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

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ロジーモデリング 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<u>球状タンパク質</u>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、<u>膜タンパク質</u>、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。
- ・<u>天然変性タンパク質</u>(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

HEM (asym_id: G)

HEM (asym_id: **H**)

🔛 🖧 IEM (asym_id: E)

HEM (asym id: F)

舗

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)C鎖
 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、これまでは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、<u>直接PDB</u>
 <u>サーバから構造を読みにいく機能("fetch")</u>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エアは対応済み



8

PDB

旧PDBフォーマットの立体構造データの例

HEADER	OX	YGEN	TRAI	NSI	PORT			07-M	AR-84	2HHB		
TITLE	TH	ECR	YSTA:	LS	STRUCTURE	OF HUMA	N DEOXYI	HAEMOGLO	BIN AT	1.74 ANG	GSTROMS	PDB
ATOM	1	Ν	VAL	А	1	6.130	16.559	4.905	1.00	41.29	Ν	
ATOM	2	CA	VAL	А	1	6.870	17.784	4.702	1.00	41.33	С	
ATOM	3	С	VAL	А	1	8.377	17.548	4.913	1.00	31.64	С	יחו אחס
ATOM	4	0	VAL	А	1	8.820	16.980	5.922	1.00	38.31	○	
ATOM	5	CB	VAL	А	1	6.345	18.763	5.731	1.00			2hhb
ATOM	6	CG1	VAL	А	1	6.745	20.188	5.356	1.00		HEM	•
ATOM	7	CG2	VAL	А	1	4.826	18.612	5.847	1.00			S HEM
ATOM	8	Ν	LEU	А	2	9.146	18.005	3.962	1.00			A STATE
ATOM	9	CA	LEU	А	2	10.599	17.914	4.153	1.00			TUPP)
ATOM	10	С	LEU	А	2	11.062	19.085	5.062	1.00			
ATOM	11	0	LEU	А	2	10.829	20.254	4.744	1.00			A鎖
ATOM	12	CB	LEU	А	2	11.269	18.078	2.776	1.00			
ATOM	13	CG	LEU	А	2	10.986	16.983	1.769	1.00	C 與 V		
ATOM	14	CD1	LEU	А	2	11.735	17.168	0.427	1.00			\mathbf{X}
ATOM	15	CD2	LEU	А	2	11.276	15.630	2.404	1.00			
ATOM	16	Ν	SER	А	3	11.674	18.766	6.158	1.00	HE		
ATOM	17	CA	SER	А	3	12.286	19.774	7.034	1.00			▶ 全当
ATOM	18	С	SER	А	3	13.529	20.322	6.334	1.00		HEM 🖤 🔍	D坝
ATOM	19	0	SER	А	3	13.995	19.754	5.344	1.00	27.40	0	
ATOM	20	СВ	SER	А	3	12.719	19.060	8.326	1.00	23.83	С	
ATOM	21	OG	SER	А	3	13.844	18.245	8.107	1.00	29.07	0	
:												
HETATM	4389	CHA	HEM	А	142	8.634	7.898	-18.334	1.00	17.24	С	
HETATM	4390	CHB	HEM	А	142	10.386	9.965	-14.276	1.00	23.17	С	
HETATM	4391	CHC	HEM	А	142	8.337	6.403	-11.611	1.00	11.04	С	
: 原子	子番号 质	[子名]	残基	名	残基番号	X座標	Y座標	Z座標	占有率	温度因子	元素名	
			釗	肖詢	別子							
L												

— 80文字の固定長 ⁻ パンチカード時代の名残り



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、<u>著者入力とデータベース管理者入力の2種が書かれる。</u>
- 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。



フォーマットです!

mmClFを使わなければならない場合

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mmClF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

185935	AND DI	DRi	English <mark>日本語</mark> 简体中文 繁體中文 한국어	
件を2022-01-12に公開中	Real Diversit Protein Data	a Bank Japan	Search pdbj.org	2021年1月12日時点で
	F 2 D Protein Data Foundation	i Bank	wwPDB_RCSB_PDB_EDBe_BMRB_Adv. Search_heig	2498エントリ
ホーム	PDB形式変換不可の PDB	<u>エントリー</u>		検索結果
トップページ	1805	SHIGA-LIKE	TOXIN COMPLEXED WITH ITS RECEPTOR	全ヒット件数: 2498
た言十十青幸長	1005	分子名称:	SHIGA-LIKE TOXIN I B SUBUNIT, alpha-D-galactopyranose-(1-4)-beta-D-galactopyranose, alpha-D-galactopyranose-(1-4)-	表示件数: 25
ルプ	👱 🐨	英者	beta-D-galactopyranose-(1-4)-beta-D-glucopyranose,	表示順 (0→9,a→z)
Q	*	登録日	1998-01-13	(0)/0 _/
問い合わせ	· · · · · · · · · · · · · · · · · · ·	公開日	1999-02-02	
DBjの引用・利用規約		最終更新日	2020-07-29	表示順
ンク集	2 0 0 Ta	実験手法	X-RAY DIFFRACTION (2.8 Å)	PDBID塁順 (0→9.a→z)
ettings	1 8 m	主引用文献	Structure of the shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3.	PDBID降順(z→a,9→0)
	0		Diochemisci y, 57, 1996	登録日の古い順
データ登録(OneDep)				登録日の新しい順
ルプ	1VVJ	Crystal Strue	cture of Frameshift Suppressor tRNA SufA6 bound to Codon CCC-G on the Ribosome	公開日の古い順
DB、EMDB、BMRBへの登録		万丁名柳; 茎者	165 rKNA, 235 rKNA, 305 ribosomai protein S10,	公開日の新しい順
The second s		登録日	2013-05-24	分解能の高い順
ダウンロード	. Chief Barry	公開日	2014-08-06	
DBアーカイブからのデータダウンロ	THE STORE OF	最終更新日	2014-12-10	Lists
·ř		実験手法	X-RAY DIFFRACTION (3.44000111574 Å)	IDリストをダウンロード
		主引用文献	Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops	
標準フォーマット	1997		PIUC.Nati.Acau.Sci.USA, 111, 2014	
				あしい PDB エントリー 廃止された全ての PDB エントリー
DBX/InitiCIF(とついて フォーフット亦通	1\/Y4	Crystal struc	ture of the Thermus thermophilus 70S ribosome in the pre-attack state of peptide bon	d 廃止された最新の PDB エントリー
		formation co	ontaining acylated tRNA-substrates in the A and P sites.	PDB形式変換不可の PDB エントリー
B形式変換不可工	ントリー 🖉 👘	分子名称:	16S Ribosomal RNA, 23S Ribosomal RNA, 30S ribosomal protein S10,	PDBグループ登録エントリー
		者百	Polikanov, Y.S. Steitz, T.A. Innis, C.A.	X線構造因子データを含む全てのPDBエン
0190000	A Contraction	豆球口	2014-05-13	
ルプ		品終更新日	2019-07-03	ス級構造因子テータを含む新しいや08エン リー
DB形式変換不可エントリー		実験手法	X-RAY DIFFRACTION (2.6 Å)	X線構造因子データが更新されたPDBエン
ループ登録エントリー	A A A A A A A A A A A A A A A A A A A	主引用文献	A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome.	U—
			Nat.Struct.Mol.Biol., 21, 2014	X線構造因子データを含む全ての廃止され

