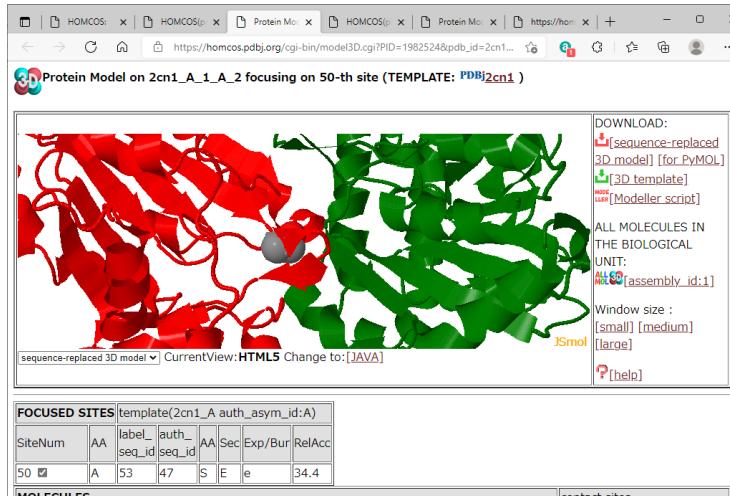
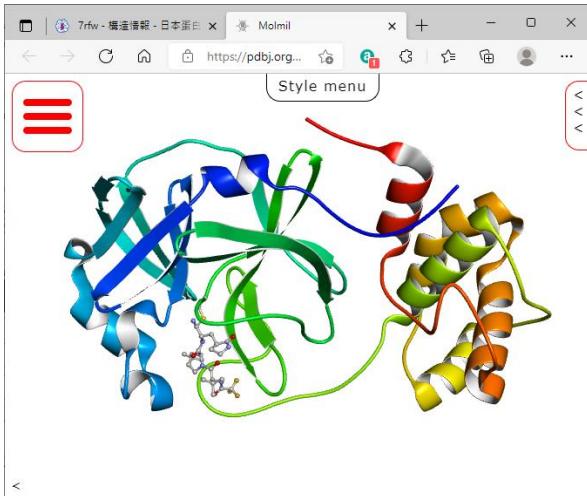


AJACSオンライン10「タンパク質のデータベースを活用する」

主催:国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター(NBDC)

PDBjを使ってタンパク質の立体構造データを見る



2022年1月20日(木)
13:40-15:20
オンライン開催

蛋白質研究奨励会 研究員・
大阪大 生命機能 招へい准教授

kawabata@prf.or.jp kawabata@protein.osaka-u.ac.jp

川端 猛

本日の話題

1. PDBデータのなりたち

どんな情報が収納されているのか？

どういうファイルフォーマットで書かれているのか？

PDBデータの検索方法

2. 分子ビューアの基本的な使い方

Molmilの使い方を紹介

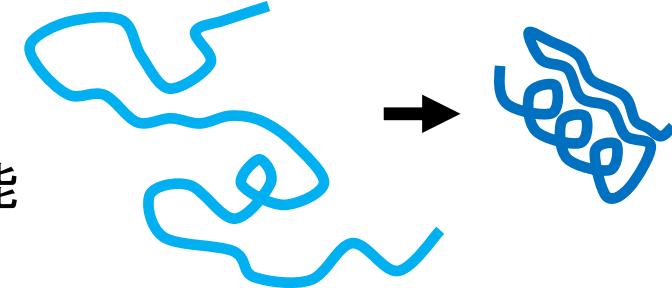
3. アミノ酸配列からの3Dモデリング

HOMCOSサーバを用いたホモジーモデリング

AlphaFold2による予測について

1. PDBデータのなりたち

Protein Data Bank (PDB)

- ・タンパク質は、アミノ酸の組成や並び方によって、様々な形状・物性・機能。天然の球状タンパク質は、生理条件下でただ一つのコンパクトな形に折りたたまる。その立体的な形状は原子スケールで決定可能
- ・PDBは1971年に米国で7つの立体構造からスタート。
- ・2003年以降、三つの拠点(米国のRCSB PDB, 欧州のPDBe, 日本のPDBj)によるworldwide PDBで、PDBの登録・配布を行っている。PDBjは、大阪大学の蛋白質研究所内に設置。
- ・ほとんどの学術雑誌が、立体構造決定を報じる研究論文の出版の条件としてPDBへの登録を義務付けている。
- ・立体構造から、分子機能や構造安定性に関する知見が得られ、分子進化や創薬の研究にも役立つ。
- ・PDBには、球状タンパク質だけではなく、膜タンパク質、補酵素や薬剤と結合したタンパク質、複数のタンパク質や核酸を含む複合体の構造も登録されている。
- ・天然変性タンパク質(生理条件下で一定の形を持たない)は、他の分子と結合し一つに定まった構造のみ登録されている

PDBのデータのなりたち

- 現在のデータ数 (2022/01/12): 1,85,935件

| X線結晶解析 | NMR | 電子顕微鏡 | その他 | 合計 |
|---------------|-------------|-------------|------------|--------|
| 162259(87.3%) | 13592(7.3%) | 9547 (5.1%) | 537 (0.3%) | 185935 |

- 各エントリには4文字の英数字のPDB_IDを持つ(2hhb, 1a5kなど)

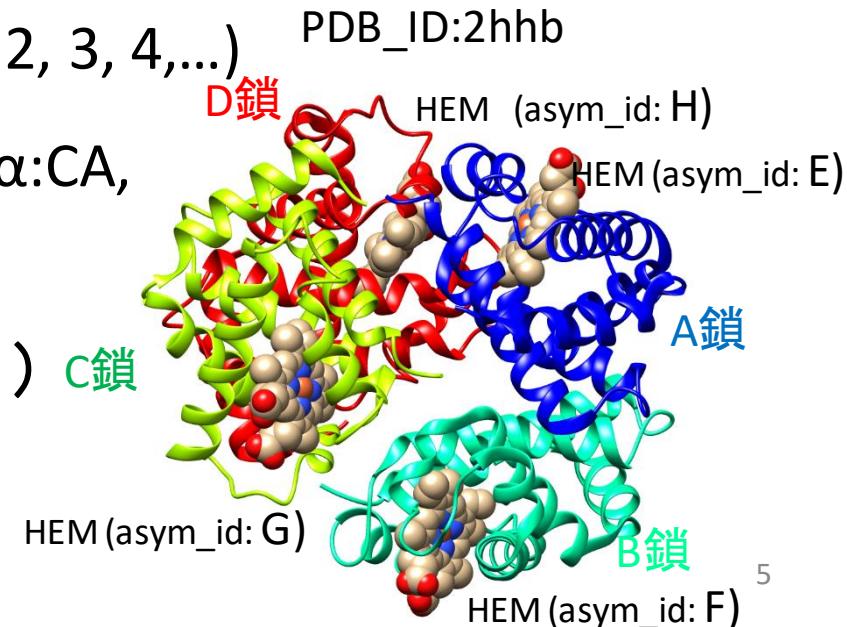
- 各高分子には、鎖識別子asym_id (A,B,C,D...)が付く

- 各残基には残基番号seq_idがつく(1, 2, 3, 4,...) PDB_ID:2hhb

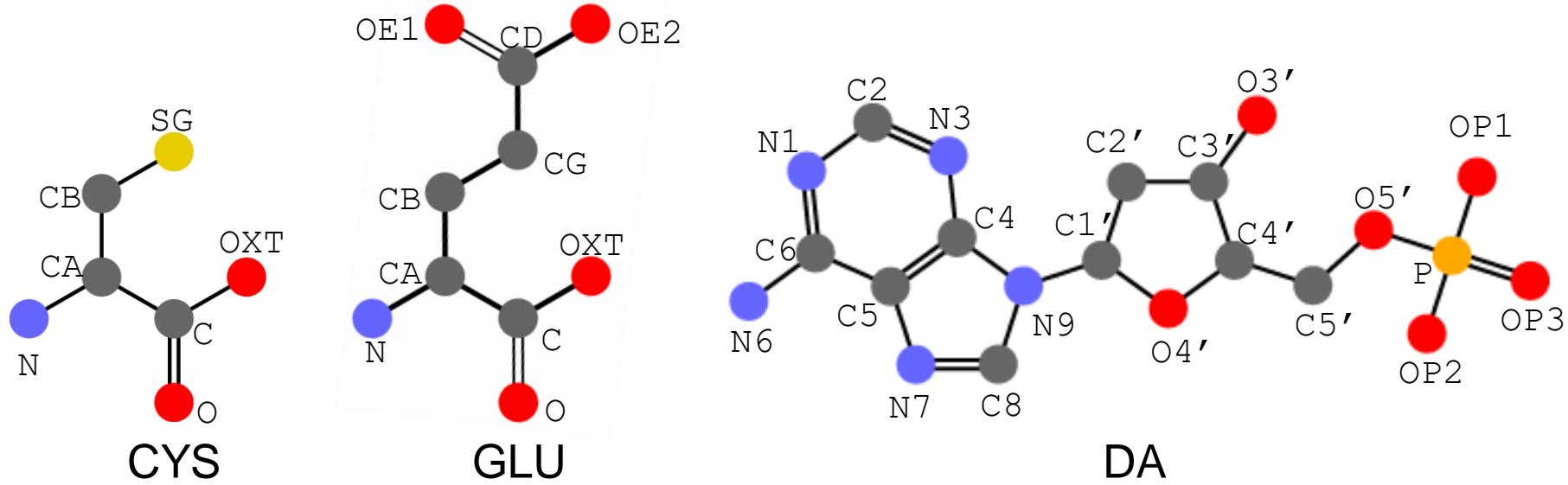
- 各原子には固有の原子名atom_id(Cα:CA, Cβ:CB, Cγ:CGなど)が付加

- 低分子化合物には化合物名(残基名) comp_id (ALA, HEM, ATP,...)が付く

- 各分子ごとにlabel_asym_idが付く



残基名と原子名



PDBで使われている原子名の例。重原子のみ示した。CYSはシステイン、GLUはグルタミン酸、DAは、DNAの構成要素となる2'-deoxyadenosine-5'-monophosphate.

※mmCIFでは、残基名はlabel_comp_idという名前、原子名はlabel_atom_idと呼ばれる

※残基名は、これまで3文字以下。しかし、今後、拡張される予定

※Chemical Components Dictionaryというデータベースで管理

※PDBjのWEBサイトで、残基名で検索すると、これらの原子名の命名法を見ることができます

PDBのファイルフォーマットについて

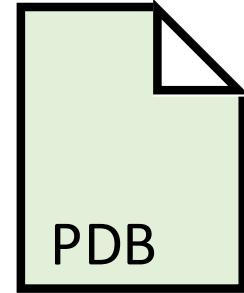
なんとなく構造を眺めるだけなら、

- PDBのウェブページをクリックして、3D構造を眺めるだけなら、ファイルフォーマットを意識する必要はありません
- PyMOL、UCSF Chimeraなど最近のビューワは、直接PDBサーバから構造を読みにいく機能("fetch")があるので、ファイルフォーマットを意識する必要はありません
- ただし、残基名、残基番号、鎖識別子などの事情はある程度理解しておいたほうがいいかもしれません

立体構造データのフォーマット

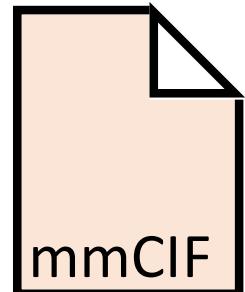
• 旧PDBフォーマット

- ・PDB創立のころ(1971年)から使われていたフォーマット
- ・1行1原子。1行80文字の固定長のフォーマット。
- ・鎖識別子は1文字なので、62鎖を超えると扱えない。
- ・原子番号は5文字なので、99999個を超える原子は扱えない。
- ・座標以外のデータの記述は場当たり的



• mmCIFフォーマット

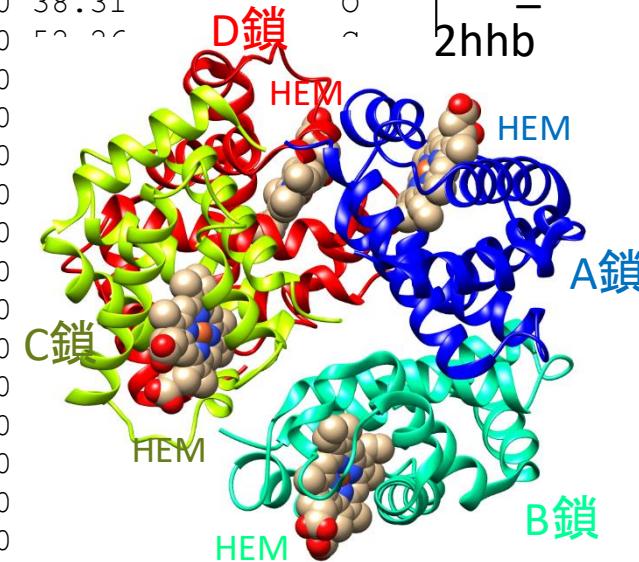
- ・STARフォーマットの一種。可変長。
- ・キー・バリュー形式と表形式を併用
- ・値の長さは可変なので、鎖や原子数の上限はない。
- ・座標以外の様々なデータも、柔軟に記述できる
- ・2014年から、PDBの正規なフォーマットとなる
- ・2019年7月から、X線のデータ登録にmmCIFが必須となる
- ・主要な分子表示、モデリング、構造決定のソフトウェアは対応済み



旧PDBフォーマットの立体構造データの例

PDB

PDB_ID:
2hhb



| HEADER | OXYGEN TRANSPORT | | | | | | | | | | 07-MAR-84 | 2Hhb |
|--------|---|-----|-----|---|-----|--------|--------|---------|------|-------|-----------|------|
| TITLE | THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS | | | | | | | | | | | |
| ATOM | 1 | N | VAL | A | 1 | 6.130 | 16.559 | 4.905 | 1.00 | 41.29 | N | |
| ATOM | 2 | CA | VAL | A | 1 | 6.870 | 17.784 | 4.702 | 1.00 | 41.33 | C | |
| ATOM | 3 | C | VAL | A | 1 | 8.377 | 17.548 | 4.913 | 1.00 | 31.64 | C | |
| ATOM | 4 | O | VAL | A | 1 | 8.820 | 16.980 | 5.922 | 1.00 | 38.31 | O | |
| ATOM | 5 | CB | VAL | A | 1 | 6.345 | 18.763 | 5.731 | 1.00 | ~ ~ | ~ | |
| ATOM | 6 | CG1 | VAL | A | 1 | 6.745 | 20.188 | 5.356 | 1.00 | | | |
| ATOM | 7 | CG2 | VAL | A | 1 | 4.826 | 18.612 | 5.847 | 1.00 | | | |
| ATOM | 8 | N | LEU | A | 2 | 9.146 | 18.005 | 3.962 | 1.00 | | | |
| ATOM | 9 | CA | LEU | A | 2 | 10.599 | 17.914 | 4.153 | 1.00 | | | |
| ATOM | 10 | C | LEU | A | 2 | 11.062 | 19.085 | 5.062 | 1.00 | | | |
| ATOM | 11 | O | LEU | A | 2 | 10.829 | 20.254 | 4.744 | 1.00 | | | |
| ATOM | 12 | CB | LEU | A | 2 | 11.269 | 18.078 | 2.776 | 1.00 | | | |
| ATOM | 13 | CG | LEU | A | 2 | 10.986 | 16.983 | 1.769 | 1.00 | | | |
| ATOM | 14 | CD1 | LEU | A | 2 | 11.735 | 17.168 | 0.427 | 1.00 | | | |
| ATOM | 15 | CD2 | LEU | A | 2 | 11.276 | 15.630 | 2.404 | 1.00 | | | |
| ATOM | 16 | N | SER | A | 3 | 11.674 | 18.766 | 6.158 | 1.00 | | | |
| ATOM | 17 | CA | SER | A | 3 | 12.286 | 19.774 | 7.034 | 1.00 | | | |
| ATOM | 18 | C | SER | A | 3 | 13.529 | 20.322 | 6.334 | 1.00 | | | |
| ATOM | 19 | O | SER | A | 3 | 13.995 | 19.754 | 5.344 | 1.00 | 27.40 | O | |
| ATOM | 20 | CB | SER | A | 3 | 12.719 | 19.060 | 8.326 | 1.00 | 23.83 | C | |
| ATOM | 21 | OG | SER | A | 3 | 13.844 | 18.245 | 8.107 | 1.00 | 29.07 | O | |
| : | | | | | | | | | | | | |
| HETATM | 4389 | CHA | HEM | A | 142 | 8.634 | 7.898 | -18.334 | 1.00 | 17.24 | C | |
| HETATM | 4390 | CHB | HEM | A | 142 | 10.386 | 9.965 | -14.276 | 1.00 | 23.17 | C | |
| HETATM | 4391 | CHC | HEM | A | 142 | 8.337 | 6.403 | -11.611 | 1.00 | 11.04 | C | |
| : | 原子番号 原子名 残基名 残基番号 鎖識別子 | | | | | X座標 | Y座標 | Z座標 | 占有率 | 温度因子 | 元素名 | |

80文字の固定長

パンチカード時代の名残り

mmCIFフォーマットの立体構造データの例

```

data_2Hhb
#
_entry.id      2Hhb
#
_struct.title   ' THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS
_struct.pdbx_descriptor 'HEMOGLOBIN (DEOXY) '
#
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num

```

項目は[カテゴリ名].[アイテム名]

※2hhbの場合58個のカテゴリがある

データベース管理者が入力(label_のアイテム)

原子座標

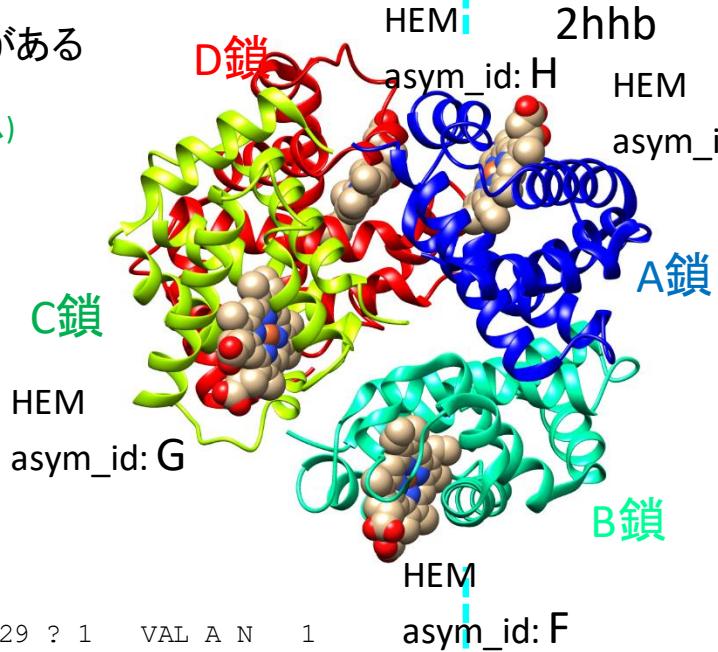
登録者が入力する(auth_のアイテム)

| | 原子番号 | 元素名 | 原子名 | 残基名 | 残基番号 | asym_id(鎖識別子) | X座標 | Y座標 | Z座標 | 占有率 | 温度因子 | 残基番号 | asym_id(鎖識別子) | 残基名 | 原子名 | | | |
|--------|------|-----|-----|-----|------|---------------|-----|-----|-----|--------|--------|---------|---------------|-------|-----|-----|-----------|---|
| ATOM | 1 | N | N | . | VAL | A | 1 | 1 | ? | 6.130 | 16.559 | 4.905 | 1.00 | 41.29 | ? | 1 | VAL A N | 1 |
| ATOM | 2 | C | CA | . | VAL | A | 1 | 1 | ? | 6.870 | 17.784 | 4.702 | 1.00 | 41.33 | ? | 1 | VAL A CA | 1 |
| ATOM | 3 | C | C | . | VAL | A | 1 | 1 | ? | 8.377 | 17.548 | 4.913 | 1.00 | 31.64 | ? | 1 | VAL A C | 1 |
| ATOM | 4 | O | O | . | VAL | A | 1 | 1 | ? | 8.820 | 16.980 | 5.922 | 1.00 | 38.31 | ? | 1 | VAL A O | 1 |
| ATOM | 5 | C | CB | . | VAL | A | 1 | 1 | ? | 6.345 | 18.763 | 5.731 | 1.00 | 52.26 | ? | 1 | VAL A CB | 1 |
| ATOM | 6 | C | CG1 | . | VAL | A | 1 | 1 | ? | 6.745 | 20.188 | 5.356 | 1.00 | 52.75 | ? | 1 | VAL A CG1 | 1 |
| ATOM | 7 | C | CG2 | . | VAL | A | 1 | 1 | ? | 4.826 | 18.612 | 5.847 | 1.00 | 58.75 | ? | 1 | VAL A CG2 | 1 |
| ATOM | 8 | N | N | . | LEU | A | 1 | 2 | ? | 9.146 | 18.005 | 3.962 | 1.00 | 27.63 | ? | 2 | LEU A N | 1 |
| ATOM | 9 | C | CA | . | LEU | A | 1 | 2 | ? | 10.599 | 17.914 | 4.153 | 1.00 | 33.62 | ? | 2 | LEU A CA | 1 |
| HETATM | 4385 | C | CHA | . | HEM | E | 3 | . | ? | 8.634 | 7.898 | -18.334 | 1.00 | 17.24 | ? | 142 | HEM A CHA | 1 |
| HETATM | 4386 | C | CHB | . | HEM | E | 3 | . | ? | 10.386 | 9.965 | -14.276 | 1.00 | 23.17 | ? | 142 | HEM A CHB | 1 |
| : | | | | | | | | | | | | | | | | | | |
| # | | | | | | | | | | | | | | | | | | |

キー・バリュー形式

mmCIF

表形式



PDB_ID
2hhb

HEM
asym_id: E

B鎖

HEM

asym_id: F

D鎖
HEM
asym_id: H

C鎖

HEM

asym_id: G

PDB_ID
2hhb

HEM
asym_id: E

A鎖

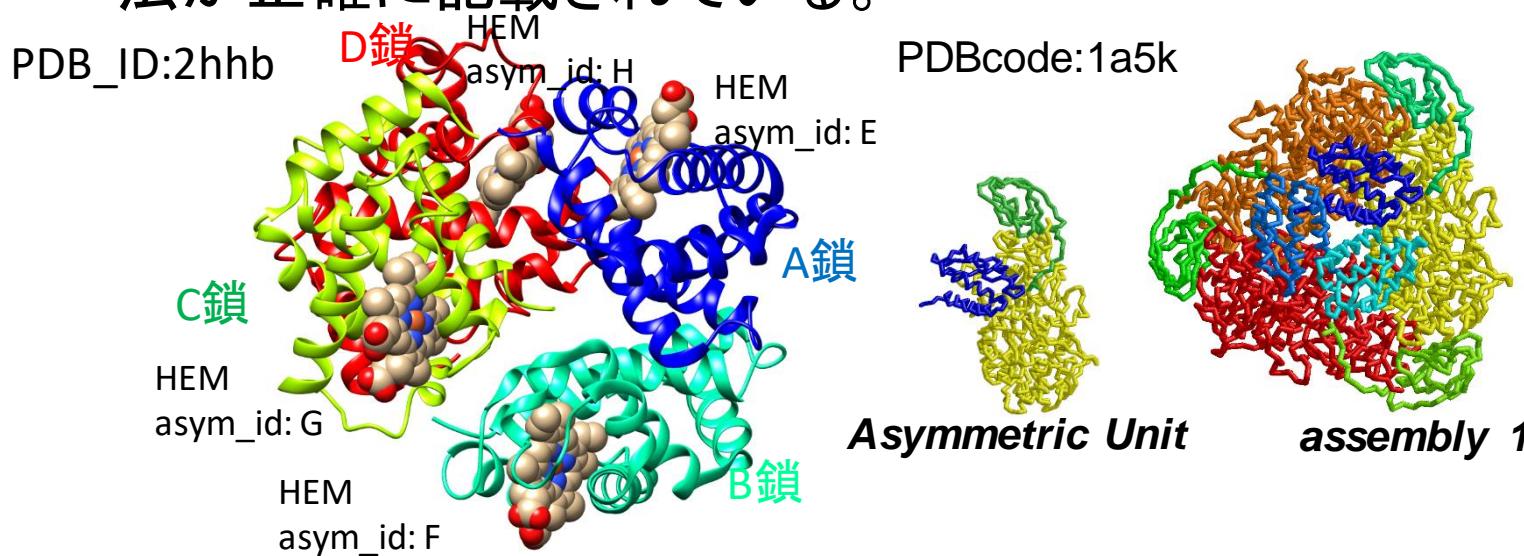
HEM

asym_id: F

mmCIFファイルの特長

- STARフォーマット。キーバリュー形式と表形式の併用で、人間にも比較的読みやすい。
- [カテゴリ名].[アイテム名]で整理された項目名
- カテゴリatom_site以外に、50-80ほどのカテゴリ(表)
- 鎖識別子、原子番号などの文字数は無制限
- 残基番号、鎖識別子については、読みやすさと統一性を両立するため、著者入力とデータベース管理者入力の2種が書かれる。
- Biological unitなど、非対称単位以外の複合体(assembly)の生成法が正確に記載されている。

