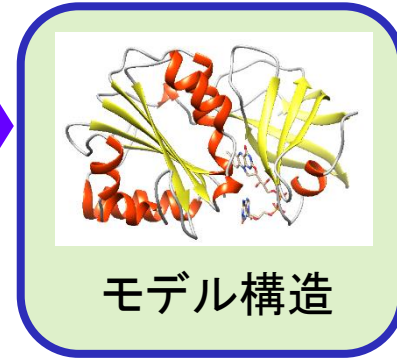


フォールド認識



ab initio法

(新規にモデルを構築)



ホモロジーモデリング

- ホモロジーモデリング (homolog modeling): 構造未知のタンパク質 (ターゲット) の構造を、構造既知のタンパク質の中からテンプレートとなる構造を選んで、それをもとに予測
 - 配列が似ていれば、構造も似ている傾向
- 主なシステム
 - Modeller, <https://salilab.org/modeller/>
 - SWISS-MODEL, <https://swissmodel.expasy.org/>
 - HHpred, <https://toolkit.tuebingen.mpg.de/tools/hhpred>
 - ESyPred3D, GENO3Dなど

ホモロジーモデリングの例

テンプレートの配列 (N8ノイラミニダーゼ)

```
TYMNNTEAICDAKGFAPFVKDNGIRIGSRGHIFVIREPFVSCSPIECRT  
FFLTQGSLLNDKHSNGTVKDRSPFRTLMSVEVGQSPNVYQARFEAVAWS  
ATACHDGKKWMTVGVTGPDSKAVAVIHYGGVPTDVVNSWAGDILRTQES  
SCTCIQGDCYWVMTDGPANRQAQYRIYKANQGRIIGQTDISFNGGHIIE  
CSCYPNDGKVECVCRDGTNRPVLVISPDLRYRVGYLCAGIPSDTPR  
GEDTQFTGSC TSPMGNQGYGVKGFGRQGTDVWVMGRTISRTRSFGFEIL  
RIKNGWTQTSKEQIRKQVVVDNLNWSGYSGSFTLPVELSGKDCLVPCFW  
VEMIRGKPEEKTIWTSSSSIVMCGVDYEVADWSWHDGAILPFDIDKM
```

ターゲットの配列 (H1N1ノイラミニダーゼ)

```
VILTGNSSLCPIISGWAIYSKDNGIRIGSKGDVVFVIREPFVSCSHLECRTF  
FLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWSASA  
ACHDGMGWLTIIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQES  
CVNGSCFTIMTDGSPNGQASYKILKIEKGKVTKSIELNAPNYHYEECS  
CYPDGTGKVMCVCRDNWHGNSRNPWVSFDQNLDYQIGYICSGVFGDNPR  
NDGTGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWD  
PNGWTE TDSSFSVRQDIVAITDWSGYSGSFVQHPPELTGLDCMRPCFW  
VELIRGQPKENTIWTSGSSISFCGVNSDTVGSWPDGAELPFSI
```

ターゲットの配列とテンプレートの配列のアラインメント

配列一致度は56.8%

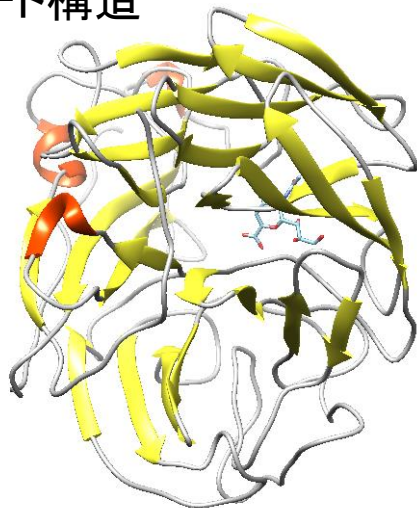
>target

```
VILTGNSSLCPIISGWAIYSKDNGIRIGSKGDVVFVIREPFVSCSHLECRTF  
FFLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWSASA  
CHDGMGWLTIIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQES  
CACVNGSCFTIMTDGSPNGQASYKILKIEKGKVTKSIELNAPNYHYEE  
CSCYPDGTGKVMCVCRDNWHGNSRNPWVSFDQNLDYQIGYICSGVFGDN  
PRNDG--TGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGF  
EMIWDPNGWTETDSSFSVRQDIVAITDWSGYSGSFVQHPPELTGLDC  
MRPCFWVELIRGQPKENTIWTSGSSISFCGVNSDTVGSWPDGAELPFSI---
```

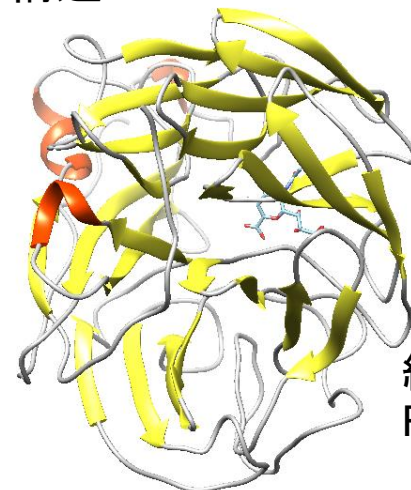
>template

```
TYMNNTEAICDAKGFAPFVKDNGIRIGSRGHIFVIREPFVSCSPIECRT  
FFLTQGSLLNDKHSNGTVKDRSPFRTLMSVEVGQSPNVYQARFEAVAWS  
ATACHDGKKWMTVGVTGPDSKAVAVIHYGGVPTDVVNSWAGDILRTQES  
SCTCIQGDCYWVMTDGPANRQAQYRIYKANQGRIIGQTDISFNGGHIIE  
CSCYPNDGKVECVCRDGTNRPVLVISPDLRYRVGYLCAGIPSDTPR  
GEDTQFTGSC TSPMGNQGYGVKGFGRQGTDVWVMGRTISRTRSFGFEIL  
RIKNGWTQTSKEQIRKQVVVDNLNWSGYSGSFTLPVELSGKDCLVPCFW  
VEMIRGKPEEKTIWTSSSSIVMCGVDYEVADWSWHDGAILPFDIDKM
```

テンプレート構造



モデル構造



結晶構造との
RMSD 1.051Å

ホモロジーモデリングの手順

1. テンプレート(鋳型)の選択

- 構造データベースに登録されているタンパク質の中からテンプレートの構造を選ぶ
 - BLAST、プロファイル、隠れマルコフモデル、配列-構造アラインメント

2. ターゲットとテンプレートとのアラインメント

- ターゲットのアミノ酸配列とテンプレート構造とのアラインメントを求める
- 場合によっては複数のテンプレートを使用する
 - 手法は1.と共通するが、局所的な対応関係、共通の機能部位を重視

3. モデル構造の構築

- テンプレート構造の情報をもとに、ターゲットのモデル構造を構築する

4. モデル構造の評価

- 構築したモデル構造が正しい構造かどうかを評価する

Modeller

- Saliらが開発したホモロジーモデリングツール
 - Windows、Mac、Unix系OSで利用可能なスタンドアロンソフトウェア
- コマンドラインインタフェースが基本
 - GUIは、EasyModeller(フリー)、Chimera(フリー)、DS Modeling(商用)など
 - <https://salilab.org/modeller/>

Modeller

Program for Comparative Protein
Structure Modelling by Satisfaction
of Spatial Restraints



```
AI LVGSMRRDGMERKDLLKANVKIFKCOGA  
SEVCPYDGGYEGPHLYHPDECIDGALCEP  
SACRPPKPPKGGD--LQALDAKGGDQCEP  
G--LAGGACKPECPVNIQQG--LYAIADDS
```

最新版は10.2

About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is [available for download](#) for most Unix/Linux systems, Windows, and Mac.

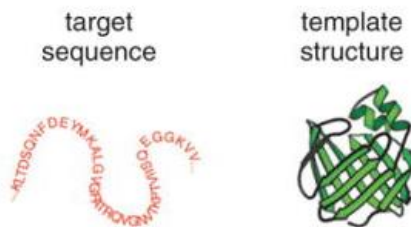
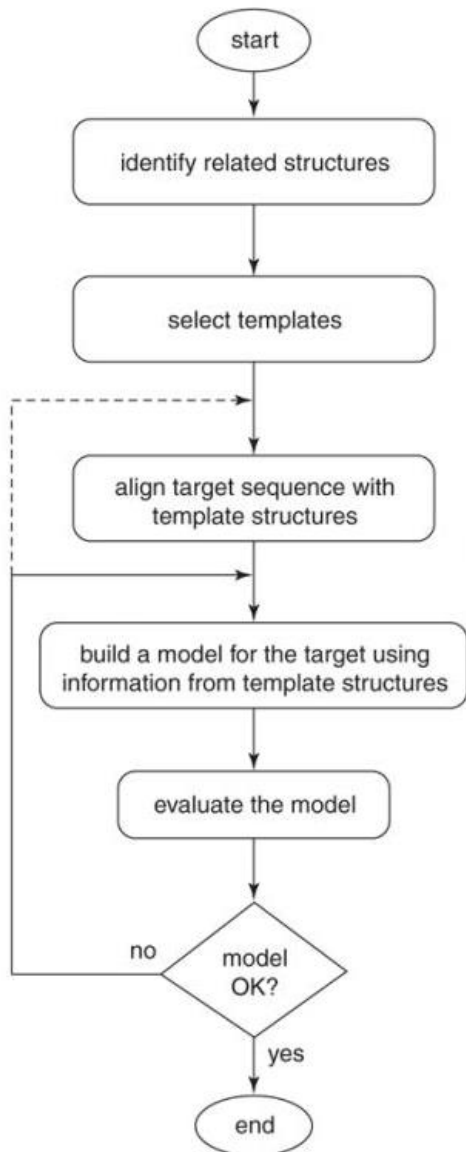
Several graphical interfaces to MODELLER are [commercially available](#). There are also many other [resources and people using Modeller](#) in graphical or web interfaces or other frameworks.

1. B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.
2. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.
4. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

The current release of Modeller is **10.2**, which was released on Nov 30th, 2021. Modeller is currently maintained by [Ben Webb](#).

About MODELLER
MODELLER News
Download & Installation
Release Notes
Data file downloads
Registration
Non-academic use
Discussion Forum
Subscribe
Browse archives
Search archives
Documentation
FAQ
Tutorial
Online manual
Wiki
Developers' Pages
Contact Us

Modellerの基本手順



テンプレートとターゲットのアラインメント

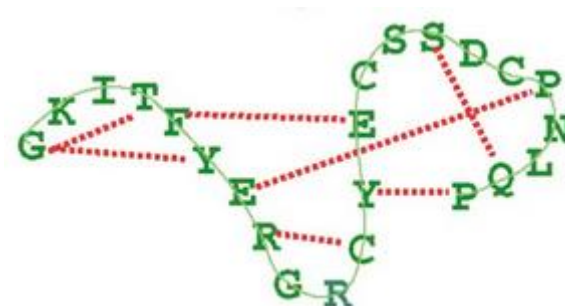
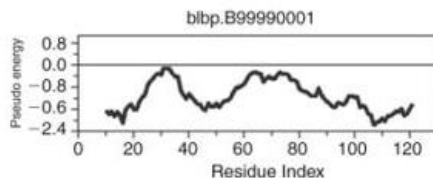
alignment

3D **GKITYERGFQGH CYESDC-----**
 SEQ **GKITYERG---RCYESDCPNLQP**

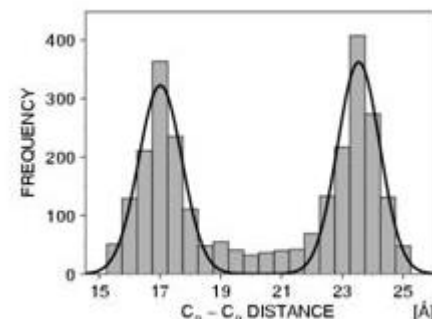
構造アラインメントやデータベースから距離、二面角などの拘束条件を抽出



様々な拘束条件を合わせて、最小化を行う



テンプレートの構造群の拘束条件をターゲットに適用



$C_{\alpha}-C_{\alpha}$ 間の距離の例

実際の分布をガウス関数で近似ある距離をとる条件確率を求め、エネルギー値に変換

SWISS-MODEL

- ExPASyサーバから利用できるホモロジーモデリングツール
- <https://swissmodel.expasy.org/>

SWISS-MODEL

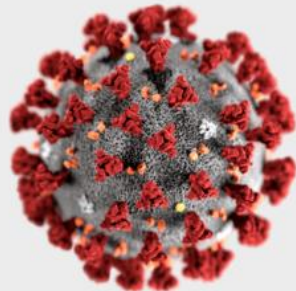
is a fully automated protein structure homology-modelling server, accessible via the **Expasy web server**, or from the program DeepView (Swiss Pdb-Viewer).

The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

[Start Modelling](#)

Repository

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?



SARS-CoV-2

Severe acute respiratory syndrome coronavirus 2, is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19).

We modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence [NC_045512](#) and annotations from [UniProt](#).

The results are available [here](#).

SWISS-MODELの基本手順

ターゲットの配列

テンプレートのアノテーション

InterProによるドメインスキャン
PSIPredによる二次構造予測
DISOPREDIによる天然変性領域予測
MEMSATによる膜貫通領域予測

テンプレートの選択

配列類似度が高いときはBLAST、
そうでないときは、PSI-BLAST、HMM-HMMIに基づく
探索 (HHblits) を行い、1つまたは複数のテンプレート
を選択

テンプレート

テンプレートの調整

マルチプルアライ
メントの利用

DeepView Project

ターゲットと複数テンプレートのマルチプル
アライメント
複数テンプレートの重ね合わせ、代表テン
プレートから外れたテンプレートは除外

異なる手法による複数の
アライメント候補の提示
と人手による修正

モデル構造の構築

Alignment
mode

Automated
mode

Project
mode

テンプレートの主鎖の平均構造をあてはめる
(配列類似度に応じて重みづけ)、フラグメン
トベースでループ構造をあてはめる
GROMOSを使ったエネルギー最小化

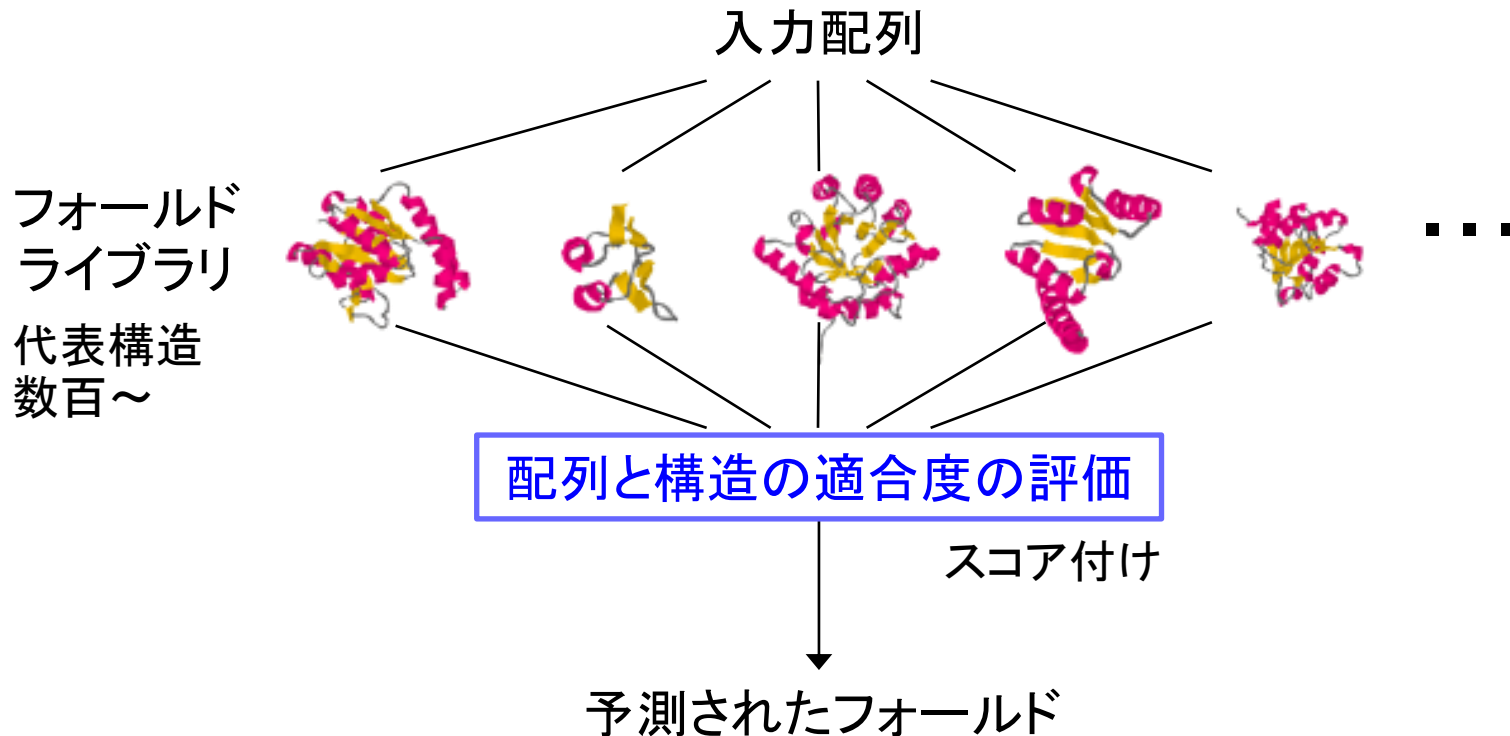
モデル構造とレポート

モデル構造の評価

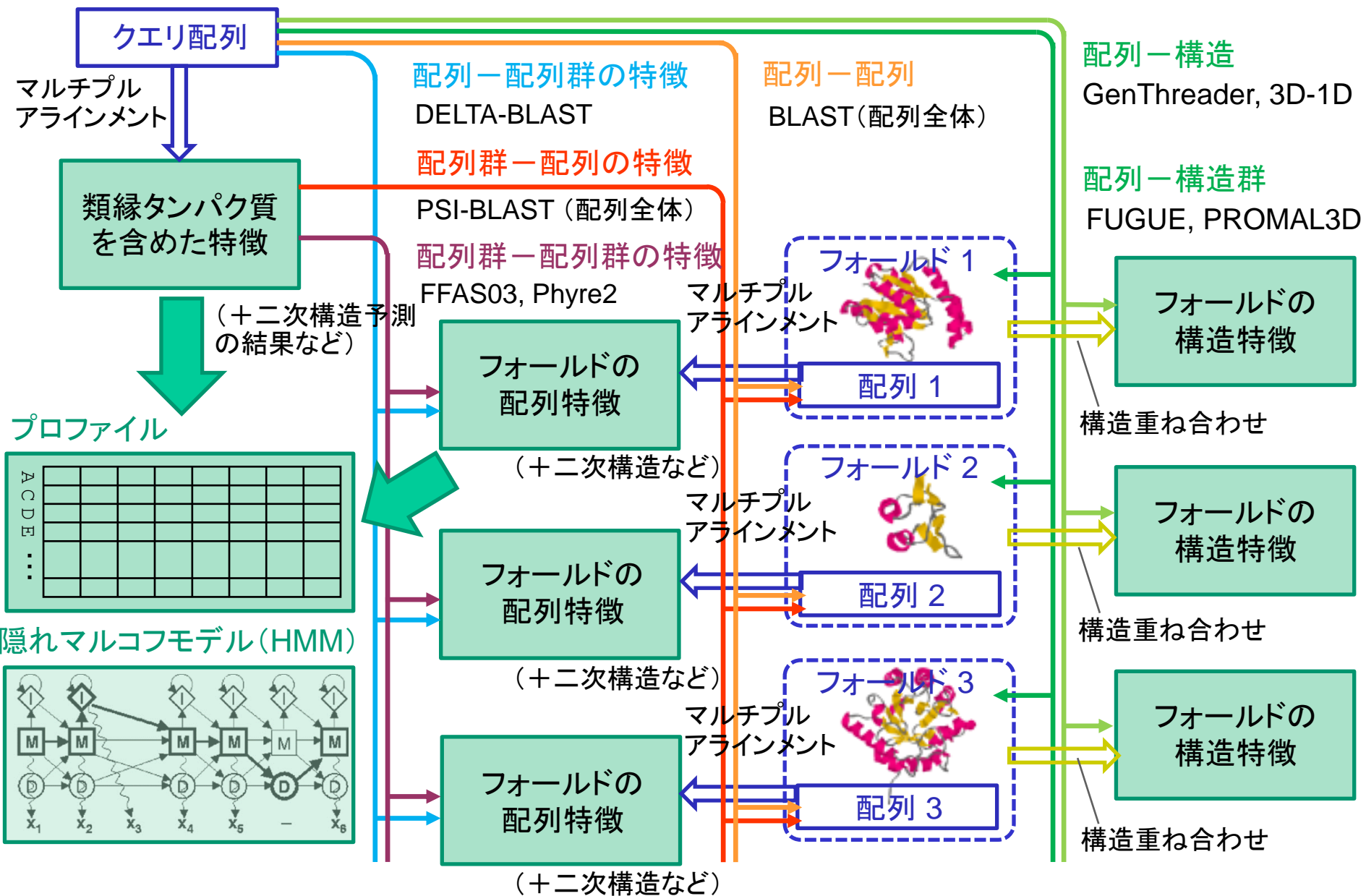
ターゲットとテンプレートの配列一致度
立体化学のチェック (WHATCHECK, PROCHEK)
グローバルモデルクオリティチェック (DFIRE)
ローカルグローバルモデルクオリティチェック (ProQres,
ANOLEAスコアなど)

フォールド認識

- 代表的なタンパク質の構造(フォールドの代表)をライブラリ化し、ターゲットの配列がどの構造に近いかを予測
- アミノ酸配列と構造との適合性を評価するスコア関数を定義し、その値によって、ターゲットに最も近い構造を見つける
 - 配列一致度が小さくてもターゲットに近い構造が得られる場合がある
- ホモロジーモデリングにおけるテンプレート探し、ターゲットとテンプレートのアラインメントの精密化に利用



配列と構造の適合度の評価



ab initio法

- ab initio法
 - 既知の構造によらず、一からモデル構造を構築
 - 新規フォールドの予測を可能とする
 - 原子レベルで伸展構造から物理エネルギーの最小化で構造モデリング
 - 構造の探索が容易ではない
 - 多大な時間を要する
 - 小さなタンパク質にしか適用できない
 - フォールディング過程の解析が目的
- ➡
- 探索の場合の数を減らす
 - エネルギー計算の効率化

ab initio法

- モデル構造の構築の効率化
 - 格子法
 - タンパク質の構造を格子で近似
 - 相互作用も隣接する頂点間などに限定
 - フラグメントアセンブリ
 - データベースに登録された局所構造(フラグメント)を組み上げてモデル構造を構築し、統計ポテンシャルで選別
 - 共変異情報をもとにコンタクトマップ(残基間距離拘束)を作成
 - 候補構造の選択、精密化において、エネルギー最小化が適用される

ab initio法

- エネルギー計算

- ポテンシャル関数

- 物理ポテンシャル → Amber, CHARMM, GROMOS, OPLSなど
 - 統計ポテンシャル → 既知のタンパク質構造から特定の構造特徴をとる傾向を統計的に計算し、ポテンシャルとして定義

- エネルギー関数の最小化計算

- 大域的な最小値を求めるのは困難
 - Simulated Annealingなどのヒューリスティックな方法
 - 多数の構造サンプリングを行い、そこから選ぶ
→ 分子動力学法(MD)、MDで得られる構造の効率的なサンプリング

統計ポテンシャル

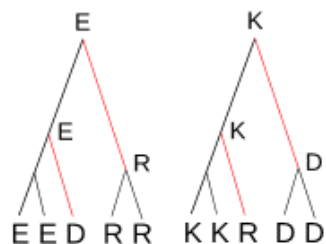
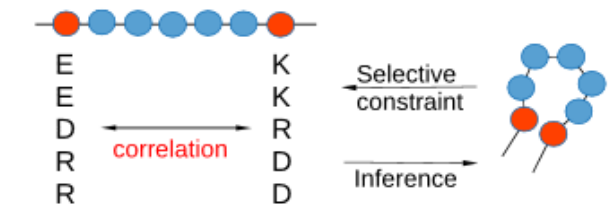
- 構造データベース (PDB) の内容から特定の構造特徴をとる傾向を確率統計的に計算し、ポテンシャルとして定義したもの
 - empirical potential, database-derived potential, knowledge-based potential, pseudo potential
 - 残基間の距離、残基の埋もれ度、回転半径などのタンパク質の構造特徴に対して、例えば、それらのとりやすさを確率の $-\log$ をとってポテンシャルの形に表したもの

- タンパク質の系において、配列 a のもとで構造 c をとる状態は、あるエネルギー関数 $E(a, c)$ のもとでのカノニカルアンサンブルになっていると考えられる (Sippl, *J. Mol. Biol.*, 1990)
 - 粒子数 N 、体積 V 、温度 T が一定のアンサンブル
 - エネルギー E の状態 x が出現する確率 $p(x)$
$$p(x) = \frac{1}{Z} \exp[-E(x)/kT] \quad \rightarrow \quad E(x) = -kT \log(p(x)) - kT \log Z \quad (Z \text{ は分配関数})$$

$p(x)$ がわかれば、 $E(x)$ を見積もることができる

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

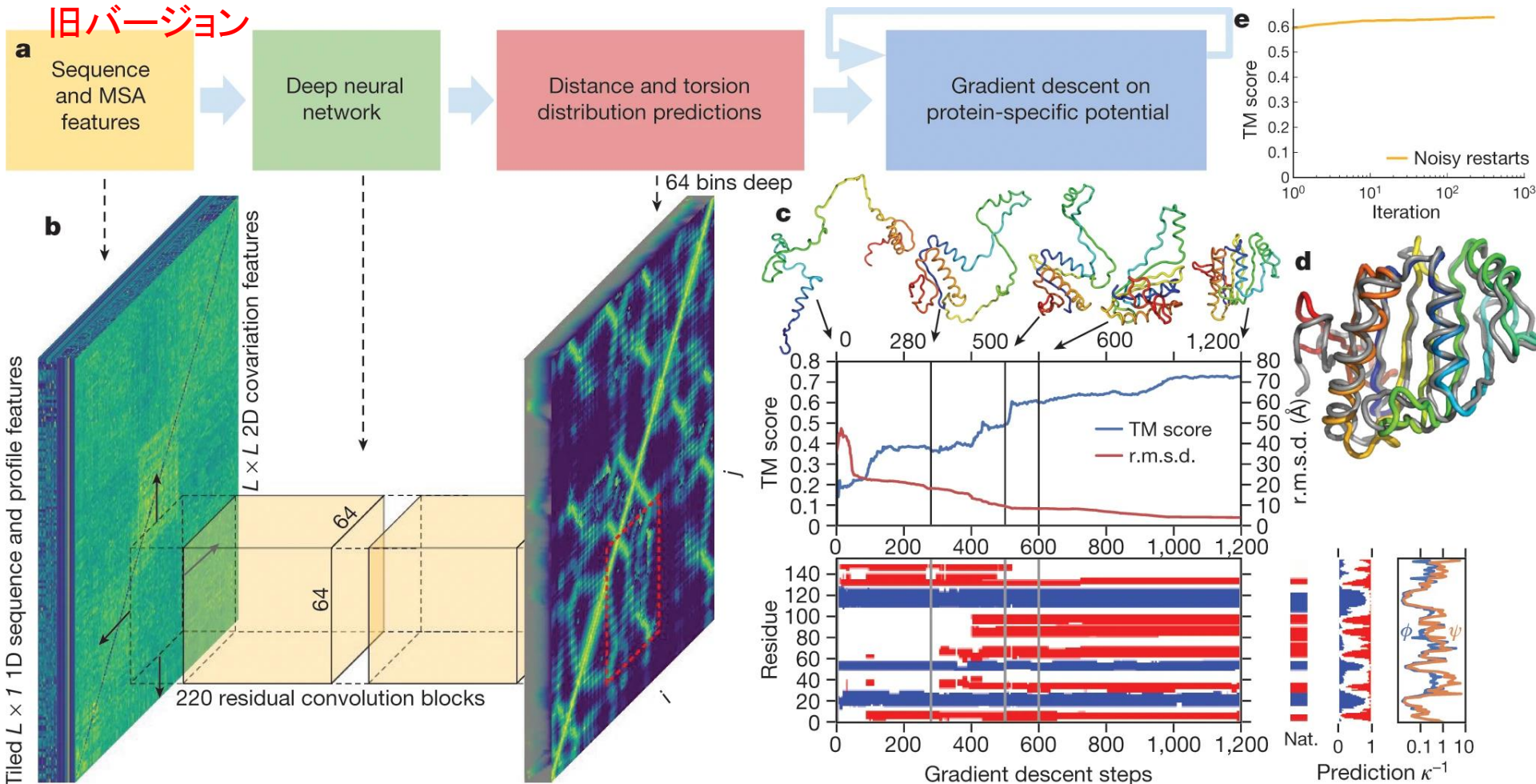


Compensatory
Co-substitution

宮澤三造, 生物物理54, 091-095 (2014)

深層学習により特徴量を抽出
コンタクト予測を行う

AlphaFoldの手法

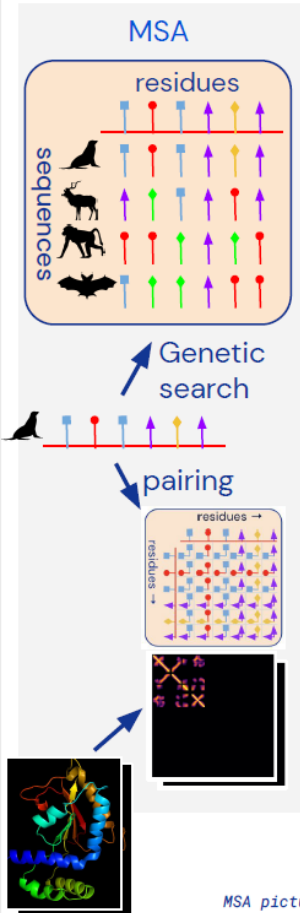


- 多数のMSAをもとにした学習
- 64残基×64残基の領域(クロップ)のそれぞれの予測結果を集約
- 残基間距離2~22Åを64に離散化したbinごとの確率分布を予測
- タンパク質固有のポテンシャルを最急降下法で最適化

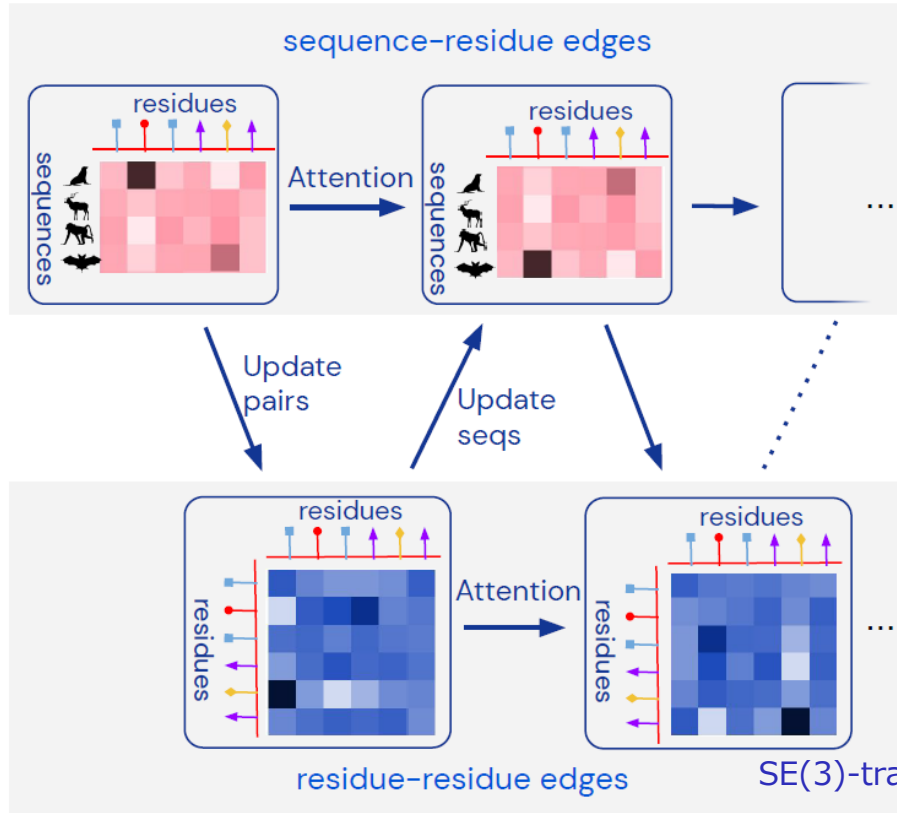
A. W. Senior, et al. Nature, 577, 706–710 (2020).

AlphaFold2の手法

Embedding

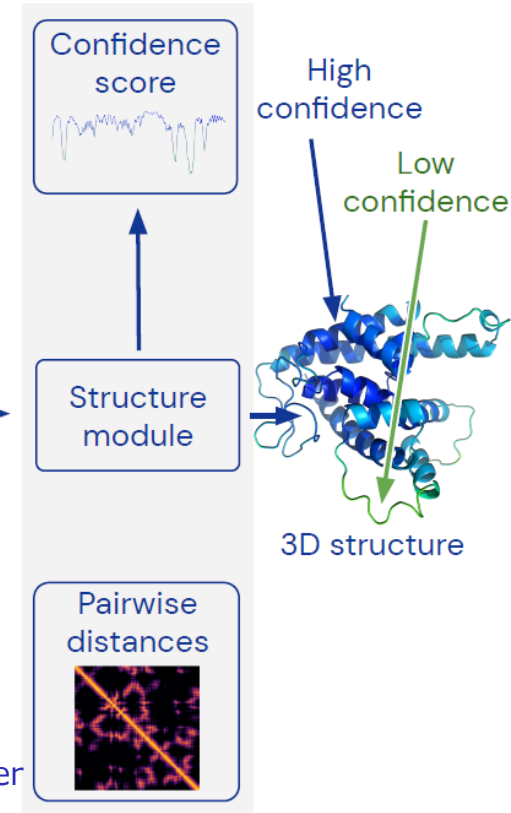


Trunk



Heads

© 2020 DeepMind Technologies Limited



MSA picture inspired by: Rieselmann, A.J., Ingraham, J.B. & Marks, D.S.,
Nature Methods (2018) doi:10.1038/s41592-018-0138-4



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

ColabFold

- AlphaFold2をGoogle Colab notebookとして実行するプログラム
 - Colaboratory (Colab) は、Google Research が提供するPythonのプログラム開発環境で、ブラウザ上でPythonを記述、実行することができる
 - Googleアカウントとブラウザ (Chrome推奨) が必要
- オープンソースソフトウェアとして利用可能
<https://github.com/sokrypton/ColabFold>
- AlphaFold2を実行
<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

ColabFoldの実行環境

- ノートブック (notebook): Pythonプログラムを入力・実行したり、テキストやコメントを記入したりするための作業場。ファイルとして保存した場合の拡張子は.ipynb
- Googleドライブの「マイドライブ」の下に「Colab Notebooks」というフォルダが作られる
- セル: ノートブックの中の入力の単位、ノートブックはコードセル (Pythonコード用) とテキストセルで構成される
- ランタイム: コードセルのPythonコードを解釈・実行するためにサーバー側で動作しているプログラム

ColabFold

Making Protein folding accessible to all via Google Colab!