13

PDB IDの探し方

(1) 文献検索 構造を決定した論文には必ずPDBIDが書いてある

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

検索サービス

PDBjのトップ画面の上のフォームに タンパク質名、ファミリー名などを入力

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 配列相同性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

ヘルプ					
PDB検索 (PDBj Mine)					
PDB詳細検索	Sequence N	lavigator			
化合物检查 (Chamaia)	Mode	PDB entry	Custom sequence		
11合物検索(Chemie)		MDMFQKVEK	GEGTYGVVYKAKNRETGQI	VALKKIRLDLEMEGVPSTAIREISLLKELKH	
BMRB検索		PNIVRLLDV HRVIHRDLK	/HNERKLYLVFEFLSODLK PONLLINELGAIKLADFGL#	XYMDSTPGSELPLHLIKSYLFQLLQGVSFCHS ARAFGVPLRTYTHEVVTLWYRAPEILLGSKFY	
Sequence-Navigator	TTAVDIWSIGCIFAEMVTRKALFPGDSEIDQLFRIFRMLGTPSEDTWPGVTQLPDYKGSF Sequence pkwrrkGleeivpnlepecrdllmollovDpsorifaktalahpyfsspppsparoyvl				
DASH	-	QRFRH			
EM Navigator					
Omokage検索	Sequence	Protein	Nucleic acid		
SeSAW	type				
	Clustering:	No o	lustering		
NBDC KDF/N-9/L	Find home	ologues		14	
未公開エントリーのステータ					

English 日本語 简体中文 繁體中文 한국어

Q

HOMCOSを利用した配列からのPDB IDの検索

П В ном	COS:相同複合体の トナキャー・/	/home	soc odbi org	- L X		(2) 粘	们至己	与啠	መኋ	ケオオ	キ ーバ		
$\leftarrow \rightarrow$ (nome	os.hanl.org	r 🕀 😩 …		(2)	以虫口	コ貝	<u></u>	ムが出来し	⊒ <i>\</i>]`		
	E.		E 6 5			複合体	の種	類为	リ(こ	表示され	15		
1	Номсс	OS:相同複合体の	検索・モデリングサーバ		$\leftarrow \rightarrow$	C A Ó	https://homcos.pd	lbj.org/cgi-bin	/prot_sch_c	onbars_demo.cgi?QUNIF	ROT_AC ର୍	ි 🚱	\$
[Go to English pag							Contact	t Molecules	for Hom	ologous Proteins			
HOMCOS(<u>HOM</u> olog 分子ペアの構造を予 にして複合体立体構	gy modeling of <u>CO</u> mplex <u>S</u> tructure) は、PC 測するためのサーバです。アミノ酸配列や化 NBを予測すスマンができます。 <i>A</i> トバク暗のM	DBに収納されてい 学構造から、PDB 即別精心性検索に	る複合体の立体構造データを利用して、分子の類似性・ 内から複合体の立体構造が解けている他の分子を探索し +pl ASTを、化合物の構造特別性検索にけVCOMPLのを目	相同性から、構造未知の たり、PDB内の構造を鋳型 BIいています		[Full Bars] [Site]] [Table]	Summa	ry Bars[0	0.0 %]	Back to Search F	Back [Back	c to H
<u><u><u></u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u></u>)「タンパク質にえ	対する	結合分子の検索」を	き選ぶ			seq_id(%): [2]	[<u>0] [30] [40]</u>	[50] [60] []	70] [80] [90] [95] [100]	tab split [show] [downloar	<u>d] [help]</u>	
4.0.0.7.0.4	タンパク質に対する結合分子の検索	アミノ酸配列	1本のアミノ酸配列をクエリとして、それと類似した。 る分子を検索します	タンパク質と結合してい	PID Qu	ueryLength PDB	us Sequence in Ur	niProt Query	ТІТ	LE			
結合分子の検索	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 パク質を検索します	合物と結合しているタン	2670477 41	19 142	PC	DTC9(NCAP	SARS2) Rei	:Name: Full=Nucleoprot ort=Protein N ;	ain ; Short=N;AltN	Name: Full=Nu	cleoc
	ホモ多量体のモデル	アミノ酸配列	1本のアミノ酸配列をクエリとして、そのホモ多量体 モデリングによって予測します	の立体構造をホモロジー	QUERYSEQ PAN BLAST file for	NGPONORNAPRI TEGOPSDSTGSNON NAAI VLOLPOOTTLPKOF VAEGSROG NI AGFAPSASAFFGMSRI GMEVTPSGT PDB] (plain) (bar) (r	IGERSGARSKORRPOGLPNNTASUR ISOASSRSSRSRNSSRNSTPOSSR WLTYTGAIKLODKOPNFKDOVILL multiple alignment)	FTALTOHGKEDLKFPRG RGTSPARMAGNGGDAALO LNKHI DAYKTFPPTEPKI [BLAST for U	NGVPINTNSSPDDA ALLLLDRLNALESKA COKKKKADETAALPA niProt: <u>(pla</u>	GYYRRATREI RGGDGKMKDLSPRIVE YYLI KSGKGGGGGGGTVTKKSAAEASKKPROKRTI RGKKGGTVTLLPAADLDDFSKGLGGSMSSI in) (bar) (multiple align)	TGPEAGLPYGANKDGI I IWATEG TKAYNVTQAFGRRGPEQTQGNFG DSTQA ment) (PSSM file)	GALNTPKDHIGTRN GDRELIRRGTDYKH	
複合体立体構造の モデリング	▲テロ多量体のモデル	2本のアミノ酸 配列	2本のアミノ酸配列をクエリとして、そのヘテロ多量 -モデリングによって予測します	体の立体構造をホモロジ	MONOMER	単量体							_
	▲ 化合物-タンパク質複合体のモデル	アミノ酸配列と 化学構造	1本のアミノ酸配列と1つの化学構造をクエリとして、 をホモロジーモデリングによって予測します	、その複合体の立体構造	L			419	e Jpdb_id	a ¹ identity[%] ²	description		_
[Download HOMCO	OS data]: Release of <u>PDB:20210908</u> , Relea	ase of <u>UniProt:2</u>	021 03 Release of KEGG:2021-09-06 3D of KEGG	COMPOUND and DRUG					82 <u>6yi3</u>	A 99.3	NCAP SARS2	Nucleoprotein	
									Segun	B 100.0	NCAP SARS2 N	Nucleoprotein	
	<u>新型コロナウィルスSARS-CoV-2に関</u>	連するタンパク質	と化合物のページを作りました!(PDB_UPDATE:202)	<u>10908)</u>	HETERO	ヘテロ]複合体		Tra.rashurid	for the homologue. 2. Identify	Ajsequence Identify I	between the query	and th
			- 0	×	1			419	pdb_id	contact mol		homologue	- 5
		t ask inn sai29/LAN/						I		a ³ description		a ⁴ identity[%]] ^p No
< 7 U	Inteps://noncos.pabj.org/cgi-bin/prot	Csch_inp.cgirot.And				•			\$ <u>7cr5[1]</u>	B monoclonal antibod	/ chain H[216 aa]	A /100.0	/7
	PI D					-			\$ <u>7cr5</u> [1]	C monoclonal antibod	γ chain L[213 aa]	A /100.0	15
		. 11° 5 55 - + -				◎	との複合	休 _					
	HOMCOS:9	ンハク資に刈	9 る結合分子の検系					419	pdb_id	contact mol		homologue	15
										a- description		100.0	28
[Go to English p									₩ <u>/act[</u> 1]	B SSRNA		A /100.0	/2
	0								\$ <u>7acs</u> [1]	B R(P*CP*AP*CP*UP	'GP*AP*C)-3')	A /99.3	/5
入力したタンノ	(a)	L		性		化合	物との複	[合体					
000011-C	(2) UNIPROTIDA	ハグミ	ノ酸���別を入力					· · · · ·	-jpdb_id	contact mol		homologue	15 14
アミノ酸配列に、		 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>		r)L					1			a identity[%]- INC
	これがの方法で入力してください。									in the		83.3	6
タンパク質配列の	DID NCAP SARS2	2 (CD	K3 HUMAN, Q00526, AAV40830.1,						₩ <u>6kl6[</u> 1]		-dimethyl-1-(5-	B /61.2	/6
(UniProt:ID/AC INSDC/RefSeg	protein id)).1)								phenylmethoxy-1H-	indol-3-yl)meth		
	LPYGANKDGIIW	VATEGALNTPKDHIG	TRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS										
アミノ酸配列 (-	-文字表記) QASSRSSRSRN QQGQTVTKKSAA WPQIACAPSAS KTFPPTEPKKDK	ISSRNSTPGSSRGTSP Easkkprokrtatka Saffgmsrigmevtps Kkkadetoalporok	ARMAGNGGDAALALLLDRLNQLESKNSGKGOO NNYTQAFGRRGPEOTQANFGDOELIRGGTDYKH GOTWITYTGAKLDDKDPHKDOYLLLNKHIDAY KQQTVTLLPAADLDDFSKQLQQSMSSADSTQA								T.J		