Relionの出力ファイルもSTAR
フォーマットです！



mmCIFを使わなければならぬ場合

1. 巨大構造を扱う場合は、mmCIFを読み込むべき

巨大な構造の場合、mmCIF形式のみで公開されています。

鎖識別子が2文字以上の場合、巨大でなくてもmmCIFだけの構造はあります

2. 立体構造を扱うプログラムの開発者は、入力をmmCIFに対応するよう改良すべき。(出力は任意)

計算結果の出力は、旧PDBフォーマットでもかまいません。

3. PDBデータベース全体に対して、統計的・包括的な解析を行う場合は、mmCIFをベースにすべき

mmCIFのデータベース管理者による残基番号、鎖識別子は、統一的な基準で決まっているので、PDBデータベース全体の統計をとる場合に向いています。

PDB形式変換不可エントリ:mmCIFだけで公開されているエントリ

PDB形式変換不可の PDB エントリ - https://pdbj.org/list/pdb-legacy-incompatible

185935 件を2022-01-12に公開中

PROTEIN DATA BANK JAPAN
NARAIKA UNIVERSITY
PDBj
Protein Data Bank Japan
Worldwide Protein Data Bank Foundation

Search pdbj.org

English 日本語 简体中文 繁體中文 한국어

wwwPDB RCSB PDB PDBe BMRB Adv. Search Search help

2021年1月12日時点で
2498エントリ

ホーム
トップページ
統計情報
ヘルプ
FAQ
お問い合わせ
PDBjの引用・利用規約
リンク集
Settings

データ登録 (OneDep)
ヘルプ
PDB、EMDB、BMRBへの登録

ダウンロード
PDBアーカイブからのデータダウンロード

標準フォーマット
PDBx/mmCIFについて
フォーマット変換

PDB形式変換不可エントリ

1BOS SHIGA-LIKE TOXIN COMPLEXED WITH ITS RECEPTOR
分子名称: SHIGA-LIKE TOXIN I B SUBUNIT, alpha-D-galactopyranose-(1-4)-beta-D-galactopyranose, alpha-D-galactopyranose-(1-4)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose, ...
著者: Ling, H., Boodhoo, A., Hazes, B., Cummings, M.D., Armstrong, G.D., Brunton, J.L., Read, R.J.
登録日: 1998-01-13
公開日: 1999-02-02
最終更新日: 2020-07-29
実験手法: X-RAY DIFFRACTION (2.8 Å)
主引用文献: Structure of the shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3. Biochemistry, 37, 1998

1VVJ Crystal Structure of Frameshift Suppressor tRNA SufA6 bound to Codon CCC-G on the Ribosome
分子名称: 16S rRNA, 23S rRNA, 30S ribosomal protein S10, ...
著者: Maehigashi, T., Dunkle, J.A., Dunham, C.M.
登録日: 2013-05-24
公開日: 2014-08-06
最終更新日: 2014-12-10
実験手法: X-RAY DIFFRACTION (3.44000111574 Å)
主引用文献: Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. Proc.Natl.Acad.Sci.U.S.A, 111, 2014

1VY4 Crystal structure of the Thermus thermophilus 70S ribosome in the pre-attack state of peptide bond formation containing acylated tRNA-substrates in the A and P sites.
分子名称: 16S Ribosomal RNA, 23S Ribosomal RNA, 30S ribosomal protein S10, ...
著者: Polikanov, Y.S., Steitz, T.A., Innis, C.A.
登録日: 2014-05-13
公開日: 2014-08-20
最終更新日: 2019-07-03
実験手法: X-RAY DIFFRACTION (2.6 Å)
主引用文献: A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. Nat.Struct.Mol.Biol., 21, 2014

検索結果
全ヒット件数: 2498
表示件数: 25
表示順: PDBID昇順 (0→9,a→z)

表示順
PDBID昇順 (0→9,a→z)
PDBID降順 (z→a,9→0)
登録日の古い順
登録日の新しい順
公開日の古い順
公開日の新しい順
分解能の高い順

Lists
IDリストをダウンロード
更新された PDB エントリー
新しい PDB エントリー
廃止された全ての PDB エントリー
廃止された最新の PDB エントリー
PDB形式変換不可の PDB エントリー
PDBグループ登録エントリー
X線構造因子データを含む全てのPDBエントリー
X線構造因子データを含む新しいPDBエントリー
X線構造因子データが更新されたPDBエントリー
X線構造因子データを含む全ての廃止された

鎖識別子(auth_asym_id)を2文字以上使うと、原子数が小さくてもmmCIFだけの公開となり、「PDB形式変換不可エントリー」とみなされる。

PDB IDの探し方

(1)文献検索 構造を決定した論文には必ずPDB IDが書いてある

GoogleやPubMedなどで、関心があるタンパク質・ファミリーネームといっしょに、3D structure, X-ray, NMR, Cryo-EMなどのキーワードとともに文献を検索

(2)PDBjのキーワード検索

PDBjのトップ画面の上のフォームに
タンパク質名、ファミリーネームなどを入力



詳細検索(PDBjのRDBを直接検索)も可能

(3)PDBjのアミノ酸配列検索

関心のあるタンパク質のアミノ酸配列から
配列相同意検索で類似した蛋白質の
立体構造を検索

配列が類似した蛋白質は構造も類似
しているので、ホモロジーモデリング
の鋳型としても使える

A screenshot of the Sequence Navigator interface on the PDBj website. On the left, there is a sidebar with various search services: 検索サービス (Search Services), ヘルプ (Help), PDB検索 (PDBj Mine), PDB詳細検索 (Detailed PDB search), 化合物検索 (Chemie), BMRB検索 (BMRB search), Sequence-Navigator (Sequence Navigator, highlighted with a red box), DASH, EM Navigator, Omokage検索 (Omokage search), SeSAW, wwPDB/RDF, NBDC RDFポータル (NBDC RDF portal), and 未公開エントリーのステータス (Status of unpublished entries). The main area is titled "Sequence Navigator". It shows a sequence entry labeled "Custom sequence" containing the amino acid sequence: MDMFQKVEKIGEGTYGVVYAKNRETGQIVALKKIRLDLEMEGVPSTAIREISLLKELKH... The area also includes settings for Sequence type (Protein selected), Clustering (No clustering selected), and a button for Find homologues.

HOMCOSを利用した配列からのPDB IDの検索

The screenshot shows the HOMCOS search interface. At the top, there is a navigation bar with a back button, forward button, home icon, and a search bar containing the URL https://homcos.pdbj.org/cgi-bin/prot_sch_inp.cgi?&LANG=ja. The main title "HOMCOS" is displayed prominently. Below it, there is a logo consisting of a blue cube and a white sphere, with the text "HOMCOS:タンパク質に対する結合分子の検索". On the left, there are links for "[Go to English page]" and "[ヘルプページ]". The main search area has a large input field labeled "(2) UniProtIDかアミノ酸配列を入力". Below this, there is a smaller input field for "アミノ酸配列". A red box highlights the UniProt ID "NCAP_SARS2" entered into the main input field. To the right of the input fields, there is explanatory text in Japanese: "性" (Species), "アミノ酸配列" (Amino acid sequence), and "のアップロードのどれかの方法で入力してください" (Please enter using one of the methods for file upload). At the bottom, there is a table with two rows. The first row contains the UniProt ID "NCAP_SARS2" and its corresponding protein information: "(CDK3_HUMAN, Q00526, AAV40830.1, NP_001249.1)". The second row contains the amino acid sequence "LPVAGWKDQIINWATEGALNTPKOHIGTRNPANNAIAVLQLPQGTTLKPGFYAE6SRGGSSQASRSRSGNSRNSTPGSSRGTSRPARMGHNQGDALALLLDRPLNQELSKM6GK6Q00QCGQTWTIKSAEAXKKPRKRATKAYNVTQGSRGPPEQTOGNGFQGQELIR0GTDYKHNPQ140FAPSSAEEFGCSTGMETPTSPCTMLTYKGAIKLQDKOPNFVKDQ01CLNRHTDAYKTFPPTEPKOKKKKAFTEGALPQR9KQ9GTVILLPAADDLDSKQLQGSMSADSTQH". A red box highlights the entire sequence.

(3) 類似蛋白質の立体構造が複合体の種類別に表示される

2. 分子ビューアの使い方

MolMilを中心として

分子ビューア: 立体構造を表示するソフト

| ソフトウェア名 | プラットフォーム | ライセンス | 特徴 |
|--------------|---|----------------------------------|---|
| PyMOL | Win/Mac/Linux用のソース・実行バイナリの公開 | 機能制限版は無料 全機能版は有料ライセンスを購入する必要 | X線・NMRの研究者のユーザーが多い。 PyMOL Wikiで様々なスクリプトが配布 |
| UCSF Chimera | Win/Mac/Linux用の実行バイナリが公開 | 非商用な使用のみ無償。商用目的では有償でライセンス契約を結ぶ必要 | 電顕の3Dマップを扱う機能が充実。最新版 Chimera Xに移行中。 |
| JSmol | ブラウザ内で起動 Java Scriptで書かれている | 無料 | ネットワークにつながったブラウザがあればインストール不要。 |
| MolMil | ブラウザ内で起動 WebGLを用いたJava Scriptで書かれている | 無料 | ネットワークにつながったブラウザがあればインストール不要。 |



Gert-jan Bekker

Molmil

- PDBjのGert-jan Bekker氏が開発した分子ビューア
- ブラウザ内で分子を表示。WebGLを使用
- インストール作業が不要
- JavaScriptで開発
- ブラウザ内の標準的なビューアJSmolに比べて、表示がきれいで動作が速い
- 自分のWEBページに3次元の分子をかざることができる
- mmCIFファイルにきちんと対応

Gert-Jan Bekker, Haruki Nakamura & Akira R. Kinjo. Molmil: a molecular viewer for the PDB and beyond. Journal of Cheminformatics 8, Article number: 42 (2016)

SARS-CoV-2の3C様プロテアーゼPDBID 7rfwの観察

(1) <https://pdbj.org>のフォームに7rfwと入力

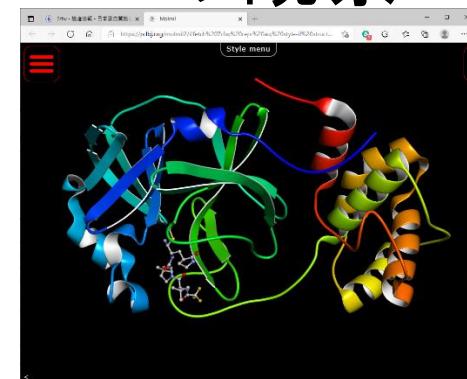
The screenshot shows the Protein Data Bank Japan (PDBj) website. At the top, there is a logo for OSAKA UNIVERSITY and the PDBj logo. Below the logo, there are links for English, 日本語 (Japanese), 简体中文 (Simplified Chinese), 繁體中文 (Traditional Chinese), and 한국어 (Korean). A search bar contains the text '7rfw'. Below the search bar, there are buttons for 'PDB' and 'Promote Elastic'. The URL 'https://www.pdbj.org/entry/7rfw' is visible at the bottom of the search bar.

(2) PDB_ID: 2ht7 のページが表示される

This screenshot shows the detailed entry page for PDB ID 2ht7. The top navigation bar includes tabs for '概要' (Summary), '構造情報' (Structure Information), '実験情報' (Experimental Information), '機能情報' (Functional Information), '相同蛋白質' (Homologous Proteins), '履歴' (History), and 'ダウンロード' (Download). The '構造情報' tab is highlighted with a red box. The main content area displays the title 'Structure of SARS-CoV-2 main protease in complex with a covalent inhibitor' and the identifier '7RFW'.

This screenshot is identical to the previous one, showing the 'Structure' tab selected for PDB ID 2ht7.

(4) ページ左の
[非対称単位を表示]
をクリックすると、molmil
で構造が表示される



(3) [構造情報]のタブをクリックすると、エントリ内の分子の一覧が表示

This screenshot shows the 'Molecules' section of the PDBj entry page. It lists various molecules associated with the entry:

| 種類 | 長さ | 分子量 | 分子数 | データベース名(アクセス番号) | 由来する生物種 |
|--|-------------|-------|---------|-----------------|---|
| kinase | polymer | 307 | 33882.6 | 1 | UniProt (P0DTD1) In PDB |
| V-((1E,2S)-1-imino-3-[(3S)-2- 1-3-((1R,2S)-2-yl)-6,6-dimethyl-3-[3- fluorooacetyl]-L-valyl]-3- 1.0]hexane-2-carboxamide | non-polymer | 501.5 | | | Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) |
| comp_id "4WI" | water | 18.0 | | | |

Below this table, there is a section for 'comp_id "4WI"':

comp_id "4WI"
Chemie (4WI) 4WI

4WIの文字にボ
インタをかざす
と構造式が表
示される

生物的単位を表示
生物学的単位を表示

3CL-PRO, 3CLp,
Main protease,
Mpro
Severe acute
respiratory
syndrome
coronavirus 2
(2019-nCoV)

3CL-PRO, 3CLp,
Main protease,
Mpro
Severe acute
respiratory syndrome
coronavirus 2 (2019-nCoV)

エンティティの一般名
3CL-PRO, 3CLp,
Main protease,
Mpro
Severe acute respiratory syndrome coronavirus 2 (2019-nCoV)

エンティティの一般名
3CL-PRO, 3CLp,
Main protease,
Mpro
Severe acute respiratory syndrome coronavirus 2 (2019-nCoV)

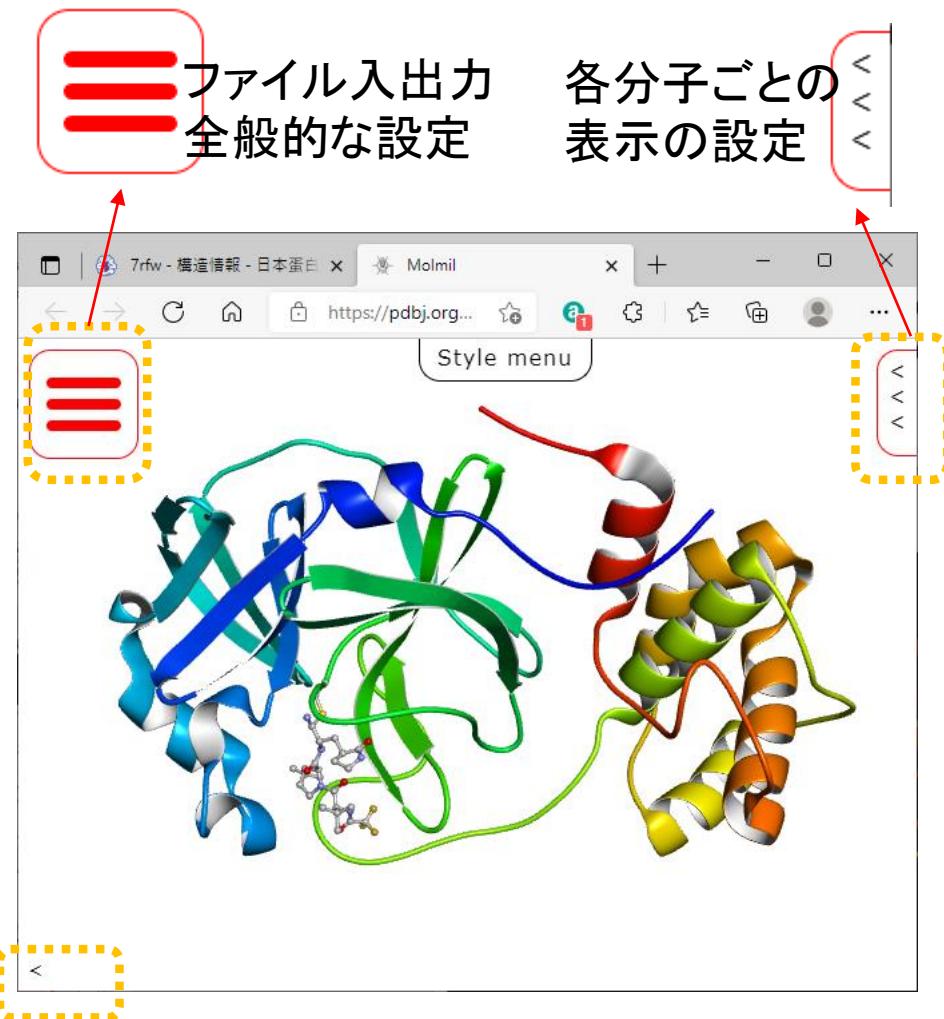
3C-likeプロテアーゼ(nsp5, 3CL-PRO, 3CLp, Main protease, Mpro) :タンパク質を切断する酵素タンパク質。ホモ二量体。コロナウィルスは、ポリプロテイン1abという7096アミノ酸からなる非常に長いタンパク質を、まず、合成。その後、その長いタンパク質を切断して、15個のタンパク質にする。そのとき、タンパク質を切断する酵素の一つ。治療薬の候補として、このタンパク質の阻害剤が開発されている。

comp_id: 4WI 別名: PF-07321332, bound form
ニルマトレルビル nirmatrelvir, bound form; Paxlovid, bound form パクスロビド

マウスの使い方

| 表示法 | <i>PyMOL</i> | <i>UCSF Chimera</i> | <i>JSmol</i> | <i>Molmil</i> |
|-------------|------------------------------------|---|--|--|
| 分子の回転 | 左ボタンで画面をドラッグ | 左ボタンで画面をドラッグ | 左ボタンで画面をドラッグ | 左ボタンで画面をドラッグ |
| 分子の並進 | 中ボタン(ホイール)で画面をドラッグ | 中ボタン(ホイール)で画面をドラッグ | Ctrlキーを押しながら、右ボタンで画面をドラッグ | (1)中ボタン(ホイール)で画面をドラッグ (2)[Shift]キー+左ボタン |
| ズームイン・アウト | 右ボタンで画面をドラッグ | (1)右ボタンで画面をドラッグ (2)ホイールを回す | (1)ホイールを回す (2)Shiftキーを押しながら、左ボタンで画面をドラッグ (3)右ボタンと左ボタンを同時に押して、画面をドラッグ | (1)右ボタンで画面をドラッグ (2)ホイールを回す |
| 断面の表示(slab) | ホイールを回す | [Tools] →[Viewing Controls]→[Side View] | コンソールからコマンド slab 50を入力 | [Settings]→[Slab]を☑、スライダで調整 |
| 残基名・原子名の確認 | 左ボタンで原子をクリックする、アッパーコントロール・ウィンドウに表示 | 画面上で原子の上にマウスポイントをかざすと、原子名のラベルが表示 | 画面上で原子の上にマウスポイントをかざすと、ポップアップでラベルが表示される | 左ボタンで原子をクリックすると、画面下に表示 |

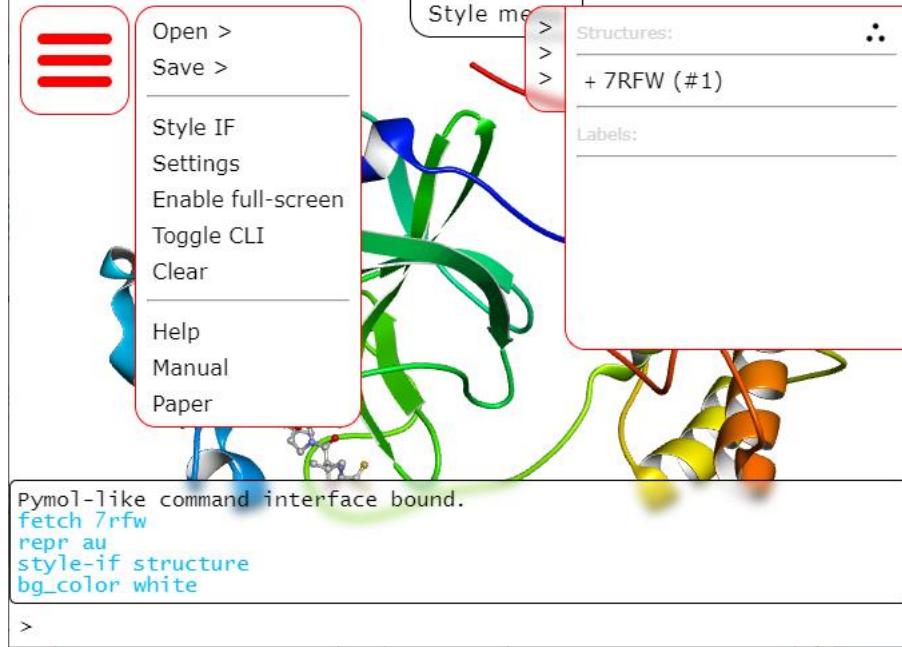
molmilの三つのインターフェース



CLI (Command Line Interface)
コマンドを入力して表示を変える

ファイル入出力
全般的な設定

各分子ごとの
表示の設定



CLI (Command Line Interface)
コマンドを入力して表示を変える

分子メニューによる表示の変更法

The diagram illustrates the use of molecular menus to change display settings for two protein structures: **分子A (3C-like protease)** and **分子B (4WI; PF-07321332; Paxlovid)**.

分子A (3C-like protease) Display Settings:

- Left-click the top-level menu icon (red box).
- Right-click the structure entry in the list (red box). A context menu appears with options: **Display**, **Color**, and **Label**.
- Left-click the **Display** option in the context menu.
- Left-click the **Ca trace** option in the expanded display menu (red box).

分子B (4WI; PF-07321332; Paxlovid) Display Settings:

- Left-click the top-level menu icon (red box).
- Left-click the structure entry in the list (red box). A context menu appears with options: **+ A**, **+ B (A)**, and **+ C (A)**.
- Left-click the **+ B (A)** option in the context menu.

Coloring Options:

- Left-click the top-level menu icon (red box).
- Left-click the structure entry in the list (red box). A context menu appears with options: **+ A**, **Chain A > Color**, and **Labels**.
- Left-click the **Chain A > Color** option in the context menu.
- Left-click the **Structure** option in the expanded color menu (red box).

