PDBj & BINDS　合同講習会
「生命科学のための立体構造データの使い方入門」

ホモロジー・モデリングと複合体の検索・モデリングサーバHOMCOSの紹介

川端　猛　（大阪大学・蛋白質研究所・寄付研究部門准教授）

kawabata@protein.osaka-u.ac.jp

２０１８年２月２０日（火）17:00～18:00
大阪大学　吹田キャンパス　銀杏会館　大会議室

今日の内容

1. ホモロジー・モデリング法とは
2. ホモロジー・モデリング演習
   - i) 配列から相同な立体構造の検索
   - ii) UCSF Chimeraによる配列と立体構造のアラインメント
   - iii) Modellerを用いたホモロジー・モデリング

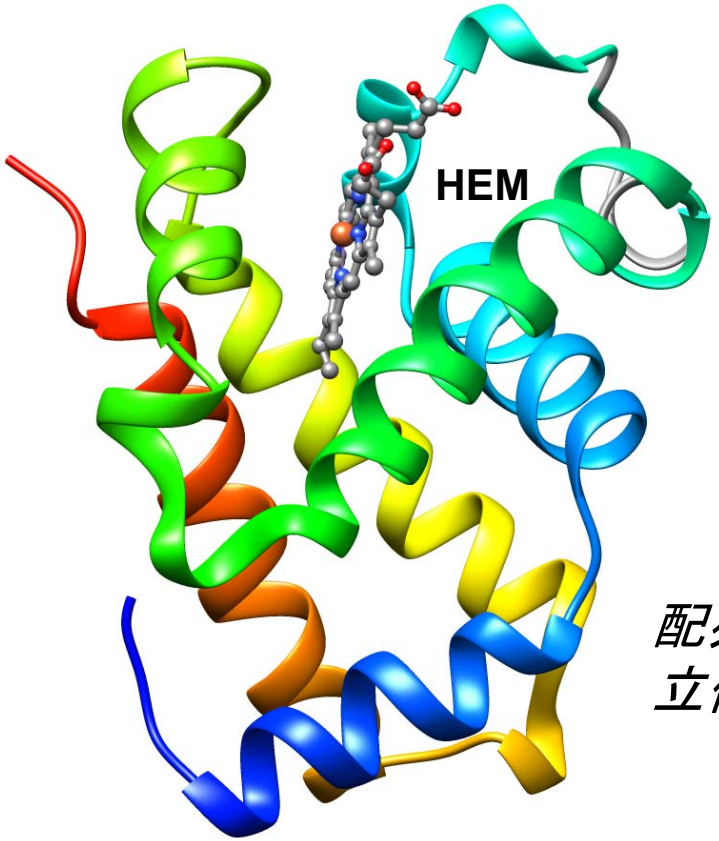
HOMCOS

3. 複合体のホモロジーモデリング
4. HOMCOSサーバを用いた演習
   - i)タンパク質の結合分子の予測
   - ii)ヘテロ複合体構造の予測
   - iii)化合物-蛋白質複合体構造の予測

付録：Modellerによる二量体のモデリングの手続き

ホモロジー・モデリング法とは

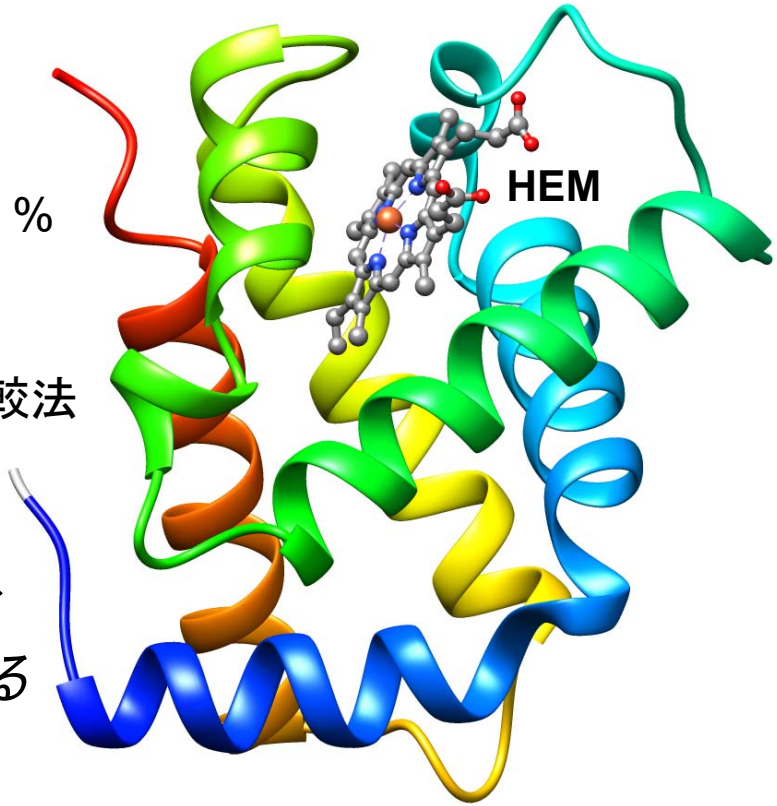
アミノ酸配列と立体構造の関係 (1)



配列一致率:46 %

標準的な配列比較法
(BLAST)

配列が似ていれば
立体構造も似ている

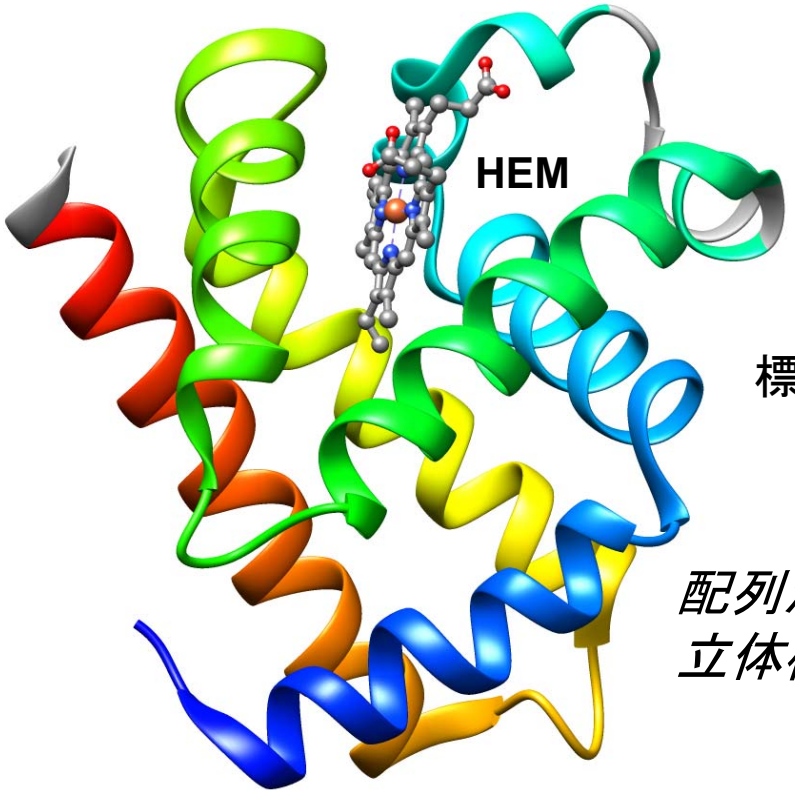


HBB_HUMAN (1bz1B)
hemoglobin beta chain, human

HBA_HUMAN (1bz1A)
hemoglobin alpha chain, human

```
HBB  4  :LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGL:76
         * *  *  * * ****     * * *** *     * *   *  * ***      *   ** *****  *

HBA  3  :LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH----GSAQVKGHGKKVADALTNAV:71

HBB  77 :AHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY:146
          ** *        ** **  ** *** ** **   *   ** *   **** * *   *  * *   *  **

HBA  72 :AHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY:141
```

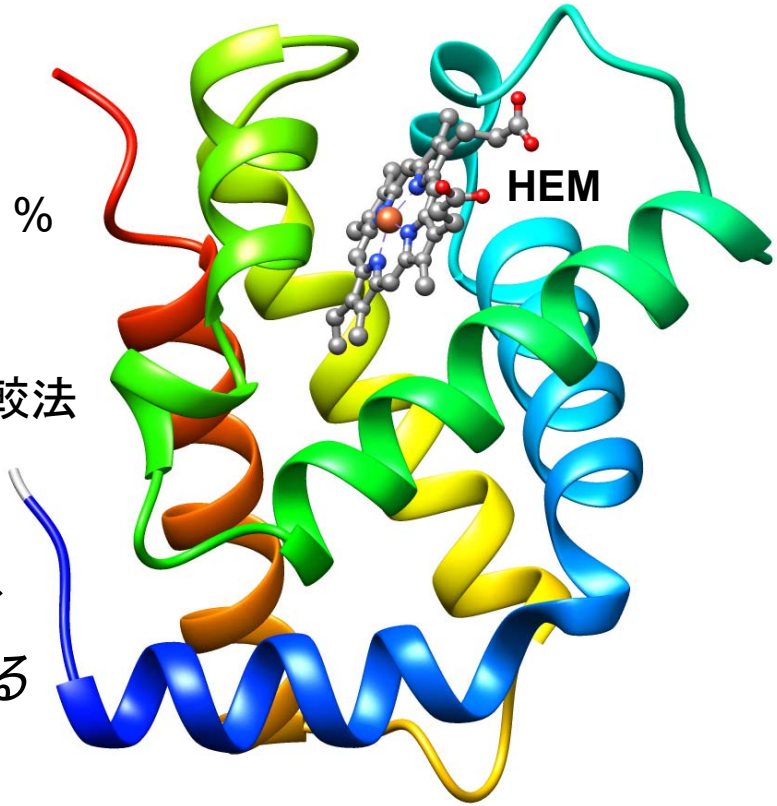
アミノ酸配列と立体構造の関係 (2)



配列一致率:27 %

標準的な配列比較法
(BLAST)

配列が似ていれば
立体構造も似ている

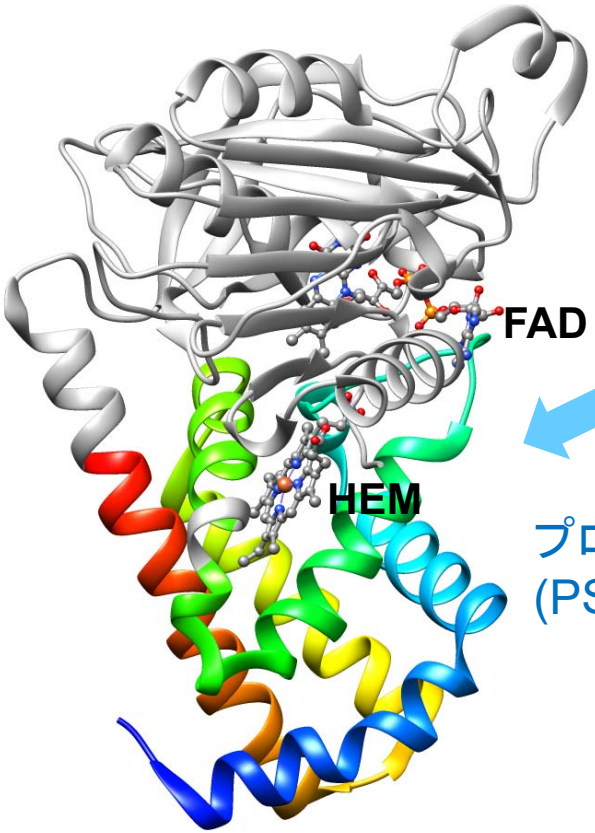


MYG_HUMAN (3rgkA)
hemoglobin beta chain, human

HBA_HUMAN (1bz1A)
hemoglobin alpha chain, human

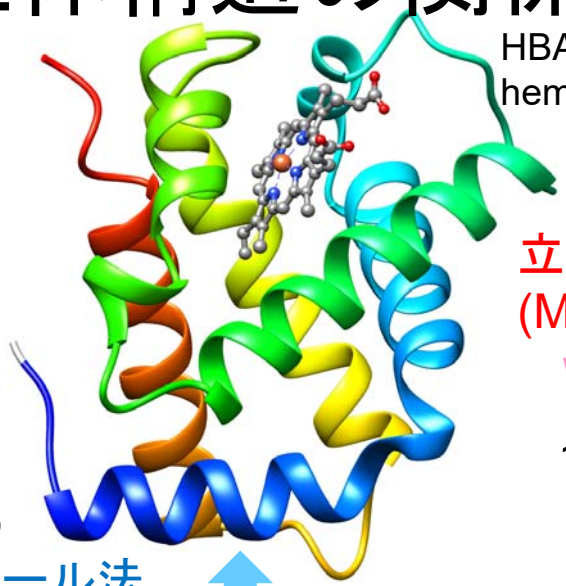
```
MYG   1:MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGG:75
         * **       *    **** *      * * * * *    * *    *  *       *    *    * **  *  **

HBA   1:MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF------DLSHGSAQVKGHGKKVADALTN:69

MYG  76:ILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYK: 148
                        *    ** *            * *     *     *  *          * *         * *

HBA  70:AVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR: 142
```

アミノ酸配列と立体構造の関係 (3)



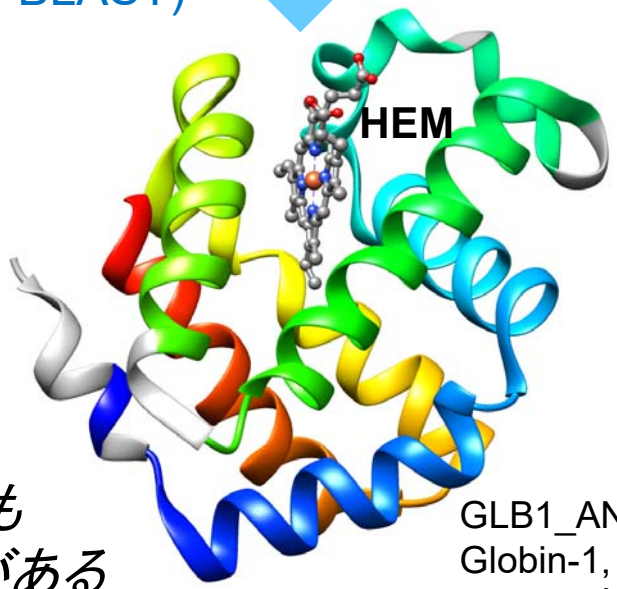
HMP_ECOLI (1gvhA)
Flavohemoprotein, hmp,
Escherichia coli
(大腸菌) 配列一致率：17%

17%
プロフィール法
(PSI-BLAST)



HBA_HUMAN (1bz1A)
hemoglobin alpha chain, human

17%

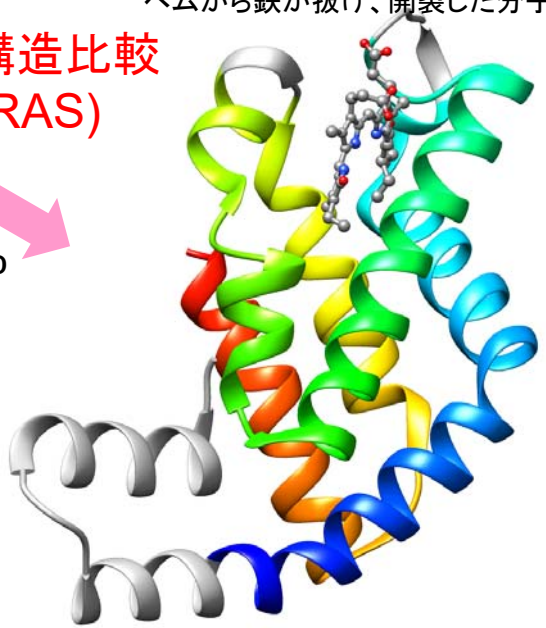


GLB1_ANAIN (3g46A)
Globin-1, *Anadara inaeqequivalivis* (二枚貝)
配列一致率：17%

立体構造比較
(MATRAS)

10%

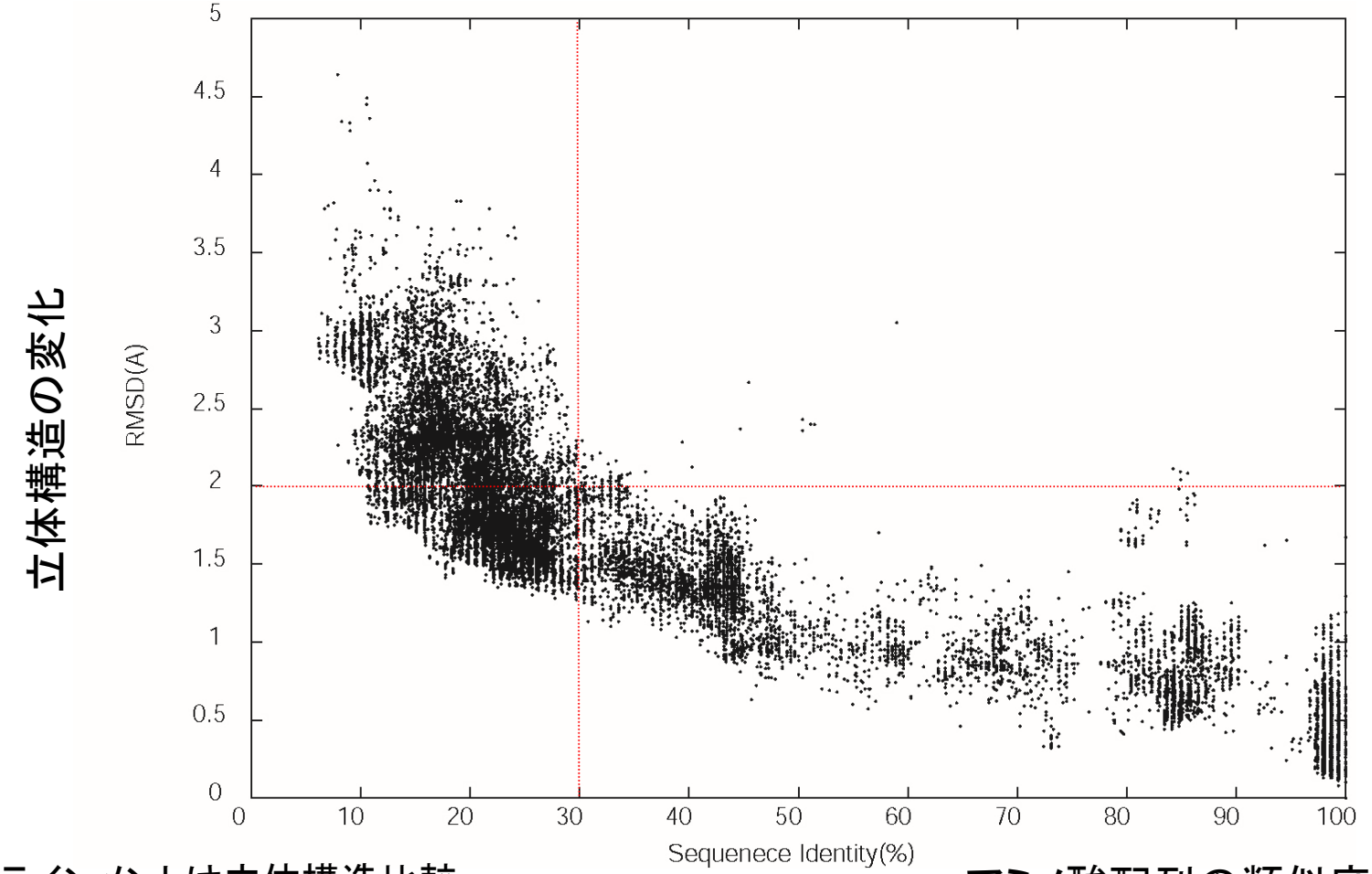
CYC: phycocyanobilin
ヘムから鉄が抜け、開裂した分子



PHCA_ARTPT (1gh0A)
C-phycocyannin alpha chain, cpcA
Arthrospira platensis (シアノバクテリア、藍藻、藍色細菌) 配列一致率:10%

配列類似性がかなり低くても
立体構造が似ている場合がある
⇒立体構造は配列よりも進化的によく保存

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)



※アラインメントは立体構造比較
プログラム(MATRAS)で計算

アミノ酸配列の類似度

立体構造の変化はアミノ酸配列の変化と相関
配列が30%以上一致していれば、RMSDは2Å以下

アミノ酸配列と立体構造の関係のまとめ

・アミノ酸配列が有意に似ていれば、多くの場合立体構造も似ている

⇒相同（進化的な起源を同じにする）タンパク質の立体構造は、おおむね似た形をしている

⇒「立体構造はアミノ酸配列より進化的によく保存する」

⇒タンパク質の機能は、その立体構造をベースとするので、大きく構造を変化させるアミノ酸置換は進化の過程で淘汰されてしまうから？

・相同なタンパク質であっても、配列類似性が低くなるほど、立体構造のずれは多くなる。

相同性（ホモロジー; homology）: 進化的に共有祖先をもつことに由来すると考えられる類似性。

立体構造予測法の二つのアプローチ

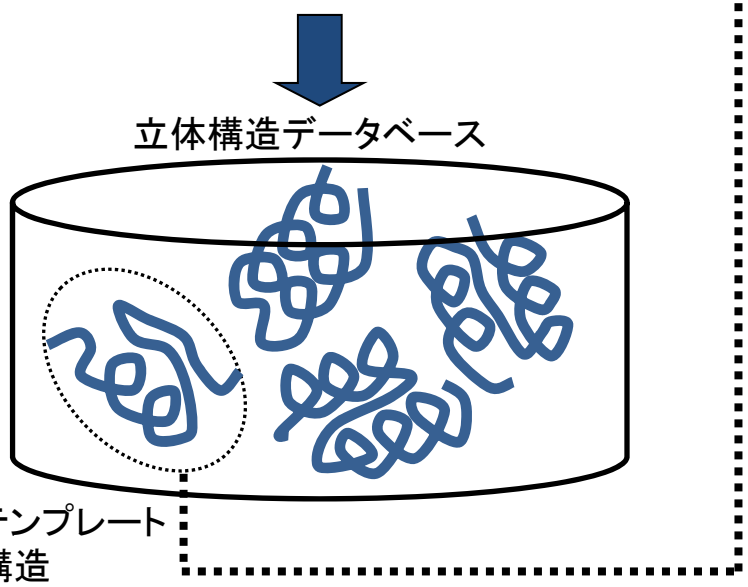
名称	ホモロジー・モデリング法 比較モデリング法 鋳型ベース予測法	非経験的予測法 Ab initio 予測法 De novo予測法
手法の概要	鋳型立体構造にできるだけ似た形で、立体構造を予測	鋳型構造を用いずに、物理化学的な原理（分子シミュレーションの技法）に基づいて立体構造を予測
鋳型立体構造	必要	不要
一般性	低い	高い
計算量	少ない	多い
予測精度	似た鋳型があれば高い	高い精度を得るには大きな計算量が必要
単体の立体構造予測	MODELLER, SWISS-MODEL, RosettaCM, 3Dzigzaw	ROSETTA, EVfold,…
蛋白質複合体予測	MODELLER,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

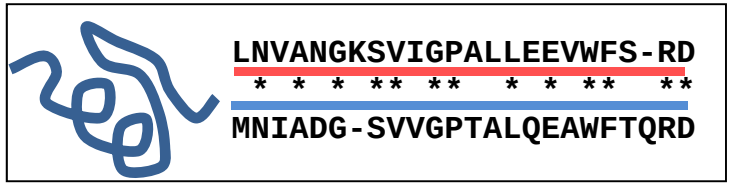
ホモロジー・モデリングによる3次構造予測

原理： 立体構造はアミノ酸配列より進化的に保存する.

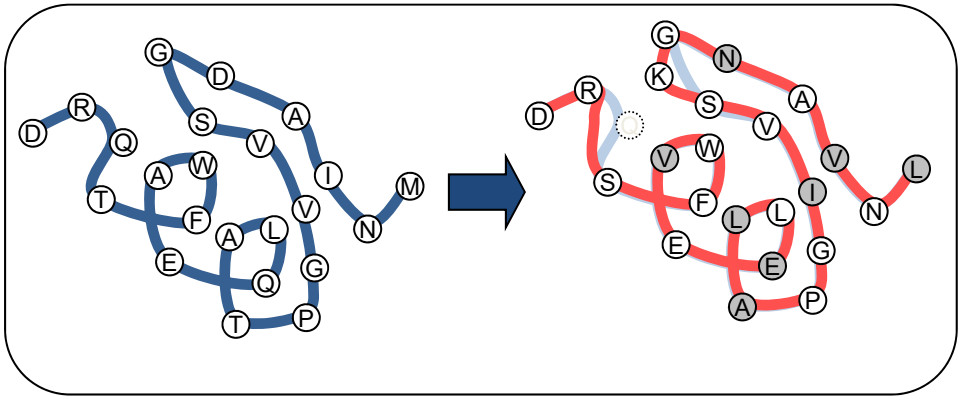
予測対象配列 （クエリ配列）

LNVANGKSVIGPALLEEVWGSRD





鋳型（テンプレート）構造とそのアラインメント



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

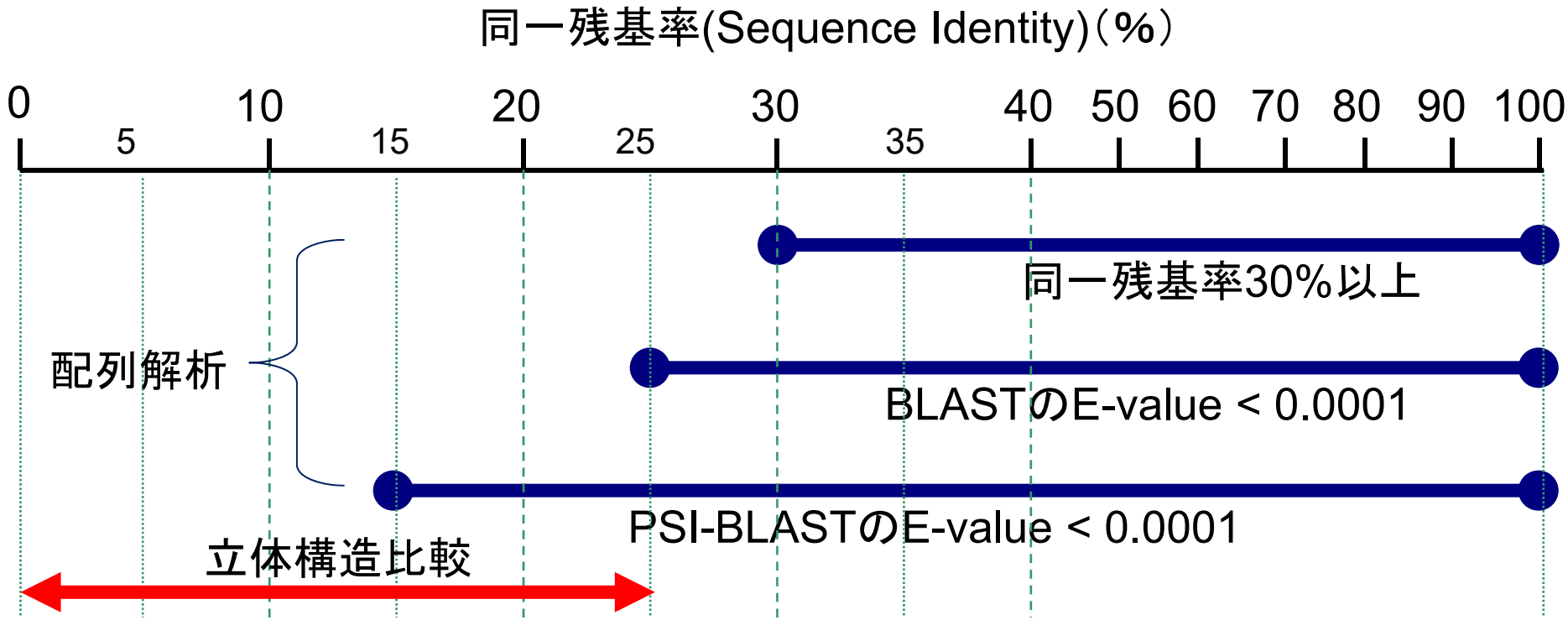
BLAST, プロフィール法, スレディング法….

