PDBjing&創薬等PF情報拠点VaProS第4回利用講習会「生命科学のための立体構造データ・ビッグデータの使い方入門」

UCSF ChimeraとMODELLERを用いたホモロジー・モデリング

川端 猛 (大阪大学・蛋白質研究所・特任研究員) kawabata@protein.osaka-u.ac.jp

2016年3月15日(火) 大阪大学 蛋白質研究所 1階講堂

今日の内容

1. ホモロジー・モデリング法とは

2. 配列から相同な立体構造の検索

3. UCSF Chimeraによる配列と立体構造のアラインメント

4. Modellerを用いたホモロジー・モデリング

立体構造予測法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	非経験的方法 Ab initio 予測法
	鋳型ベース予測法	De novo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,

ホモロジー・モデリングによる3次構造予測

<u>原理: 立体構造はアミノ酸配列より保存しやすい.</u>

立体構造データベース

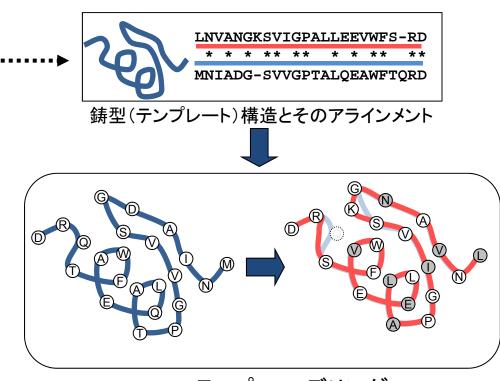
テンプレート
構造

予測対象配列 (クエリ配列)

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....



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)挿入ループ部を構築

MODELLER, SWISS-MODEL, RosettaCM, 3Dzigzaw

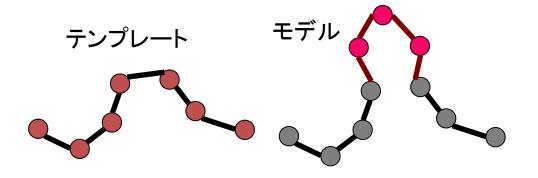
モデリング

鋳型(テンプレート)構造を元にした全原子の構築

(1)ループの構築

Sequence ALIMSTKGFVS

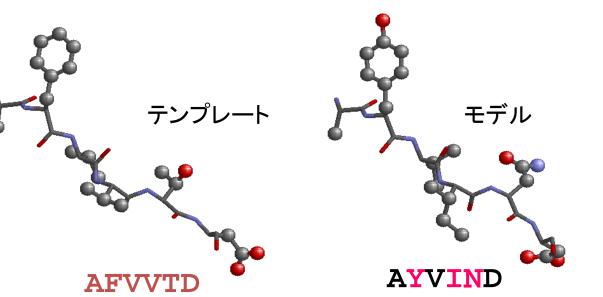
Structure LLLM---GFIT



(2)側鎖原子の構築

Sequence AYVIND

Structure AFVVTD

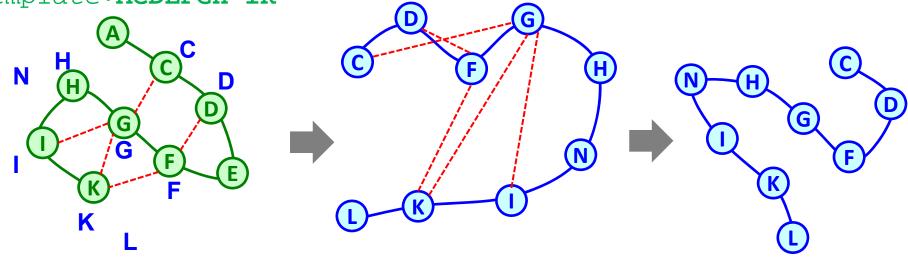


MODELLER, SWISS-MODEL, RosettaCM, 3Dzigzaw

MODELLERの手続き:空間拘束の充足

Target :-CD-FGHNIKL

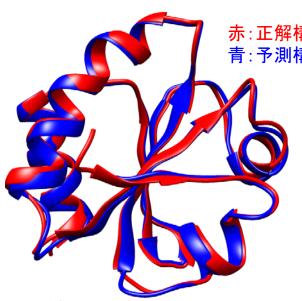
Template: ACDEFGH-IK-



- アラインメントする
- 1. 標的配列を鋳型構造に 2. 空間的拘束(spatial constraint) 3. 空間的拘束を満たすような を抽出する
 - 構造を探す
- 空間的拘束 = 相同な鋳型構造から得られる拘束 + 立体化学的な拘束
- ・空間的拘束は、CHARM22の力場と似た形式の目的関数(ポテンシャルエネルギー)に変換される
- ・構造探索は、1) 可変目的関数法+共役勾配法, 2)焼きなまし法+分子動力学法 で行う。
- ・ループの構築や側鎖の構築の手続きは、この「空間的拘束の充足」の手続きに含まれる。
- B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 5.6.1-5.6.32, 2014.
- A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

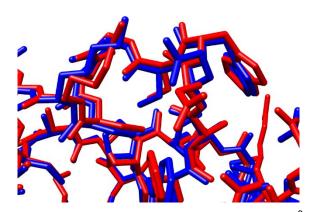
モデル結果の例: 予測対象 2trx_A Thioredoxin (*Escherichia coli*)