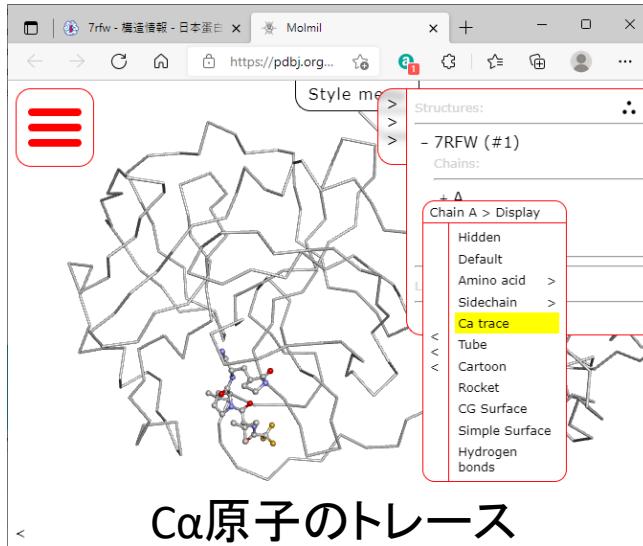
Information Panel:

comp_id: **4WI** 別名: **PF-07321332**, bound form
nirmatrelvir, bound form;
Paxlovid, bound form

ニルマトレルビル
パクスロビド

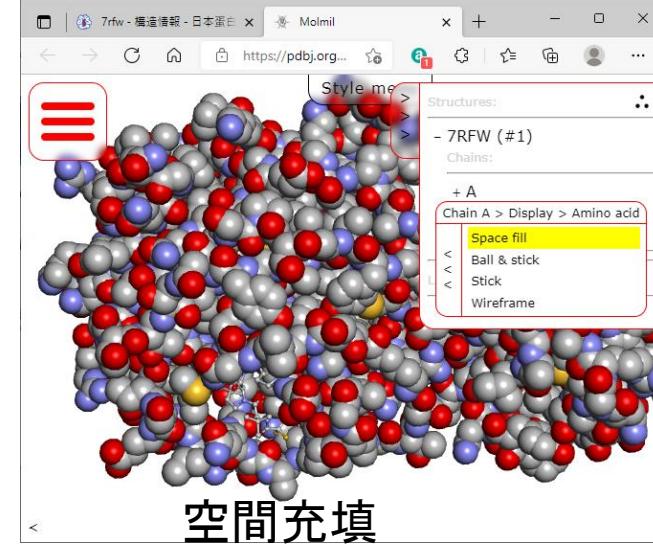
分子メニューによる表示の変更例

[Chain A]→[Display]→[Ca trace]



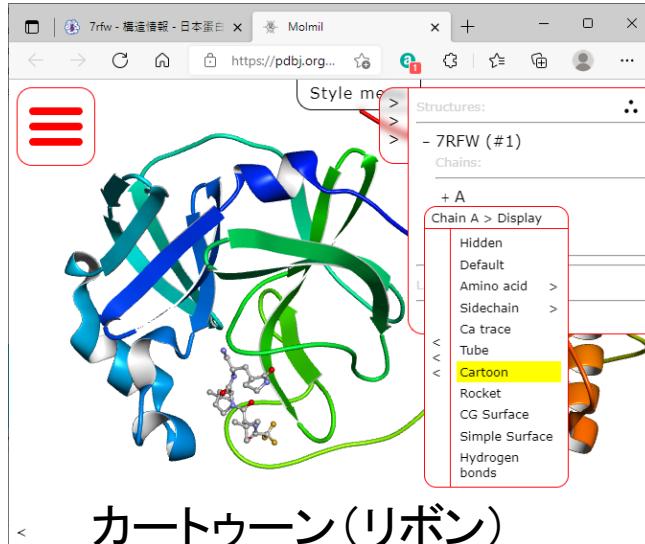
Ca原子のトレース

[Chain A]→[Display]→[Amino acid]→[Space fill]



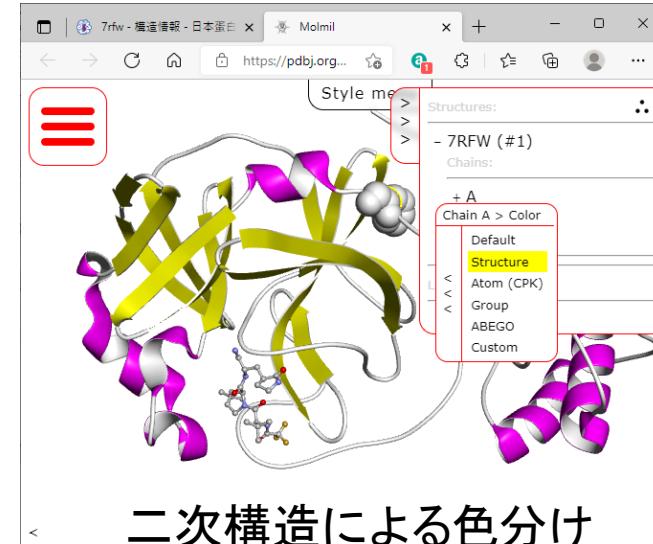
空間充填

[Chain A]→[Display]→[Cartoon]



カートゥーン(リボン)

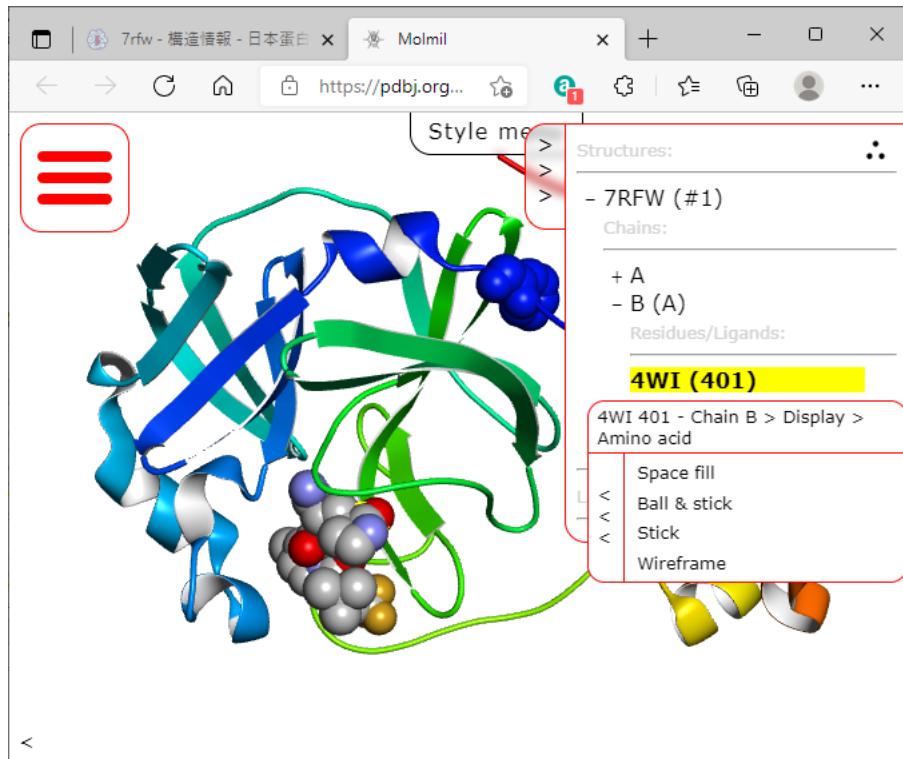
[Chain A]→[Color]→[Structure]



二次構造による色分け

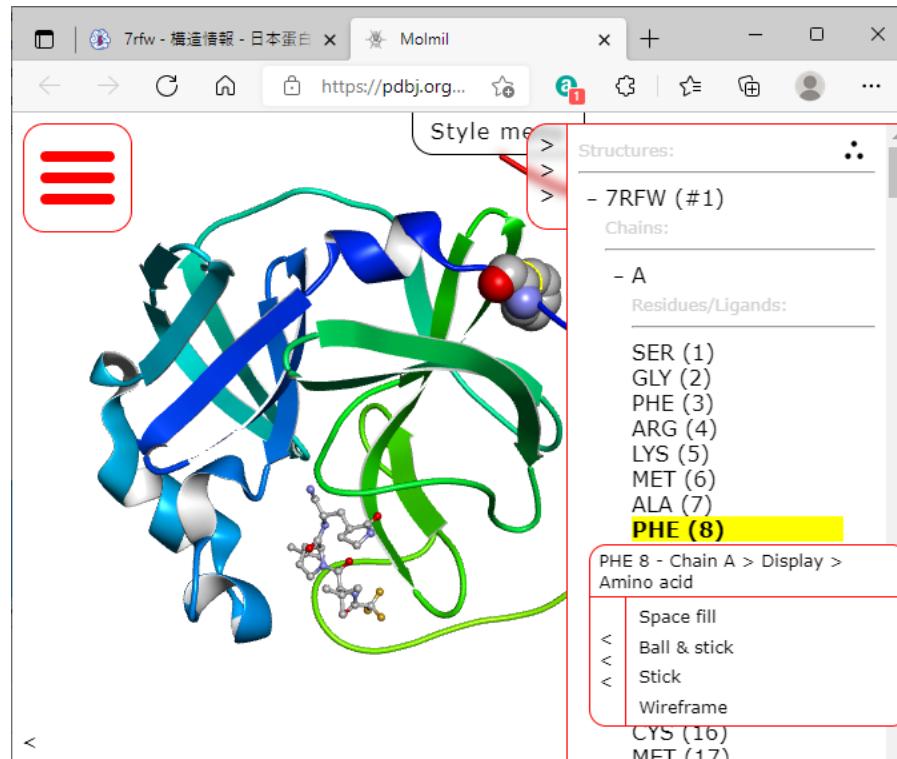
分子メニューによる分子・残基ごとの表示の変更

[4WI – Chain A]→[Display]→[Amino acid]→[Spacefill]



化合物4WIを空間充填表示

[PHE 8 Chain A]→[Display]→[Amino acid]→[Space fill]



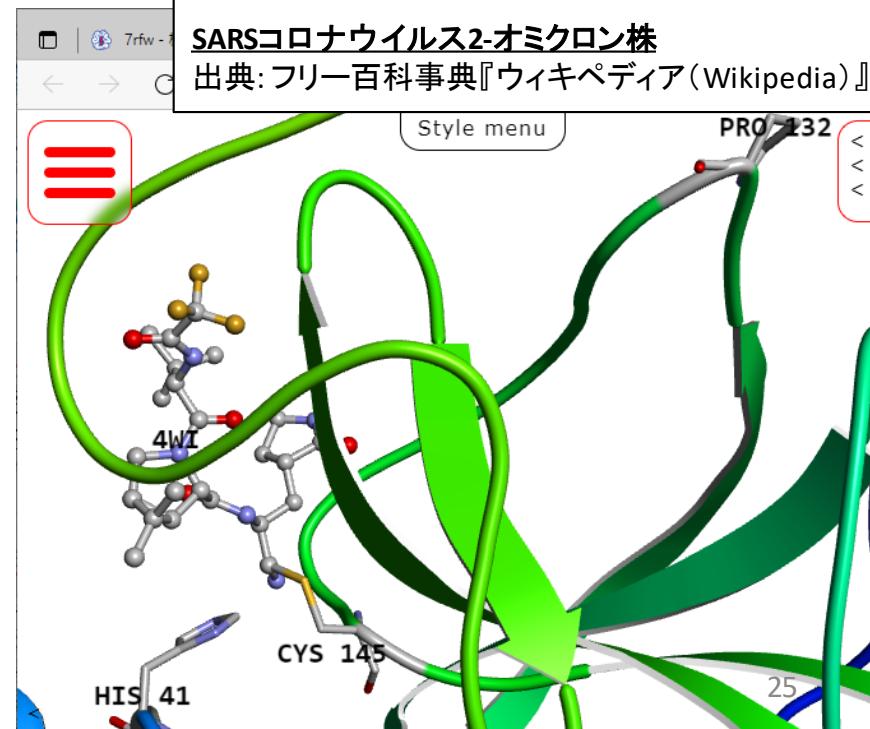
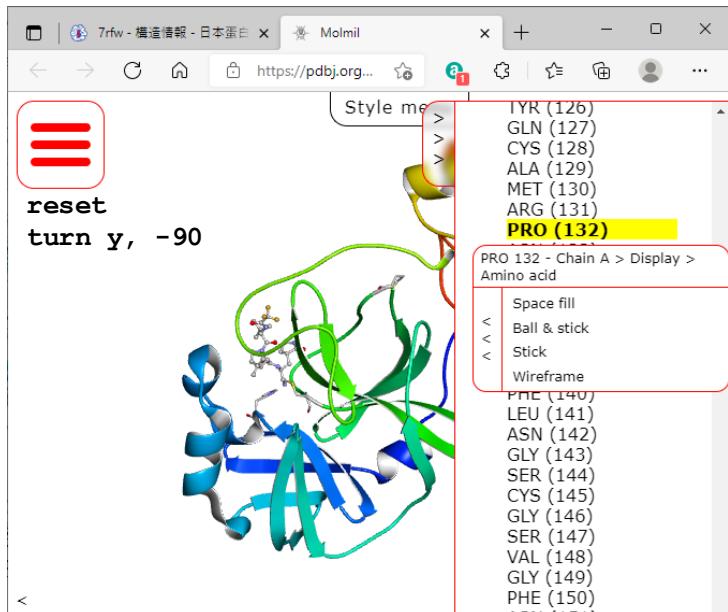
PHE8を空間充填表示

特定の残基を指定して表示(7rfw)

変位部位: P132H

活性部位: His41, Cys145

| | | | |
|----|--|---|----------|
| ID | R1AB_SARS2 | Reviewed; | 7096 AA. |
| AC | P0DTD1; | | |
| DE | RecName: Full=Replicase polyprotein 1ab; | | |
| DE | Short=pp1ab; | | |
| FT | CHAIN | 3264..3569 | |
| FT | | /note="3C-like proteinase nsp5" | |
| FT | | /evidence="ECO:0000250 UniProtKB:P0C6V3" | |
| FT | | /id="PRO_0000449623" | |
| FT | ACT_SITE | 3304 His41 | |
| FT | | /note="For 3CL-PRO activity" | |
| FT | ACT_SITE | 3408 Cys145 | |
| FT | | /note="Nucleophile; for 3CL-PRO activity" | |



スパイクプロテイン: A67V, Δ69-70, T95I, G142D, Δ143-145, Δ211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, [要説明], G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F
30のうちの半分(15)の変更はレセプター結合ドメイン-RBDに位置する(residues 319–541)

ORF1ab
nsp3: K38R, V1069I, Δ1265, L1266I, A1892T
nsp4: T492I
nsp5: P132H
nsp6: Δ105-107, A189V
nsp12: P323L
nsp14: I42V

コロナウイルスエンベローププロテイン: T91

Membrane protein: D3G, Q19E, A63T

Nucleocapsid protein: P13L, Δ31-33, R203K, G204R

SARSコロナウイルス2-オミクロン株

出典: フリー百科事典『ウィキペディア(Wikipedia)』

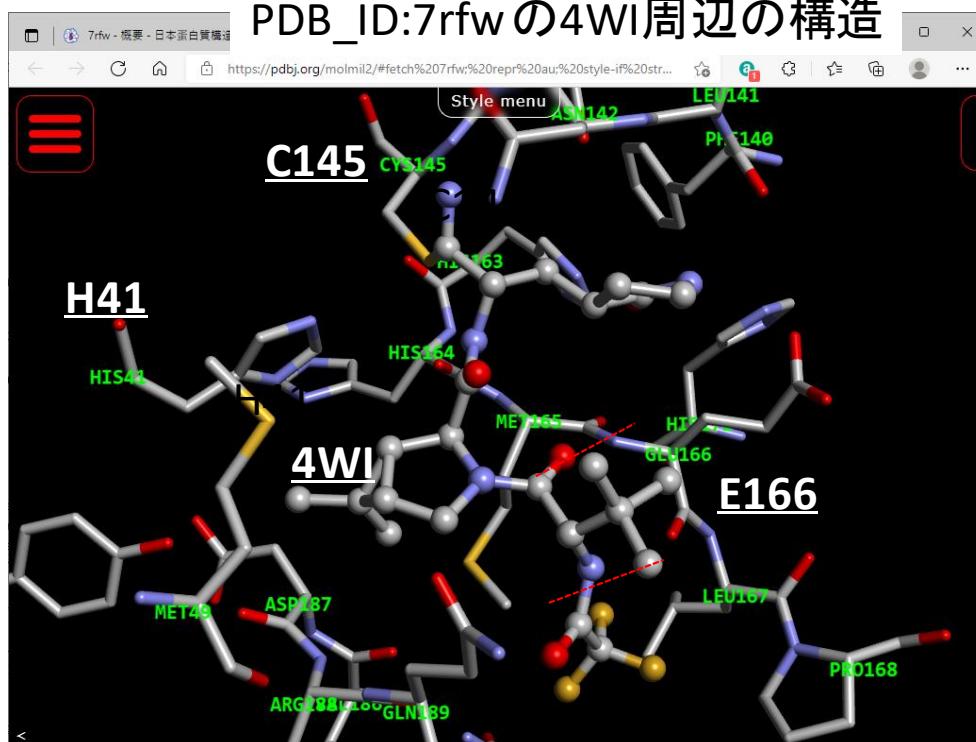
[PRO 132 Chain A]→[Display]→[Amino acid]→[stick]
[PRO 132 Chain A]→[Color]→[Atom (CPK)]
[PRO 132 Chain A]→[Label]

コマンド入力・リガンド周辺原子の表示(7rfw)

スペースを必ず入れること!

コマンドラインインターフェースに以下のコマンドを入力する

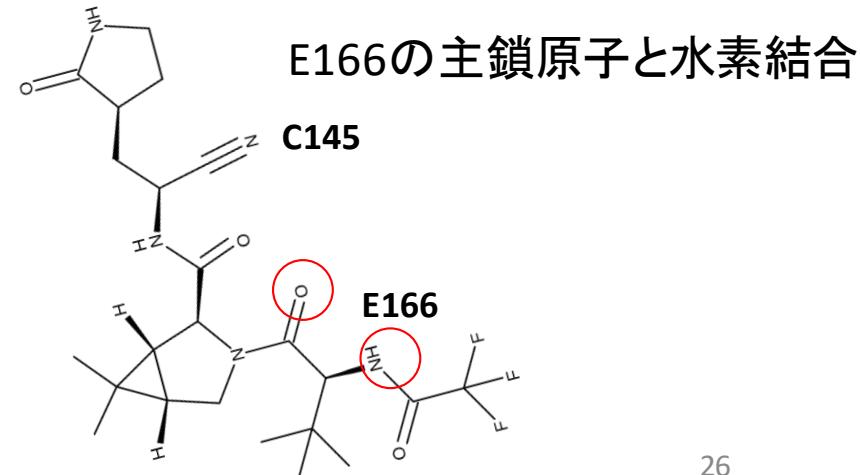
```
hide cartoon, all          #カートゥーン表示をオフ  
hide ball_stick, all       #ステイック表示をオフ  
show sticks, byres (resn 4WI around 5) #分子4WIの周辺5Åの残基を選択  
show ball_stick, resn 4WI #分子4WIを、ボール&ステイックで表示  
label byres (resn 4WI around 5) and name CA, "%s%s"%(resn, resi)  
#分子4WIの周辺5Åの残基のCα原子に残基名、残基番号のラベルを表示
```



原子をマウス左ボタンでクリックすると、その残基名・残基番号・原子名が、左下に表示

H41, C145はプロテアーゼの活性部位

C145と4WIのニトリル基が共有結合している



MolmilコマンドとPyMOLコマンドの違い

デフォルトの蛋白質がカートゥーン表示のときに以下のコマンドを実行する場合

1. “,all”が必要

- ✗ `hide cartoon` 何も起きない...
- `hide cartoon,all` カートゥーンが消えてくれる！

2. “color”と“cartoon_color”を区別

- ✗ `color red,all` 何も起きない...
- `cartoon_color red,all` カートゥーンの色が赤くなる！

Molmil コマンド早見表 : 選択コマンド

| 選択コマンドの書式 | 例 | 意味 |
|--------------------------------|--|----------------------|
| [実行], all | show sticks, all | 全ての原子を棒表示 |
| [実行], chain [鎖] | cartoon color red, chain A | A鎖のカートゥーンを赤に |
| [実行], resn [残基名] | show sticks, resn CYS | システィンを棒表示 |
| [実行], name [原子名] | show spheres, name CB | Cβ原子を球表示に |
| [実行], symbol [元素名] | show spheres, symbol S | 硫黄元素を球表示に |
| [実行], entity [エンティティ番号] | show spheres, entity 3 | エンティティ3を球表示に |
| [実行], resi [番号] | show spheres, resi 104 | 104番目を球表示に |
| [実行], resi [番号]-[番号] | show spheres, resi 104-212 | 104~212番目を球表示に |
| [実行], [条件] around [距離] | show sticks, resn ATP around 5 | ATPから5Å以下の原子を棒表示に |
| [実行], byres ([条件] around [距離]) | show sticks, byres (resn ATP around 5) | ATPから5Å以下の残基を棒表示に |
| [実行], [条件] and [条件] | show spheres, chain A and resi 104 | A鎖の104番目を球表示に |
| [実行], [条件] or [条件] | show sticks, resn SER or resn THR | セリンかスレオニンを棒表示に |
| [実行], hetatm | show spheres, hetatm | HETATMを球表示に |
| [実行], hydro | show spheres, hydro | 水分子を球表示に |
| [実行], snfg | show cartoon, snfg | SNFG表示可能な糖鎖をカートゥーンに |
| select [選択原子名], [条件] | select actsite, resi 104+212 | 104と212番目をactsiteと命名 |
| [実行], [選択原子名] | show spheres, actsite | 選択したactsiteを球表示に |

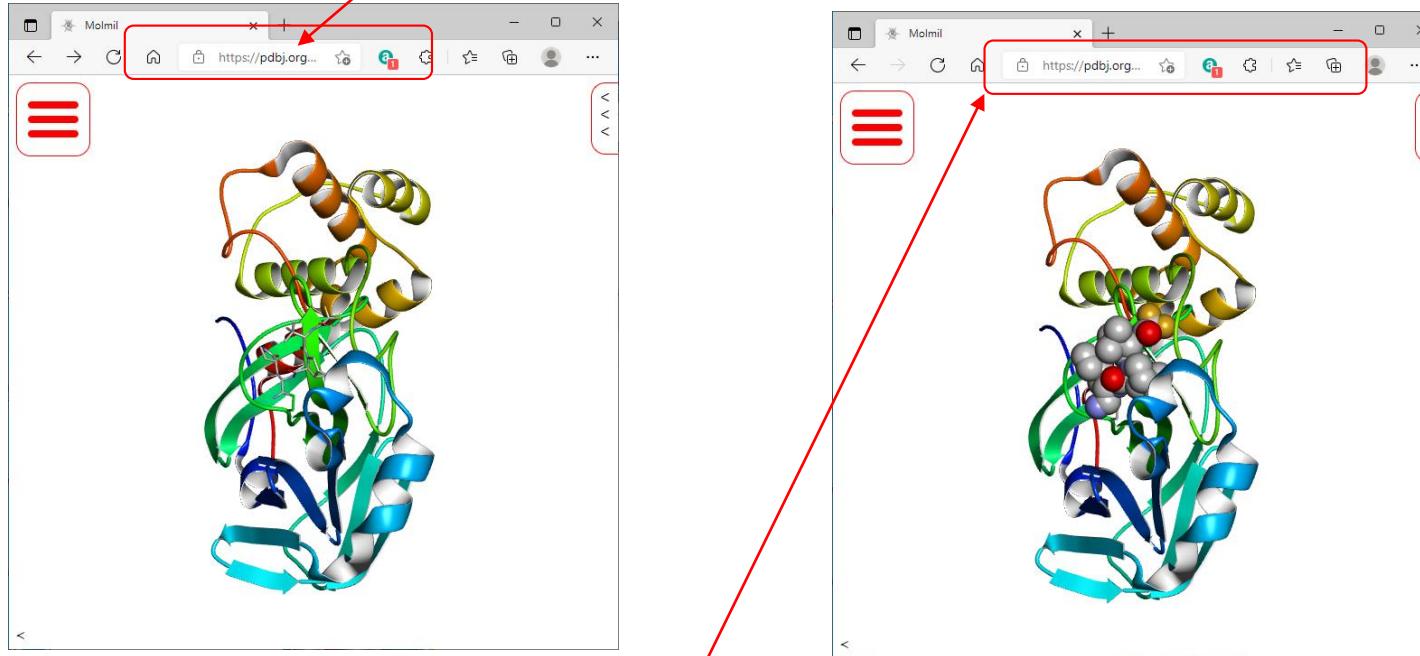
Molmil コマンド早見表 : 実行コマンド

| 実行コマンドの書式 | 例 | 意味 |
|--|-------------------------------------|---|
| <code>show [表示法]</code> | <code>show spheres, all</code> | 球モデルの非表示 |
| <code>hide [表示法]</code> | <code>hide spheres, all</code> | 球モデルの表示 |
| ※ [表示法]は、球:spheres、線:lines、スティック:sticks、ボール&スティック:ball_stick、カートゥーン:cartoon、バックボーン:ca-trace、SNFGとスティック:snfg-icon | | |
| <code>color [色]</code> | <code>color blue, all</code> | 青色にする |
| <code>cartoon_color [色]</code> | <code>cartoon_color red, all</code> | カートゥーンの色を赤にする |
| <code>color [r,g,b]</code> | <code>color [0,255,255],all</code> | シアン色[0,255,255]にする |
| <code>color cpk</code> | <code>color cpk, all</code> | 元素ごとに色分けする |
| ※次の[色]も使用できる。N末からC末へ青から赤:group、二次構造による色分け:structure、温度因子による彩色:bfactor | | |
| <code>bg_color [色]</code> | <code>bg_color white</code> | 背景を白に |
| <code>turn [xyz], [回転角(°)]</code> | <code>turn y, 180</code> | Y軸のまわりに180° 回転 |
| <code>reset</code> | <code>reset</code> | 分子を元の向きに戻す |
| <code>set cif_use_auth, [onかoff]</code> | <code>set cif_use_auth, off</code> | label_asym_id を鎖識別子に、label_seq_idを残基番号に |

WEBアドレス1行で構造が表示できる

WEBブラウザのアドレスに以下のコマンドを打つだけで構造を表示できます

```
https://pdbj.org/molmil2/#fetch 7rfw;
```



```
https://pdbj.org/molmil2/#fetch 7rfw;cartoon_color group,all;show spheres, resn 4WI;
```

#の後に、Molmilのコマンドをセミコロン(;)区切りで入力することができます

電子メールやブログなどで、簡単に指定した表示の立体構造へのリンクを示すことができる

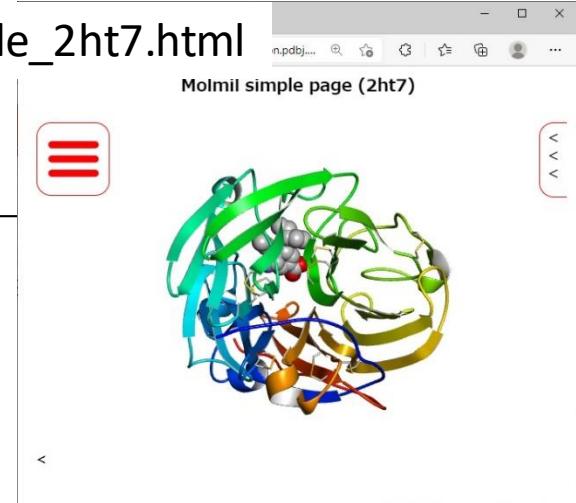
Molmilを組み込んだシンプルなWEBページ

https://numon.pdbj.org/binds_workshop_202109/molmil_simple_2ht7.html

```
<html>
  <head>
    <script src="https://pdbj.org/molmil2/molmil.js"></script>
    <script>
      var molmil_settings = {src: "/molmil2/"}, canvas;

      function initViewer() {
        canvas = document.getElementById("molmilViewer");
        molmil.autoSetup(); PDBjサーバから2ht7を取得
        molmil.loadPDB("2ht7", null, null, canvas.molmilViewer);
        canvas.commandLine.run("cartoon_color group, all;show spheres, resn G39;"); Molmilのコマンドをここに記入
      }
    </script>
  </head>

  <body onload="initViewer();">
    <center>
      <h3>Molmil simple page (2ht7)</h3>
      <span class="molmil_UI_container">
        <canvas id="molmilViewer" width="600" height="400"></canvas>
      </span>
    </center>
  </body>
</html>
```