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

相同なタンパク質の発見する方法(フォールド認識)



・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。

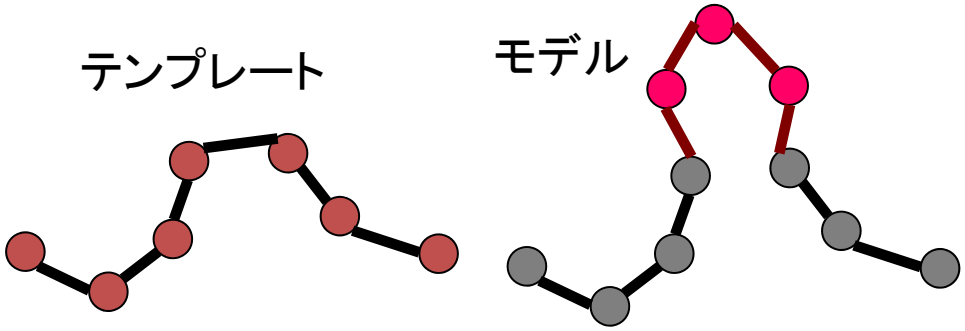
・立体構造比較は、複合体のホモロジー・モデリングのために役に立つことがある。

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

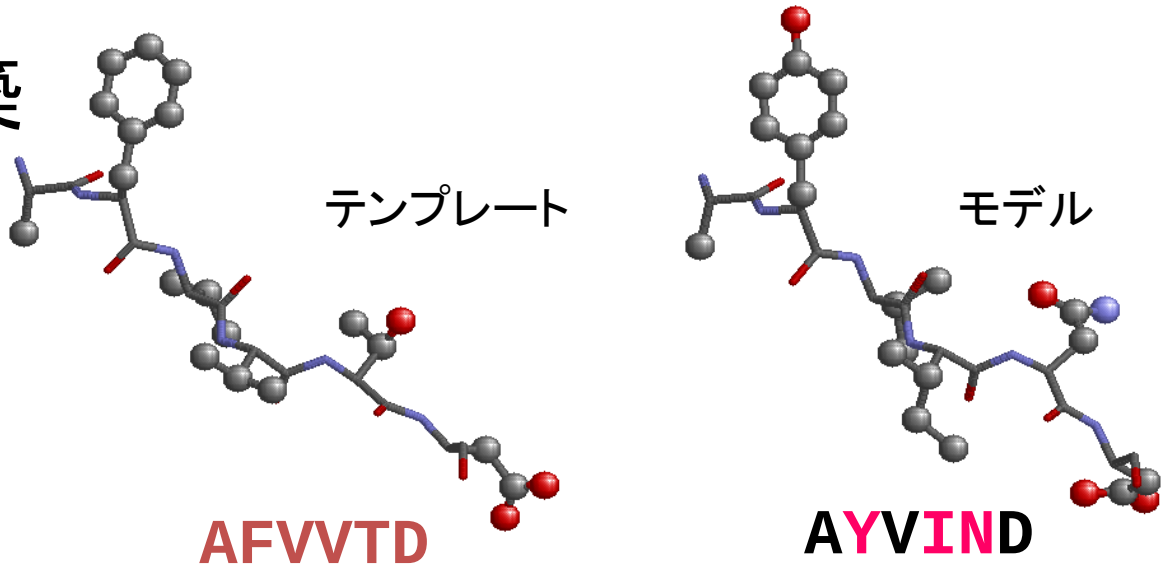
(1)ループの構築

Sequence ALIMSTKGFVS
Structure LLLM---GFIT



(2)側鎖原子の構築

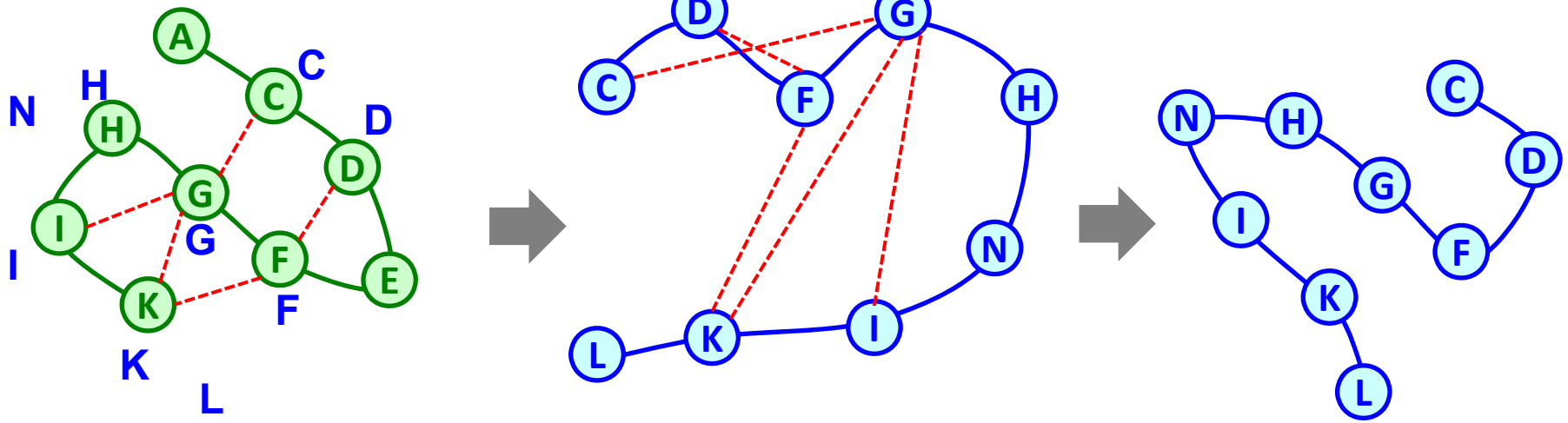
Sequence AYVIND
Structure AFVVTD



MODELLER, SWISS-MODEL, RosettaCM, 3Dzigzaw

MODELLERの手続き:空間拘束の充足

Target :-CD-FGHNIKL
Template:ACDEFGH-IK-



1. 標的配列を鋳型構造にアラインメントする　2. 空間的拘束(spatial constraint)を抽出する　3. 空間的拘束を満たすような構造を探す

・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。

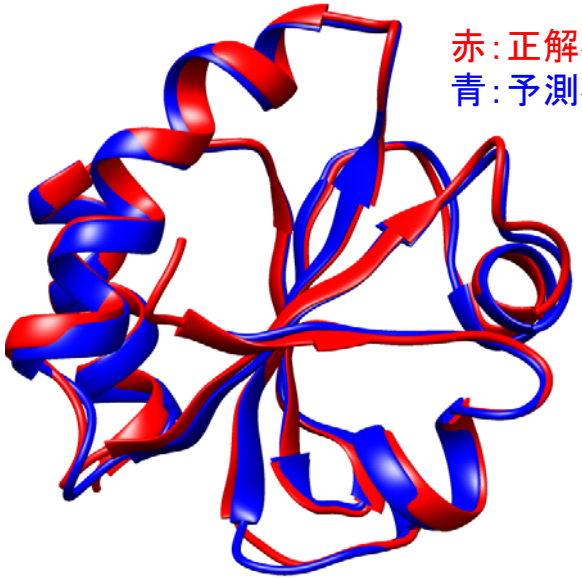
・ループの構築や側鎖の構築の手続きは、この「空間的拘束の充足」の手続きに含まれる。

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.

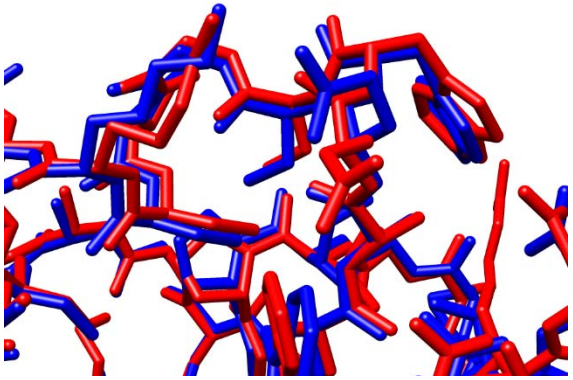
モデル結果の例: 予測対象　2trx_A Thioredoxin (*Escherichia coli*)

アラインメント:BLAST, モデリング:MODELLER

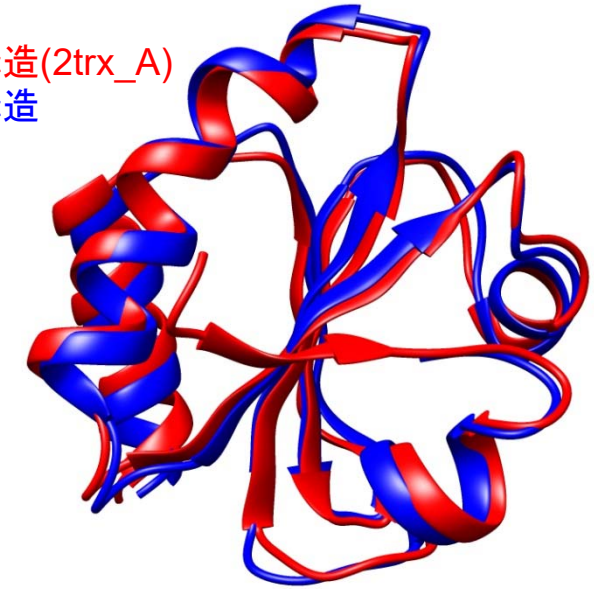
赤:正解構造(2trx_A)
青:予測構造



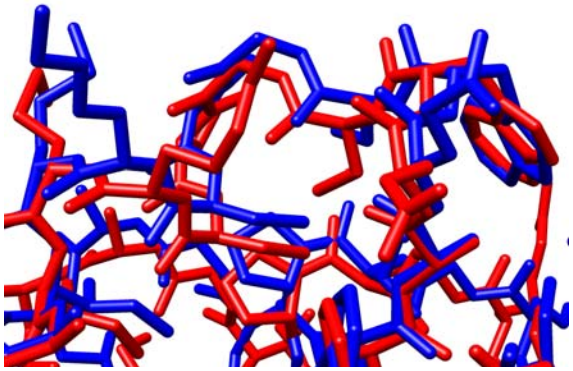
鋳型とのSeqID=86.5%
RMSD(Cα)=0.64 Å
鋳型:2yn1_A LGPCA Thioredoxin



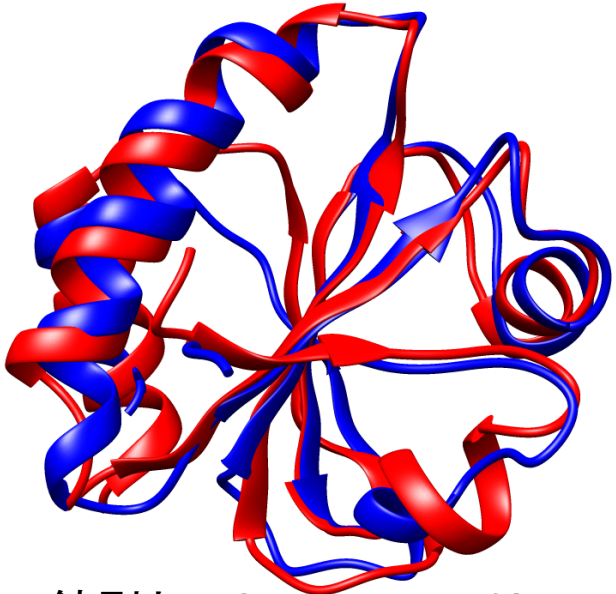
RMSD(全重原子)=1.29 Å



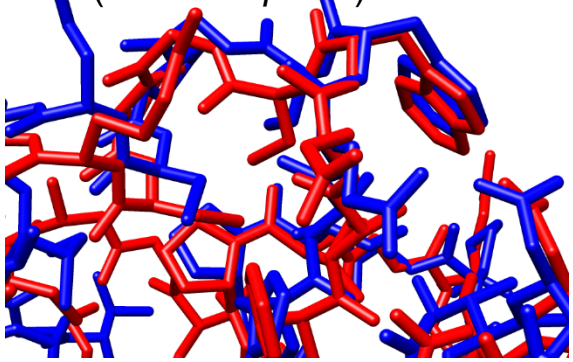
鋳型とのSeqID=52.5%
RMSD(Cα)=1.08 Å
鋳型:1fb0_A Thioredoxin M
(*Spinacia oleracea (spinach)*)



RMSD(全重原子)=1.81 Å



鋳型とのSeqID=25.3%
RMSD(Cα)=2.30 Å
鋳型:1x5e_A Thioredoxin
domain containing protein
(*Homo sapiens*)



RMSD(全重原子)=2.83 Å

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: [Q0NZT1]: 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

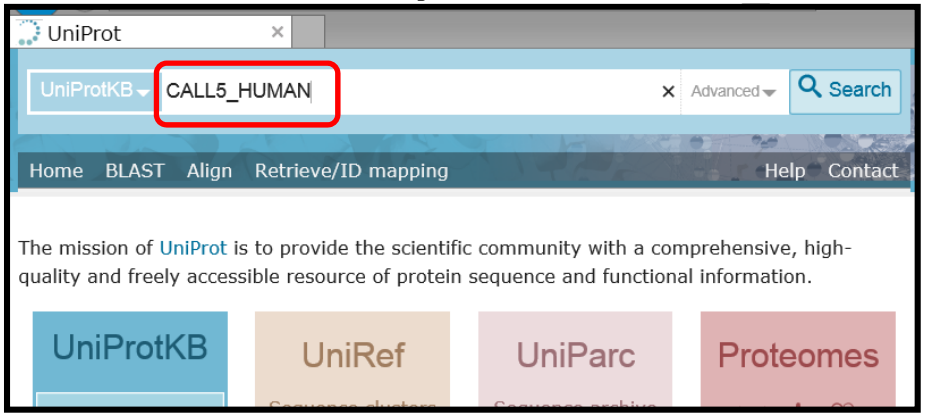
```
VARIANT 58 58 S -> G
        (polymorphism confirmed at protein FT level;dbSNP:rs11546426).
VARIANT 74 74 K -> R
        (polymorphism confirmed at protein FT level; dbSNP:rs10904516).
```

※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。

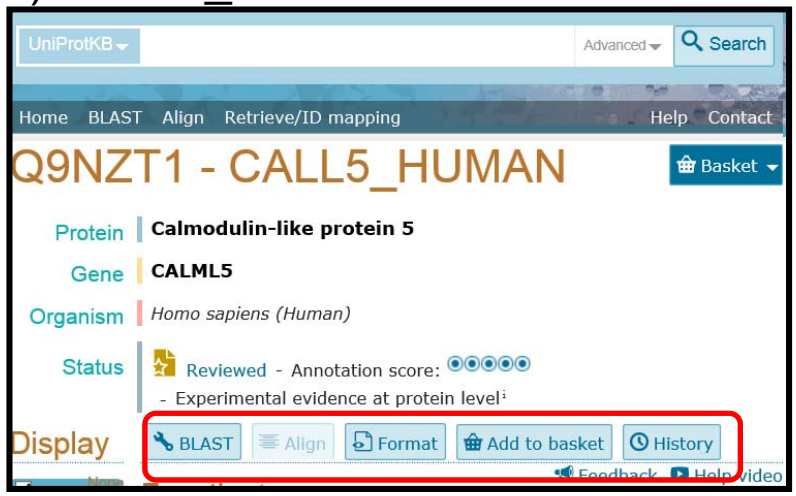
アミノ酸配列の取得と検索

1) Googleで”UniProt“と入力

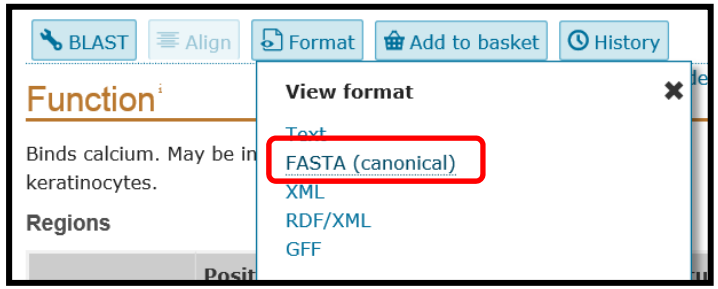
2) UniProtのページのフォームに”CALL5_HUMAN“と入力



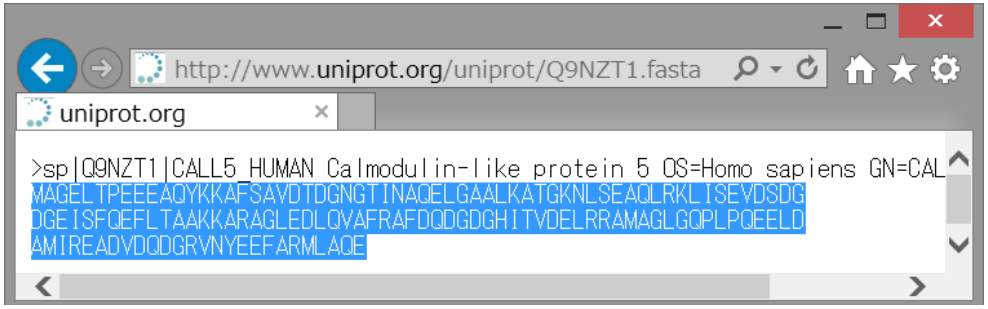
3) CALL5_HUMANのページ



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



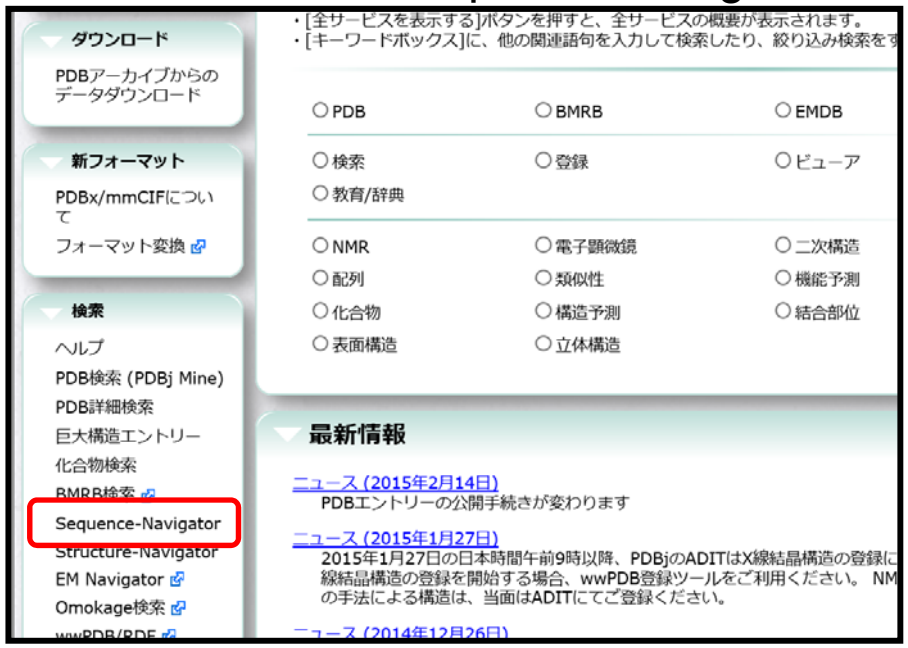
5)一文字表記のアミノ酸配列が表示される。
これをマウスで選択し、コピーする。



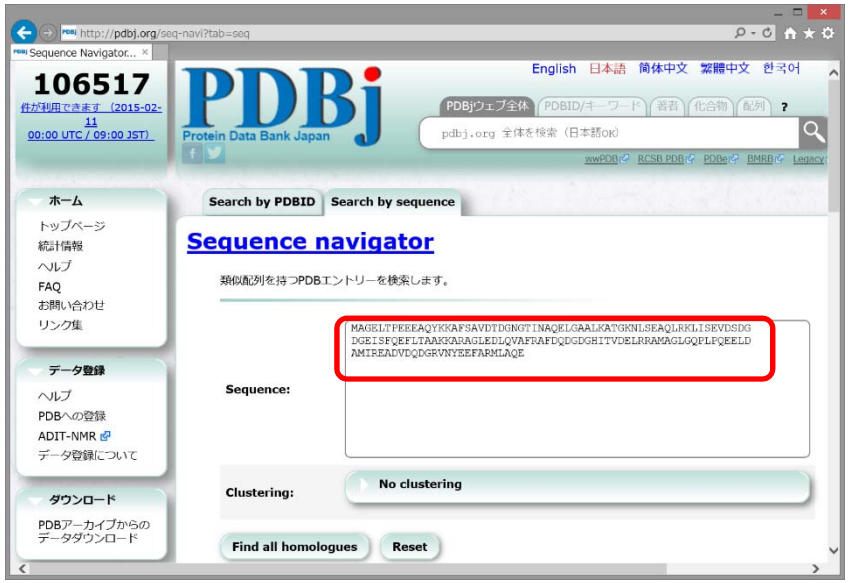
PDBjによる相同な立体構造（鋳型構造）の検索

1) Googleで”PDBj“と入力

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

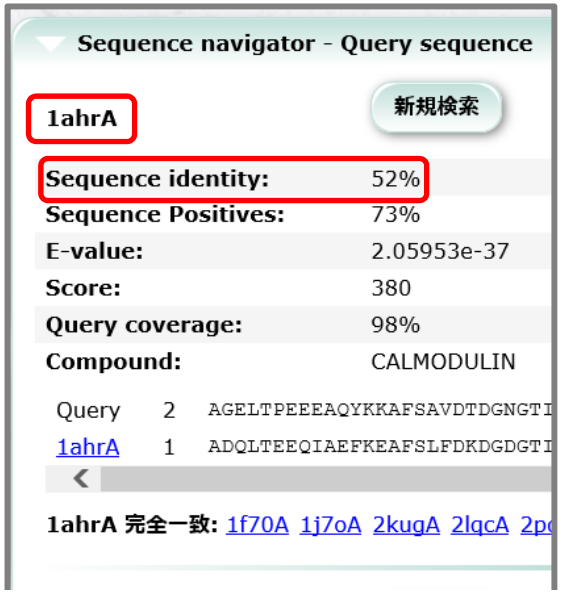


3) [Search by sequence]のタブを選び、フォームに、UniProtのページでコピーしたCALL5_HUMANの配列をペースト



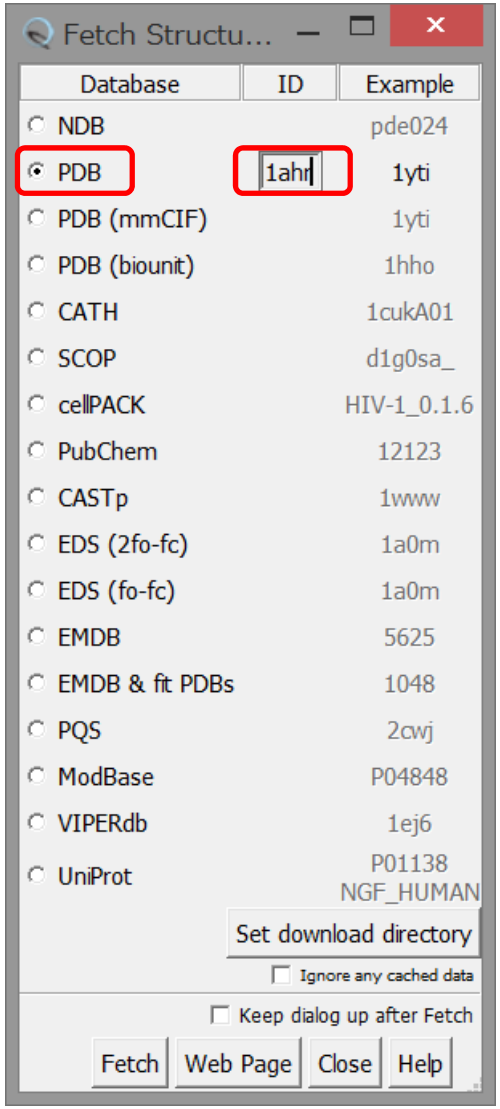
4) 対PDBのBLAST検索の結果が表示される。PDBコード“**1ahr**”の**A**鎖が、sequence identity 52%でヒット。