アラインメント: BLAST, モデリング: MODELLER

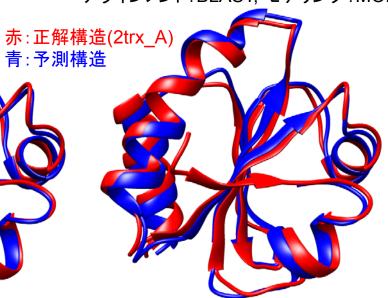


鋳型との**SeqID=86.5%** RMSD(Ca)=0.64 Å

鋳型: 2yn1_A LGPCA Thioredoxin

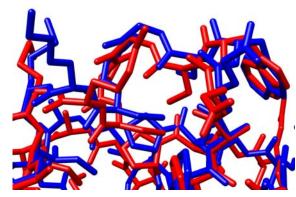


RMSD(heavy atom)=1.29 Å

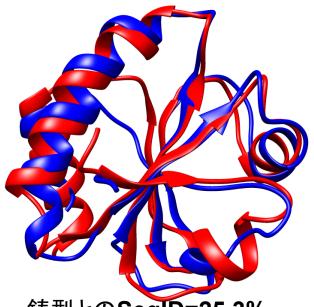


鋳型との**SeqID=52.5%** RMSD(Ca)=1.08 Å

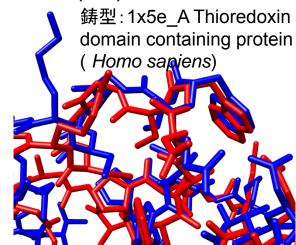
鋳型:1fb0_A Thioredoxin M (Spinacia oleracea (spinach))



RMSD(heavy atom)=1.81 Å

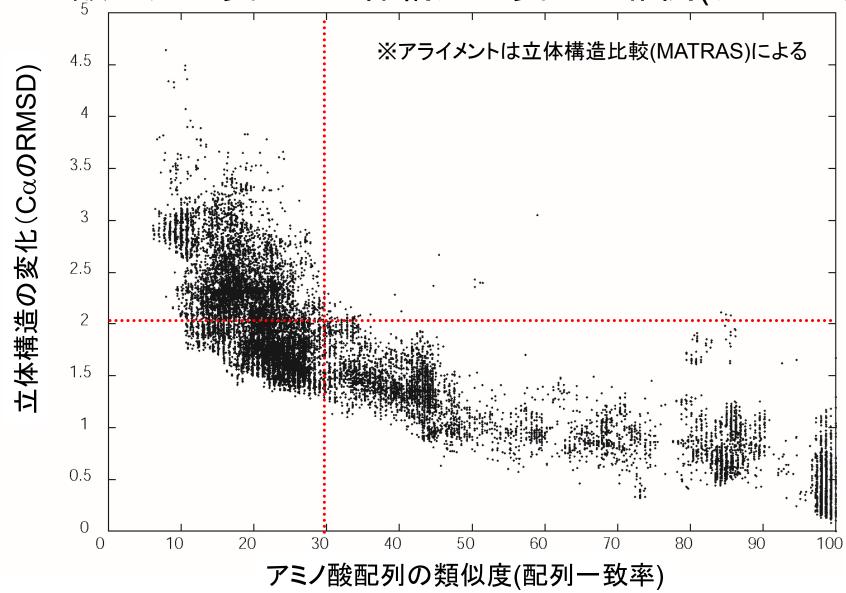


鋳型との**SeqID=25.3%** RMSD(Ca)=2.30 Å



RMSD(heavy atom)=2.83 Å

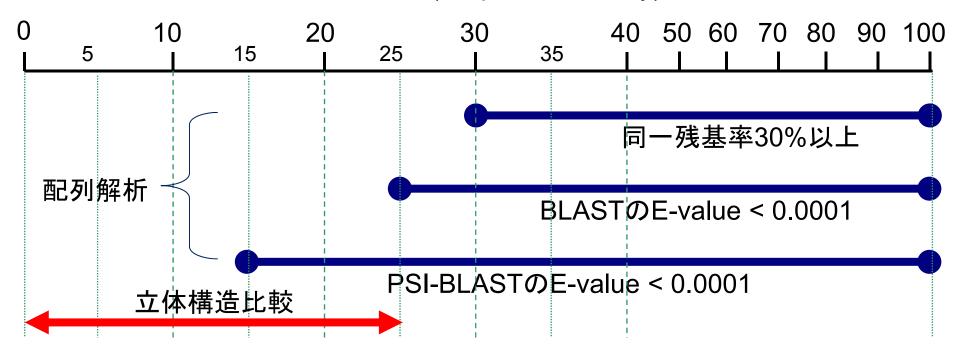
アミノ酸配列の変化と立体構造の変化の相関(グロビン族)



立体構造の変化はアミノ酸配列の変化と相関。配列一致率が30%以上であれば、2.0 Å以下のずれ

相同なタンパク質の発見する方法

同一残基率(Sequence Identity)(%)



- ・BLASTでヒットしない場合でも、プロフィール法(PSI-BLAST, HMMer, HHsearch, HHBlitsなど)では、ヒットする場合がある。
- ・配列類似度が低い場合、鋳型の相同タンパク質が見つかっても、予測精度には限界があるので、使用目的には注意が必要。
- ・立体構造比較は、複合体のホモロジー・モデリングでは役に立つことがある。