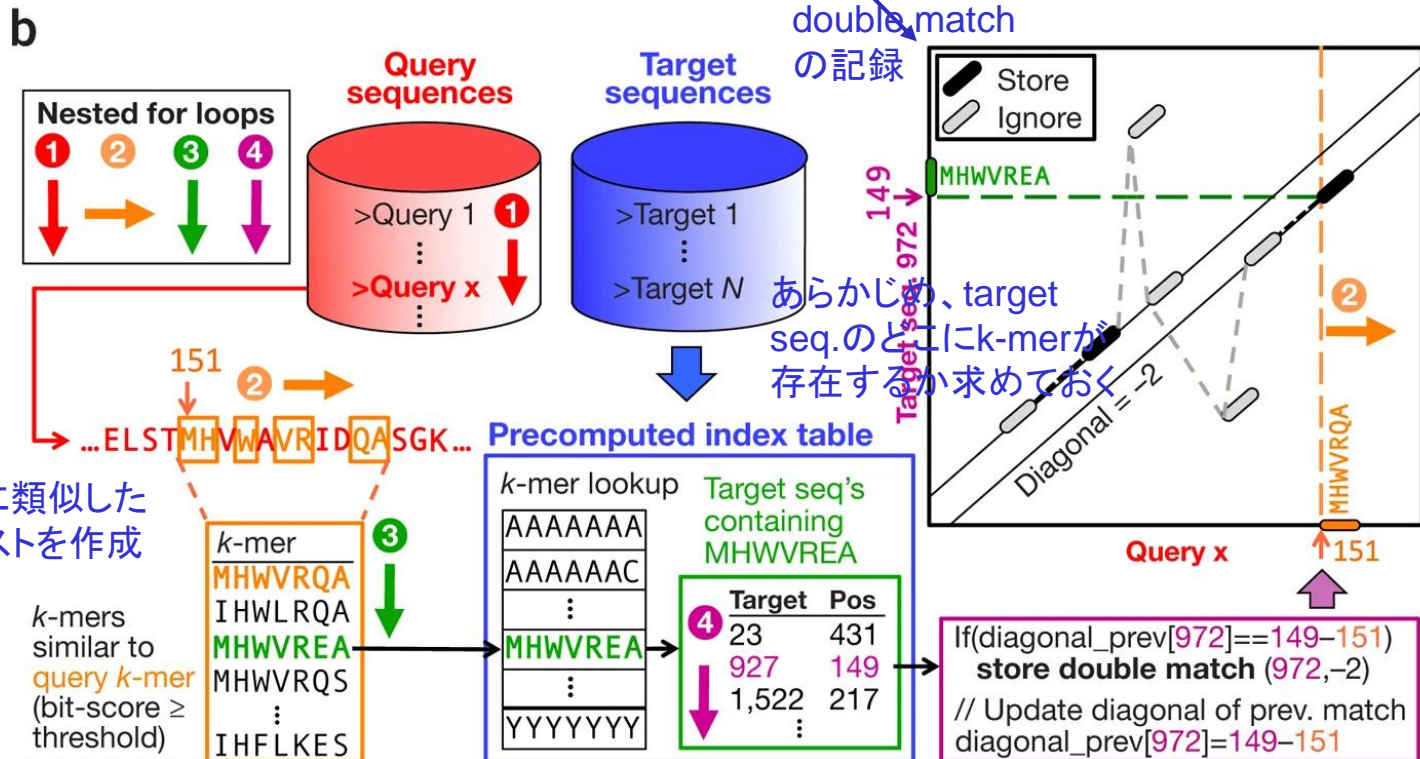
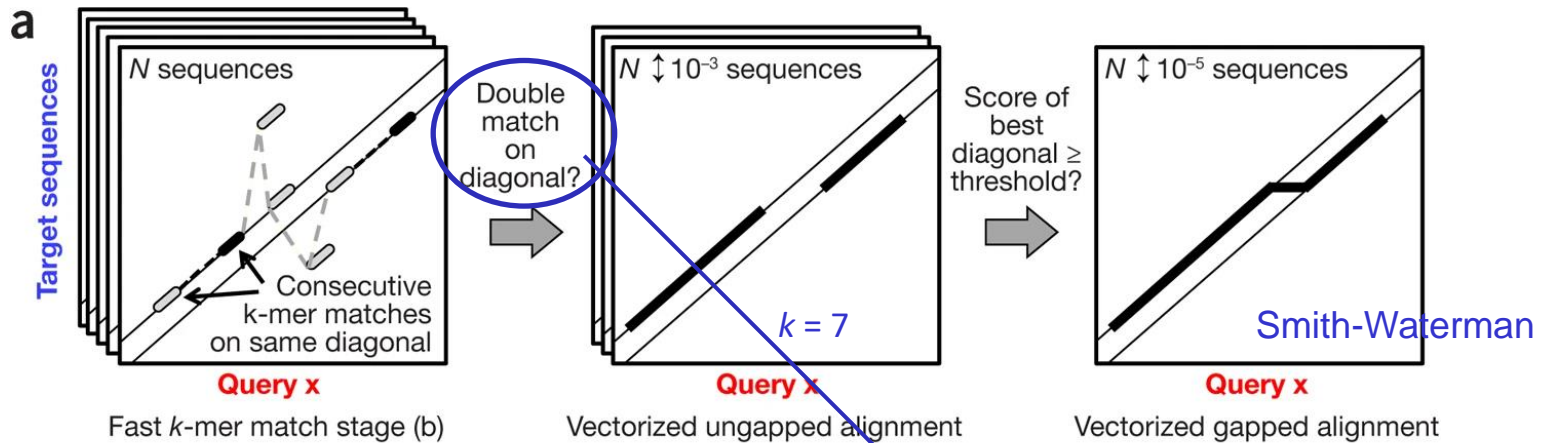
Notebooks	monomers	complexes	mmseqs2	jackhmmer	templates
AlphaFold2_mmseqs2	Yes	Yes	Yes	No	Yes
AlphaFold2_batch	Yes	Yes	Yes	No	Yes
RoseTTAFold	Yes	No	Yes	No	No
AlphaFold2 (from Deepmind)	Yes	Yes	No	Yes	No
BETA (in development) notebooks					
AlphaFold2_advanced	Yes	Yes	Yes	Yes	No
OLD retired notebooks					
AlphaFold2_complexes	No	Yes	No	No	No
AlphaFold2_jackhmmer	Yes	No	Yes	Yes	No
AlphaFold2_noTemplates_noMD					
AlphaFold2_noTemplates_yesMD					

MMseqs2

- MMseqs2 (Many-against-Many sequence searching): 多数のアミノ酸配列、塩基配列の検索とクラスタリングを行うソフトウェア群
- マルチコアで並列計算、優れたスケーラビリティをもつ
- BLASTの10000倍の速度で実行、100倍の速度でも、ほぼ同じ感度を実現
 - PSI-BLASTと同じ感度のプロフィール検索を400倍以上の速度で行うことが可能

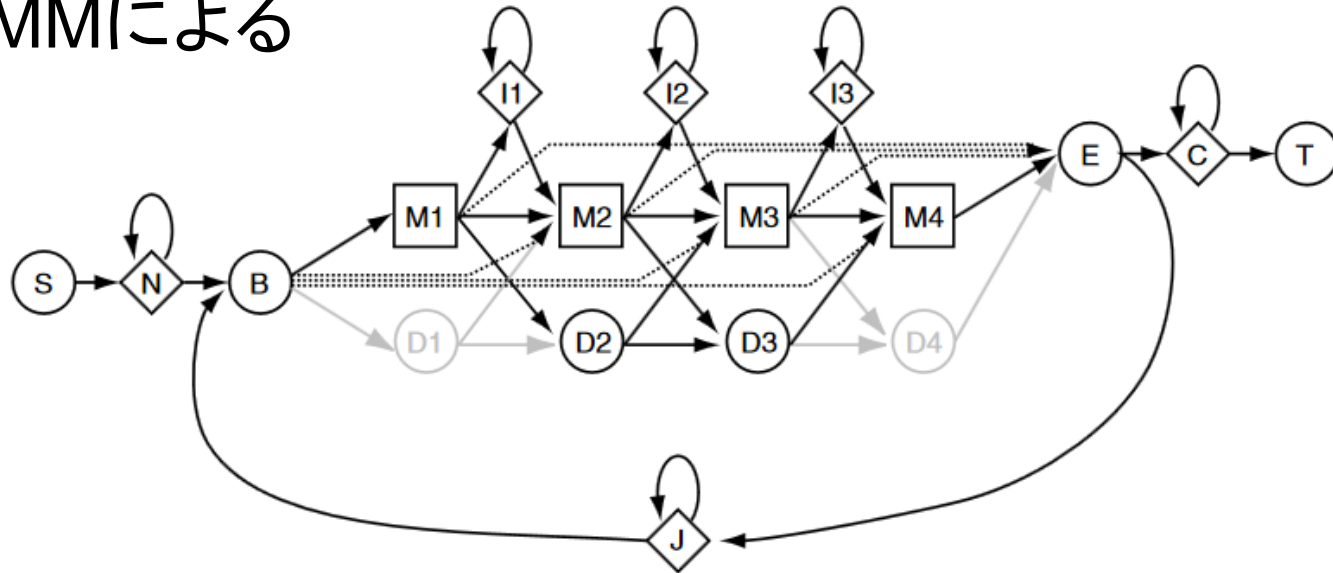
[Steinegger M and Soeding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology, doi: 10.1038/nbt.3988 \(2017\).](#)
[Steinegger M and Soeding J. Clustering huge protein sequence sets in linear time. Nature Communications, doi: 10.1038/s41467-018-04964-5 \(2018\).](#)
[Mirdita M, Steinegger M and Soeding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics, doi: 10.1093/bioinformatics/bty1057 \(2019\).](#)
[Mirdita M, et al.: Fast and sensitive taxonomic assignment to metagenomic contigs. Bioinformatics, doi: 10.1093/bioinformatics/btab184 \(2021\).](#)

MMseqs2



HMMER

- HMMER: 隠れマルコフモデル (HMM, hidden Markov model) を用いて、配列データベースに対して検索を行い、アラインメントを求めるためのソフトウェア
- 挿入、欠失を含む柔軟な配列パターンを表すプロフィールHMMによる



M_x : 長さLのモチーフに対してL個のM状態をもつ。モチーフのx番目の文字に一致。各文字の出力確率が定義される。
 I_x : x番目の文字の後に挿入があることを示す。バックグラウンドの出力確率に応じて文字を出力。
 D_x : x番目の文字が削除されたことを示す。文字は出力しない。
S: 開始状態。文字を出力しない。
N: N末端のアラインメントされていない状態。遷移時に出力。

B: モチーフの開始状態。文字を出力しない。
E: モチーフの終了状態。文字を出力しない。
C: C末端のアラインメントされていない状態。遷移時に出力。
J: モチーフを連結する状態。遷移時に出力。
T: 終了状態。

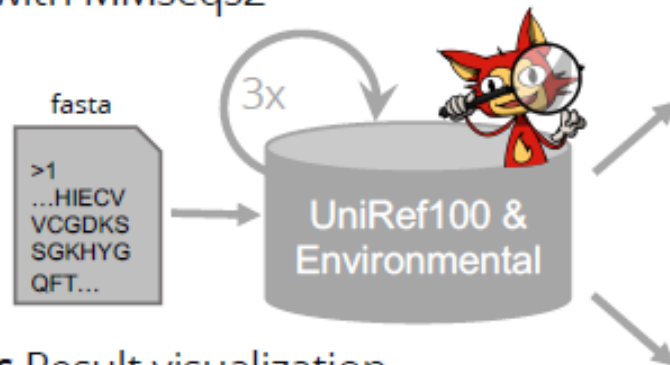
Sun Y., Buhler J. Designing patterns for profile HMM search Bioinformatics (2007) 23: e36–e43.

HMMを用いた探索

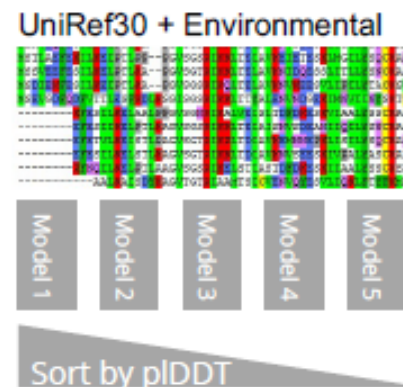
- phmmer: アミノ酸配列データベースに対して、1つの配列を探索 (BLASTPのような使用法)
- jackhmmer: アミノ酸配列データベースに対して、1つの配列を繰り返し探索 (PSI-BLASTのような使用法)
- hmmsearch: アミノ酸配列データベースに対して、プロファイルHMMを探索
- hmmscan: プロファイルHMMに対してアミノ酸配列を探索

ColabFoldの手法

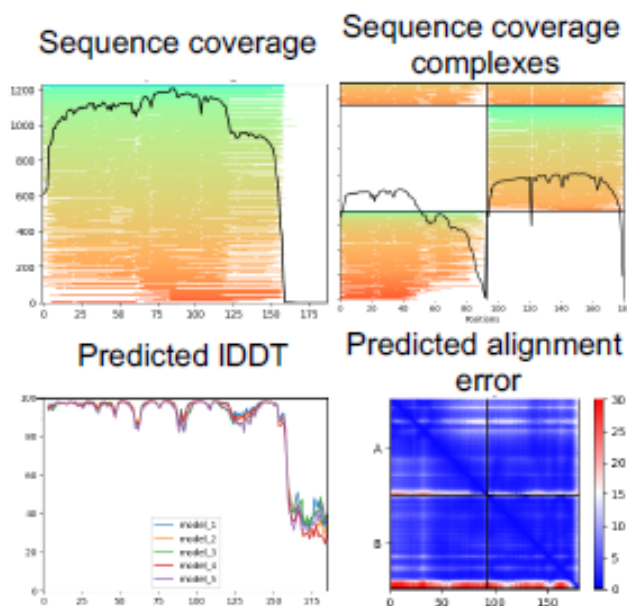
a Multiple sequence alignment with MMseqs2



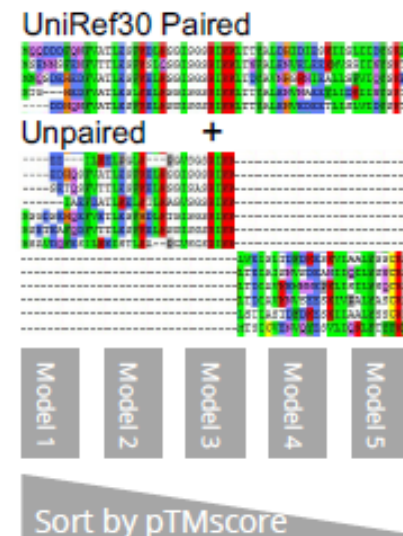
b1 Structure prediction



c Result visualization



b2 Complex prediction



Environmental: BFDおよびMGnifyメタゲノムデータベースをMMseqs2とクラスタリングを用いて冗長性を排除し、数を減らしたもの

UniRef100: UniProtKBの11残基以上の配列で100%一致しているものをクラスタリング

UniRef90: UniRef100の配列をクラスタリング、クラスターの中の最長のものと90%一致し、80%以上重なっているものからなるクラスターを形成、代表はアノテーションの質、モデル生物などを優先

M. Mirdita, et al. *BioRxiv*, 10.1101/2021.08.15.456425 (2021).

使い方の説明

AlphaFold2の利用

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

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv, 2021](#)



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

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

```
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" ← ジョブ名、結果のファイルの名前などに反映される

use_amber: ← Amber力場を用いて構造を緩和

template_mode: none ← none: テンプレートを使用しない

"none" = no template information is used, **pdb70** = detect templates in **pdb70**, **custom** = upload and search own templates (more format)

custom: 利用者がPDB形式のファイルで指定

MSA options (custom MSA upload, single sequence, pairing mode)

```
msa_mode: MMseqs2 (UniRef+Environmental) ← マルチプルアラインメント (MSA) のモード
```

```
pair_mode: unpaired+paired ← MSAは、“unpaired”と“paired”の両方がデフォルト
```

"unpaired+paired" = pair sequences from **single species + unpaired MSA**, **unpaired** = separate MSA for each chain, **paired** - only use paired sequences.

paired: 配列ペアのMSA

AlphaFold2の利用

AlphaFold2.ipynb

共有

+ コード + テキスト ドライブにコピー

▶ **Advanced settings**

model_type: auto ← autoがデフォルト(単体はAlphaFold2-ptm、複合体はAlphaFold2-multimer-v2)

num_recycles: 3

ptm: predicted TMスコア、AlphaFoldに組み込みの性能指標

AlphaFold2-ptm: 単体用

AlphaFold2-multimer-v2: 複合体用

• "auto" = protein structure prediction using "AlphaFold2-ptm" and complex prediction "AlphaFold-multimer-v2". For complexes "AlphaFold-multimer-v[1.2]" and "AlphaFold-ptm" can be used.

save_to_google_drive ← 結果をGoogleドライブに保存する → チェックを入れた方がよい

• if the save_to_google_drive option was selected, the result zip will be uploaded to your Google Drive

Don't forget to hit Runtime -> Run all after updating the form.

コードの表示

▶ **Install dependencies** 使用するパッケージのインストールなど

コードの表示

▶ **Run Prediction**

コードの表示

▶ **Display 3D structure**

rank_num: 1 ← 1~5が選択可能

color: IDDT ← chain: チェインごとの色づけ

show_sidechains:

show_mainchains:

IDDT: pLDDT値(モデルの質を表す)、デフォルト rainbow: N末からC末のレインボー表示

H1N1ノイラミニダーゼの構造予測(1)

- ターゲット: H1N1ノイラミニダーゼ(タンパク質p1)
 - 講義のページ
 - アミノ酸配列は「[p1.fasta](#)」
 - 正解構造: [PDB ID: 3b7e](#) → 構造が分からないと仮定
 - 4量体であることが知られているが、ここでは単体として予測
- ColabFoldで実行

H1N1ノイラミニダーゼの構造予測(2)

AlphaFold2.ipynb

共有 編集

RAM ディスク

ドライブにコピー

「ドライブにコピー」を指定

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv, 2021](#)



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

```
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 "
```

```
use_amber: 
```

```
template_mode: none
```

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (mmCIF format)

[コードの表示](#)

✓ [2] MSA options (custom MSA upload, single sequence, pairing mode)

```
msa_mode: MMseqs2 (UniRef+Environmental)
```

```
pair_mode: unpaired+paired
```

- "unpaired+paired" - pair sequences from same species. Unpaired MSA. "unpaired" - generate MSA for each chain. "paired" - only use paired sequences.

H1N1ノイラミニダーゼの構造予測(3)



AlphaFold2.ipynb のコピー ☆ 「のコピー」は削除しておく



+ コード + テキスト

接続 編集

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv, 2021](#)



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

```
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"
```

```
use_amber: 
```

```
template_mode: none
```

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (mmCIF format)

[コードの表示](#)

▶ MSA options (custom MSA upload, single sequence, pairing mode)

```
msa_mode: MMseqs2 (UniRef+Environmental)
```

```
pair_mode: unpaired+paired
```

- "unpaired+paired" - pair sequences from same species + unpaired MSA. "unpaired" - separate MSA for each chain. "paired" - only use paired sequences

H1N1ノイラミニダーゼの構造予測(4)

+ コード + テキスト

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv. 2021](#)



ターゲットの配列を入力(配列の部分だけ入力する)

Input protein sequence(s), then hit Runtime -> Run a [p1.fasta](#)

query_sequence: VILTGNSLCPISGWAIYSKDNIGRIGSKGDVFIREFPFISCSHLECRFTFFLTQGALLNDKHSNGTVKDRSPYRTLMSCP VGEAPSPYNSRFESVAWSASACHDGI

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

jobname: p1 ジョブ名を入力

use_amber:

template_mode: none

- "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (mmCIF format)

[コードの表示](#)

MSA options (custom MSA upload, single sequence, pairing mode)

msa_mode: MMseqs2 (UniRef+Environmental) MMseq2(デフォルト)

pair_mode: unpaired+paired unpaired+paired(デフォルト)

H1N1ノイラミニダーゼの構造予測(5)

AlphaFold2.ipynb のコピー ☆

コメント 共有 接続 編集

+ コード + テキスト

Advanced settings

model_type: auto

num_recycles: 3

- "auto" = protein structure prediction using "AlphaFold2-ptm" and complex prediction "AlphaFold-multimer-v2". For complexes "AlphaFold-multimer-v[1,2]" and "AlphaFold-ptm" can be used.

save_to_google_drive: **Googleドライブに結果を保存しておいた方がよい**

- if the save_to_google_drive option was selected, the result zip will be uploaded to your Google Drive

Don't forget to hit Runtime -> Run all after updating the form.

コードの表示

Install dependencies

コードの表示

Run Prediction

コードの表示

Display 3D structure

rank_num: 1

color: IDDT

show_sidechains:

show_mainchains:

H1N1ノイラミニダーゼの構造予測(6)

The screenshot shows a Google Colab notebook interface. At the top, the title is "AlphaFold2.ipynb". The "Runtime" menu is open, showing options like "Run all cells", "Run previous cells", "Run current cell", "Run selected cells", "Run all cells below", "Interrupt runtime", "Restart runtime", "Restart runtime and run all cells", and "Reset runtime to default timeout". A red circle highlights the option "Change runtime type".

Below the menu, the notebook content includes a code cell with the following parameters:

```
query_sequence: EFPIKSCSHLECRFTFFLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWSASACHDGI
```

- Use : to specify a jobname (supports homo- and hetero-oligomers)
- jobname: "p1"
- use_amber:
- template_mode: none

Below this, the "MSA options (custom MSA upload, single sequence, pairing mode)" section is visible:

```
msa_mode: MMseqs2 (UniRef+Environmental)
```

```
pair_mode: unpaired+paired
```

On the right side, the "Notebook settings" dialog is open. It shows the "Hardware accelerator" set to "GPU" and the "Runtime type" set to "Standard". There are two checkboxes: "Background execution" and "Exclude outputs from code cells when saving notebook", both of which are currently unchecked. A red circle highlights the "Save" button at the bottom right of the dialog.

Red text overlay: GPUの使用、メモリサイズ、バックグラウンド実行 (Colab Proのみ) を指定

H1N1ノイラミニダーゼの構造予測(7)

AlphaFold2.ipynb

コメント 共有

接続 編集

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

ランタイムを再起動 Ctrl+M

再起動してすべてのセルを実行

ランタイムを出荷時設定にリセット

ランタイムのタイプを変更

セッションの管理

ランタイムログの表示

すべてのセルを実行 → 実行が開始される

「警告: このノートブックは Google が作成したものではありません。」が表示されても無視 → 「このまま実行」をクリック

Input protein sequence

```
query_sequence:   
   
 • Use : to specify a specific protein sequence   
 jobname: "p1"   
 use_amber:    
 template_mode: none   
   
 • "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (mmCIF format)   
   
 コードの表示   
   
 MSA options (custom MSA upload, single sequence, pairing mode)   
 msa_mode: MMseqs2 (UniRef+Environmental)   
 pair_mode: unpaired+paired   
   
 - "unpaired+paired" - pair sequences from same species + unpaired MSA "unpaired" - generate MSA for each chain "paired" - only use paired sequences
```



H1N1ノイラミニダーゼの構造予測(8)

途中経過が表示される

セルのコードの表示(通常は不要)

```
[3] コードの表示
[4] Install dependencies
    コードの表示

Looking in links: https://storage.googleapis.com/jax-releases/jax\_releases.html
Requirement already satisfied: jax[cuda]<0.3.0 in /usr/local/lib/python3.7/dist-packages (0.2.28)
Requirement already satisfied: scipy>=1.2.1 in /usr/local/lib/python3.7/dist-packages (from jax[cuda]<0.3.0) (1.4.1)
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.7/dist-packages (from jax[cuda]<0.3.0) (1.21.5)
Requirement already satisfied: opt-einsum in /usr/local/lib/python3.7/dist-packages (from jax[cuda]<0.3.0) (3.3.0)
Requirement already satisfied: absl-py in /usr/local/lib/python3.7/dist-packages (from jax[cuda]<0.3.0) (0.13.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from jax[cuda]<0.3.0) (3.10.0.2)
Collecting jaxlib==0.1.76+cuda11.cudnn82
  Downloading https://storage.googleapis.com/jax-releases/cuda11/jaxlib-0.1.76%2Bcuda11.cudnn82-cp37-none-manylinux2010\_x86\_64.whl (153.1 MB)
Requirement already satisfied: flatbuffers<3.0,>=1.12 in /usr/local/lib/python3.7/dist-packages (from jaxlib==0.1.76+cuda11.cudnn82->jax[cuda]<0.3.0) (2.0)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from absl-py->jax[cuda]<0.3.0) (1.15.0)
Installing collected packages: jaxlib
  Attempting uninstall: jaxlib
    Found existing installation: jaxlib 0.3.2+cuda11.cudnn805
    Uninstalling jaxlib-0.3.2+cuda11.cudnn805:
      Successfully uninstalled jaxlib-0.3.2+cuda11.cudnn805
  Successfully installed jaxlib-0.1.76+cuda11.cudnn82
```

Run Prediction 実行中のステップ
コードの表示

Display 3D structure このステップは実行予定であることを示す
rank_num: 1
color: IDDT

H1N1ノイラミニダーゼの構造予測(9)

AlphaFold2.ipynb のコピー ☆

コメント 共有 編集

RAM ディスク

+ コード + テキスト

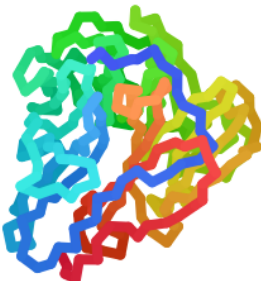
```
[4] Uninstalling jaxlib-0.3.2+cuda11.cudnn805:
      Successfully uninstalled jaxlib-0.3.2+cuda11.cudnn805
      Successfully installed jaxlib-0.1.76+cuda11.cudnn82
```

Run Prediction

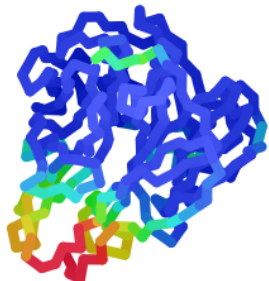
コードの表示

```
Downloading alphafold2 weights to .: 100% |██████████| 3.47G/3.47G [00:38<00:00, 95.6MB/s]
2022-04-08 10:34:01,642 Running colabfold 1.2.0 (0a7b11137dda693263f2eebd8c83f7b91af901f0)
2022-04-08 10:34:01,647 Found 5 citations for tools or databases
2022-04-08 10:34:07,881 Query 1/1: p1_906ba (length 385)
COMPLETE: 100% |██████████| 150/150 [elapsed: 00:02 remaining: 00:00]
2022-04-08 10:34:10,236 Running model_1
2022-04-08 10:38:47,455 model_1 took 272.1s (3 recycles) with pLDDT 88.1
```

colored by N→C



colored by pLDDT



5つのモデルが表示される

```
2022-04-08 10:39:21,547 Running model_2
```

Display 3D structure

rank_num: 1

color: IDDT

show_sidechains:

show_mainchains:

H1N1ノイラミニダーゼの構造予測(10)

AlphaFold2.ipynb ☆

コメント 共有 編集

RAM ディスク

+ コード + テキスト

[0] Display 3D structure

rank_num: 1

color: IDDT

show_sidechains:

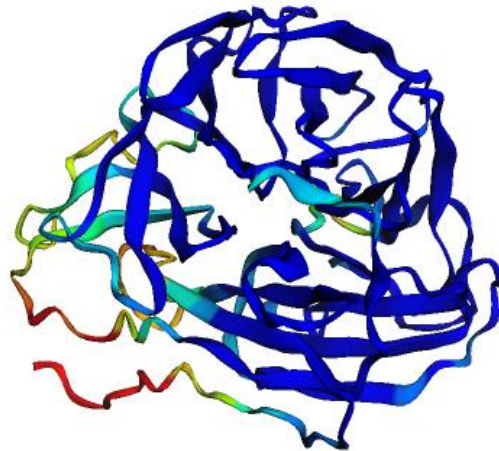
show_mainchains:

コードの表示

デフォルトで5つのモデル構造を生成
ランクが選べるが、ランク1の構造がデフォルトで表示される

IDDTによる色づけ(デフォルト)

上で選択したランクの構造を表示



■ Very high (pLDDT > 90)
■ Confident (90 > pLDDT > 70)
■ Low (70 > pLDDT > 50)
■ Very low (pLDDT < 50)

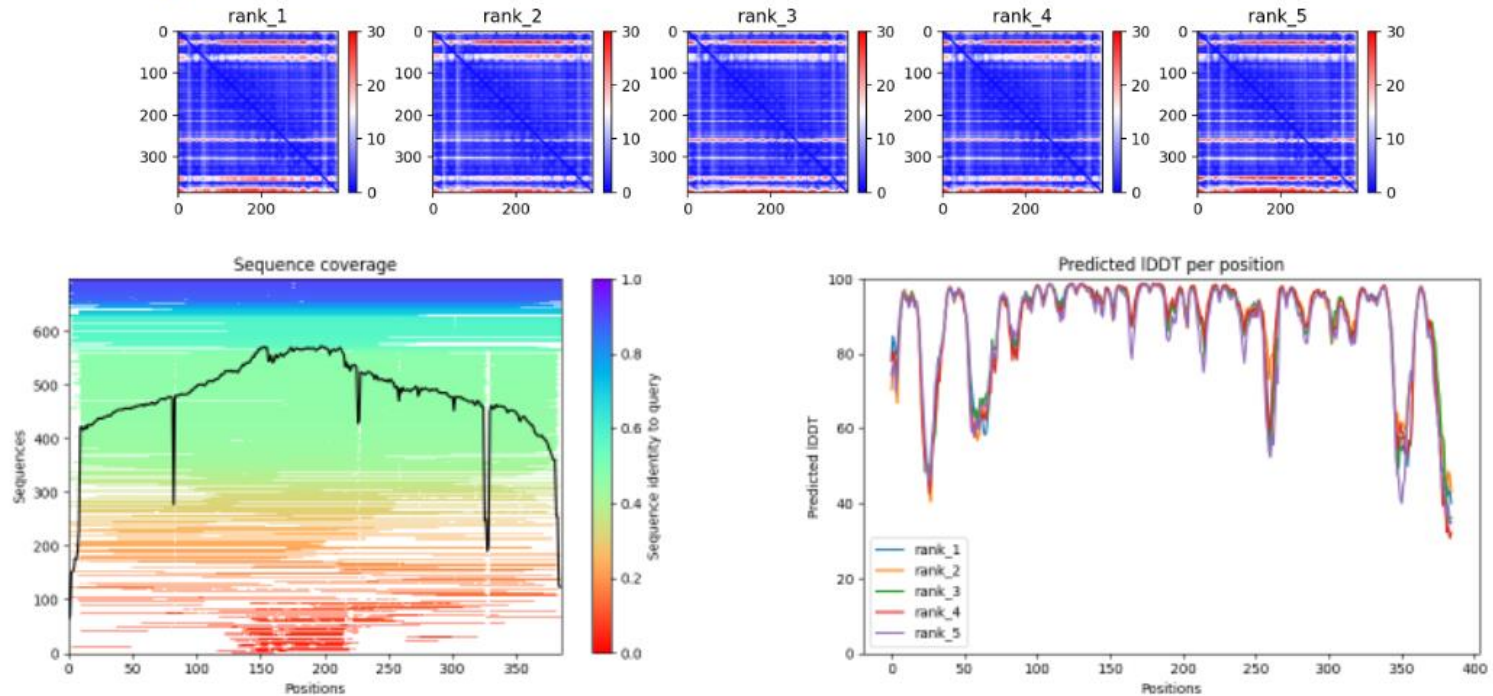
H1N1ノイラミニダーゼの構造予測(11)

+ コード + テキスト

[7]

Plots for p1_906ba

PAEの分布



配列のカバー率と一致度

各残基の予測されたIDDT (pLDDT) 値

[8] Package and download results

If you are having issues downloading the result archive, try disabling your adblocker and run this cell again. If that fails click on the little folder icon to the left, navigate to file: jobname.result.zip, right-click and select "Download" (see [screenshot](#)).

コードの表示