→これを鋳型とする



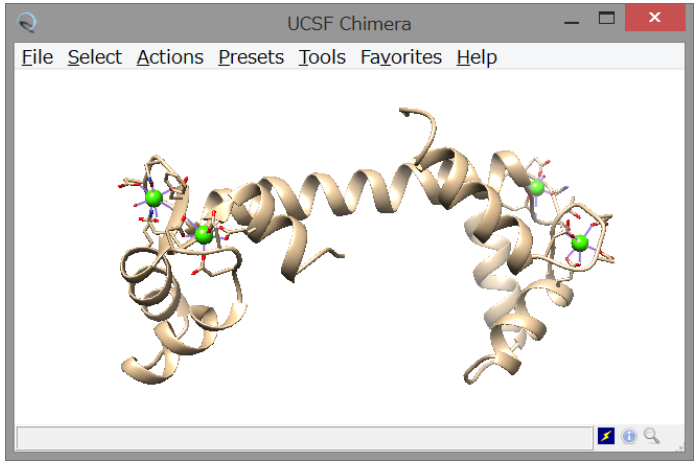
Chimera：鋳型構造の読み込み

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

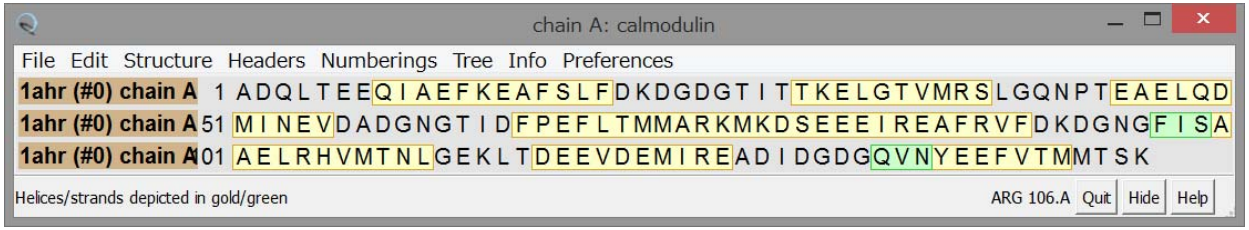


2) [PDB]を選択、IDのフォームに”**1ahr**”と入力し、[Fetch]をクリック。

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。

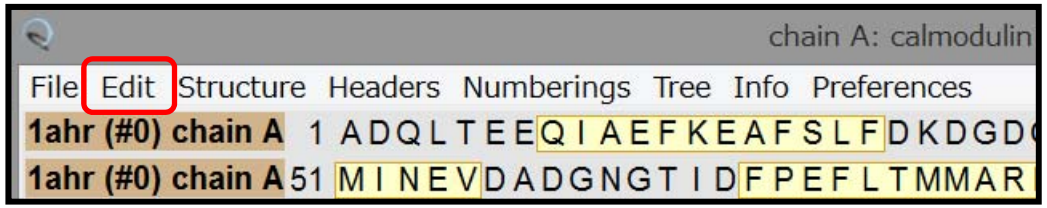


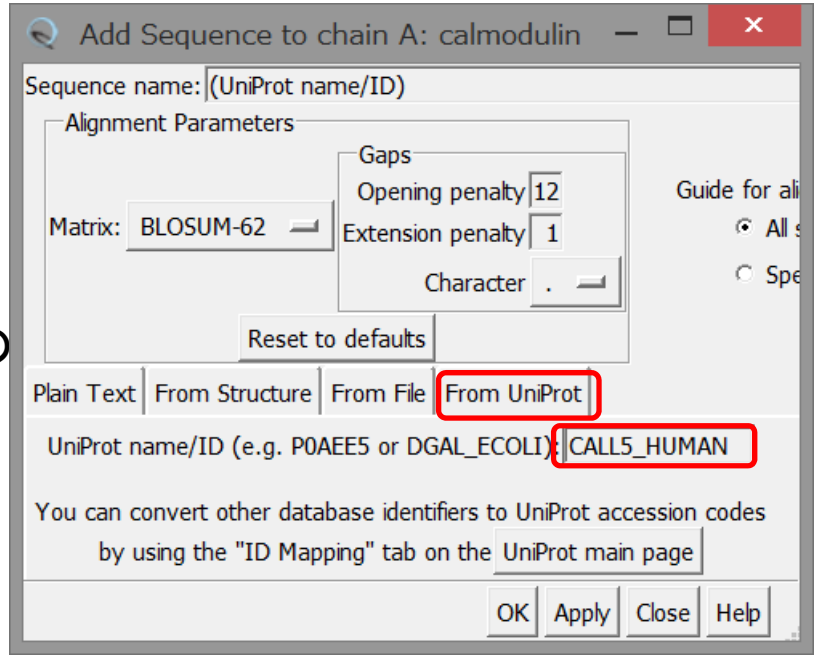
4) メニューから**[Favorites]→[Sequence]** を選ぶ。と以下のSequenceウィンドウが表示される。



Chimera：標的配列の読み込み

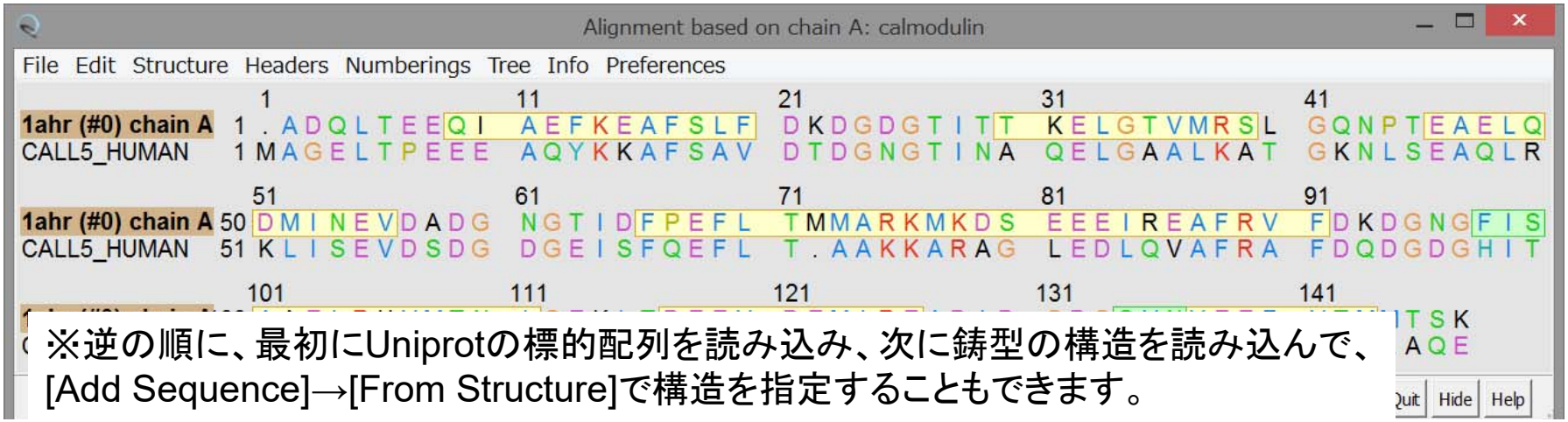
1) Sequenceウインドウのメニューから **[Edit]→[Add Sequence…]** を選ぶとAdd Sequenceのウィンドウが表示される。





2) Add Sequence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。



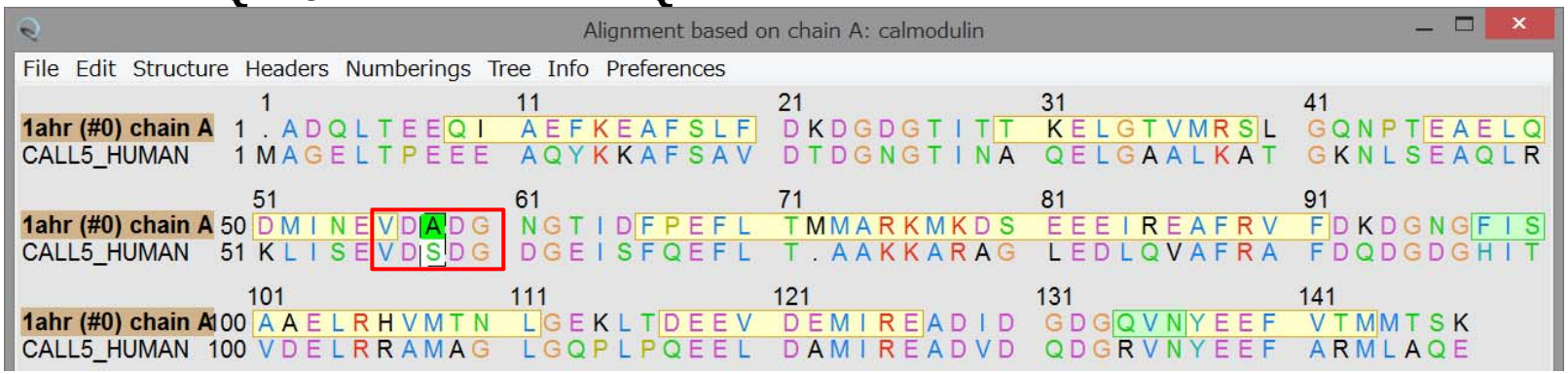
※逆の順に、最初にUniprotの標的配列を読み込み、次に鋳型の構造を読み込んで、[Add Sequence]→[From Structure]で構造を指定することもできます。

Chimera:変異箇所の立体構造の確認

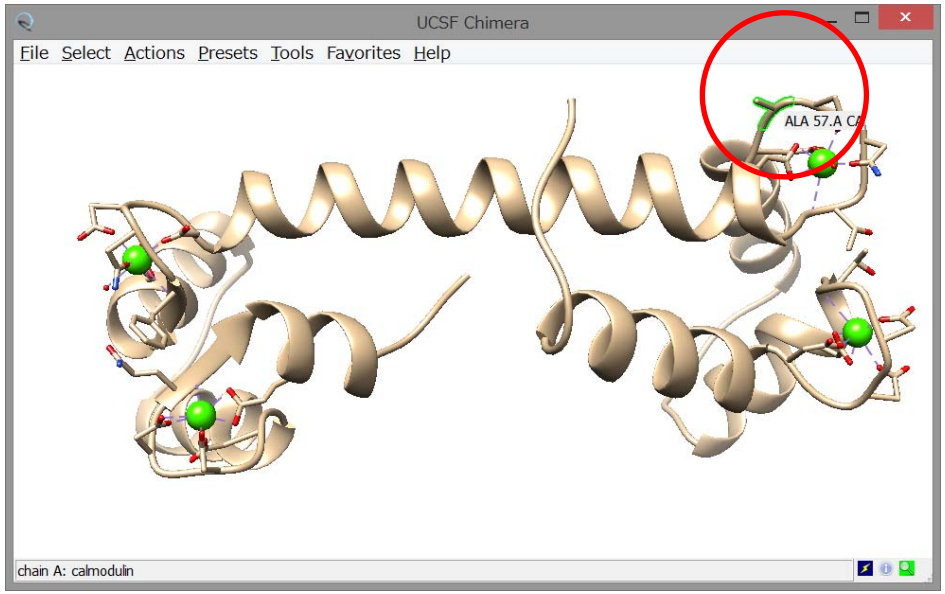
SNPが報告されている 58番目のS (S->G)の立体構造上の位置を確認してみる。

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 (VD**S**DG)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。



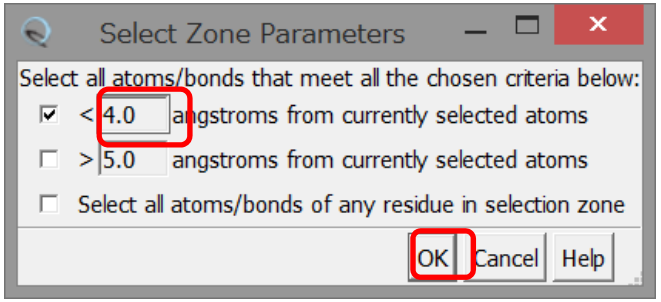
2) 選択された状態で、
[Actions]→[Atoms/Bonds] →[Show]
とすると、選択された構造部位がスティック表示される。

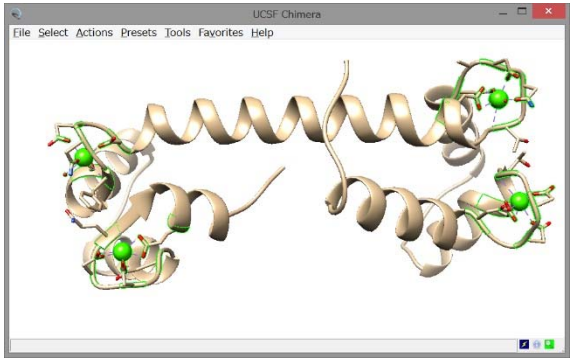
※同様に74番目の**K -> R**
の位置も確認してみる

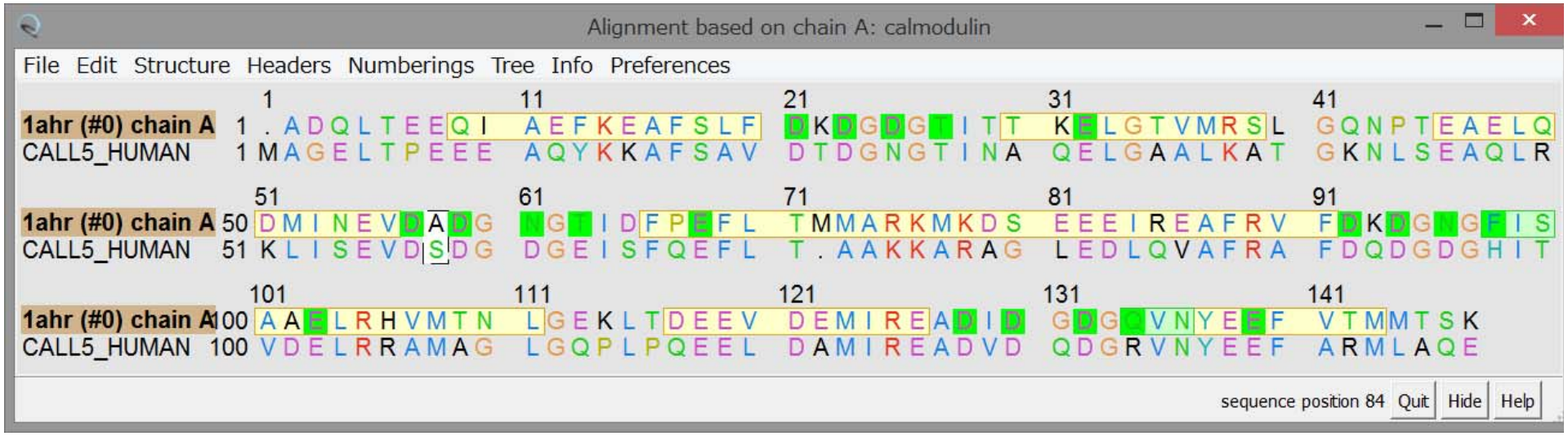
Chimera:標的配列のCa^{2+}結合部位の推定

鋳型立体構造(1ahr)のCa^{2+}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、Ca^{2+}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、Ca^{2+}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、Ca^{2+}結合部位が緑色で強調表示されている。

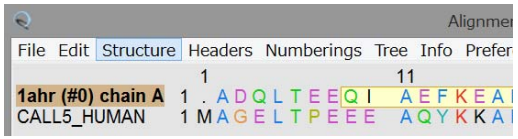




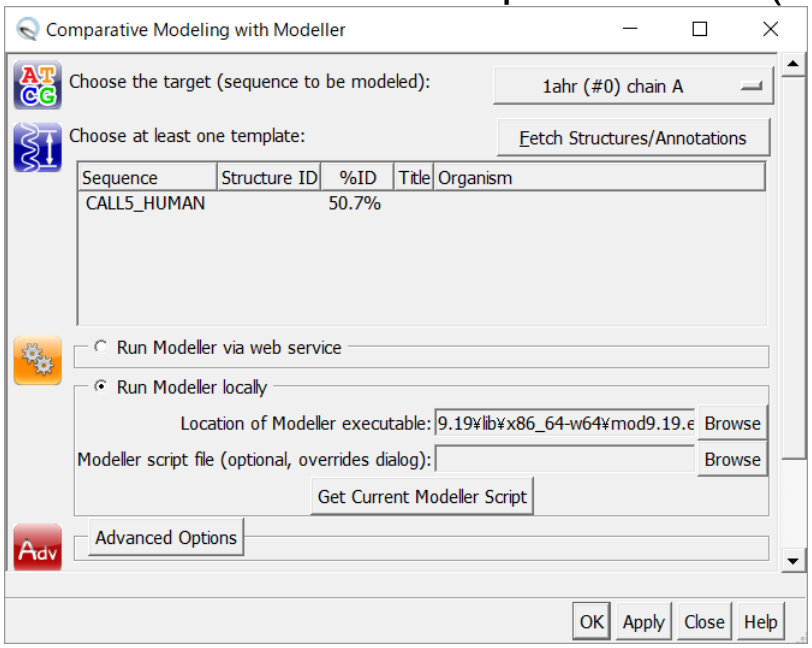


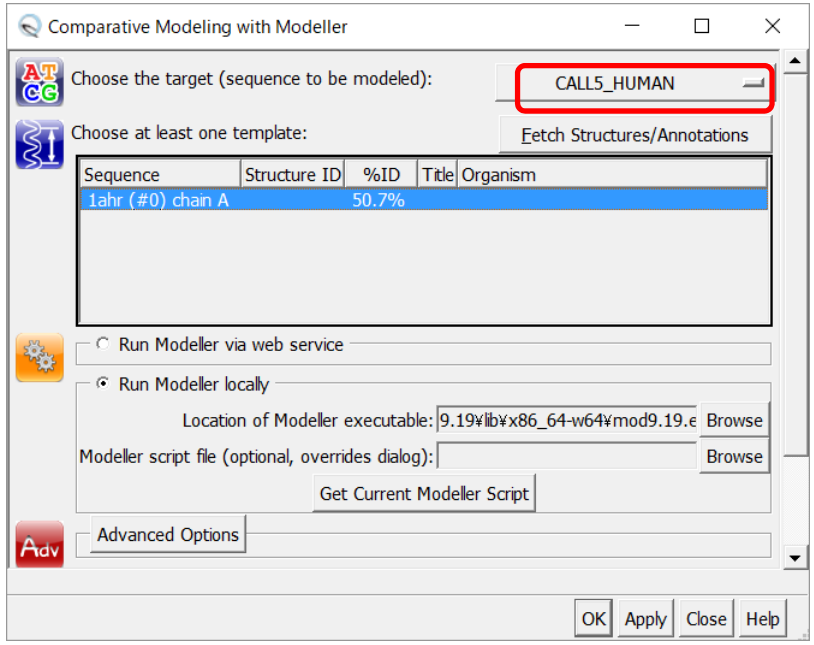
Modellerによるホモロジーモデリング(1)

1) sequenceウィンドウの **[Structure]→[Modeller(homology)…]** を選択



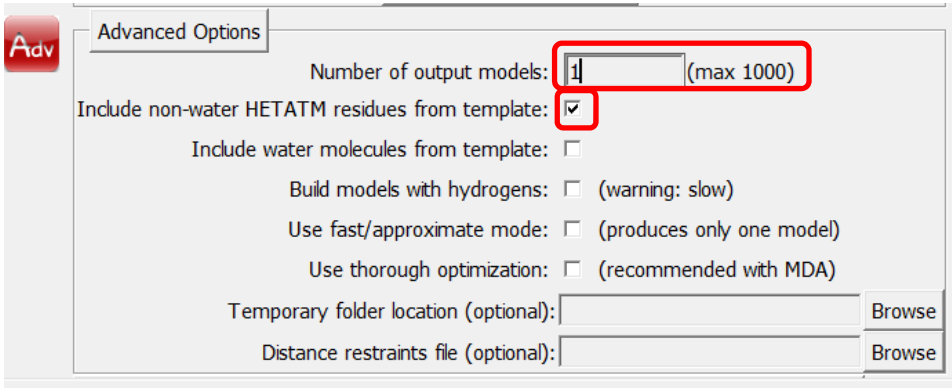
2) Modellerウィンドウの、Choose the targetを”CALL5_HUMAN”とし、Choose at least one template：を1ahr(#1) chainAを選択する。





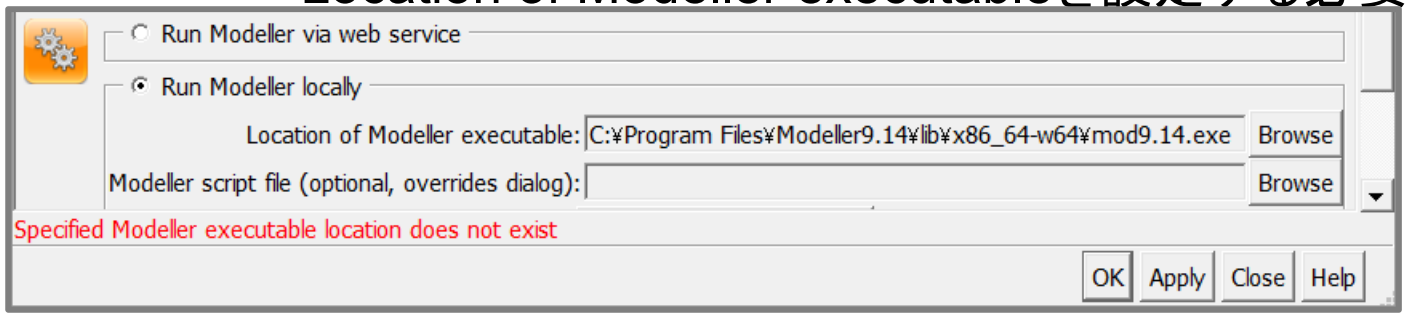
3) [Advanced Options]をクリックし、[Number of output models](出力するモデル構造の数)を1とし、

[Include non-water HETATM residues from template](低分子化合物・イオンをモデルに含めるかどうか)をチェック☑する。



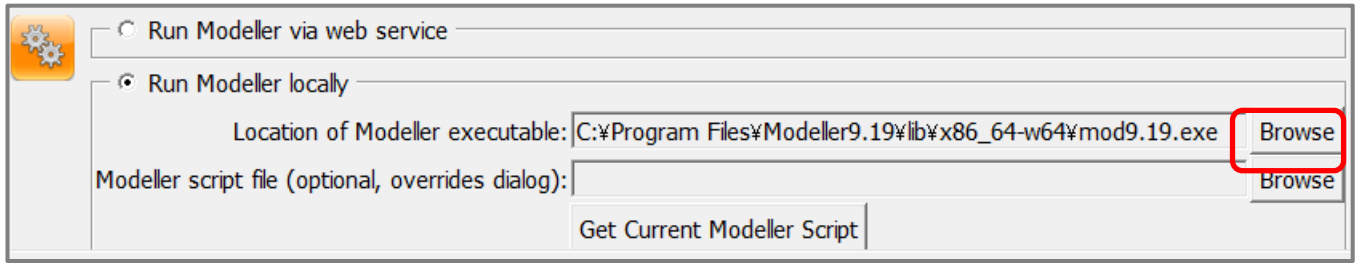
ローカルにModellerを起動するための注意

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



デフォルトでは間違った場所に設定されているかもしれません。

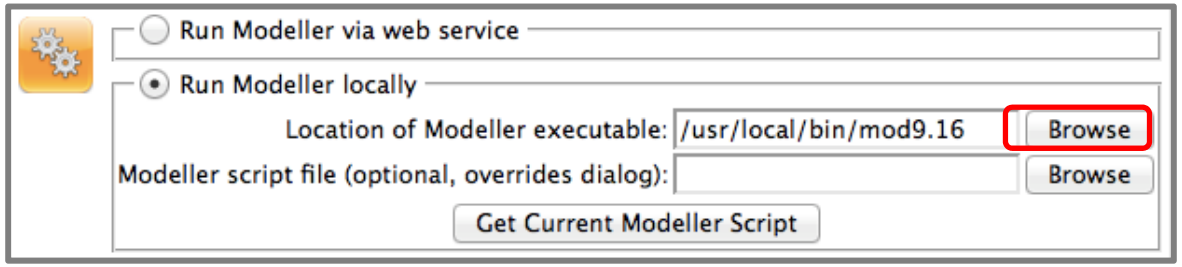
Windowsの場合の設定例



[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¥Modeller9.19¥lib¥x86_64-w64¥mod9.19.exe

Macintoshの場合の設定例



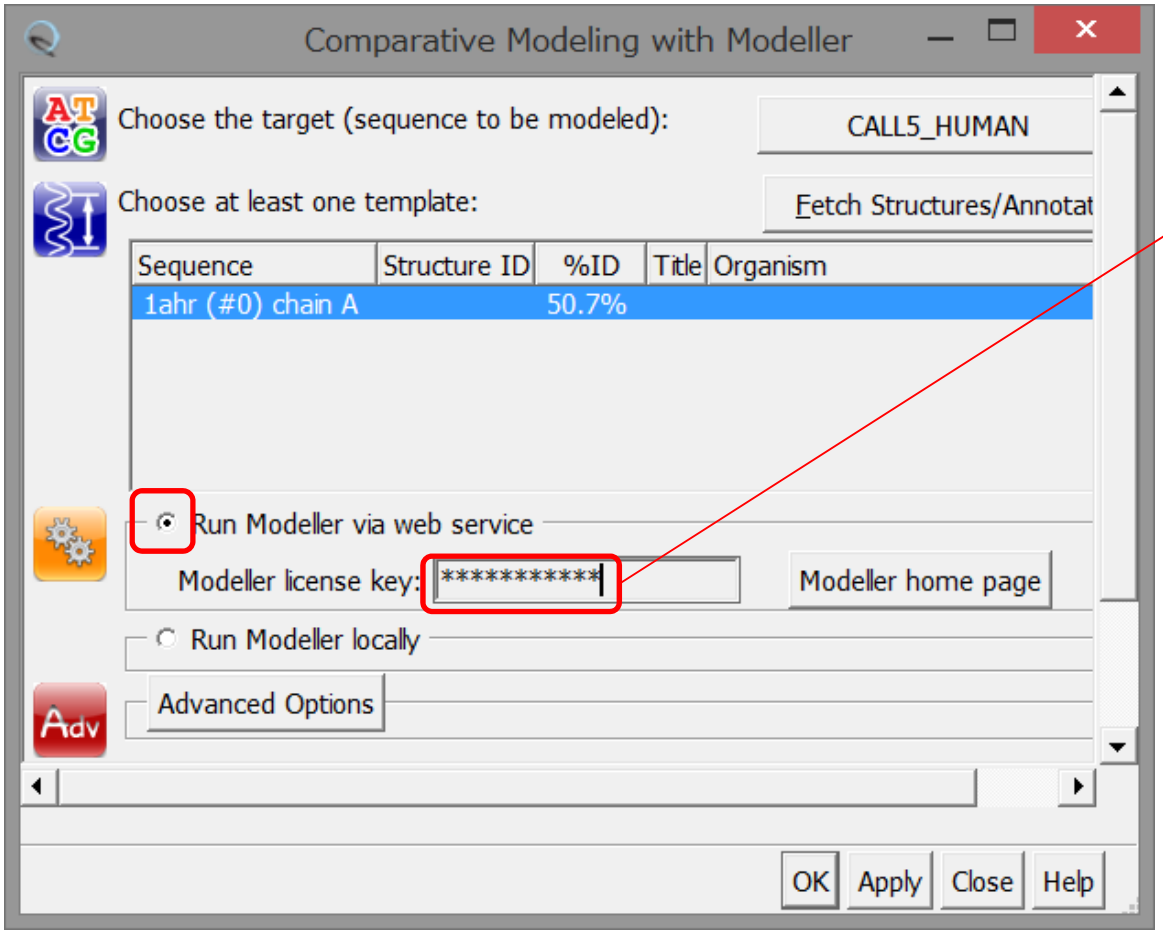
[Browse]をクリックしてフォルダを移動し、Modellerの実行ファイルを選択

/usr/local/bin/mod9.19

※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。



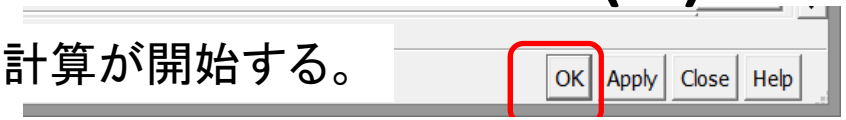
アカデミックライセンスの<u>ライセンスキー文字列</u>を入力する必要があります。

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。

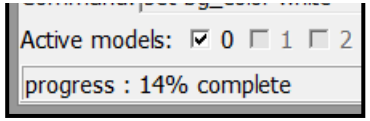
※Web Serviceを利用した場合も、ローカルに起動した場合も以後の手続きは同じです。

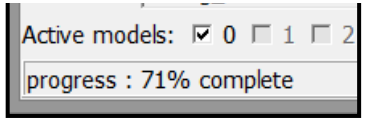
Modellerによるホモロジーモデリング (2)