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

ソフトウエア名	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。

Molmil



Gert-jan Bekker

- ・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)

SAR	S-CoV-2	Ø 3C7	様プロ	テア-	ーゼ	PDBID) 7rfwの	観察
(1) https	s://pdbj.org0)フォーム	に 7rfw と入	力			C (c) Proc. Natarillo - 55 (2 million) × (a) - honori ← → O (b) → (b) - (b) - honorign by projection (2 million) ACCO → (b) -	x =
РОВЈСОИ (2)PDB	Pose Data Bank Jap rr Pose Data Bank Pose Data Pose Data Bank Pose Data Pos	English 日本語 简体中: 7 rfw uerres and imks B <u>Promode Elastic</u> ····································	* 繁忙中文 한국어 *	(4)ペ- [非対 をクリ で構造	ージ左の 称単位 ックする もが表示) を表示] っと、moln こされる	nil	
^{概要} 機造情報 実 構造情報 7RFW	陵情報 機能情報 相同蛋白質 屆歷 4	∞► 3)[構造情	報]のタブを	ミクリック	すると、	エントリア	りの分子の一覧	覧が表示
Structure of inhibitor	f SARS-CoV-2 main p	^{概要構造情報 構造情報 7RFW}	機能情報 相同蛋白質 履歷 夕	ウンロード				 ■ PDBx/mmCF ■ PDBML (ヘッダのみ (no-atom)) ■ PDB形式 (全でり情報) ■ ● 検証レポート (PDF) ■ ■ EDMap 2fo-fc (MTZ) More
7RFW の概要 エントリーDOI 分子名称	<u>10.2210/pdb7rfw/pdb</u> 3C-like proteinase, (1R.2S,5S) yl]propan-2-yl)-6,6-dimethyl-	Structure of SAF	S-CoV-2 main prote	ease in complex	k with a coval	ent inhibitor	Severe a cute respiratory syndrome	構造 非対称単位を表示 非対称単位を表示
^{機能のキーワード} 3C-likeプロ 3CLp, Main	azabicyclo[3.1.0]hexane-2-car protease sars-cov-2 covalent hydrolase/inhibitor テアーゼ(nsp5, 3CL protease, Mpro):タ	ァィティ SC-like protea -PRO, ンパク質	ase (nsp5, 3CL-	PRO, 3CLp,	Main pro 化学式里 分子数 33882.5 1	tease, Mpro テータベース名(アクセス衝号) UniProt (<u>PODTD1</u>)	coronavirus 2 2019-nCoV) 由来する生物種 名 Severe acute resouratory syndrome ちいのの、 Science Severe acute resouratory syndrome ちいのの、 Science Severe acute resouratory syndrome	
を切断する 体。コロナ 1abという7	」酵素タンパク質。7 ウィルスは、ポリプ 096アミノ酸からなる	トモニ量 コテイン [₩] ₩ る非常に	E,2S)-1-imino-3-[(3S)-2- [propan-2-yl]-6,6-dimethyl-3-[3- non- exane-2-carboxamide wate	comp_id "4	501.5 1 18.0	Chemie (<u>4WI</u>)	coronavirus 212019 nCovi 4WIの文字にホ インタをかざす	生物学的単位を表示 生物学的単位を表示 生物学的単位を表示
長いタンパ と、その長 個のタンパ パク質を切	り員を、より、ロル いタンパク質を切断 パク質にする。そのと 〕断する酵素の一つ	た。そのめ fiして、15 cき、タン の。治療薬 □	C-like proteinase)				~、 示される	
の候補とし 剤が開発さ	て、このタンパク質 されている。	の阻害		comp_id	l: 4WI 別名 e lvir , boun	፤ : PF-07321 d form; Pax	. 332 , bound form I ovid , bound form	パクスロビド

マウスの使いフ	与
---------	---

表示法	PyMOL	UCSF Chimera	JSmol	Molmil
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 インタをかざす と、原子名のラ ベルが表示	画面上で原子の上にマウス ポインタをかざすと、、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 20

molmilの三つのインタフェース



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例

[Chain A]→[Display]→[Ca trace]



$[Chain A] \rightarrow [Display] \rightarrow [Amino acid] \rightarrow [Space fill]$



$[Chain A] \rightarrow [Display] \rightarrow [Cartoon]$



[Chain A]→[Color]→[Structure]



分子メニューによる分子・残基ごとの表示の変更

 $[4WI - Chain A] \rightarrow [Display] \rightarrow [Amino acid] \rightarrow [Spacefill]$

[PHE 8 Chain A] \rightarrow [Display] \rightarrow [Amino acid] \rightarrow [Space fill]



化合物4WIを空間充填表示

PHE8を空間充填表示



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 #スティック表示をオフ 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&スティックで表示 label byres (resn 4WI around 5) and name CA, "%s%s"%(resn, resi) #分子4WIの周辺5Åの残基のCa原子に残基名、残基番号のラベルを表示



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 何も起きない… hide cartoon X カートゥーンが消えてくれる! hide cartoon,all \cap