```
adding: config.json (deflated 47%)  
adding: p1_906ba_predicted_model.zip (deflated 91%)
```

H1N1ノイラミニダーゼの構造予測(12)

+ コード + テキスト

RAM
ディスク

編集

Instructions

Quick start

1. Paste your protein sequence(s) in the input field.
2. Press "Runtime" -> "Run all".
3. The pipeline consists of 5 steps. The currently running steps is indicated by a circle with a stop sign next to it.

Result zip file contents

zipファイルの内容

1. PDB formatted structures sorted by avg. pLDDT and complexes are sorted by pTMscore. (unrelaxed and relaxed if `use_amber` is enabled).
2. Plots of the model quality.
3. Plots of the MSA coverage.
4. Parameter log file.
5. A3M formatted input MSA.
6. A `predicted_aligned_error_v1.json` using [AlphaFold-DB's format](#) and a `scores.json` for each model which contains an array (list of lists) for PAE, a list with the average pLDDT and the pTMscore.
7. BibTeX file with citations for all used tools and databases.

At the end of the job a download modal box will pop up with a `jobname.result.zip` file. Additionally, if the `save_to_google_drive` option was selected, the `jobname.result.zip` will be uploaded to your Google Drive.

MSA generation for complexes

For the complex prediction we use unpaired and paired MSAs. Unpaired MSA is generated the same way as for the protein structures prediction by searching the UniRef100 and environmental sequences three iterations each.

The paired MSA is generated by searching the UniRef100 database and pairing the best hits sharing the same NCBI taxonomical identifier (=species or sub-species). We only pair sequences if all of the query sequences are present for the respective taxonomical identifier.

Using a custom MSA as input

To predict the structure with a custom MSA (A3M formatted): (1) Change the `msa_mode` to "custom". (2) Wait for an upload box to appear at the

0秒 完了時間: 11:05

p1_906ba.result.zip

結果のzipファイル → 解凍する

Googleドライブに結果を保存する指定しておけば、そこからzipファイルをダウンロードできる

H1N1ノイラミニダーゼの構造予測(13)

zipファイルの内容

cite.bibtex	→	参考文献	
config.json	→	指定内容のjsonファイル	
p1_906ba.a3m	→	MSA	
p1_906ba_coverage.png		}	MSAのカバー状況、PAE、pLDDTの結果の画像
p1_906ba_PAE.png			
p1_906ba_plddt.png			
p1_906ba_predicted_aligned_error_v1.json		}	モデル構造のPDBファイル pLDDT値によってランクづけされている PAE、pLDDT、pTMの結果を格納したjsonファイル
p1_906ba_unrelaxed_rank_1_model_3.pdb			
p1_906ba_unrelaxed_rank_1_model_3_scores.json			
p1_906ba_unrelaxed_rank_2_model_1.pdb			
p1_906ba_unrelaxed_rank_2_model_1_scores.json			
p1_906ba_unrelaxed_rank_3_model_4.pdb			
p1_906ba_unrelaxed_rank_3_model_4_scores.json			
p1_906ba_unrelaxed_rank_4_model_5.pdb			
p1_906ba_unrelaxed_rank_4_model_5_scores.json			
p1_906ba_unrelaxed_rank_5_model_2.pdb			
p1_906ba_unrelaxed_rank_5_model_2_scores.json			

結果の見方

TM-score

- TM-score (Template modeling score): タンパク質構造のトポロジー的な類似度を評価するための指標
- タンパク質全長にわたる比較に使用
 - 例えば、天然構造とモデル構造との構造類似度の評価
- アミノ酸残基レベルの距離に基づく指標
- RMSDの問題に対する対処
 - 残基ペア間の距離が大きいものの影響を減じる
 - グローバルなフォールド類似性に対する感度を上げる
 - 長さに依存しないよう、誤差を正規化する
- 値は、(0, 1]で、値が大きいほど類似していることを示し、値1は完全一致を示す
- PDB構造の分布により、0.17以下のスコアはランダムに選ばれた無関係なタンパク質に対応し、0.5以上のスコアはSCOP/CATHで概ね同じフォールドと考えられる
- 計算式

$$\text{TM-score} = \max \left(\frac{1}{L_T} \sum_{i=1}^{L_C} \frac{1}{1 + \left(\frac{d_i}{d_0(L_T)}\right)^2} \right)$$

- L_T : ターゲットタンパク質の構造のアミノ酸の配列長
- L_C : ターゲットタンパク質とアラインされたアミノ酸残基の数
- d_i : i 番目ペアのアミノ酸残基間距離、
- $d_0(L_T) = 1.24 \sqrt[3]{L_T - 15} - 18$
- maxは、最適に重ね合わせられた構造を示す

pTM

- TM-scoreの計算式は、グローバルアライメントで最適解を求めるのは困難 → 残基ペアの距離をもとに計算
- モデリングにおけるTM-scoreの予測では、その距離の分布を予測して計算

predicted TM-score (pTM) の計算

e_{ij} : 正解構造と予測構造が残基*i*でアラインされているとき、
予測構造の残基*j*のC α 原子の誤差

→ この確率分布をニューラルネットワークで予測
0~31.5 Å を 0.5 Å のbinで64段階で予測

$$\text{pTM} = \max_i \frac{1}{N_{res}} \sum_j \mathbb{E}[f(e_{ij})]$$

構造の一部 \mathcal{D} の残基に対して計算

$$\text{pTM}(\mathcal{D}) = \max_{i \in \mathcal{D}} \frac{1}{|\mathcal{D}|} \sum_{j \in \mathcal{D}} \mathbb{E} \left[\frac{1}{1 + \left(\frac{e_{ij}}{d_0(|\mathcal{D}|)} \right)^2} \right]$$

$f(e_{ij})$: ドメインのパッキングを示す指標になる

GDT-TS

- GDT (Global Distance Test): CASPで用いられている構造類似度スコア
- タンパク質全長にわたる比較に使用
 - 例えば、天然構造とモデル構造との構造類似度の評価
- GDT_P n : C $_{\alpha}$ 原子が誤差 $n\text{\AA}$ 以内にある割合
- GDT_TS (GDT Total Score):
$$\text{GDT_TS} = \frac{1}{4} (\text{GDT_P}_1 + \text{GDT_P}_2 + \text{GDT_P}_4 + \text{GDT_P}_8)$$
- GDT_HA (High Accurate GDT):
$$\text{GDT_HA} = \frac{1}{4} (\text{GDT_P}_{0.5} + \text{GDT_P}_1 + \text{GDT_P}_2 + \text{GDT_P}_4)$$
- 値は、(0, 1]で、値が大きいほど類似していることを示し、値1は全残基が最も厳しい誤差範囲 (GDT_TSは 1\AA 、GDT_HAは 0.5\AA 以内) に入っているとき

IDDTとpLDDT

- IDDT (The Local Distance Difference Test)
- 参照構造をモデル構造がいかに再現しているかを示す
- 参照構造の距離 R_0 以内にある原子ペアを計算対象とする
 - デフォルトは $R_0 = 15 \text{ \AA}$
- これらの原子ペアの距離がモデル構造において、許容されるしきい値 L 以内におさまれば、「保存されている」とみなし、その「保存されている」ペアの割合をスコアとする
- IDDTスコアは、 $L = 0.5 \text{ \AA}$, 1 \AA , 2 \AA , 4 \AA のスコアの平均
- 計算は全原子、 C_α 原子、主鎖原子など
- 重ね合わせによらないスコアが計算される

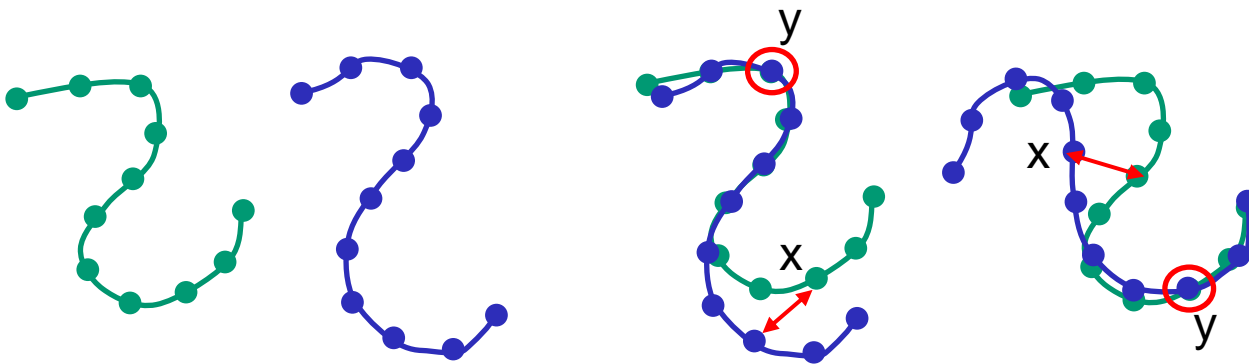


AlphaFold 2では、 C_α -IDDT値の予測値 (pLDDT) を評価に使用

- pLDDTは0~100の値をとる
- pLDDT < 50の領域は天然変性領域、複合体形成時に構造をもつようになる領域の可能性

PAE

- PAE (Predicted Aligned Error): 参照構造とモデル構造が残基 y でアラインされたときの残基 x の予測位置誤差 (Å)
 - x : scored residue, y : aligned residue
 - 2つの異なるドメインの残基 x と残基 y に対して、 (x, y) のPAEが一貫して低い場合 (5Å以下)、相対的なドメイン位置について確らしいと判断できる
 - 誤差の計算は C_α 、N、C原子を使用、0~31.75Å

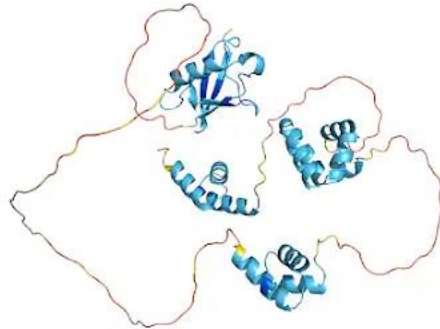


残基間距離ではなく、期待される距離の誤差
(expected distance error)

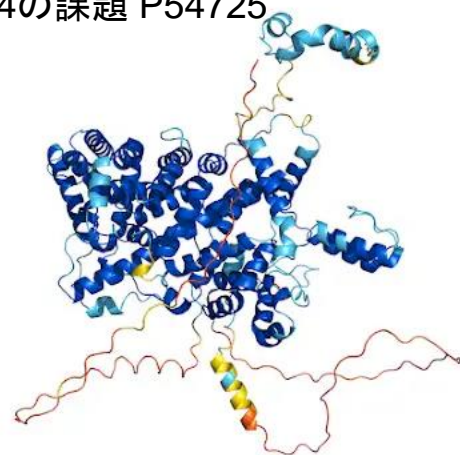
マルチドメインタンパク質の予測結果

ドメインリンカを含めた予測の例

Per residue confidence



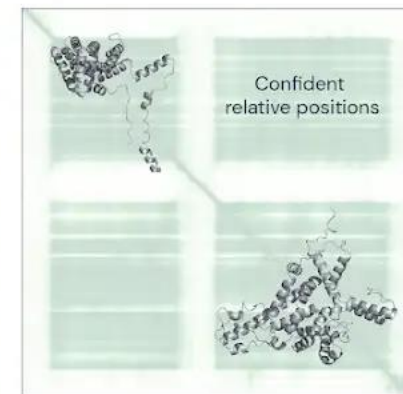
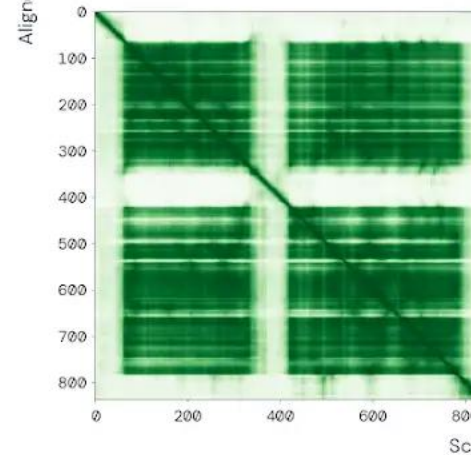
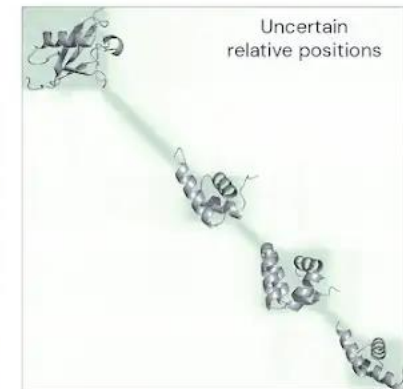
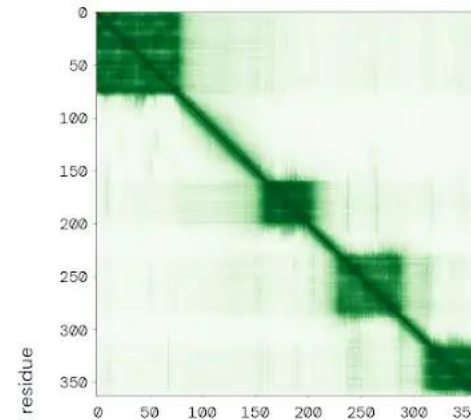
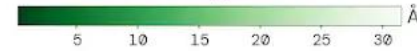
CASP14の課題 P54725



CASP14の課題 Q5VSL9

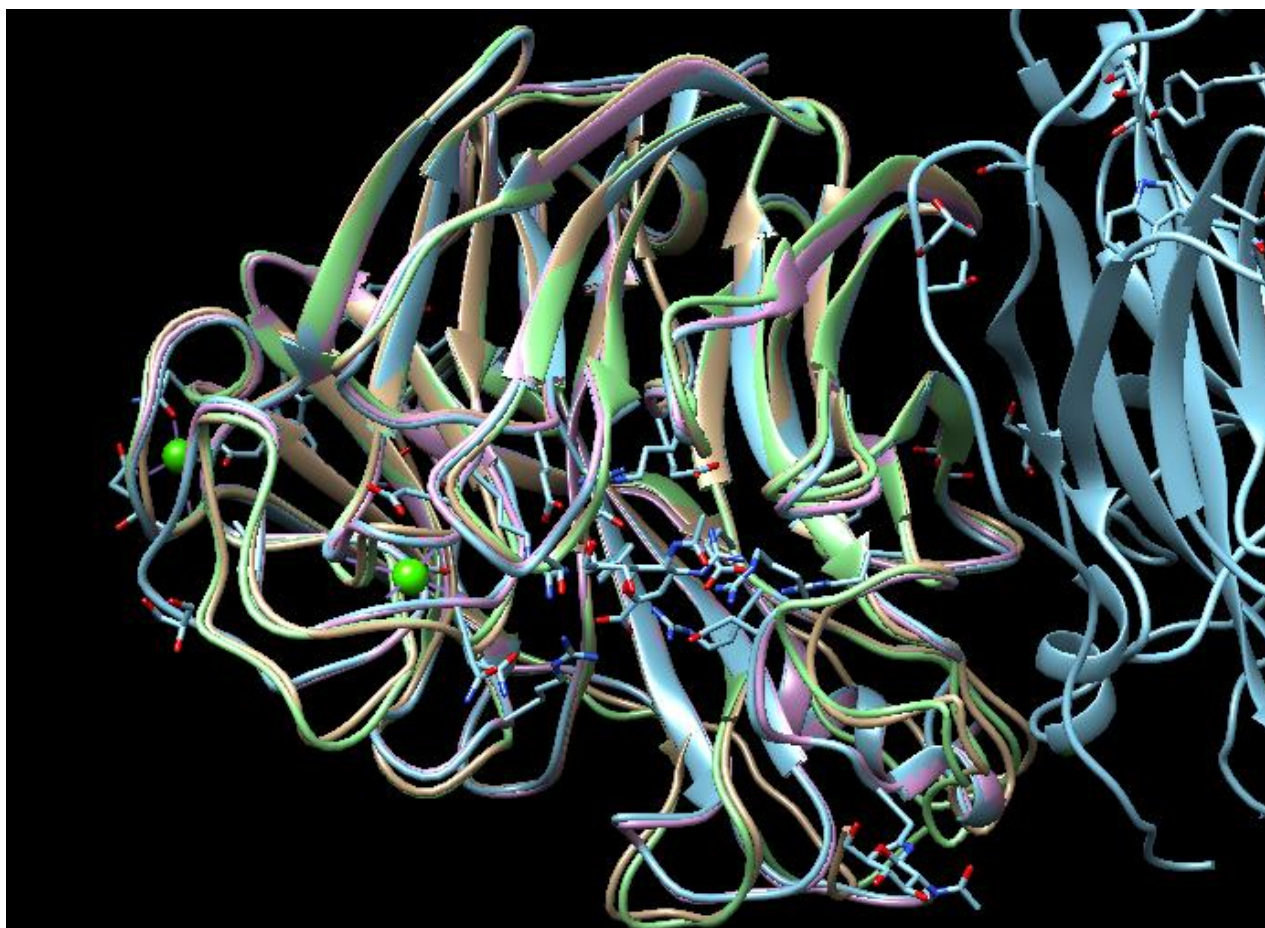
ドメインの位置関係もある程度信頼性がある(現在では構造決定されている)

Predicted Aligned Error



構造予測の結果

- モデル構造と正解構造(結晶構造、PDB ID: 3b7e)の構造重ね合わせ



AlphaFold2 (relaxed, template) → 0.252Å
(全体)

AlphaFold2 (unrelaxed, template) → 0.699Å
(336残基)、1.579Å
(全体)

水色: 結晶構造 (PDBID: 3b7e)

茶色: オプションなし → 0.734Å (329残基)、1.946 (全体)

ピンク: テンプレートを使用 → 0.258Å (385残基 = 全体)

緑色: Amber力場で構造を緩和 → 0.649Å (323残基)、2.282 (全体)

プリオンタンパク質の構造予測(1)

- ターゲット: ヒトのプリオンタンパク質 (タンパク質 p2) → [講義のページ](#)
 - アミノ酸配列は「[p2.fasta](#)」
 - 正解構造: [PDB ID: 1qix](#) → 構造がわからないと仮定
- ColabFoldで実行

プリオンタンパク質の構造予測(2)

+ コード + テキスト

「ランタイム」→「すべてのセルを実行」

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv. 2021](#)



ターゲットの配列を入力(配列の部分だけ入力する)

Input protein sequence(s), then hit Runtime -> Run all [p2.fasta](#)

query_sequence: `GSKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGGWGQPHGGGGWGQPHGGGGWGQPHGGGGWGQGGGTHSQWNKPSKPKTNMKHMAC`

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

jobname: `p2` **ジョブ名を入力**

use_amber:

template_mode: `none`

- `"none"` = no template information is used, `"pdb70"` = detect templates in pdb70, `"custom"` - upload and search own templates (mmCIF format)

[コードの表示](#)

MSA options (custom MSA upload, single sequence, pairing mode)

msa_mode: `MMseqs2 (UniRef+Environmental)`

pair_mode: `unpaired+paired`

- `"unpaired+paired"` = pair sequences from same species + unpaired MSA, `"unpaired"` = separate MSA for each chain, `"paired"` - only use paired sequences.

[コードの表示](#)

プリオンタンパク質の構造予測(3)

+ コード + テキスト

RAM ディスク 編集

[5] Run Prediction

コードの表示

```
Downloading alphafold2 weights to .: 100%|██████████| 3.47G/3.47G [00:32<00:00, 116MB/s]
2022-04-14 14:13:37,222 Running colabfold 1.2.0 (f3f924e4d0acc69ebab7083ca895339976e57f12)
2022-04-14 14:13:37,224 Found 5 citations for tools or databases
2022-04-14 14:13:42,821 Query 1/1: p2_5224d (length 210)
COMPLETE: 100%|██████████| 150/150 [elapsed: 00:01 remaining: 00:00]
2022-04-14 14:13:44,566 Running model_1
2022-04-14 14:17:53,720 model_1 took 245.3s (3 recycles) with pLDDT 62.3
```

colored by N→C

colored by pLDDT