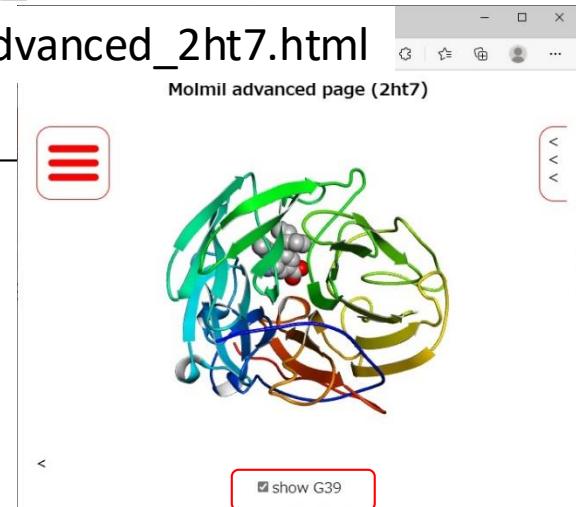
Molmilを組み込んだボタン付きWEBページ

https://numon.pdbj.org/binds_workshop_202109/molmil_advanced_2ht7.html

```
<html>
<head>
<script src="https://pdbj.org/molmil2/molmil.js"></script>
<script>
var molmil_settings = {src: "/molmil2/"}, canvas;

function initViewer() {
    canvas = document.getElementById("molmilViewer");
    molmil.autoSetup();      Webサーバに置いたPDBファイルを読み込む
    canvas.molmilViewer.loadStructure("2ht7_in_this_server.pdb", "pdb");
    canvas.commandLine.run("cartoon_color group, all; show spheres, resn G39;");
}
function show_G39(boolEnable) {
    if (boolEnable){ canvas.commandLine.run("show spheres, resn G39;"); }
        else{ canvas.commandLine.run("hide spheres, resn G39;"); }
}
</script>          G39を球表示したり、隠したりする関数
</head>

<body onload="initViewer();">
<center>
    <h3>Molmil advanced page (2ht7)</h3>
    <span class="molmil_UI_container">
        <canvas id="molmilViewer" width="600" height="400"></canvas>
    </span>
    <BR> <input type="checkbox" onclick="show_G39(this.checked)" checked>show G39
</center>          checkboxをクリックすると、関数show_G39()を呼び出すように設定
</body>
</html>
```



Molmilを活用したWEBページ

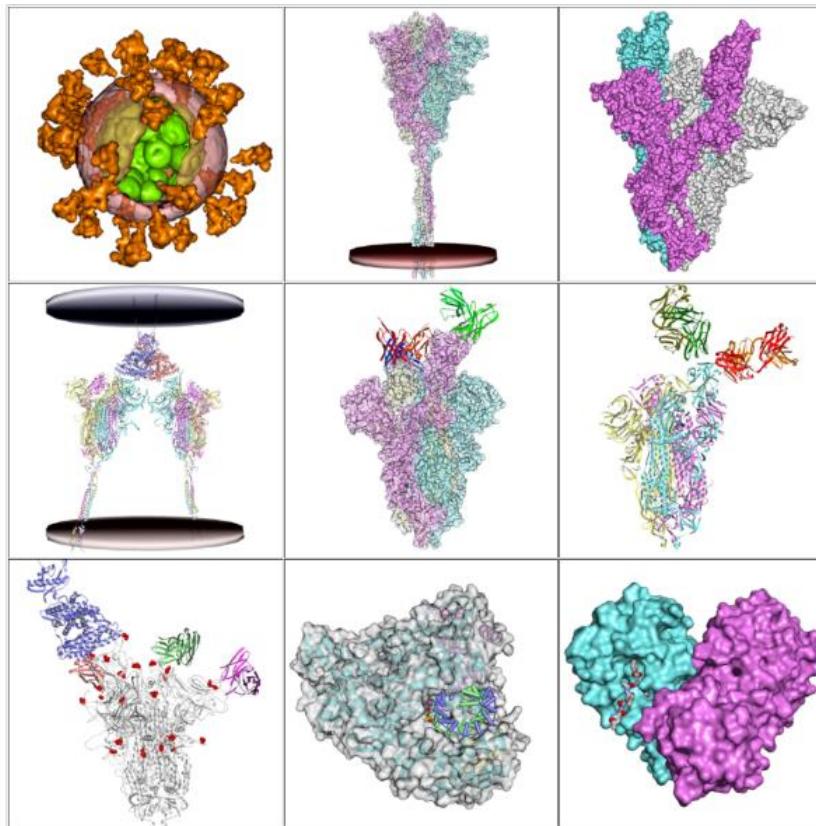
一般向けコロナウィルスの解説ページ

生命をささえるタンパク質の「かたち」 - 新型コロナウィルスの立体構造 -

<https://numon.pdbj.org/covid19>

YouTubeに解説動画もあります

<https://www.youtube.com/user/PDBjmovie>



3. アミノ酸配列からの 3Dモデリング

HOMCOS

と

AlphaFold2

3Dモデリング(立体構造予測)の二つのアプローチ

| | | |
|----------------|---|---|
| 名称 | ホモロジー・モデリング法 比較モデリング法 鋳型ベース予測法 | 非経験的予測法 Ab initio 予測法 デノボ予測法 |
| 手法の概要 | 鋳型立体構造にできるだけ似た形で、立体構造を予測 | 鋳型構造を用いずに、物理化学的な原理(分子シミュレーションの技法)に基づいて立体構造を予測 |
| 鋳型立体構造 | 必要 | 不要 |
| 一般性 | 低い | 高い |
| 計算量 | 少ない | 多い |
| 予測精度 | 似た鋳型があれば高い | 高い精度を得るには大きな計算量が必要 |
| 単体の立体構造予測 | MODELLER, SWISS-MODEL, RosettaCM, 3Dzizgaw | ROSETTA, EVfold,... |
| 蛋白質複合体予測 | MODELLER, HOMCOS | ZDOCK, HADDOCK,... |
| 低分子-タンパク質複合体予測 | MODELLER, HOMCOS, fkcombu | DOCK, AutoDock, sievgene, Glide,... |

※AlphaFold2やRoseTTAFoldは、鋳型があれば使うが、なければデノボで予測する

ホモロジー・モデリングによる3次構造予測

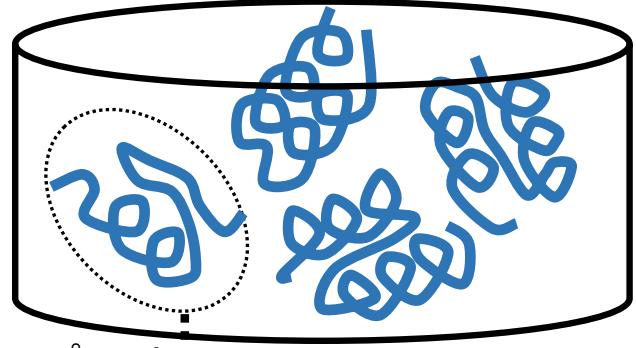
原理：立体構造はアミノ酸配列より進化的に保存する。

予測対象配列（クエリ配列）

LNVANGKSVIGPALLEEVWGSRD



立体構造データベース



テンプレート構造

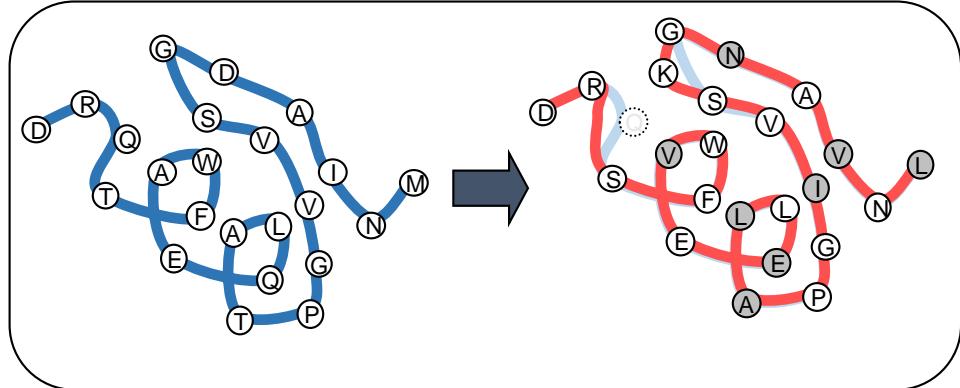
ステップ1：フォールド認識

立体構造データベースの中から、クエリ配列に最も適合する「鋳型構造（テンプレート構造）」を探す

BLAST, プロフィール法, スレディング法...



鋳型（テンプレート）構造とそのアライメント



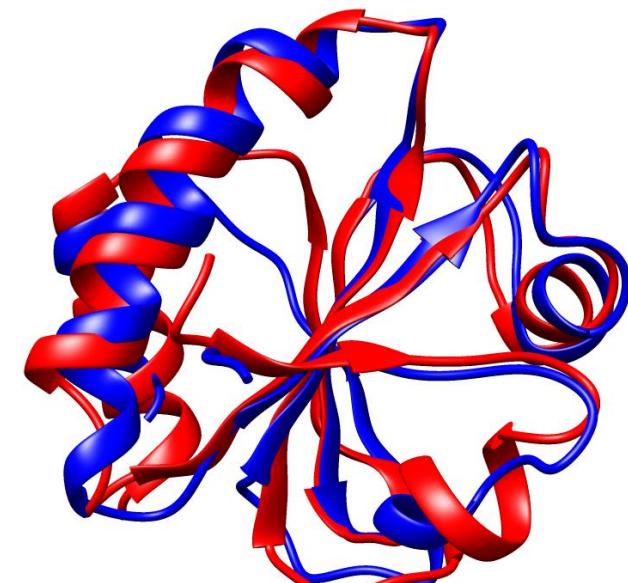
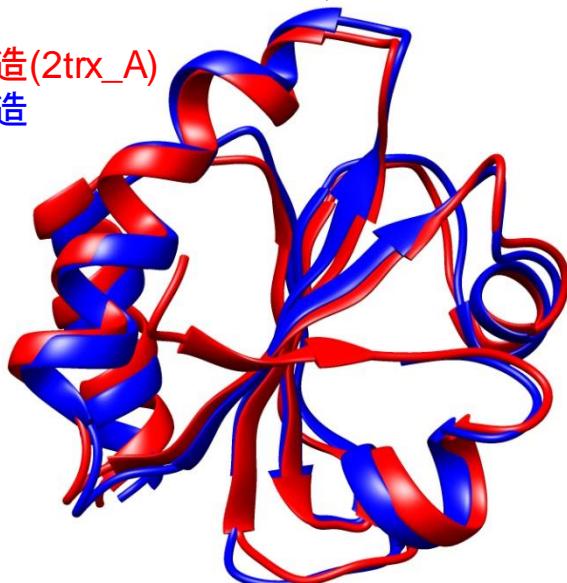
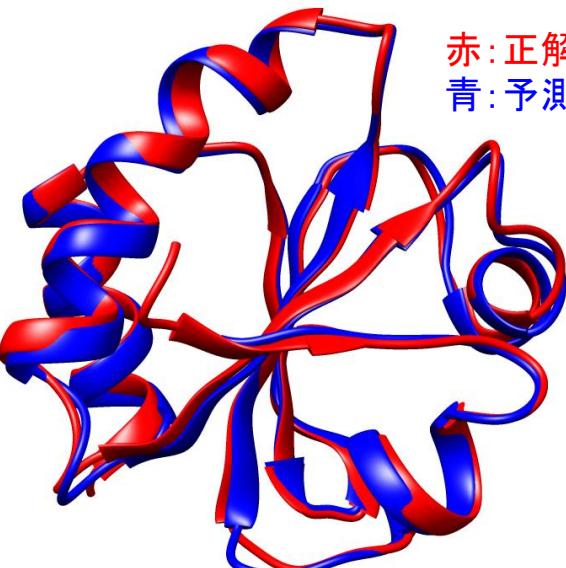
ステップ2：モデリング

鋳型（テンプレート）構造に従って全原子を構築
(1) 側鎖原子の構築
(2) 揿入ループ部を構築

MODELLER, SWISS-MODEL,
RosettaCM, 3Dzizgaw

モデル結果の例: 予測対象 2trx_A Thioredoxin (*Escherichia coli*)

アラインメント:BLAST, モデリング:MODELLER



鑄型とのSeqID=86.5%

RMSD($C\alpha$)=0.64 Å

鑄型: 2yn1_ALGPCA Thioredoxin

鑄型とのSeqID=52.5%

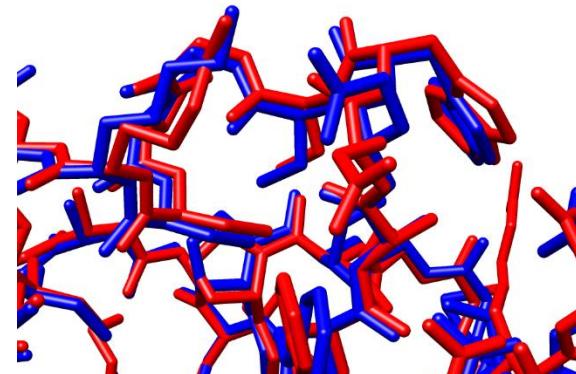
RMSD($C\alpha$)=1.08 Å

鑄型: 1fb0_A Thioredoxin M
(*Spinacia oleracea (spinach)*)

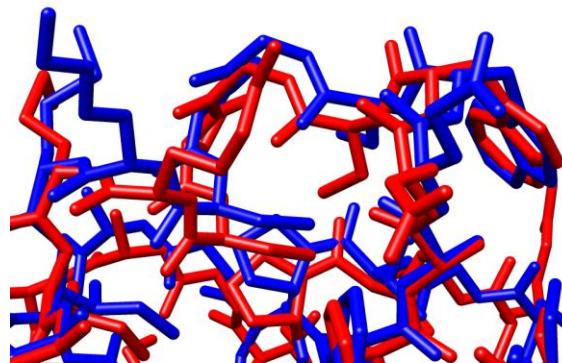
鑄型とのSeqID=25.3%

RMSD($C\alpha$)=2.30 Å

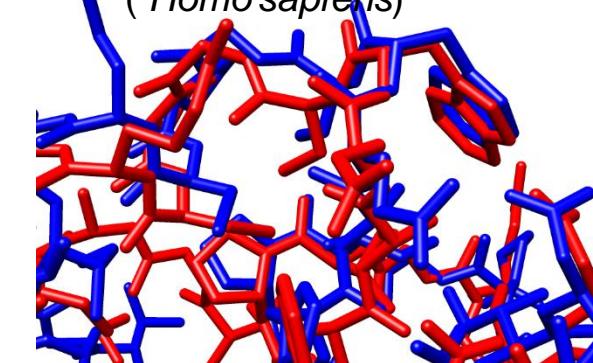
鑄型: 1x5e_A Thioredoxin
domain containing protein
(*Homo sapiens*)



RMSD(全重原子)=1.29 Å

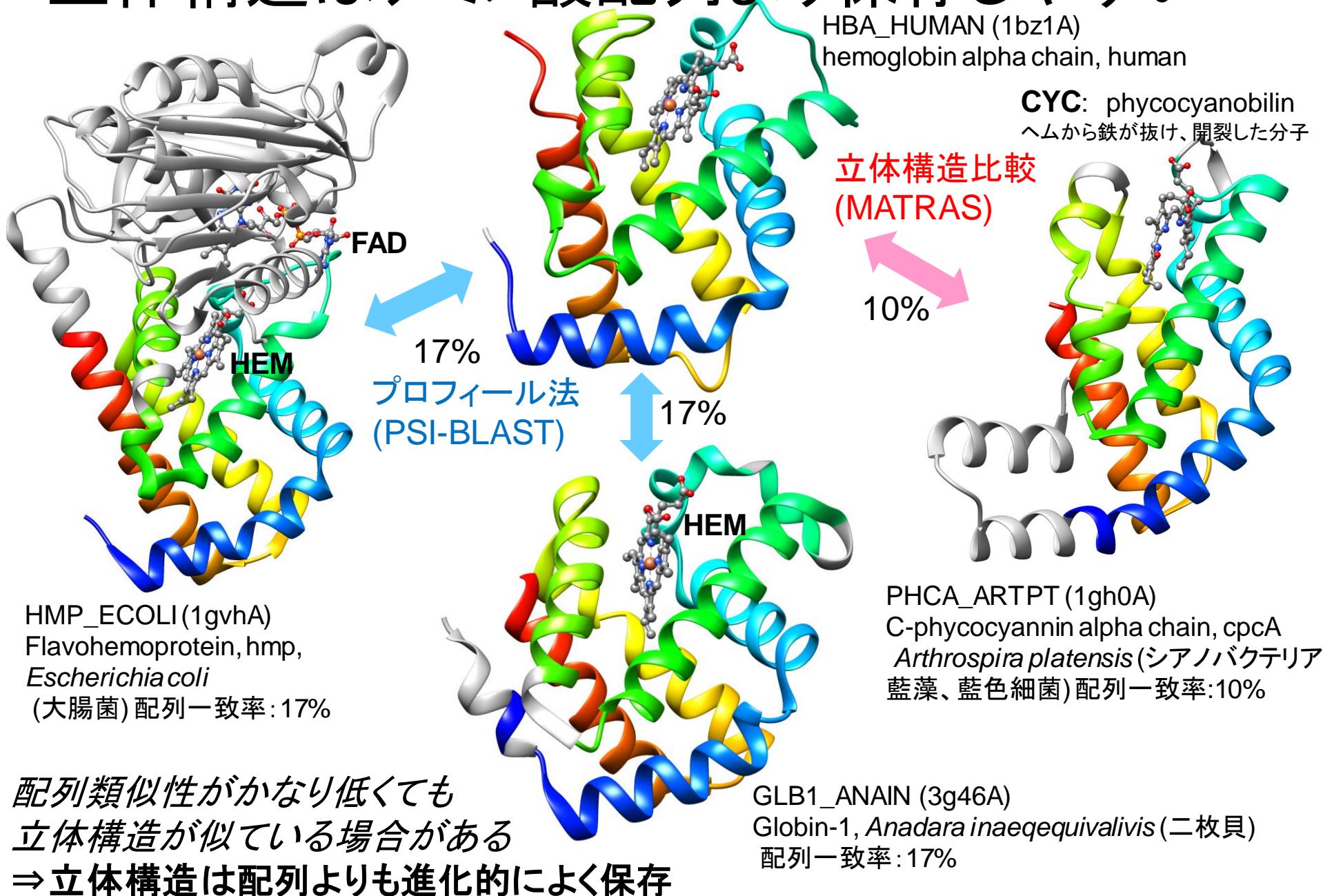


RMSD(全重原子)=1.81 Å

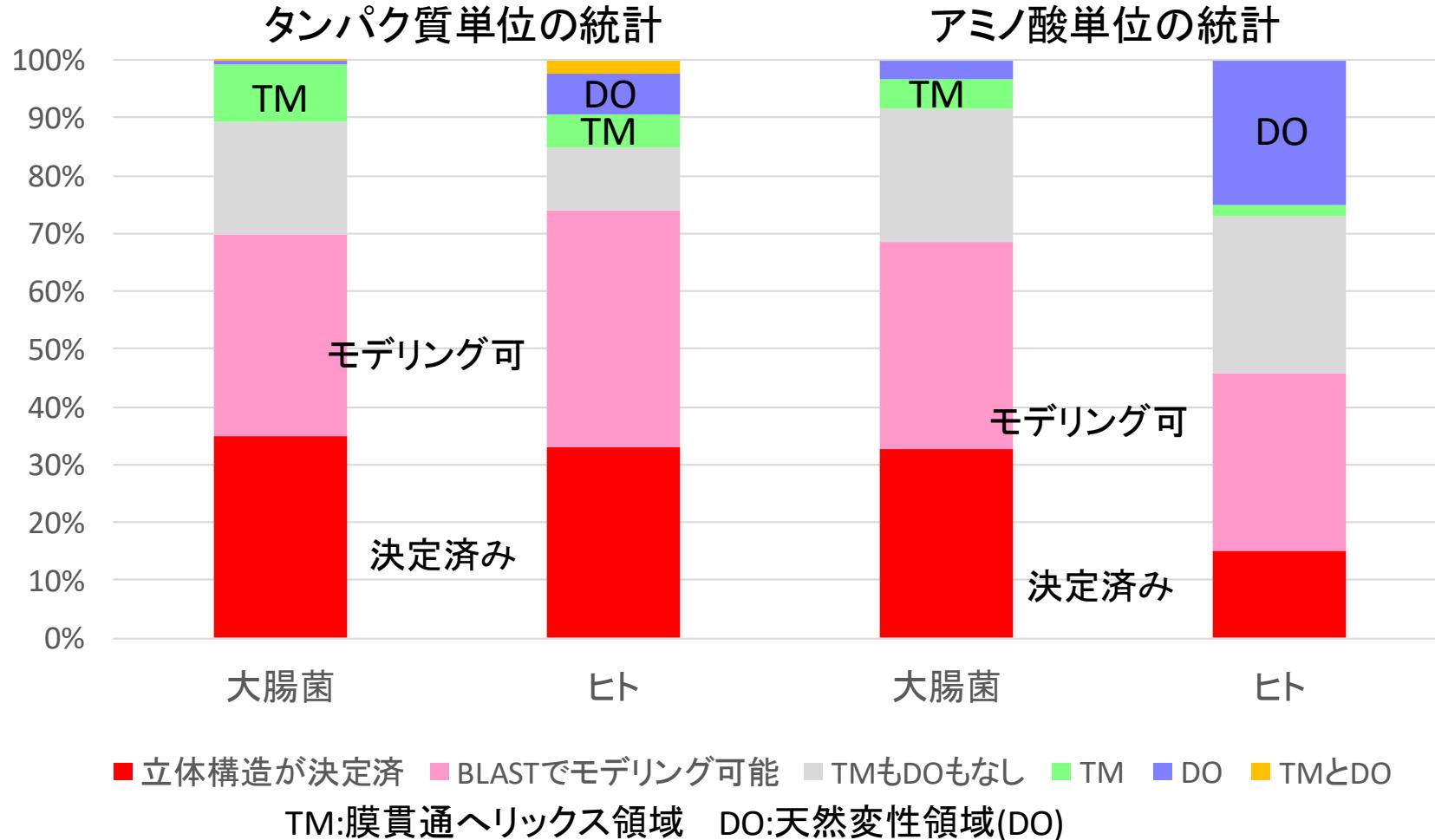


RMSD(全重原子)=2.83 Å

立体構造はアミノ酸配列より保存しやすい



立体構造が同定・予測できる割合



アミノ酸配列はUniProt 2019_07に登録されている4,519個の大腸菌のタンパク質、20,431個のヒトのタンパク質を用いた。2019/08/07のPDBを用いた。プログラムBLASTを用いて、大腸菌・ヒトのタンパク質と類似しているPDBに登録されている立体構造既知のタンパク質のアミノ酸配列を比較。赤は、配列一致率95%以上の構造既知のタンパク質が見つかった場合、すなわち、ほぼそのタンパク質そのものの立体構造が入手可能な場合の割合。ピンクは、配列一致率は95%未満だが、統計有意に類似した立体構造既知のタンパク質が見つかった場合、すなわちホモロジー・モデルリングが可能な場合の割合。残りは、現状では立体構造未知でホモロジー・モデルリングも不可能な場合の割合を示す。タンパク質単位の比較では、1部分でも立体構造とアラインメントされた場合は、そのタンパク質はモデルリング可能と見なした。膜貫通ヘリックス領域はUniProtの注釈を、天然変性領域はDISOPRED3.16による予測を用いた。

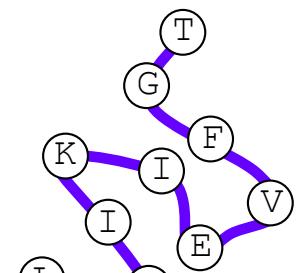
ホモロジー・モデリング法の使い方の留意点

- **鋳型構造の選択とアラインメントが予測精度をほぼ決定**
 - ・鋳型構造の選択・アライメントの不具合は、「モデリング」の過程では修正されない。
 - ・同じような鋳型構造が複数ある場合、解像度・結合リガンドなどを考慮して選択する。
※鋳型構造の検討にはHOMCOSの「タンパク質に対する結合分子の検索」が役立つ。
 - ・特に、配列類似性が低い鋳型構造を使う場合、BLASTよりも、PSI-BLAST, HMMerなどのプロフィール法のほうが、正確なアラインメントを与える。
 - ・配列モチーフなどが一致するようにアライメントの手修正が必要な場合もある。
- **鋳型構造とアラインメントされていない部分（ループ部・挿入部）の構造を決めるのは一般に困難**
 - ・特にアミノ酸長が長くなると、挿入部の信頼性は著しく低くなる。
 - ・どうしても、ループ部の構造を使う必要がある場合、複数の構造を出力させて、候補構造集団として、取り扱ったほうがよい。

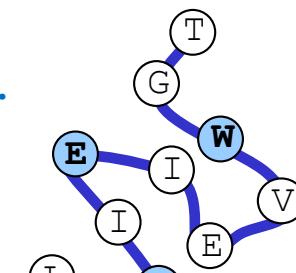
鋳型ベースのモデリング:Template-based Modeling

複合体のホモロジー・モデリング

タンパク質単量体

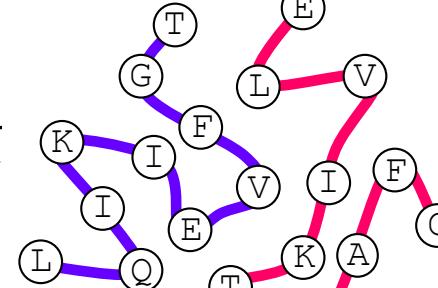


TGWVEIEINL..