モデルの精度とその使い方

鋳型との配列一致率 (Sequence Identity)	モデルの精度 CaのRMSD	使い方 (Applications)
50 ~ 100 %	1.0 Å	触媒機構の研究 リガンドの設計・改変 高分子のドッキング 結合蛋白質の予測 仮想スクリーニング・低分子のドッキング
30 ~ 50 %	1.5 Å	抗体のエピトープの同定 X線結晶解析の分子置換 キメラ体の設計 より安定で結晶化容易な変異体の設計 部位特異的変異体の解釈 NMR構造の精密化 低解像度電子密度マップへのフィッティング
30 %以下	3.5 Å	疎な実験データからのモデリング 立体構造類似性からの機能推定 保存された表面残基のパッチの同定 3Dモチーフによる機能部位の発見

Baker, D., Sali, A. Science (2001), 294, 93-96

UCSF ChimeraとModellerを用いた ホモロジー・モデリング

配列から相同立体構造の取得

標的(予測対象)とするアミノ酸配列: UniProtの CALL5_HUMAN

CALL5_HUMAN: Calmodulin-like protein 5

SQ SEQUENCE 146 AA; 15893 MW; 70746291268494CC CRC64; MAGELTPEEE AQYKKAFSAV DTDGNGTINA QELGAALKAT GKNLSEAQLR KLISEVDSDG DGEISFQEFL TAAKKARAGL EDLQVAFRAF DQDGDGHITV DELRRAMAGL GQPLPQEELD AMIREADVDQ DGRVNYEEFA RMLAQE

二つの変異体がUniProtに記載されている

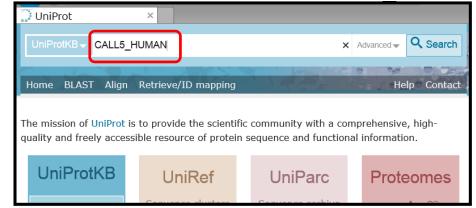
※ちなみに、ヒトの有名なカルモジュリンは、CALM_HUMANで、CALL5とは50%ほどの配列一致率。