```
2022-04-14 14:18:11,183 Running model_2
2022-04-14 14:18:38,106 model_2 took 24.9s (3 recycles) with pLDDT 63.3
```

colored by N→C

colored by pLDDT



プリオンタンパク質の構造予測(4)

+ コード + テキスト

[8] Display 3D structure

rank_num: 1

color: IDDT

show_sidechains:

show_mainchains:

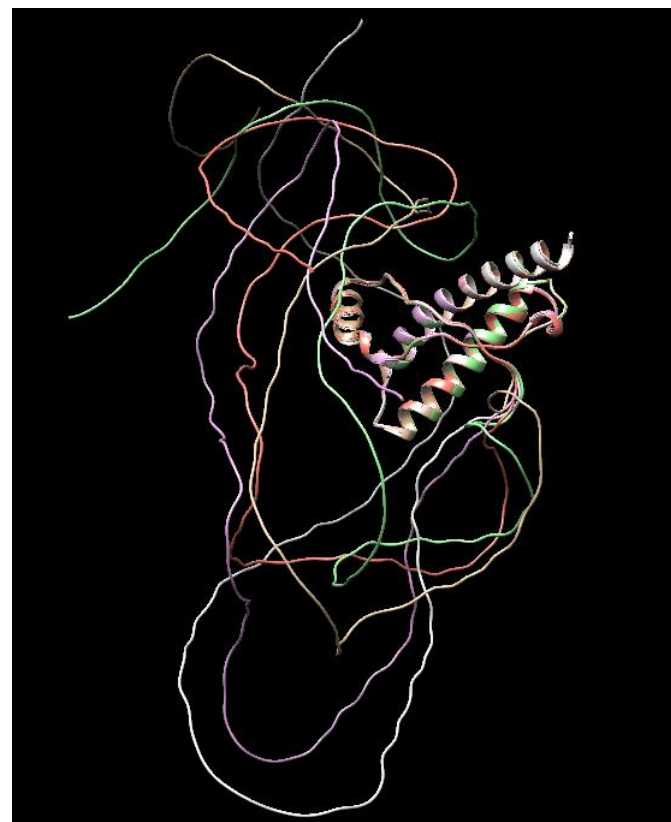
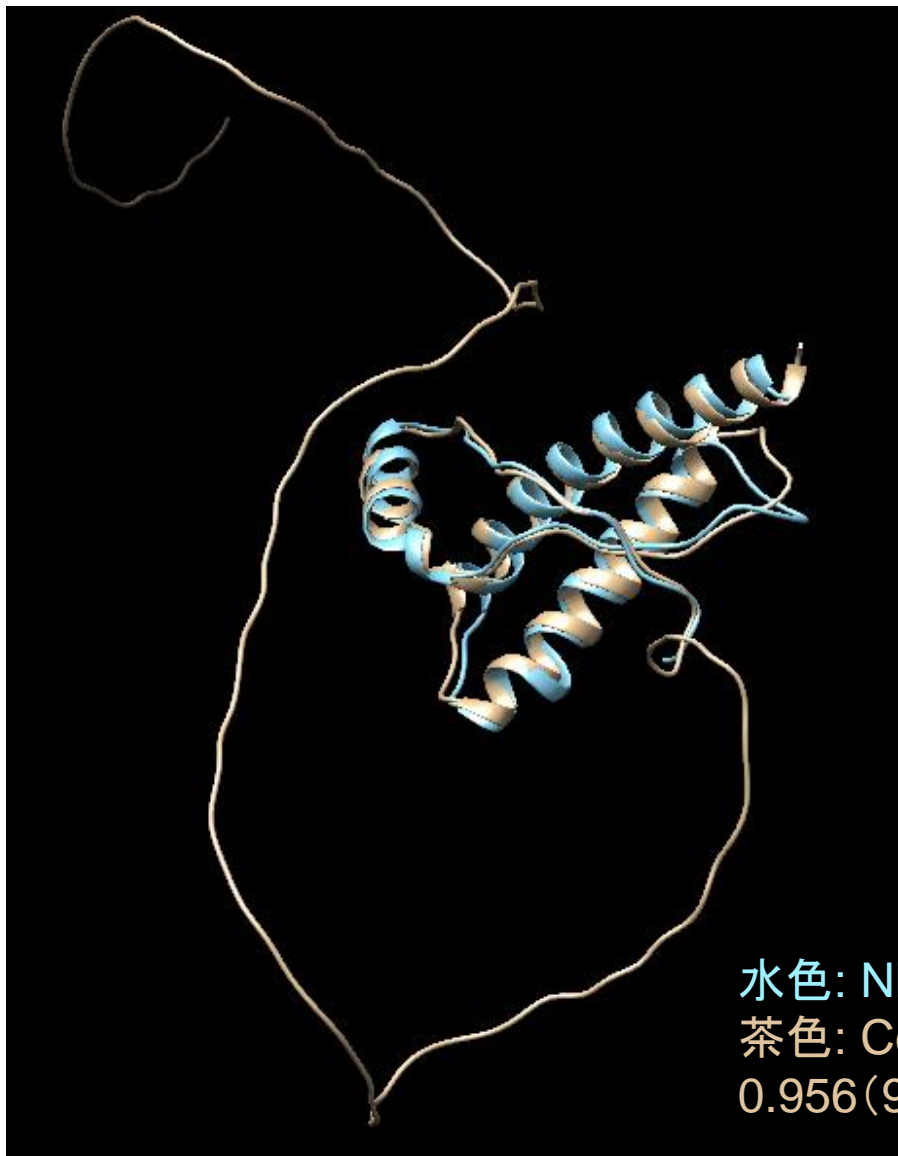
[コードの表示](#)



構造予測の結果

- モデル構造と正解構造(NMR構造、PDB ID: 1qlx)の構造重ね合わせ

予測した5の構造の重ね合わせ



水色: NMR構造 (PDBID: 1qlx)


茶色: ColabFold → 1.533Å (104残基)、
0.956 (91残基)

Ras-Raf複合体の構造予測(1)

- ターゲット: Ras-Raf複合体 → 講義のページ
 - アミノ酸配列は「[pp1.fasta](#)」
 - 正解構造: [PDB ID: 3kud](#) → 構造がわからないと仮定
- ColabFoldで実行

Structure Summary | 3D View | Annotations | Experiment | Sequence | Genome | Ligands | Versions

Biological Assembly 1 ?



3KUD
Complex of Ras-GDP with RafRBD(A85K)
PDB DOI: [10.2210/pdb3KUD/pdb](#)
Classification: **GTP BINDING PROTEIN/TRANSFERASE**
Organism(s): *Homo sapiens*
Expression System: *Escherichia coli*
Mutation(s): Yes ⓘ

Deposited: 2009-11-27 Released: 2010-03-23
Deposition Author(s): Filchtinski, D., Sharabi, O., Rueppel, A., Vetter, I.R., Herrmann, C., Shifman, J.M.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.15 Å
R-Value Free: 0.264
R-Value Work: 0.237
R-Value Observed: 0.238

wwPDB Validation ⓘ

Metric	Percentile Ranks	Value
Rfree		0.272
Clashscore		38
Ramachandran outliers		4.6%
Sidechain outliers		20.4%
RSRZ outliers		37.3%

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Display Files | Download Files

Contact Us

Ras-Raf複合体の構造予測(2)

AlphaFold2.ipynb

ランタイム

「ランタイム」→「すべてのセルを実行」

ColabFold: 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: [v1.0](#), [v1.1](#), [v1.2](#), [v1.3](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv. 2021](#)



ターゲットの配列を入力

pp1.fasta

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

query_sequence: MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDITAGQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSI

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

jobname: Ras-Raf ジョブ名を入力

use_amber:

template_mode: none

• "none" = no template information is used, "pdb70" = detect templates in pdb70, "custom" - upload and search own templates (mmCIF format)

コードの表示

[2] MSA options (custom MSA upload, single sequence, pairing mode)

msa_mode: MMseqs2 (UniRef+Environmental)

pair_mode: unpaired+paired

• "unpaired+paired" = pair sequences from same species + unpaired MSA, "unpaired" = separate MSA for each chain, "paired" - only use paired sequences.

Ras-Raf複合体の構造予測(3)

+ コード + テキスト

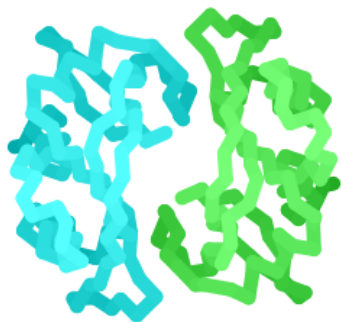
接続 編集

Run Prediction

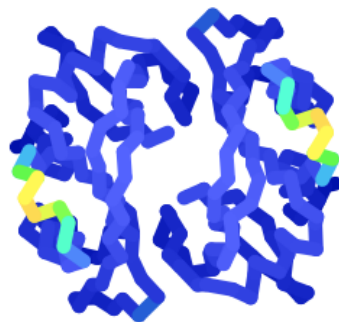
[コードの表示](#)

2022-04-11 18:27:43,392 Running model_3
2022-04-11 18:28:02,585 model_3 took 19.2s (3 recycles) with pLDDT 94.6 and ptmscore 0.881

colored by chain

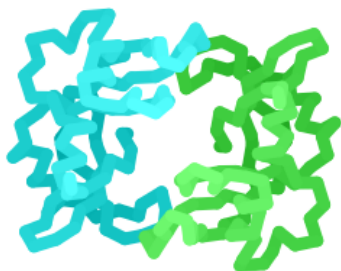


colored by pLDDT

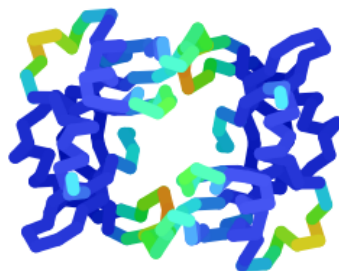


2022-04-11 18:28:22,459 Running model_4
2022-04-11 18:28:41,668 model_4 took 19.2s (3 recycles) with pLDDT 87.7 and ptmscore 0.551

colored by chain



colored by pLDDT



Ras-Raf複合体の構造予測(4)

+ コード + テキスト

Display 3D structure

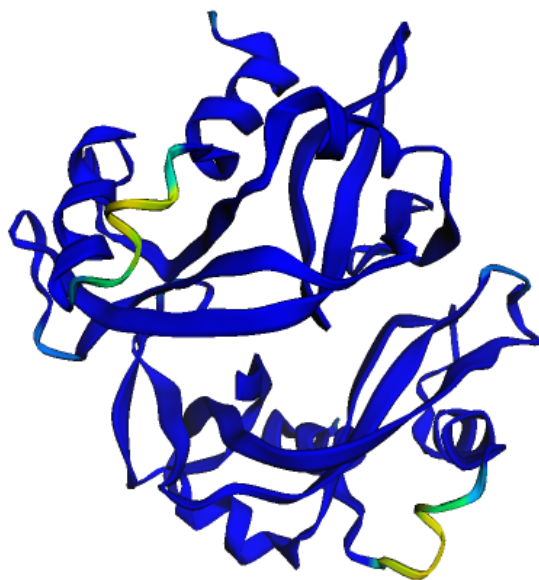
rank_num: 1

color: IDDT

show_sidechains:

show_mainchains:

[コードの表示](#)



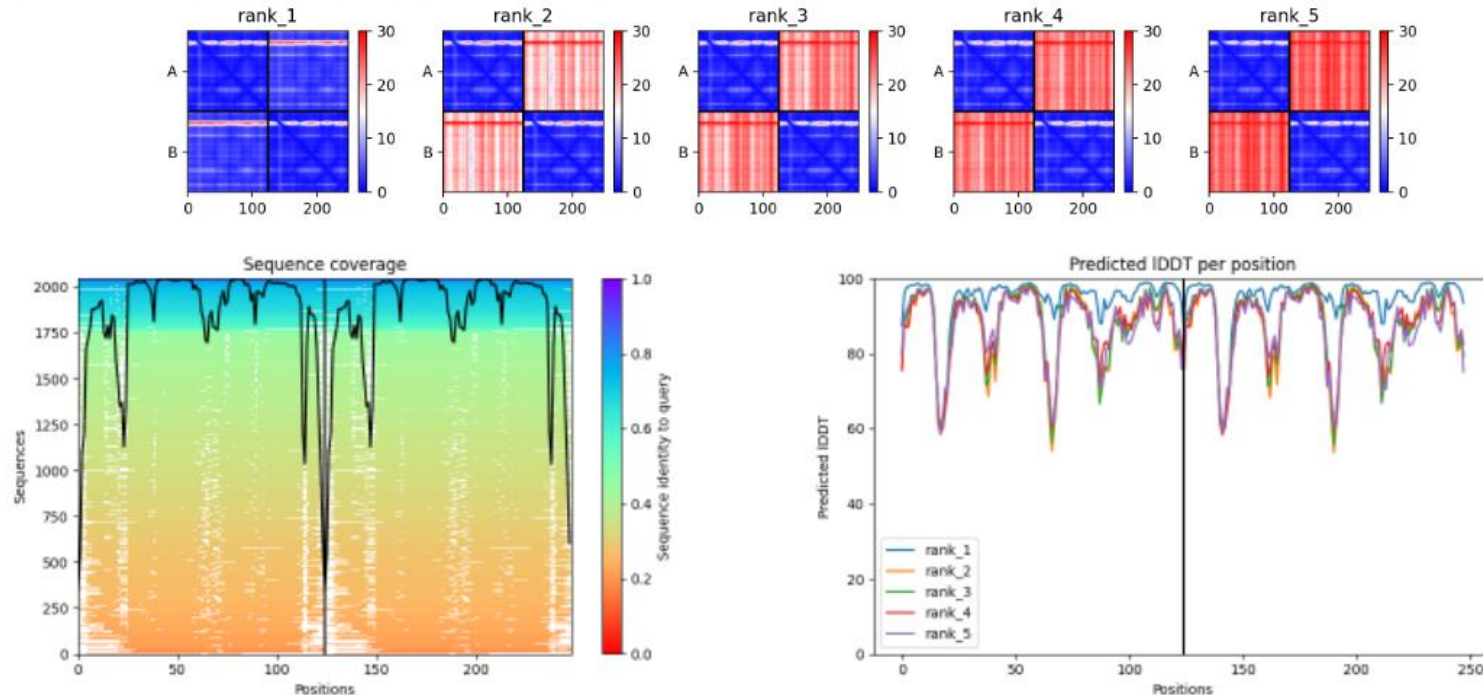
Ras-Raf複合体の構造予測(5)

+ コード + テキスト

Plots

コードの表示

Plots for ribonuclease_d730b



Package and download results

Ras-Raf複合体の構造予測(6)

+ コード + テキスト

RAM ディスク 編集

Result zip file contents

1. PDB formatted structures sorted by avg. pLDDT and complexes are sorted by pTMScore. (unrelaxed and relaxed if `use_amber` is enabled).
2. Plots of the model quality.
3. Plots of the MSA coverage.
4. Parameter log file.
5. A3M formatted input MSA.
6. A `predicted_aligned_error_v1.json` using [AlphaFold-DB's format](#) and a `scores.json` for each model which contains an array (list of lists) for PAE, a list with the average pLDDT and the pTMScore.
7. BibTeX file with citations for all used tools and databases.

At the end of the job a download modal box will pop up with a `jobname.result.zip` file. Additionally, if the `save_to_google_drive` option was selected, the `jobname.result.zip` will be uploaded to your Google Drive.

MSA generation for complexes

For the complex prediction we use unpaired and paired MSAs. Unpaired MSA is generated the same way as for the protein structures prediction by searching the UniRef100 and environmental sequences three iterations each.

The paired MSA is generated by searching the UniRef100 database and pairing the best hits sharing the same NCBI taxonomical identifier (=species or sub-species). We only pair sequences if all of the query sequences are present for the respective taxonomical identifier.

Using a custom MSA as input

To predict the structure with a custom MSA (A3M formatted): (1) Change the `msa_mode` to "custom", (2) Wait for an upload box to appear at the end of the "MSA options ..." box. Upload your A3M. The first fasta entry of the A3M must be the query sequence without gaps.

As an alternative for MSA generation the [HHblits Toolkit server](#) can be used. After submitting your query, click "Query Template MSA" -> "Download Full A3M". Download the A3M file and upload it in this notebook.

Using custom templates

To predict the structure with a custom templates (mmCIF formatted): (1) change the `template_mode` to "custom" execture cell and (2) wait for an upload box to appear at the end of the "Input Protein" box. Select and upload your templates (multiple choices are possible). Templates need to be in mmCIF format, its names must follow the PDB naming (four letters) and must contain `_entity_poly_seq` and

✓ 0秒 完了時間: 1:09

✕

RasRaf_3f734.result.zip

結果のzipファイル → 解凍する

すべて表示

✕

構造予測の結果

- 複数チェーンの重ね合わせ

Tools → Structure

Comparison → MatchMaker

RasRafモデル構造のBチェーンと
結晶構造3kudのAチェーン

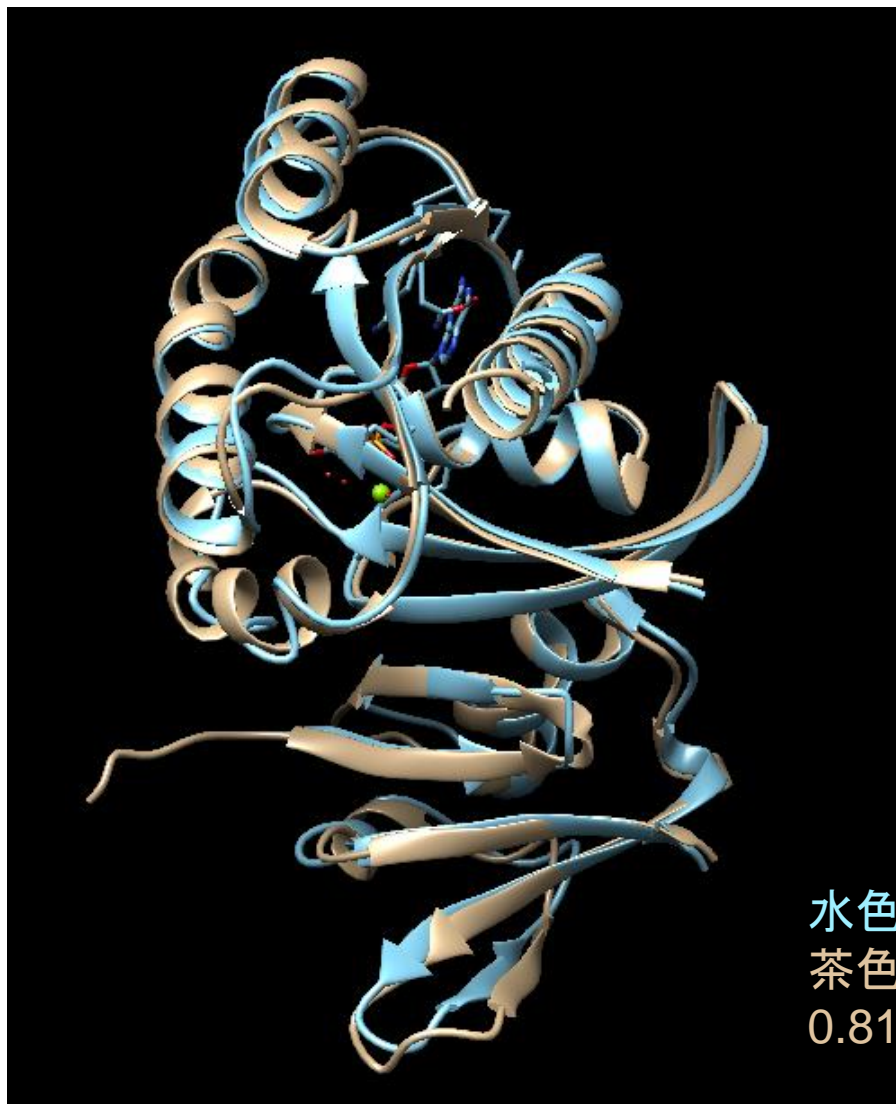
RasRafモデル構造のCチェーンと
結晶構造3kudのBチェーン
の両方を重ね合わせる

特定の(複数の)チェーン

The screenshot shows the MatchMaker software interface. The 'Reference chain:' list on the left contains 'RasRaf_3f734_unrelaxed_rank_1_model_1.pdb (#0) chain B', 'RasRaf_3f734_unrelaxed_rank_1_model_1.pdb (#0) chain C', '3KUD (#1) chain A', and '3KUD (#1) chain B'. The 'Chain(s) to match:' list on the right contains 'ref: RasRaf_3f734_unrelaxed_rank_1_model_1.pdb (#0) chain B', '3KUD (#1) chain A', 'ref: RasRaf_3f734_unrelaxed_rank_1_model_1.pdb (#0) chain C', and '3KUD (#1) chain B'. The 'Chain pairing' section has three radio button options: 'Best-aligning pair of chains between reference and match structure', 'Specific chain in reference structure with best-aligning chain in match structure', and 'Specific chain(s) in reference structure with specific chain(s) in match structure', which is selected and circled in red. The 'Alignment algorithm:' is set to 'Needleman-Wunsch', 'Matrix:' is 'BLOSUM-62', 'Gap opening penalty' is 12, and 'Gap extension penalty' is 1. Other options include 'Include secondary structure score (30%)', 'Compute secondary structure assignments', 'Show pairwise alignment(s)', and 'Iterate by pruning long atom pairs until no pair exceeds: 2 angstroms'. Buttons for 'Save settings', 'Reset to defaults', 'OK', 'Apply', 'Cancel', and 'Help' are visible at the bottom.

構造予測の結果

- モデル構造と正解構造(結晶構造、PDB ID: 3kud)の構造重ね合わせ



DockQ (Quality Measure for Protein-Protein Docking Models)、Fnat、LRMS、iRMSなどの指標についても調べてみよう

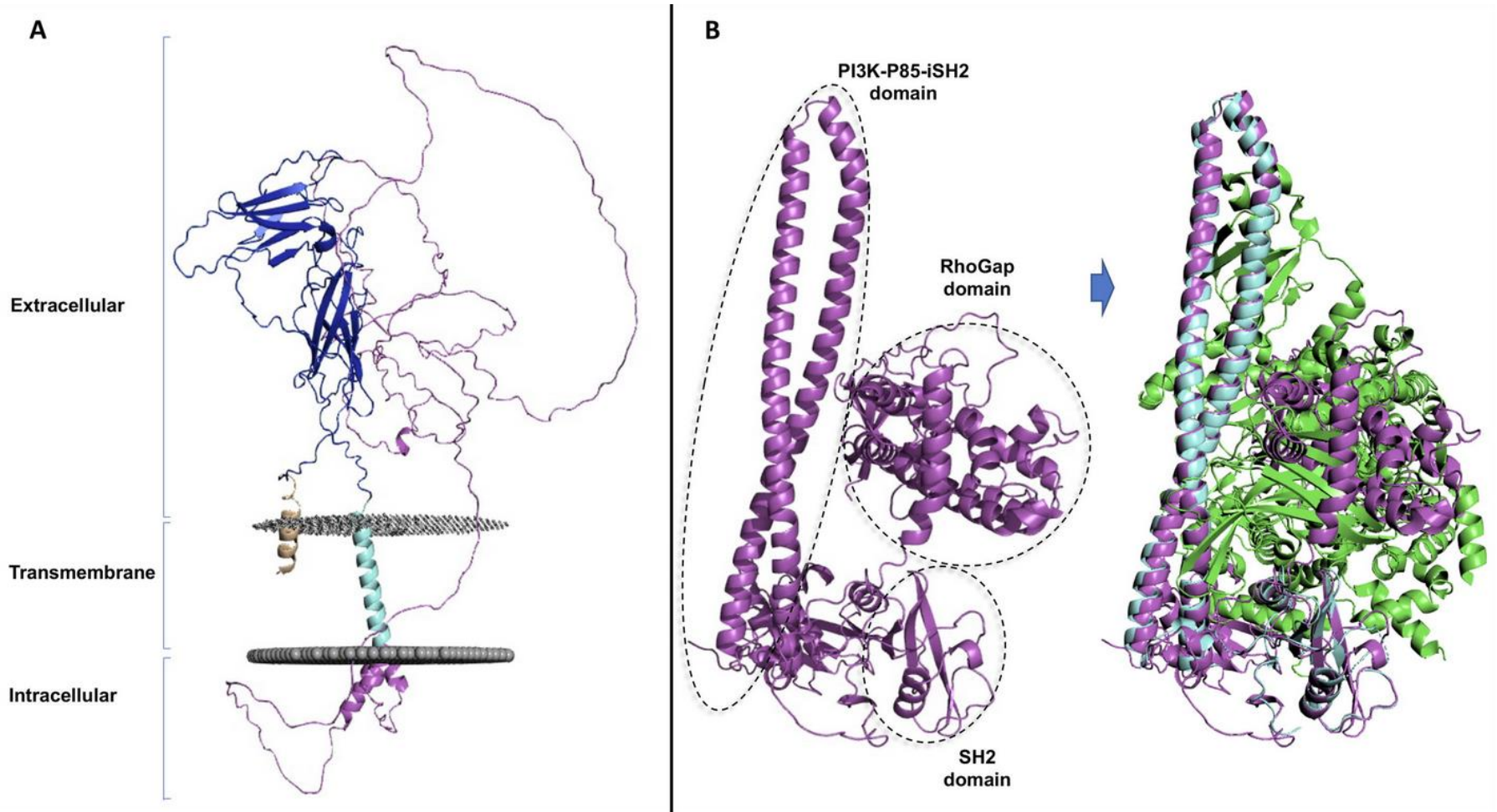
水色: 結晶構造 (PDBID: 3kud)

茶色: ColabFold → 1.007Å (241残基)、
0.810 (231残基)

AlphaFoldの限界

- フォールディング過程に関する情報は用いていない
- 複数のコンフォメーションをとるタンパク質の構造を予測できていない
- 膜貫通領域を意識したモデリングは行っていない
- 他の分子(核酸、低分子化合物、金属など)との結合構造を予測することができない
- タンパク質複合体構造、ドメインの相対的な位置の予測、の予測の精度はまだ高くない
- 生体内で、変異によってフォールドがしにくくなっているタンパク質も、構造をモデリングしてしまう

AlphaFoldの限界



成長ホルモン受容体 (GHR、UniProt P10912)
構造化されていない長い細胞内テール (残基
289-638) の配置が誤っている

PIK3R1 (UniProt P27986) のモデル構造 (マゼンタ)
、右側は実験によって得られた PIK3R1 (水色) と
PIK3CD (緑) の複合体構造

[A David, et al. The AlphaFold Database of Protein Structures: A Biologist's Guide. J. Mol. Biol. 2022 434:167336.](#)

受容体型チロシンキナーゼの予測結果

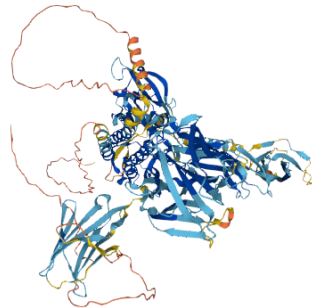
- UniProtKB P36888 (FLT3_HUMAN)

Structure

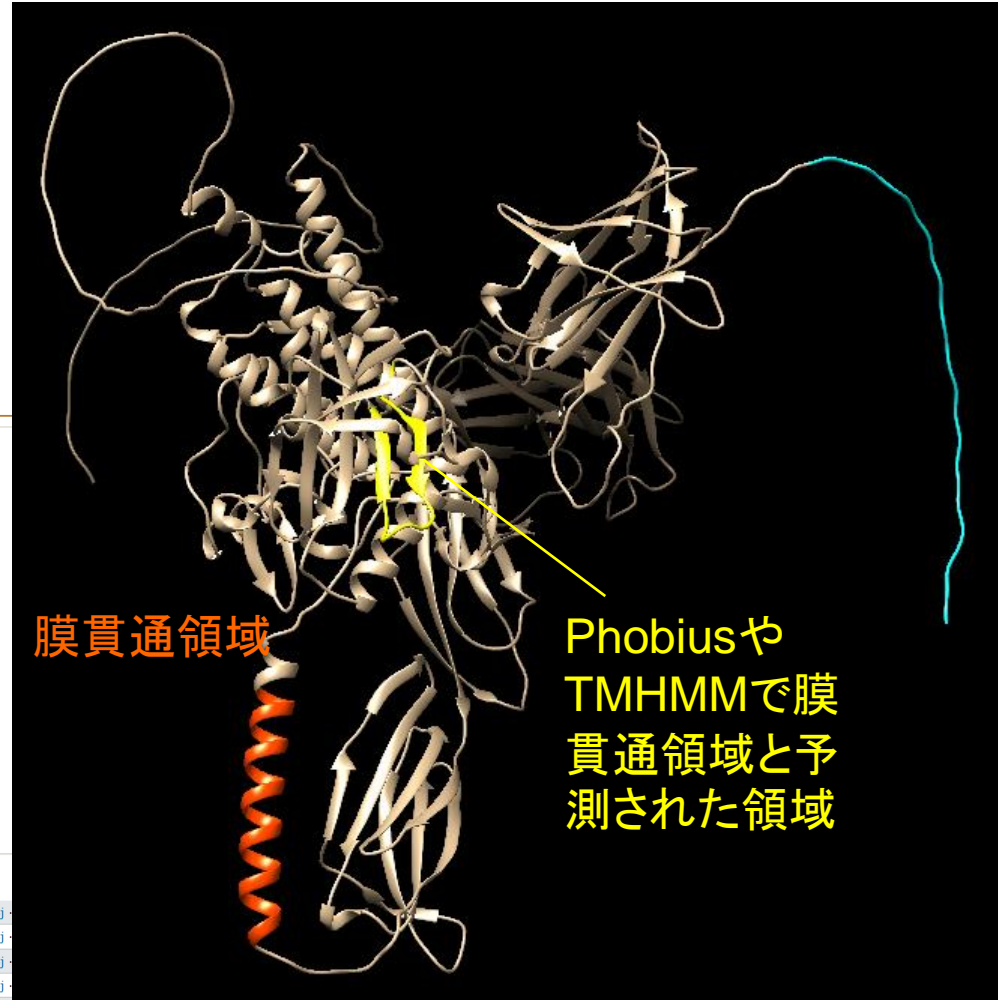
Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
PDB	1RJB	X-ray	2.10 Å	A	564-958	PDBe · RCSB-PDB · PDBJ
PDB	3QS7	X-ray	4.30 Å	E/F/G/H	27-436	PDBe · RCSB-PDB · PDBJ
PDB	3QS9	X-ray	7.80 Å	E/F/G/H	27-540	PDBe · RCSB-PDB · PDBJ
PDB	4RT7	X-ray	3.10 Å	A	564-958	PDBe · RCSB-PDB · PDBJ
PDB	4XUF	X-ray	3.20 Å	A/B	600-947	PDBe · RCSB-PDB · PDBJ · PDBsum
PDB	5X02	X-ray	2.40 Å	A	564-958	PDBe · RCSB-PDB · PDBJ · PDBsum
PDB	6IL3	X-ray	2.50 Å	A	564-958	PDBe · RCSB-PDB · PDBJ · PDBsum
PDB	6JQR	X-ray	2.20 Å	A	571-951	PDBe · RCSB-PDB · PDBJ · PDBsum
AlphaFold	AF-P36888-F1	Predicted			1-993	AlphaFold

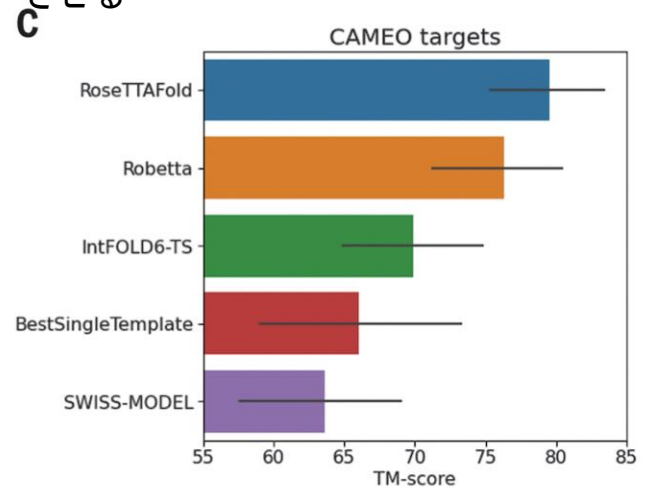
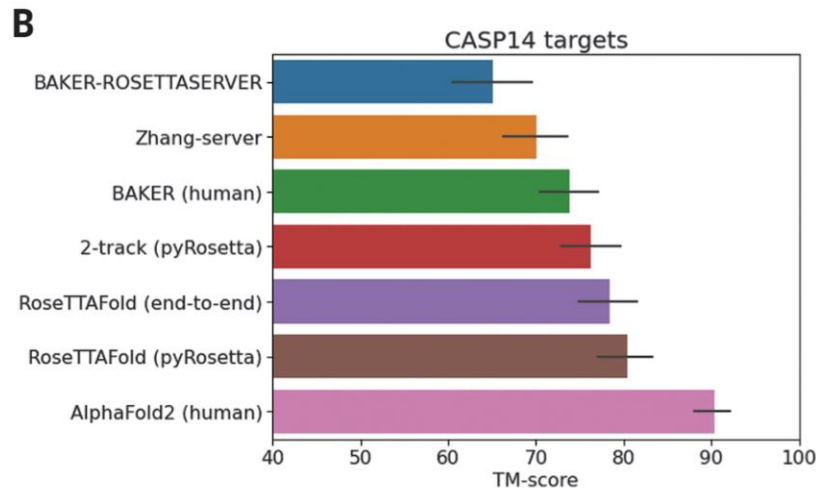
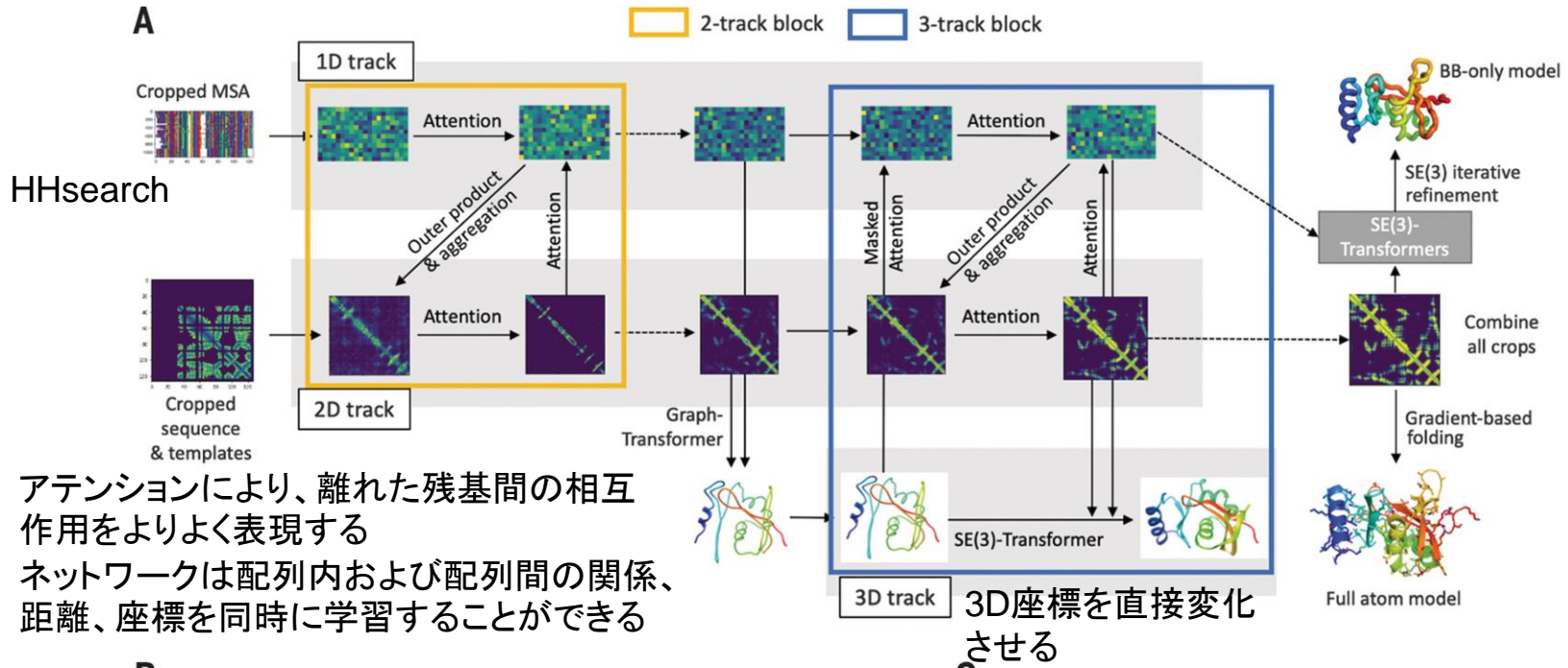


膜貫通領域

Phobiusや
TMHMMで膜
貫通領域と予
測された領域

參考資料

RoseTTAFold



Baek, M., et al. Science (2021) 373: 871–876

SWISS-MODEL

- ExPASyサーバから利用できるホモロジーモデリングツール
- <https://swissmodel.expasy.org/>

SWISS-MODEL

is a fully automated protein structure homology-modelling server, accessible via the **Expasy web server**, or from the program DeepView (Swiss Pdb-Viewer).

The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

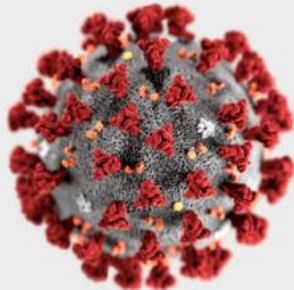
[Start Modelling](#)

Repository

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?



SWISS-MODELによる構造予測



SARS-CoV-2

Severe acute respiratory syndrome coronavirus 2, is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19).

We modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence [NC_045512](#) and annotations from [UniProt](#).

The results are available [here](#).

SWISS-MODEL Repository

Search SWISS-MODEL Repository

Fetch by UniProtKB AC or Entry Name:


SWISS-MODEL Homology Models: *B6VTU2*, *A0A4D6MRN2_VIGUN*, *A0A0R3P6F1*, *SYE_MIGAN*

Experimental Structures: *P11838*, *Q84II6_JANS3*, *Q5H5J0*, *AOX_TRYBB*

Or search using free text: *Hydroxymethylbilane synthase*, *Twin-arginine translocation pathway signal sequence domain protein*, *Mannitol-1-phosphate 5-dehydrogenase*, *Tyrosine-protein kinase Lck*

タンパク質の全配列のうち、モデルがどこまでカバーできているか










The SWISS-MODEL Repository is a database of annotated 3D protein structure models generated by the SWISS-MODEL homology-modelling pipeline.

Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017). The SWISS-MODEL Repository - new features and functionality *Nucleic Acids Res.* 45(D1):D313-D319.  [doi](https://doi.org/10.1093/nar/nkx104)

The aim of the SWISS-MODEL Repository is to provide access to an up-to-date collection of annotated 3D protein models generated by automated homology modelling for relevant model organisms and experimental structure information for all sequences in UniProtKB. Regular updates ensure that target coverage is complete, that models are built using the most recent sequence and template structure databases, and that improvements in the underlying modelling pipeline are fully utilised. It also allows users to assess the quality of the models using the latest QMEAN results. If a sequence has not been modelled, the user can build models interactively via the SWISS-MODEL workspace.

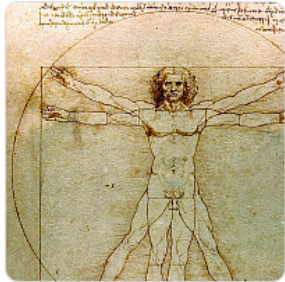
Currently the repository contains 2,211,366 models from SWISS-MODEL for UniProtKB targets as well as 166,639 structures from PDB with mapping to UniProtKB.

We currently provide models for the **reference proteomes** of the following model organisms, based on UniProtKB release 2021_02. If you want to download a large number of models, please contact us.

	Proteome Size	Sequences Modelled	Models	Seq Coverage	Metadata (Homology models and experimental structures)	Coordinates (Homology models only)
<i>Homo sapiens</i>	20,614	18,093	43,135		↓ 13.7 MB	↓ 4.4 GB
<i>Mus musculus</i>	21,990	19,422	43,359		↓ 8.0 MB	↓ 3.1 GB
<i>Caenorhabditis elegans</i>	19,818	13,837	24,078		↓ 3.8 MB	↓ 1.3 GB
<i>Escherichia coli</i>	4,391	3,624	6,002		↓ 1.7 MB	↓ 442.2 MB
<i>Arabidopsis thaliana</i>	27,468	21,606	38,434		↓ 5.8 MB	↓ 2.2 GB
<i>Drosophila melanogaster</i>	13,821	10,628	20,638		↓ 3.3 MB	↓ 1.3 GB
<i>Saccharomyces cerevisiae</i>	6,050	4,933	8,397		↓ 2.0 MB	↓ 497.0 MB
<i>Schizosaccharomyces pombe</i>	5,138	4,259	7,635		↓ 1.1 MB	↓ 454.5 MB
<i>Caulobacter vibrioides</i>	3,720	3,067	5,113		↓ 747.7 KB	↓ 346.6 MB

SWISS-MODEL Repository

Homo sapiens (Human)



wikimedia.org

Homo sapiens (Latin: "wise man") is the species name for humans. Believed to have originated in Africa, *Homo sapiens* is the only surviving member of the *Homo* genus.

The first complete draft of the human genome was completed in 2001 with an estimated cost of \$300million. With the advances in DNA sequencing methods, this price has dropped dramatically down to almost \$1000 per genome. By combining the individual genome information with other data, such as protein structure models, personalised medicine becomes viable.

"Homo sapiens", [Wikipedia: The Free Encyclopedia](#)

Protein models in Repository

From left to right: i) The number of proteins in the reference proteome of *Homo sapiens*, ii) the number of unique protein sequences for which at least one model is available, iii) the total number of models and iv) a coverage bar plot is shown.

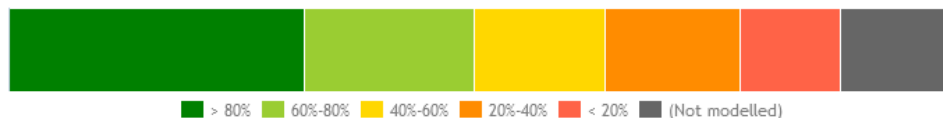
The bar plot shows the coverage for every protein in the reference proteome of *Homo sapiens* for which there is at least one model. Different colours (dark green to red boxes) represent the coverage of the targets. Targets with high coverage are represented in dark green (more than 80% of the target's length is covered by models), whereas low coverage is shown in red. The size of each box is proportional to the number of target sequences with a given coverage.

For information on the latest proteome for *Homo sapiens*, please visit [Proteome](#).

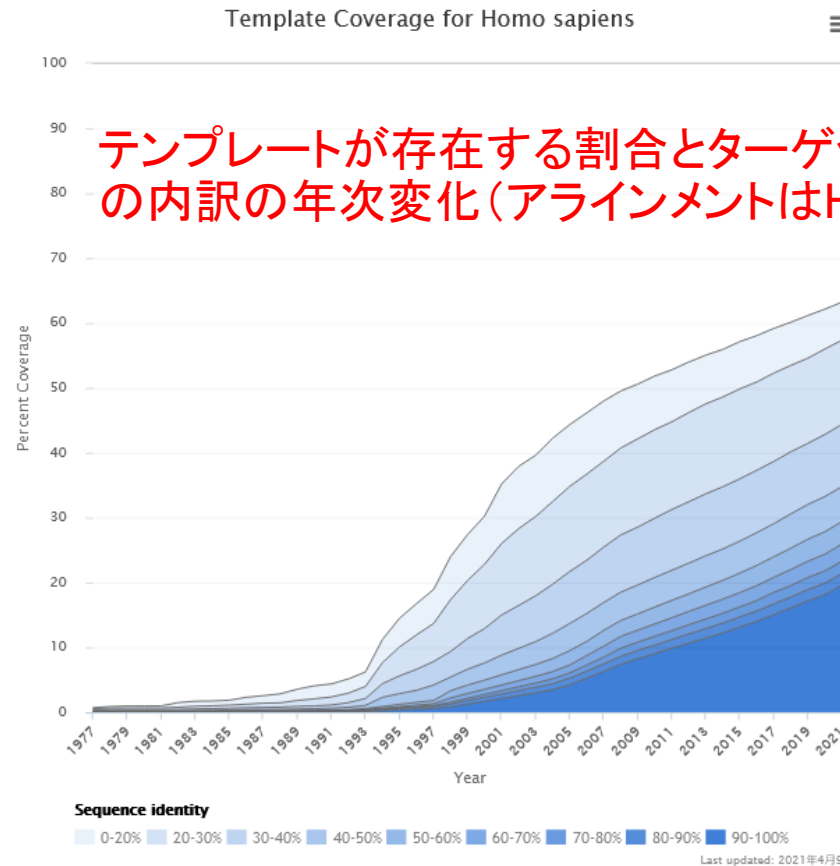
You can easily download the latest protein sequences for *Homo sapiens* proteome [here](#). Please note this download is for the current UniProtKB release, which may be different to release 2021_02 that was used for the most up to date SWISS-MODEL Repository.

Proteins in proteome	Sequences modelled	Models
20,614	18,093	43,135

Detailed coverage numbers are obtained by hovering the mouse over one of the boxes.

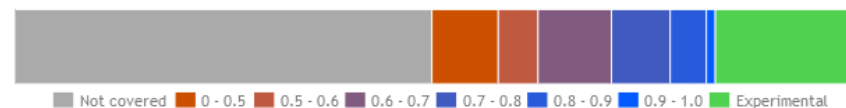


SWISS-MODEL Repository



Residue Coverage

This chart shows the percentage of residues in the *Homo sapiens* proteome which are covered by experimental structures and the enhancement of coverage by homology modelling by the SWISS-MODEL pipeline. Experimental residue coverage is determined using SIFTS mapping. For residues which are not covered by experimental structures (including where there are no atom records in SIFTS mapping) the model coverage bars are coloured by QMEANDisCo local quality score.



ヒトのタンパク質全体でモデルおよび実験でカバーできている割合(数値はQMEANで表されるモデルの質)

SWISS-MODELの基本手順

ターゲットの配列

テンプレートのアノテーション

InterProによるドメインスキャン
PSIPredによる二次構造予測
DISOPREDIによる天然変性領域予測
MEMSATによる膜貫通領域予測

テンプレートの選択

配列類似度が高いときはBLAST、
そうでないときは、PSI-BLAST、HMM-HMMIに基づく
探索(HHblits)を行い、1つまたは複数のテンプレート
を選択

テンプレート

テンプレートの調整

マルチプルアライ
メントの利用

DeepView Project

ターゲットと複数テンプレートのマルチプル
アライメント
複数テンプレートの重ね合わせ、代表テン
プレートから外れたテンプレートは除外

異なる手法による複数の
アライメント候補の提示
と人手による修正

モデル構造の構築

Alignment
mode

Automated
mode

Project
mode

テンプレートの主鎖の平均構造をあてはめる
(配列類似度に応じて重みづけ)、フラグメン
トベースでループ構造をあてはめる
GROMOSを使ったエネルギー最小化

モデル構造とレポート

モデル構造の評価

ターゲットとテンプレートの配列一致度
立体化学のチェック(WHATCHECK, PROCHEK)
グローバルモデルクオリティチェック(DFIRE)
ローカルグローバルモデルクオリティチェック(ProQres,
ANOLEAスコアなど)

SWISS-MODELによる構造予測(1)

The screenshot shows the SWISS-MODEL web interface for starting a new modelling project. The interface includes a header with the SWISS-MODEL logo and navigation links (Modelling, Repository, Tools, Documentation, Login, Create Account). The main section is titled 'Start a New Modelling Project' and contains several input fields and buttons:

- Target Sequence(s):** A text input field with a placeholder 'Paste your target sequence(s) or UniProtKB AC here'. Below it are instructions: '(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)'. There is a green '+ Upload Target Sequence File...' button and a grey 'Validate' button.
- Supported Inputs:** A dropdown menu on the right with four options: 'Sequence(s)', 'Target-Template Alignment', 'User Template', and 'DeepView Project'. The last three options are circled in red.
- Project Title:** A text input field containing 'Untitled Project'.
- Email:** A text input field containing 'Optional'.
- Buttons:** Two large blue buttons at the bottom: 'Search For Templates' and 'Build Model'. Both are circled in red.

Red annotations and Japanese text are overlaid on the screenshot to explain the workflow:

- ターゲットとテンプレートのアラインメントをユーザが指定する** (User specifies alignment of target and template) - points to the 'Supported Inputs' dropdown.
- ユーザがテンプレートを指定** (User specifies template) - points to the 'Target-Template Alignment' and 'User Template' options in the dropdown.
- ユーザが構造を見ながらターゲットとテンプレートのアラインメントを調整する** (User adjusts alignment of target and template while viewing structure) - points to the 'Build Model' button.
- テンプレートの探索を行う** (Perform template search) - points to the 'Search For Templates' button.
- モデル構造を構築する** (Construct model structure) - points to the 'Build Model' button.
- (既知構造のモデリングを行うと、その構造そのものをテンプレートとしてモデリングしてしまう)** (If you perform modeling of known structures, you will end up modeling that structure as a template) - points to the 'Build Model' button.

At the bottom of the interface, there is a disclaimer: 'By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.'

SWISS-MODELによる構造予測(2)

「Build Model」を選択すると、PDBに登録されている構造既知のタンパク質をモデリングするとき、その構造がテンプレートとして自動的に選ばれてしまうので、「Search For Templates」を実行する

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

```
>p1
VILTGNSSLCPISGWAIYSKDNIGIRIGSKGQVVFVIREPFISCSHLECRTEFLTQGALLNDKHSNGIVKDRSPYRILMSCP
VGEAPSPYNSRFESVAWSASACHDGMGWLITIGISGPNDAVAVLKYNGIITDIKSWRNINILRQISECACVNGSCFTIM
TDGPSIQQASYKILKIEKGKVTKSIELNAPNYHYEECSCHPDTGKVMCVCRDNNHGSNRPFVWVDFQNLDYQIGYICSGVF
GDNPRENDGTGSCGPFVSSNGANGIKGFSTYDNGVWIGRTKSTSSRSGFEMIWDPNGWTEIDSSFSVRQDIVAITDWSGY
SGSFVQHPPELITGLDCMRPCFWVELIRGQPKENIWTSGSSISFCGVNSDITVGWSWPDGAELPFSTI
```

ターゲットの配列を入力
p1.fastaを入力

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

Project Title:

Untitled Project
Optional

Search For Templates

Build Model



入力すると、カラー表示に変化

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

```
Target: VILTGNSSLCPISGWAIYSKDNIGIRIGSKGQVVFVIREPFISCSHLECRTEFLTQGALLNDKHSNGIVKDRSPYRILMSCP 110
Target: IGISGPNDAVAVLKYNGIITDIKSWRNINILRQISECACVNGSCFTIMTDGPSIQQASYKILKIEKGKVTKSIELNAPNYHYEECSCHPDTGKVMCVCRDNNHGSNR 220
Target: WVSFDQNLDYQIGYICSGVFGDNPRENDGTGSCGPFVSSNGANGIKGFSTYDNGVWIGRTKSTSSRSGFEMIWDPNGWTEIDSSFSVRQDIVAITDWSGYSGSFVQHPPEL 330
Target: IGLDCMRPCFWVELIRGQPKENIWTSGSSISFCGVNSDITVGWSWPDGAELPFSTI 385
```

Add Hetero Target Reset

Project Title:

Untitled Project

Email:

Optional

Search For Templates

Build Model

SWISS-MODELによる構造予測(3)

All Projects

モデリング中に表示される画面

Untitled Project Created: today at 18:34

Summary

Templates

Models

📄 ⬇️ ✕

Template Results

The search for templates matching your target sequence is currently running. Please wait.