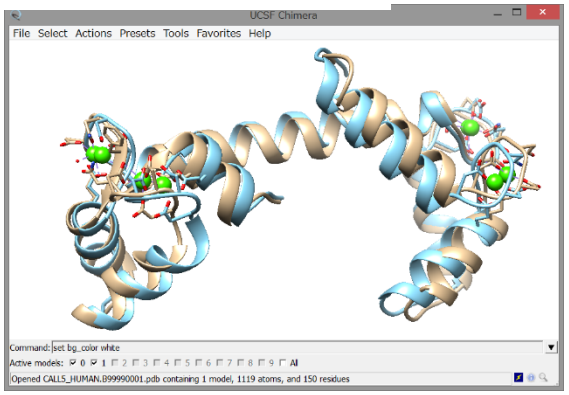
4) Modellerウィンドウの下の[OK]をクリックすると、計算が開始する。



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。





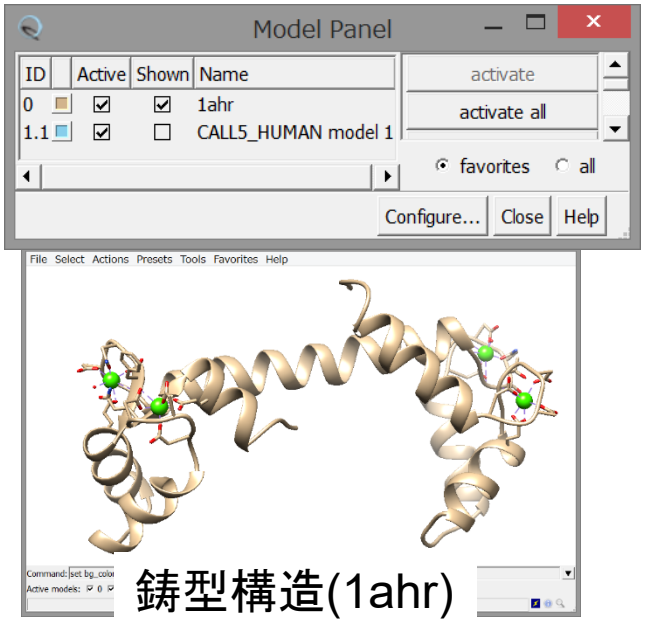


5) 計算が終了すると、鋳型構造とモデル構造が表示される。

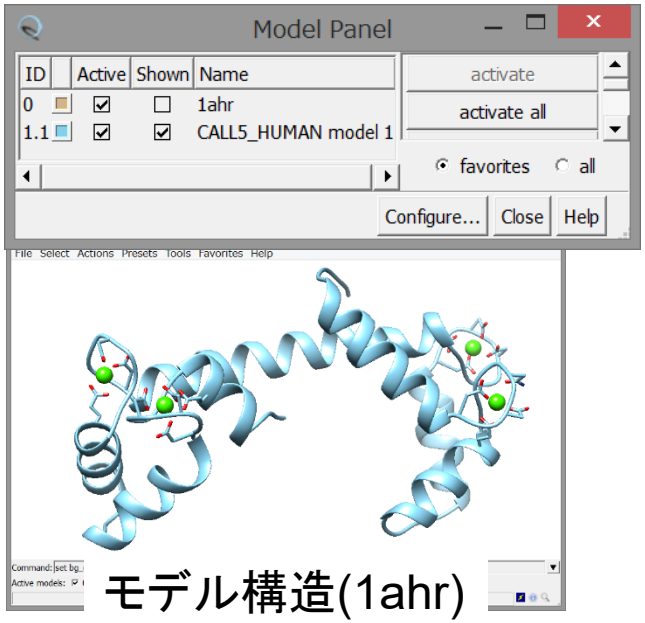
6) **[Favorites]→[Model Panel]** を選択

※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。



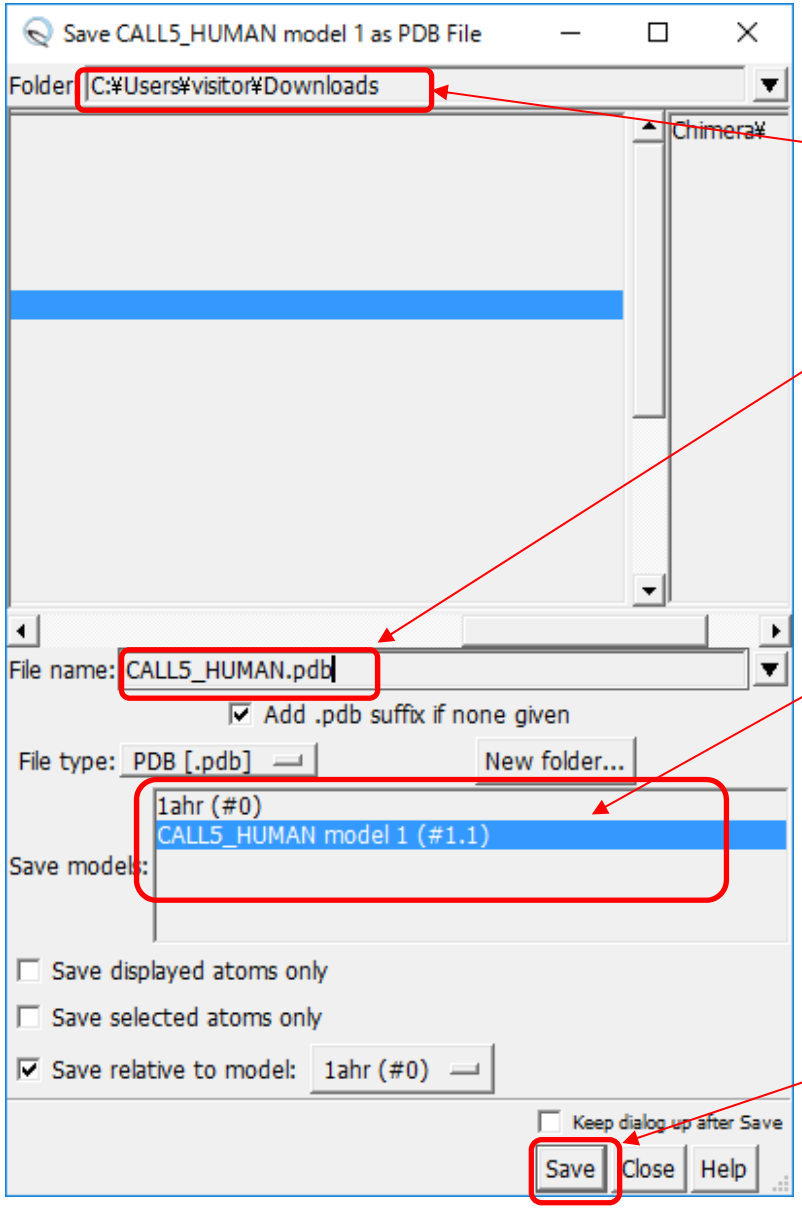
鋳型構造(1ahr)



モデル構造(1ahr)

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、

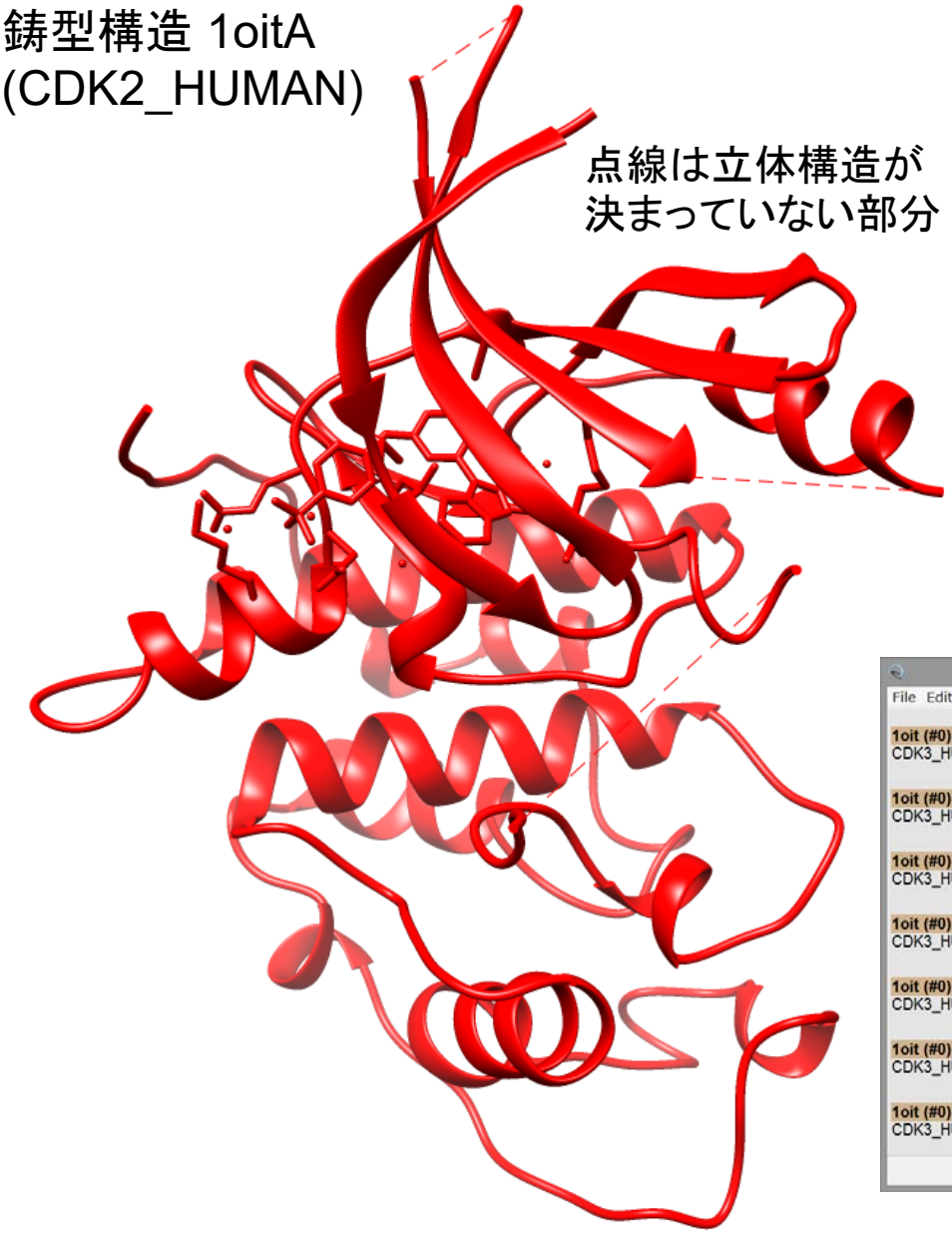
8) 適当なフォルダとファイル名（”CALL5_HUMAN.pdb”）を入力

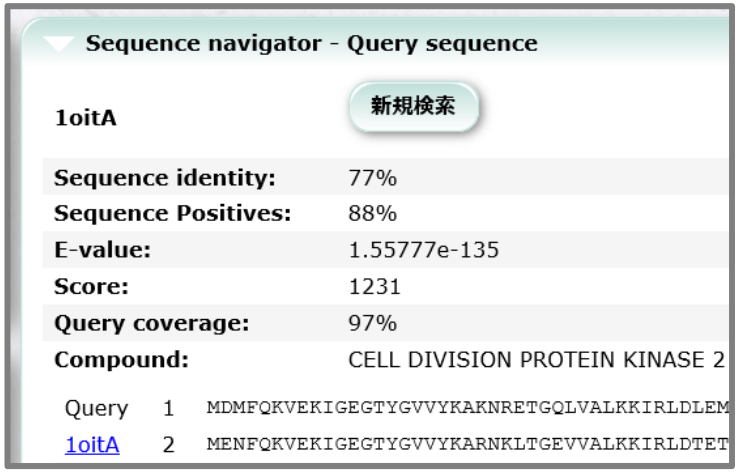
9)Save models:
では、保存したモデル（”CALL5_HUMAN　model 1 (#1.1)”）を選択

10) [Save]をクリック

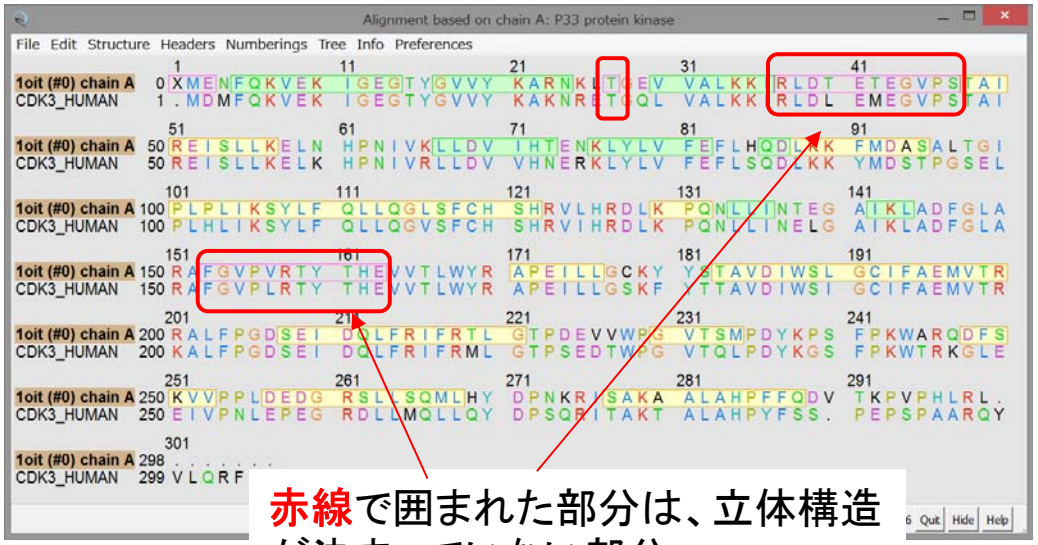
CDK3_HUMANのモデリング例

鋳型構造 1oitA
(CDK2_HUMAN)



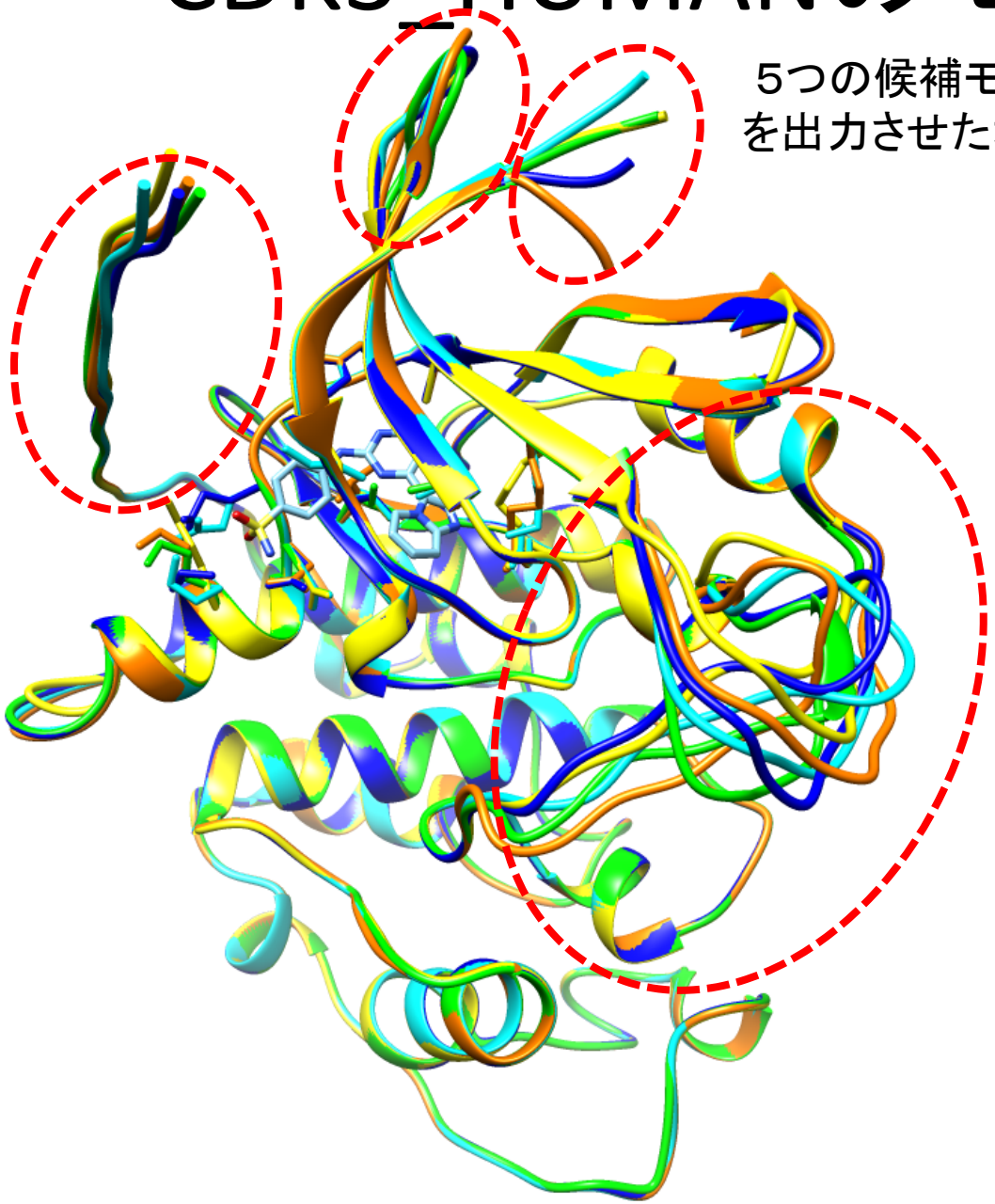


鋳型と標的配列のアラインメント



赤線で囲まれた部分は、立体構造が決まっていない部分

CDK3_HUMANのモデリング例



５つの候補モデル構造
を出力させた場合

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

・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。

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、
mod9.19 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。

alignment.aliの内容

```
>P1;query
sequence:query:2: :144: : : : :
AGELTPEEEAQYKKAFSAVDTDGNGTINAQELGAALKATGKNLSEAQLRKLISEVDSDGD
GEISFQEFLTA-AKKARAGLEDLQVAFRAFDQDGDGHITVDELRRAMAGLGQPLPQEELD
AMIREADVDQDGRVNYEEFARMLA--
*
>P1;1ahr
structureX:1ahr:1:A:148:A: : : :
ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGN
GTIDFPEFLTMMARKMKDSEEEIREAFRVFDKDGNGFISAAELRHVMTNLGEKLTDEEVD
EMIREADIDGDGQVNYEEFVTMMTSK
*
```

model.pyの内容

```
from modeller import *
from modeller.automodel import *
log.verbose()
env = environ()
env.io.atom_files_directory = ['.', '../atom_files']
env.io.hetatm = True
a = automodel(env,
              alnfile  = 'alignment.ali',
              knowns   = '1ahr',
              sequence = 'query')
a.starting_model= 1
a.ending_model  = 1
a.make()
```

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]→[Sequence]
・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

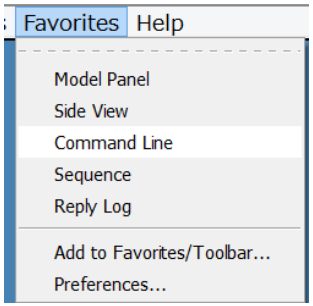
・ペプチド・化合物・核酸の構造構築 [Tools]→[Structure Editing]→[Build Structure]

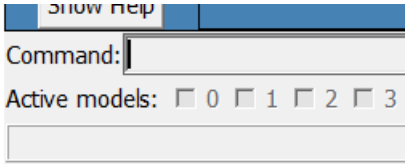
・水素原子の付加 [Tools]→[Structure Editing]→[AddH]

・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]→[Command Line]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 と ~display		原子の表示と非表示
ribbon と ~ribbon		リボンモデルの表示と非表示
surface と ~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].[鎖]@[原子名]	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

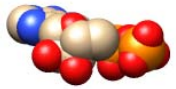
UCSF Chimeraの操作法などが載っている本

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

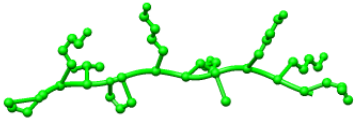
HOMCOSサーバを用いた複合体の検索とモデリング

複合体立体構造は
単量体構造より機能情報が豊富

Cyclin-dependent protein kinase (CDK2)

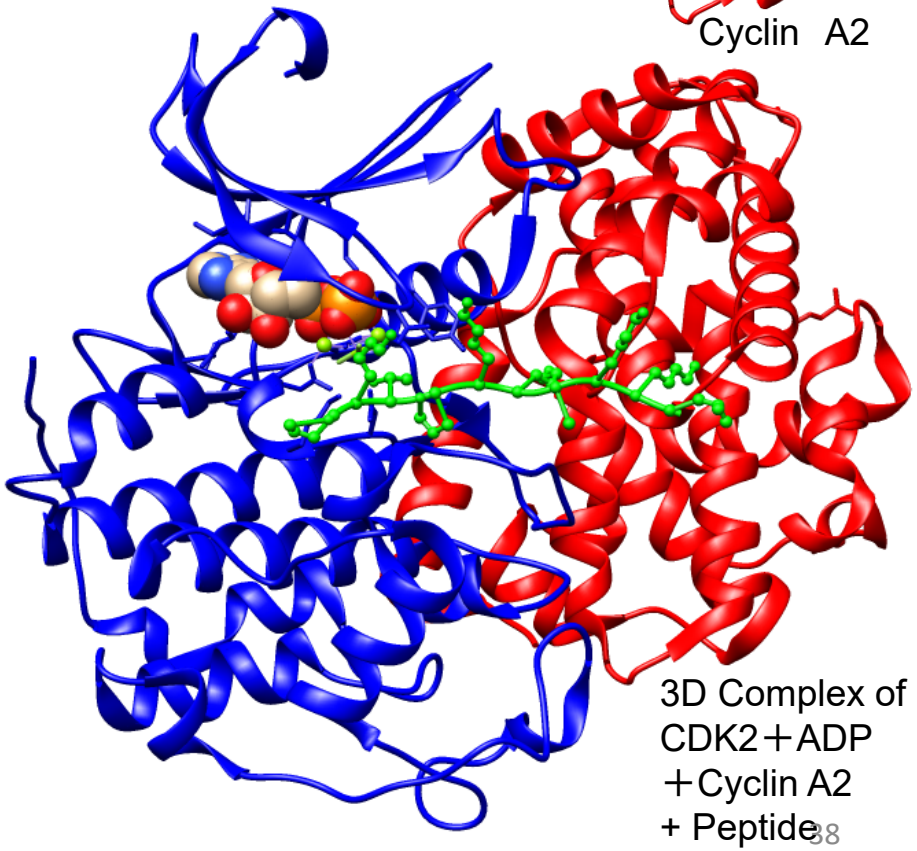


ADP



Substrate Peptide
PKTPKKAKKL

Cyclin A2



3D Complex of
CDK2＋ADP
＋Cyclin A2
+ Peptide
(PDBcode:3qhw)

複合体立体構造から以下のことがわかる

(1)他の分子との結合部位

→変異体の解釈・設計

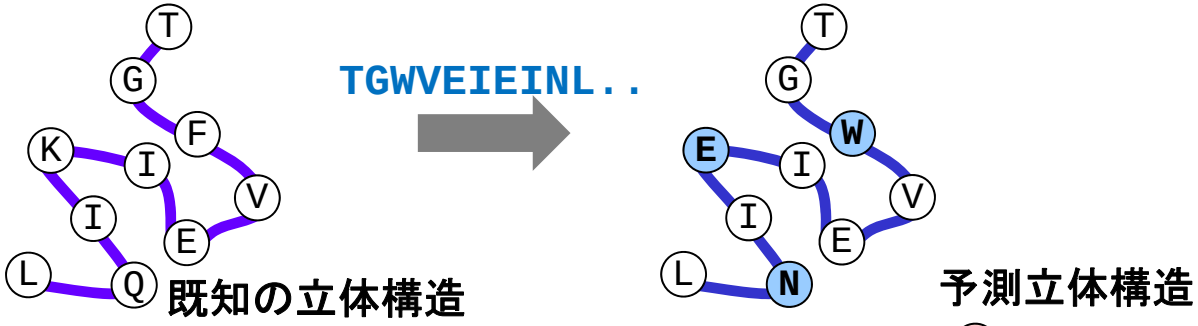
→ 阻害剤の設計・改変

(2)結合・反応のメカニズムの理解

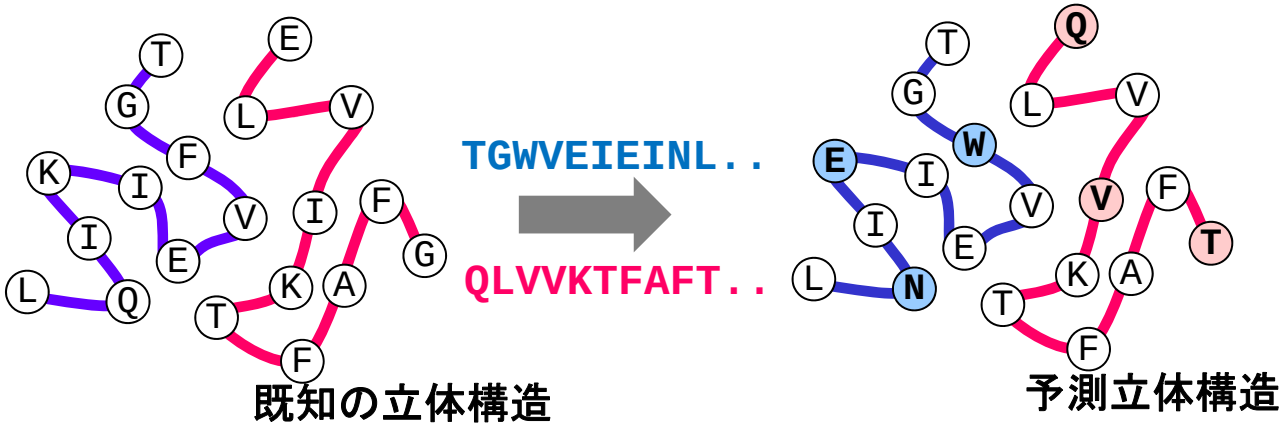
鋳型ベースのモデリング:Template-based Modeling

複合体のホモロジー・モデリング

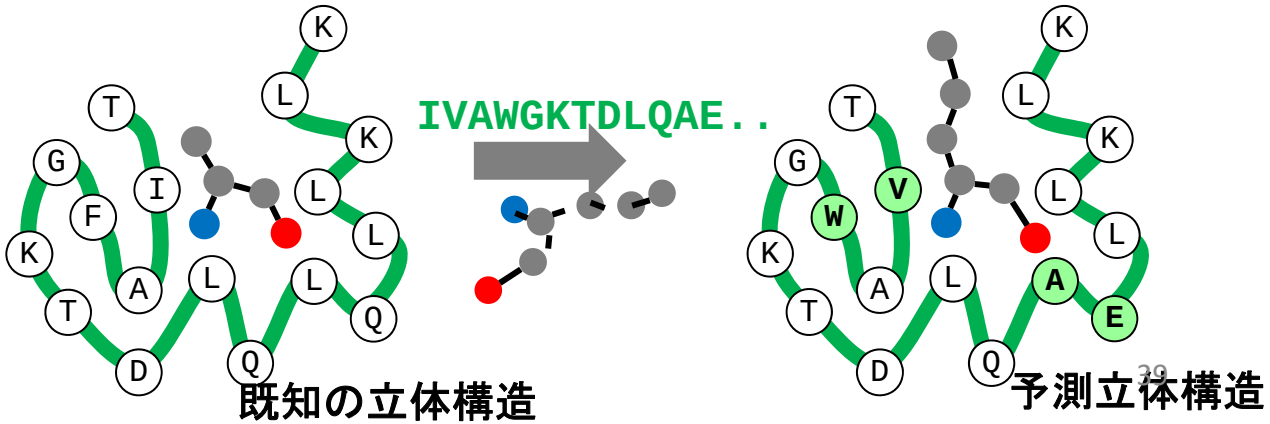
タンパク質単量体



タンパク質ータンパク質
複合体



化合物ータンパク質
複合体



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

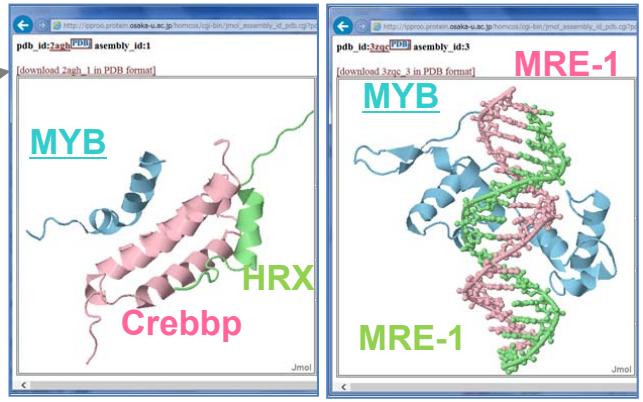
・配列相同性検索は**BLAST**、化学構造類似性検索は**KCOMBU**を使用