アミノ酸配列の取得と検索

1) Googleで"UniProt"と入力

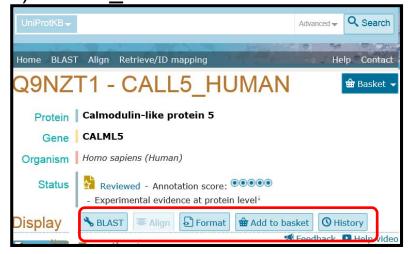


2) UniProtのページのフォームに"CALL5_HUMAN"と入力



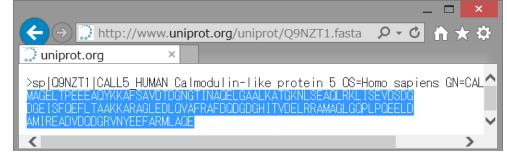
4) メニューの[Format]から"FASTA(canonical)"を選ぶ

3) CALL5 HUMANのページ





5)一文字表記のアミノ酸配列が表示される。 これをマウスで選択し、コピーする。



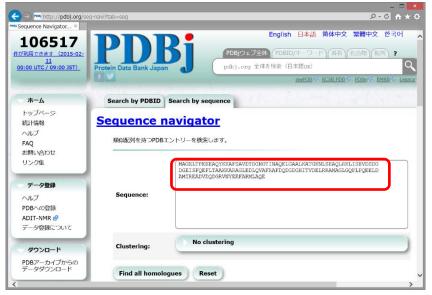
PDBjによる相同な立体構造(鋳型構造)の検索

1) Googleで"PDBj"と入力

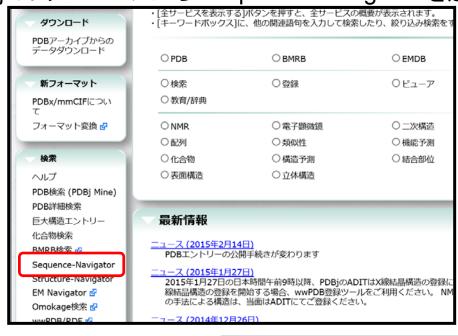


3) [Search by sequence]のタブを選び、フォームに、UniProtのページでコピーしたCALL5_HUMANの配列をペースト

2015年1月27日の日本時間午前9時以降、PDBiのADITはX線結晶構

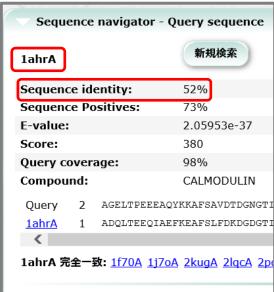


2) PDBjのトップページから"Sequence Navigator"を選択



4) 対PDBの BLAST検索の結果が表示される。 PDBコード"**1ahr**" の**A**鎖が、 sequence identity 51%でヒット。

→これを鋳型とする

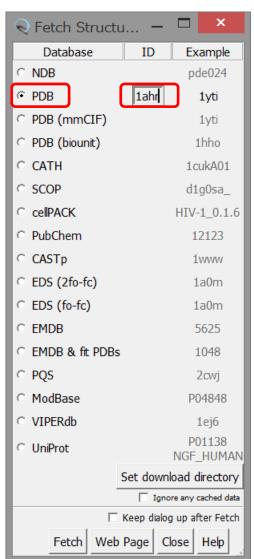


マウスの操作の方法の確認

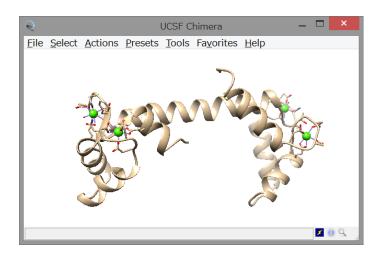
	RasMol	Jmol	UCSF Chimera
分子の回転	左ボタンで画面をドラッグ	左ボタンで画面をドラッグ	左ボタンで画面をドラッグ
分子の並進	右ボタンでドラッグ	Ctrlキーを押しながら、右ボ タンで画面をドラッグ	ホイール(中ボタン)で画 面をドラッグ
ズームイン・ア ウト	Shiftキーを押しながら、左 ボタンで画面をドラッグ	Shiftキーを押しながら、左 ボタンで画面をドラッグ あるいはホイールをまわす	右ボタンでドラッグ ある いは、ホイールをまわす
分子の断面表 示	Ctrlキーを押しながら、左 ボタンで画面をドラッグ		マウス操作だけではできない。[Tools]→[Viewing Controls]→[Side View]
マウスによる 原子名の確認	画面上で原子をクリックすると、原子名がコマンドラインウィンドウに表示される	画面上で原子をクリックすると、原子名がコンソール ウィンドウに表示される	画面上で原子の上にマウスポインタをしばらくかざ しておくと、原子名のラベルが表示される
その他		右ボタンドラッグでメニュー が表示される。	

Chimera: 鋳型構造の読み込み

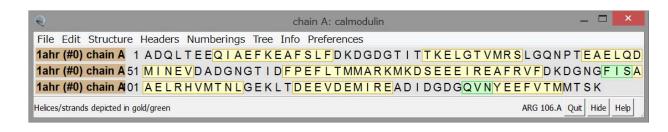
1) Chimeraを起動して、メニューから [File]→[Fetch by ID...]を選ぶ。



- 2) [PDB]を選択、IDのフォームに"1ahr"と入力し、[Fetch]をクリック。
- 3) 左図のような構造が表示されるはず。緑色の球はカルシウムイオン。



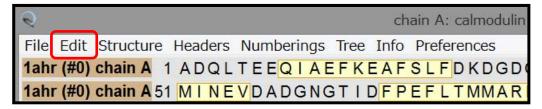
4) メニューから[Tool]→[Sequence]→[Sequence] を選ぶ。と以下のようなSequenceウィンドウが表示される。



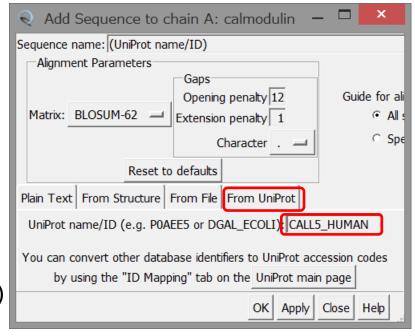
Chimera:標的配列の読み込み

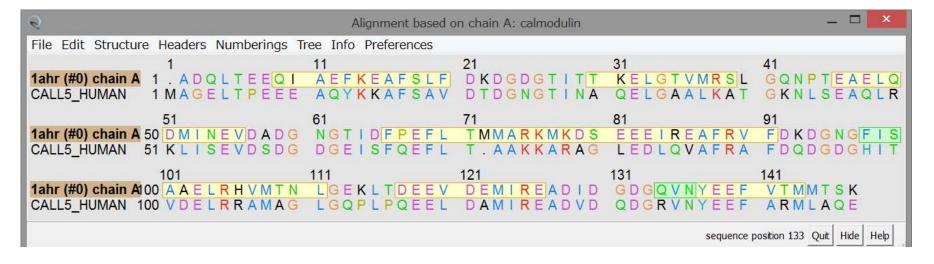
1) Sequenceウインドウのメニューから [Edit]→[Add Sequence...]を選ぶと

Add Sequenceのウィンドウが表示される。



- 2) Add Sequeceウィンドウから、[From UniProt]の タブを選択し、UniProt name/IDのフォームに、 "CALL5_HUMAN"と入力し、[OK]をクリック。
- 3) 以下のような構造(1ahr)と配列(CALL5_HUMAN) のアラインメントが表示される。

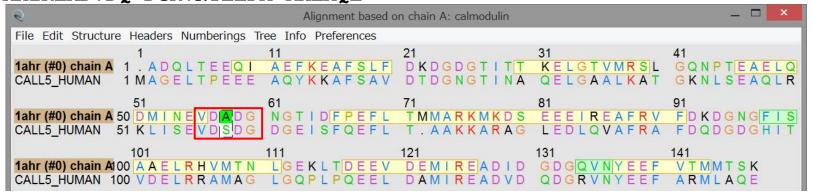




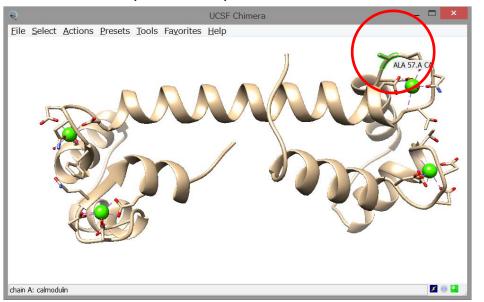
Chimera:変異箇所の立体構造の確認

SNPが報告されている 58番目のS (S->G)の立体構造上の位置を確認してみる。 VARIANT 58 S -> G (polymorphism confirmed at protein FT level).