```
...running HHblits against SMTL
...reusing cached query alignment
...predicting residue burial status with ACCpro
...searching PDB profile database with previously built query profile
...running BLAST against SMTL
...extracting distance constraints from 178 templates
...filtering list of templates
...structurally superpose templates
...predicting oligomeric state conservation
```

If you want to come back later, bookmark this link:

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

```
VILTGNSSLCPISGWAIYSKDNGIRIGSKGDVVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWSASACHDG
MGWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGSCFTIMTDGSPNGQASYKILKIEKGKVTKSIELNAPNYHYEECSYPTDGKVMCVC
RDNWHGNSRNPWVSFDQNLDYQIGYICSGVFGDNPRPNDGTGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWDPNGWTETDSSFSVRQDIVAIT
DWSGYSGSFVQHPELTGLDCMRPCFWVELIRGQPKENTIWTSGSSISFCGVNSDVTVGWSWPDGAE LPPSI
```

SWISS-MODELによる構造予測(4)

All Projects

Untitled Project Created: today at 19:57

登録されたテンプレートからの選択 → 「Build Models」ボタンを押す

Summary **Templates 27** Models Project Data ▾

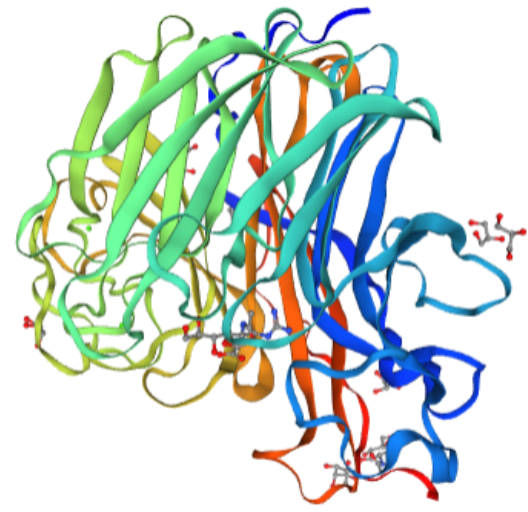
Template Results

デフォルトは1位でヒットした既知構造
(今回はチェックをはずす)

Templates	Quaternary Structure	Sequence Similarity	Alignment	More ▾			
<input checked="" type="checkbox"/> 3b7e.1.A Neuraminidase <i>Neuraminidase of A/Brevig Mission/1/1918 H1N1 strain in complex with zanamivir</i>		1.02	1.00	100.00	X-ray, 1.4Å	homo-tetramer	2 x NAG-FUC, 2 x NAG ^{CE} , 8 x CA ^{CE} , 4 x ZMR ^{CE}
<input type="checkbox"/> 6d96.1.A Neuraminidase <i>Structure of influenza neuraminidase from strain A/BrevigMission/1/1918(H1N1) expressed in HEK-293F cells</i>		0.81	0.92	48.04	X-ray, 2.9Å	hetero-12-mer	4 x NAG-NAG-BMA-MAN-MAN-MAN, 4 x NAG-NAG, 4 x NAG ^{CE} , 4 x CA ^{CE}
<input type="checkbox"/> 1ncc.1.A INFLUENZA A SUBTYPE N9 NEURAMINIDASE <i>CRYSTAL STRUCTURES OF TWO MUTANT NEURAMINIDASE-ANTIBODY COMPLEXES WITH AMINO ACID SUBSTITUTIONS IN THE INTERFACE</i>		0.83	0.92	47.50	X-ray, 3.5Å	hetero-12-mer	4 x NAG-NAG-BMA-MAN-MAN-MAN, 4 x NAG ^{CE} , 4 x CA ^{CE}
<input type="checkbox"/> 1nma.1.A N9 NEURAMINIDASE <i>N9 NEURAMINIDASE COMPLEXES WITH ANTIBODIES NC41 AND NC10: EMPIRICAL FREE-ENERGY CALCULATIONS CAPTURE SPECIFICITY TRENDS OBSERVED WITH THE NATURAL INFLUENZA DATA</i>		0.78	0.75	48.17	X-ray, 3.0Å	hetero-12-mer	4 x NAG-NAG-BMA-MAN-MAN-MAN, 4 x NAG ^{CE}

Build Models 1
Clear Selection

チェックした構造が表示される



このリストからテンプレートを選ぶ場合は、左のボックスにチェックを入れて、右のグラフィックスの上の「Build Models」を選択する。

今回は、N8ノイラミニダーゼが見当たらないので、全体の候補を見るため、末尾のfull listを指定する

The full list of templates matching your target sequence includes the following templates which are not in the list above. The full template list is available in **text** or **html** format.

1a4g.1.A, 1b9v.1.D, 1inf.1.A, 1ing.1.A, 1inh.1.A, 1iv1.1.A, 1ive.1.A, 1ivg.1.A, 1i7g.1.A, 1mwe.1.C, 1nca.2.C, 1ncb.1.A, 1ncc.1.A, 1ncd.1.A, 1nma.1.A, 1nmb.1.A, 1nna.1.D, 1nvj.1.A, 2aep.1.A, 2bat.1.A, 2b8h.1.A, 2bat.1.A, 2ht7.1.A, 2htv.1.A, 2hu0.1.A, 2qwd.1.A, 3b7e.1.A, 3cl0.1.A, 3cl2.1.A, 3cye.1.A, 3f14.1.A, 3k38.1.A, 3k38.1.A, 3nn9.1.A, 3o9j.1.A, 3sal.1.A, 3tia.1.A, 4b7m.1.C, 4b7q.1.A, 4b7r.1.A, 4cpl.1.A, 4cpo.1.A, 4d8s.1.A, 4fvk.1.A, 4gb1.1.A, 4gdi.1.A, 4gdi.1.A, 4gdj.1.A, 4gez.1.A, 4gzo.1.A, 4gzs.1.A, 4h52.1.A, 4h53.1.A, 4h53.1.D, 4hzv.1.A, 4hzy.1.A, 4hzz.1.A, 4k1j.1.A, 4k3y.1.A, 4k3y.1.C, 4ks5.1.A, 4m3m.1.A, 4mc7.1.A, 4nju.1.A, 4mju.1.A, 4mju.1.A, 4mwj.1.B, 4nn9.1.A, 4qn3.1.A, 4qn4.1.A, 4qn4.1.A, 4qnp.1.A, 4wa5.1.D, 5hug.1.A, 5huk.1.A, 5hum.1.A, 5hun.1.A, 5i14.1.A, 5nn9.1.A, 5nwe.1.A, 5nz4.1.A, 5nze.1.A, 5nzf.1.A, 5nzn.1.A, 6br5.1.A, 6ard.1.A, 6ard.1.B, 6ard.1.C, 6ard.1.D, 6d96.1.A, 6g01.2.D, 6hfv.1.D, 6ho0.1.C, 6lxi.1.A, 6lxk.1.A, 6n4d.1.A, 6nb8.1.A, 6nn9.1.A, 6oze.1.A, 6ozf.1.A, 6ozw.1.A, 6o20.1.A, 6o23.1.A, 6v4n.1.A, 6v4o.1.A, 7cm1.1.A

Cartoon ▾

SWISS-MODELによる構造予測(5)

189 Unfiltered Template Results

Models	Name	Description	GMQE	QSQE	Seq Id	Coverage	Range	Method	Resolution	Oligo-state	Ligands	Found by	Seq Similarity
Build Homomer Build Monomer	3b7e.1.A	Neuraminidase Neuraminidase of A/Brevig Mission/1/1918 H1N1 strain in complex with zanamivir	1.02	1.00	100.00	1.00	1-385	X-ray	1.45	homo-tetramer	2 x NAG, 8 x CA, 4 x ZMR, 2 x NAG-FUC	0.64	HHblits
Build Homomer Build Monomer	3cye.1.A	Neuraminidase Crystal structure of the native 1918 H1N1 neuraminidase from a crystal with lattice-translocation defects	1.01	1.00	100.00	1.00	1-385	X-ray	1.65	homo-tetramer	2 x NAG, 8 x CA, 2 x NAG-NAG-MAN-MAN	0.64	HHblits
Build Homomer Build Monomer	6d96.1.A	Neuraminidase Structure of influenza neuraminidase from strain A/BrevigMission/1/1918(H1N1) expressed in HEK-293E cells	1.02	1.00	100.00	1.00	1-385	X-ray	2.15	homo-tetramer	9 x NAG, 8 x CA, 2 x NAG-NAG-FUC, 1 x NAG-FUC	0.64	HHblits
Build Homomer Build Monomer	6lxi.1.A	Neuraminidase Crystal structure of Z2B3 Fab in complex with influenza virus neuraminidase from A/Brevig Mission/1/1918 (H1N1)	1.01	1.00	100.00	1.00	1-385	X-ray	2.50	homo-tetramer	8 x CA, 4 x NAG-NAG-BMA-MAN-FUC	0.64	BLAST
Build Homomer Build Monomer	6lxi.1.A	Neuraminidase Crystal structure of Z2B3 Fab in complex with influenza virus neuraminidase from A/Brevig Mission/1/1918 (H1N1)	1.01	1.00	100.00	1.00	2-385	X-ray	2.50	homo-tetramer	8 x CA, 4 x NAG-NAG-BMA-MAN-FUC	0.64	HHblits
Build Homomer Build Monomer	3o9j.1.A	Neuraminidase Influenza NA in complex with compound 5	0.87		56.92	0.99	3-385	X-ray	2.00	homo-tetramer	4 x RP6, 4 x CA, 4 x NDG	0.49	BLAST
Build Homomer Build Monomer	4d8s.1.A	Neuraminidase Influenza NA in complex with antiviral compound	0.86		56.92	0.99	3-385	X-ray	2.40	homo-tetramer	4 x CA, 4 x 0HX	0.49	BLAST
Build Homomer Build Monomer	4gb1.1.A	Neuraminidase Synthesis and Evaluation of Novel 3-C-alkylated-Neu5Ac2en Derivatives as Probes of Influenza Virus Sialidase 150-loop flexibility	0.87		56.92	0.99	3-385	X-ray	2.62	homo-tetramer	4 x 0LP, 4 x CA	0.49	BLAST
Build Homomer Build Monomer	4wa5.1.D	Neuraminidase The crystal structure of neuraminidase from a H3N8 influenza virus isolated from New England harbor seals in complex with zanamivir	0.87		58.07	1.00	2-385	X-ray	1.95	homo-tetramer	4 x ZMR, 4 x CA, 4 x NAG-FUC	0.48	HHblits
Build Homomer Build Monomer	2ht7.1.A	Neuraminidase N8 neuraminidase in open complex with oseltamivir	0.85		56.66	0.99	3-385	X-ray	2.60	homo-tetramer	4 x G39	0.48	BLAST
Build Homomer Build Monomer	5hun.1.A	Neuraminidase The crystal structure of neuraminidase from A/Indonesia/Westport/09/2009 influenza virus	0.88		56.51	1.00	2-385	X-ray	2.30	homo-tetramer	4 x CA	0.48	HHblits
Build Homomer Build Monomer	5hun.1.A	Neuraminidase The crystal structure of neuraminidase from A/gyrfalcon/Washington/4/1088-6/2014 influenza virus	0.87		57.56	0.98	9-385	X-ray	2.30	homo-tetramer	4 x CA	0.49	BLAST

N8ノイラミナーゼの2ht7.1.Aを選択
「Build Model (Homomer)」


SWISS-MODELによる構造予測(6)

All Projects

Untitled Project Created: today at 08:14

Summary Templates 27 Models 1

Model Results



Model 01

Oligo-State: Homo-tetramer (requested by user)

Ligands: None

Global Quality Estimate

QMEAN	-2.26
Cβ	-1.32
All Atom	-2.28
solvation	-1.46
torsion	-1.46

Local Quality Estimate

Comparison

Template: 2ht7.1.A Seq Identity: 56.51% Coverage: [Progress bar]

Description: Neuraminidase

Model-Template Alignment

```

Model_01:A VILTGNSSLCPISGWAIYSKDNGIRIGSKGDV FVIREPPI SC SHLE CRTFFLTQCALLNDKHSNGT VKDRSPYR TLMSCP 80
Model_01:B VILTGNSSLCPISGWAIYSKDNGIRIGSKGDV FVIREPPI SC SHLE CRTFFLTQCALLNDKHSNGT VKDRSPYR TLMSCP 80
Model_01:C VILTGNSSLCPISGWAIYSKDNGIRIGSKGDV FVIREPPI SC SHLE CRTFFLTQCALLNDKHSNGT VKDRSPYR TLMSCP 80
Model_01:D VILTGNSSLCPISGWAIYSKDNGIRIGSKGDV FVIREPPI SC SHLE CRTFFLTQCALLNDKHSNGT VKDRSPYR TLMSCP 80
2ht7.1.A -YNNNTERRCDARGFAPFSKDNGIRIGSRG H FVIREPPI SC SPIECRTFFLTQGSLLNDKHSNGT VKDRSPYR TLMSCP 80
Model_01:A VGEAPSPYNSRFESVAVWSASACHDGMGLTIGISGPDNGAVAVLKYNGI ITDTIKSWRNILRTOES E CACVNGSCFTIM 160
Model_01:B VGEAPSPYNSRFESVAVWSASACHDGMGLTIGISGPDNGAVAVLKYNGI ITDTIKSWRNILRTOES E CACVNGSCFTIM 160
Model_01:C VGEAPSPYNSRFESVAVWSASACHDGMGLTIGISGPDNGAVAVLKYNGI ITDTIKSWRNILRTOES E CACVNGSCFTIM 160
Model_01:D VGEAPSPYNSRFESVAVWSASACHDGMGLTIGISGPDNGAVAVLKYNGI ITDTIKSWRNILRTOES E CACVNGSCFTIM 160
2ht7.1.A VGSSEPIYQARFEVAVWSA ACQDGGK VNTIGY GPDSAVAVLH G G ITD V VSWAGDILRTOES S C C GDCVYIN 160
Model_01:A TDGPNNGQASYKILKIEKGVTKTSIELNAPNYHYE ECSCYPDT GKVMCVCRDNWHGSNRPVWS PDQNLDYQIGYICSGVF 240
Model_01:B TDGPNNGQASYKILKIEKGVTKTSIELNAPNYHYE ECSCYPDT GKVMCVCRDNWHGSNRPVWS PDQNLDYQIGYICSGVF 240
Model_01:C TDGPNNGQASYKILKIEKGVTKTSIELNAPNYHYE ECSCYPDT GKVMCVCRDNWHGSNRPVWS PDQNLDYQIGYICSGVF 240
Model_01:D TDGPNNGQASYKILKIEKGVTKTSIELNAPNYHYE ECSCYPDT GKVMCVCRDNWHGSNRPVWS PDQNLDYQIGYICSGVF 240
2ht7.1.A TDGPNRQA VYRIYK G R IEGD D S FNGGH I ECSCYP D GKVCVCRD WIG I NRP V D PDL VYR V G Y CAG I P 240
Model_01:A GDNPRPNDG --TGSCGPVSSNANGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWD PNCWTET DSSFSVRQDIVAITDWS 318
Model_01:B GDNPRPNDG --TGSCGPVSSNANGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWD PNCWTET DSSFSVRQDIVAITDWS 318
Model_01:C GDNPRPNDG --TGSCGPVSSNANGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWD PNCWTET DSSFSVRQDIVAITDWS 318
Model_01:D GDNPRPNDG --TGSCGPVSSNANGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWD PNCWTET DSSFSVRQDIVAITDWS 318
2ht7.1.A SDI PR G E D I O T G S C T S P M G N Q G Y G V K G F S F R D S T V W V G S T I S R T S R S G F E I E R N G W T Q T S K E I R K Q V V V D N I S 320
Model_01:A GYSGSFVQHP ELTGLDCMRPCFWVELLRGQPKENT IWTSGSSISFCGVNSDITVGWSWPDGAELPFSTI 385
Model_01:B GYSGSFVQHP ELTGLDCMRPCFWVELLRGQPKENT IWTSGSSISFCGVNSDITVGWSWPDGAELPFSTI 385
                    
```

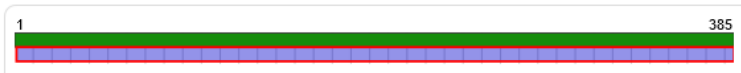
Order by: GMQE

GMQE: 0.85 QMEAN: -2.26

PDB形式の座標のダウンロード

DeepViewer/Swiss-PdbViewer表示用の座標のダウンロード (ターゲットとテンプレートとの重ね合わせ)

モデルに関する情報



SWISS-MODELによる構造予測(7)

Oligo-State: Homo-tetramer (requested by user)
Ligands: None

Global Quality Estimate

QMEAN	-2.26
C β	-1.32
All Atom	-2.28
solvation	-1.46
torsion	-1.46

Model 01

- ↓ PDB Format
- ↓ mmCIF Format beta
- ↓ JSON Format
- ↓ DeepView Format
- ↓ Model Report
- PDB Format (Display)
- JSON Format (Display)
- Model Report (Display)**
- Send to SwissDock
- Modelling Logs
- Delete Model

結果のレポート
の表示

モデル構造のデータ
はここからもダウン
ロードできる

Model Building Report

This document lists the results for the homology modelling project "Untitled Project" submitted to SWISS-MODEL workspace on Dec. 17, 2020, 8:14 a.m. The submitted primary amino acid sequence is given in Table T1.

If you use any results in your research, please cite the relevant publications:

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46(W1), W296-W303 (2018). [PMID](#) [DOI](#)
- Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, S162-S173 (2009). [PMID](#) [DOI](#)
- Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res.* 45, D313-D319 (2017). [PMID](#) [DOI](#)
- Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics* 36, 1765-1771 (2020). [PMID](#) [DOI](#)
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7 (2017). [PMID](#) [DOI](#)

Results

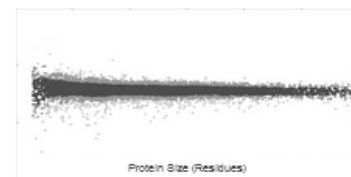
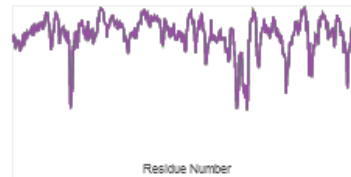
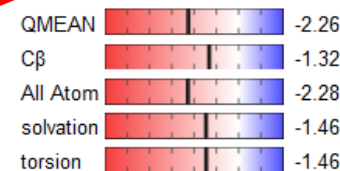
The SWISS-MODEL template library (SMTL version 2020-12-16, PDB release 2020-12-11) was searched with BLAST (Camacho et al.) and HHblits (Steinegger et al.) for evolutionary related structures matching the target sequence in Table T1. For details on the template search, see Materials and Methods. Overall 189 templates were found (Table T2).

Models

構造評価の結果

The following model was built (see Materials and Methods "Model Building"):

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN
	PDB	ProMod3 3.2.0	homo-tetramer (requested by user)	None	0.85	-2.26



Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
2ht7_1 A	56.51	homo-		HHblits	X-ray	2.60Å	0.48	2 - 385	1.00	Neuraminidase

SWISS-MODELによる構造予測(8)

テンプレートに結合しているリガンドのうち、保存性のあるものは、モデルに付加され、そうでないものは、外される

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
2ht7.1.A	56.51	homo-tetramer	-	HHblits	X-ray	2.60Å	0.48	2 - 385	1.00	Neuraminidase

Excluded ligands

Ligand Name.Number	Reason for Exclusion	Description
G39.1	Binding site not conserved.	(3R,4R,5S)-4-(ACETYLAMINO)-5-AMINO-3-(PENTAN-3-YLOXY)CYCLOHEX-1-ENE-1-CARBOXYLIC ACID
G39.2	Binding site not conserved.	(3R,4R,5S)-4-(ACETYLAMINO)-5-AMINO-3-(PENTAN-3-YLOXY)CYCLOHEX-1-ENE-1-CARBOXYLIC ACID
G39.3	Binding site not conserved.	(3R,4R,5S)-4-(ACETYLAMINO)-5-AMINO-3-(PENTAN-3-YLOXY)CYCLOHEX-1-ENE-1-CARBOXYLIC ACID
G39.4	Binding site not conserved.	(3R,4R,5S)-4-(ACETYLAMINO)-5-AMINO-3-(PENTAN-3-YLOXY)CYCLOHEX-1-ENE-1-CARBOXYLIC ACID

Target VILTGNSLCPISGWA IYSKDNGIRIGSKGDV FVIREPFISCSHLECR TFFLTQGALLNDKHSNGTVKDRSPYRTLMSCP
2ht7.1.A -YMNTEAICDAKGFAPFSKDN GIRIGSRGHIFVIREPFVSCSPIECRTFFLTQGSLLNDKHSNGTVKDRSPFRTLMSVE

Target VGEAPSPYNSRFESVAWSASACHDGMWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESEACAVNGSCFTIM
2ht7.1.A VGGSPNVYQARFEAVAWSATACHDGKKWMTVGVTPGDSKAVAVIHYGGVPTD VVNSWAGDILRTQESSCTCIQGDICYWWM

Target TDGPSNGQASYKILKIEKGKVTKSIELNAPNYHYEECSYPTDGKVMCVCRDNW HGSNRPWV SFDQNL DYQIGYICSGVF
2ht7.1.A TDGPANRQAQYRIYKANQGR IIGQTDISFN GGHIEECSCYPNDGKVECVCRD GWTGTNRPLVVISPDLSYRVGYLCAGIP

Target GDNPRPNDG--TGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRS GFEMI WDPNGW TETDSSFSVRQDIVAITDWS
2ht7.1.A SDTPRGEDTQFTG SCTSPMGNQGYGVKGFGRQGT D VWMGRTISRTSRSGFEILRIKNGW TQTSKEQIRKQVVDNLNWS

Target GYSGSFVQHP ELTGLDCMRPCFVVELIRGQPKENTI WTS GSSISFCGVNSD TVGWSWPDGAELPF SI
2ht7.1.A GYSGSFTLPVELSGKDCLVPCFVWEMIRGKPEEKTI WTS SSSIVMCGVDYEVADWSWHDGAILPFDI

Target VILTGNSLCPISGWA IYSKDNGIRIGSKGDV FVIREPFISCSHLECR TFFLTQGALLNDKHSNGTVKDRSPYRTLMSCP
2ht7.1.D -YMNTEAICDAKGFAPFSKDN GIRIGSRGHIFVIREPFVSCSPIECRTFFLTQGSLLNDKHSNGTVKDRSPFRTLMSVE

Target VGEAPSPYNSRFESVAWSASACHDGMWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESEACAVNGSCFTIM
2ht7.1.D VGGSPNVYQARFEAVAWSATACHDGKKWMTVGVTPGDSKAVAVIHYGGVPTD VVNSWAGDILRTQESSCTCIQGDICYWWM

Target TDGPSNGQASYKILKIEKGKVTKSIELNAPNYHYEECSYPTDGKVMCVCRDNW HGSNRPWV SFDQNL DYQIGYICSGVF
2ht7.1.D TDGPANRQAQYRIYKANQGR IIGQTDISFN GGHIEECSCYPNDGKVECVCRD GWTGTNRPLVVISPDLSYRVGYLCAGIP

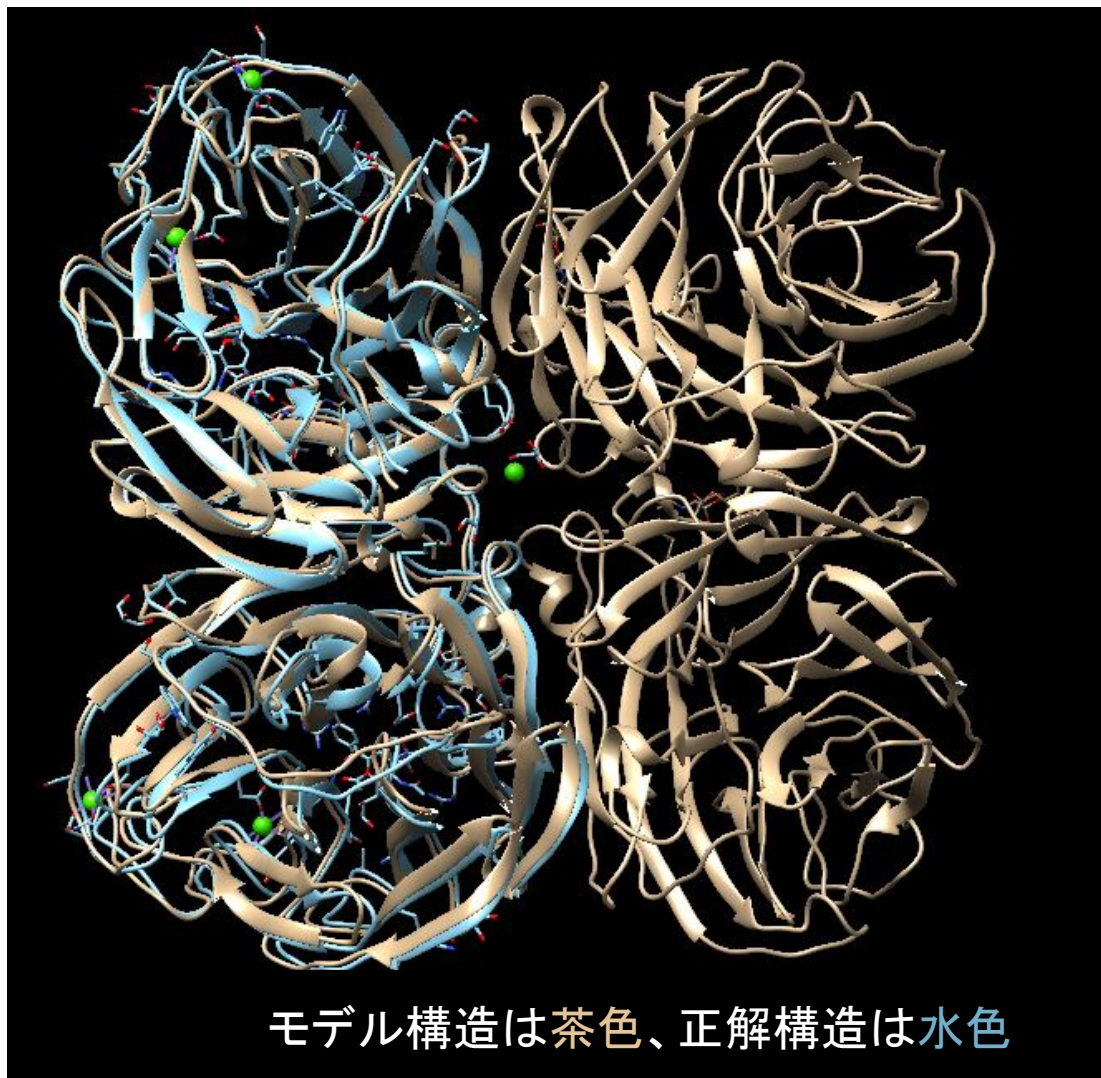
Target GDNPRPNDG--TGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRS GFEMI WDPNGW TETDSSFSVRQDIVAITDWS
2ht7.1.D SDTPRGEDTQFTG SCTSPMGNQGYGVKGFGRQGT D VWMGRTISRTSRSGFEILRIKNGW TQTSKEQIRKQVVDNLNWS

Target GYSGSFVQHP ELTGLDCMRPCFVVELIRGQPKENTI WTS GSSISFCGVNSD TVGWSWPDGAELPF SI
2ht7.1.D GYSGSFTLPVELSGKDCLVPCFVWEMIRGKPEEKTI WTS SSSIVMCGVDYEVADWSWHDGAILPFDI

モデリングに用いた
ターゲットとテンプレート
のアラインメント

SWISS-MODELによる構造予測(9)

- モデル構造と正解構造(結晶構造、PDB ID: 3b7e)の構造アラインメントを実行 → Chimeraを利用



テンプレートの2ht7が4量体であることから、モデル構造も4量体、3b7eは2量体

A鎖の重ね合わせで C_{α} RMSDは、
全長385残基で1.051 Å、
363残基で0.711 Å
(Chimera 1.5の結果)

BLASTによるテンプレート検索(1)

blastn **blastp** blastx

NCBIのBLASTを利用したテンプレートの検索とアラインメントの取得

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. [more...](#)

Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
GDNPRPNDGTGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFE
MIWDPNGWTEEDSSFSVRQDIVAITDWSGY
SGSFFVQHPELTGLDCMRPCFVWELIRGQPKENTIWTSGSSISFCGVNSDTV
GWSWPDGAELPFSI
```

タンパク質p1「p1.fasta」の配列を入力

Or, upload file ファイルが選択されていません p1

Align two or more sequences

- Non-redundant protein sequences (nr)
- RefSeq Select proteins (refseq_select)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot (swissprot)
- Patented protein sequences (pataa)
- Protein Data Bank proteins (pdb)**
- Metagenomic proteins (env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

Choose Search Set

Database exclude

Organism Optional exclude

Exclude Optional Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

データベースは「Protein Data Bank proteins (pdb)」を選択

Program Selection

Algorithm blastp (protein-protein BLAST) PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

「blastp」を選択

Search database **pdb** using **Blastp** (protein-protein BLAST) Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

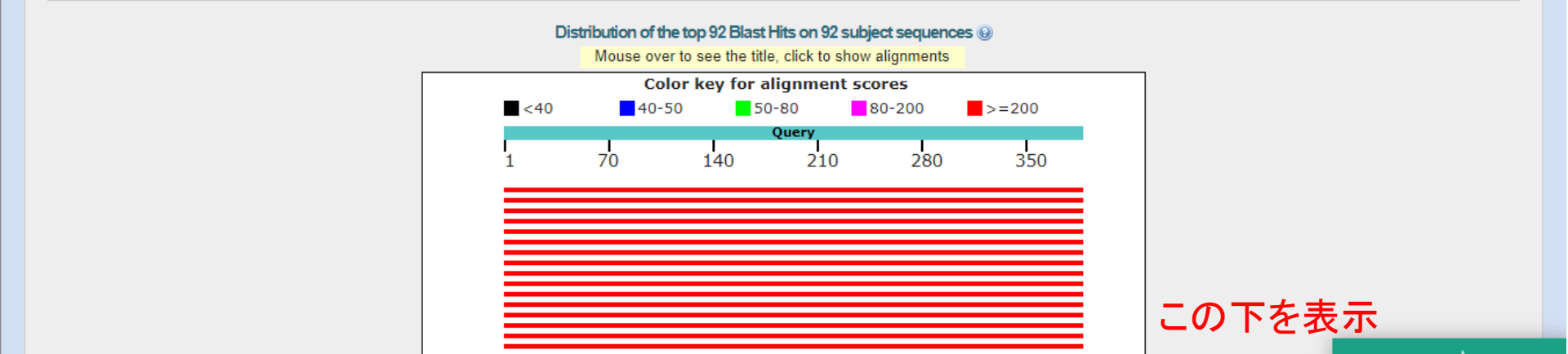
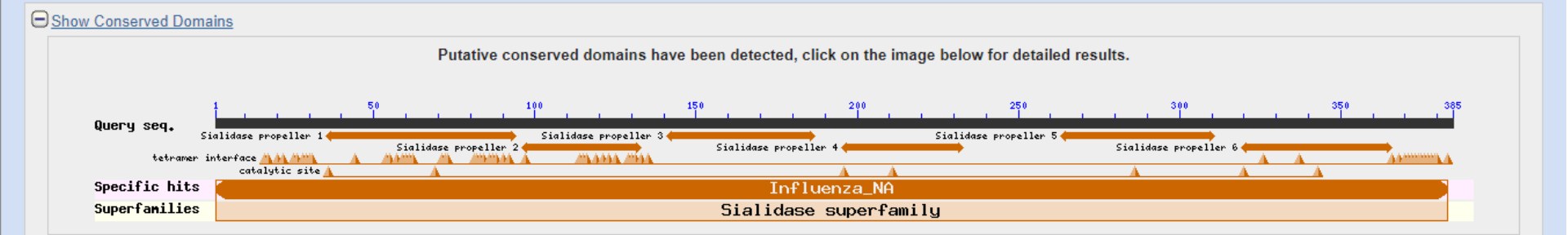
参考 BLASTによるテンプレート検索(2)

BLAST® » blastp suite » RID-W040ETAT016 Home Recent Results Saved Strategies Help

BLAST Results
Edit and Resubmit Save Search Strategies Formatting options Download YouTube How to read this page Blast report description NEW Click here to use the new BLAST results page
Job title: p1

RID W040ETAT016 (Expires on 12-21 12:09 pm)
Query ID lcl|Query_41543 Database Name pdb
Description p1 Description PDB protein database
Molecule type amino acid Program BLASTP 2.12.0+ Citation
Query Length 385
Other reports: Search Summary Taxonomy reports Distance tree of results Multiple alignment MSA viewer
Analyze your query with SmartBLAST

Graphic Summary



この下を表示

BLASTによるテンプレート検索(3)

<input type="checkbox"/>	Influenza Neuraminidase in complex with a stereomutated analogue of Oseltamivir carboxylate [Influenza A virus (A/duck/Ukraine/1/1963(H3N8))]	469	469	99%	3e-165	56.88%	4M3M_A
<input type="checkbox"/>	Influenza Neuraminidase in complex with a novel antiviral compound [Influenza A virus (A/duck/Ukraine/1/1963(H3N8))]	468	468	99%	4e-165	56.88%	4MJU_A
<input type="checkbox"/>	Influenza Neuraminidase in complex with a novel antiviral compound [Influenza A virus (A/duck/Ukraine/1/1963(H3N8))]	468	468	99%	7e-165	56.62%	4MJV_A
<input type="checkbox"/>	The crystal structure of neuraminidase from a H3N8 influenza virus isolated from New England harbor seals [Influenza A virus (A/harbor seal/Massachusetts/1/2011(H3N8))]	466	466	99%	3e-164	57.77%	4WA3_A
<input type="checkbox"/>	Influenza NA in complex with Compound 9 [Influenza A virus (A/duck/Ukraine/1/1963(H3N8))]	466	466	99%	5e-164	56.62%	3O9J_A
<input type="checkbox"/>	Chain A_Neuraminidase [Influenza A virus (A/duck/Ukraine/1/1963(H3N8))]	465	465	99%	7e-164	56.62%	4D8S_A
<input type="checkbox"/>	N8 Neuraminidase [Influenza A virus]	463	463	99%	6e-163	56.36%	2HT5_A
<input type="checkbox"/>	The crystal structure of neuraminidase from A/						5HUN_A
<input type="checkbox"/>	Crystal Structure of Influenza A Virus Neurami						3SAL_A
<input type="checkbox"/>	The crystal structure of neuraminidase from A/						5HUM_A
<input type="checkbox"/>	Crystal structure of Neuraminidase N6 [Influen						4QN4_A
<input type="checkbox"/>	Structure of Neuraminidase from English duck sub						1V0Z_A
<input type="checkbox"/>	Crystal structure of influenza A neuraminidase N3-H274Y [Influenza A virus (A/swine/Missouri/2124514/2006(H2N3))]	356	356	97%	9e-121	47.77%	4HZY_A
<input type="checkbox"/>	The crystal structure of influenza A neuraminidase N3 [Influenza A virus (A/swine/Missouri/2124514/2006(H2N3))]	355	355	97%	1e-120	48.04%	4HZV_A
<input type="checkbox"/>	Crystal structure of influenza neuraminidase N3-H274Y complexed with oseltamivir [Influenza A virus (A/swine/Missouri/2124514/2006(H2N3))]	355	355	97%	1e-120	47.77%	4HZZ_A
<input type="checkbox"/>	A/NWS/whale/Maine/1/84 (H1N9) reassortant influenza virus neuraminidase [Influenza A virus]	355	355	97%	2e-120	48.81%	2B8H_A
<input type="checkbox"/>	N9 Neuraminidase Complexes With Antibodies Nc41 And Nc10: Empirical Free-Energy Calculations Capture Specificity Trends Observed With Mutant Binding Data [Influenz	355	355	97%	2e-120	48.81%	1NMA_N
<input type="checkbox"/>	Refined Crystal Structure Of The Influenza Virus N9 Neuraminidase-Nc41 Fab Complex [Influenza A virus (A/whale/Maine/1/84(H13N9))]	355	355	97%	3e-120	48.81%	1NCD_N
<input type="checkbox"/>	Refined Crystal Structure Of The Influenza Virus N9 Neuraminidase-Nc41 Fab Complex [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	354	354	97%	4e-120	48.55%	1NCA_N
<input type="checkbox"/>	Chain N_N9 NEURAMINIDASE [Influenza A virus]	357	357	97%	5e-120	48.81%	1NMB_N
<input type="checkbox"/>	Crystal Structures Of Two Mutant Neuraminidase-Antibody Complexes With Amino Acid Substitutions In The Interface [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	353	353	97%	5e-120	48.55%	1NCC_N
<input type="checkbox"/>	Refined Atomic Structures Of N9 Subtype Influenza Virus Neuraminidase And Escape Mutants [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	353	353	97%	7e-120	48.55%	4NN9_A
<input type="checkbox"/>	THREE-DIMENSIONAL STRUCTURE OF INFLUENZA A N9 NEURAMINIDASE AND ITS COMPLEX WITH THE INHIBITOR 2-DEOXY 2,3-DEHYDRO-N-ACETYL NEURAM	353	353	97%	7e-120	48.55%	1NNA_A
<input type="checkbox"/>	COMPLEX BETWEEN NC10 ANTI-INFLUENZA VIRUS NEURAMINIDASE SINGLE CHAIN ANTIBODY WITH A 5 RESIDUE LINKER AND INFLUENZA VIRUS NEURAMINII	353	353	97%	7e-120	48.55%	1A14_N
<input type="checkbox"/>	Refined Atomic Structures Of N9 Subtype Influenza Virus Neuraminidase And Escape Mutants [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	353	353	97%	1e-119	48.55%	5NN9_A
<input type="checkbox"/>	Crystal Structures Of Two Mutant Neuraminidase-Antibody Complexes With Amino Acid Substitutions In The Interface [Influenza A virus (A/tern/Australia/G70C/1975(H11N9))]	352	352	97%	2e-119	48.28%	1NCB_N
<input type="checkbox"/>	Crystal structure of Neuraminidase N7 [Influenza A virus (A/mallard/ALB/196/1996(H10N7))]	352	352	98%	2e-119	46.51%	4QN3_A
<input type="checkbox"/>	Chain A_neuraminidase [Influenza A virus]	352	352	97%	2e-119	48.28%	1L7H_A
<input type="checkbox"/>	Anhui N9 [Influenza A virus]	352	352	97%	3e-119	48.55%	4MWJ_A
<input type="checkbox"/>	Refined Atomic Structures Of N9 Subtype Influenza Virus Neuraminidase And Escape Mutants [unidentified influenza virus]	352	352	97%	3e-119	48.28%	3NN9_A
<input type="checkbox"/>	Chain A_Tetrabrachion_Neuraminidase [synthetic construct]	355	355	97%	3e-119	48.55%	6CRD_A