"**HOMCOS**"でグーグル検索
http://homcos.pdbj.org

VaProSからの検索では、**3D Interaction**に相当

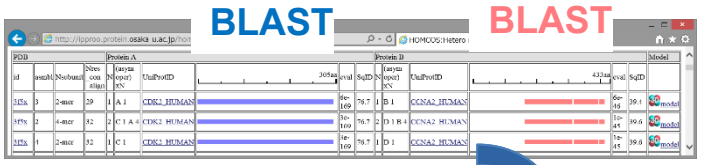
サービス	HOMCOS	入力1	入力2
PDB内の結合分子の検索	タンパク質に対する結合分子の検索	アミノ酸配列	
	化合物に対する結合分子の検索	化合物構造	
複合体立体構造のホモロジーモデリング	ホモ多量体モデル	アミノ酸配列	
	ヘテロ多量体モデル	アミノ酸配列A	アミノ酸配列B
	化合物ータンパク質複合体のモデル	アミノ酸配列	化合物構造

タンパク質に対する結合分子検索

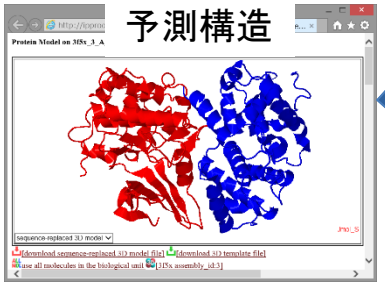


ヘテロ多量体のモデリング

アミノ酸配列2本を入力、それぞれ、PDBに対するBLASTを実行

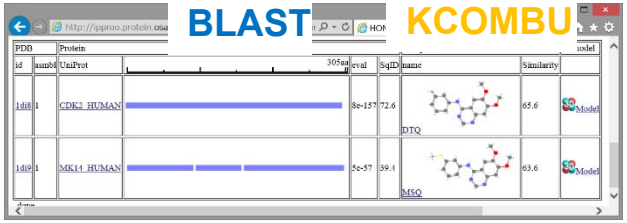


鋳型構造

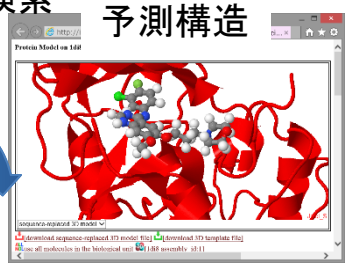


化合物ータンパク質複合体のモデル

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

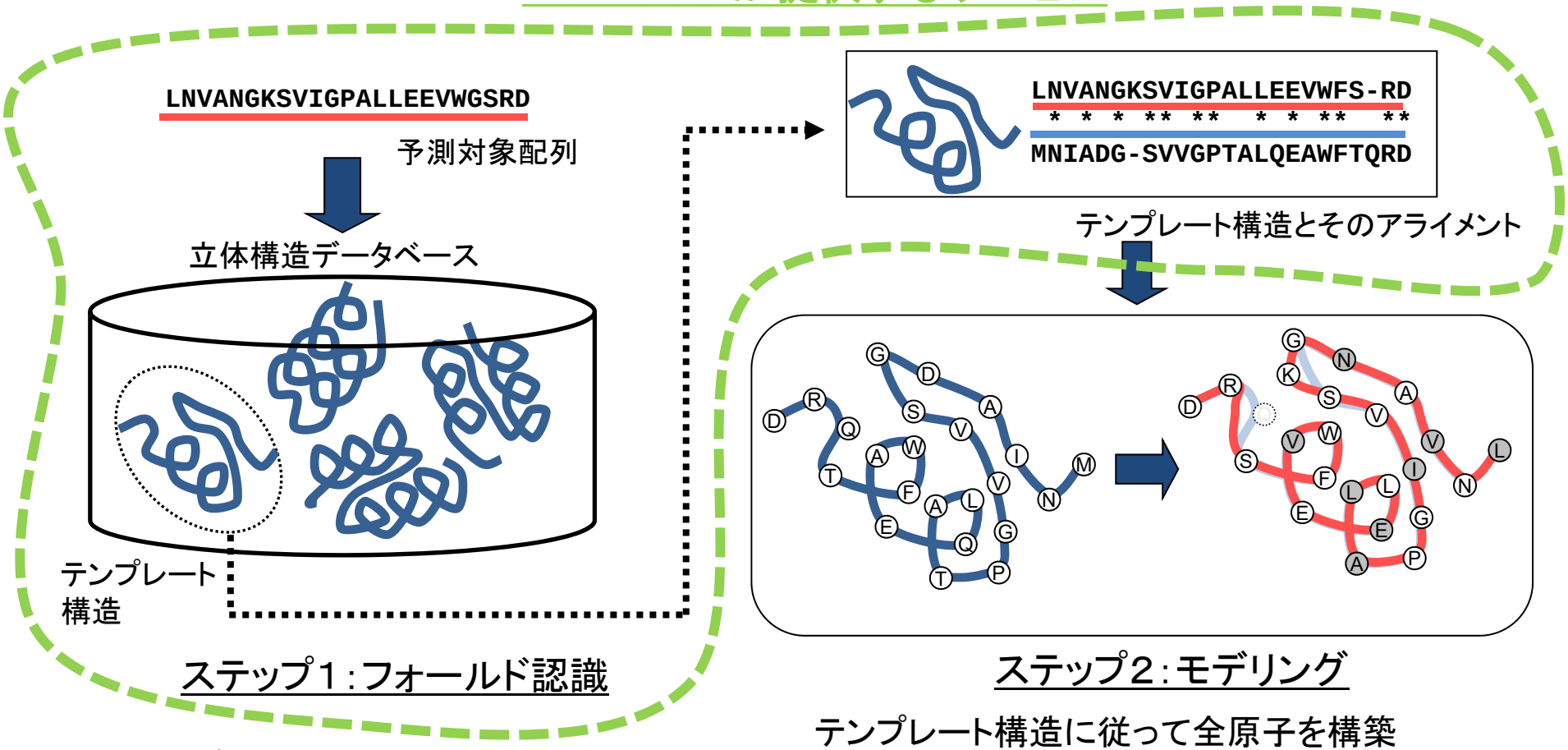


鋳型構造



ホモロジー・モデリングによる3次構造予測

HOMCOS が提供するサービス



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構造（テンプレート構造）を探す

ステップ2：モデリング

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

HOMCOSは
BLASTを採用

HOMCOSは、
（1）簡易モデルと（2）MODELLERの入力ファイルのみを提供

HOMCOSを用いた
蛋白質の結合分子の予測

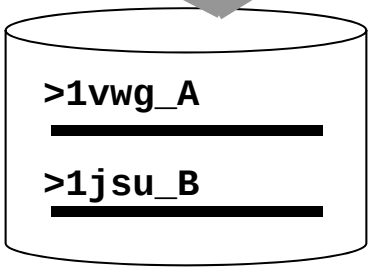
タンパク質に対する結合分子の検索

問い合わせ配列

TGWVEIEINL…

BLAST検索

PDBに登録された
アミノ酸配列の
データベース

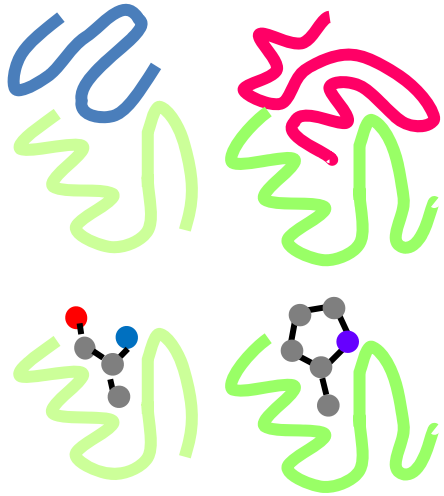


相同なタンパク質のリスト

```
2g9xA
1w98A
1fq1B
:
```

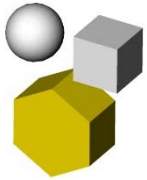
```
1vwg_1 A1 B1
2g9x_1 A1 B1
:
```

コンタクトしている分子の表

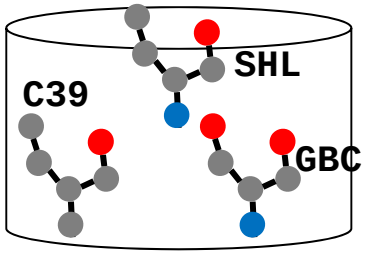


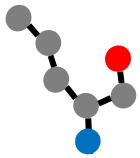
問い合わせ配列とコンタクトする
分子の予測リスト

化合物に対する結合タンパク質の検索



PDBに登録された
化合物の
データベース





問い合わせ化合物

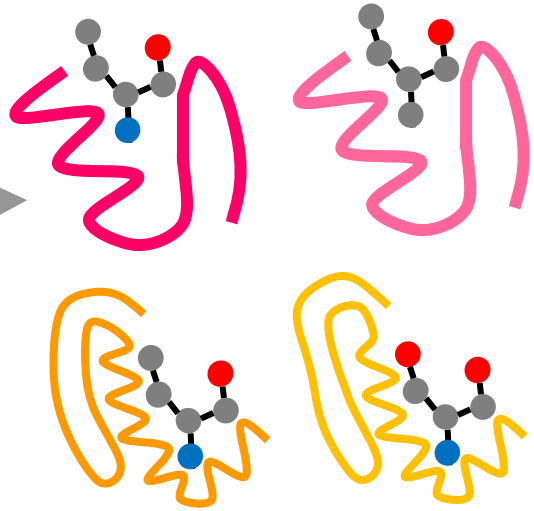
KCOMBU検索

```
SHL
C39
GBC
:
```

類似した化合物のリスト

```
1vwg_1 A1 B1
2g9x_1 A1 B1
:
```

コンタクトしている分子の表



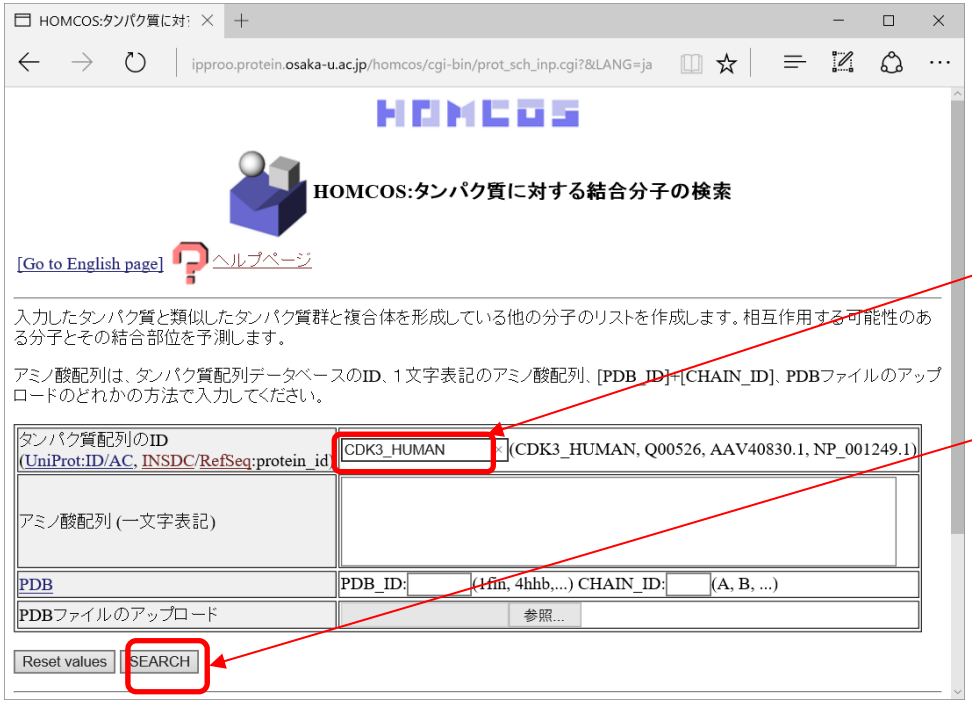
問い合わせ化合物とコンタクトする
分子の予測リスト

CDK3を題材をした結合分子予測

CDK3_HUMAN（Cyclin-dependent kinase 3）

1) Googleで"**HOMCOS**"と入力

2)「タンパク質に対する検索」を選ぶ



3) UniProtIDのフォームに
"**CDK3_HUMAN**"と入力して、
[SEARCH]をクリックする。

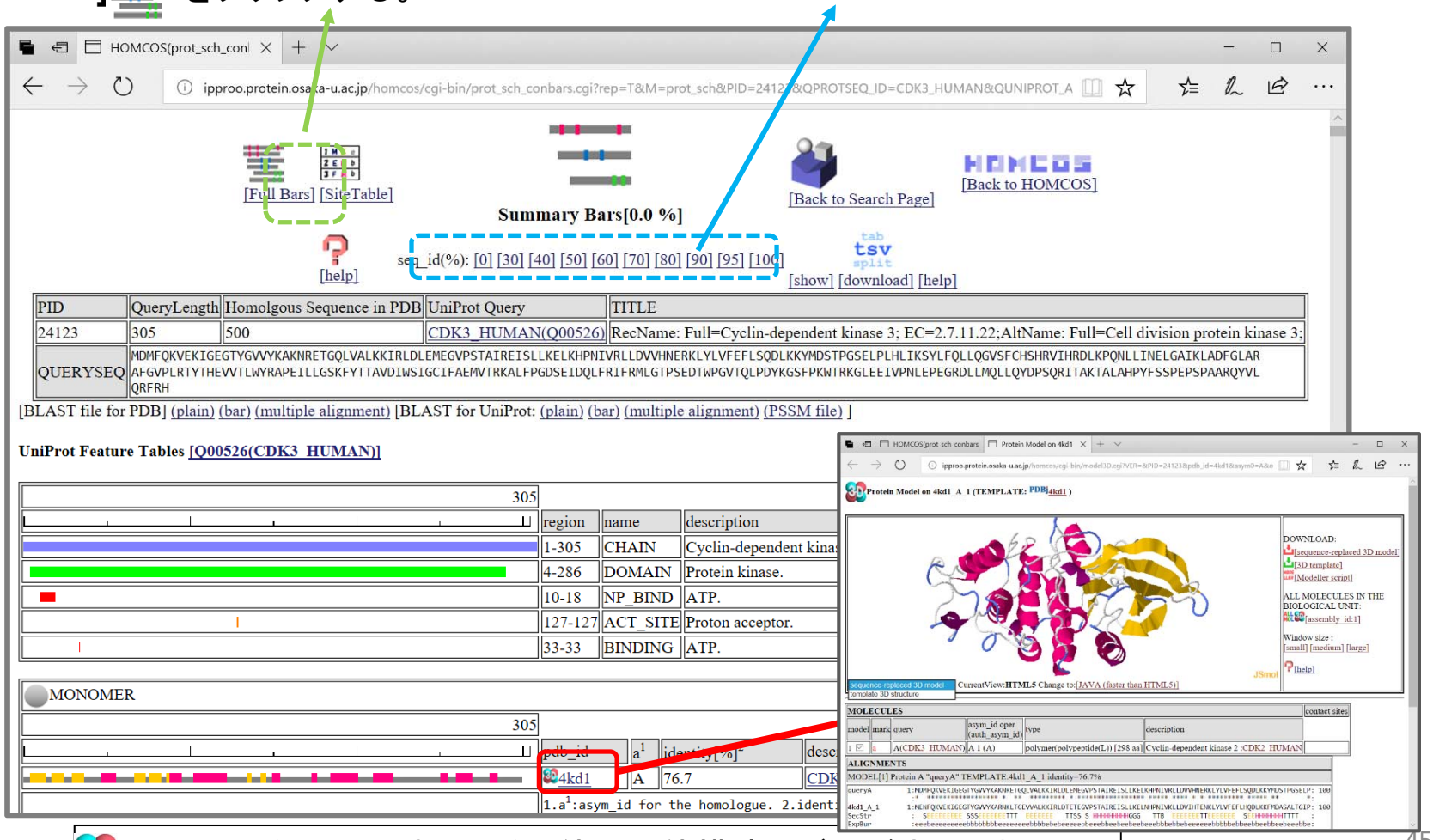
問い合わせ蛋白質の配列は以下の4通りで入力可
(i) PDB_ID+鎖 (ii) PDBファイルのアップロード
(iii) UniProt ID (iv) アミノ酸配列

「タンパク質に対する検索」結果のトップ画面(CDK3)

・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。

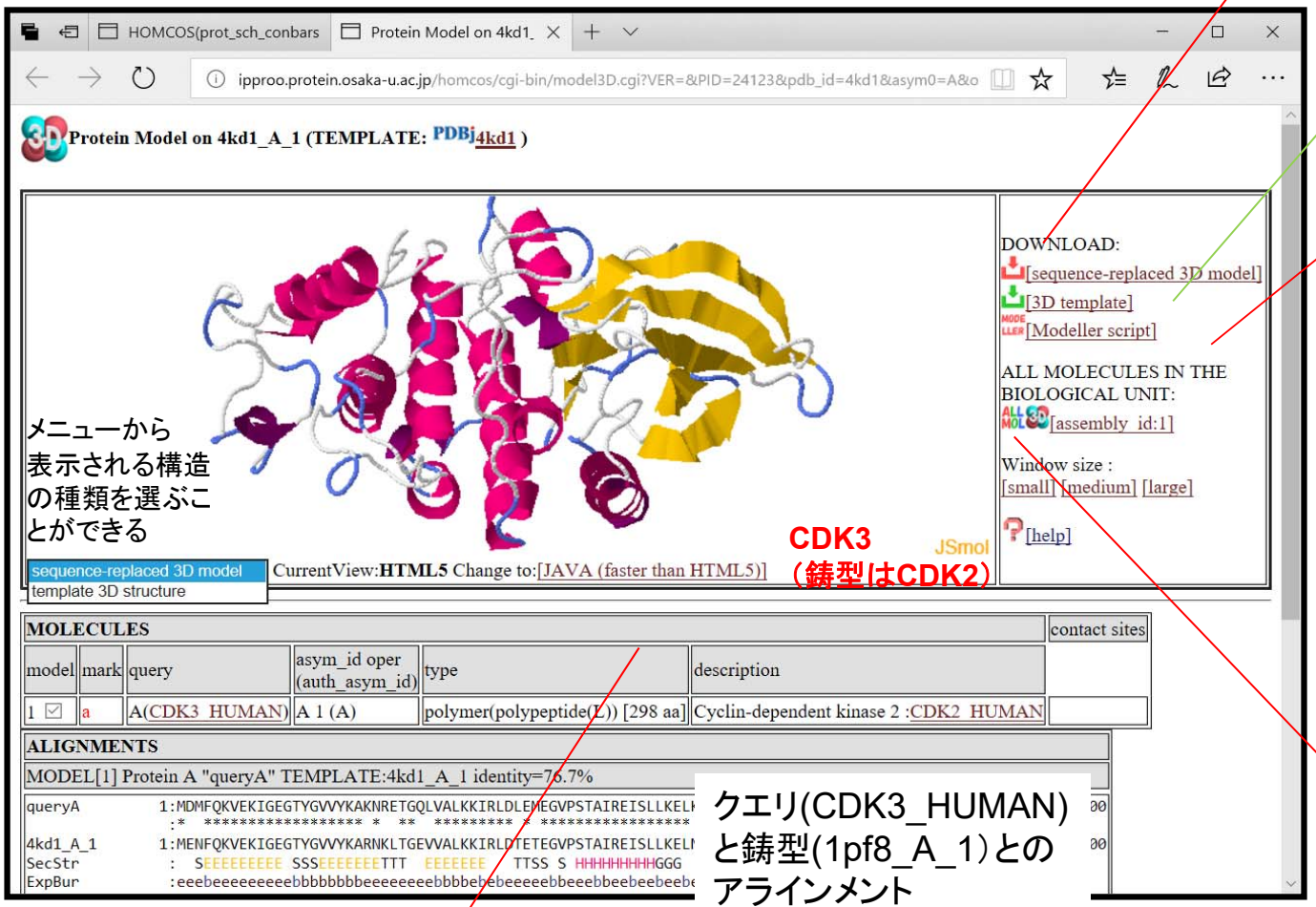
・相同性のしきい値は、デフォルトではE-value<0.001だけで、同一残基率は0%に設定。よりしきい値を上げれば(30%,40%,…,95%)、候補構造は減るが、予測の信頼性は向上。



3Dのアイコンをクリックすると単量体の立体構造モデルが表示される

単量体立体構造の表示(CDK3)

Sequence-replaced 3D model（簡易ホモロジーモデル構造）：
鋳型構造と座標は同じ。残基名と残基番号を標的配列と入れ替えてある。側鎖原子や挿入残基は正しくモデリングされていない。



メニューから表示される構造の種類を選ぶことができる

CDK3
（鋳型はCDK2）

クエリ(CDK3_HUMAN)と鋳型(1pf8_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。

簡易ホモロ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: 4kd1_A_1.pdb
4. execute command:

```
mod9 model1.py
```

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

マウスの使い方

表示法	UCSF Chimera	JSmol
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
分子の並進	ホイール(中ボタン)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
ズームイン・アウト	(1)ホイールを回す (2)右ボタンをドラッグ	(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
マウスによる残基名・原子名の確認	画面上で、カーソルを原子の上に置き、しばらく待つと、ポップアップでラベルが表示される	画面上で、カーソルを原子の上に置き、しばらく待つと、ポップアップでラベルが表示される

Contact Barのヘテロ複合体の画面(CDK3)

HETERO

305	pdb_id	contact mol		homologue			
		a[3]	description	a[4]	identity [%][5]	Ncon[6]	description
	5uq1 [4]	D	SPDYA_HUMAN Speedy protein A[139 aa]	C	92.3 /78.3	26 /26	CDK2_HUMAN Cyclin-dependent kinase 2
	1fq1 [1]	A	CDKN3_HUMAN CYCLIN-DEPENDENT KINASE INHIBITOR 3[183 aa]	B	94.1 /78.2	17 /17	CDK2_HUMAN CELL DIVISION PROTEIN KINASE 2
	4bcm [1]	D	CCNA2_HUMAN CYCLIN-A2[236 aa]	C	84.4 /78.2	32 /32	CDK2_HUMAN CYCLIN-DEPENDENT KINASE 2
	2cci[1]	E	CDC6_HUMAN CELL DIVISION CONTROL PROTEIN 6 HOMOLOG[19 aa]	A	91.7 /78.2	12 /12	CDK2_HUMAN CELL DIVISION PROTEIN KINASE 2
	2cci[1]	F	CDC6_HUMAN CELL DIVISION CONTROL PROTEIN 6 HOMOLOG[14 aa]	C	88.9 /78.2	9 /9	CDK2_HUMAN CELL DIVISION PROTEIN KINASE 2
	1qmz [4]	F	SUBSTRATE PEPTIDE[7 aa]	C	92.3 /78.2	13 /13	CDK2_HUMAN CELL DIVISION PROTEIN KINASE 2

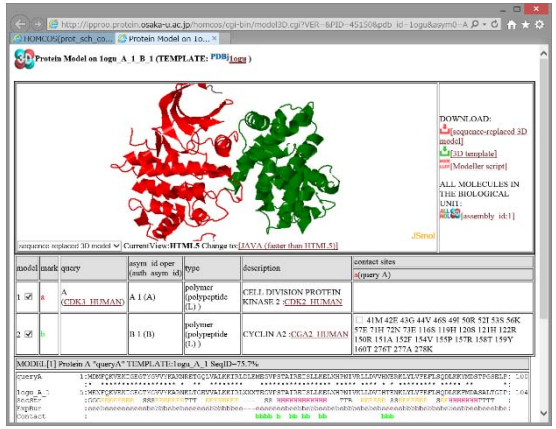
3.a[3]:asym_id for the contact molecule. 4.a[4]:asym_id for the template homologue. 5.identity[%][5]:sequence identity between the query and the template homologue only for the contact residues. Number after the slash / is sequence identity for all the aligned region. 6.Ncon[6]:number of aligned contact residues for the query. Number after the slash / is number of contact residues in the template homologue.