MAGELTPEEE AQYKKAFSAV DTDGNGTINA QELGAALKAT GKNLSEAQLR KLISEVDSDG DGEISFQEFL TAAKKARAGL EDLQVAFRAF DQDGDGHITV DELRRAMAGL GQPLPQEELD AMIREADVDQ DGRVNYEEFA RMLAQE



1) 58番目のS (VDSDG)を探し、それに対応する構造部位(この場合はA)をマウスで選択する。



2) 選択された状態で、 [Actions]→[Atoms/Bonds] →[Show] とすると、選択された構造部位がスティック表示される。

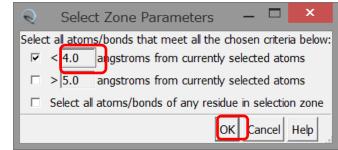
※同様に74番目のK -> R の位置も確認してみる

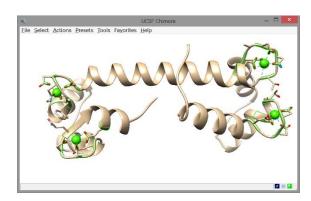
Chimera:標的配列のCa²⁺結合部位の推定

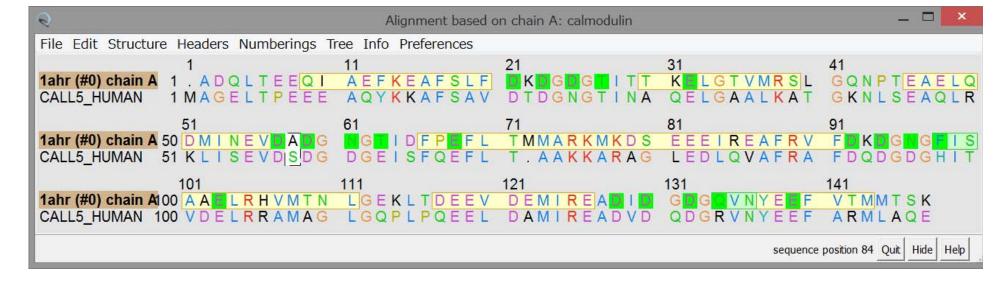
鋳型立体構造(1ahr)のCa²⁺イオンの結合部位を求め、sequenceウィンドウで対応する標

的配列の部位を確認すればよい。

- 1) メニューから[Select]→[Residue] →[CA]を選択し、 Ca²⁺イオンを選択。
- 2) メニューから[Select]→[Zone...] を選択する。
- 3) Select Zone Parameterのウィンドウが表示される。 一番上のフォームの"5.0"を"**4.0**"に書き直して、[OK] をクリックする。
- 4) 選択された状態で、[Actions]→[Atoms/Bonds]
- →[Show] で、Ca²⁺結合部位がスティック表示される。
- 5) 選択された状態で、sequenceウィンドウを確認すると、Ca²⁺結合部位が緑色で強調表示されている。

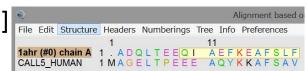




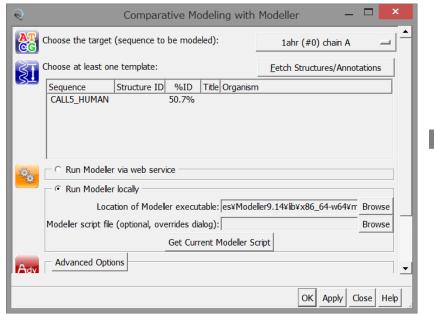


Modellerによるホモロジーモデリング(1)

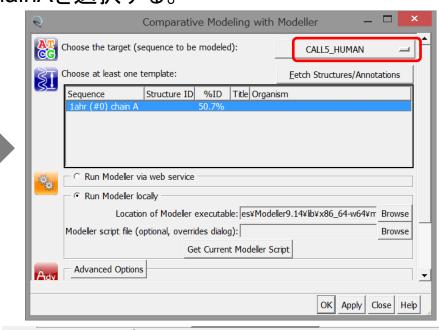
1) sequenceウィンドウの[Structure]→[Modeller(homology)...] File Edit Structure Headers Numberings Tree Info Preferences を選択



2) Modellerウィンドウの、Choose the targetを"CALL5 HUMAN"とし、 Choose at least one template:を1ahr(#1) chainAを選択する。



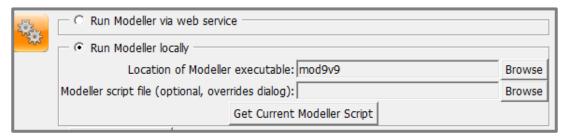
3) [Advanced Options]をクリックし、 [Number of output models]を1とし、 [Include non-water HETATM residues from template]を☑する。



Advanced Options Number of output models: 1 (max 1000) Include non-water HETATM residues from template: Include water molecules from template: Build models with hydrogens: ☐ (warning: slow) Use fast/approximate mode: ☐ (produces only one model) Use thorough optimization: (recommended with MDA) Temporary folder location (optional): **Browse** Distance restraints file (optional): **Browse**