2. "color"と"cartoon_color"を区別

× | color red,all

 \cap

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	システインを棒表示
[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
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
show [表示法]	show spheres, all	球モデルの非表示
hide [表示法]	hide spheres, all	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_sti	ck, カートゥーン:cartoon, バック
ボーン: ca-trace, SNFGとスティック: snf	fg-icon	
color [色]	color blue, all	青色にする
cartoon_color [色]	cartoon_color red, all	カートゥーンの色を赤にする
color [r,g,b]	color [0,255,255],all	シアン色[0,255,255]にする
color cpk	color cpk, all	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	5赤:group、二次構造による色分け:structu	ıre、温度因子による彩色:bfactor
bg_color [色]	bg_color white	背景を白に
turn [xyz], [回転角(°)]	turn y, 180	Y軸のまわりに180°回転
reset	reset	分子を元の向きに戻す
set cif_use_auth, [on \hbar `off]	<pre>set cif_use_auth,off</pre>	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
https://numon.pdbj.org/binds_workshop_202109/molmil_simple_2ht7.html
                                                                     Molmil simple page (2ht7)
                                                                                      V V V
<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, 3Dzigzaw	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³⁵



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)構造に従って全原子を構築 (1)側鎖原子の構築 (2)挿入ループ部を構築

MODELLER, SWISS-MODEL, ₃₆ RosettaCM, 3Dzigzaw
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

赤:正解構造(2trx_A)

青:予測構造





鋳型との**SeqID=52.5%** RMSD(Ca)=1.08Å 鋳型:1fb0_AThioredoxin M (*Spinacia oleracea (spinach*))



RMSD(全重原子)=1.81Å





立体構造が同定・予測できる割合

1000/	タンパ	パク質単位の	の統計	アミ	ノ酸単位の約	充計	
100%				ТЛЛ			
90%	 I IVI		TM			DO	
80%							
70%							
60%							
50%	 -	゠デリング可					
40%				-	Fデリング可		
30%					• • • •		
20%		<u>↓</u> → → → →					
10%		沢定済み			決定済み		
0%							
	大腸菌		ヒト	大腸菌		ヒト	

■立体構造が決定済
 ■ BLASTでモデリング可能
 ■ TMもDOもなし
 ■ TM
 ■ DO
 ■ TMとDO
 TM:膜貫通へリックス領域
 DO:天然変性領域(DO)

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リング可能と見なした。膜貫通へリックス領域は 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。



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<u>複合体</u>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











^{98.2} awryk 11: KAYUMAROWE I YAKAKOZAL UNIOWINI DA AWXROCOWILLING I I ZECKEL TAL HAMPIE I PARTY

11:KSSVRUKIPTRVEETICGLUKGGAALDITTOFINITLSPFSYNDKROPTOHNTUNKLVTDEOFRKLUULKEDYTATEVDPVLTVEEKFPYNVEINTKS:



→「タンパク質に対する検索」結果(ASAH1_HUMAN)

HOMCOS(prot sch conbars.cgi): X P Protein Model on 51/7z B 1 A 1	C Protein Mode	el on 5u	81 A 1 × +				- A X	
	nham ogi2 Totta			_T0	т.			
The second	nbars.cgirrep=1800	=prot_ II		= 1 &pssm:	1/39.8	1/41	lacid amidase subuni	
	88 <u>6dy3[</u> 4]	в	Q_{9GUI1} CAEEL N-acylethanolamine-hydrolyzing acid amidase beta-s	A	45.5 /35.3	33 /46	Q9GUI1_CAEEL N-acylethanolamine-hydrolyzing acid amidase alpha	
	3.a ³ :asym_id t the contact res slash / is numl	for th sidue ber o	le contact molecule. 4.a ⁴ :asym_id for the template homologue. 5.identit s. Number after the slash / is sequence identity for all the aligned regior f contact residues in the template homologue.	:y[%] ⁵ :s n. 6.Ncor	equence identity 1 ⁶ :number of alig	between ined cont	the query and the template homologue only for act residues for the query. Number after the	
▲ 化合物との複合	体の積	ŧì	<u>生</u> 目	le e	8	⊉ற்	アイコンをクリックする	
	pdb_id	COM		nc	imologue	노는	休樓浩エデルがまテ	
	<u> </u>	a ³	description	a4	identity[%]	<u>2</u> <u>7</u>	仲伸迫て アルか衣小	
	€2 <u>5u7z[</u> 3]	к	NAG 2-acetamido-2-deoxy-beta-D-glucopyranose[14 atoms]	A	100.0 /98.2	2 /2	A8K0B6_HUMAN Acid ceramidase	
_	82 <u>5u7z[</u> 3]	ĸ	NAG 	В	100.0 /99.6	3 /3	A8K0B6_HUMAN Acid ceramidase	
	80 <u>5u7z[</u> 3]	L		В	100.0 /99.6	2 /2	A8K0B6_HUMAN Acid ceramidase	
	82 <u>6mhm</u> [2]	м	hexylcarbamic acid[9 atoms]	В	100.0 /99.6	8 /8	ASAH1 HUMAN Acid ceramidase subunit beta	
	■ <mark>825u84[</mark> 3]	G	130 131 132 133 134 5-amino-2,4,6-modobenzene-1,3-dicarboxylic acid	□ ⊡ HOMCOS ← → C Protein Mo	(prot_uch_conbars.cg): x ① Protein	Model on 6mhm_8_1_ .org/cgi-bin/model3 TEMPLATE: PDI		
	■ \$ <u>\$5u84[</u> 1]	н	I3C 	DOWNLOAD: U(32)template U(32)templ				
	■ <mark>82</mark> 5u84[2]	I	I3C I I I I I I I I I I I I I I I I I I I	equence-replaced counterceptaced emplate 3D structur IOLECOLES	3D model CurrentView:HTM	Change to:	JAVA] JSmol [mailton [market] Ontact sites Contact sites	
	82 <u>5u84[1]</u>	J	I3C 1 5-amino-2,4,6-triiodobenzene-1,3-dicarboxylic acid	a A(2 b LIGNMENTS	ery (auth_asym (sauth_asym (sah1_HUMAN) B 1 (B) M 1 (B)	polymer(po [253 aa] non-polym	jdescription e[query A) alypeptide(Li) Acid ceramidase submit beta <u>Acid Lit HuMAN</u> 143C 161M 162D 163F 211L 224N er(IRY) hexylcarbamic acid 215E 320N (identity: 100.0 %/99.6 %)	
u , , , ,	■ \$\$ <u>5u84[</u> 1]	L	I3C 5-amino-2,4,6-triiodobenzene-1,3-dicarboxylic acid	WDEL[1] Problem ieryλ 143:6 coStr 3 coStr 3 intact 3 <t< th=""><th>An A "queryA" TEMPLATE: 60 75 (VAEDKKeLINSWOO GVILGINING TO VAEDKKELINSWOO GVILGINING EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE SUMMER OF LITEVLENS TS ELEANLINK SHAMMANGES SAMAN OF LITEVLENS TS YEEANNLINK SHAMMANGES ENHANCES SHAMMANGES ENHANCES SHAMMANG</th><th>nnm_B_1 identi reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLCS</th><th>VY-99-09 VY-99-09 MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION OF INTERNATIONOO OF INTERNATIONOO</th></t<>	An A "queryA" TEMPLATE: 60 75 (VAEDKKeLINSWOO GVILGINING TO VAEDKKELINSWOO GVILGINING EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE TS ELECTRIC EEELE SUMMER OF LITEVLENS TS ELEANLINK SHAMMANGES SAMAN OF LITEVLENS TS YEEANNLINK SHAMMANGES ENHANCES SHAMMANGES ENHANCES SHAMMANG	nnm_B_1 identi reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLTVHLDFG reviteGuxPLCS	VY-99-09 VY-99-09 MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION OF INTERNATION MENTIVESER AVENUES (INVESTIGATION OF INTERNATION OF INTERNATIONOO	
	\$ <u>6dxx[</u> 6]	J	TON 2-{2-[4-(1,1,3,3- TETRAMETHYI BUTYI)PHENOXYIETHOXYJE	ntact : HeryA 343:5 thm.B_1 201:5 cStr : pBur :4	ISOGENISFETNYDVLSTKPULNKLTVYTLLO ISOGENISFETNYDVLSTKPULNKLTVYTLLO ISOGENISFETNYDVLSTKPULNKLTVYTLLO ISOGENISFETNYDVLSTKPULNKLTVYTLLO ISOGENISFETNYDVLSTKPULNKLTVYTLLO	VTKGGFETYLRDCPDPC VTKGGFETYLRDCPDPC TTTTEEEEEE 88 sbeeseebebeebebeebebeebe	iar: 2000 Ilacid amidase subuni	
		-	WTE					