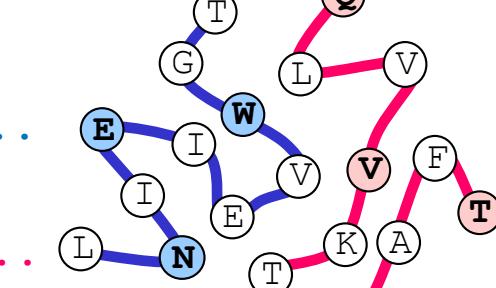


予測立体構造

タンパク質－タンパク質
複合体

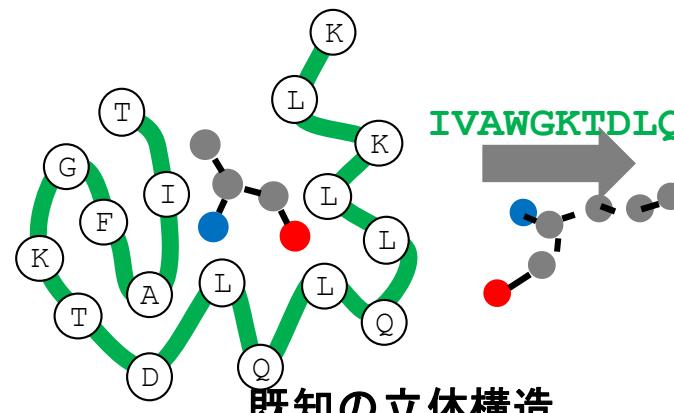


TGWVEIEINL..
QLVVKTFQFT..

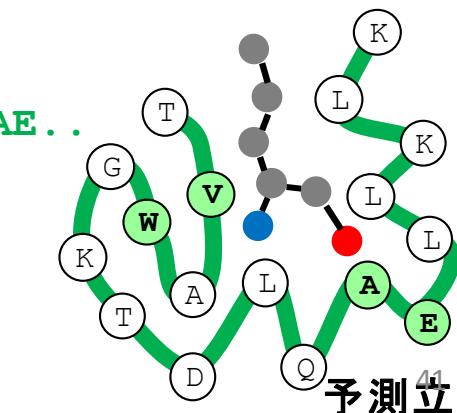


予測立体構造

化合物－タンパク質
複合体



IVAWGKTDLQAE..



予測立体構造

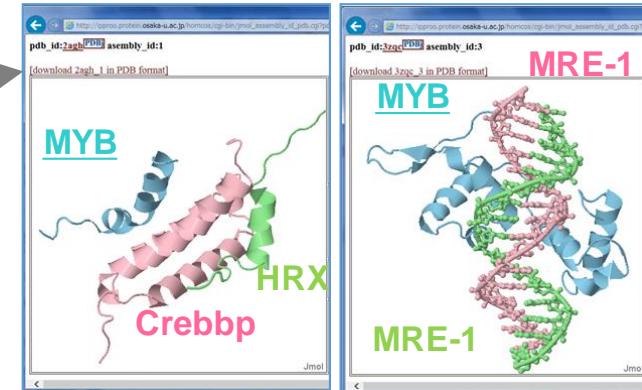
HOMCOS :複合体立体構造の検索・ホモロジーモデリングのサーバ

- PDB内の**複合体**の立体構造データを検索し、それを鋳型にモデリングする
- 配列相同性検索は**BLAST**、化学構造類似性検索は**KCOMBU**を使用

“HOMCOS”でグーグル検索 <http://homcos.pdbj.org>

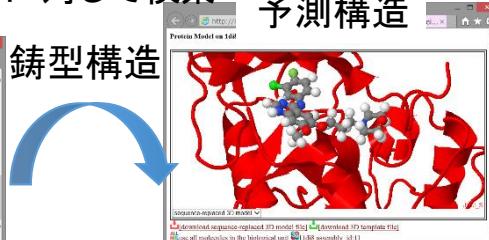
| サービス | HOMCOS | | 入力1 | 入力2 |
|--------------------|---------------|------------------|---------|---------|
| PDB内の結合分子の検索 | | タンパク質に対する結合分子の検索 | アミノ酸配列 | |
| | | 化合物に対する結合分子の検索 | 化合物構造 | |
| 複合体立体構造のホモロジーモデリング | | ホモ多量体モデル | アミノ酸配列 | |
| | | ヘテロ多量体モデル | アミノ酸配列A | アミノ酸配列B |
| | | 化合物—タンパク質複合体のモデル | アミノ酸配列 | 化合物構造 |

タンパク質に対する結合分子検索



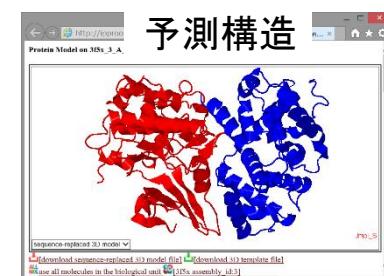
ヘテロ多量体のモデリング
アミノ酸配列2本を入力、それぞれ、PDBに対するBLASTを実行

化合物—タンパク質複合体のモデル
アミノ酸配列と化合物構造を入力、アミノ酸配列はBLASTで、
化学構造はKCOMBU でPDBに対して検索

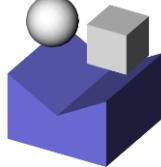


鋳型構造

予測構造



鋳型構造

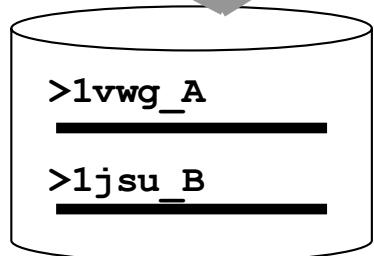


タンパク質に対する結合分子の検索

問い合わせ配列
TGWVEIEINL...

PDBに登録された
アミノ酸配列の
データベース

BLAST検索

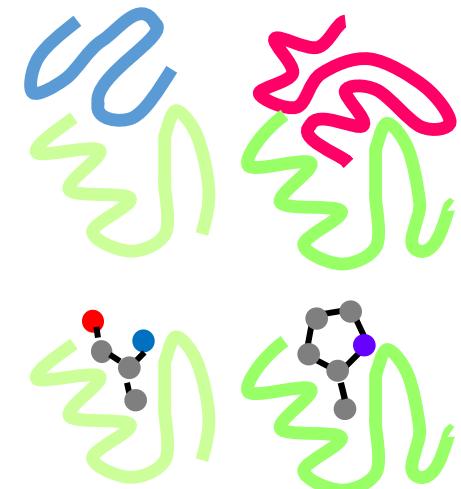


相同的なタンパク質のリスト

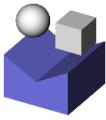
2g9xA
1w98A
1fq1B
:

1vwg_1 A1 B1
2g9x_1 A1 B1
:

コンタクトしている分子の表



問い合わせ配列とコンタクトする
分子の予測リスト



HOMCOSを用いた立体構造予測・結合分子予測

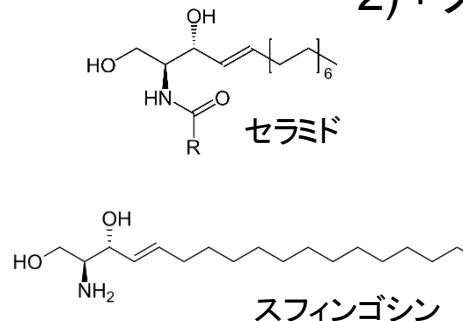
ASAHI_HUMAN (Acid ceramidase)

酸性セラミダーゼ(ASAHI)：セラミドから脂肪酸を切り出してスフィンゴシン(SPH)を生成する酵素

2) 「タンパク質に対する検索」を選ぶ

1) Googleで“HOMCOS”を検索

Google 検索結果
約 24,800 件 (0.33 秒)
<https://homcos.pdbj.org> ...
HOMCOS : 相同複合体の検索・モデリングサーバ
 HOMCOS(HOMology modeling of COnplex Structure)は、PDBに収納されている構造データを利用して、分子の類似性・相同性から、構造未知の分子ペアの構造データを予測することができます。タンパク質の配列類似性検索にはBLASTを、化合物の構造類似性検索にはCOMBUIを用いています。



【Go to English page】 ? ヘルプページ

HOMCOS(HOMology modeling of COnplex Structure)は、PDBに収納されている複合体の立体構造データを利用して、分子の類似性・相同性から、構造未知の分子ペアの構造データを予測することができます。タンパク質の配列類似性検索にはBLASTを、化合物の構造類似性検索にはCOMBUIを用いています。

| サービス | クエリ / 質問 | 説明 |
|---------------|--|---|
| 結合分子の検索 | <input checked="" type="checkbox"/> タンパク質に対する結合分子の検索 | 1つのアミノ酸配列をクエリとして、それと類似したタンパク質と結合している分子を検索します。 |
| | <input type="checkbox"/> 化合物に対する結合タンパク質の検索 | 1つの化合物構造をクエリとして、それと類似した化合物と結合しているタンパク質を検索します。 |
| 複合体立体構造のモデリング | <input type="checkbox"/> ホモ多量体のモデル | 1本のアミノ酸配列をクエリとして、そのホモ多量体の立体構造をホモモーテーリングによって予測します。 |
| | <input type="checkbox"/> ヘテロ多量体のモデル | 2本のアミノ酸配列をクエリとして、そのヘテロ多量体の立体構造をホモモーテーリングによって予測します。 |
| | <input type="checkbox"/> 化合物-タンパク質複合体のモデル | 1本のアミノ酸配列と1つまたは複数の化合物構造をクエリとして、その複合体の立体構造をホモモーテーリングによって予測します。 |

(Download HOMCOS_data): Release of PDB:20211208, Release of UniProt:2021_04 Release of KEGG:2021-12-06 3D of KEGG_COMPOUND_and_DRUG

新型コロナウイルスSARS-CoV-2に関するタンパク質と化合物のページを作りました!(PDB_UPDATE:20211208)

【Go to English page】 ? ヘルプページ

入力したタンパク質と類似したタンパク質群と複合体を形成している他の分子のリストを作成します。相互作用する可能性のある分子とその結合部位を予測します。

アミノ酸配列は、タンパク質配列データベースのID、1文字表記のアミノ酸配列、[PDB_ID]+[CHAIN_ID]、PDBファイルのアップロードのどれかの方法で入力してください。

| | |
|---|---|
| タンパク質配列のID (UniProt:ID/AC, INSDC/RefSeq:protein_id) | <input type="text" value="ASAHI_HUMAN"/> (CDK3_HUMAN, Q00526, AAV40830.1, NP_001249.1) |
| アミノ酸配列 (一文字表記) | <input type="text"/> |
| PDB | PDB_ID: <input type="text"/> (1fin, 4hhb,...) CHAIN_ID: <input type="text"/> (A, B, ...) |
| PDBファイルのアップロード | <input type="button" value="ファイルの選択"/> <input type="button" value="ファイルが選...されていません"/> |
| <input type="button" value="Reset values"/> <input style="border: 2px solid red; border-radius: 5px; padding: 2px 10px;" type="button" value="SEARCH"/> | |

3) UniProtIDのフォームに
“ASAHI_HUMAN”と入力して、
 [SEARCH]をクリックする。

ABHD5_HUMAN, ADRO_HUMAN, AMACR_HUMAN, CATD_HUMAN,
 CBPA6_HUMANなどでもよい

問い合わせ蛋白質の配列は以下の4通りで入力可

- (i) UniProt ID
- (ii) アミノ酸配列
- (iii) PDB_ID+鎖
- (iv) PDBファイルのアップロード

「タンパク質に対する検索」結果(ASAHI_HUMAN)

- ・単量体、複合体構造は、デフォルトでは代表構造だけがバー表示されている。アライメント領域・相互作用部位によって代表を決めている。
 - ・全ての相同な立体構造を表示する場合は、**[Full Bars]**  をクリックする。

- ・相同性のしきい値は、デフォルトではE-value<0.001だけで、同一残基率は0%に設定。よりしきい値を上げれば(30%,40%,...,95%)、候補構造は減るが、予測の信頼性は向上。

左側の画面は、ホモジンアセチル化酵素A (ASA1) のプロトコルスクリーンショットです。右側の画面は、ASA1の単量体構造と複合体構造の表示例です。

左側の画面 (ホモジンアセチル化酵素A)

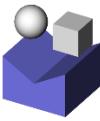
- Summary Bars [0.0 %]**: バー図で、各アミノ酸の構成割合を示す。各バーの下に%値が記載されている。
- seq_id(%): [0] [30] [40] [50] [60] [70] [80] [90] [95] [100]**: セレクション範囲を指定するためのボタン。
- [show] [download] [help]**: ボタン。

右側の画面 (ASA1の構造)

ASA1の構造は、二つの表示モードがあります。

- ASA1の単量体構造 (赤色)**: ASA1の単量体構造を表示。構造は主に赤色で示されています。
- ASA1と他の蛋白質との複合体構造 (緑色)**: ASA1と他の蛋白質との複合体構造を表示。構造は主に緑色で示されています。

これらの構造は、ASA1のアミノ酸構成割合 (Summary Bars) によって色分けされています。



単量体立体構造の表示(ASAHI_HUMAN)

Sequence-replaced 3D model(簡易ホモジーモデル構造) :

鋳型構造と座標は同じ。残基名と残基番号を標的配列に入れ替えてある。側鎖原子や挿入残基は正しくモデリングされていない。

簡易ホモジーモデル構造の ダウンロード

鋳型構造の ダウンロード

MODELLER をクリックすると
Modellerの入力ファイルを
ダウンロード可能。
Modellerをインストールす
れば、全原子のモデルが構
築可能

Modeller scripts for each subunit

Subunit 1

1. Download modeller script: [model1.py](#)
 2. Download alignment: [alignment1.ali](#)
 3. Download template PDB file: [5u81_A_1.pdb](#)
 4. execute command:

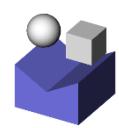
mod9 model11.py

ALL
MOL

**ALL
MOL** をクリックすると、
生物学的単位に含まれる
全ての分子が表示される

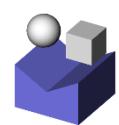
クエリ(ASAH1_HUMAN)と 鋳型(5u81_A_1)との アラインメント

デフォルトではHTML5(Jsmol)による分子の表示。



「タンパク質に対する検索」結果(ASAHI_HUMAN)

3Dのアイコンをクリックすると立体構造モデルが表示



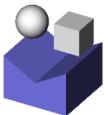
化合物複合体立体構造の表示(ASAHI_HUMAN)

複合体のSequence-replaced model、化合物付きの鋳型構造のダウンロード、Modellerの入力ファイルのダウンロードも同様に可能

**ALL
MOL** 生物学的
単位内の全分
子を表示

予測接触残基 標的配列 (ASAHI_HUMAN) の残基番号、残基 名。カッコ内は鋳型 のアミノ酸

接触残基が'b'の文字で示されている。



Site Table

Contact Bar(summary)[0.0 %]

seq_id(%): [0] [30] [40] [50] [60] [70] [80] [90] [95] [100]

[bars:full] [SiteTable]

[\[Back to Search Page\]](#)

コンタクトバー表示の画面上のこのアイコンをクリック

Variant

Sites by Variants [help]

Site Table[0.0 %]

seq_id(%): [0] [30] [40] [50] [60] [70] [80] [90] [95] [100]

[Summary Bars] [Full Bars]

[Back to Search Page] [HOMCOS] [Back to HOMCOS]

tab tsv split [show] [download] [help]

SITEをクリックすると特定のサイトのまとめのページが表示される

二次構造

(**H**: α ヘリックス、**E**: β シート)

[BLA]

[e]

[contact_mols]:predicted binding molecules

[variant]:UniProt Human Variant.

溶媒露出度(%)

相同配列群のアミノ酸頻度。頻度順にソート。出現したアミノ酸だけ表示

UniProtのアノテーション
(Feature Table)

| | a | s | e | acc | pdb | contact_mols |
|--------|---|---|------|---------|-----|---|
| SITE 1 | M | E | 25.5 | Q4fx3_A | | compound HHT precipitant |
| SITE 2 | D | T | 76.4 | Q4fx3_A | | hetero CCNA2 HUMAN CG2A HUMAN compound HHT homo precipitant |
| SITE 3 | M | T | 35.8 | Q4fx3_A | | hetero CCNA2 HUMAN compound HHT precipitant |
| SITE 4 | F | E | 15.3 | Q4fx3_A | | compound HHT precipitant |
| SITE 5 | Q | E | 10.7 | Q4fx3_A | | precipitant |
| SITE 6 | K | E | 45.8 | Q4fx3_A | | compound HHT |

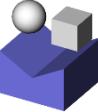
結合分子のサマリー

| | observed aa | feature table | variant |
|--|--------------------|---|---------|
| | MLIVQ | | |
| | DENKGOST | | |
| | RDKNQELTVICMSFHAG | | |
| | YFWL | DOMAIN Protein kinase. DOMAIN Protein kinase. | |
| | EQKDRTVINHLACMSY | DOMAIN Protein kinase. DOMAIN Protein kinase. | |
| | IKVLRNDEFGMQACHPST | DOMAIN Protein kinase. DOMAIN Protein kinase. | |

(1) 埋もれている部位(溶媒露出度accが小さい部位)に変異が入ると、天然構造が不安定になり、機能を失活しやすい。

(2) 相同タンパク質群で観察されるアミノ酸の割合(observed aa)が大きい(よく観察される)アミノ酸に変異した場合、機能への影響は小さい。逆に、稀にしか観察されないアミノ酸に変異した場合は、機能を失いやすい。SIFT scoreなど多くのプログラムがこの原理に基づく。

ASAHI1_HUMANのSite_Table



1 M E e
2 E H b
3 F H b

Site Table[0.0 %]



HOMCOS
[Back to Search Page]

Variant
[Sites by Variants] [help]

seq_id(%): [0] [30] [40] [50] [60] [70] [80] [90] [95] [100]

tab
tsv
split

[show] [download] [help]

| PID | QueryLength | Homologous Sequence in PDB | UniProt Query | TITLE |
|----------|-------------|----------------------------|----------------------|--|
| 3946607 | 395 | 33 | Q13510(ASAHI1_HUMAN) | RecName: Full=Acid ceramidase ; Short=AC; Short=ACDase ; Short=Acid CDase; EC=3.5.1.23 ;AltName: Full=Acylsphingosine deacylase;AltName: Full=N-acylethanolamine hydrolase ASAHI1 ; EC=3.5.1.- ;AltName: Full=N-acylsphingosine amidohydrolase;AltName: Full=Putative 32 kDa heart protein; Short=PHP32;Contains: RecName: Full=Acid ceramidase subunit alpha ;Contains: RecName: Full=Acid ceramidase subunit beta ;Flags: Precursor; |
| QUERYSEQ | | | | MPPRSRQVALVLLAAAVSGCAVAOHAPWITEDKPSGTYRGAIVYTTNLDLPPVKRWHMLDKAPVVKTVTNSLNKINTFVPSGKIMOVVDEKLPGGLGNFPGPFEEMKGIAAVTDIPLGETISFNIFYELFTICTS1VAED KGKHLTHGRNMDFGFLGLINNDTHVITTEQLKIPVTLNLDFQRNNKTKVEKASSFAFGVQMLTGFKPQFSLTLNERFSINGGYGLILEWILGKQVMWIGFLTRTVLLENSTYEAKNLLTKTKLAPAYFILGGNGEGOGVITDRKE SLDVYELDAKQGRIVVWTQNYDRWIKHPFFLDRTPAKMCLNRTSQENISFETMYDVL8TLPVNLNTVYTTLIDVTKQGFETYLRCDCPDC1GM |

[BLAST file for PDB] (plain) (bar) (multiple alignment) [BLAST for UniProt: (plain) (bar) (multiple alignment) (PSSM file)]

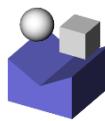
[n]:site number of query sequence. [a]:amino acid of query sequence. [s]:predicted secondary structure.
[e]:predicted exposed/buried. [acc]:predicted relative accessibility(%). [pdb]:PDB code of homologous structure.
[contact_mols]:predicted binding molecules [observed aa]:Observed amino acids among homologous sequences. [feature table]:UniProt Feature Table
[variant]:UniProt Human Variant.

| n | a | s | e | acc | pdb | contact_mols | observed aa | feature table | variant |
|----------|---|---|------|--------|--------------|---|-------------------|---|--|
| SITE 1 | M | - | - | - | - | | M | SIGNAL DISORDER predicted by DISOPRED SIGNAL | |
| SITE 2 | P | - | - | - | - | | LPR | SIGNAL DISORDER predicted by DISOPRED SIGNAL | |
| SITE 3 | G | - | - | - | - | | G | SIGNAL DISORDER predicted by DISOPRED SIGNAL | |
| SITE 138 | E | T | b | 0.0 | 5u81_A | hetero Q9GUI1_CAEEL A8K0B6_HUMAN ASAHI1_HUMAN compound I3C metal IOD | ED | | E->V:(0.0 %):LP/P Farber lipogranulomatosis (FRBRL) dbSNP:rs137853594 [MIM:228000] |
| SITE 139 | L | T | b | 13.5 | 5u81_A | hetero A8K0B6_HUMAN ASAHI1_HUMAN compound I3C | ILAMVFSY | | |
| SITE 140 | F | S | e | 20.1 | 5u81_A | hetero A8K0B6_HUMAN ASAHI1_HUMAN precipitant | SFTA | | |
| SITE 141 | T | E | b | 1.3 | 5u81_A | | ATRV | MUTAGEN /note="T->A: Decreased rate of autocatalytic processing." Decreased rate of autocatalytic processing." | |
| SITE 142 | I | E | b | 13.0 | 5u81_A | metal CL | FATMV | | |
| SITE 143 | C | E | b | 0.0 | 5u81_A | compound WTF HJA JRY precipitant | C | ACT SITE /note="Nucleophile" ECO:0000269 PubMed:30525581" autocatalytic processing. Loss of ceramidase activity." ACT SITE /note="Ceramidase activity" ECO:0000269 PubMed:30525581" MUTAGEN /note="C->ceramidase activity." | |
| SITE 144 | T | E | b | 0.0 | 5u81_A | | I | | |
| SITE 145 | S | E | b | 0.0 | 5u81_A | | S | | |
| SITE 146 | I | E | b | 0.0 | 5u81_A | | IVA | | |
| SITE 147 | V | E | b | 0.0 | 5u81_A | | VI | | |
| SITE 148 | A | E | b | 0.0 | 5u81_A | | ATI | | |
| SITE 149 | E | E | b | 2.5 | 5u81_A | metal CL | QED | | |
| SITE 150 | D | b | 15.4 | 5u81_A | compound I3C | DNAEFGIQLPQRSTV | | | |
| SITE 151 | K | T | e | 61.3 | 5u81_A | precipitant | SKNEGADFILPQRTV | | |
| SITE 152 | K | T | e | 88.7 | 5u81_A | precipitant | KRAQDQEFGILNPSTV | | |
| SITE 153 | G | b | 14.3 | 5u81_A | | | GAFISWFH1KMNPORTY | | |

138番目のサイトのまとめページ

SITE Summary for the 138-th Site(E)

| | | | |
|--|-------------|-----------|--|
| PID | QueryLength | FocusSite | TITLE |
| 3946607 | 395 | 138 E | RecName: Full=Acid ceramidase ; Short=AC; Short=ACDase ; Short=Acid CDase; EC=3.5.1.23 ;AltName: Full=Acylsphingosine deacylase;AltName: Full=N-acylethanolamine hydrolase ASAHI1 ; EC=3.5.1.- ;AltName: Full=N-acylsphingosine amidohydrolase;AltName: Full=Putative 32 kDa heart protein; Short=PHP32;Contains: RecName: Full=Acid ceramidase subunit alpha ;Contains: RecName: Full=Acid ceramidase subunit beta ;Flags: Precursor; |
| UniProt Information | | | |
| AC/ID | | | |
| AC:Q13510 ID:ASAHI1_HUMAN | | | |
| VARIANT: /note="E -> V (in FRBRL: loss of ceramidase activity; dbSNP:rs137853594)" | | | |
| Feature Table for 138-th site | | | |
| ECO:0000269 PubMed:10993217" /id="VAR_021582" | | | |
| HELIX: | | | |
| CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_0000002312" | | | |
| VARIANT for 138-th site | | | |
| E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL)" [MIM:228000] | | | |
| Evolutionary Information | | | |
| Percentages of Amino Acids in Homologous Proteins [show alignment] | | | |
| E989 0:10X | | | |
| 3D Structure Information | | | |
| Template For Monomer predicted SecStr predicted ExpBur Predicted Relative Acc(%) | | | |
| 5u81 T (Hbond turn) b (buried) 0.0 | | | |
| 3D Complex Information | | | |
| Predicted Bind Molecules | | | |
| hetero:9 compound:1 metal:1 | | | |
| Templates for 3D complexes | | | |
| hetero [52802:ABK0B6_HUMAN] 5u7z_A_1_B_1 5u7z_C_1_D_1 5u7z_C_2_D_2 6mhm_A_1_B_1 6mhm_C_1_D_1 | | | |
| [49694:Q9GUI1_CAEEL] 5dy3_A_1_B_1 5dy3_C_1_D_1 5dy3_E_1_F_1 5dy3_G_1_H_1 compound [I3C] 5u84_A_1_L_1 | | | |
| metal [IOD] 5u84_B_1_VA_1 | | | |



138番目の部位のまとめ(ASAHL H)

https://homcos.pdbj.org/cgi-bin/prot_sch_eachsite.cgi?con=T&pssm=T&precalc_blt=T&q...