相互作用部位

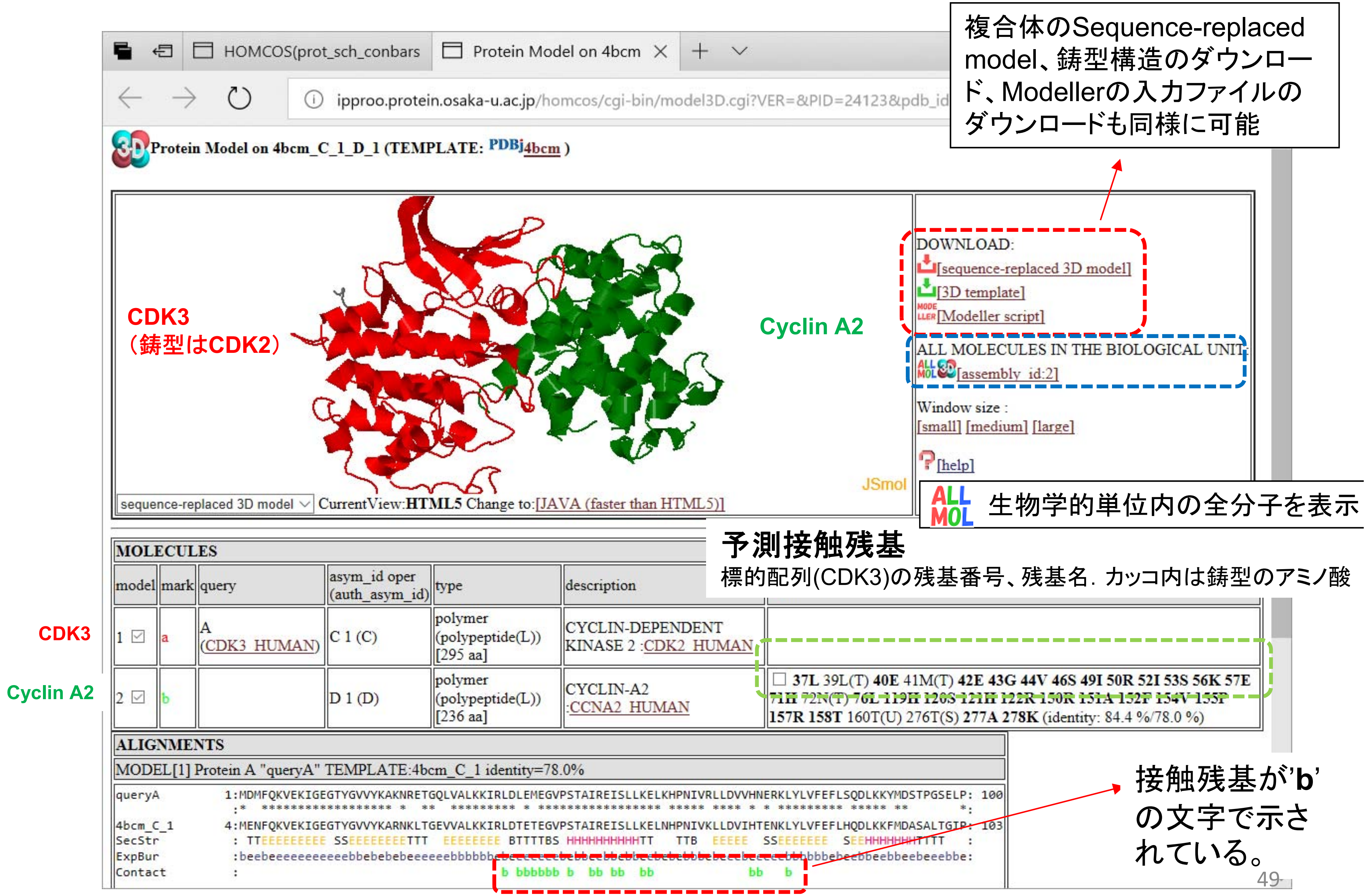
コンタクトしている別の
タンパク質の分子名

標的分子と鋳型の同一残基率(%)
値が高いほど予測の信頼性が高い。
最初の数字は、相互作用面の同一残基率、
二番目の数字は全体の同一残基率

のアイコンをクリックすると
ヘテロ複合体の
立体構造モデルが
表示される

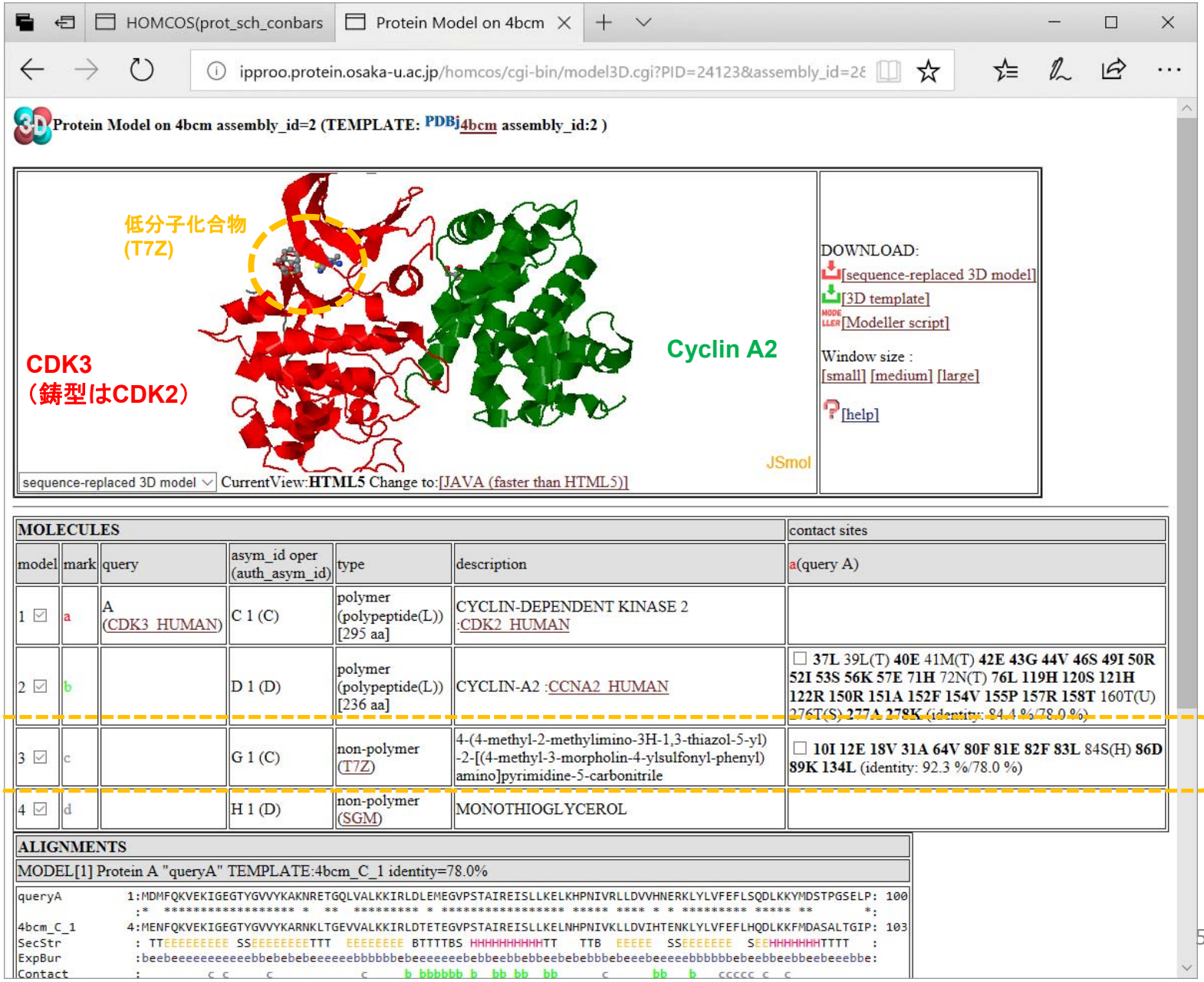


ヘテロ複合体立体構造の表示(CDK3)



生物学的単位内の全分子を表示した場合

特にDNA二重鎖を扱うときに ALL MOL が重要。



CDK3

Cyclin A2

低分子化合物 (T7Z)

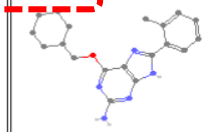
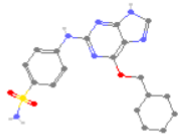
Contact Barの化合物複合体の画面(CDK3)

4QE, 4SPなどはPDBの3文字表記の分子名

3.a[3]:asym_id for the contact molecule. 4.a[4]:asym_id for the template homologue. 5.identity[%][5]:sequence identity between the query and the template homologue only for the contact residues. Number after the slash / is sequence identity for all the aligned region. 6.Ncon[6]:number of aligned contact residues for the query. Number after the slash / is number of contact residues in the template homologue.

COMPOUND

コンタクトしている
化合物の分子名

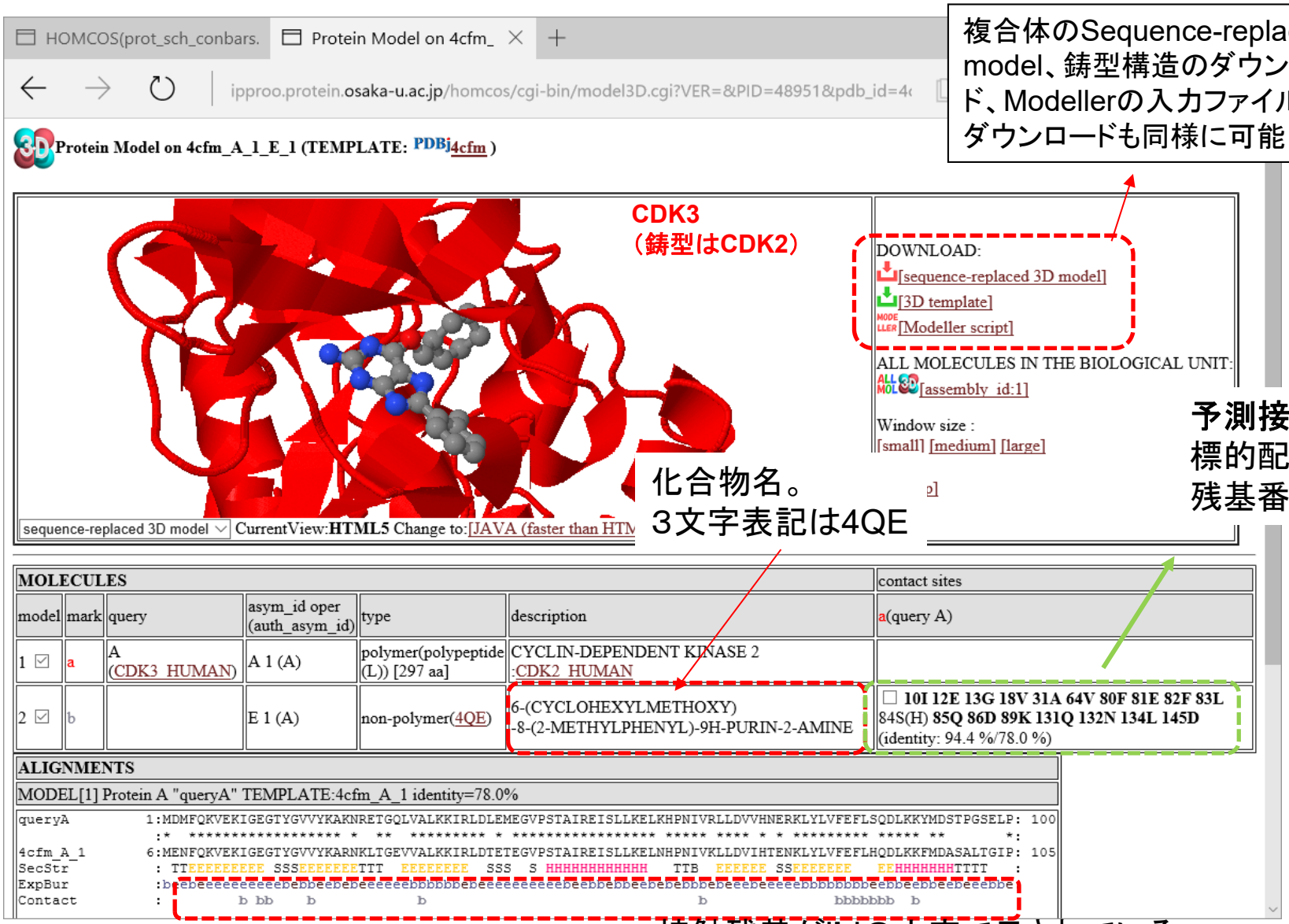
305	pdb_id	contact mol		homologue			
		a[3]	description	a[4]	identity [%][5]	Ncon[6]	description
	4cfm [2]	E	4QE  6-(CYCLOHEXYLMETHOXY) -8-(2-METHYLPHENYL)-9H-PURIN-..	A	94.4 /78.2	18 /18	CDK2_HUMAN CYCLIN-DEPENDENT KINASE 2
	1h1s [11]	E	4SP  O6-CYCLOHEXYLMETHOXY-2-(4'-SULPHAMOYLANILINO) PURI..	A	94.7 /78.2	19 /19	CDK2_HUMAN CELL DIVISION PROTEIN KINASE 2

相互作用部位

3D のアイコンをクリックすると
化合物ータンパク質複合体の
立体構造モデルが
表示される

標的分子と鋳型の同一残基率(%)
値が高いほど予測の信頼性が高い。
最初の数字は、相互作用面の同一残基率、
二番目の数字は全体の同一残基率

化合物ータンパク質複合体(CDK3)



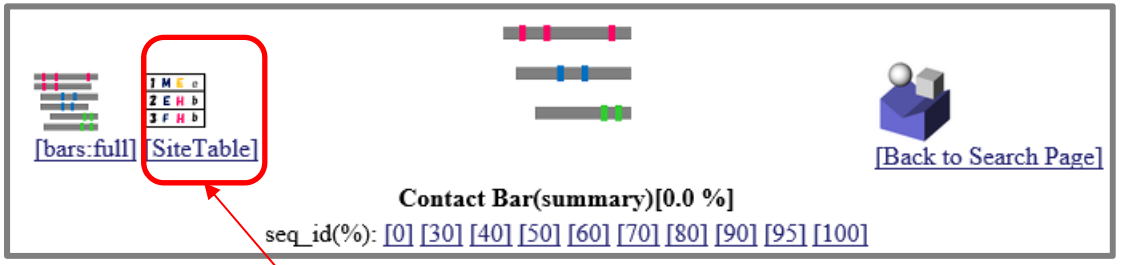
複合体のSequence-replaced model、鋳型構造のダウンロード、Modellerの入力ファイルのダウンロードも同様に可能

化合物名。
3文字表記は4QE

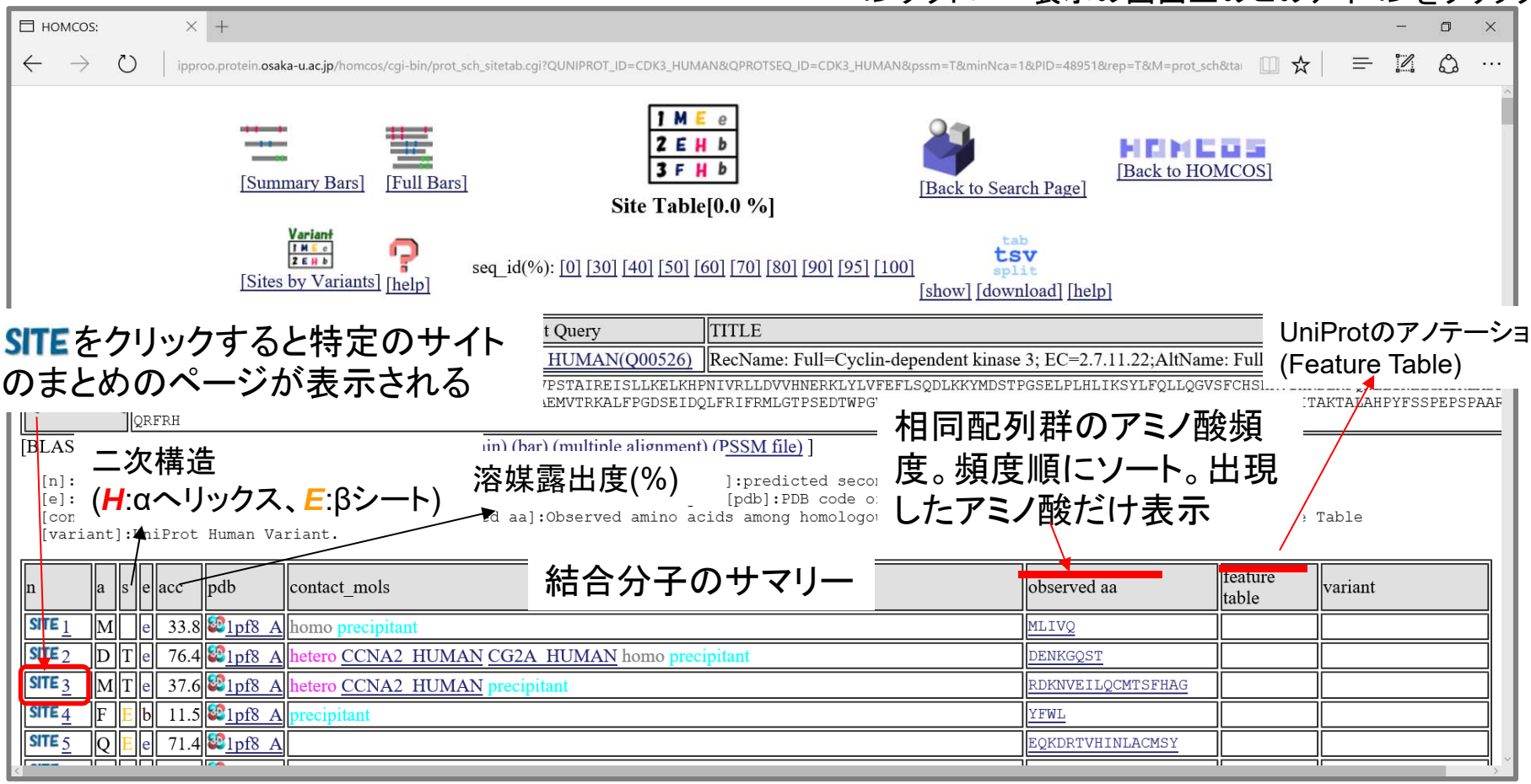
予測接触残基
標的配列(CDK3)の
残基番号、残基名

接触残基が'**b**'の文字で示されている。

Site Table



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



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

UniProtのアノテーション (Feature Table)

二次構造 (*H*:αヘリックス、*E*:βシート)

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

結合分子のサマリー

(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。

3番目の部位のまとめ(CDK3)

HOMCOS: | HOMCOS(prot_sch_each

ipproo.protein.osaka-u.ac.jp/homcos/cgi-bin/prot_sch_eachsite.cgi?QUNIPROT_ID=CDK3_H

SITE Summary for the 3-rd Site(M)

PID	QueryLength	FocusSite	TITLE
48951	305	3 M	RecName: Full=Cyclin-dependent kinase 3; EC=2.7.11.22;AltName: Full=Cell division protein kinase 3;

UniProt Information

AC/ID	AC:Q00526 ID:CDK3_HUMAN
Feature Table for 3-th site	CHAIN: Cyclin-dependent kinase 3. /FTId=PRO_0000085776.

Evolutionary Information

Percentage of Amino Acids in Homologous Protein
R:18% D:15% K:14% N:7% V:6% Q:5% E:5% I:5% L:5% C:4% M:4% T:4% S:3% H:2% F:2% A:1% G:1%

3D Structure Information

Template For Monomer	predicted SecStr	predicted ExpBur	Predicted Relative Acc(%)
1pf8	T (*Hbond turn*)	e (*exposed*)	37.6

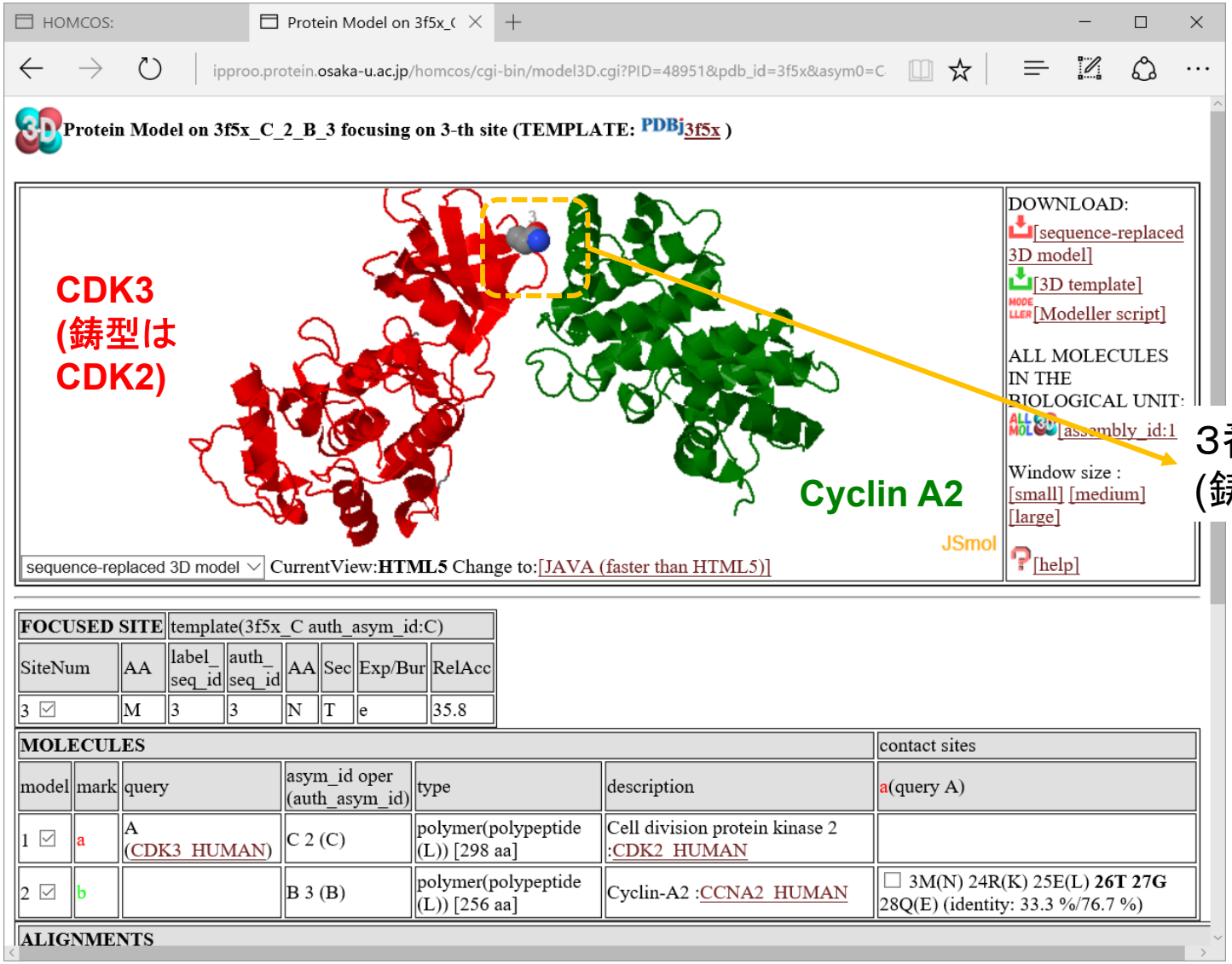
3D Complex Information

Predicted Bind Molecules
precipitant:1 hetero:8
Templates for 3D complexes
precipitant [**EDO**] 3qu0_A_1_E_1 **hetero** [**4052:CCNA2_HUMAN**] 3f5x_C_2_B_3 3f5x_A_4_D_1 2v22_A_1_D_1 2v22_C_1_B_1 5if1_C_1_B_1 1jst_A_1_D_1 3f5x_C_1_B_4 3f5x_A_3_D_2

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

これらのPDBのIDをクリックすると、この部位（3番目の部位）を結合サイトとする複合体立体構造のモデルが表示される。

3番目の部位がタンパク質間相互作用部位となる例(CDK3)



3番目のMet
(鋳型ではAsn)

ARSA_HUMANの場合のSite Tableの画面

ARSA_HUMAN(Arylsulfatase A)

HOMCOS:

ippro.protein.osaka-u.ac.jp/homcos/cgi-bin/prot_sch_sitetab.cgi?QUNIPROT_ID=ARSA_HUMAN&QPROTSEQ_ID=ARSA_HUMAN&pssm=1

[Summary Bars] [Full Bars]

Site Table[0.0 %]

[Back to Search Page] [Back to HOMCOS]

[Sites by Variants] [help]

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

[show] [download] [help]

PID	QueryLength	Homolgous Sequence in PDB	UniProt Query	TITLE
50092	507	66	ARSA_HUMAN (P15289)	RecName: Full=Arylsulfatase A; Short=ASA; EC=3.1.6.8;AltName: Full=Cerebroside-sulfatase;Contains: RecName: Full=Arylsulfatase A component B;Contains: RecName: Full=Arylsulfatase A component C;Flags: Precursor;

QUERYSE

溶媒露出度(%)

結合分子のサマリー

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)

UniProtの変異体に関する情報(VARIANT)

[BLAST file

[n]:site

[e]:predicted exposed/buri

[contact_mols]:binding oth

[variant]:UniProt Human Variant.