Site Table		ummary)[0.0 %] [60] [70] [80] [90] の画面上のこ	[95] [100] のアイコン	to Search Page] をクリック
SUMMARY Bars [Full Bars] Summary Bars [Full Bars]		Back to HOMCOS Back to HOMCOS 22;AltName: Full=Cell di 22;AltName: Full=Cell di	™ UniProtの (Feature 1	アノテーション 「able)
□ a s e acd pdb contact_mols		observed aa	feature table	variant
STE1_M_e 29.5 24fx3_A compound HHT precipitant		MLIVQ		
STE 2 D T e 76.4 4fx3 Antetero CCNA2 HUMAN CG2A HUMAN compound HHT	homo precipitant			
STE4 F E b 15.3 经4fx3 A compound HHT precipitant STE4 F E b 15.3 经4fx3 A compound HHT precipitant 結合分子	そのサマリー		DOMAIN Protein kinase. DOMAIN Protein kinase.	
STE 5 Q E b 10.7 & 4fx3 A precipitant		EQKDRTVINHLACMSY	DOMAIN Protein kinase. DOMAIN Protein kinase.	
SITE <u>6</u> K E e 45.8 $\underbrace{\& 4fx3 \ A}$ compound <u>HHT</u>		IKVLRNDEFGMQACHPST	DOMAIN Protein kinase. <u>DOMAIN</u> Protein kinase.	
(4) 押まね オいて 如佳 ()の世帯山 声。。。 ど	(2) 扣回ね、パク哲理で組図され	スマこ /融の宝	ll合(abaan	(adaa) + (+)

(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。

ASAH1_HUMANတSite_Table 1 M E e 2 E H b 壨 -----HCMCOS 3 F H b [Back to HOMCOS] [Summary Bars] [Full Bars] [Back to Search Page] Site Table[0.0 %] Variant 1 M <u>c</u> c 2 E H b P

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

tsv [show] [download] [help]

[Sites by Variants] [help]

PID	QueryLength	Homolgous Sequence in PDB	UniProt Query	TITLE			
3946607	395 33 Q13510(ASAH1_HUMAN) RecName: Full=Acid ceramidase ; Short=AC; Short=ACDase ; Short=AcDase ; Short=Acid CDase; EC=3.5.1.23 ;AltName: Full=Acylsphingosine deacylase;AltName: Full=N-acylsphingosine deacylase;AltName: Full=N-acylsphingosine deacylase;AltName: Full=N-acylsphingosine amidohydrolase;AltName: Full=Acid ceramidase is Short=PHP32;Contains: RecName: Full=Acid ceramidase subunit alpha ;Contains: RecName: Full=Acid ceramidase subunit beta ;Flags: Precursor;						
WPGRSCVALVLLAANSCAWAGHAPPUTEDCORKSTYPPSGPTYRGAVPIUTILDLPPYKBWEINDUARPULKVTWISLANIITHTYPSGKTMOVVDEKLPGL_GHPGPFEEEMIGIAAVTDIPLGETISFNIFVELFTICTSIVAED QUERYSEQ KKGHLIHGRMIDFGVFLGWNINDTWITEQLKPLTVNLDFORNNKTVFKASSFAGYVGMLTGFKPGLFELTLENFSINGGVLGUEWILGKKDWWIGFLTRTVLENSTSVEEAKNLLTKTKILAPAYFILGGNGSGEGCVITRORKE SLDVYELDAKGGRVVVGTNURDRKHPFLDDRTPAMICLANISTEVENULTIVTLIDVTKGGFTVLEDCOPPCIGW							