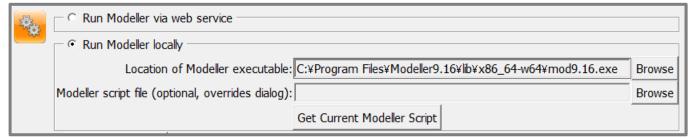
ローカルにModellerを起動するための注意

Location of Modeller executableを設定する必要があります。



デフォルトではmod9v9に なっていますが、このままでは 動きません。

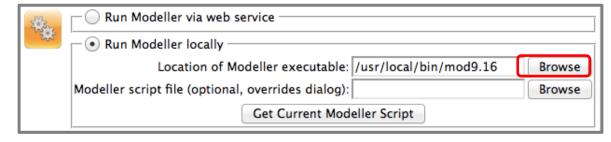
Windowsの場合の設定例



[Browse]をクリックして フォルダを移動し、 Modellerの 実行ファイルを選択

C:\Program Files\Modeller9.16\Plib\x86_64-w64\Prod9.16.exe

Macintoshの場合の設定例



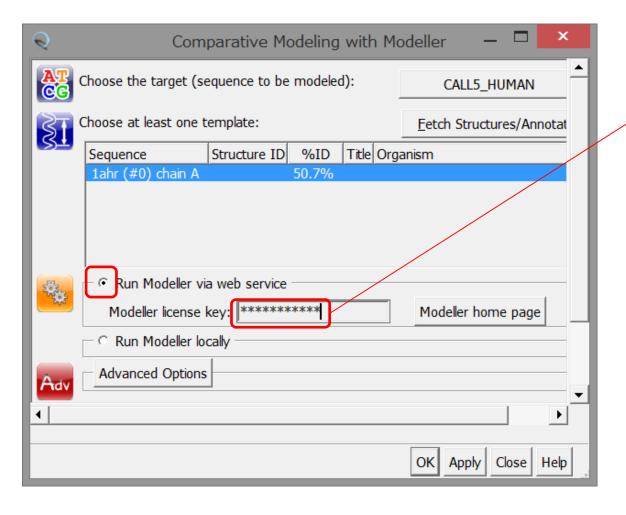
[Browse]をクリックして フォルダを移動し、 Modellerの 実行ファイルを選択

/usr/local/bin/mod9.16

※バージョンやインストール場所によって詳細は異なります。各自の設定に合わせてください。

Web Serverを利用する場合

ローカルのModellerを起動できない場合、Chimeraの開発グループが用意したWeb serverを利用することができます。



アカデミックライセンスの <u>ライセンスキー文字列</u>を 入力する必要があります。

アカデミックの方が、ライセンスキーを取得するには
http://salilab.org/modellerにアクセスし、[Registration]から、ユーザー情報を入力してください。しばらくすると、ライセンスキーの文字列が電子メールで送付されます。

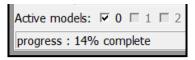
※Web Serviceを利用した場合も、ローカルに起動した場合も以後の手続きは同じです。

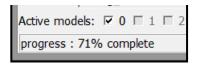
Modellerによるホモロジーモデリング (2)

4) Modellerウィンドウの下の[OK]をクリックすると、計算が開始する。



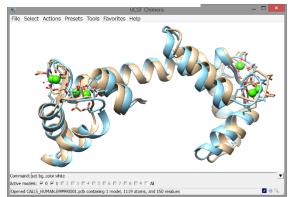
計算終了までは1分~数分かかる。計算進行状況は、画面左下に表示される。



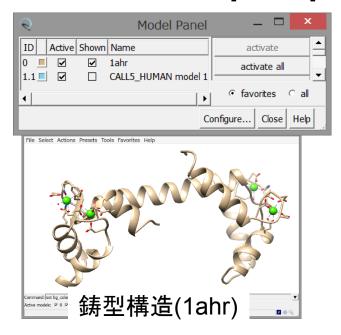


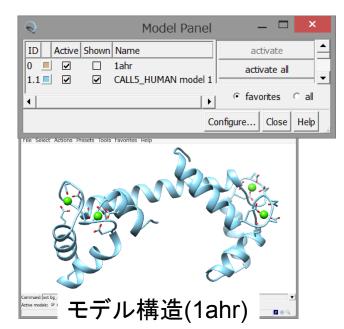
- 5) 計算が終了すると、鋳型構造とモデル構造が表示される。
- 6) [Favorites]→[Model Panel] を選択