選択

クエリと同じものは除く

N8ノイラミニダーゼ2HT5のAチェーンを選択
 → アラインメントのセクションに移動
 そこで、2HT5と同じ配列をもつ構造を表示できる

BLASTによるテンプレート検索(4)

BLAST検索の結果の検討

Download ▾ GenPept Graphics

Chain A, Neuraminidase [Influenza A virus]
 Sequence ID: 2HT5_A Length: 390 Number of Residues: 1
 See 5 more titles

配列一致度が高く、ア
 ラインメントされた部分
 も長いことも要チェック

- Chain A, N8 Neuraminidase In Open Complex With Oseltamivir
Sequence ID: [pdb12HT7IA](#)
- Chain A, N8 Neuraminidase In Complex With Oseltamivir
Sequence ID: [pdb12HT8IA](#)
- Chain A, N8 Neuraminidase In Complex With Zanamivir
Sequence ID: [pdb12HTQIA](#)
- Chain A, N8 Neuraminidase In Complex With Dana
Sequence ID: [pdb12HTRIA](#)
- Chain A, N8 Neuraminidase In Complex With Peramivir
Sequence ID: [pdb12HTUIA](#)

Range 1: 3 to 387 GenPept Graphics ターゲットとテンプレートの配列一致度56%

Score	Expect	Method	Identities	Positives	Gaps
463 bits(1191)	4e-163	Compositional matrix adjust.	217/385 (56%)	279/385(72%)	2/385(0%)
Query 3	LTGNSSSLCPISGWAIYSKDNNGIRIGSKGDVVFIREPFISCSHLECRFFLTQGGALLNDKH				62
Sbjct 3	MNNTAICDAKGFAPFSKDNNGIRIGSRGHI FVIREPFVSCSPIECRTFFLTQGSLLNDKH				62
Query 63	SNGTVKDRSPYRILMSPVGEAPSPYNSRFESVAWSASACHDGMGWLTIIGISGPDNGAVA				122
Sbjct 63	SNGTVKDRSP+RTILMS VG++P+ Y +RFE+VAWSA+ACHDG W+I+G++GPD+ AVA				122
Query 123	VLKYNGLIITDIKSWRNILRTQESACVNGSCFTIMIDGPSNGQASYKILKIEKGKVT				182
Sbjct 123	V+ Y G+ TD + SW +ILRTQES C C+ G C+ +MTDGP+N QA Y+I K +G++				182
Query 183	KSIELNAPNYHYEECSYCPDTGKVMCVCRDNWHGNSRNPWVSFDQNLDYQIGYICSGVFGD				242
Sbjct 183	+++ H EECSCYP+ GKV CVCRD W G+NRP + +L Y++GY+C+G+ D				242
Query 243	NPRPNDG--TGSCGPVSSNGANGIKGFSFRYDNGVWIGRIKSTSSRSGFEMIWDPNGWTE				300
Sbjct 243	PR D TGSC N G+KGF FR VW+GRI S +SRSGFE++ NGWT+				302
Query 301	TDSSFSVRQDIVAITDWSGYSGSFVQHPPELIGLDCMRPCFWVELIRGQPKENTIIWISGSS				360
Sbjct 303	T +Q +V +WSGYSGSF EL+G DC+ PCFWVE+IRG+P+E TIWTS SS				362
Query 361	ISFCGVNSDITVGWSWPDGAELPFSI 385				
Sbjct 363	I CGV+ + WSW DGA LPF I 387				

単体構造は2HT5だが、
 同じリガンド(ザナミビル)との複合体構造である「2HTQ」を選択することができる
 そのほか、解像度の良い構造を選ぶなどの工夫が考えられる

このアラインメントを利用(モデリングソフトに入力するには、データ形式の変換が必要)

HHpredの利用(1)

H1N1ノイラミニダーゼp1の配列を入力として、フォールドライブラリを検索

<https://toolkit.tuebingen.mpg.de/tools/hhpred>



MPI
Bioinformatics
Toolkit

Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

HHblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

HHpred ?

ID Date Tool

Input Parameters

```
>p1
VILTGMSSLCPIISGWAIIYSKDNIGIRIGSKGDVFIREFPISCSHLECRFFLTQGALLNDKHSNGTVKDRSPYRTLMSCP
VGEAPSPYNSRFESVAWSASACHDGMGLTIGISGPDINGAVAVLKYNGIITDTIKSWRNILRTQESECACVNGSCFTIM
TDGPSNGQASYKILKIEKGKVTKSIELNAPNYHYEECCYPTDGKVMCVRDNHGSNRPWVSFDQNLQDYQIGYICSGVF
GDNPRPNDGTGSCGPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWDPNGWTETDSSFVSRQDIVAITDWSGY
SGSFVQHPPELTGLDCMRPCFWELIRGQPKENTIWTSGSSISFCGVNSDTVGSWPDGAELEPFSI
```

タンパク質p1の配列
([p1.fasta](#))を入力

Paste Example Upload File

Protein FASTA

Align two sequences/MSAs

配列(のアルインメント)を与えることも可能

Select structural/domain databases

Select proteomes

PDB_mmCIF70_29_Nov

検索対象のデータベースが選択可能

Custom Job ID

Submit

「Submit Job」をクリック

HHpredの利用(2)



Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

HHblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

ID	Date	Tool
5699189		HHPR

HHpred ?

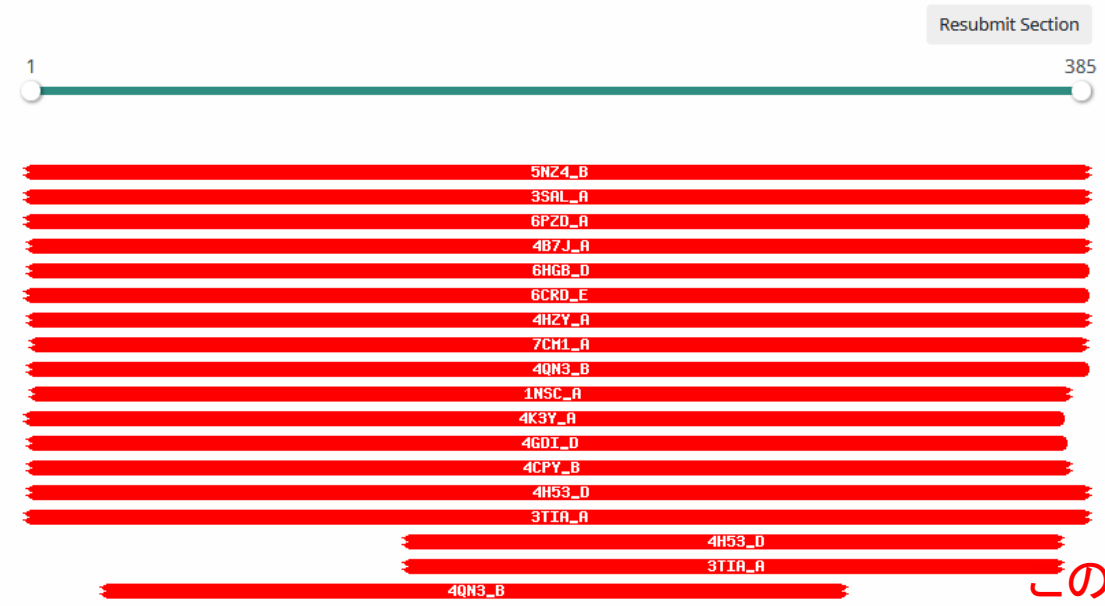
Job ID: 5699189, Created: 4 hours ago

Input Parameters Results Raw Output Probability Plot Query Template MSA Query MSA

Vis Hits Aln Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap Seqs

Number of Hits: **18**
Query MSA diversity (Neff): **4.62482**

Visualization



この下を見ると...

HHpredの利用(3)

Vis
Hits
Aln
Select All
Forward
Forward Query A3M
Model using selection
Download HHR
Color Seqs
Wrap Seqs

Hitlist

Show 25 Entries

Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/> 1	5NZ4_B	neuraminidase; neuraminidase, influenza, complex, inhibitor, hydrolase; HET: BMA, NAG, G39, MAN, EDO; 1.36A {unidentifie	100	5e-161	1201.5	43.5	385	388
<input type="checkbox"/> 2	3SAL_A	Neuraminidase (E.C.3.2.1.18); 6-BLADED BETA-PROPELLER, HYDROLASE, CALCIUM BINDING; HET: GOL, NAG, BMA; 1.5A {Influenza A	100	4.9e-158	1183.02	42.7	385	395
<input type="checkbox"/> 3	6PZD_A	Neuraminidase (E.C.3.2.1.18); antibody, inhibition mechanism, VIRAL PROTEIN; HET: BMA, MES, EDO, MAN, NAG; 1.12A {Influe	100	3.4e-157	1173.52	41.6	383	393
<input type="checkbox"/> 4	4B7J_A	NEURAMINIDASE; HYDROLASE, NEURAMINIDASE INHIBITOR, NAI, NAIS; HET: G39, NAG; 2.417A {INFLUENZA A VIRUS} SCOP: b.68.1.0;	100	1e-156	1193.81	41.7	384	469
<input type="checkbox"/> 5	6HGB_D	Neuraminidase (E.C.3.2.1.18); Influenza, A, Native, N6, Neuraminidase; HET: MAN, PO4, PEG, BMA, NAG, GOL; 1.5A {Influenz	100	3.9e-156	1167.44	41.9	382	389
<input type="checkbox"/> 6	6CRD_E	Tetrabrachion, Neuraminidase (E.C.3.2.1.18); NEURAMINIDASE, SIALIDASE, HYDROLASE(O-GLUCOSYL), HYDROLASE-HYDROLASE, ARTIFI	100	1.9e-155	1181.07	42.6	383	473
<input type="checkbox"/> 7	4HZY_A	Neuraminidase; neuraminidase, HYDROLASE; HET: FUC, NAG; 1.598A {Influenza A virus} SCOP: I.1.1.1, b.68.1.0; Related PDB	100	8.9e-155	1155.58	41.4	379	388
<input type="checkbox"/> 8	7CM1_A	Neuraminidase (E.C.3.2.1.18); Lattice-	100	2.5e-152	1137.66	40.9	375	387

この下を見ると...



HHpredの利用(4)

5nz4_Bとクエリ配列
のアラインメント

Vis Hits Aln | Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap Seqs

Alignments

Template alignment | Template 3D Structure | PDBE

5NZ4_B neuraminidase; neuraminidase, influenza, complex, inhibitor, hydrolase; HET: BMA, NAG, G39, MAN, EDO; 1.36A (unidentified influenza virus) SCOP: b.68.1.0; Related PDB entries: 6D96_A 6D96_D 6D96_G 6D96_H 6D96_B 6D96_C 6D96_F 6D96_E 5HUG_A 3CYE_A 3CYE_B 2HTY_A 2HTY_D 2HTY_G 2HTY_H 2HTY_B 2HTY_C 2HTY_F 2HTY_E 2HU0_A 2HU0_G 2HU0_H 2HU0_B 2HU0_C 2HU0_F 2HU0_E 2HU0_A 2HU0_D 2HU0_G 2HU0_H 2HU0_B 2HU0_C 2HU0_F 2HU0_E 3B7E_A 3B7E_B 3BEQ_A 3BEQ_B 3CKZ_A 3CL2_A 3CL2_D 3CL2_G 3CL2_H 3CL2_B 3CL2_C 3CL2_F 3CL2_E 3CL0_A 6Q23_D 6Q23_A 6Q23_B 6Q23_C 6HP0_A 6HP0_D 6HP0_B 6HP0_C 3NSS_A 3NSS_B 5NZ4_A 5NZE_A 5NZE_B 3TI3_A 3TI3_B 3TI4_A 3TI4_B 3TI5_A 3TI5_B 3TI6_A 3TI6_B 4B7R_A 4B7R_D 4B7R_B 4B7R_C 6G01_A 6G01_B 6G02_A 6G02_B 4QNP_A 4QNP_B 2HTW_A 2HTV_A 2HTV_B 5NWE_A 5NWE_D 5NWE_B 5NWE_C 5NZF_A 5NZF_D 5NZF_B 5NZF_C 5NZN_A 5NZN_D 5NZN_B 5NZN_C

Probability: 100%, E-value: 5e-161, Score: 1201.5, Aligned cols: 385, Identities: 89%, Similarity: 1.585, Template Neff: 4.6

```

Q ss_pred          CccCCCCCCCCcEEecCceeeCCCCEEECcEeeCCceeeeeeccccCCCCcCCcCCcCCcCCcEEec
Q 1                VILTGNSLCPISGWAISYKDNIGIRIGSKGDVVFVIREPFISCSHLECRFTFLTQGALLNDKHSNGTVKDRSPYRTLHMSCP  80 (385)
Q Consensus        1      ~~~~~C~i~g~w~~~~~kdn~r~g~~~~i~v~ReP~vsC~~~~cr~FaL~qg~l~n~khsngT~DRs~r~L~s~  80 (385)
                   +|+++++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|
T Consensus        2      ~~~~~C~i~g~w~~~~~kdn~r~g~~~~i~v~ReP~vsC~~~~cr~FaL~qg~l~n~khsngT~DRs~r~L~s~  81 (388)
T 5NZ4_B           2      VKLAGNSLCPVSGWAISYKDNISVIRIGSKGDVVFVIREPFISCSPLECRFTFLTQGALLNDKHSNGTKDRSPYRTLHMSCP  81 (388)
T ss_dssp          CCCCCCSCBCCCSEEEEEECCHHHHTTSCCEEEEEEEESCSEEEEEEEETTSGGGTTTCCCTTCCEEEEE
T ss_pred          cccccCCCCcEEecCceeeCCCCcEEcEeeCCeEeCCchheeeccccCCCCcCCcCCcCCcCCcEEec
  
```

```

Q ss_pred          CCCCCcCCCCcEEEEEccccCCCCcEEEEECCCCCEEEEECEEEcceeHHhccccCCcEeeCCEEEEEE
Q 1                VGEAPSPYNSRFESVAWSASACHDGMGLTIGISGPDNGAVAVLKYNIGIITDIKSWRNILRTQSEACVNGSCFTIM  160 (385)
Q Consensus        81      l~g~p~~~~n~~~~c~v~a~w~S~s~a~C~h~D~G~k~w~i~i~G~d~A~a~i~Y~g~t~d~i~S~w~n~I~L~R~T~Q~E~S~e~c~v~C~n~G~C~w~m  160 (385)
                   ||++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|
T Consensus        82      l~g~p~~~~n~~~~c~a~w~S~s~a~C~h~D~G~k~w~i~i~G~d~A~a~i~Y~g~t~d~i~S~w~n~I~L~R~T~Q~E~S~e~c~v~C~n~G~C~w~  161 (388)
T 5NZ4_B           82      IGEVPSPYNSRFESVAWSASACHDGINMLTIGISGPDNGAVAVLKYNIGIITDIKSWRNVLRTQSEACVNGSCFTVM  161 (388)
T ss_dssp          TTSCCTTTCEEEEECSEEEEECSSSEEEEEESCSTSCSEEEEEETEEEEEECSSSSCBCCSSCEETEEEEEE
T ss_pred          CCCCCcCCCCcEEEEEccccCCCCcEEEEECCCCCEEEEECEEEcceeHHhcCCCCcCCcEeeCCEEEEEE
  
```

```

Q ss_pred          ecCCCCcCceEEEEEcCeeeeeEeCCCCccccccccCCCCCEEEeCCCCcCCCCEEecCCCCcccccccc
Q 1                TDGPSNGQASYKILKIEKGVKTSIELNAPNYHYEECSYPTDGKVMCVRDNWHGSNRPWVFDQNLDYQIGYICSGVF  240 (385)
Q Consensus        161      TDGpa~~~~a~ri~i~e~G~K~I~i~~~~~h~E~E~C~S~C~y~~~~i~C~v~C~R~D~N~W~g~s~n~R~P~l~~~~~Y~i~C~S~g~  240 (385)
                   |||++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|
T Consensus        162      tDgpa~~~~a~ri~i~e~G~I~i~~~~~h~E~E~C~S~C~y~~~~i~C~v~C~R~D~N~W~g~s~n~R~P~l~~~~~Y~i~C~S~g~  241 (388)
T 5NZ4_B           162      TDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSYPTDSEITCVRDNWHGSNRPWVFNQNLQYIGYICSGVF  241 (388)
T ss_dssp          EECSSSCCEEEEEETEEEEEECCCTCCCEEEEEETEEEEEECCSSCSSEEEEECTTCEEEEECCSSC
T ss_pred          ecCCCCcCceEEEEEcCeEeeEeCCCCccccccccCCCCCEEEeCCCCCCCCcEEecCCCCcccccccc
  
```

```

Q ss_pred          CCCCCCCCCCCCCCCCCcEEeEeCCcEeEeEeCCCCcEEEECCCCcCCCCcCCcCCcEEeEeCCcccc
Q 1                GDNPRPNNDGTGSCGVPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIDPNGWGTEDSSFSVRQDIVAIDWISGY  320 (385)
Q Consensus        241      ~DtPrP~D~t~c~~~~~g~G~V~K~G~f~n~f~g~w~G~R~T~i~s~S~R~S~G~F~E~m~~~~~G~w~~~~~q~i~v~~~~~W~S~G~Y  320 (385)
                   +|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|++|
T Consensus        242      ~DtPrP~D~t~c~~~~~g~G~V~K~G~f~n~f~g~w~G~R~T~i~s~S~R~S~G~F~E~m~~~~~G~w~~~~~q~i~v~~~~~W~S~G~Y  321 (388)
T 5NZ4_B           242      GDNPRPNNDKGTGSCGVPVSSNGANGIKGFSFKYGVNGVWIGRTKTSISSRNGFEMIDPNGWGTDNFISKQDIVGINEHSYG  321 (388)
  
```

HHpredの利用(5)

5nz4_Bをテンプレートとしてモデリング
アラインメントはHHpredの結果を利用

HHpred 5699189, Created: 4 hours ago

Number of Hits: 18
Query MSA diversity (Neff): 4.62482

Visualization

Resubmit Section

1 385

5NZ4_B
3SAL_A
6PZD_A
4B7J_A
6HGB_D

Hitlist

Show 25 Entries Search:

Nr	Hit	Name	Probability	E-value	SS	Cols	Target Length
1	<input checked="" type="checkbox"/>	5NZ4_B neuraminidase; neuraminidase, influenza, complex, inhibitor, hydrolase; HET: BMA, G39, EDO, MAN, NAG; 1.36A	100	3.1e-160	44.3	385	388

「Model using selection」を指定

5nz4_Bをテンプレートとして選択
(チェックを入れる)

HHpredの利用(6)

MPI Bioinformatics Toolkit

Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

HHblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

HHpred-TemplateSelection ? Job ID: 3957180, Parent Job ID: 5699189, Created: 4 minutes ago

ID	Date	Tool
3957180		HTMP
5699189		HHPR

Results Summary

```
>P1;UKNP
sequence:UKNP:1 :A:385 :A:::
VILTGNSSLCPISGWA IYSKDN GIRIGSKGDV FVIREPFISCSHLECR TFFLTQ GALLNDKHSNGT VKDRSPYR TLMSCP VGEAPSPYNSRFESVAWSASACHDGMGWL TIGISGPDNGAVAVLKYNGII
>P1;5NZ4
structure:5NZ4:83 :B:467 :B::unidentified influenza virus:1.36:
VKLAGNSSLCPVSGWA IYSKDN SVRIGSKGDV FVIREPFISCSPLECR TFFLTQ GALLNDKHSNGT IKDRSPYR TLMSCP IGEVPSYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGII
```

ターゲットとテンプレートの配列が表示される

Forward to MODELLER

「Forward to Modeller」をクリック

If you use HHpred-TemplateSelection on our Toolkit for your research, please cite as appropriate:

HHpredの利用(7)

The screenshot shows the HHpred web interface. At the top, there are navigation tabs: Search, Alignment, Sequence Analysis, Zary Structure, 3ary Structure (selected), Classification, and Utils. Below these are sub-tabs for MODELLER and SamCC. On the left, there is a logo for MPI Bioinformatics Toolkit and a table listing tools:

ID	Date	Tool	
3957180		HTMP	✕
5699189		HHPR	✕

The main content area is titled "MODELLER" and contains an "Input" section. A red annotation reads: "Modellerの入力形式のアラインメントが生成され、表示される" (Modeller's input format alignment is generated and displayed). The input text is as follows:

```
>P1;UKNP
sequence:UKNP:1 :A:385 :A::::
VILTGNSSLCPISGWAIIYSKDNIGIRIGSKGDFVIREPFISCSHLECRFTFFLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWSASACHDGMGLTIGISGPDNGAVAV
LKYNGIITDTIKSWRNILRTQESECACVNGSCFTIMTDGSPNGQASYKILKIEKGKVTKSIELNAPNYHYEECSYPDTGKVMCVRDNWHGNSRPWVSFDQNLQDYIGYICSGVFGDNPRP
NDGTGSCGVPVSSNGANGIKGFSFRYDNGVWIGRTKSTSSRSGFEMIWDPNGWTETDSSFSVRQDIVAITDWSGYSGSFVQHPELTGLDCMRPCFWVELIRGQPKENTIWTSGSSISFCGVNSD
TVGWSWPDGAELPFSI*
>P1;5NZ4
structure:5NZ4:83 :B:467 :B::unidentified influenza virus:1.36:
VKLAGNSSLCPVSGWAIIYSKDNVIRIGSKGDFVIREPFISCSPLECRFTFFLTQGALLNDKHSNGTIKDRSPYRTLMSCPVGEVPSYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAV
LKYNGIITDTIKSWRNWVLRQESECACVNGSCFTVMTDGPNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSYPDSSEITCVRDNWHGNSRPWVSFNQNLLEYIGYICSGIFGDNPRP
NDKTGSCGVPVSSNGANGVKGFSFKYGNVWIGRTKSISSRNGFEMIWDPNGWTGTDNNSFIKQDIVGNEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTIWTSGSSISFCGVNSD
TVGWSWPDGAELPFTI*
```

Below the input text, there are links for "Paste Example" and "Upload File", and a "Protein PIR" button. A red annotation reads: "最初は、Modellerのキーが要求される(今回は入力済み)" (Initially, the Modeller key is required (this time it is already entered)). Below this, it says "MODELLER-key is stored in your profile." Another red annotation reads: "「Submit job」をクリック" (Click "Submit job"). The "Submit" button is circled in red.

At the bottom, there are links for Help, FAQ, Privacy Policy, Imprint, Contact Us, Cite Us, and Recent Updates. The footer text is: © 2008-2020, Dept. of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen

HHpredの利用(8)

MPI Bioinformatics Toolkit

Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

MODELLER SamCC

Job ID: 8717910, Parent Job ID: 3957180, Created: a minute ago

MODELLER ?

ID	Date	Tool	
8717910		MODL	×
3957180		HTMP	×
5699189		HHPR	×

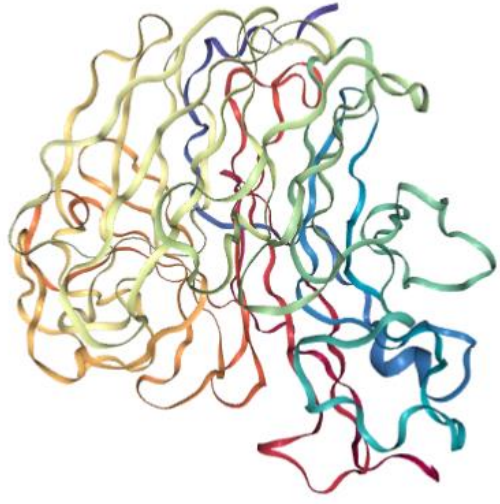
Download PDB File

PDB形式(座標データ)のファイルのダウンロード

生成されたモデル構造のグラフィックス表示

モデル構造と正解構造3b7e Aチェーンとの比較
RMSD between 385 pruned atom pairs is 0.272 angstroms

If you use MODELLER on our Toolkit for your research, please cite as appropriate:



Phyreの利用(1)

H1N1ノイラミニダーゼp1の配列を入力として、フォールドライブラリを検索

Standard Mode | [Switch to Expert Mode](#) | [View past jobs](#) | [Phyre Alarm](#) | Welcome shimizu@bi.a.u-tokyo.ac.jp | [My account](#) | [Not you?](#) Retrieve Phyre Job Id

<http://www.sbg.bio.ic.ac.uk/phyre2/>

Phyre²

Protein Homology/analogY Recognition Engine V 2.0

Subscribe to Phyre at Google Groups

Email:

[Visit Phyre at Google Groups](#)

[Follow @Phyre2server](#)



Please do not use 'intensive mode' unless you are an experienced user and understand its pitfalls (and your search has already failed with 'normal mode'). For most users, most of the time, 'normal mode' will give you the answer you require

If you have more than 5 or 6 sequences to model, it is easier for you (and better for everyone!) if you use "batch" mode, which is available under the Expert Mode after you log in (top left of the interface). If you haven't registered for a Login [i](#), you can do so on the Login page.

Current Phyre2 server load = 18% (normal running) [i](#)

E-mail Address	<input type="text" value="shimizu@bi.a.u-tokyo.ac.jp"/>
Optional Job description	<input type="text" value="p1"/>
Amino Acid Sequence i	<pre>VIL TGNSSLCPISGWA IYSKDNIGIRIGSKGDV FVIREPFISCSHLECR TFFLTQ GALL NDKHSNGTVKDRSPYR TLMSCP VGEAPSPYNSRFESVAWSASACHDGMGLTIGISGPDNGAVAVLKYNGIITDTIKSWR NNILRTQSEACACVNGSCTIM TDGPSNGQASYKILKIEKGKVTKSIELNAPNYHYEECCYPTDGKVMCVCRDNW HGSN RPWVSFDQNLDYQIGYICSGVF GDNPRPNDGTGSCGPVSSNGANGIKGFSFRYDNGWIGRTKSTSSRSGFEMIDPNGW TETDSSFSVRQDIVAITDWSGY SGSFVQHPELTGLDCMRPCFWELIRGQPKENIWTSGSSISFCGVNSDTVGSWSPDG AELPFSI</pre>
	Or try the sequence finder
Modelling Mode i	Normal <input checked="" type="radio"/> Intensive <input type="radio"/> Test <input type="checkbox"/>
Please tick as appropriate. i	[NOT for Profit <input type="radio"/> FOR Profit (Commercial) <input type="radio"/> Other <input checked="" type="radio"/>
	<input type="button" value="Phyre Search"/> <input type="button" value="Reset"/>

タンパク質p1の配列
(p1.fasta)を入力

「Phyre Search」をクリック

Phyre2の利用(2)

Phyre²

Job Status

Email	shimizu@bi.a.u-tokyo.ac.jp
Job Description	p1
Unique Job ID	987ea2f7c07b5b06
Date	Fri Apr 24 13:57:57 BST 2020

Estimated total processing time: 2.5 hours ± 2.2 hours

2. Building hidden Markov model of sequence

A link to results will be mailed to you when the job is finished
Or bookmark this page to return to it at any time

Tip: Submit a set of sequences (up to 1000)

This page auto-refreshes every 30 seconds



1. Finding homologues with PSI-Blast
2. Building hidden Markov model of sequence
3. Checking for transmembrane helices
4. Constructing models
5. Loop modelling
6. (Multiple template modelling with Poing, intensive mode only)
7. Generating results page

Phyre²

Retrieve Phyre Job ID: [Fetch]

Email	jimu@bi.a.u-tokyo.ac.jp
Description	p1
Date	Tue Apr 26 05:04:34 BST 2022
Unique Job ID	19656ef9bab6e2a0
Sequence	VILTGNSLSC ... Download FASTA
Job Type	normal
Job Expiry	30 days

[Download zip of all results](#)

Summary

Top model

Model (left) based on template [c2hu4D](#)

Top template information

PDB header: hydrolase
Chain: D: **PDB Molecule:** neuraminidase;
PDB title: n1 neuraminidase in complex with oseltamivir 2
PDB Entry: [PDBa](#) [RCSB](#) [PDBj](#)

Confidence and coverage	
Confidence:	100.0%
Coverage:	100%

385 residues (100% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

[3D viewing](#)
[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

Image coloured by rainbow N → C terminus
 Model dimensions (Å): X:51.941 Y:53.683 Z:55.458

Sequence analysis

[View PSI-Blast Pseudo-Multiple Sequence Alignment](#) [Download FASTA version](#)

Secondary structure and disorder prediction [\[Show\]](#)

Domain analysis [\[Show\]](#)

Detailed template information [\[Hide\]](#)

Phyreの利用(3)

配列p1と構造が似ているものと予想される順に表示

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	1zupA			100.0	92	PDB header: hydrolase Chain: D: PDB Molecule: neuraminidase; PDBTitle: n1 neuraminidase in complex with oseltamivir 2 PDB Entry: PDBe RCSB PDBj
2	c4b7qD			100.0	89	PDB header: hydrolase Chain: D: PDB Molecule: neuraminidase; PDBTitle: h1n1 2009 pandemic influenza virus: resistance of the i223r2 neuraminidase mutant explained by kinetic and structural analysis PDB Entry: PDBe RCSB PDBj
3	c3salB			100.0	59	PDB header: hydrolase Chain: B: PDB Molecule: neuraminidase; PDBTitle: crystal structure of influenza a virus neuraminidase n5 PDB Entry: PDBe RCSB PDBj
4	c2htuA			100.0	57	PDB header: hydrolase Chain: A: PDB Molecule: neuraminidase; PDBTitle: n8 neuraminidase in complex with peramivir PDB Entry: PDBe RCSB PDBj
5	1im8lv			100.0	49	PDB header: complex (hydrolase/immunoglobulin) Chain: N: PDB Molecule: n9 neuraminidase; PDBTitle: the structure of a complex between the nc19 antibody and influenza2 virus neuraminidase and comparison with the overlapping binding site3 of the nc41 antibody PDB Entry: PDBe RCSB PDBj
6	d1f8ea			100.0	48	Fold: 6-bladed beta-propeller Superfamily: Sialidases Family: Sialidases (neuraminidases) PDB entry: PDBe RCSB PDBj

「c2htuA」を選択 (N8ノイラミニダーゼのリガンド(ペラミビル)との複合体構造) アラインメントを求める

モデル構造のクオリティなどを解析

Phyreの利用(4)

Phyre²

Job Description	p1	Date	Tue Apr 26 05:04:34 BST 2022
Confidence	100.00%	Aligned Residues	383
Rank	4	Template	c2htuA_
% Identity	57%	PDB info	PDB header: hydrolase Chain: A: PDB Molecule: neuraminidase; PDBTitle: n8 neuraminidase in complex with peramivir PDB Entry: PDBe RCSB PDBj
Resolution	2.20 Å	Model Dimensions (Å)	X:46.529 Y:51.294 Z:53.898