]:UniProt Feature Table

n	a	s	e	acc	pdb	contact_mols	observed aa	feature table	variant
SITE 1	M	-	-	-	-		M	SIGNAL	
SITE 2	G	-	-	-	-		AKG	SIGNAL	
SITE 27	F	E	b	0.5	1n2l_A		LMFIVT		
SITE 28	A	E	b	0.0	1n2l_A		ATVML		
SITE 29	D	S	b	0.0	1n2l_A	compound CSN metal CA MG MN ZN NA precipitant	D	METAL Calcium.	D->N:(0.0 %):Disease Leukodystrophy metachromatic (MLD) dbSNP:rs199476346 [MIM:250100]
SITE 30	D	S	b	1.2	1n2l_A	compound CSN metal CA MG NA precipitant	DQTH	METAL Calcium.	D->H:(2.0 %):Disease Leukodystrophy metachromatic (MLD) dbSNP:rs199476340 [MIM:250100]
SITE 31	L		b	0.0	1n2l_A		LQMVG		
SITE 32	G		b	0.0	1n2l_A		GDRANE		G->S:(0.0 %):Disease Leukodystrophy metachromatic (MLD) dbSNP:rs199476350 [MIM:250100]

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

(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。

29番目の部位のまとめ(ARSA)

HOMCOS: | HOMCOS(prot_sch_eacl

ipproo.protein.osaka-u.ac.jp/homcos/cgi-bin/prot_sch_eachsite.cgi?QUNIPROT_I

SITE Summary for the 29-th Site(D)

PID	QueryLength	FocusSite	TITLE
50092	507	29 D	RecName: Full=Arylsulfatase A; Short=ASA; EC=3.1.6.8;AltName: Full=Cerebroside-sulfatase;Contains: RecName: Full=Arylsulfatase A component B;Contains: RecName: Full=Arylsulfatase A component C;Flags: Precursor;

UniProt Information

AC/ID	AC:P15289 ID:ARSA_HUMAN
Feature Table for 29-th site	METAL: Calcium. VARIANT: D -> N (in MLD; infantile-onset; causes a severe reduction of enzyme activity; dbSNP:rs199476346). {ECO:0000269|PubMed:15326627}. /FTId=VAR_054165. STRAND: {ECO:0000244|PDB:1AUK}. VAR_SEQ: Missing (in isoform 2). {ECO:0000303|PubMed:14702039}. /FTId=VSP_046190. CHAIN: Arylsulfatase A component B. /FTId=PRO_0000033418. CHAIN: Arylsulfatase A. /FTId=PRO_0000033417.
VARIANT for 29-th site	D->N Disease dbSNP:rs199476346 "Leukodystrophy metachromatic (MLD) " [MIM:250100]

D→Nの変異は、Leukodystrophy metachromatic(MLD)[リソソーム病(異染性白質ジストロフィー)]という病気と関連がある。

Evolutionary Information

Percentage of Amino Acids in Homologous Protein
D:100%

29番目のDの進化的保存は極めてよい(100%)

3D Structure Information

Template For Monomer	predicted SecStr	predicted ExpBur	Predicted Relative Acc(%)
1n2l	S (*bend*)	b (*buried*)	0.0

3D Complex Information

Predicted Bind Molecules

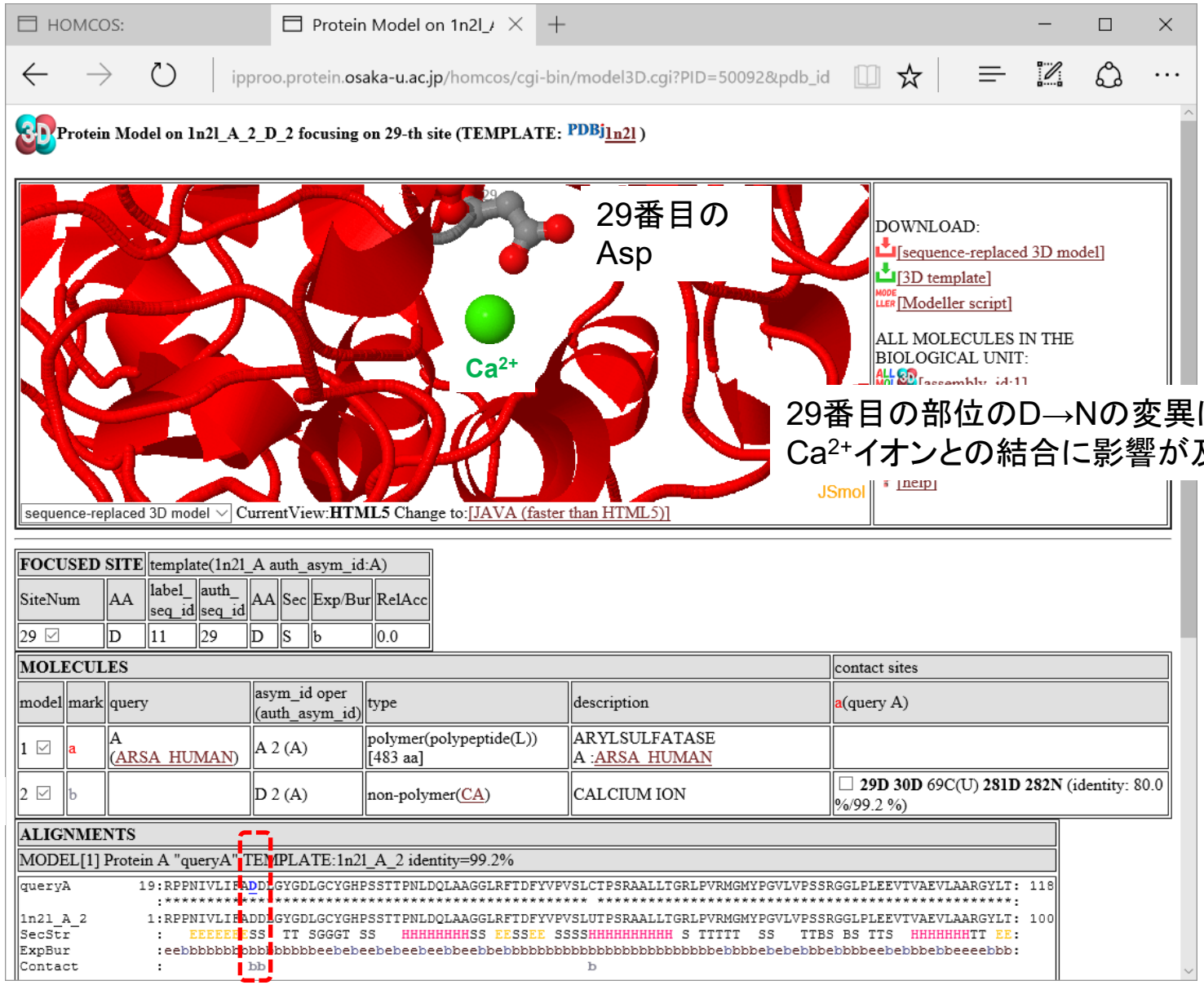
precipitant:3 metal:76 compound:2

Templates for 3D complexes

precipitant [SO4] 5aj9 B 1 Q 1 1hdh A 1 D 1 1hdh B 1 I 1 metal [ZN] 4upi A 2 B 2 4upi A 1 B 1 4upl B 1 D 1 4upl A 1 C 1 [NA] 3ed4 A 1 G 1 3ed4 B 1 L 1 3ed4 C 1 Q 1 3ed4 D 1 T 1 [MG] 4uph C 1 I 1 1auk A 4 C 4 1auk A 1 C 1 4uph D 1 K 1 1auk A 2 C 2 1e2s A 2 B 2 1e2s A 1 B 1 1e1z A 2 C 2 1auk A 6 C 6 1e3c A 1 C 1 1e33 A 1 C 1 4uph A 1 E 1 1auk A 5 C 5 4uph B 1 G 1 1e1z A 1 C 1 1e33 A 2 C 2 1auk A 7 C 7 1auk A 8 C 8 1auk A 3 C 3 1e3c A 2 C 2 [MN] 4ug4 G 1 O 1 4ug4 B 1 J 1 4ug4 E 1 M 1 4ug4 E 1 N 1 4ug4 G 2 O 2 4ug4 C 3 K 3 4ug4 C 1 K 1 4ug4 F 2 N 2 4ug4 H 1 P 1 4ug4 D 1 L 1 4ug4 A 1 I 1 4ug4 A 3 I 3 [CA] 5aj9 A 1 D 1 1n2l A 2 D 2 1hdh B 1 H 1 4cys B 1 I 1 1n2k A 1 D 1 1p49 A 1 F 1 4miv E 1 FA 1 4cxs B 1 G 1 4cxu A 1 C 1 4cyr A 1 J 1 4fdj B 1 I 1 4cxk A 1 C 1 4fdj A 1 E 1 1p49 A 3 F 3 1n2k A 2 D 2 1fsu A 1 D 1 1hdh A 1 C 1 5aj9 B 1 P 1 4cxs A 1 C 1 4miv B 1 O 1 4miv C 1 T 1 4miv F 1 IA 1 4cxk B 1 F 1 1n2l A 1 D 1 4cyr B 1 P 1 4cxu B 1 E 1 4mhx B 1 H 1 4fdi A 1 E 1 4cys A 1 C 1 4miv G 1 NA 1 4miv D 1 Z 1 4miv A 1 I 1 1p49 A 2 F 2 4miv H 1 RA 1 4fdi B 1 J 1 4mhx A 1 C 1 compound [CSN] 1e2s A 1 D 1 1e2s A 2 D 2

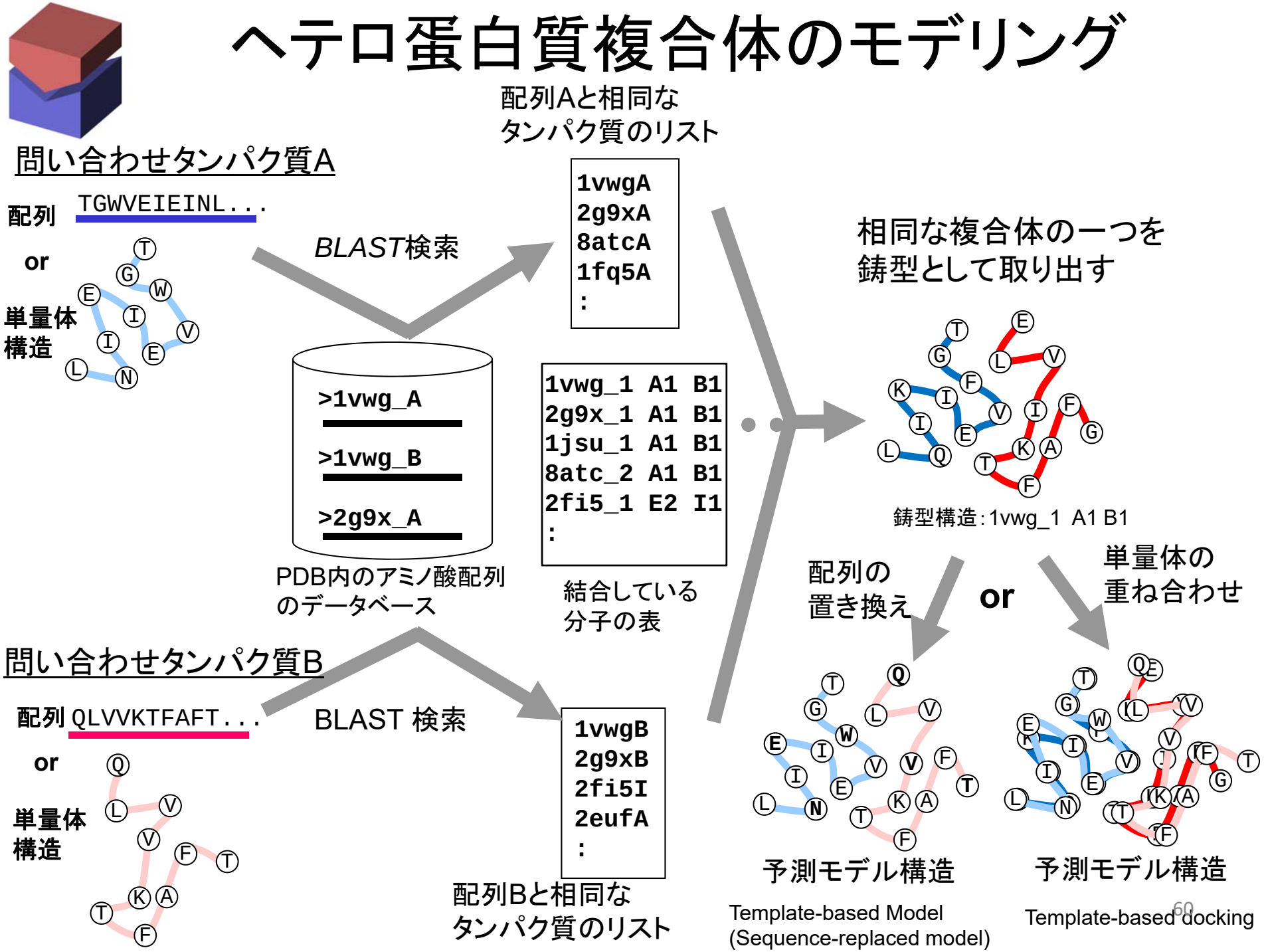
カルシウムイオン(Ca^{2+})との複合体構造が予測されている。

29番目の部位に注目した
Ca^{2+}イオン－タンパク質複合体の予測結合構造



HOMCOSを用いた
ヘテロ蛋白質複合体の予測

ヘテロ蛋白質複合体のモデリング



ヘテロ多量体のモデリング(2本の配列から)

1) Googleで"**HOMCOS**"と入力

2)「ヘテロ多量体のモデル」を選ぶ

3) タンパク質AのUNIPROT_IDに**CDK5_HUMAN**を
タンパク質BのUNIPROT_IDに**CCNB1_HUMAN**を入力

CDK5_HUMAN: Cyclin-dependent proten kinase 5
CCNB1_HUMAN : G2/mitotic-specific cyclin B1

問い合わせ蛋白質の配列は以下の4通りで入力可
(i) PDB_ID+鎖 (ii) PDBファイルのアップロード
(iii) UniProt ID (iv) アミノ酸配列

ヘテロ多量体のモデリング（2本の配列から）

配列Aと配列Bについて、対PDBの***BLAST*** 検索が実行される

sequence-replaced model
template 3D structure
の二つを、表示・ダウンロード可能

複合体のsequence-replaced 3D model

HOMCOS:Hetero multin

ippro.protein.osaka-u.ac.jp/homcos/cgi-bin/hetero_model_res.cgi?M=hetero_mode

proteinA	TITLE	RecName: Full=Cyclin-dependent-like kinase 5; EC=2.7.11.1;AltName: I catalytic subunit; Short=TPKII catalytic subunit;
	UniProt_ID for query	CDK5_HUMAN
	QUERYSEQ (length 292)	MQKYEKLEKIGEGTYGTVFKAKNRETHEIVALKRVRLDDDDEGVPSSALREICLLKELK FGIPVRCYSAEVVTLWYRPPDVLFGAKLYSTSIDMWSAGCIFAELANAGRPLFPGNDVD
	BLAST	Nsbjseq 500 [show plain BLAST file]
proteinB	TITLE	RecName: Full=G2/mitotic-specific cyclin-B1;
	UniProt_ID for query	CCNB1_HUMAN
	QUERYSEQ (length 433)	MALRVTRNSKINAENKAKINMAGAKRVPTAPAATSKPGLRPRTALGDIGNKVSE LAVNDVDAEDGADPNLCSEYVKDIYAYLRQLEEEQAVRPKYLLGREVTGNMRAI LPLHFLRRASKIGEVDVEQHTLAKYLMELTMLDYDMVHFPPSQIAAGAFCLALK
	BLAST	Nsbjseq 224 [show plain BLAST file]

PID	50341	precalc_blt	T	minNca	1	minRca	0	Nmodel	208

Homologous hetero-multimer found by BLAST

asmbl:assembly_id Npro:number of proteins homologo
identity[%]:sequence identity only for the contact re
Ncon:number of aligned contact residues. Number afte

CDK5_HUMAN is sequence identity for all the aligned region. residues in the template homologue. CCNB1_HUMAN

PDB				Protein A						Protein B				
id	asmbl	Npro	N	asym oper	UniProtID	292aa	identity [%]	Ncon	N	asym oper	UniProtID	433aa	identity[%]	Ncon
2jgz	1	2-mer	1	A 1	CDK2_HUMAN		60.0/ 61.2	30/30	1	B 1	CCNB1_HUMAN		100.0/100.0	29/29
5lqb	1	2-mer	1	A 1	CDK1_HUMAN		62.5/ 58.0	24/24	1	B 1	CCNB1_HUMAN		100.0/ 98.9	23/23
4y72	1	2-mer	1	A 1	CDK1_HUMAN		62.5/ 58.0	24/24	1	B 1	CCNB1_HUMAN		100.0/ 98.9	23/23
4yc3	1	2-mer	1	A 1	CDK1_HUMAN		61.5/ 57.1	26/26	1	B 1	CCNB1_HUMAN		100.0/ 98.9	25/25
1w98	1	2-mer	1	A 1	CDK2_HUMAN		71.4/ 60.7	35/35	1	B 1	CCNE1_HUMAN		27.6/ 40.5	29/39
4bco	1	2-mer	1	A 1	CDK2_HUMAN		66.7/ 60.7	30/30	1	B 1	CCNA2_BOVIN		44.8/ 39.8	29/30

HOMCOS:Hetero m

ippro.protein.osaka-u.ac.jp/homcos/cgi-bin/model3D.cgi?pdb_id=2jgz&PID=50341&sbjidA=2jgz_A

Protein Model on 2jgz_A_1_B_1 (TEMPLATE: PDB 2jgz)

sequence-replaced 3D model CurrentView:HTML5 Change to:[JAVA (faster than HTML5)]

JSmol

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

ALL MOLECULES IN THE BIOLOGICAL UNIT:
[assembly_id:1]

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

[help]

MOLECULES						contact sites	
model	mark	query	asym_id oper (auth_asym_id)	type	description	a(query A)	b(query B)
1	a	A(CDK5_HUMAN)	A 1 (A)	polymer (polypeptide (L)) [289 aa]	CELL DIVISION PROTEIN KINASE 2 :CDK2_HUMAN		167C 169E 170Y 171V 173D 174I 175Y 177Y 178L 181L 219L 221E 253F 257K 258Y 259E 260E 261M 265E 266I 283R 286E 287M 290L 291R 295F 296G 297L 298G (identity: 100.0 %/100.0 %)
2	b	B (CCNB1_HUMAN)	B 1 (B)	polymer (polypeptide (L)) [260 aa]	G2/MITOTIC-SPECIFIC CYCLIN-B1 :CCNB1_HUMAN	40D(E) 41D(T) 42E 43G 44V 46S 49L(I) 50R 52I 53C(S) 57E 69V 71H 115G(A) 118H 119S 120R(H) 121N(R) 149R 150A 151F 153I (V) 156R 157C(T) 158Y 159S(U) 272V(N) 276S 277A 278E(K) (identity: 60.0 %/61.2 %)	

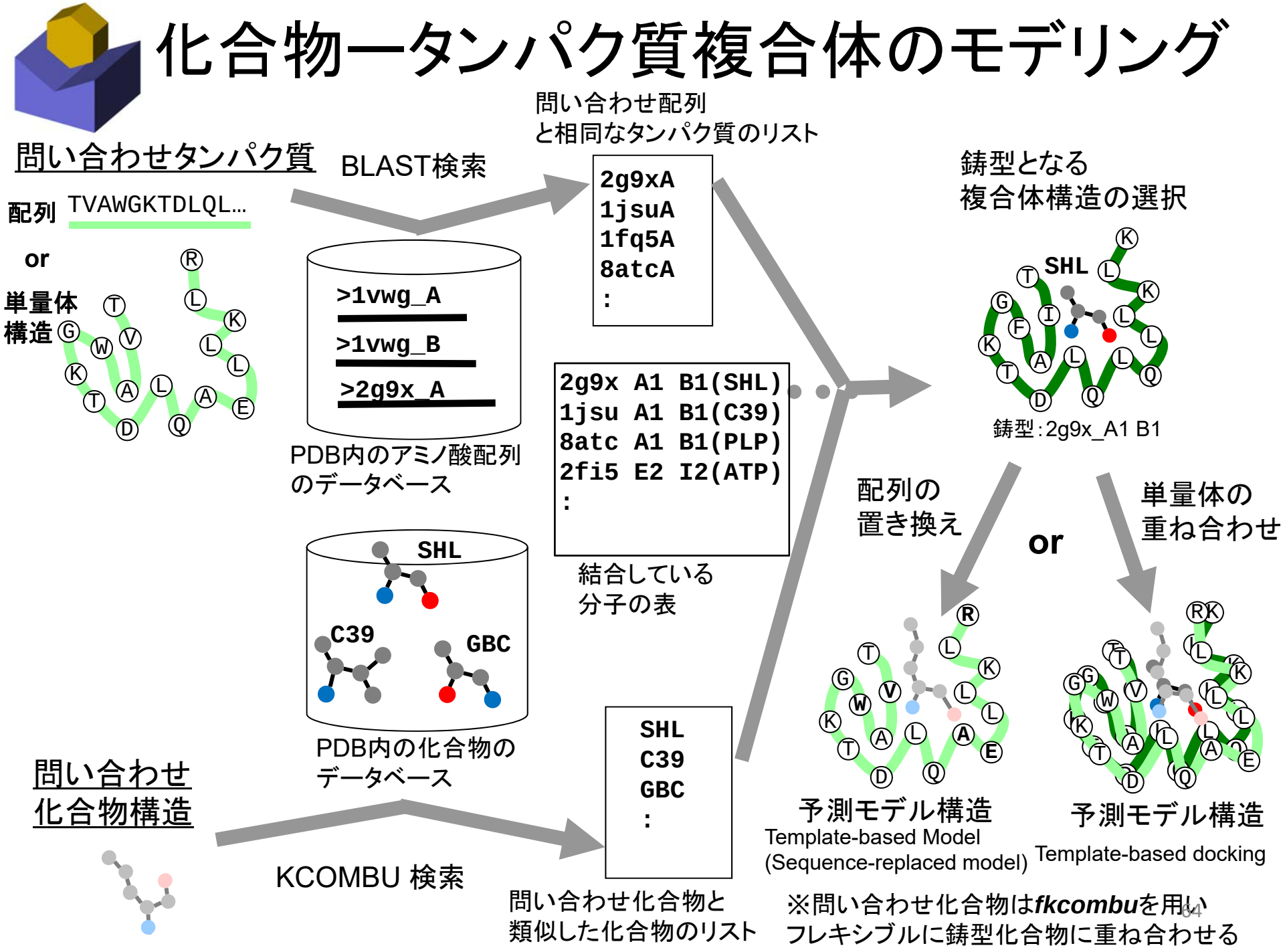
ALIGNMENTS

MODEL[1] Protein A "CDK5_HUMAN" TEMPLATE:2jgz_A_1 identity=61.2%

```
CDK5_HUMAN     1:MQKYEKLEKIGEGTYGTVFKAKNRETHEIVALKRVRLDDDDEGVPSSALREICLLKELKHKNIVRLHDVLHSDKKLTLVFEFCDQDLKKYFDSCNGDLDP: 100
                 :*     * ********* * ** *   * * ****  ***   ***** * *** ***** * *** * ** *   ** *****  *****  *       *:
2jgz_A_1       2:MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIREISLLKELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIP: 101
SecStr          : TTeeeeeeee SSSeeeeeeeTTT  eeeeeeeSSS TTSS HHHHHHHHHHH   TTB  eeeeeeeTTeeeeeee  SSHHHHHHHTTTS   :
ExpBur          :beebeeeeeeeebeeeebbebebeeeeeebbbbbbeeeeeeeeebbbebbebbeebebebbbebeeebeeeeebbbbbbebeebbeeebeeebeeeeebe:
Contact         :                                        bbbbb b  bb bb   b          b b
```

HOMCOSを用いた
化合物ー蛋白質複合体の予測

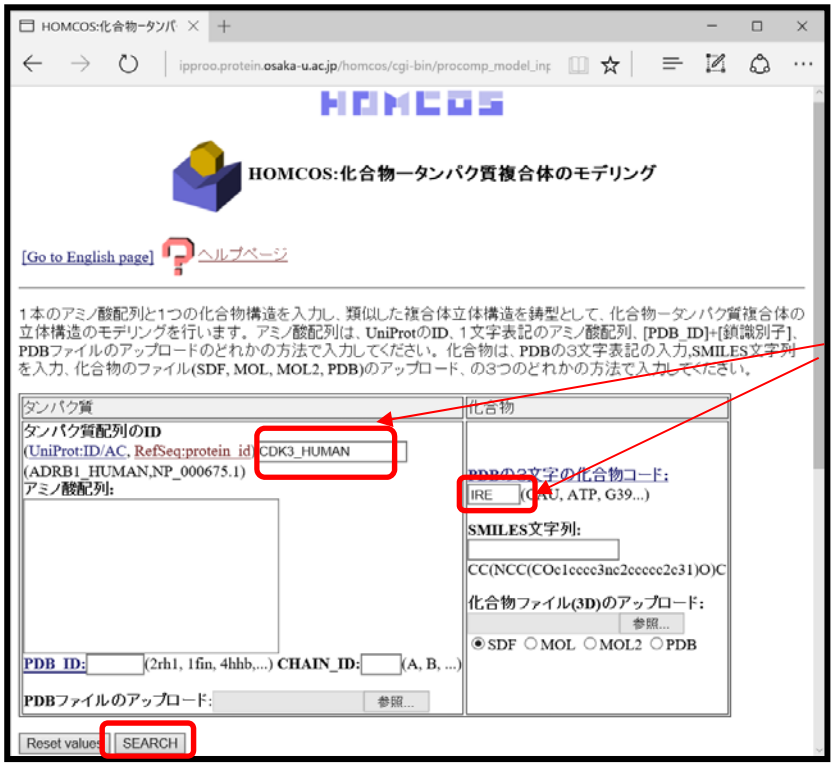
化合物ータンパク質複合体のモデリング



化合物タンパク質複合体モデリングのページ

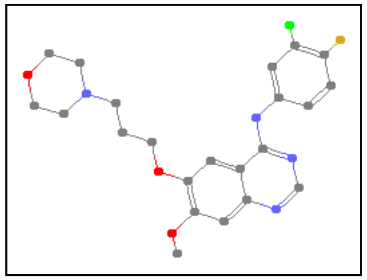
1) Googleで"**HOMCOS**"と入力

2)「化合物タンパク質複合体のモデル」を選ぶ



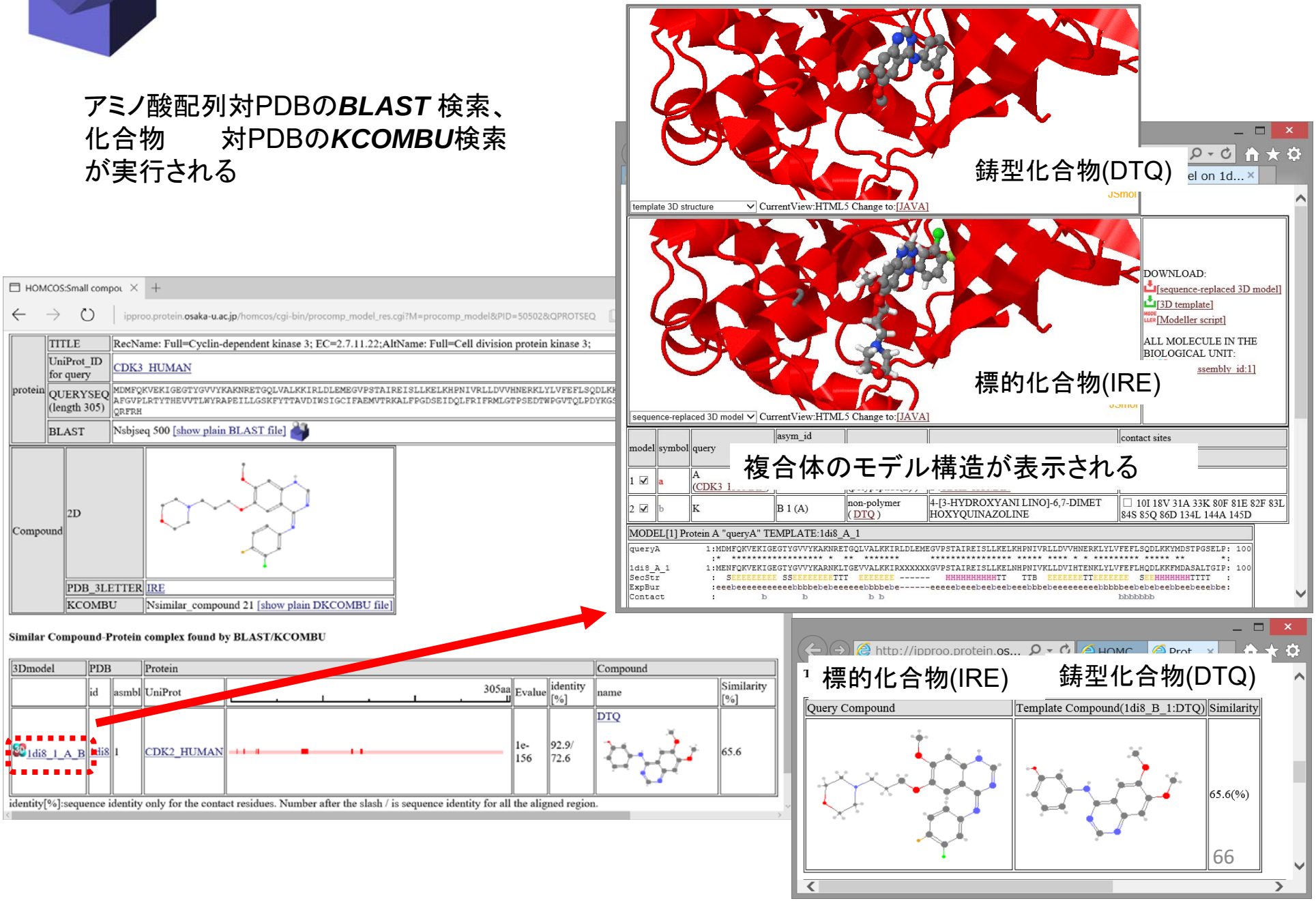
3) PROTEINのUNIPROT_IDには**CDK3_HUMAN**を
COMPOUNDのPDB three letter ligand codeには**IRE**を入力