[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]:redicted exposed/buried. [acc]:predicted relative accesssibility(%). [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	а	s e	acc	pdb	contact_mols	observed aa	feature table		variant			
SITE <u>1</u>	М		-	-		M	SIGNAL DISORDER predicted by DISOPRED SIGNAL					
SITE 2	Ρ		-	-		LPR	SIGNAL DISORDER predicted by DISOPRED SIGNAL					
SITE 3	G		-	-		G	SIGNAL DISORDER predicted by DISOPRED SIGNAL					
SITE <u>13</u>	<u>8</u> E	ΤŁ	0.0	₿ <u>5u81</u> A	hetero <u>Q9GUI1 CAEEL</u> A8K0B6 HUMAN ASAH1 HUMAN compound <u>I3C</u> metal <u>IOD</u>	ED			E->V:(0.0 %): <u>LP/P</u> Farber lipogranulomatosis (FRBRL) <u>dbSNP:rs137853594</u> [MIM:228000]			
SITE <u>13</u>	<u>9</u> L	ΤŁ	13.5	89 <u>5u81 A</u>	hetero <u>A8K0B6_HUMAN</u> ASAH1_HUMAN compound <u>I3C</u>	ILAMVFSY		138番目の1	ナイトのまとめページ			
SITE <u>14</u>	<u>0</u> F	se	20.1	₿ <u>5u81</u> A	hetero <u>A8K0B6_HUMAN</u> ASAH1_HUMAN precipitant	<u>SFTA</u>		□ □ HOMCOS: × □ HOMCOS(prot,sch, · × ▷ P ← → C (a) https://homcos.pdbj.org/cgi-bin/p	stein Model on 6 · x D Protein Model on 5 · x H - O X = rot_sch_eachsite.cgitcon=T&pssm=T&precak_bit=T&q f_0 0 1 f= 0 0 f= 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			
SITE <u>14</u>	<u>1</u> T	E	1.3	₿ <u>5u81</u> A		ATRV.	MUTAGEN /note="T->A: Decreased rate of autocatalytic poceased rate of autocatalytic processing."	PID QueryLength FocusSite TITLE	Summary for the 138-th Site(E)			
SITE <u>1</u> 4	2 I	ΕŁ	13.0	<mark>₿25u81 A</mark>	metal <u>CL</u>	FAIMV		RecName: Full=Acid	eramidase ; Short=AC; Short=ACDase ; Short=Acid CDase; EC=3.5.1.23			
SITE <u>14</u>	<u>3</u> C	EŁ	0.0	₿ <u>5u81</u> A	compound WTE HJA JRY precipitant	<u>c</u>	ACT_SITE /note="Nucleophile" ECO:0000269 PubMed:30 autocatalytic processing. Loss of ceramidase activity." AC ECO:0000269 PubMed:30525581" <u>MUTAGEN</u> /note="C-> ceramidase activity."	3946607 395 138 E EC-a3.5.1.: ;AltName protein; Short=PHP3 Full=Acid ceramidase UniProt Information AC:Q13510 ID:ASAH1_HUMAN	Julipsanie Georpas, Facturater, Luip Leptenbuodine TryJousse Rohm 1, Julie-Nacyslani Ingsine amiddhydrolase; AltName: Fuli =Putetter 22 Kba heart ;;Contains: RecName: Fuli=Acid ceramidase subunit alpha ;Contains: RecName: subunit beta ;Flags: Precursor;			
SITE 14	4 T	Et	0.0	<mark>≌5u81 A</mark>		L		Feature Table for 138- ECO:0000269 PubMed:10993	RBRL; loss of ceramidase activity; dbSNP:rs137853594)" <u>17</u> " /id="VAR_021582"			
SITE 14	<u>5</u> S	EŁ	0.0	<u>₿5u81 A</u>		<u>s</u>		th site HELIX: CHAIN: /note="Acid ceramida:	e subunit alpha" /id="PRO_0000002312"			
SITE 14	<u>6</u> I	EŁ	0.0	<mark>₿5u81 A</mark>		IVA		VARIANT for 138-th site E->V LP/P dbSNP:rs13785359	"Farber lipogranulomatosis (FRBRL) " [MIM:228000]			
SITE 14	zν	EŁ	0.0	<mark>₿5u81 A</mark>		VI		Evolutionary Information				
SITE 14	8 A	EŁ	0.0	₿ <u>5u81</u> A		ATI		E:90% D:10%	s (show alignment)			
SITE 14	9 E	E	2.5	<u>₿5u81 A</u>	metal <u>CL</u>	QED		3D Structure Information				
SITE 15	0 D	Ľ	15.4	₿ <u>5u81</u> A	compound <u>I3C</u>	DNAEFGIKLPORSTV		Implate for Honomer predicted Second predicted Exp Implate for Honomer T (Hbond turn) b (buried)				
SITE 15	1 K	Те	61.3	<u>₿5u81 A</u>	precipitant	SKNEGADFILPORTV		3D Complex Information				
SITE <u>15</u>	<u>2</u> K	Te	88.7	₿ <u>5u81</u> A	precipitant	KRADQEFGILNPSTV		Tended Bind Polecies Petero:9 compound:1 metal:1 Templates for 3D complexes hetero [52802:A3K0B6 HUMAN] 經5u7z A 1 B 1 [49694:Q9GUI CAEEL] 迎后dy3 A 1 B 1 迎后dy3 C metal [IOD 增5u54 B 1 V 1	₩5u7z C 1 D 1 ₩5u7z C 2 D 2 ₩6mhm A 1 B 1 ₩6mhm C 1 D 1 1 D 1 ₩6dy3 E 1 F 1 ₩6dy3 G 1 H 1 compound [I3C] ₩5u84 A 1 L 1			
SITE 15	3 G	11	14 3	25u81 A		GAELSYDEHIKMNPORTY		,				