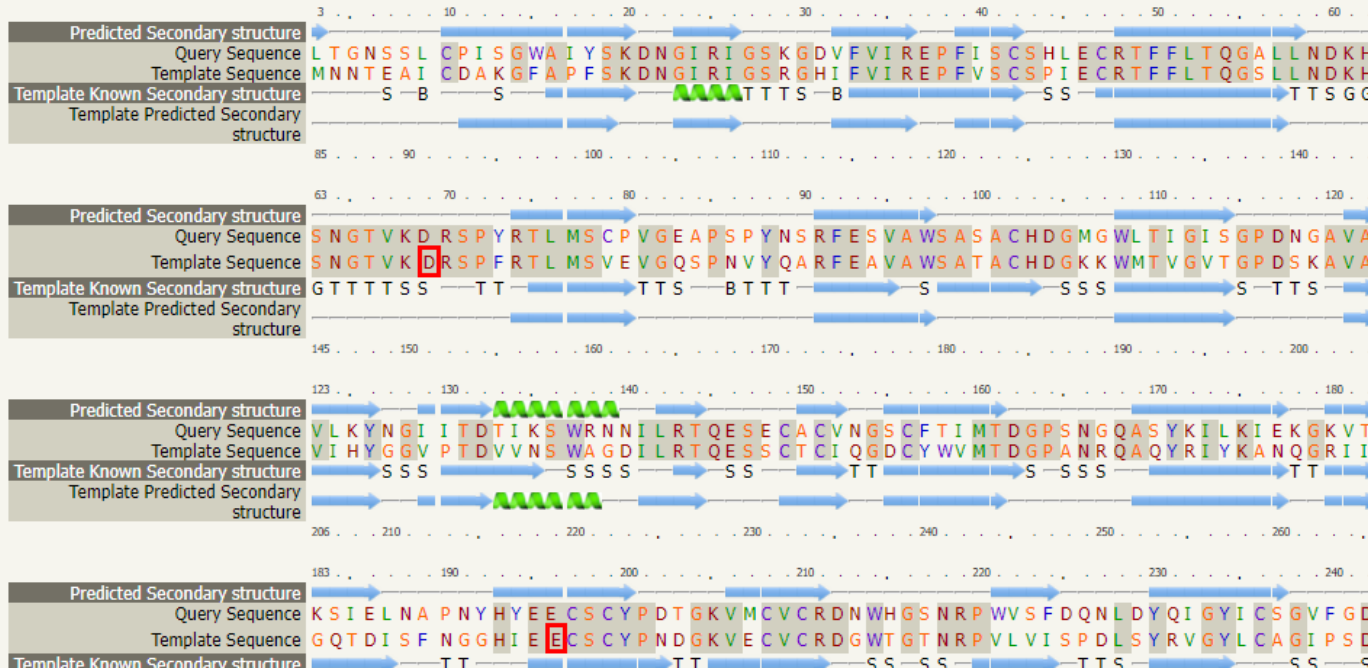
Show / Hide SS confidence

Show / Hide Conservation and Alignment quality

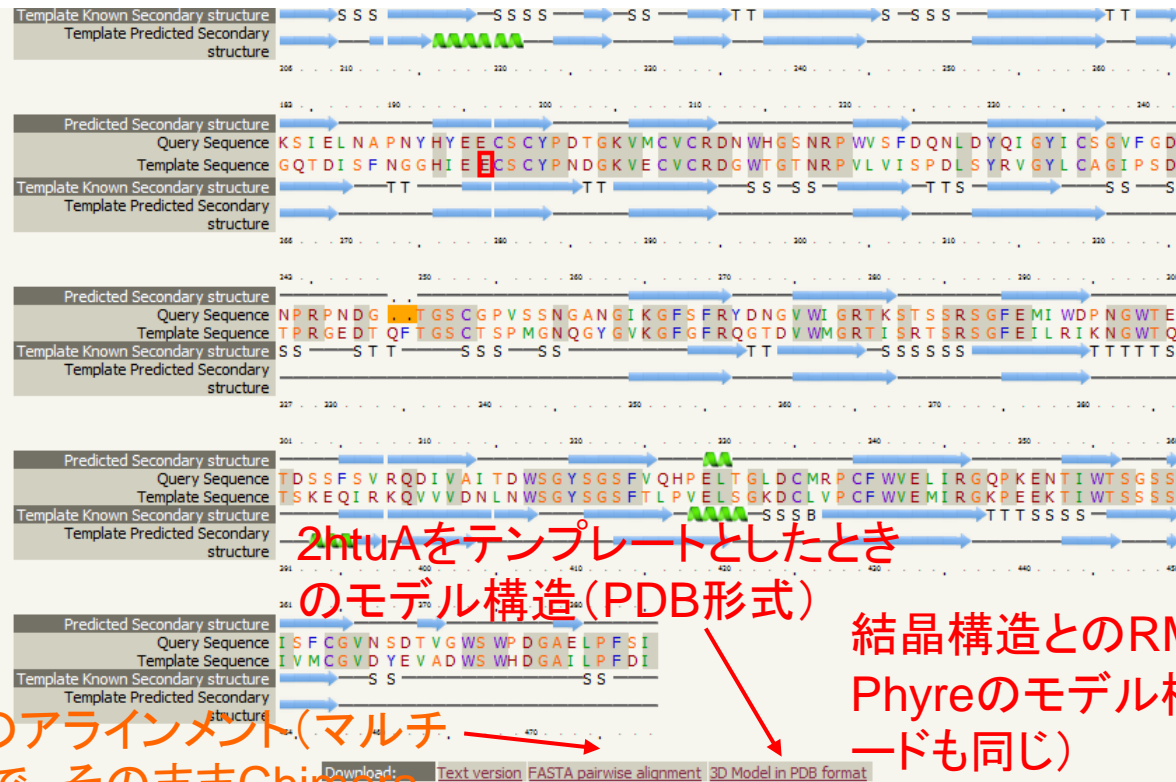
- Insertion relative to template
- Deletion relative to template
- Catalytic residue from the [CSA](#)

「c2htuA」とクエリ配列とのアラインメント

この例では、BLASTと同じだが、別なアラインメントが得られる可能性もある



Phyreの利用(5)



2htuAをテンプレートとしたとき
のモデル構造(PDB形式)

ターゲットとのアラインメント(マルチFASTA形式で、そのままChimeraのModellerで利用できる)



[View in JSmol](#)

[Send structure to FirstGlance for more viewing options](#)

結晶構造とのRMSD値
Phyreのモデル構造(intensiveモードも同じ)

RMSD between 360 pruned atom pairs is 0.669 angstroms;
(across all 383 pairs: 1.123)

Modellerのモデル構造
(アラインメントはPhyre)

RMSD between 367 pruned atom pairs is 0.671 angstroms;
(across all 385 pairs: 0.923)

モデル構造の比較(1)

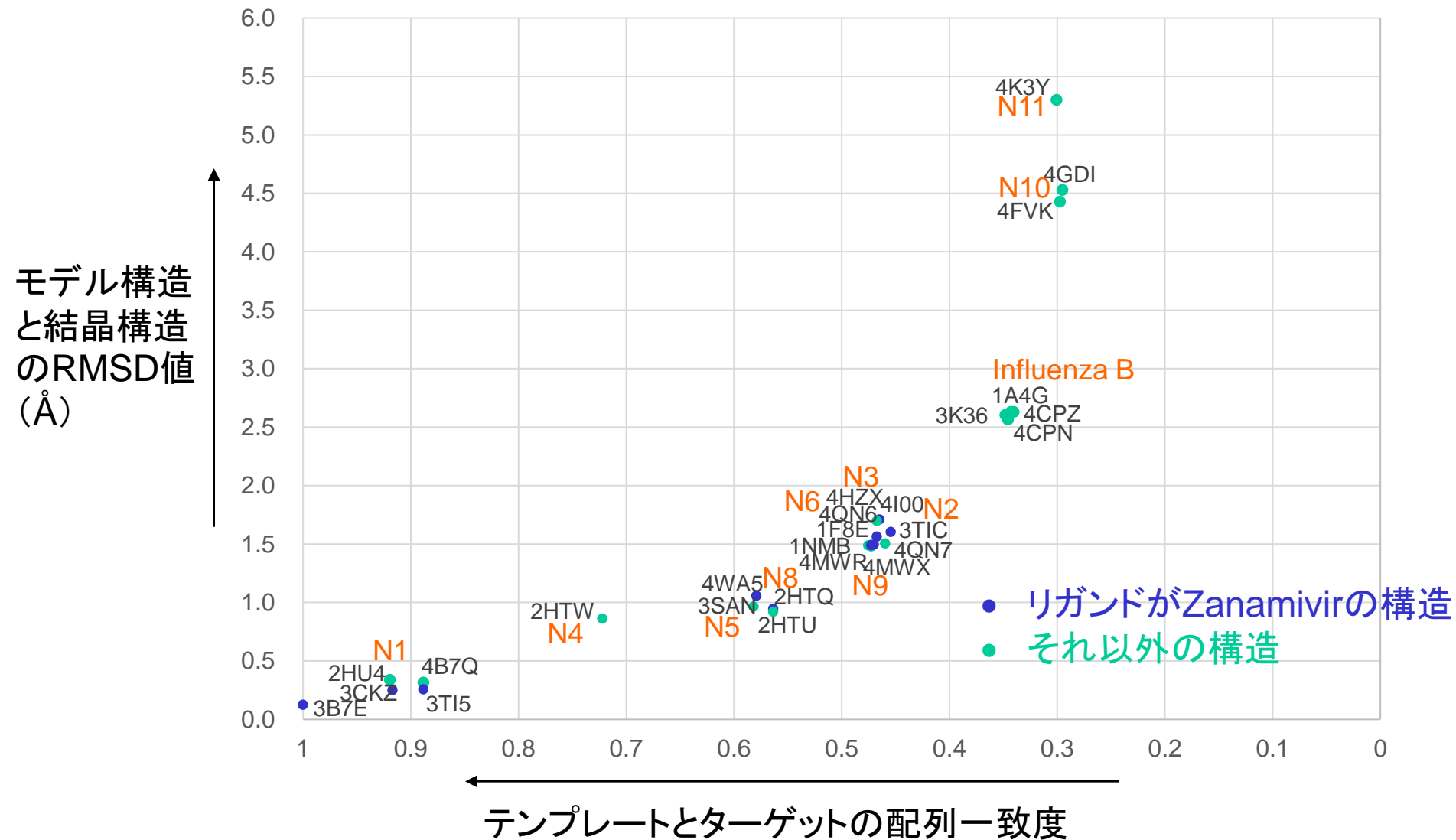
アラインメントの違いによるモデル構造の比較

テンプレート	BLAST	PSI-BLAST	COBALT	Phyre2	HHpred	SWISS-MODEL
2HU4_A 92.0%	0.339 0.339 (385)		0.339 0.339 (385)	0.315 0.315 (385)*	ターゲットと 同じフォール ドに属し、選 択できない	0.647 0.434 (382)
2HTU_A 56.4%	0.923 0.671 (367)		0.919 0.681 (368)	0.935 (383) 0.669 (365)		1.045 0.731 (363)
3SAL_A 58.2%	1.241 0.710 (358)	1.103 0.653 (362)	1.158 0.685 (362)	1.074 (383) [†] 0.644 (366) [†]	1.057 0.647 (363)	配列一致度の 低いものは候 補のテンプレ ートとして表 れない
1F8E_A 48.8%	3.791 0.784 (353)	1.318 0.806 (359)	1.489 0.787 (352)	1.350 (379) 0.769 (352)	1.389 [‡] 0.769 (351) [‡]	
4K3Y_A 34.2%	7.512 0.925 (253)	6.305 0.920 (264)	5.301 0.943 (263)	2.744 (350) 0.941 (274)	6.769 0.998 (248)	

* 2HU4_D, † 3SAL_B, ‡ 1F8D_A, || 2HT5_A
SWISS-MODELのアラインメントは、BLASTと
HHblitsを組み合わせによる

モデル構造の比較(2)

• テンプレートとモデル構造の関係



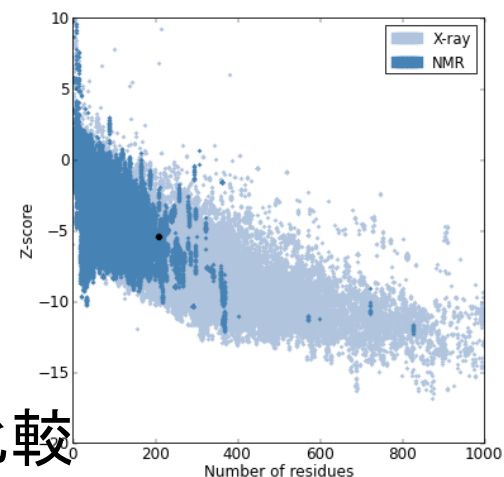
タンパク質構造評価プログラム

- タンパク質構造の評価
 - 構築した構造の「タンパク質らしさ」の評価
 - タンパク質らしくない部位の同定
 - それをもとに修正を加えることを可能にする
- 用途 → 構造予測、構造決定、コンピュータによる変異実験など
- PROCHECK (Thornton, et al.)
 - 主鎖・側鎖の二面角、主鎖の原子間の距離、角度
- ProSA (Sippl, et al.)
 - 原子間・残基間コンタクト、溶媒露出度
- ProQ (Wallner, et al.)
 - 原子間・残基間コンタクト、溶媒露出度
- WHATCHECK (Griend, et al.)
 - 対称性、幾何学的性質、構造等総合的評価
- ~~Verify3D (Eisenberg, et al.)~~
 - ~~埋没面積、極性度、二次構造~~
- 実験で決定された構造のスコアの分布との比較

ProSA-webの結果の例

Overall model quality [HELP](#)

Z-Score: **-5.48**



タンパク質の二次構造予測

- タンパク質のアミノ酸配列から、各アミノ酸が形成する二次構造を予測
 - 3分類予測と8分類予測 (DSSPによる分類)

8分類	3分類
α -helix (H)	helix (H)
3_{10} -helix (G)	
β -bridge (B)	extended strand (E)
extended strand (E)	
hydrogen bonded turn (T)	coil (C)
bend (S)	
other (' ' or C)	
π -helix (I)	

タンパク質の二次構造予測

- タンパク質の重要な構造特徴を予測することで、タンパク質の立体構造・機能に関する知見を得る
- 類似配列のマルチプルアラインメントをもとに、特に周辺の配列特徴を用いて予測



- AlphaFold2の利用
 - 多数のタンパク質の配列情報を用いることで、立体構造モデリングの精度が向上し、二次構造予測の結果も向上
 - 構築したモデル構造に対してDSSPで二次構造を判定
 - 従来の二次構造予測手法では、予測を誤ることがあった「通常と外れた」配列特徴をもつカメレオン配列、discordant helixなどの予測精度が向上

二次構造予測のサイト

- Jpred4
 - <http://www.compbio.dundee.ac.uk/jpred/>
- PSIPRED
 - <http://bioinf.cs.ucl.ac.uk/psipred/>
- PredictProtein
 - <https://www.predictprotein.org/>
- NetsurfP-2.0
 - <https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0>

二次構造予測手法の発展

- Chou-Fasman法

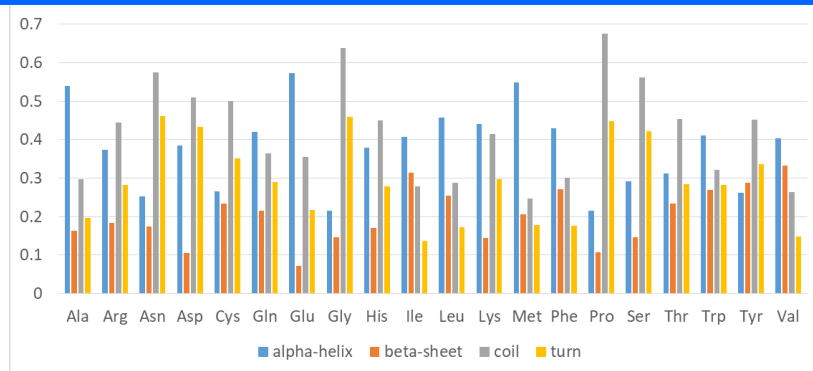
- 各アミノ酸の二次構造の傾向性をもとに、配列に従って二次構造を構築

- GOR (Garnier, Osguthorpe, Robson) 法

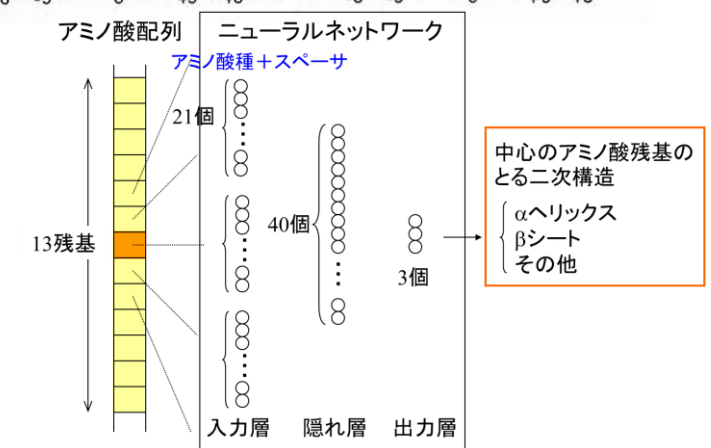
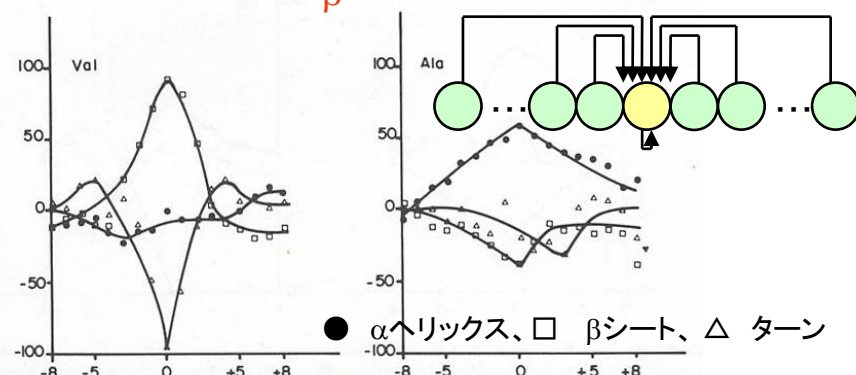
- 各残基が形成する二次構造に周辺残基が及ぼす情報量を計算

- ニューラルネットワーク

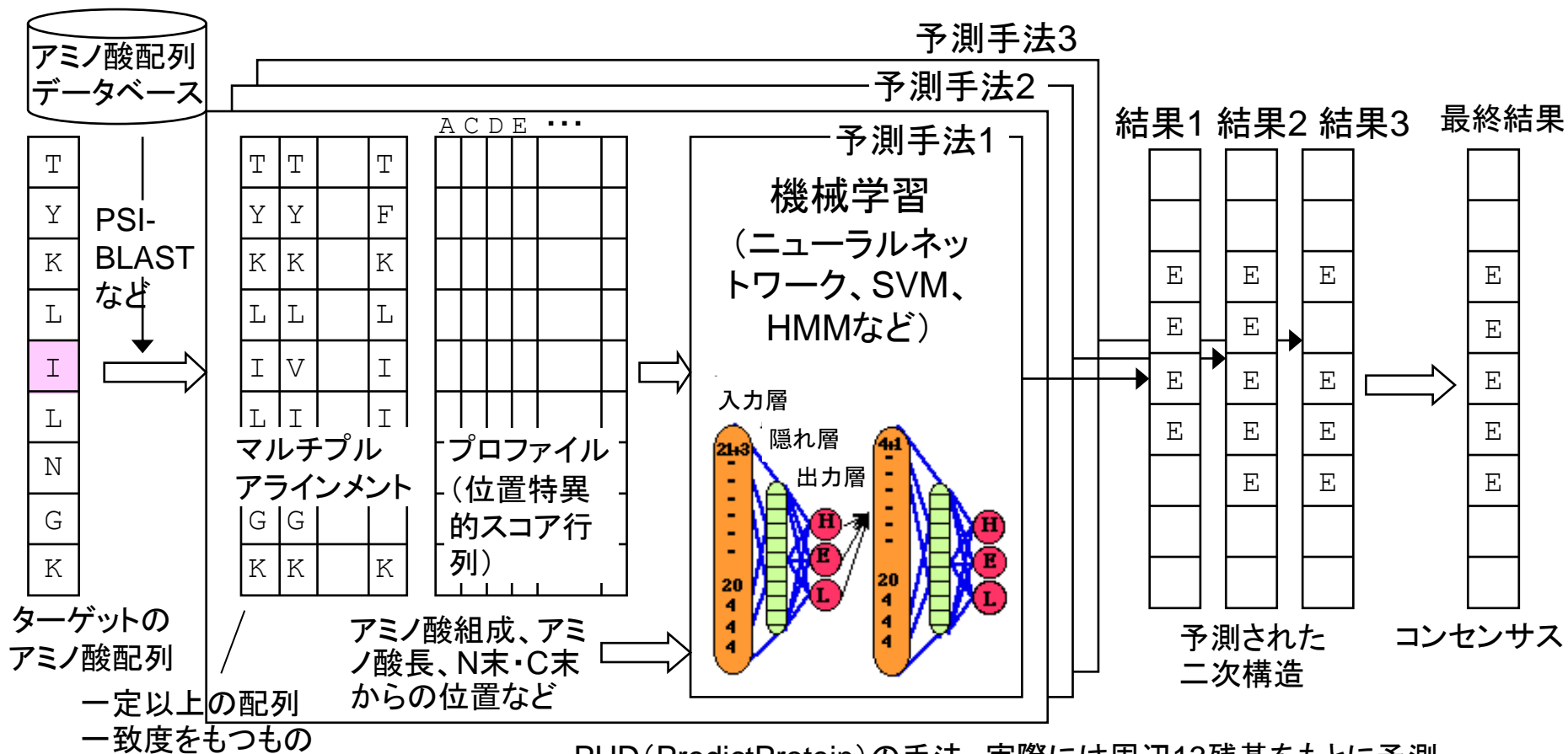
- 各残基が形成する二次構造を周辺残基の傾向をもとに学習させる



PISGWAIY**YSKDNGIR**IGSKG**DVFVIRE**PF**I**



タンパク質の二次構造予測手法の例

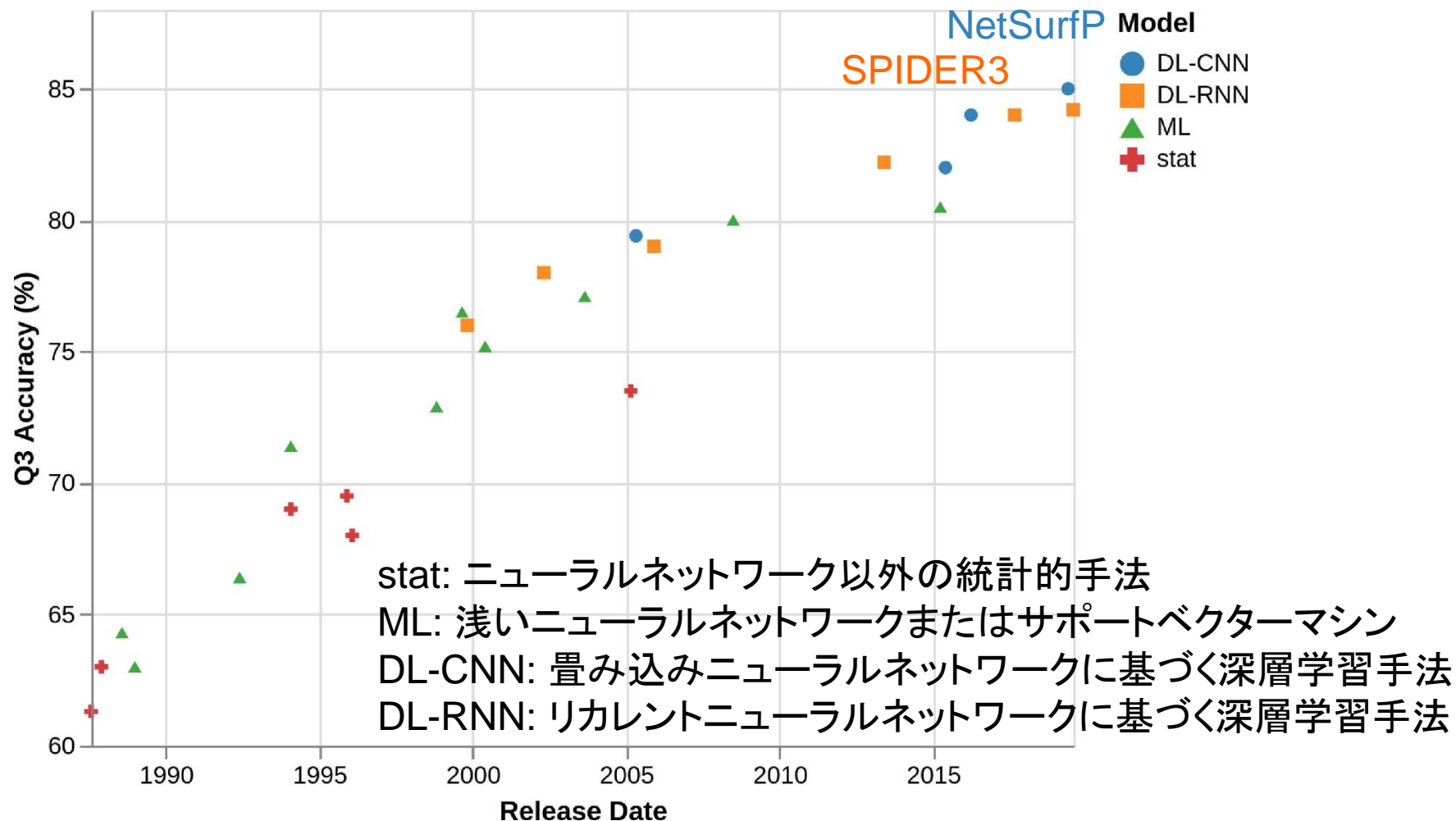


PHD (PredictProtein) の手法、実際には周辺13残基をもとに予測

類似の手法は、配列からの相互作用部位予測、ドメイン予測、天然変性領域予測、溶媒露出度予測などにも利用されている

二次構造予測の精度

主な二次構造予測手法の精度 (Q₃値)



[M. Torrìsi, et al. Deep learning methods in protein structure prediction, Computational and Structural Biotechnology Journal, 20, 2020](#)

Jpred4の利用

Jpred 4
Incorporating Jnet
<http://www.compbio.dundee.ac.uk/jpred/>
A Protein Secondary Structure Prediction Server

Home About News F.A.Q. Help & Tutorials Contact Publications

Input sequence(?)
RV
AGWWEAVSETGVDAFVRIIGFDNVRQVQLISFIAYKPPGC]

ターゲットの配列を入力
フラボドキシンの配列
「p0.fasta」を使用

Make Prediction

Jpred 4
Incorporating Jnet
A Protein Secondary Structure Prediction Server

Home About News F.A.Q. Help & Tutorials Contact Publications

Job Status
Your job (2FCR) started at 02:51 on 07/01/2015, the current time is 02:52 on 07/01/2015.

10% complete...

See below for a more detailed log of progress (use your back button to return):
Log file

If the job is taking too long it may be worth deleting and resubmitting it using an e-mail address: delete job

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton G.J. Paper in preparation
Previous: Cole C, Barber JD & Barton G.J. J Mol Biol. 2008; 35 (suppl 2): W197-W201 [PMID]

Jpred 4
Incorporating Jnet
A Protein Secondary Structure Prediction Server

Home REST API About News F.A.Q. Help & Tutorials Monitoring Contact Publications

Match found in PDB

The sequence you submitted is similar to those with known structure. These may provide a more accurate secondary structure assignment than a JPred prediction.

If you still want to carry out a Jpred prediction click **continue**

Hits found

Show 25 entries

PDB	Chain	Description	Blast E-value
2fcr	A	FLAVODOXIN	2e-28
3esy	D	Flavodoxin	2e-28
3esy	C	Flavodoxin	2e-28
3esy	B	Flavodoxin	2e-28

PDBに登録済みの既知構造が見つかった。それでも予測を実行するなら「continue」ボタンをクリック

Jpred 4
Incorporating Jnet
A Protein Secondary Structure Prediction Server

Home About News F.A.Q. Help & Tutorials Contact Publications

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

- View results summary in SVG - displayed below:

- View results in Jalview (Links to a separate page with the Jalview Java applet)
- View full results in HTML
- View simple results in HTML
- View results in PS
- View results in PDF
- View full multiple sequence alignment with gaps and insertions
- View full multiple sequence alignment without gaps and insertions
- View everything in a results directory
- Get all (but PS) files in TAR.GZ archive

This Jpred prediction was made with following:
Jnet version: 2.3.1
UniRef50 release: 2014_07_09-Jul-2014

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton G.J. Paper in preparation
Previous: Cole C, Barber JD & Barton G.J. J Mol Biol. 2008; 35 (suppl 2): W197-W201 [PMID]

結果のページ

Jpred4の予測結果

Jpred 4 Incorporating Jnet

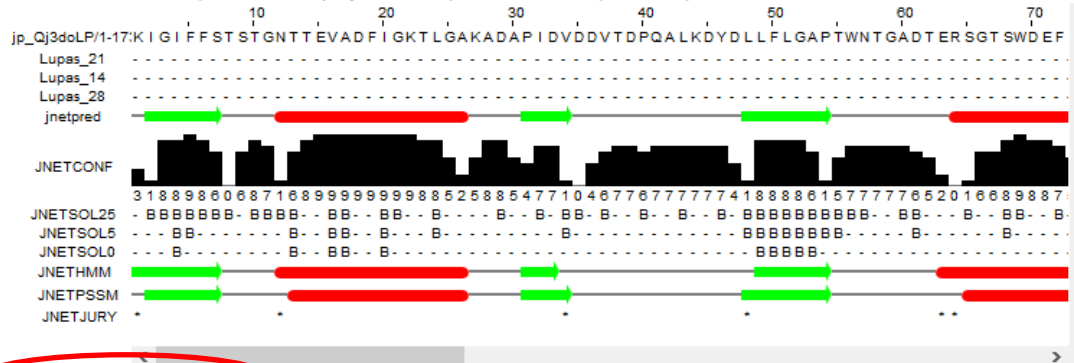
A Protein Secondary Structure Prediction Server

[Home](#) [REST API](#) [About](#) [News](#) [F.A.Q.](#) [Help & Tutorials](#) [Monitoring](#) [Contact](#) [Publications](#)

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

- [View results summary in SVG - displayed below \(details on acronyms used\):](#)



- [View full results in HTML](#)

- [View simple results in HTML](#)
- [View results in PDF](#)
- [View results in Jalview](#) (Link to a separate page with the Jalview Java Desktop application)
- [View everything in a results directory](#) (details on data each file contains are available through [README file](#))
- [Get all \(but PS\) files in TAR.GZ archive](#)
- [View results using in-browser Jalview Java applet](#) (light version, limited functionality w.r.t. Jalview Desktop version linked above)

Jpred4の予測結果の詳細

UniRef90_K9TFB1	: KIGLFFSTGTNTTEVADF IGKTLGAKADAP IDVDDVTPDQAL KDYDLLFLGAPTWNNTGADTERTSGTSWDFLYDKLPEVDMKDL PVAIFGLGDAEGYDPNFCDAIEEIHDFCAFAGKAGKPVGFSNPDDYDYEEESKSVRDKFLGLPLDMVNDQIPMEKRVAGWVEAVVSETGV
UniRef90_UP100034D1500	: KIGLFFGTGTGNTESVAELIQKEFGGESVYVSEEIENADPDSDFENYDCIIVGCPITWNVGELQSDWEFGYEDL----DAIDFTGKVVAYFGAGDQVGYADNFQDAMGILEEKISLGGKTVGYWPTDGYDFDGSKAVKNGKVFVGLAIDEDNQSDLTEGRITKSWVAQIKNEFSL
UniRef90_UP1000376B90C	: KIGLFFGTGTGNTESI AQTIQKELGGDSVVELQDVAQAEVEDLA EY EYL IIGCPTWNI GELQSDW----EGLYDDLDDIDFTGKIKIAYFGAGDQISYDNFQDAMGILEEKISLGGKTVGYWPTDNYDFNESKAVRNGKVFVGLAIDEDNQSDLTEGRITKSWVAQIKNEFSL
UniRef90_P31158	: Q1GLFYGSGTGTEDI AERI QEALGDDVYTLHEI SEAE N-SDFEQYDNL I IACPTWD I GELQSDW----DGYPPELDEVDVDFSGKTVAYFGTGDQIGYADNFQDAMGILEEKISLGGKTVGYWPTDGYDFNESKAVKNGKVFVGLAIDEDNQSDLTEGRITKSWVAQIKNEFSL
UniRef90_L8M7Y3	: KIGLFFGTGTGNTTEELAAQIQAAFFGGSDIVELFDVAEVDIEALRDFDQL IIGCPTWNVGELQSDWEALY----DLDLDDVDFSGKTIAYFGAGDQVGYADNFQDAMGILEEKISLGGKTVGYWPTDGYDFNESKAVRNGKVFVGLAIDEDNQPELTAERIQAQVVAQLKPAFGL
UniRef90_UP10002FE6068	: KVGLFYGSGTGTKTESAAEMI QSEFGGSSVYTLHEIADVSDSDFEQYDFL IIGAPTWD I GEL----QADWDGFYNDLDSIDFSGKTVAYFGTGDQLGYEENFQDAMGILEEKISLGGKTVGYWPTDGYEHAESKA EKNKGFVGLALDEDNQSELTESRITKSWVAQIKNEFSGT
UniRef90_D2RL1T	: KIGLFFGTGTGNTQTEAELIQKEFGGDSVVDIYDISKVESSDFENYNY I IIGCPTWNI GELQYDWNF----FDELNIIDFNGKIVAYFGVGDGNGYDPSFQDAMGILEEKISLGGKTVGYWPTDGYEHAESKA EKNKGFVGLALDEDNQSELTEGRITKSWVAQIKNEFSGT
UniRef90_UP100037CB4F4	: -VLLVYASMYGNTTESTASVLAALKAEKMTNITAMYDVSKTDDAFKYSHLVLASVYTNLGIYPKMLNLF-----EDMKALNLRKRIVGI IENGSWAC-----
UniRef90_A4RRP5	: -VLIYVSSRVDETQKIAELI AEGVRRHQVEVKTASQIETEEDLSGFDVAYFGSPTYHGEMLPMPKQVLFMAERAK-----LEGKPGGAFGAYGWEANKRIFDTMNYIFKMKMVSQPLMIKASHVEDGV-----
UniRef90_HI24T3	: -MHAEDALVAEGCCVFLQSTYGDGEPDTSSDFVYHCASDGRMPDLLENVTFVYFGLGNRC--YEQFNAAAKMVHAKALVDLGAQPLLLKHLGDDD-----QCLEQDFEN-----WIEAFWPAF-----
UniRef90_UP10004762B0B	: -ILITVYTKHGSTADI AWTIRNSFFDA-GYKADVKKIQDAEDIRQYSLI I IGTPIYEGKLM-----ETEEFVKLHRNYLNKNTALFITG--YSLRNKSPAGIQKAELAKLARHVDLVTGGGKLDAKNIPVYKEKISSLFRREGKIPGQYDWRIGEWADSLKE-----
UniRef90_D4CSL3	: -VLVAYASALGSTREIAQHMASRMAYVLGE-VECRSVEEVEAVSRYEAVVYVGSIAHNQAWLPPALLFF-----KHARELANRPVWAFVSVGMADALPKPFRGAALQOERLA-----
UniRef90_E5V4A3	: KVNIVYVYFTGNTLRMVKAFKELQEAAGVSFKSYVVELKNDDEAFDCEILALAS--PANQTEAIEKEYQOPMFRKNAERFKNKKIYLFGTFGWGTGMV----MSHWI KEVEELGAKI V ELPMAC-----
UniRef90_R6QW52	: KILLAYSSKTKNTKVAEAIYNEIKSLANVLLDIIKROKTKLQEQYDLYILGA-----WTDKTTANKMKQDFVNIQEI KDKKVALFLTCGVPREHYHADD SINNYIDFMKERGNDVLKTFVCGG-----
UniRef90_S0FYU9	: -VLIYASISYGGTENAANILASELAKGKIKNIAMHPSYI ISEAFRCSHI VFAAPTYNAGIFTPM-----ETLLDLLAAHNLQNRVYAFIENGSWAPISAKLMGDIYAKMKNMTVLAASKATLKSTVKDAQREELK-----
UniRef90_D5QDH9	: KVLVYFSQSGNTEKIAKACEAAKTSYIDLKLEELTPDMVAKYDCIFMGSPHSGSLAAPVKECLSVLKST-----SGQQMAGFITHMAYPEQDMDAFEEMKTAOREKG---IEYRGCFDQGFLEAMHEPVQKLGIDDE-----
UniRef90_Q2PCA9	: -ISIVYHSGYGHARQAEAAVEAGARRVAGADVKLISIGDWDTELEASDAIFGSPTYN-GAISAKFKFMEASTKTAWEQKWINKIAGFTNSGAQH-----
UniRef90_RICCM4	: KVV IAYGSAYGYTKHMAEKI AEGVKSVMGVVRSYDVAWLKDFGDAKGLLLGPTLVADAIPPMMI IACNLNPIYIHC-----DRYISCFGSHGWS-----
UniRef90_RICCM4	: KVFIPYYSAYGNTAKLAEKIAEGIKRAGDIDVDVNIILEERVEKSTAI IIGSPTINQNTLLPIYKLFVAVINPITN-----RGKLA AAFGSGYGSWGEVGI IESHKLNKLKI IEGGPRIGFVYEE-----
OrigSeq	: 1-----11-----21-----31-----41-----51-----61-----71-----81-----91-----101-----111-----121-----131-----141-----151-----161-----171
Jnet	: -EEEEEE--H-----E---E---E-----E-----H-----E-----H-----E-----H-----H-----H-----H-----H-----
jhm	: EEEEEEE--H-----E---E---E-----E-----H-----E-----E-----H-----E-----H-----H-----H-----H-----
jpsm	: -EEEEEE--H-----E---E---E-----E-----H-----E-----H-----E-----E-----E-----E-----E-----E-----
Lupas 14	: -----
Lupas 21	: -----
Lupas 28	: -----
Jnet_25	: -BBBBBBB-BBBB--BB--BB--B--B--B-BBBB-BBBB--BB--B--BB--B--B--B-B-B-B-B-BBBB-BBBB--BBB--BB--BB--B-B-BBB--B--B--B--B--BBB--B--B--BB--BB--B--
Jnet_5	: --BB--B--BB--B--B--B--B--B-BBBB-BBBB--
Jnet_0	: --B--
Jnet Rel	: 31899860687168999999998525895477104677677777418888615777776520166898875257777777771578743677777701325899999998517716572215776456013332113211147788877627899999999875189

Notes
Key:
 Colour code for alignment:
 Blue - Complete identity at a position
 Shades of red - The more red a position is, the higher the level of conservation of chemical properties of the amino acids
 Jnet - Final secondary structure prediction for query
 jalign - Jnet alignment prediction
 jhm - Jnet hmm profile prediction
 jpsm - Jnet PSIBLAST pssm profile prediction
 Lupas - Lupas Coil prediction (window size of 14, 21 and 28)
 Note on coiled coil predictions - = less than 50% probability
 c = between 50% and 90% probability
 C = greater than 90% probability
 Jnet_25 - Jnet prediction of burial, less than 25% solvent accessibility
 Jnet_5 - Jnet prediction of burial, less than 5% exposure
 Jnet_0 - Jnet prediction of burial, 0% exposure
 Jnet Rel - Jnet reliability of prediction accuracy, ranges from 0 to 9, bigger is better.