Iressa/Gefitinib (IRE)



化合物ータンパク質複合体のモデリング

アミノ酸配列対PDBの***BLAST*** 検索、
化合物　　対PDBの***KCOMBU***検索
が実行される



鋳型構造が全く見つからない場合は？

１．剛体ドッキングプログラムを試す。

高分子なら、ZDOCK,　HADDOCK、低分子なら、DOCK,　AutoDock vina, sievgene, GLIDE　などがよく使われる。もちろん、単体の立体構造が入手できる場合に限る。

⇒　予測精度の限界を留意。結合部位などの実験情報をできるだけ考慮したほうがよい。

２．分子動力学法などの高度なサンプリング法。

結合に伴って大きな構造変化がある場合は、分子動力学法などの分子変形を伴うサンプリング法が必須なはず。

⇒　計算量が膨大。専門家の助言を求めたほうがよい。

３．手作業で組み立てる

⇒　実験情報がある程度あるなら、全く不可能でもない。

USCF Chimeraの[Favorites]→[Model Panels]でモデルパネルを立ち上げ、[A]の☑をオン・オフしながら、分子の姿勢を変えていく。

結合の強さの予測について

1. 結合の強さの予測は、結合構造の予測より困難

・物理化学な手法による強さの予測は、正しい結合立体構造が必要。
結合構造に誤りがあると強さは正確に予測できない。
・生体分子間の相互作用は一般に弱く、1アミノ酸置換でも結合定数が
変わる可能性。精密な見積もりが要求される。

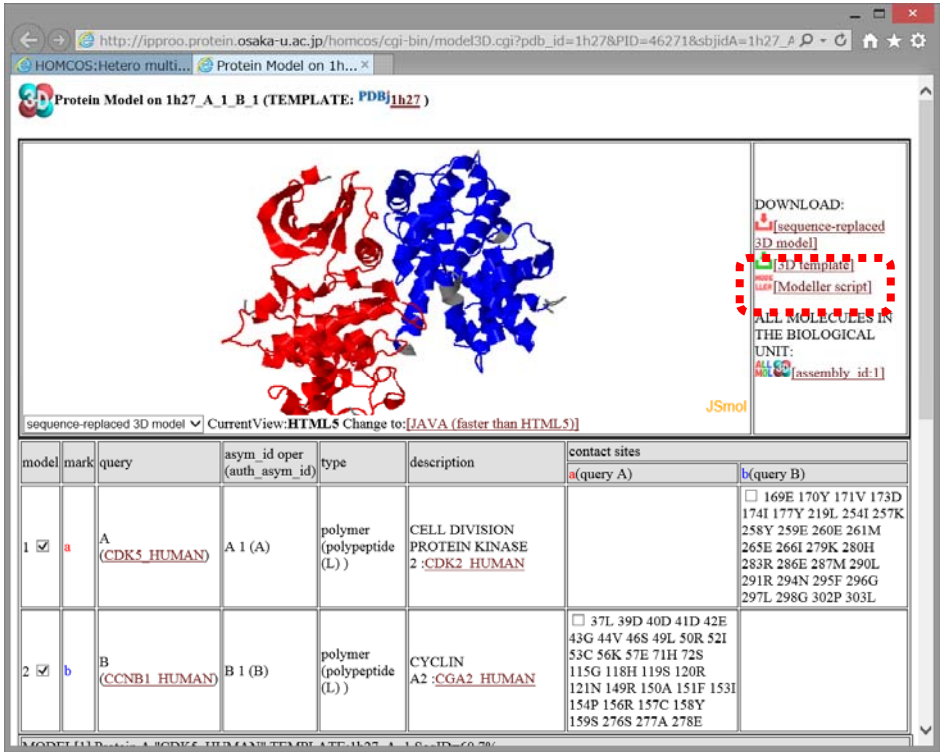
2. 鋳型ベースで複合体予測ができても、本当に結合する保証はない

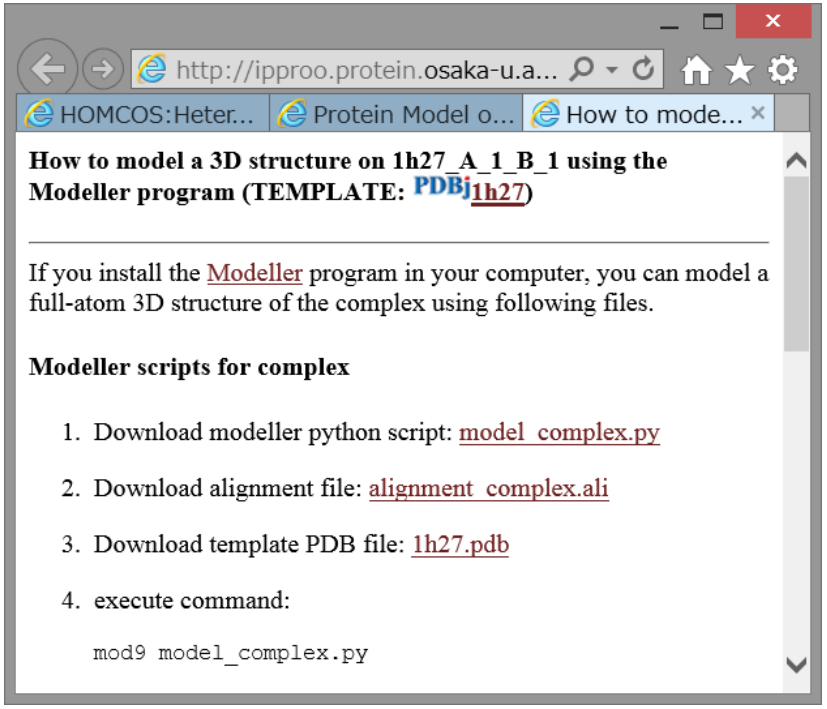
・相同な複合体の鋳型構造があっても、相互作用するとは限らない。
・特にキナーゼ、GPCRなど多遺伝子族では、その特異性の予測は困難
・「もし結合するとしたら、どのように結合するか」の予測と考えるべき。
・相互作用に重要そうなアミノ酸の保存の確認まではできる

付録：Modellerによる二量体のモデリングの手続き

Modellerによる二量体のモデリング [1]

HOMCOSのヘテロ多量体モデリングで、CDK5_HUMANとCCNB1_HUMANを入力し、適当な鋳型構造(1h27)を選んで、以下のモデル3D構造のウィンドウが表示されたとする。





（1）モデル3D構造のウィンドウで MODELLER をクリックする

（2）Modellerのスクリプトのページが表示される。

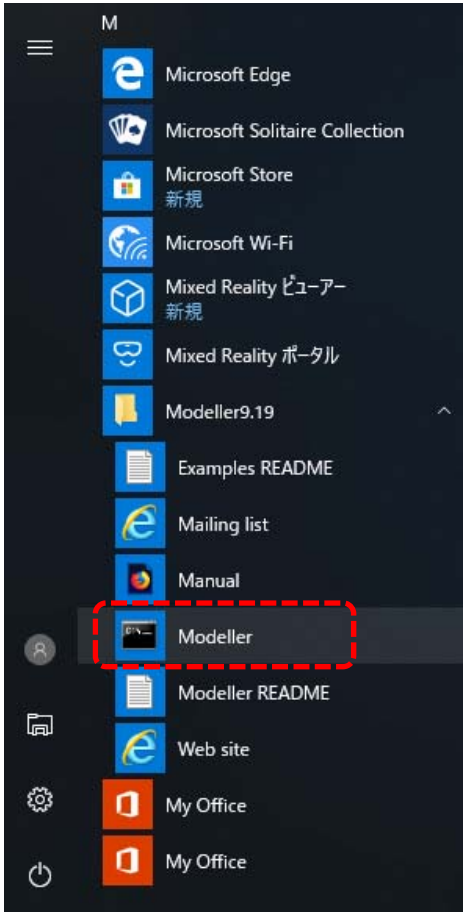
（3）スクリプトファイル (**model_complex.py**)、アラインメントファイル(**alignment_complex.ali**), 鋳型構造のファイル(**1h27.pdb**)の三つを自分のパソコンにダウンロードする。

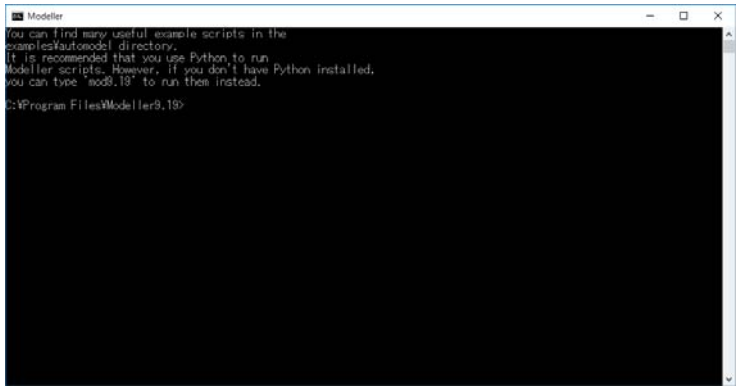
今回は、Windowsでは、**C:¥Users¥visitor¥Downloads** というディレクトリに、
Macでは、 **/Users/guest/Downloads** というディレクトリに保存する。

Modellerによる二量体のモデリング(Win10)[2]

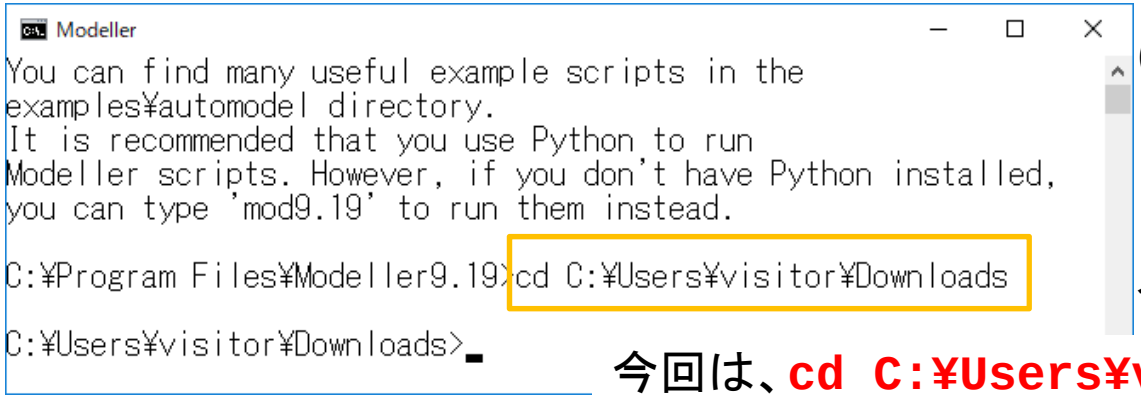
(4)スタート画面の矢印をクリックすると、インストールされたプログラムの一覧が表示される。

(5)アルファベットのMにある、“Modeller 9.19”というフォルダ内にある”Modeller”というコマンドプロンプトのアイコンをクリック





(6)このようなコマンドプロンプトのウィンドウが表示される。



(7) コマンドのウィンドウ内で
cd ［ディレクトリ名］
と入力し、「モデリング用ディレクトリ」（前頁で三つのファイルをコピーしたディレクトリ）に移動する。

今回は、**cd C:¥Users¥visitor¥Downloads**と入力する。

Modellerによる二量体のモデリング(Win10)[3]

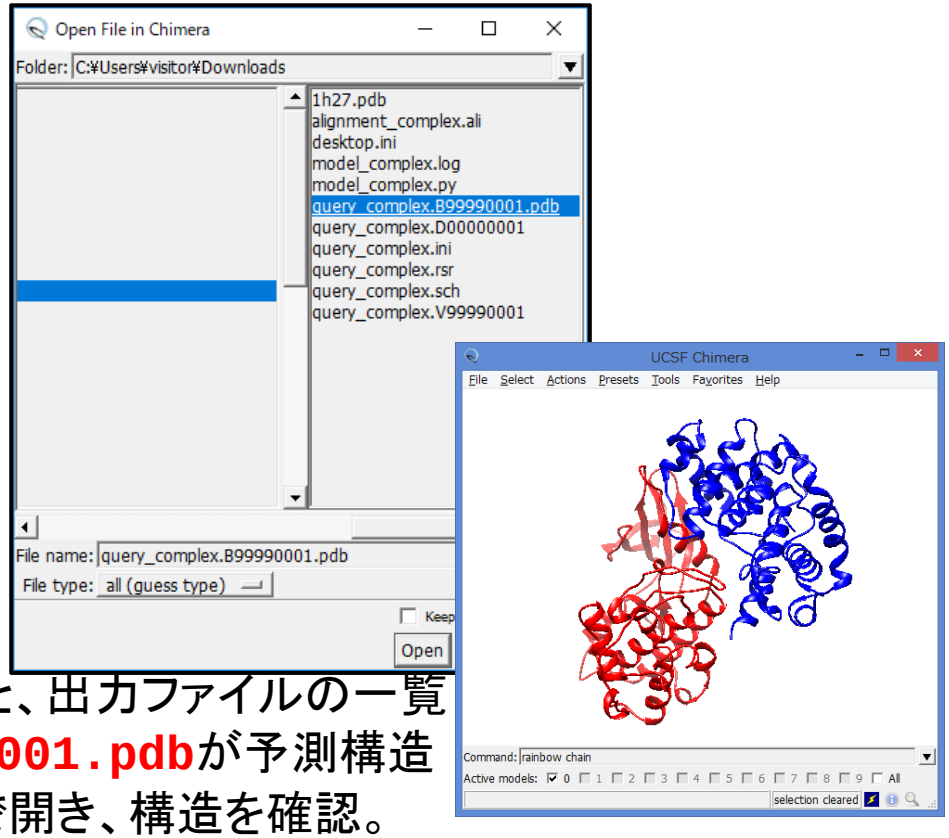
```
C:¥Users¥visitor¥Downloads>dir
 ドライブ C のボリューム ラベルは OS です
 ボリューム シリアル番号は 66B5-94BF です

 C:¥Users¥visitor¥Downloads のディレクトリ

2018/02/19  15:37    <DIR>          .
2018/02/19  15:37    <DIR>          ..
2018/02/19  15:37           363,723 1h27.pdb
2018/02/19  15:37             1,258 alignment_complex.ali
2018/02/19  15:37             1,129 model_complex.py
               3 個のファイル             366,110 バイト
               2 個のディレクトリ  441,980,542,976 バイトの空き領域

C:¥Users¥visitor¥Downloads>
```

(8)コマンド**dir**を入力し、現在のディレクトリ内のファイルの一覧を表示。ダウンロードした三つのファイルがあることを確認。

```
C:¥Users¥visitor¥Downloads>mod9.19 model_complex.py
```

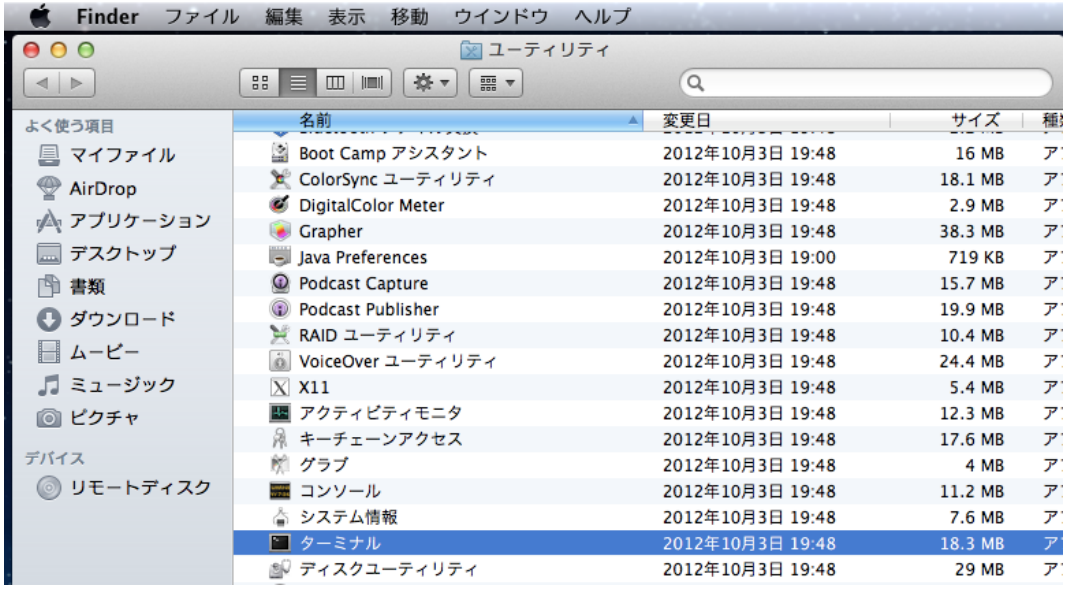
(9) コマンド**mod9.19 [スクリプトファイル]**を入力し、Modellerを実行する。今回は、**mod9.19 model_complex.py**と入力する。この後、計算終了までには1分～数分程度の時間がかかる。

```
C:¥Users¥visitor¥Downloads>dir
 ドライブ C のボリューム ラベルは OS です
 ボリューム シリアル番号は 66B5-94BF です

 C:¥Users¥visitor¥Downloads のディレクトリ

2018/02/19  15:41    <DIR>          .
2018/02/19  15:41    <DIR>          ..
2018/02/19  15:37           363,723 1h27.pdb
2018/02/19  15:37             1,258 alignment_complex.ali
2018/02/19  15:41            48,993 model_complex.log
2018/02/19  15:37             1,129 model_complex.py
2018/02/19  15:41           344,655 query_complex.B99990001.pdb
2018/02/19  15:41            20,930 query_complex.D00000001
2018/02/19  15:40           344,455 query_complex.ini
2018/02/19  15:40         3,805,818 query_complex.rsr
2018/02/19  15:40             3,704 query_complex.sch
2018/02/19  15:41           267,108 query_complex.V99990001
              10 個のファイル           5,201,773 バイト
               2 個のディレクトリ  442,001,444,864 バイトの空き領域
```

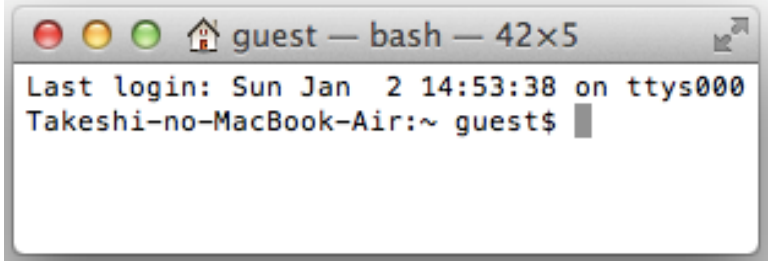


(10)計算終了後、再びコマンド**dir**を入力すると、出力ファイルの一覧を表示。このうち、**query_complex.B99990001.pdb**が予測構造のPDBファイル。このファイルをChimeraなどで開き、構造を確認。

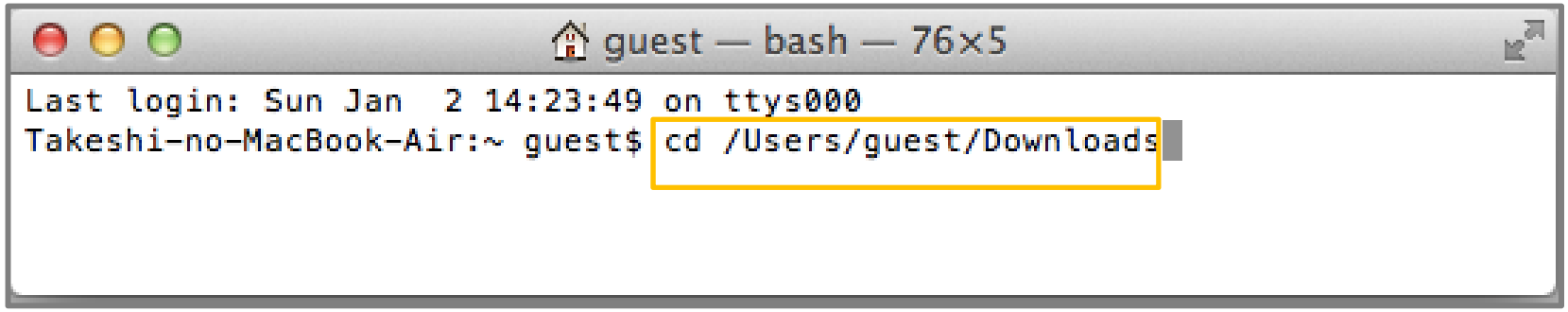
Modellerによる二量体のモデリング(Mac)[2]



(4)Finderから、
[移動]→[ユーティリティ]を選び
[ターミナル]を起動する。

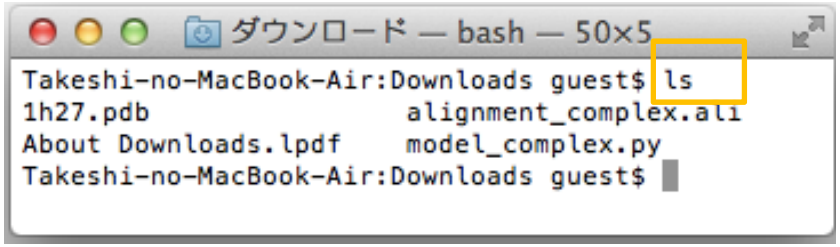


(5)ターミナルのウィンドウ内で
cd [ディレクトリ名] と入力し、
「モデリング用ディレクトリ」(前頁で三つのファイルをコピーしたディレクトリ)に移動する。

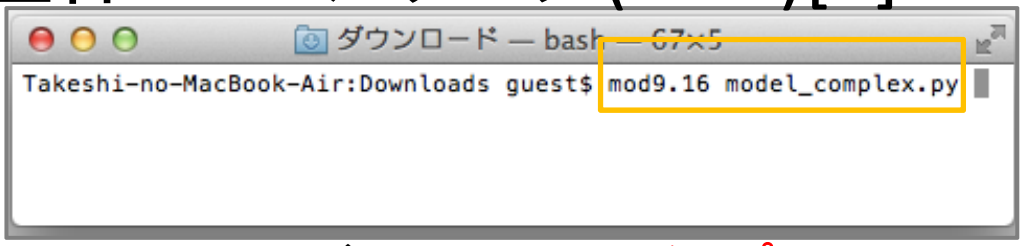


今回は、**cd /Users/guest/Downloads**と入力する。

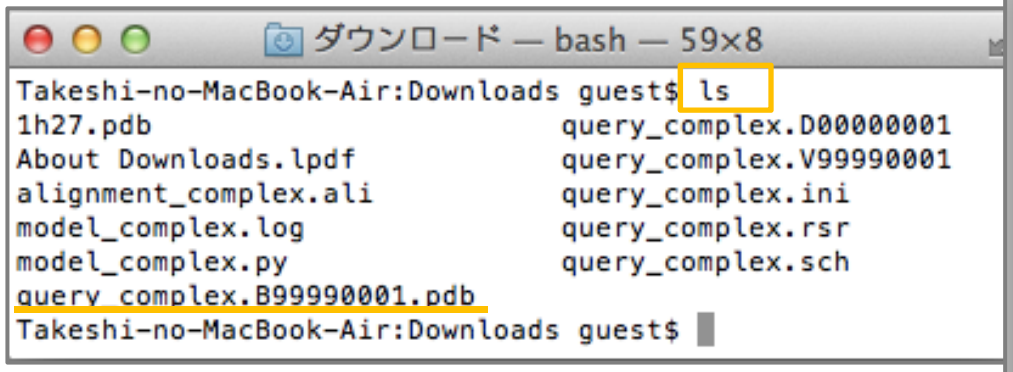
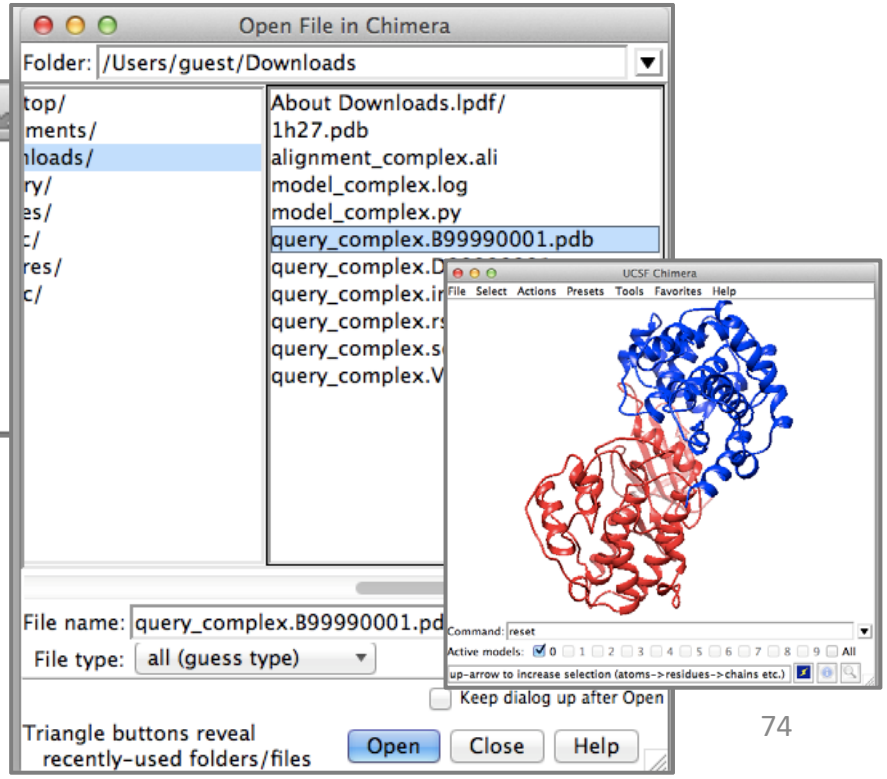
Modellerによる二量体のモデリング(Mac)[3]



(6)コマンド**ls**を入力すると、現在のディレクトリにあるファイルの一覧を表示。ダウンロードした三つのファイルがあることを確認。



(7) コマンド**mod9.16 [スクリプトファイル]**を入力し、Modellerを実行する。今回は、**mod9.16 model_complex.py**と入力する。この後、計算終了までには1分～数分程度の時間がかかる。



(8)計算終了後、再びコマンド**ls**を入力すると、出力ファイルの一覧を表示。このうち、**query_complex.B99990001.pdb**が予測構造のPDBファイル。このファイルをChimeraなどで開き、構造を確認。