SITE Summary for the 138-th Site(E)

| PID | QueryLength | FocusSite | TITLE |
|---------|-------------|-----------|---|
| 3946607 | 395 | 138 E | RecName: Full=Acid ceramidase ; Short=AC; Short=ACDase ; Short=Acid CDase ; AltName: Full=Acylsphingosine deacylase; AltName: Full=N-acylethanolamine hydrolase; AltName: EC=3.5.1.- ; AltName: Full=N-acylsphingosine amidohydrolase; AltName: Full=Putative Ceramide Glucosidase; Contains DocName: Full=Acid ceramidase subunit alpha |

[変異体のサマリーの表示\(ASAHL_HUMAN\)](#)

Focused Site

query sp | Q13510 | ASAHL
sp | A5A6P2 | ASAHL
sp | Q60H4 | ASAHL
sp | ADA383ZFX3 | ASAHL
sp | Q17QB3 | ASAHL
sp | AOAOP6JG37 | ASAHL
sp | Q6P7S1 | ASAHL
sp | Q9WV54 | ASAHL
sp | Q45686 | ASAHL CAEEL [1]
sp | Q09551 | ASAHL CAEEL [1]
sp | Q907V9 | NAAA_MOUSE [1]
sp | Q5KTC7 | NAAA RAT [1]
sp | Q9GU11 | NAAA CAEEL [1]
sp | HOVCJ6 | NAAA CAVP0 [1]
sp | G1T7U7 | NAAA RABIT [1]
sp | Q02083 | NAAA HUMAN [1]
sp | Q5UR76 | YR632_MIMIV [1]

130:YNIFYELFTYC: 140
137:LNLFYEMSRFC: 147
121:VNLAYEASAFC: 131
121:VNLAYEASAFC: 131
103:MNLYDVAAFC: 122
113:VNLAYEASAFC: 123
117:VNLEYYSAF: 127
116:VNLEYESSVFC: 126
86:LQLVYEITAAD: 96

変異体の情報

: E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL)" [MIM:228000]

UniProt Information

AC/ID: Q13510 ID: ASAHL_HUMAN

VARIANT: /note="E -> V (in FRBRL; loss of ceramidase activity; dbSNP:rs137853594)"
ECO:0000269|PubMed:10993717|/id="VAR_021582"
HELIX:
CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_0000002312"

Feature Table for 138-th site

VARIANT for 138-th site

E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL)" [MIM:228000]

Evolutionary Information 相同配列群のアミノ酸頻度

Percentages of Amino Acids in Homologous Proteins [show alignment]

E:90% D:10%

3D Structure Information

Template For Monomer: predicted SecStr: predicted ExpBur: Predicted
5u81 T (Hbond turn) b (buried) 0.0

3D Complex Information

Predicted Bind Molecules
hetero:9 compound:1 metal:1

Templates for 3D complexes

hetero [52802:A8K0B6_HUMAN] 5u7z A 1 B 1 5u7z C 1 D 1 5u7z C 2 D 2 6mhm A 1 B 1 6mhm C 1 D 1
[49694:Q9GUI1_CAEEL] 6dy3 A 1 B 1 6dy3 C 1 D 1 6dy3 E 1 F 1 6dy3 G 1 H 1 compound [I3C] 5u84 A 1 L 1
metal [IOD] 5u84 B 1 VA 1

これらをクリックすると、この結合サイトとする複合体。

https://homcos.pdbj.org/cgi-bin/model3D.cgi?PID=3946607&pdb_id=5u84&easy...

Protein Model on 5u84_A_1_L_1 focusing on 138-th site (TEMPLATE: PDB:5u84)

DOWNLOAD:
sequence-replaced 3D model [for Pymol]
3D template
Modeller script

ALL MOLECULES IN THE BIOLOGICAL UNIT:
assembly_id:1

Window size :
[small] [medium] [large]

Sequence-replaced 3D model CurrentView:HTML5 Change to:[JAVA]

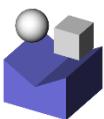
FOCUSSED SITES template(5u84_A_auth_asym_id:A)
SiteNum AA label auth_seq_id seq_id AA Sec_Ex/Bur RelAcc
138 E 127 138 E G e 20.6

MOLECULES

| model | mark | query | asym_id oper (auth_asym_id) | type | description |
|-------|------|-----------------|-----------------------------|----------------------------------|--|
| 1 | ☒ | a (ASAHL_HUMAN) | A 1 (A) | polymer(polypeptide(L)) [352 aa] | Acid ceramidase :ASAHL_BALAS |
| 2 | ☒ | b | L 1 (A) | non-polymer(I3C) | 5-amino-2,4,6-triiodobenzene-1,3-dicarboxylic acid |

contact sites
a(query A)
134N 135I 138E 139LF
223L 252L(I) 275I 276L
277A (identity: 77.8 %/83.1 %)

51



変異体のサマリーの表示(ASAHI_HUMAN)



[\[Summary Bars\]](#)



[\[Full Bars\]](#)

| | | | |
|---|---|---|---|
| 1 | M | E | e |
| 2 | E | H | b |
| 3 | F | H | b |

[Site Table \[0.0 %\]](#)



[\[Back to Search Page\]](#)

Variant

| | | | |
|---|---|---|---|
| 1 | M | E | e |
| 2 | E | H | b |



[seq_id\(%\): \[0\] \[30\] \[40\] \[50\] \[60\] \[70\] \[80\] \[90\] \[95\] \[100\]](#)

tab
tsv
split

[\[show\]](#) [\[download\]](#) [\[help\]](#)

SiteTable表示の画面上のこのアイコンをクリック

Statistics of sites in view of Disease classification

UniProtに記載されていた変異体のサマリーが表示される

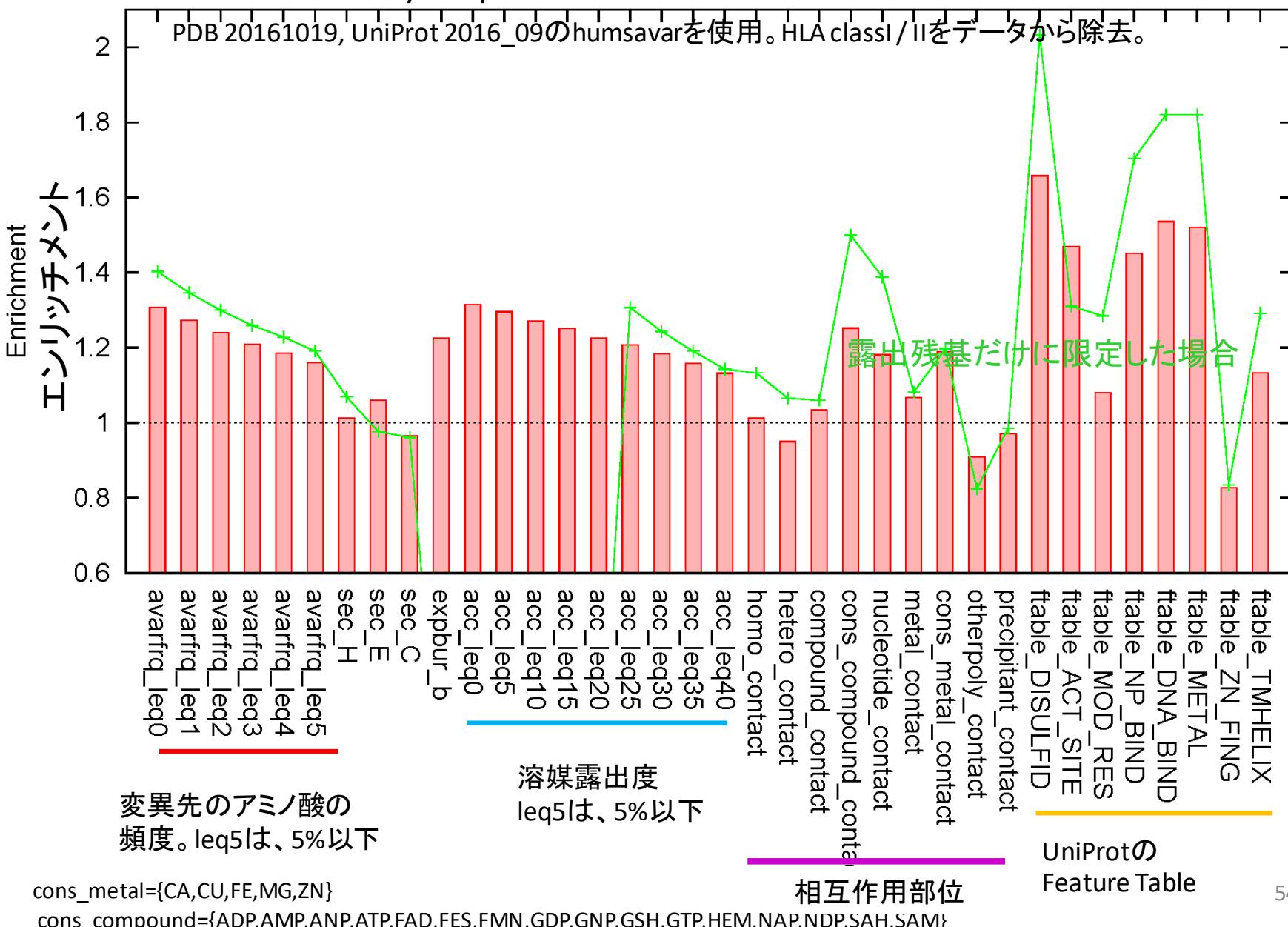
| | | | |
|------------------------|---|--|--------------------------------|
| All | 159950 Spinal muscular atrophy with progressive myoclonic epilepsy (SMAPME) | 228000 Farber lipogranulomatosis (FRBRL) | LB/B (likely benign or benign) |
| Number of sites | 395 | 2 | 14 |
| Buried or Exposed | Buried | 57.7 (%) [213] | 83.3 (%) [10] |
| | Exposed | 42.3 (%) [156] | 16.7 (%) [2] |
| Ave relacc | 23.6 % | 46.9 % | 8.1 % |
| SD relacc | 26.28 % | 41.74 % | 11.79 % |
| Contact Mol | hetero | 39.5 (%) [156] | 57.1 (%) [8] |
| | nucleotide | 0.0 (%) [0] | 0.0 (%) [0] |
| | compound | 20.3 (%) [80] | 35.7 (%) [5] |
| | metal | 25.3 (%) [100] | 42.9 (%) [6] |
| | otherpoly | 10.4 (%) [41] | 7.1 (%) [1] |
| | homo | 5.8 (%) [23] | 0.0 (%) [0] |
| | precipitant | 24.1 (%) [95] | 28.6 (%) [4] |
| Number of variants | 25 | 3 | 15 |
| N_Freq(AAvariant)==0 % | | 33.3 % [1] | 86.7 % [13] |
| N_Freq(AAvariant)>0 % | | 66.7 % [2] | 13.3 % [2] |
| Ave Freq(AAvariant) | | 6.7 % | 0.9 % |
| SD Freq(AAvariant) | | 8.73 % | 2.75 % |

病気の原因になるアミノ酸変異について

- (1) 埋もれている部位(溶媒露出度accが小さい部位)に変異が入ると、天然構造が不安定になり、機能を失活しやすい。
- (2) 相同タンパク質群で観察されるアミノ酸の割合が小さい(あまり観察されない)アミノ酸に変異した場合、機能を失いやすい。
※SIFT scoreなど多くのプログラムがこの原理に基づく。
- (3) 活性部位、SS結合部位、基質結合部位、DNA結合部位、イオン結合部位に変異が入ると当然失活しやすい
→ ただし、完全に失活する変異を持った個体は致死になりやすいため、集団に残りにくい

特徴量ごとの病因性識別能の比較

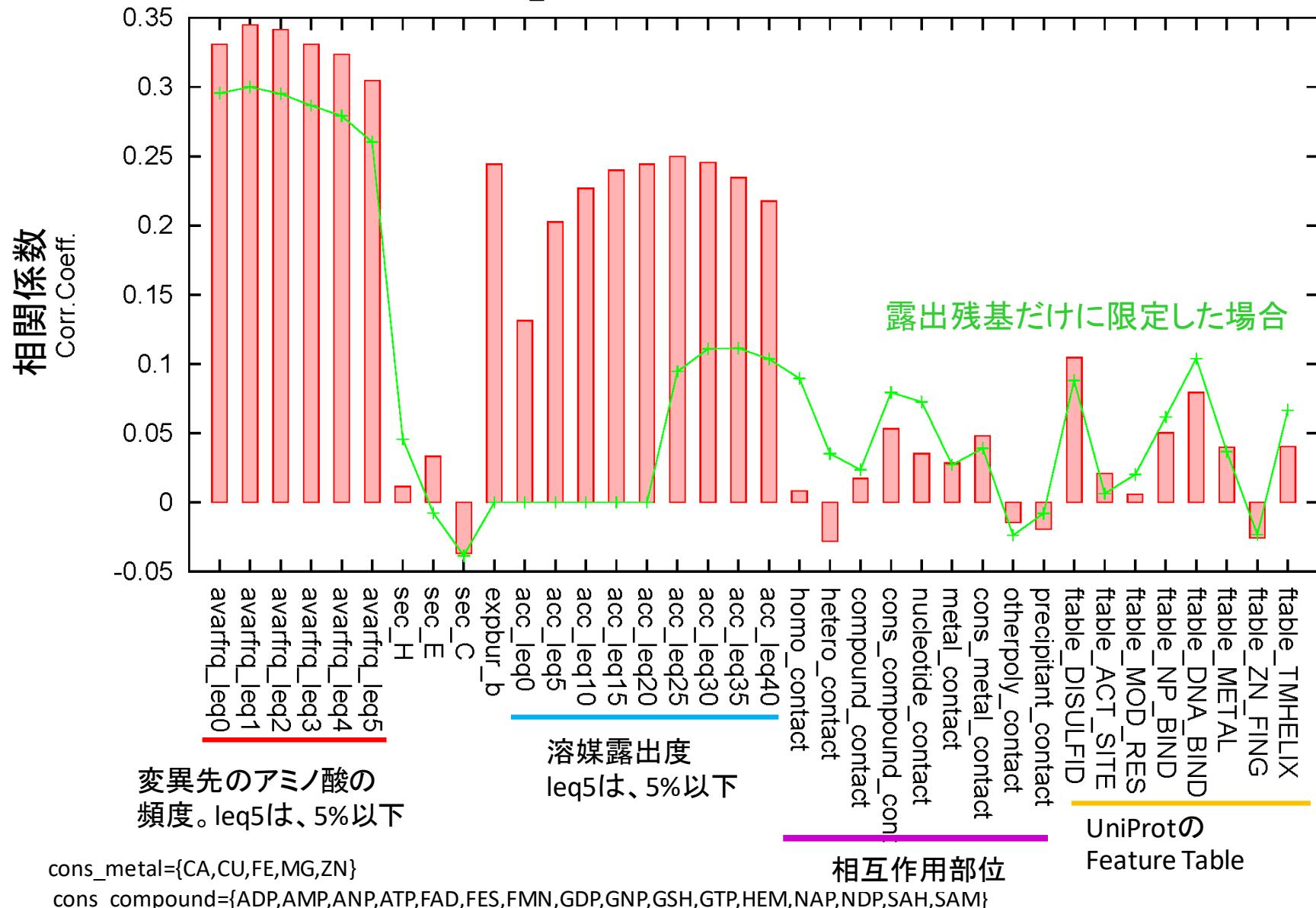
“Disease”と“Polymorphism”の変異の識別性能を、エンリッチメントで評価



特徴量ごとの病因性識別能の比較

“Disease”と“Polymorphism”的変異の識別性能を、点相関係数で評価

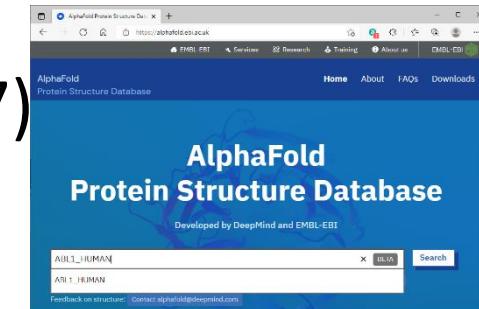
PDB 20161019, UniProt 2016_09のhumsavarを使用。HLA classI / IIをデータから除去。



AlphaFold2について

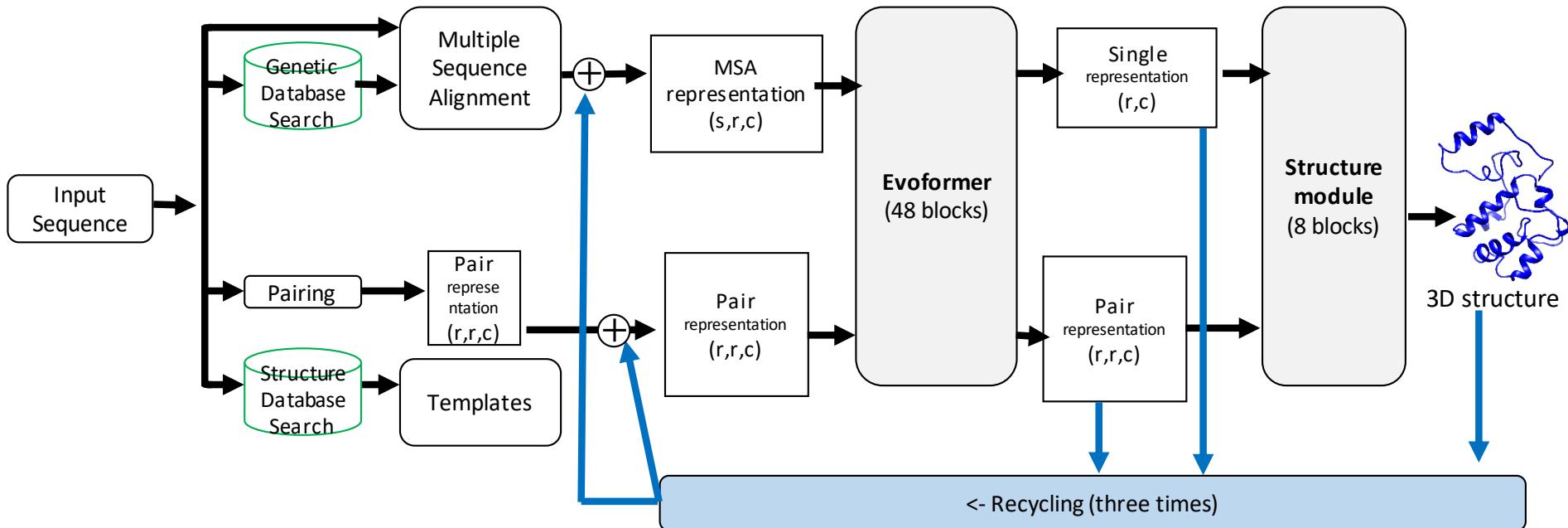


- DeepMind社(Google系列)が開発
- 立体構造予測コンテストCASP14で高成績(2020)
- 詳細な解説の論文がNatureに発表(62ページのSupp.Info.)
(2021/07)
- 予測プログラムのソースも公開(2021/07)
- データベースAlphaFold DBが公開(2021/07)



- 鑄型を使うが、鑄型がない場合でも高精度で予測される
- 最新版ではタンパク質多量体の予測にも対応
- 膜蛋白質も、人工設計タンパク質も、それなりにうまく予測

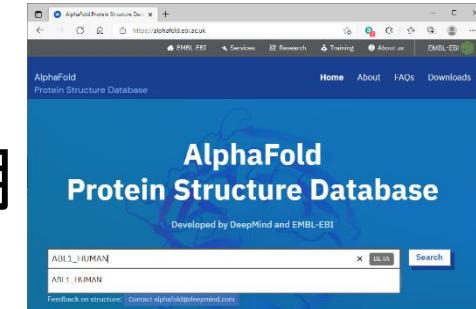
AlphaFold2の手法の概要



- 「注意機構付きTransformer」(2017)を使用(*Google翻訳*で使われている深層学習モデル)
- 共進化によるコンタクト予測を行うため、マルチプル配列アラインメント(MSA)を作成、ネットワークの入力とする
- 鑄型予測を行うため、PDBとの配列相同性検索も内部で作成、ネットワークの入力とする
- 各残基の局所構造(single representation)と接触残基ペア(pair representation)を推定する”**Evoformer**”というブロック
- それらを、原子の3次元座標を生成する”**Structure module**”というブロックに通して、予測構造を出力
- 一度出力された予測構造を再びEvoformerの入力し、反復的に予測を繰り返すことで、構造を改善させる(Recycling)
- 鑄型構造を使用しなくても、性能はさほど低下しないとされ、デノボ法としても高い性能を持つ。
- 二次構造予測、共進化を利用したコンタクトマップ予測、局所構造予測、主鎖原子の座標生成、側鎖原子の座標生成がすべて一連のニューラルネットワークの中に組み込まれている。
- アミノ酸配列から立体構造までほぼ全てニューラルネット(「エネルギー」や「確率」という概念を使わない)。⁵⁷

AlphaFold2を使用するには？

(1)予測構造データベースAlphaFold DBを利用する
<https://alphafold.ebi.ac.uk/>



UniProtのヒトのタンパク質 + 主要なモデル生物のタンパク質
単量体構造のみ。タンパク質多量体、核酸、低分子化合物などは含まれない

(2)予測サービスのcolab Webサーバを利用する

<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>
タンパク質多量体の予測も可能

予測計算の計算量制限あり。タンパク質の長さ、予測精度に制限がある

(3)学習済み予測プログラムをインストールして使う

CPUだけでも動作するが、GPUがあったほうがよい。

プログラム以外に大きな配列データベース(1-2TB)をダウンロードする必要

予測計算は実行できるが、学習計算はできない。

予測プログラムのソースはオープンだが、パラメータセットは商用目的での使用は不可

AlphaFold2の予測構造の見方

- ・ 残基ごとの予測構造の信頼性pLDDTスコアを確認
 - PDB/mmCIFファイルの温度因子の欄に記入される
 - Webページの3Dビューアでは、信頼性が高いほど青、低いほど赤で表示される
- ・ 残基の相対位置の信頼性”Predicted Aligned Error”を確認
 - Webページでは2次元マップとして表示
 - ドメインごとにバラバラに鋳型予測され、ドメイン間の配置を予測推定した場合、ドメイン間の残基の”Predicted Aligned Error”は大きくなる
- ・ 天然変性領域は、便宜的に伸びきった予測構造が与えられる
 - 天然変性領域は、生理条件下で一定の構造に定まらない
 - 天然変性領域は、pLDDTスコアが極めて低くなる
 - 天然変性領域の伸びきった予測構造を、その通りに解釈してはならない