NetSurf 2.0による二次構造予測

DTU.dk

> Departments and Centers

> Shortcuts

Contact

Dansk

Search for text or person

DTU



DTU Health Tech

<https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0>

NEWS

EDUCATION

RESEARCH

COLLABORATION

SERVICES AND PRODUCTS

ABOUT US

NetSurfP - 2.0

Protein secondary structure and relative solvent accessibility

Server predicts the surface accessibility, secondary structure, disorder, and phi/psi dihedral angles of amino acids in an amino acid sequence.

There has been some portability issues for the output. This will be taken care for later. It is possible to see and export the output. Notice: it is a slow service.

Submission

Abstract

Instructions

Dataset

Downloads

Submit data

Paste in FASTA sequences or choose a file from your computer below. For detailed instructions, see "Help" tab above. Only amino acid input is accepted, maximum 100 sequences or a total of 100,000 residues .

For an input of less than 10 sequences, the HHblits method is used. For 10 and more sequences MMseqs is used to generate the sequence profiles.

For an overview of the methods, performance data and citation information is found under the Abstract/Cite tab above.

Sequence submission: paste the sequence(s) *and/or* upload a local file

```
>2FCR:A|PDBID|CHAIN|SEQUENCE
KIGIFFSTSTGNTTEVADF IGTKLGAKADAPIDVDDVTDPAQALKDYDLLFLGAPTWNTGADTERS
GT
SWDEFLYDKLPEV
DMKDLPVAVIFGLGDAEGYPDNFCDAIEEIHDCFAKQGAKPVGFSNPDDYDYEESKSVRDGKFLGL
PL
DMVNDQIPMEKRV
AGWVEAVVSETGV
```

← ターゲットの配列を入力

フラボドキシシン2fcrの配列

「[p0.fasta](#)」を使用

For example sequences [Click here](#)

Format directly from your local disk: ファイルが選択されていません。

参考 NetSurf 2.0による二次構造予測の結果

NetSurfP-2.0 フラボドキシン(2fcr)の適用例

NetSurfP server predicts the surface accessibility, secondary structure, disorder, and phi/psi dihedral angles of amino acids in an amino acid sequence.

Submission Help Abstract/Cite Data **Server Output**

Export All

Showing 1 Prediction

Below is a graphical representation of 173 residue predictions across 1 sequence. Running time was 94 seconds (94 seconds per sequence). Hover your mouse over a sequence position to see all outputs.

Relative Surface Accessibility: Red is exposed and blue is buried, thresholded at 25%.

Secondary Structure: Helix, Strand, Coil.

Disorder: Thickness of line equals probability of disordered residue.

データで、二次構造の8分類
予測、溶媒接触表面積、二
面角の予測結果が得られる

2FCR_A_PDBID_CHAIN_SEQUENCE

Export 2FCR_A_PDBID_CHAIN_SEQUENCE

K I G I F F S T S T G N T T E V A D F I G K T L G A K A D A P I D V D D V T D P Q A L K D Y D L L F L G A P T W N T G A D T E R S G T S W D E F L Y D K L P E V



D M K D L P V A I F G L G D A E G Y P D N F C D A I E E I H D C F A K Q G A K P V G F S N P D D Y D Y E E S K S V R D G K F L G L P L D M V N D Q I P M E K R V



A G W V E A V V S E T G V



二次構造予測結果のまとめ

アミノ酸配列 (2fcr)

KIGIFFSTSTGNTTEVADFIGKTLGAKADAPIDVDDVTDQPALKDYDLLFLGAPTWNTGADTERSGETSWDEFYDKLPEVDMKDLVVAIFGLGDAEGYP
DNFCDAIEEIHDCFAKQGAKPVGFSNPDDYDYEESKSVRDGKFLGLPLDMVNDQIPMEKRVAGWVEAVVSETGV

DSSPの結果

CEEEEECCSSSHHHHHHHHHHHHHGGGBCCEEGGGCSCGGGGGCSEEEEEEECCSTTCSSCCSCSTHHHHHHHTGGGCCCTTCEEEEEEEECTTTCT
TSTTTTHHHHHHHHHHHHTTCEEECCBCGGGSCCSCCTTEETTESSEEEETTTCSSCHHHHHHHHHHHHHHHHTC

3状態に変換

CEEEEECCCCCHHHHHHHHHHHHHHHHECCCEHHHCCCHHHHHHCCEEEEEEECCCCCCCCCCCCCHHHHHHHCHHHCCCCCEEEEEEECCCCC
CCCCCHHHHHHHHHHHHCCCEEECCCHHHCCCCCCCCCECCCECCCEEECCCCCCCCCHHHHHHHHHHHHHHHHHCC

PredictProtein

LEEEEEELLLLLHHHHHHHHHHHHHHLLLLLEEEEEELLLLLLLLLLLLLLEEEEEELLLLLLLLLLHHHHHHHHHHHHLLLLLLLLLLEEEEEELLLLLLLL
HHHHHHHHHHHHHHHHHHLLLLLEELLLLLLLLLLHHHHHHHHHHHHLLLLLEEEEEELLLLLLLLLLHHHHHHHHHHHHHHHHLL

PSIPRED

CEEEEECCCCCHHHHHHHHHHHHHCCCCCEEECCCCCHHHHHCCCEEEEEEECCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCEEEEEEECCCCCCC
CCHHHHHHHHHHHHHHCCCEEEEEEECCCCCCCCCHHEECCEEECCCCCCCCCHHHHHHHHHHHHHHHHHCC

Jpred4 (Jnet)

-EEEEEE----HHHHHHHHHHHHHHH----EEE-----EEEEEE-----HHHHHHHHHHH-----EEEEEE-----
--HHHHHHHHHHHHHHH----EEE-----HHHHH-HHHH-----HHHHHHHHHHHHH---

Jpred4 (jhmm)

EEEEEE----HHHHHHHHHHHHHHH----EEE-----EEEEEE-----HHHHHHHHHHH-----EEEEEE-----
-HHHHHHHHHHHHHHH----EEE-----HHHHHHHHHHH-----HHHHHHHHHHHHH--

Jpred4 (jpssm)

-EEEEEE----HHHHHHHHHHHHHHH----EEE-----EEEEEE-----HHHHHHHHH-----EEEEEE-----
---HHHHHHHHHHHHH---EEEEEE-----EEEE-----HHHHHHHHHHH---

NetSurfP

CEEEEEEECCCCCHHHHHHHHHHHHHHHCCCCEEEECCCCCCCCCCCCCEEEEEEECCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCEEEEEEECCCCC
HHHHHHHHHHHHHHHHHCCCCEEEECCCCCCCCCCCCCHCHHCCHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHHHCCC

NetSurfP-2.0

CEEEEECCSTSHHHHHHHHHHHHTGCCSSCEEGGGCCCGGGGGGCSEEEEEEECCCTTCCCCTTTTCCHHHHHHHHSTTCCTTSCEEEEEEECCCTTCT
TCHHHHHHHHHHHHHHTTCEEEEEECGGGCCHHHHHTEETTSEEEEECCCTTCCSCHHHHHHHHHHHHHHHHTC

NetSurfP-2.0 3状態に変換

CEEEEECCCCCHHHHHHHHHHHHHHHCCCCCEHHHCCCHHHHHHCCEEEEEEECCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCEEEEEEECCCCCCC
CCHHHHHHHHHHHHHHCCCEEEEEEECCCCCHHHHHHCCECCCEEEEECCCCCCCCCHHHHHHHHHHHHHHHHCCC

AlphaFold2

CEEEEECCSSSHHHHHHHHHHHHHGGGBCCEEGGGCSCGGGGGCSEEEEEEECCSTTCSSCCSCSHHHHIIIIIGGGCCCTTCEEEEEEECCCTTST
TSTTTTHHHHHHHHHHHHTTCEEECCBCGGGSCCSCCTTEETTESSEEEETTTCSSCHHHHHHHHHHHHHHHHTC

AlphaFold2 3状態に変換

CEEEEECCCCCHHHHHHHHHHHHHHHHECCCEHHHCCCHHHHHHCCEEEEEEECCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCEEEEEEECCCCCCC
CCCCCHHHHHHHHHHHHCCCEEECCCHHHCCCCCCCCCECCCECCCEEECCCCCCCCCHHHHHHHHHHHHHHHHCCC

二次構造予測の正答率

全体的な正答率
 Q_3 値

$$Q_3 = \frac{P_\alpha + P_\beta + P_C}{T}$$

α ヘリックス β シート それ以外

accuracyを示す

- P_i : 二次構造*i*の残基を二次構造*i*と正しく予測した数
- T : 全残基数

$T = 173$

手法	正答数	Q_3
PredictProtein	118	0.682
PSIPRED	123	0.711
Jpred4 (jnet)	123	0.711
Jpred4 (jhmm)	120	0.694
Jpred4 (jpssm)	131	0.757
SPIDER3	140	0.809
NetSurfP-2.0	144	0.832
NetSurfP-2.0 (Q8値)	129	0.746
Alphafold 2	169	0.977
Alphafold 2 (Q8値)	164	0.948

二次構造の予測例

discordant helixを含む例 → ヒトの α -ラクトアルブミン(1b9o)

アミノ酸配列(1b9o)

KQFTKCELSQLLKDIDGYGGIALPELICTMFHTSGYDTQAIVENDESTHEYGLFQISNKLWCKSSQVPQSRNICDISC
DKFLDDDDITDDIMCAKKILDIKGIDYWLAHKALCTEKLEQWLCEKL

DSSP

CBCCHHHHHHHTGGGTGGGCC**HHHHHHHHHHHH**TTBTTCEECSSECEETTTTEETTTTSBCTTCTTCCCTTCSBG
GGGSSCCHHHHHHHHHHHHTTTHHHHTTCCTTCSGGGGSCCC

PredictProtein

HHHHHHHHHH EEEEEEE EEE EEEEE EEE
HHHHH HHHHHHHHHHH HHHHHHHHHHHHHHHHHHH

Jpred4

----HHHHHHHHHH-----EEEEEE-----EEEEEE-----
HHHH----HHHHHHHHHHHHHH-----H--HH-----HHHHHH--

PSIPRED

CCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHCCCCCEEEECSCCCCCCEEEEECHHHHCCCCCCCCCCCCCCCC
CHHCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHCCCCCHHHHHCCC

NetSurf2

CEECHHHHHHHHHHTTTCTTCCCHHHHHHHHHHHHTCCCTCEEETTSCEEEEEEECHGGHCCTTCCCCCCTTSCCH
GGHTCSCHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHTTCCCHHHHHCC

AlphaFold2

CBCCHHHHHHHTGGGTGGGCCCHHHHHHHHHHHHTTBTTCEECSSECEETTTTEETTTTSBCTTCTTCCCTTCSBG
GGGSSCCHHHHHHHHHHHHHHHCGGGSHHHHHHSSSCCGGGCCCCC

溶媒露出度予測

- 多くは、二次構造予測のサイトで予測可能
 - Jpred4
 - PredictProtein (PROFacc)
 - PSIPRED
- 残基の疎水性が重要、周辺残基を含めた予測、類縁タンパク質の傾向を取り入れることにより、予測精度が向上
 - 予測手法として、機械学習がよく用いられる
 - 二次構造予測と共通した手法の枠組み
- 溶媒露出の状態の定義の例
 - 2状態
 - Buried (B): solvent accessibility < 16%
 - Exposed (E): solvent accessibility \geq 16%
 - 3状態
 - Buried (B): solvent accessibility < 9%
 - Intermediate (I): $9\% \leq$ solvent accessibility < 36%
 - Exposed (E): solvent accessibility \geq 36%

タンパク質の天然変性領域

- 天然変性領域 (disorder region): 生体内で一定の構造をとらない領域 (natively unfolded region, intrinsically disordered region, unstructured region)
 - タンパク質全体が天然変性領域のものは天然変性タンパク質 (disordered protein) と呼ばれる
- 構造がフレキシブルで、変動が大きい
- 全体的に荷電アミノ酸が多く、疎水性アミノ酸が少ない
- 高等生物にとくに多く見られる
- 転写調節に関するタンパク質や、DNA結合タンパク質などに多く見られ、機能的にも重要な場合がある
- X線結晶構造で、座標が欠失 (PDBのREMARK 465に記載)、変動が大きいループ領域など

タンパク質天然変性領域予測

- 塩基組成の特徴を学習させる
 - 構造全体の天然変性と局所的な天然変性を別々に予測
 - 局所的な予測では、類縁タンパク質の傾向のプロファイル化などが効果的
- 機械学習(ニューラルネットワーク、SVM、深層学習)などが用いられる
- 主な予測サイト
 - DISOPRED
 - 配列類似のタンパク質をもとに、プロファイルを形成し、アミノ酸配列の隣接15残基のウィンドウ幅でSVMにより予測
 - <http://bioinf.cs.ucl.ac.uk/psipred/>
 - PSIPREDの一部として公開
 - DisEMBL
 - コイル、hot loop (B-factorが大きいループ)、座標が欠失した領域を予測する3つのニューラルネットワークの統合、ウィンドウ幅は3~51
 - <http://dis.embl.de/>
 - PONDR
 - 特定のアミノ酸の組成、疎水性、配列の複雑度などを隣接9~21残基のウィンドウ幅でニューラルネットワークにより予測
 - <http://www.pondr.com/>

天然変性データベース

- DisProtデータベース
 - 天然変性領域を登録したデータベース
 - <https://www.disprot.org/>

The screenshot shows the DisProt website interface. At the top, there is a navigation bar with links for 'Browse', 'Release notes', 'Download', 'Help', and 'About'. A search bar is located on the right side of the header. The main content area features the DisProt logo and the title 'Intrinsically disordered proteins'. Below the title, a brief description states: 'DisProt is a database of intrinsically disordered proteins. Disordered regions are manually curated from literature. DisProt annotations cover both structural and functional aspects of disorder detected by specific experimental methods. Annotation concepts and detection methods are encoded in the Disorder Ontology. Read [more about DisProt](#).' A search bar is provided for searching within the database. Below the search bar, there are 'Examples' listed: P53, CTNNB1, SARS-CoV-2 Spike glycoprotein, ORF3a protein, and Replicase polyprotein 1ab. The page is divided into two main sections: 'Proteins per organism' and 'Datasets'. The 'Proteins per organism' section lists various organisms with their corresponding protein counts: *H. sapiens*: 636, *M. musculus*: 111, *R. norvegicus*: 57, *S. cerevisiae*: 137, *E. coli*: 74, *A. thaliana*: 40, *D. melanogaster*: 34, and *C. elegans*: 21. The 'Datasets' section lists 'Viruses: 144' and 'Unicellular toxins and antitoxins'. At the bottom, there is a 'How to cite' section with a citation for the 2020 release of DisProt.

DisProt

Version: 8.1
Release: 2020_12

Intrinsically disordered proteins

DisProt is a database of intrinsically disordered proteins. Disordered regions are manually curated from literature. DisProt annotations cover both structural and functional aspects of disorder detected by specific experimental methods. Annotation concepts and detection methods are encoded in the Disorder Ontology. Read [more about DisProt](#).

Search in DisProt ... Search

Examples P53 CTNNB1
SARS-CoV-2 Spike glycoprotein ORF3a protein Replicase polyprotein 1ab

Proteins per organism

<i>H. sapiens</i> : 636	<i>M. musculus</i> : 111	<i>R. norvegicus</i> : 57	<i>S. cerevisiae</i> : 137
<i>E. coli</i> : 74	<i>A. thaliana</i> : 40	<i>D. melanogaster</i> : 34	<i>C. elegans</i> : 21

Datasets

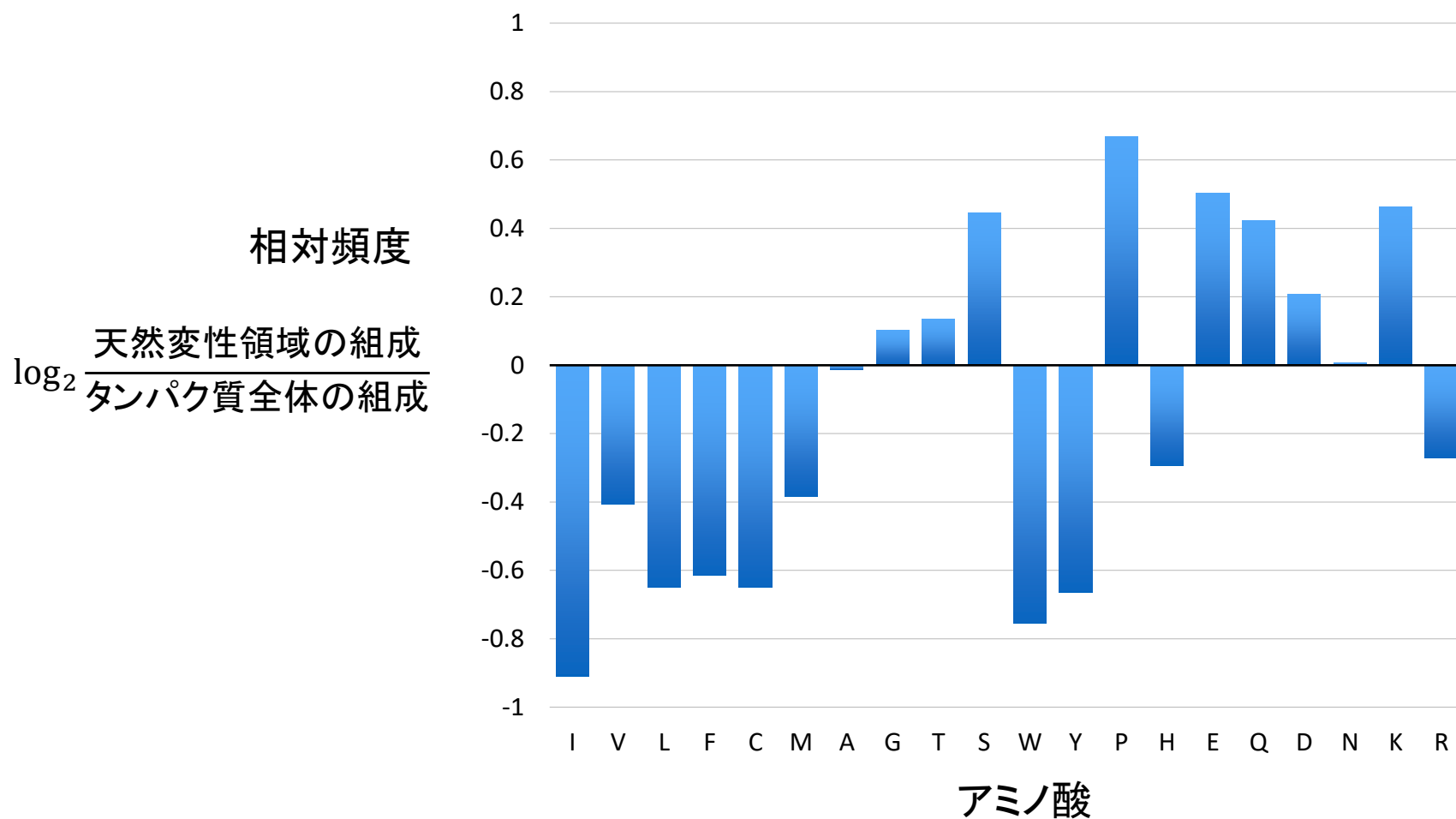
Viruses: 144	Unicellular toxins and antitoxins
Fungi: 171	

How to cite

DisProt: Intrinsic protein disorder annotation in 2020 Hatos, A., Hajdu-Soltész, B., Monzon, A.M., Palopoli, N., Álvarez, L., Aykac-Fas, B., Bassot, C., Benítez, G.I., Bevilacqua, M., Chasapi, A., Chemes, L., Davey, N.E., Davidović, R., Dunker, A.K., Elofsson, A., Gobeil, J., Foutel, N.S.G., Sudha, G., Guharoy, M., Horvath, T., Iglesias, V., Kajava, A.V., Kovacs, O.P., Lamb, J., Lambrugh, M., Lazar, T., Leclercq, J.Y., Leonardi, E., Macedo-Ribeiro, S., Macossay-Castillo, M., Maiani, E., Manso, J.A., Marino-Buslje, C., Martínez-Pérez, E., Mészáros, B., Mičetić, I., Minervini, G., Murvai, N., Necci, M., Ouzounis, C.A., Pajkos, M., Paladin, L., Pancsa, R., Papaleo, E., Parisi, G., Pasche, E., Barbosa Pereira, P.J., Promponas, V.J., Pujols, J., Quaglia, F., Ruch, P., Salvatore, M., Schad, E., Szabo, B., Szaniszló, T., Tamana, S., Tantos, A., Veljkovic, N., Ventura, S., Vranken, W., Dosztányi, Z., Tompa, P., Tosatto, S.C.E., Piovesan, D. (2020) *Nucleic Acids Research*, 48 (D1), pp. D269-D276. [PubMed](#) [NAR](#)

天然変性領域のアミノ酸組成

- 天然変性領域のアミノ酸組成
 - DisProtデータベースの配列から計算



プリオンタンパク質の天然変性予測(1)

- プリオンタンパク質 (p2.fasta) の天然変性予測 → [DISOPRED](#)を利用

PSIPRED

UCL Department of Computer Science: Bioinformatics Group

UCL

MAIN NAVIGATION

- Site Links
- Server Links
- PSIPRED Workbench
- PSIPRED Overview
- Server Citation
- Help & Tutorials
- News
- Contact
- History
- Store
- PSIPRED Github

Data Input

Select input data type

Sequence Data PDB Structure Data

Choose prediction methods

Popular Analyses

- PSIPRED 4.0 (Predict Secondary Structure)
- DISOPRED3 (Disopred Prediction)
- MEMSAT-SVM (Membrane Helix Prediction)
- pGenTHREADER (Profile Based Fold Recognition)

Contact Analysis

Submission details

Protein Sequence

```
>p2
GSKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGGWGQPHGGGGWGQPHGG
GWGQGGGTHSQWN
KPSKPKTNMKHMAGAAAAGAVVGLGGYMLGSAMSRPIHFSDYEDRYRENMHRYPNQVYYRPMDEYSNQNNF
```

Help...
If you wish to test these services follow this link to retrieve a test fasta sequence.

Job name

p2

Email (optional)

Email (optional)

Reset Submit

Required Options

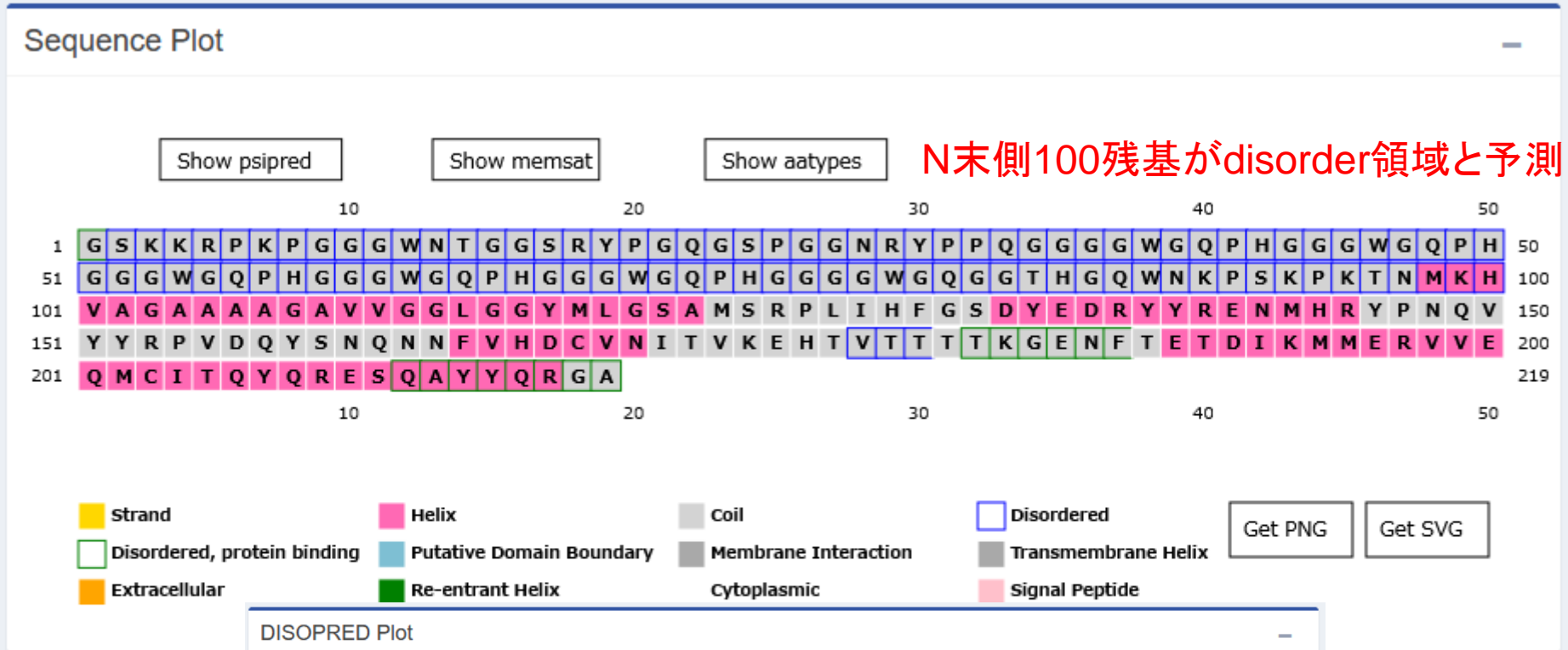
ヒトのプリオンタンパク質の配列
(p2.fasta)を入力

Emailは入力した方がよい(未記入だと促される)

与えられた配列からdisorder領域(残基)を予測
指定したアドレスにメールで結果が返すことも可能

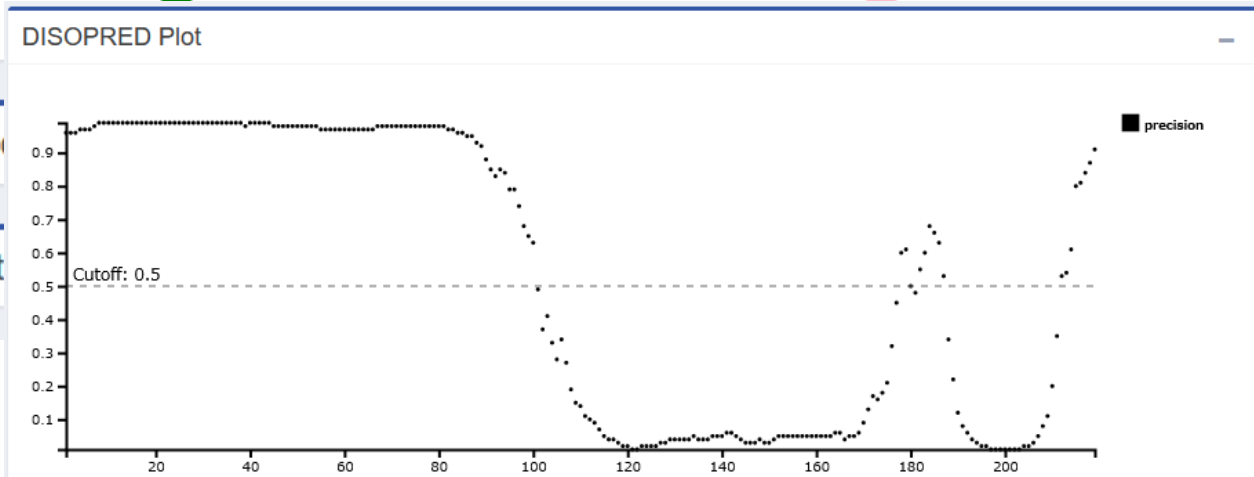
プリオンタンパク質の天然変性予測(2)

DISOPREDの予測結果

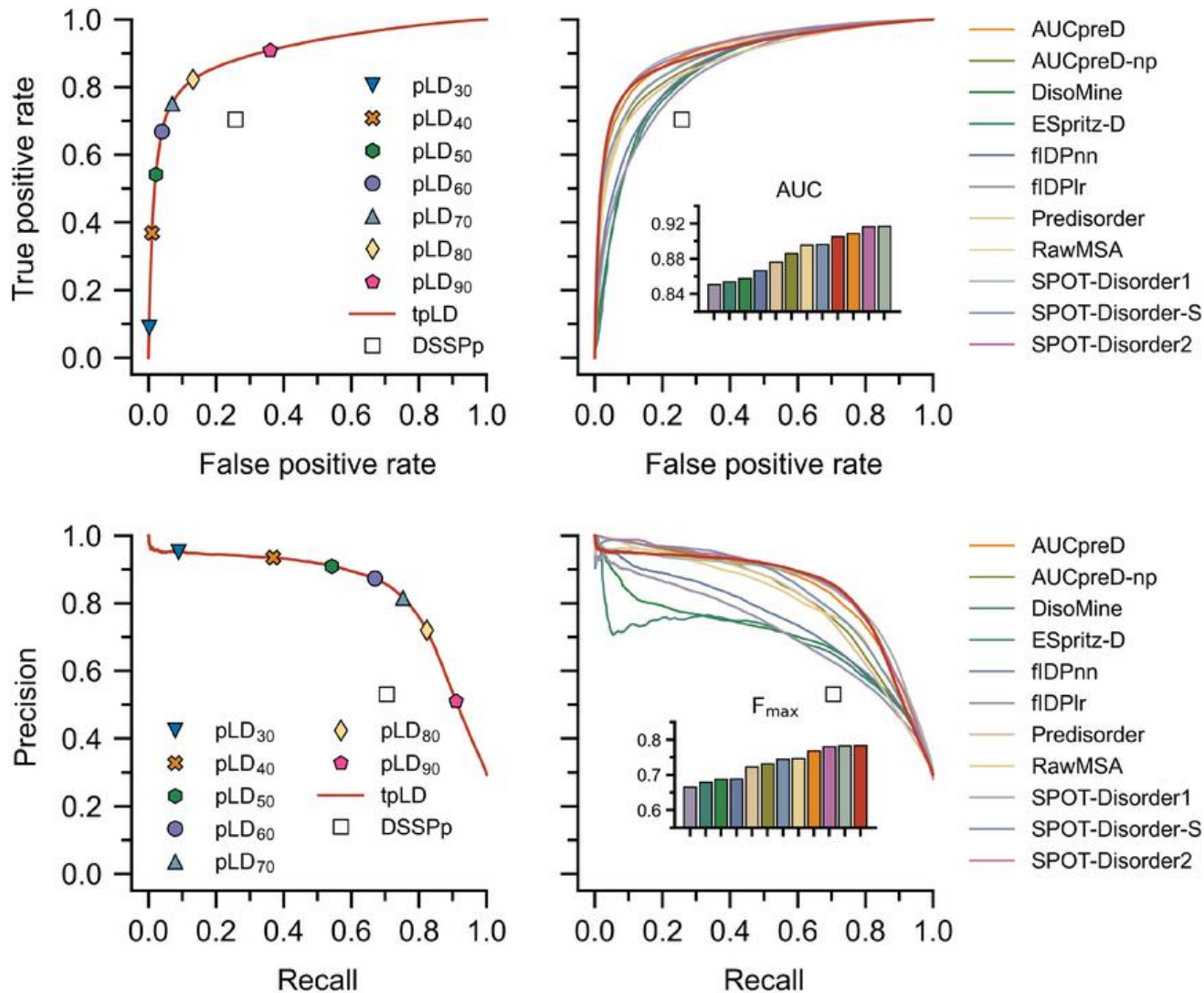


PSIPRED Carto

DISOPRED Plot



AlphaFold2による天然変性領域の予測



[Wilson, C.J., et al. AlphaFold2: A Role for Disordered Protein/Region Prediction? Int. J. Mol. Sci. \(2022\) 23: 4591.](#)