Import Information Import Information Import Information RecKamer Juli-Acylphingosine decylase; Althome: Full-N-acylethanolamine to the second se	21 13	8番目の部位のまとめ(ASA	
Import Information All Name: Full - Acid ceramidase c; Short - AC; Short - ACDase ; Short - Acd Case ; Sho			
Image: Second			statistics of also be deer of Obsease cases/finations: Uneit/Popelに記載されてした:変異体のサマリーが最高される
DetAil Model on 5 Protein Model on 5			
Image: Section in the producted back is decided as a subset in the producted back is a subset in the producted back in the producted back is a subset in the producted back in the producted	🗖 💾 номсоs:	x 🗅 HOMCOS(prot_sch_c x 🗅 Protein Model on 6 x 🗅 Protein Model on 5 x 🗅 Protein Model on 5 x	s x sp A5A6P2 ASAH1
SUFFE Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) PID QueryLength FocusSite TITLE 3946007 395 Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 3946007 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) 394607 1mage: Summary for the 138-th Site(E) Image: Summary for the 138-th Site(E) Arribus: Information E-> V	\leftarrow \rightarrow C a	🖞 https://homcos.pdbj.org/cgi-bin/prot_sch_eachsite.cgi?con=T&pssm=T&precalc_blt=T&q 🏠 🚱	sp A0A3832FX3 A g170B3 ASAH1
PID QueryLength FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to FocusSite TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to Focus TITLE 3946607 395 IBE ECNIMINARY BIT ACCESS to Focus TITLE 3946607 SE SE VILVE ACCESS to Focus TITLE 3946607 SE SE VILVE ACCESS to Focus SE 4010 ECNIMINARY BIT ACCESS to Focus SE SE SE SE 4010 ECNIMINARY BIT ACCESS to Focus SE SE SE SE		CITE Summary for the 138-th Site(E)	sp ADAOP6JG37 A sp DGP751 ASAH1
PID QueryLength Focussite ITTLE 3946607 395 Ital RecName: Full=Acid ceramidase ; Short=AC; Short=ACDase ; Short=Acid Coase Short=Acid Coase 3946607 395 Ital RecName: Full=Acid spiningosine deacylase; AltName: Full=N-acylethanolamine h Short=Acid Coase 3946607 395 Ital RecName: Full=Acylsphingosine deacylase; AltName: Full=N-acylethanolamine h Short=Acid Coase JaiProt Information Ital RecName: Full=N-acylsphingosine amidohydrolase; AltName: Full=N: JmiProt Information :E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL)" [MIM:228000] Ac/ID Acia::scie:::scie::scie:::scie::scie:::scie:::scie:::scie::sc		SILE summary for the 150 th Site(L)	
Image: Section of the section of th	PID Ouervl ength	FocusSite TITLE	SP 009071 ASAHZ CAEEL[1] 137:LNLFTEMSRAFC: 147 SP 09D7V9 NAAA MOUSE[1] 121:VNLAYEASAFC: 131 SP 05KTC7 NAAA RATE1 121:VNLAYEASAFC: 131
3946607 395 18 E :AltName: Full=Acylsphingosine deacylase; AltName: Full=N-acylethanolamine h () <th></th> <th>RecName: Full=Acid ceramidase : Short=AC: Short=ACDase : Short=Acid CE</th> <th>sp 090111 NAAA CAEEL[1] 103:MNIL/DVAAFC: 122 Dase sp HOVCJ6 NAAA CAYPO[1] 113:VNI AYFASAFC: 123</th>		RecName: Full=Acid ceramidase : Short=AC: Short=ACDase : Short=Acid CE	sp 090111 NAAA CAEEL[1] 103:MNIL/DVAAFC: 122 Dase sp HOVCJ6 NAAA CAYPO[1] 113:VNI AYFASAFC: 123
3946607 395 138 E EC=3.5.1.: ;AltName: Full=N=acytsphingosine amidohydrolase;AltName: Full=PLise Control Proc Name (Part Alter		;AltName: Full=Acylsphingosine deacylase;AltName: Full=N-acylethanolamin	ne h sp G117U7 NAAA RABIT 1 117:VNLAYEYSAFC: 127 sp Q02083 NAAA HUMAN 1 116:VNLAYESSVFC: 126
文ipirot Information 変異体の情報 AC/ID E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:228000] MAC/ID MAIANT: /note="E-> V (in FRBRL; loss of ceramidase activity; dbSNP/rs137853594) Feature Table for 138- th site CO:0000269 PubMed:10993717" /id="VAR_021582" HLIX: CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_000002312" VARIANT for 138-th is E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:228000] Volutionary Information 相同配列群のアミノ酸頻度 Percentages of Amino Acids in Homologous Proteins [mills] [how alignment] E:30K D:10K BD Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted & 5u81 To (<i>Hoon turn</i>) b (<i>buried</i>) 0.0	3946607 395	138 E EC=3.5.1 ;AltName: Full=N-acylsphingosine amidohydrolase;AltName: Full=	=PL sp <u>Q5UR76 YR632_MIMIV[1]</u> 86:LQLVY <u>E</u> ITAAC:96
JaiProt Information :E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:228000] AC/ID MAILANT: /note="E-> V (in FRBRL; loss of ceramidase activity; dbSNP/rs137853594)" Feature Table for 138- th site MAILANT: /note="C-> V (in FRBRL; loss of ceramidase activity; dbSNP/rs137853594)" VARIANT for 138-th site E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:22800] VARIANT for 138-th site E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:22800] Evolutionary Information H同配列群のアミノ酸頻度 Percentages of Amino Acids in Homologous Proteins mill [show alignment] E:30% D:10% Chalper Information BD Structure Information Chalper Predicted ExpBuP Predicted ExpBuP Predicted ExpBuP redicted ExpBu		変異体の情報	
AC/ID PC-S_213310 10.ASAFT_FORMAN Feature Table for 138- th site VARIANT: /note="E -> V (in FRBRL; loss of ceramidase activity; dbSNP/rs137853594)" EC0:0000269]PubMed: <u>10993717</u> "/id="VAR_021582" HELIX: CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_000002312" 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 (Sign Show alignment) E:00% 0:10% BD Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted BD Complex Information Predicted Bind Molecules	UniProt Information	: E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis	(FRBRL)"[MIM:228000]
Feature Table for 138- th site Image: Character Control of the	AC/ID		· · · · · · · · · · · · · · · · · · ·
Feature Table for 138- th site ECO:0000269[PubMed: <u>10993717</u> "/id="VAR_021582" HELLX: CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_000002312" 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 Site: [Show alignment] E:30X 0:10X Chalms: / (Hbond turn) b (buried) 0.0 BD Structure Information Chalber for Monomer predicted SecStr predicted ExpBur Predicted Chabry Arb		VARIANT: /note="E -> V (in FRBRL; loss of ceramidase activity; dbSNP:rs137853594)	Y
Interview CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_000002312" VARIANT for 138-th site E->V LP/P dbSNP:rs137853594 "Farber lipogranulomatosis (FRBRL) " [MIM:228000] Evolutionary Information Homologous Proteins E:00% D:10% Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted Schow alignment] 8D Structure Information Chicket ExpBur Predicted Template For Monomer predicted SecStr predicted ExpBur Predicted Schow Alignment 8D Complex Information Course Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted Schow Alignment 8D Complex Information 0.0 Predicted Bind Molecules 0.0	Feature Table for 138-	ECO:0000269 PubMed: <u>10993717</u> " /id="VAR_021582" HELIX:	C C
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 () (State 1) E:90% D:10% BD Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted State 1 T (Hbond turn) b (buried) 0.0 BD Complex Information D.0 Recent Bind Molecules Counce State 100 (Counce Counce Acids Counce Counce Acids Coun		CHAIN: /note="Acid ceramidase subunit alpha" /id="PRO_000002312"	Protein Model on Su84_A_1_L_1 focusing on 138-th site (TEMPLATE: PDB/Su84)
site Evolutionary Information 相同配列群のアミノ酸頻度 Percentages of Amino Acids in Homologous Proteins E:30% D:10% 3D Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted 路合サイトとする複合体 3D Complex Information Predicted Bind Molecules Predicted Bind Molecules	VARIANT for 138-th	E->V LP/P dbSNP:rs137853594 "Farber linograpulomatosis (FRBRL) " [MIM:228000]	Useruencereslaced 3D modell (for PMOL)
Evolutionary Information 相同配列群のアミン酸頻度 Percentages of Amino Acids in Homologous Proteins (1) 20 X 20 Structure Information 3D Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted (2) Su81 T (Hbond turn) b (buried) 0.0 3D Complex Information 0.0 BD Complex Information 0.0 Predicted Bind Molecules 0.0	site		(3D temolate)
Percentages of Amino Actos in Homologous Proteins E:30% D:10% 3D Structure Information Template For Monomer predicted SecStr predicted ExpBur Predicted & 5u81 T (Hbond turn) b (buried) 0.0 BD Complex Information Predicted Bind Molecules Predicted Bind Molecules	Evolutionary Informat	ion 相同配列群のアミノ酸頻度	ALL MOLECULES IN THE BIOLOGICAL UNIT: Millionarchite
20 Structure Information これらをクリックすると、 Template For Monomer predicted SecStr predicted ExpBur Predicted 結合サイトとする複合体 シュリン T (Hbond turn) b (buried) 0.0 BD Complex Information D.0 Note that the sector of the secto	Fercentages of Amino A	Class II Homologous Proteins area [snow alignment]	Window size : [small] [medium] [large]
Template For Monomer predicted SecStr predicted ExpBur Predicted Predicted ExpBur Predicted たんちをクリクワクすると、て FocuseD striss template(Suid A auth, asym_ud.A) SuB1 T (Hbond turn) b (buried) 0.0 BD Complex Information Predicted Bind Molecules Image: Dispan: Disp	E:90% D:10% 3D Structure Informat		Isequence-replaced 3D model Current View: HTMLS Change to: [<u>AVA</u>]
	Template For Monomer	predicted SecStr predicted ExpBur Predicted 会社 ステレーティング ると、	FOCUSED SITES (template(5v84_A auth_asym_id:A)) SiteNum AA Bell_auth_asym_id:A) SiteNum AA Set Exp/Bur RelAcc
BD Complex Information model mark query asym_id oper (auti, asym_id) type description aquery A) Predicted Bind Molecules 1 a A(ASAHI HUMAN) A 1 (A) polymer(polypeptide(L)) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) A (A) acception of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Acid ceramidase Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) Image: Application of the polymer (auti, asym_id) <td< td=""><th><mark>\$25u81</mark></th><td>T (Hbond turn) b (buried) 0.0 応告リイトと9 る復告1本</td><td>I38 Ø E G e 20.6 MOLECULES contact sites contact sites</td></td<>	<mark>\$25u81</mark>	T (Hbond turn) b (buried) 0.0 応告リイトと9 る復告1本	I38 Ø E G e 20.6 MOLECULES contact sites contact sites
Predicted Bind Molecules	3D Complex Informat	on	model mark query asym_id oper (auth_asym_id) type description a(query A)
5-amino-2.4.6-	Predicted Bind Molecule	'S	1 Z a A(ASAH1 HUMAN) A 1 (A) polymer(polypeptide(L)) Acid ceramidase [352 aa] :ASAH1 BALAS
hetero:9 compound:1 metal:1	hetero:9 compound:1 r	netal:1	2 Z b L1 (A) non-polymer(<u>I3C</u>) 5-arnino-2,4,6- triiodobenzene-1,3- dicarboxylic acid
Templates for 3D complexes	Templates for 3D comp	lexes	
hetero [52802: <u>A8K0B6_HUMAN]</u> ²² 507z A 1 B 1 ²² 507z C 1 D 1 ²² 507z C 2 D 2 ²² 6mhm A 1 B 1 ²² 6mhm C 1 / 1	hetero [52802: <u>A8K08</u>	B <u>6 HUMAN]</u> [©] <u>5u7z A 1 B 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 2 D 2</u> [©] <u>6mhm A 1 B 1</u> [©] EL 1 [©] 5u7z A 1 B 1 [©] 5u7z C 1 D 1 [©] 5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] EL 1 [©] 5u7z A 1 B 1 [©] 5u7z C 1 D 1 [©] 5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] 5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] 5u7z C 1 D 1 [©] 5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] 5u7z C 1 D 1 [©] 5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] 5u7z C 1 D 1 [©] <u>5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 2 D 2 [©] <u>6mhm A 1 B 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u>5u7z C 2 D 2</u> [©] <u>6mhm A 1 B 1</u> [©] <u>5u7z C 1 D 1</u> [©] <u></u></u></u></u>	
retal [IOD] @5u84 B 1 VA 1	metal [IOD] \$5u84 [EEJ®ouys A I B I ©ouys C I D I ©ouys E I F I ©ouys G I H I compound 3 1 VA 1	51 S1