※Modellerでは水素原子は生成されません。



Model Panelウィンドウの[Shown]の☑のオン・オフで、オブジェクトの表示・非表示を選択可能。

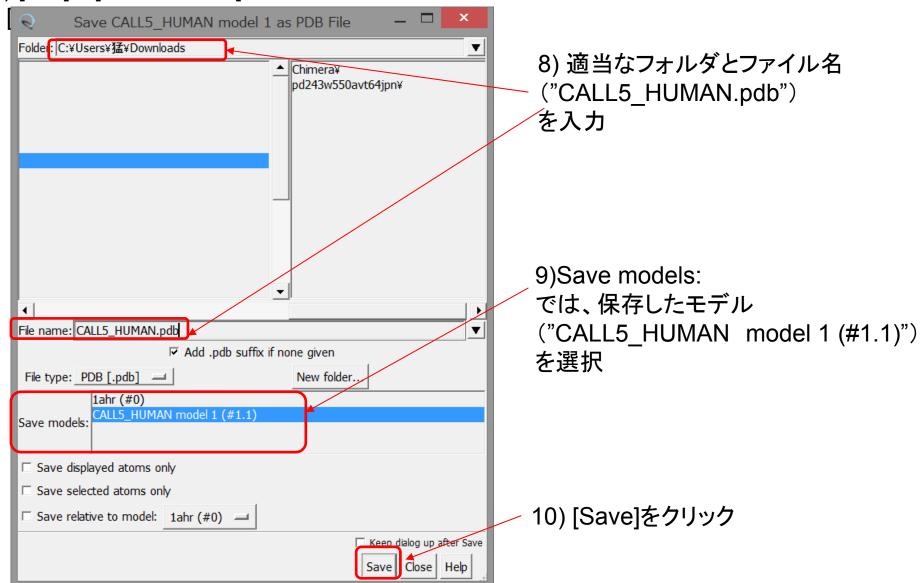




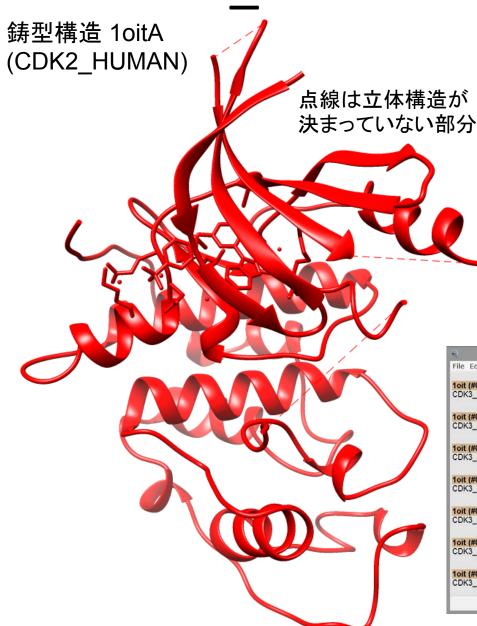
Modellerによるホモロジーモデリング (3)

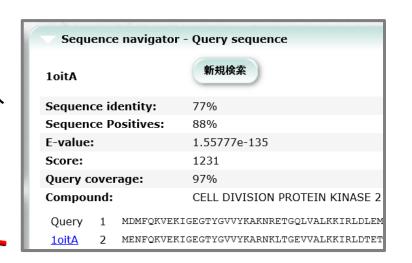
最後に、モデリングした構造だけを、PDB形式のファイルに保存する

7) [File]→[Save PDB…] を選択すると

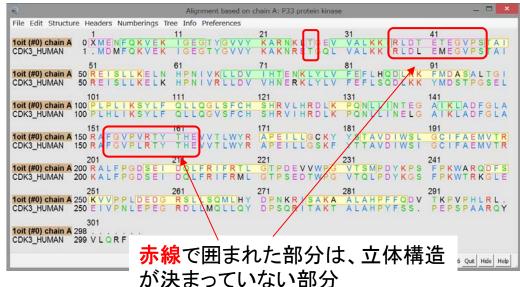


CDK3_HUMANのモデリング例

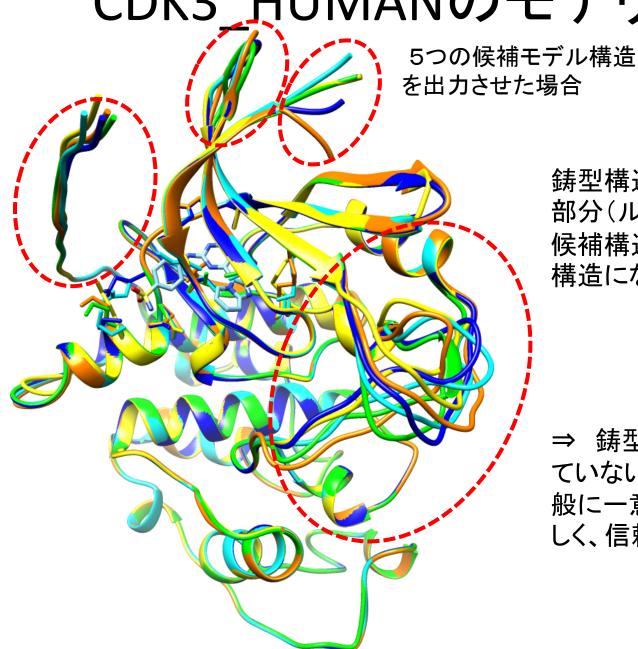




鋳型と標的配列のアラインメント



CDK3 HUMANのモデリング例



鋳型構造とアライメントされない 部分(ループ部、挿入部)は、 候補構造ごとにかなり異なった 構造になる

⇒ 鋳型とのアラインメントされていない部分の予測構造は、一般に一意に構造を決めるのが難しく、信頼性が低い場合が多い。

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 定してしまう
- ・鋳型構造の選択の不具合やアライメントの不具合が「モデリング」の過程で修正されること はない。
- ・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鋳型を選択
- ・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、 いくつかの可能性があり得るとして取り扱ったほうがよい。