ABL1_HUMAN :Tyrosine-protein kinase ABL1

HOMCOS

UniProt Feature Tables [P00519(ABL1_HUMAN)]

| 1130 | | | |
|------|----------|----------|------------------|
| | region | name | description |
| | 1-1130 | CHAIN | /note="Tyrosine- |
| ■ | 61-121 | DOMAIN | /note="SH3" |
| ■ | 127-217 | DOMAIN | /note="SH2" |
| | 242-493 | DOMAIN | /note="Protein k |
| ■ | 248-256 | NP_BIND | /note="ATP" |
| | 316-322 | NP_BIND | /note="ATP" |
| ■ | 1-60 | REGION | /note="CAP" |
| | 518-996 | REGION | /note="Disorder |
| | 869-968 | REGION | /note="DNA-bind" |
| | 953-1130 | REGION | /note="F-actin-b |
| ■ | 533-549 | COMPBIAS | /note="Basic and |
| ■ | 579-601 | COMPBIAS | /note="Basic and |
| ■ | 622-638 | COMPBIAS | /note="Basic and |
| ■ | 685-703 | COMPBIAS | /note="Polar res |
| ■ | 735-753 | COMPBIAS | /note="Polar res |
| ■ | 758-772 | COMPBIAS | /note="Basic and |
| | 870-894 | COMPBIAS | /note="Basic and |
| | 363-363 | ACT_SITE | /note="Proton ac |
| ■ | 271-271 | BINDING | /note="ATP" |
| ■ | 1-1130 | DISORDER | predicted by DIS |

予測天然変性領域

天然変性領域予測

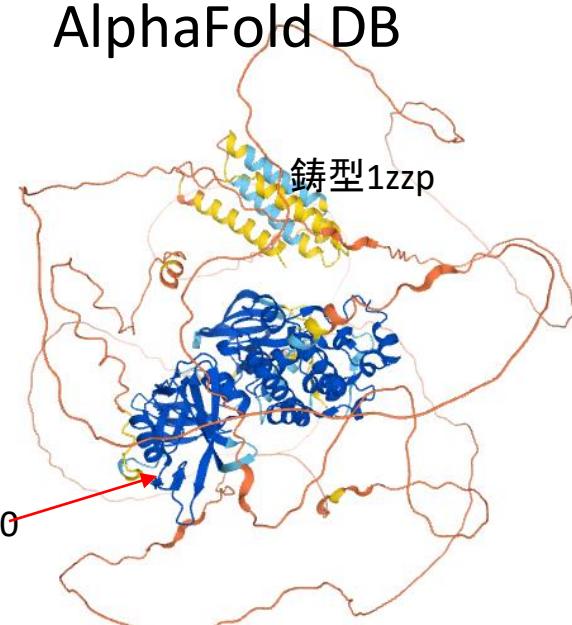
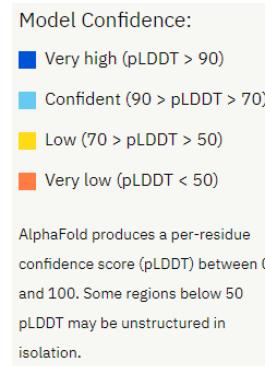
MONOMER

鑄型2fo0

1130

| pdb_id | a ¹ | identity[%] ² | description |
|--------|----------------|--------------------------|-------------|
| 2fo0 | A | 99.8 | ABL1_HUMAN |
| 6amv | A | 99.1 | ABL1_HUMAN |
| 1zzp | A | 99.1 | ABL1_HUMAN |

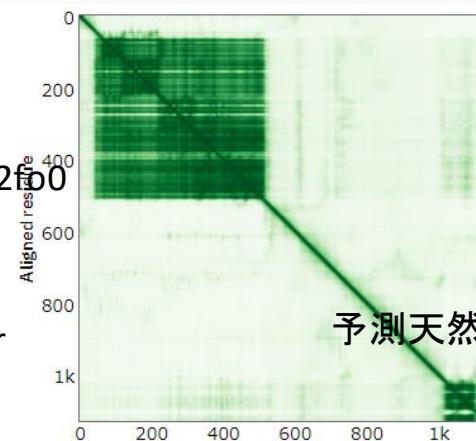
鑄型1zzp



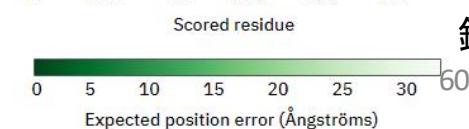
鑄型2fo0

予測天然変性領域

鑄型2fo0
Predicted
Aligned Error



予測天然変性領域



鑄型1zzp

Expected position error (Ångströms)

HHIP_HUMAN :Hedgehog-interacting protein

(AC: Q96QV1)

HOMCOS

AlphaFold DB

7pgk:(公開日:2021-12-15)

UniProt Feature Tables [Q96QV1(HHIP_HUMAN)]

| | | region | name | description |
|--|--|---------|----------|---|
| | | 1-17 | SIGNAL | |
| | | 18-700 | CHAIN | /note="Hedgehog-interacting protein" |
| | | 607-634 | DOMAIN | /note="EGF-like 1" |
| | | 635-667 | DOMAIN | /note="EGF-like 2" |
| | | 376-388 | REGION | /note="Interaction with disordered regions" |
| | | 383-383 | METAL | /note="Zinc; shared with HHIP_HUMAN" |
| | | 1-683 | DISORDER | predicted by DISOPRED |

MONOMER

3ho5:(公開日:2009-06-23)

| | | | description |
|--|------|---|------------------|
| | 3hos | B | 99.8 HHIP_HUMAN |
| | Zpgk | A | 100.0 HHIP_HUMAN |

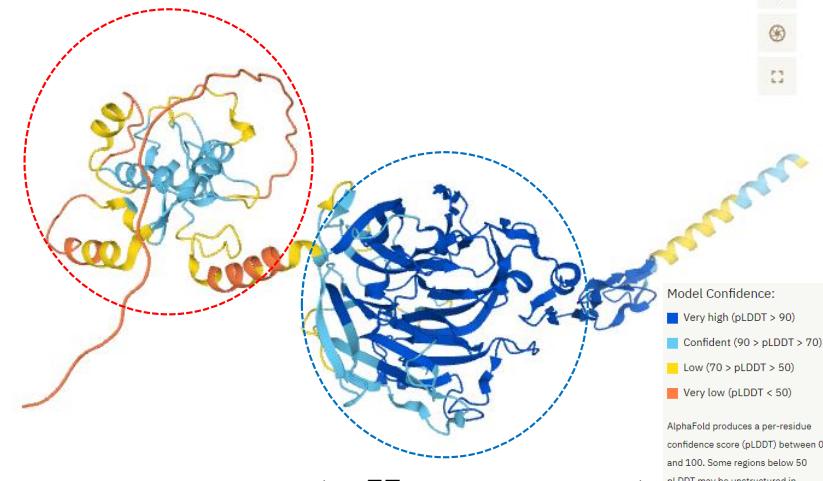
1.asym_id for the homologue. 2.identity[%] 3.sequence

7pgk:(公開日:2021-12-15)

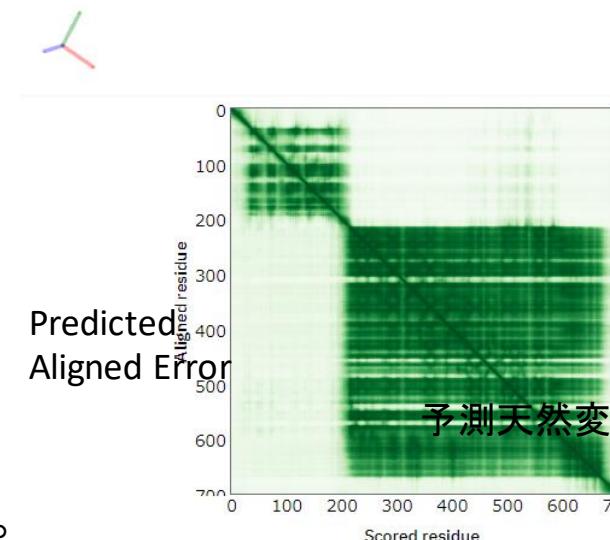
AlphaFold is trained on protein chains in the PDB released before 2018-04-30. Predictions can also make use of up to 4 templates released before **2021-02-15**.
<https://alphafold.ebi.ac.uk/faq>

→AlphaFold DBは、2021-02-15のPDBを用いている。

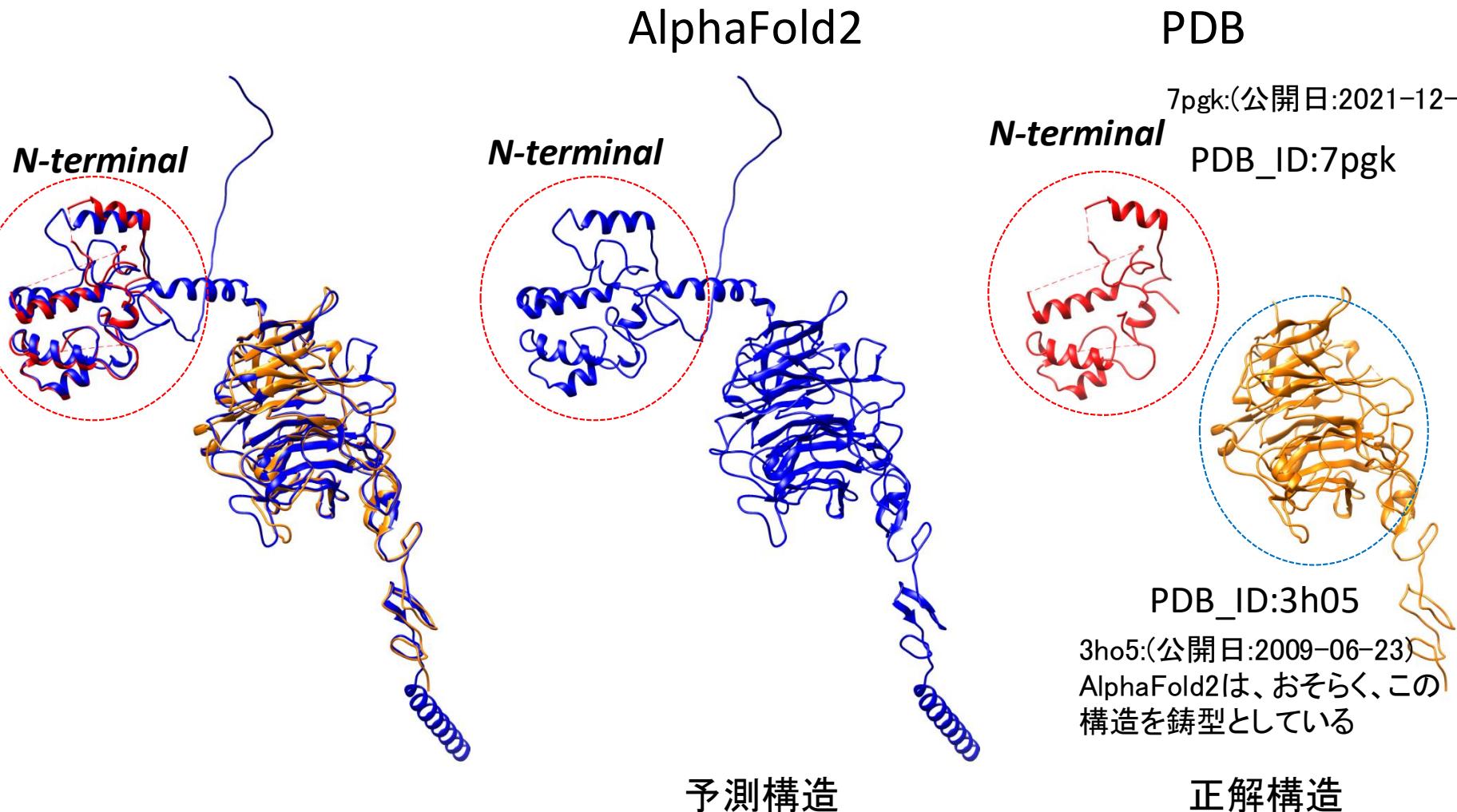
それ以後、決定された構造を用いて、予測精度の評価ができるはず



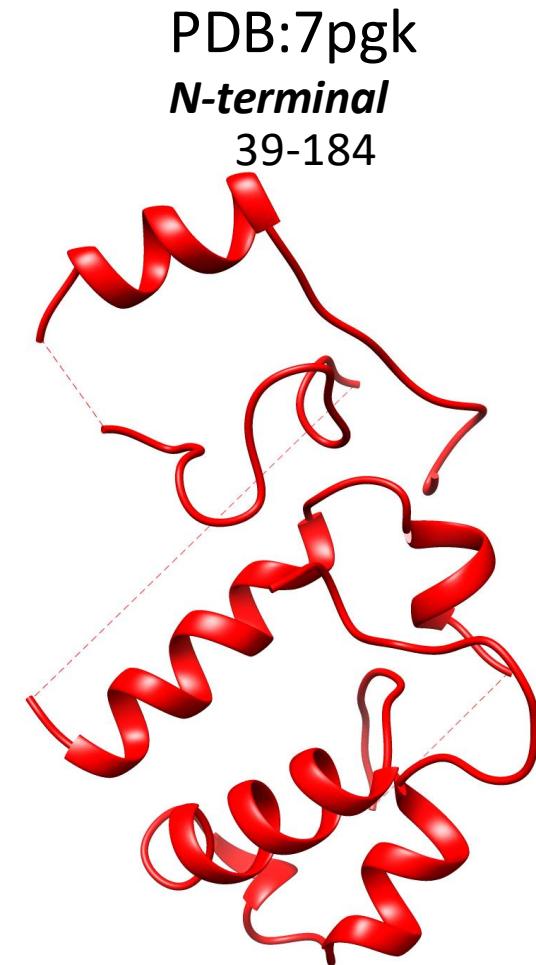
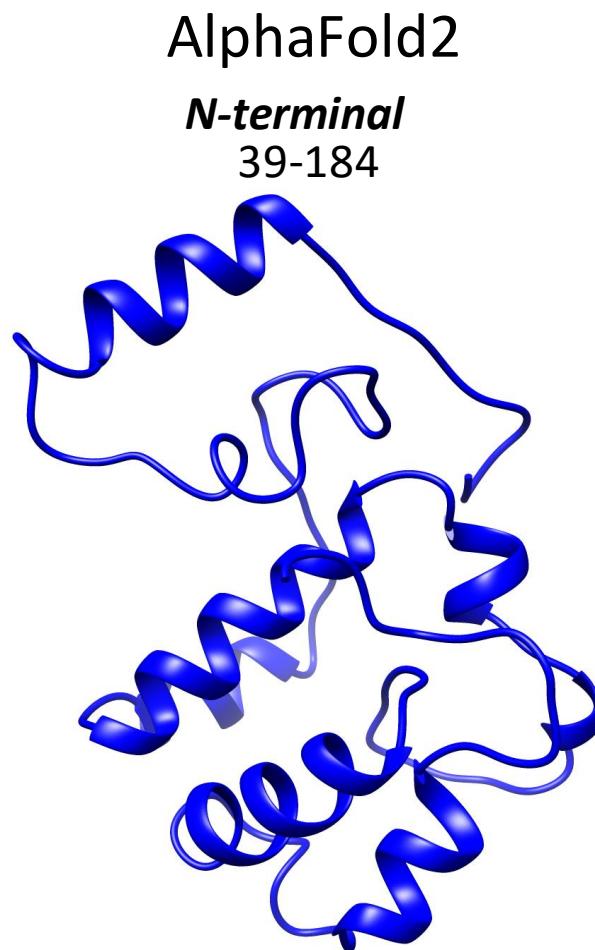
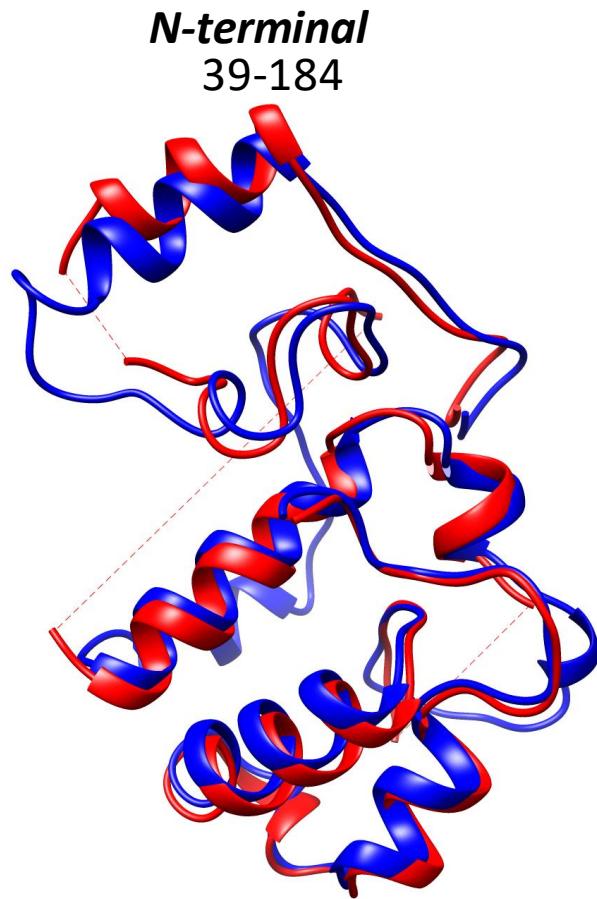
3ho5:(公開日:2009-06-23)



HHIP_HUMAN :Hedgehog-interacting protein (AC: Q96QV1)



HHIP_HUMAN :Hedgehog-interacting protein (AC: Q96QV1)



予測構造

RMSD(Calpha)=2.01 Å

正解構造

※PSI-BLASTを用いてもこの領域は何もヒットしない...

GP158_HUMAN :Probable G-protein coupled receptor 158 (AC: Q5T848)

HOMCOS

| | | | |
|----------|-----------|----------|------------------------|
| 予測天然変性領域 | 975-998 | COMPBIAS | /note="Polar residues" |
| | 1149-1164 | COMPBIAS | /note="Polar residues" |
| | 1-1145 | PRED | |

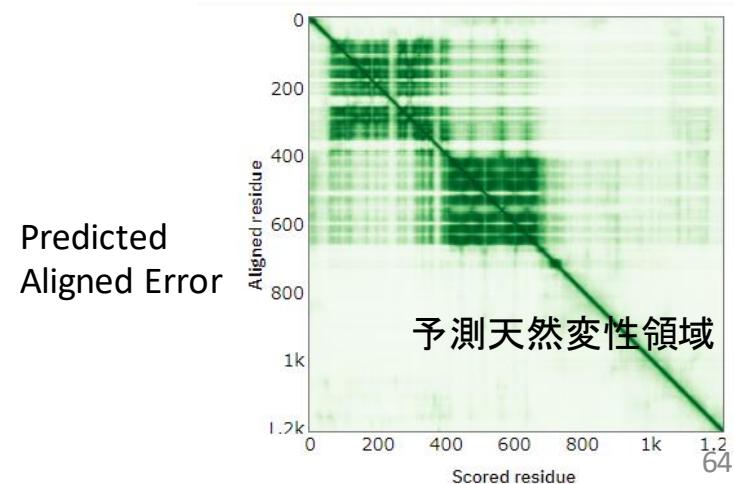
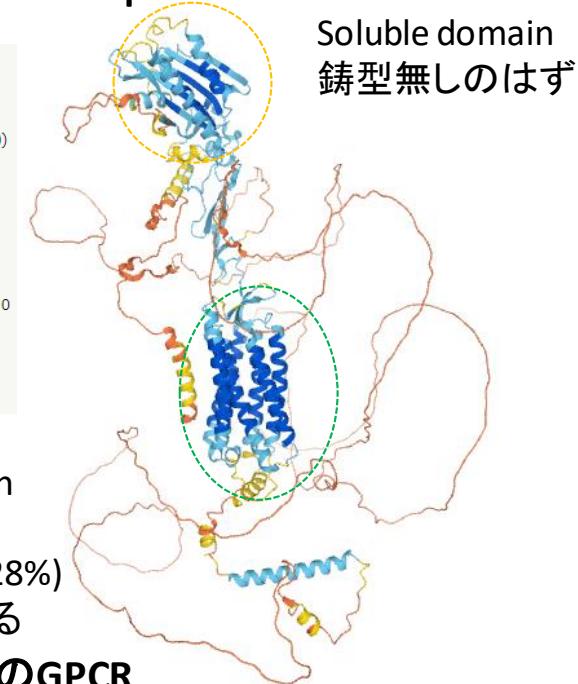
| MONOMER | 1215 | pdb_id | a ¹ | identity[%] ² | description |
|---------|------|--------|----------------|--------------------------|-------------------------|
| | | 7ewr | C | 100.0 | GP158_HUMAN Probable |
| | | 7ewp | A | 100.0 | GP158_HUMAN Probable |
| | | 7she | B | 100.0 | GP158_HUMAN G-protei |
| | | 7she | A | 100.0 | |
| | | 7ewr | D | 100.0 | |
| | | 7ewp | B | 100.0 | |
| | | 7ewl | A | 100.0 | |
| | | 7ewl | B | 100.0 | GP158_HUMAN Probable |
| | | 7shf | D | 100.0 | GP158_HUMAN G-protei |
| | | 7shf | C | 100.0 | GP158_HUMAN G-protei |
| | | 7dge | B | 27.7 | GRM1_HUMAN Metabot |
| | | 7dge | A | 27.7 | GRM1_HUMAN Metabot |
| | | 7dgd | A | 27.7 | GRM1_HUMAN Metabot |
| | | 7dgd | B | 27.7 | GRM1_HUMAN Metabot |
| | | 6n51 | A | 25.9 | GRM5_HUMAN Metabot |
| | | 6n51 | D | 25.9 | GRM5_HUMAN Metabot |
| | | 4or2 | B | 23.9 | C562_ECOLX GRM1_HU |
| | | 4or2 | A | 24.1 | C562_ECOLX GRM1_HU |
| | | 7fd9 | B | 26.3 | GRM5_HUMAN Metabot |
| | | 7fd9 | A | 26.3 | GRM5_HUMAN Metabot |
| | | 6n52 | B | 26.7 | GRM5_HUMAN Metabot |
| | | 6n52 | A | 26.7 | GRM5_HUMAN Metabot |
| | | 7fd8 | B | 26.3 | GRM5_HUMAN Metabot |
| | | 7fd8 | A | 26.3 | GRM5_HUMAN Metabot |
| | | 7epd | B | 27.5 | Isoform 3 of Metabotrop |

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.