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	7		
Buired or	Buried	57.7 (%) [213]	50.0 (%) [1]	83.3 (%) [10]	57.1 (%) [4]		
Exposed	Exposed	42.3 (%) [156]	50.0 (%) [1]	16.7 (%) [2]	42.9 (%) [3]		
Ave relacc		23.6 %	46.9 %	8.1 %	30.7 %		
SD relacc		26.28 %	41.74 %	11.79 %	27.81 %		
	hetero	39.5 (%) [156]	50.0 (%) [1]	57.1 (%) [8]	28.6 (%) [2]		
	nucleotide	0.0 (%) [0]	0.0 (%) [0]	0.0 (%) [0]	0.0 (%) [0]		
	compound	20.3 (%) [80]	0.0 (%) [0]	35.7 (%) [5]	0.0 (%) [0]		
Contact Mol	metal	25.3 (%) [100]	0.0 (%) [0]	42.9 (%) [6]	14.3 (%) [1]		
	otherpoly	10.4 (%) [41]	0.0 (%) [0]	7.1 (%) [1]	14.3 (%) [1]		
	homo	5.8 (%) [23]	0.0 (%) [0]	0.0 (%) [0]	28.6 (%) [2]		
	precipitant	24.1 (%) [95]	50.0 (%) [1]	28.6 (%) [4]	0.0 (%) [0]		
Number of variants		25	3	15	7		
N_Freq(AAvariant)==0 %			33.3 % [1]	86.7 % [13]	28.6 % [2]		
N_Freq(AAvariant)>0 %			66.7 % [2]	13.3 % [2]	71.4 % [5]		
Ave Freq(AAvariar	nt)		6.7 %	0.9 %	14.6 %		
SD Freq(AAvarian	:)		8.73 %	2.75 %	12.00 %		

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



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、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。



cons_compound={ADP,AMP,ANP,ATP,FAD,FES,FMN,GDP,GNP,GSH,GTP,HEM,NAP,NDP,SAH,SAM}

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)



DeepMind

- ・ 鋳型を使うが、 鋳型がない場合でも 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





- ・「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
- ・ 鋳型予測を行うため、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)
- 鋳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 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

AlphaFold

Protein Structure Database

(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 AlphaFold DB 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
L	248-256	NP_BIND	/note="ATP"
I	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
1	622-638	COMPBIAS	/note="Basic and
	685-703	COMPBIAS	/note="Polar res
	735-753	COMPBIAS	/note="Polar res
1	758-772	COMPBIAS	/note="Basic and
	870-894	COMPBIAS	/note="Basic and
	363-363	ACT_SITE	/note="Proton a
<u> </u>	271-271	BINDING	/note="ATP"
	1-1130	DISORDER	predicted by DIS
	天伏	変性領	國家副
MONOMER	>>m;		

結开りつfoO				1130					
业内主2100	1				pc	lb_id	a1	identity[%] ²	description
					82	2fo0	Α	99.8	ABL1 HUMA
					82	<u>6amv</u>	Α	99.1	ABL1 HUMA
					88	<u>1zzp</u>	Α	99.1	ABL1 HUMA
[- (4			

Model Confidence:

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

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







Expected position error (Ångströms)

HHIP_HUMAN :Hedgehog-interacting protein (AC: Q96QV1) AlphaFold DB HOMCOS

UniProt Feature Tables [Q96QV1(HHIP_HUMAN)]

700			
	region	name	description
	1-17	SIGNAL	
	18-700	CHAIN	/note="Hedgehog-intera
-	607-634	DOMAIN	/note="EGF-like 1"
-	635-667	DOMAIN	/note="EGF-like 2"
	376-388	REGION	/note="Interaction with
I	383-383	METAL	/note="Zinc; shared wit
	1-683	DISORDER	predicted by DISOPRED
● ^{MONOMER}]日:20	009–06	-23) escription
	8 <u>3ho5</u>	B 99.8	HHIP HUMAN
	8 <u>7pgk</u>	A 100.0	HHIP HUMAN
	1.a ¹ :asym_i	id for the hom	wologue. 2.identity[%] ² :sequer



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。 ⁷⁰⁰ ¹⁰⁰ ²⁰⁰ 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





HHIP_HUMAN :Hedgehog-interacting protein (AC: Q96QV1)



HHIP_HUMAN :Hedgehog-interacting protein (AC: Q96QV1)



RMSD(Calpha)=2.01Å

※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





Soluble domain 鋳型無しのはず

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 TM domain TM domain TM domain 412-668 412-668 412-668 7回膜貫通型のGPCR TM domain は相同な (identity=28%) 鋳型がある 予測構造 正解構造

RMSD(Calpha)=1.57Å

GP158_HUMAN : Probable G-protein coupled receptor 158



RMSD(Calpha)=3.84Å

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

MKS3 HUMAN : Meckelin (AC: Q5HY8)

MONOMER



Expected position error (Ångströms)

C

۲

MKS3_HUMAN : Meckelin (AC: Q5HY8)





MKS3_HUMAN : Meckelin (AC: Q5HY8)



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