コマンドラインでのModellerの使用法

- (1)アラインメントファイル (alignment.ali)
- (2) 鋳型のPDBファイル (1ahr.pdb)
- (3) スクリプトファイル (model.py)
- の三つのファイルを用意し、 コマンドラインで、 mod9.16 model.py というコマンドを実行すれば よい。
 - ⇒HOMCOSサーバで これらのファイルを生成 することも可能
 - ⇒alignment.aliファイル をエディタで修正すれば、 アラインメントを手直しする ことができる。

```
| SP1;query | sequence:query:2::144::::: | AGELTPEEEAQYKKAFSAVDTDGNGTINAQELGAALKATGKNLSEAQLRKLISEVDSDGD | GEISFQEFLTA-AKKARAGLEDLQVAFRAFDQDGDGHITVDELRRAMAGLGQPLPQEELD | AMIREADVDQDGRVNYEEFARMLA-- | *
| SP1;1ahr | structureX:1ahr:1:A:148:A:::: | ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGN | GTIDFPEFLTMMARKMKDSEEEIREAFRVFDKDGNGFISAAELRHVMTNLGEKLTDEEVD | EMIREADIDGDGQVNYEEFVTMMTSK | *
```

UCSF Chimeraだけで実行できる解析

「見てわかる構造生命科学」に記載されている機能

- ・リガンド分子と近接している残基の同定 [Select]→[Zone...]
- •指定した原子間の距離の計測 [Tools]→[Structure Analysis]→[Distance]
- 静電ポテンシャルによる分子表面の色付け

[Tools]→[Surface/BindingAnalysys]→[CoulombicSurfaceColoring]

- ・アミノ酸配列と立体構造とのアラインメント [Tools]→[Sequnece]
- ・進化的保存が高い部位の立体構造上の位置の観察 [Tools]→[Sequence]
- •1アミノ酸置換構造のモデリング [Tools]→[Structure Editing]→[Rotamer]
- ・相同な二つの立体構造の比較 [Tools]→[Structure Comparison]→[MatchMaker]
- ・モーフィングアニメーション

[Tools]→[Structure Comparison]→[MorphConfomation]

それ以外の機能

- ・ペプチド・化合物・核酸の構造構築 [Tools]→[Structure Editing]→[Build Structure]
- ・水素原子の付加 [Tools]→[Structure Editing]→[AddH]
- ・部分電荷の付加 [Tools]→[Structure Editing]→[Add Charge]
- ・低分子ドッキングプログラム Auto Dock Vina の実行

[Surface/Binding Analysis]→[AutoDock Vina]

・ドッキング候補ポーズの解析 [Surface/Binding Analysis]→[ViewDock]

UCSF Chimeraの実行コマンド一覧

Favorites Help	
Model Panel Side View	Command:
Command Line	Active models:
Sequence Reply Log	
Add to Favorites/Toolbar Preferences	:

Command: ☐ 0 ☐ 1 ☐ 2 ☐	
Active models: ☐ 0 ☐ 1 ☐ 2 ☐	
	3

[Favorites]→[CommandLine]で 画面下部にCommand:という コマンドを打ち込む窓が表示される。

コマンドを打ち込むことで、より細やかな設定が可能になる。

書式	例	意味
display & ~display		原子の表示と非表示
ribbon & ~ribbon		リボンモデルの表示と非表示
surface & ~surface		分子表面の表示と非表示
repr [表示法]	repr sphere	原子を空間充填モデルで
[表示法]は、球:sphere、線:wire、スティック:stick、ボール&スティック:bsが使える。		
color [色]	color blue	青色にする
color byelement		元素ごとに色分けする
rainbow		N末からC末へ虹色に
rainbow chain		鎖ごとに虹色に
set bg_color [色]	set 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		分子を元の向きに戻す

UCSF Chimeraの選択コマンド一覧

書式	例	意味
[実行]:.[鎖]	color red :.A	A鎖を赤に
[実行]:[残基名]	color red :CYS	システインを赤に
[実行] @[原子名]	color red @CB	Cb原子を赤に
[実行]:[残基名]@[原子名]:[鎖]	color red :CYS.A@CB	A鎖のシステインのCb原子を赤 に
[実行]:[番号]	color red :104	104番目を赤に
[実行]:[番号],[番号]	color red :104,212	104番目と212番目を赤に
[実行]:[番号]-[番号]	color red :104-212	104~212番目を赤に
[実行]:[番号]-[番号].[鎖]	color red :104-212.A	A鎖の104~212番目を赤に
[実行] [条件] za<[距離]	color red :ATP za<5	ATPから5Å未満の原子を赤に
[実行] [条件] zr<[距離]	color red :ATP zr<5	ATPから5Å未満の残基を赤に
[実行][条件]&&[条件]	color red :.A && :104	A鎖の104番目を赤に
[実行] [条件] [条件]	color red :SER :THR	セリンかスレオニンを赤に
[実行] protein	color red protein	タンパク質を赤に
[実行] nucleic acid	color red nucleic acid	核酸を赤に
[実行] ligand	color red ligand	リガンド分子を赤に

UCSF Chimeraの操作法が載っている本



見てわかる 構造生命科学 —生命科学 研究へのタンパク質構造の利用— 中村春木 編 化学同人 税抜5000円 RasMol, UCSF Chimera, PyMOLの 使い方を解説