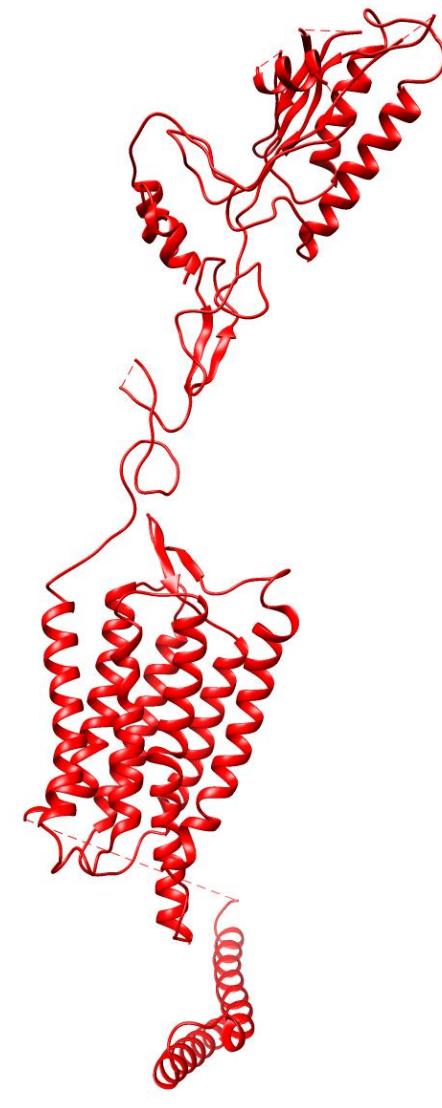
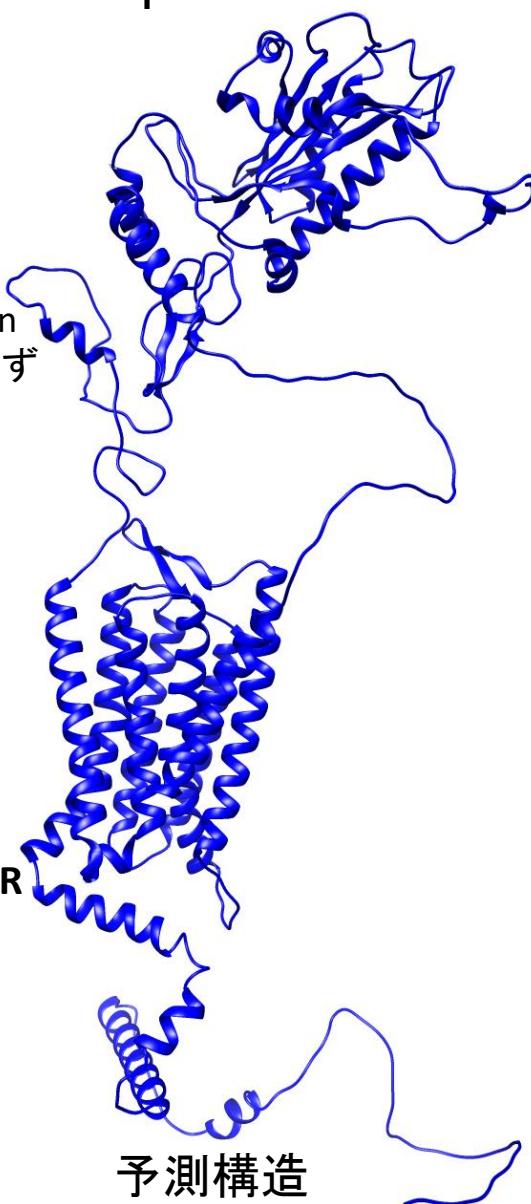
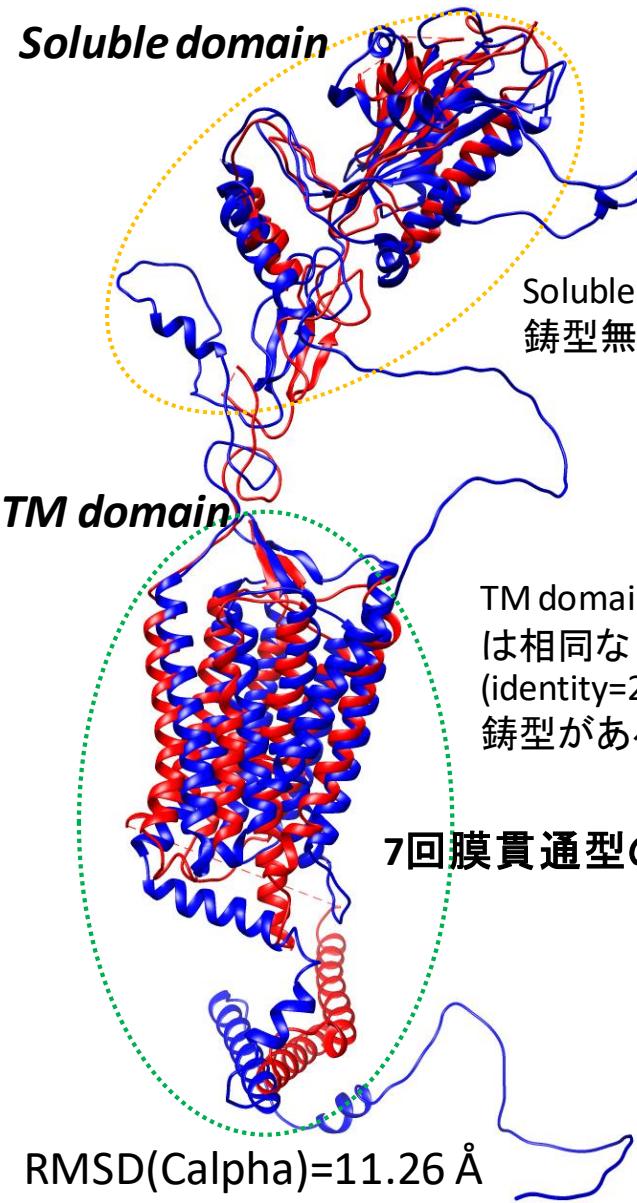
TM domain
は相同な
(identity=28%)
鋳型がある
7回膜貫通型のGPCR



GP158_HUMAN :Probable G-protein coupled receptor 158 (AC: Q5T848)

AlphaFold2

PDB:7ewp

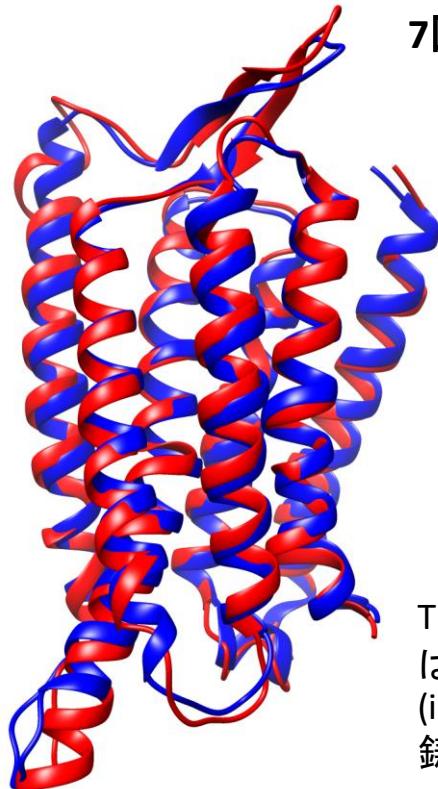


GP158_HUMAN :Probable G-protein coupled receptor 158 (AC: Q5T848)

AlphaFold2

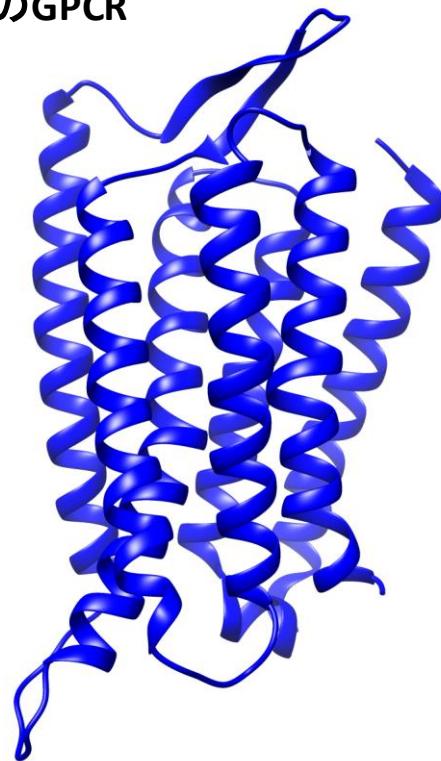
PDB:7ewp

TM domain
412-668



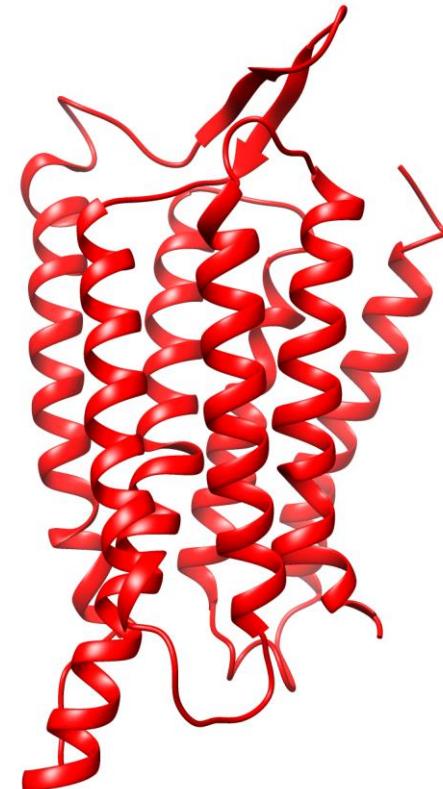
TM domain
は相同な
(identity=28%)
鑄型がある

TM domain
412-668



RMSD(Calpha)=1.57 Å

TM domain
412-668



GP158_HUMAN :Probable G-protein coupled receptor 158

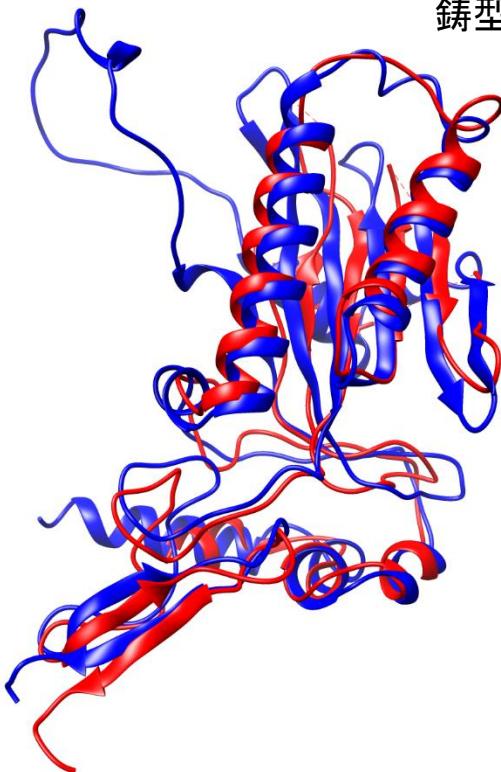
AlphaFold2

PDB:7ewp

Soluble domain

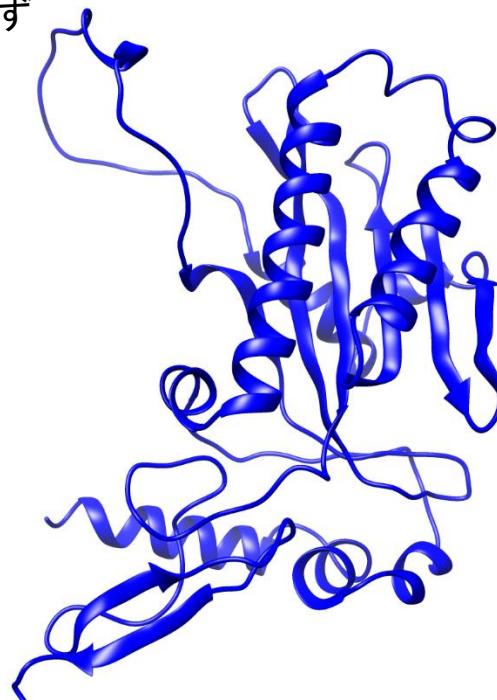
60-358

Soluble domain
鑄型無しのはず



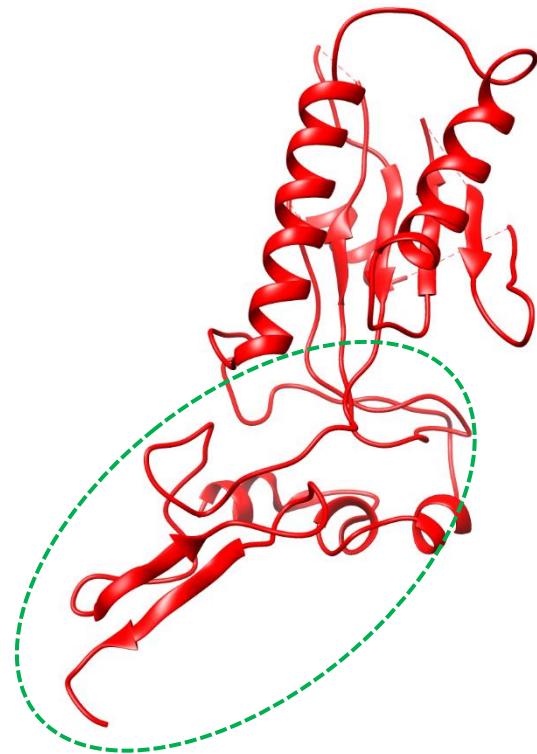
Soluble domain

60-358



Soluble domain

70-358



予測構造

正解構造

RMSD(Calpha)=3.84 Å

※PSI-BLASTで284-412が2e4xにはidentiy=17.8%でヒットする。なので、C末部分は鑄型ベース?⁶⁷

MKS3_HUMAN :Meckelin (AC: Q5HY8)

HOMCOS

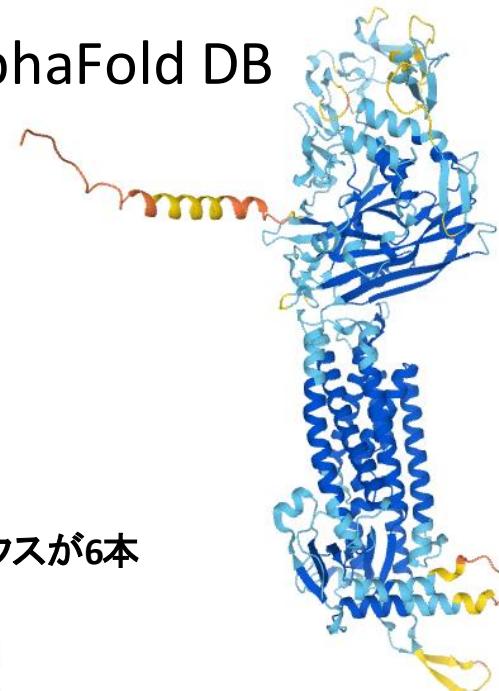
UniProt Feature Tables [Q5HYA8(MKS3_HUMAN)]

| 995 | | | |
|-----|---------|----------|-----------------------------------|
| | region | name | description |
| ■ | 1-35 | SIGNAL | |
| ■ | 36-995 | CHAIN | /note="Meckelin" /id="PRO_000022" |
| ■ | 526-546 | TRANSMEM | /note="Helical" |
| ■ | 570-590 | TRANSMEM | /note="Helical" |
| ■ | 609-629 | TRANSMEM | /note="Helical" |
| ■ | 689-709 | TRANSMEM | /note="Helical" |
| ■ | 734-754 | TRANSMEM | /note="Helical" |
| ■ | 939-959 | TRANSMEM | /note="Helical" |
| ■ | 1-990 | DISORDER | predicted by DISOPRED |

| MONOMER | | | |
|---------|--------|----------------|--------------------------|
| 995 | | | |
| | pdb_id | a ¹ | identity[%] ² |
| ■ | 7fh1 | A | 100.0 |

7fh1:(公開日:2021-12-01)

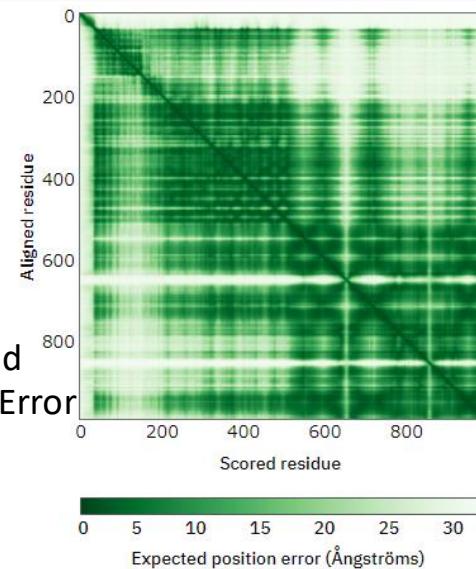
AlphaFold DB



UniProtでは
膜貫通ヘリックスが6本

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.

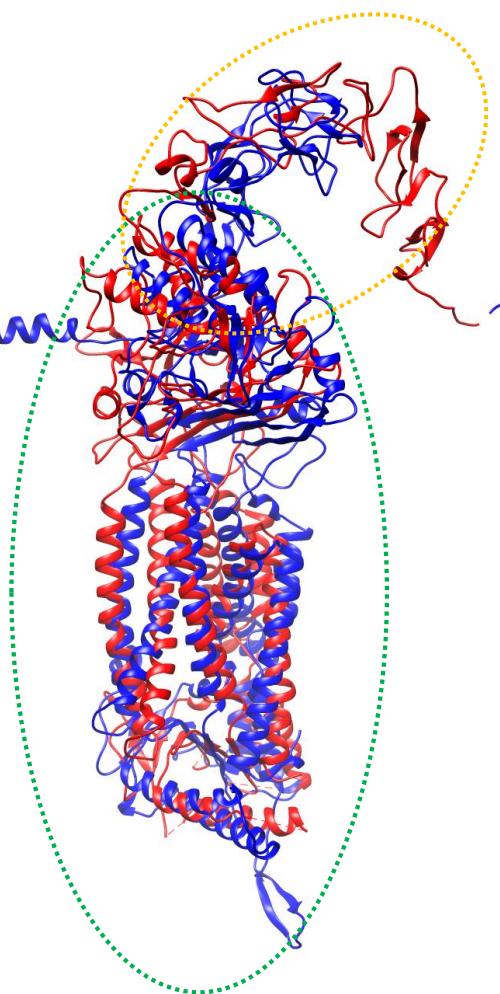


MKS3_HUMAN :Meckelin (AC: Q5HY8)

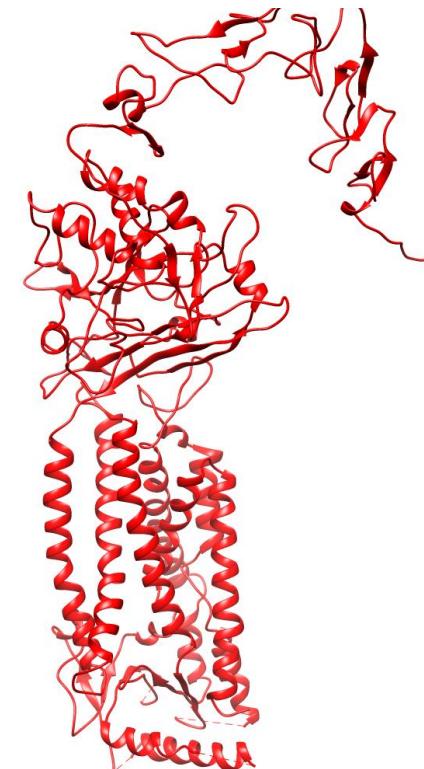
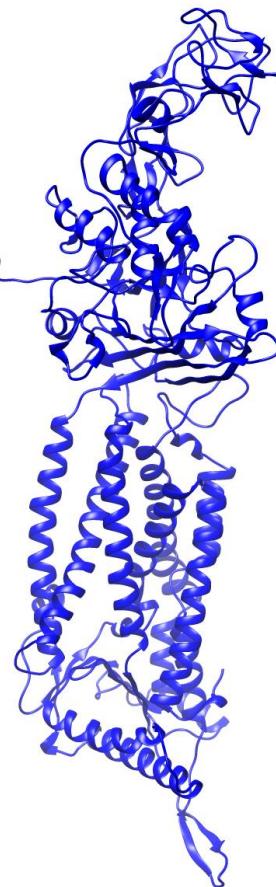
AlphaFold2

PDB:7fh1

7fh1:(公開日:2021-12-01)



予測構造



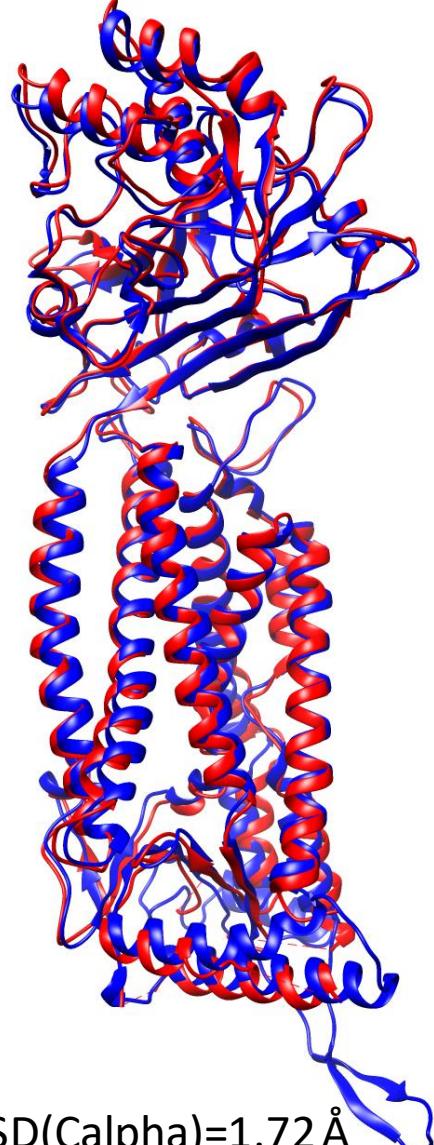
正解構造

RMSD(Calpha)=13.61 Å

MKS3_HUMAN :Meckelin

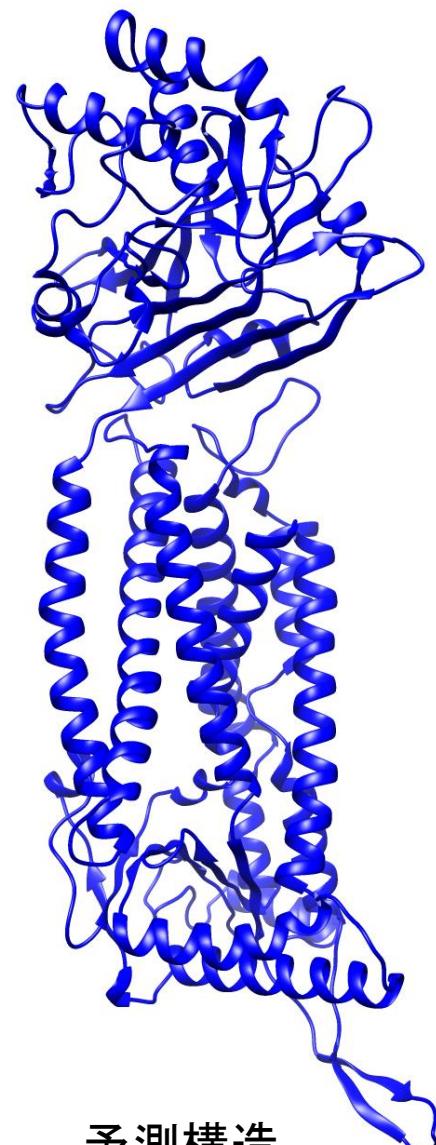
(AC: Q5HY8)

209-989



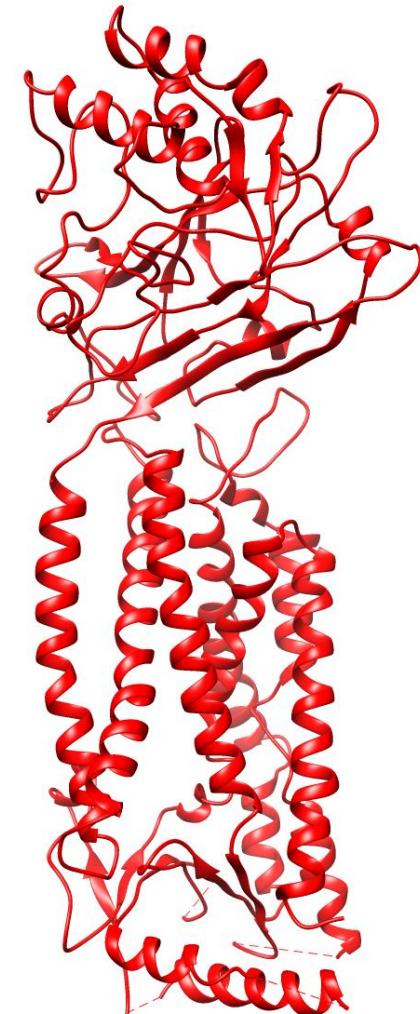
AlphaFold2

209-989



PDB:7fh1

209-989



7fh1:(公開日:2021-12-01)

MKS3_HUMAN :Meckelin (AC: Q5HY8)

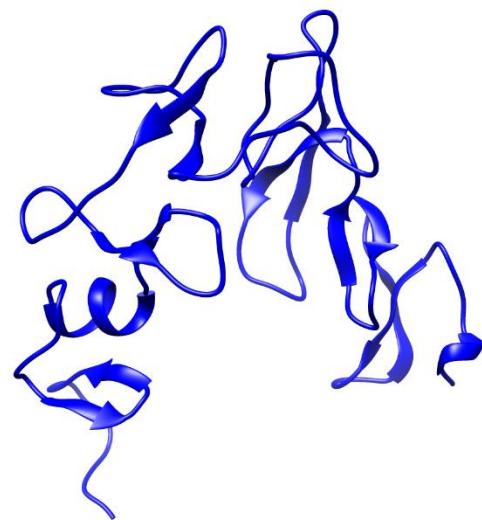
44-208



RMSD(Calpha)=11.08 Å

AlphaFold2

44-208



予測構造

PDB:7fh1

44-208



7fh1:(公開日:2021-12-01)

正解構造

AlphaFold DBの予測構造のまとめ

- ・他の蛋白質、核酸、化合物などの分子は含まれない
- ・天然変性領域は予測されない(低いpLDDTスコア)
- ・ドメイン間の配置は正確に予測されない場合がある
(低いPredicted Align error)
- ・修飾アミノ酸は考慮されない
- ・鋳型構造がない部分でも、驚くほど精確に予測されることがある
- ・予測原理は「遠縁の相同鋳型構造を探す」ではない。
「最小自由エネルギー構造を探す」でもない。
- ・PDBデータベースから期待される「タンパク質らしい構造」を生成するが予測原理？ 言語モデル(英語らしさ、日本語らしさ)に相当するモデルが巧妙に実現されている？

ご清聴ありがとうございました

